SHORT COMMUNICATION

Phylogenetic analysis of forkhead transcription factors in the Panarthropoda

Christoph Schomburg[1,2] · Ralf Janssen[3] · Nikola-Michael Prpic[2]

Abstract

Fox genes encode transcription factors that contain a DNA binding domain, the forkhead domain, and are known from diverse animal species. The exact homology of the Fox genes of different species is debated and this makes inferences about the evolution of the Fox genes, and their duplications and losses difficult. We have performed phylogenetic analyses of the Fox gene complements of 32 panarthropod species. Our results confirm an ancestral complement of FoxA, FoxB, FoxC, FoxD, FoxF, FoxG, FoxJ1, FoxJ2/3, FoxK, FoxL1, FoxL2, FoxN1/4, FoxN2/3, FoxO, FoxP, and FoxQ2 in the Arthropoda, and additionally FoxH and FoxQ1 in the Panarthropoda (including tardigrades and onychophorans). We identify a novel Fox gene sub-family, that we designate as FoxT that includes two genes in *Drosophila melanogaster*, *Circadianly Regulated Gene* (*Crg-1*) and *forkhead domain 3F* (*fd3F*). In a very recent paper, the same new Fox gene sub-family was identified in insects (Lin et al. 2021). Our analysis confirms the presence of FoxT and shows that its members are present throughout Panarthropoda. We show that the hitherto unclassified gene *CG32006* from the fly *Drosophila melanogaster* belongs to FoxJ1. We also detect gene losses: FoxE and FoxM were lost already in the panarthropod ancestor, whereas the loss of FoxH occurred in the arthropod ancestor. Finally, we find an ortholog of FoxQ1 in the bark scorpion *Centruroides sculpturatus*, confirmed not only by phylogenetic analysis, but also by forming an evolutionarily conserved gene cluster with FoxF, FoxC, and FoxL1. This suggests that FoxQ1 belongs to the ancestral Fox gene complement in panarthropods and also in chelicerates, but has been lost at the base of the mandibulate arthropods.

Keywords Fox genes · Phylogeny · Gene duplication · Forkhead domain · Panarthropods

Introduction

Phenotypic diversity of all organisms is achieved through changes in developmental genetic programs. These processes are governed by genetic networks, which usually have transcription factors at the nodes of these networks. For new

✉ Ralf Janssen
  ralf.janssen@geo.uu.se

1 Fachgebiet Botanik, Institut Für Biologie, Universität Kassel, Heinrich-Plett-Straße 40, 34132 Kassel, Germany

2 Institut Für Allgemeine Zoologie Und Entwicklungsbiologie, AG Zoologie Mit Dem Schwerpunkt Molekulare Entwicklungsbiologie, Justus-Liebig-Universität Gießen, Heinrich-Buff-Ring 38, 35392 Gießen, Germany

3 Department of Earth Sciences, Uppsala University, Villavägen 16, 75236 Uppsala, Sweden

genetic networks to arise, existing genes are either co-opted from other networks, or new functions are introduced by the expansion of existing gene families through duplication and subsequent neo-functionalization (Lynch and Conery 2000). One family of genes that have been expanded in particular are the forkhead box genes (Fox genes). They are present with at least one family member found in opisthokont lineages such as Ichthyosporea (Suga et al. 2013), but have more than 40 members in mammals (e.g., Katoh and Katoh 2004).

The first forkhead domain gene to be identified was *fork head* itself in the fly *Drosophila melanogaster* (Weigel et al. 1989), followed shortly by one of its homologs in the rat (Lai et al. 1990). Both genes were found to code for a similar helix-turn-helix motif DNA binding domain, which was termed winged-helix (Li and Tucker 1993). Subsequently, it could be shown that this 110 amino acid forkhead domain is widely conserved among different taxa (Kaufmann and Knöchel 1996) and that genes containing this domain are widespread throughout the animal kingdom. The identified

forkhead genes have been shown to fulfill diverse roles during embryonic development, cell fate decisions, morphogenesis, cell cycle control, metabolism, signal transduction, or the change of chromatin state (e.g., Carlsson and Mahlapuu 2002; Pohl and Knöchel 2005).

With the identification of more and more forkhead genes, it became apparent that many of these independently discovered genes actually form groups of homologs shared between deuterostomes, lophotrochozoans, ecdysozoans, and non-bilaterian metazoans (Larroux et al. 2008; Magie et al. 2005; Mazet et al. 2003). This discovery led to the introduction of a unified nomenclature for forkhead gene sub-families using the letters of the Latin alphabet to denote homologous genes in diverse organisms. Initially, sub-families were named from FoxA to FoxO (Kaestner et al. 2000), but new sub-families were subsequently added up to FoxT, some subgroups have been further subdivided (e.g., FoxQ1 and FoxQ2), and some Fox genes are difficult to place into a sub-family (e.g., *fd3F*). Therefore, the exact number of Fox families is still debated, ranging from 17 to well over 20 (Larroux et al. 2008; Lin et al. 2021; Mazet et al. 2003; Hannenhalli and Kaestner 2009; Shimeld et al. 2010a, b; Tu et al. 2006). Newly identified forkhead domain genes were usually compared to the Fox repertoire of well-studied model organisms, such as the fly *Drosophila melanogaster* or the mouse *Mus musculus*. Thus, many genes which were not found to be homologs of any known Fox family were assigned an "orphan state," for instance in the nematode *Caenorhabditis elegans* (Hope et al. 2003), the hemichordate *Saccoglossus kowalewskii* (Fritzenwanker et al. 2014), and several lophotrochozoans (e.g., Yang et al. 2014). Other approaches focused on the Fox gene repertoire in a single lineage and classified forkhead homologs only in *Drosophila melanogaster* (Lee and Frasch 2004), or among vertebrates (Kaestner et al. 2000). Several studies have tried to unravel the phylogenetic history of forkhead genes and link their emergence to the introduction of evolutionary novelties (e.g., Larroux et al. 2008; Mazet et al. 2003; Shimeld et al. 2010b). These studies are all dependent on the correct initial identification of forkhead genes and therefore these studies potentially overlooked orphan genes, which might belong to a previously unidentified new family or derived members of an existing family, due to the lack of comparable sequences. With the ever-increasing availability of sequenced genomes and embryonic transcriptomes from a wide variety of species, it is now possible to compare the Fox gene repertoire of many different taxa at the same time.

The existence of clustered genes in a genome has been reported for several gene families, the most prominent example being the Hox cluster (e.g., Garcia-Fernàndez 2005; Lemons and McGinnis 2006). Clusters of genes of the same family are thought to arise via tandem duplication from one ancestral gene (e.g., Shimeld et al. 2010b). In the case of Fox genes, a close association has been observed for the two paralogs of *sloppy-paired* in *Drosophila melanogaster* (Cadigan et al. 1994). In addition, the genes FoxQ1, FoxF, FoxC, and FoxL1 form a cluster in the genomes of many different lineages (Shimeld et al. 2010a; Mazet et al. 2006; Wotton et al. 2008; Yu et al. 2008a, b). This suggested that this cluster had evolved by tandem duplication of an ancestral gene and was already present in the bilaterian ancestor (Mazet et al. 2006). Traces of this cluster can be found throughout the Bilateria, but the reason for its maintenance remains unknown. It has been proposed that the co-linearity of genomic arrangement and expression patterns may act as a selective force for the maintenance of such a cluster, as seen for the Hox genes (Monteiro and Ferrier 2006), or that their co-expression in the same tissue might explain their presence at the same chromosomal location, where they form a regulatory block of chromatin (Shimeld et al. 2010a).

Our aim was to identify members of the forkhead family in members of panarthropods (Arthropoda, Tardigrada, and Onychophora), to assign them to their respective subfamilies by phylogenetic sequence analysis and to analyze their chromosomal location (if known) and identify possible clusters of Fox genes in order to gain insights into the evolutionary history of this gene family.

Results and discussion

Fox gene complements in diverse panarthropods

We have searched the genomic or transcriptomic sequence resources of 32 panarthropod species including representatives of the Tardigrada and Onychophora (see list of species in Table 1) for the presence of genes that contain a forkhead domain. The number of Fox genes identified in the genome/transcriptome sequences varied between 10 genes and a maximum of 50 genes. Usually, panarthropods have at least 16 Fox genes (Fig. 2, rightmost column). However, a few species have a smaller complement of Fox genes. Only 10 Fox genes have been found in the locust *Locusta migratoria*, 12 Fox genes have been found in the woodlouse *Armadillidium vulgare*, and 14 Fox genes have been found in the tardigrade *Ramazzottius varieornatus*, the dipluran *Catajapyx aquilonaris*, the springtail *Folsomia candida*, and the honey bee *Apis mellifera*. It is difficult to judge whether these low numbers represent a genuinely reduced Fox gene complement in these species, or are an artifact caused by incomplete genome assembly in these species. At least in *Locusta migratoria*, the unusually low Fox gene number suggests the latter, whereas the tardigrade genome has been shown previously to have undergone strong reduction including significant gene loss (Bemm et al. 2017; Yoshida et al. 2017;

Table 1 List of panarthropod species used in the analysis. A total of 32 species was used, and for each species, the scientific name and a common name are given

Tardigrades
- *Ramazzottius varieornatus* (water bear)

Onychophorans
- *Euperipatoides kanangrensis* (velvet worm)

Chelicerates
- *Limulus polyphemus* (horseshoe crab)
- *Centruroides sculpturatus* (bark scorpion)
- *Parasteatoda tepidariorum* (cobweb spider)
- *Varroa destructor* (bee mite)
- *Tetranychus urticae* (spider mite)

Myriapods
- *Strigamia maritima* (coastal centipede)
- *Glomeris marginata* (pill millipede)

Crustaceans
- *Eurytemora affinis* (copepod)
- *Daphnia magna* (water flea)
- *Penaeus* (or *Litopenaeus*) *vannamei* (king prawn)
- *Hyalella azteca* (amphipod)
- *Armadillidium vulgare* (woodlouse)

Apterygote hexapods
- *Folsomia candida* (white springtail)
- *Catajapyx aquilonaris* (forcepstail)

Basal pterygote insects
- *Ephemera danica* (mayfly)
- *Ladona* (or *Libellula*) *fulva* (dragonfly)

Hemimetabolous insects
- *Medauroidea extradentata* (stick insect)
- *Blatella germanica* (cockroach)
- *Locusta migratoria* (migratory locust)
- *Pediculus humanus* (body louse)
- *Frankliniella occidentalis* (flower thrips)
- *Laodelphax striatellus* (brown planthopper)
- *Halyomorpha halys* (marmorated shield bug)
- *Rhopalosiphum maidis* (corn aphid)

Holometabolous insects
- *Apis mellifera* (honeybee)
- *Tribolium castaneum* (flour beetle)
- *Limnephilus lunatus* (caddisfly)
- *Papilio xuthus* (swallowtail butterfly)
- *Ctenocephalides felis* (cat flea)
- *Drosophila melanogaster* (vinegar fly)

Guijarro-Clarke et al. 2020), thus making the evolutionary loss of Fox genes likely.

By contrast, a few species show a very high number of Fox genes. The scorpion *Centruroides sculpturatus* and the cat flea *Ctenocephalides felis* have more than 30 Fox genes, and the horseshoe crab *Limulus polyphemus* has the maximum number of all studied panarthropod species with 50 Fox genes. In the scorpion and the horseshoe crab, this increase in Fox genes is likely to be linked to genome duplications that have been detected in these animal groups (e.g., Kenny et al. 2016; Schwager et al. 2017; Harper et al. 2020). The increase of Fox genes in the cat flea may instead be caused by extensive tandem duplications of single Fox genes (Driscoll et al. 2020).

Phylogeny of panarthropod Fox genes

We then used the Fox gene sequences from the 32 panarthropod species supplemented with selected additional species from diverse groups of the Opisthokonta (see Supplementary Table S1 for a list of all selected species, and see Supplementary Table S2 for a list of all Fox genes used in the analyses) for phylogenetic reconstruction in order to be able to place the panarthropod Fox sequences into the established Fox gene sub-families. We have reconstructed the phylogeny of all Fox genes from all of the selected taxa based on the sequence of their forkhead domains (Fig. S1), and we have also reconstructed the phylogeny of all Fox genes from the selected panarthropods only, both based on the sequence of the forkhead domain (Fig. S2), and based on the entire

gene sequences (Fig. 1, Fig. S3). The detailed results of Fox gene sub-family assignments of all panarthropod Fox gene sequences are shown in Table S3 and a summary overview is shown in Fig. 2. In most cases, the assignment of a Fox gene to a certain sub-family was consistent across all three phylogenetic analyses; these results are shown with the dark green background color in Fig. 2. In some cases, only two of the three phylogenetic analyses agreed on the sub-family assignment; these cases are indicated by the light green background color in Fig. 2. Finally, in some cases the assignment to sub-family differed in all three analyses; these Fox gene sequences could therefore not be assigned to a sub-family and are included in the column "orphan" in Fig. 2.

Based on the results of the phylogenetic sequence analysis, the ancestral complement of arthropod Fox genes includes at least one member of FoxA, FoxB, FoxC, FoxD, FoxF, FoxG, FoxJ1, FoxJ2/3, FoxK, FoxL1, FoxL2, FoxN1/4, FoxN2/3, FoxO, FoxP, and FoxQ2. In addition, there is one further monophyletic group that only comprises panarthropod Fox gene sequences. In the sequence analysis, together with non-panarthropod opisthokonts, this monophyletic group is nested within the FoxN1/4 clade, but in the panarthropod-only analyses this group forms a well-separated monophylum, related to but clearly distinct from FoxN1/4. This new subgroup of Fox genes was also very recently identified by another research group, and was named FoxT, in accordance with the recommendations for a unified nomenclature of Fox genes (Lin et al. 2021). Our new data now reveal that FoxT is not insect-specific (Lin et al. 2021), but that it has members in other arthropods and

Fig. 1 Unrooted phylogenetic cladogram of all Fox genes from representatives of the Panarthropoda, based on the entire sequence of the conceptually translated proteins. The colors denote the Fox-gene subfamilies. "Clade I" and "Clade II" indicate the principal subdivision of Fox genes into a group with intronless forkhead domain (Clade I) and a group with the forkhead domain interrupted by at least one intron (Clade II). Species and sequence accession numbers are omitted for lack of space, but are included in Supplementary Figure S3. Numbers at the tree edges indicate summarized bootstrap values according to the Majority Rule

even an onychophoran species (summarized in Fig. 2). This means that FoxT was present in the last common ancestor of Panarthropoda, but was apparently lost independently in several lineages of arthropods. In most lineages of insects, however, FoxT has been retained, suggesting an important and conserved function of this Fox gene in insects. The function of FoxT is currently only known for the *Drosophila*

melanogaster homolog of the FoxT sub-family, *fd3F* (see also below), which is required for the development of the mechanosensory cilium of chordotonal neurons (Newton et al. 2012). The development of mechanosensory cilia, however, is not a panarthropod-specific process, and therefore the origin of FoxT may correlate with a previously unidentified common role in the biology of Panarthropoda.

	A	B	C	D	E	F	G	H	J1	J2/3	K	L1	L2	M	N1/4	N2/3	O	P	Q1	Q2	T	Orphan	Total
Ramazzottius varieornatus	1	1	1	-	-	1	1	-	-	-	1	1	1	-	1	-	2	1	1	1	-	-	14
Euperipatoides kanangrensis	1	1	1	-	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1	1	1	1	20
Limulus polyphemus	4	3	6	3	-	2	3	-	5	3	2	2	1	-	2	3	5	4	-	2	-	-	50
Centruroides sculpturatus	2	1	3	2	-	5	2	-	1	1	1	2	3	-	3	1	2	2	1	1	1	-	34
Parasteatoda tepidariorum	2	1	1	1	-	2	1	-	1	1	2	1	1	-	1	1	2	3	-	1	-	-	22
Varroa destructor	1	1	1	1	-	1	1	-	1	1	2	1	1	-	1	-	1	1	-	1	-	-	16
Tetranychus urticae	1	1	1	1	-	1	3	-	1	-	1	-	-	-	3	1	1	-	-	1	-	-	17
Strigamia maritima	1	2	1	1	-	1	1	-	1	1	1	1	1	-	2	1	1	1	-	1	-	-	18
Glomeris marginata	1	1	1	1	-	1	1	-	1	1	1	1	1	-	1	1	1	1	-	1	-	-	16
Eurytemora affinis	1	2	1	1	-	1	1	-	1	-	1	-	2	-	1	2	3	-	3	-	-	1	21
Daphnia magna	1	1	1	1	-	1	1	-	1	-	1	-	1	-	1	2	1	-	1	1	1	1	17
Penaeus vannamei	1	4	1	2	-	1	2	-	1	-	1	1	1	-	1	1	1	-	1	2	-	-	21
Hyalella azteca	1	4	1	2	-	1	2	-	-	-	1	1	1	-	1	1	2	-	1	1	-	-	20
Armadillidium vulgare	1	3	-	-	-	-	-	-	-	-	1	1	1	-	1	-	2	-	-	-	-	-	12
Folsomia candida	1	1	1	1	-	1	1	-	1	-	1	-	-	-	1	-	1	-	1	1	-	1	14
Catajapyx aquilonaris	1	1	1	1	-	1	1	-	-	-	1	-	-	-	1	-	2	1	-	1	1	-	14
Ephemera danica	1	2	1	1	-	1	1	-	1	1	1	1	1	-	1	1	1	1	-	1	1	-	19
Ladona fulva	1	1	2	1	-	1	1	-	1	1	1	-	1	-	1	1	1	1	-	1	1	-	17
Medauroidea extradentata	1	2	1	1	-	1	1	-	1	1	1	1	1	-	1	-	1	1	-	1	1	-	17
Blattella germanica	1	2	1	1	-	1	1	-	1	1	1	1	1	-	2	1	1	1	-	1	1	-	19
Locusta migratoria	-	-	-	-	-	-	1	-	1	1	1	1	1	-	-	1	-	2	-	-	1	-	10
Pediculus humanus	1	2	1	1	-	1	1	-	1	-	1	1	-	-	1	2	-	1	-	1	1	-	17
Frankliniella occidentalis	1	1	1	1	-	1	3	-	1	-	1	1	1	-	1	1	2	1	-	1	1	-	20
Laodelphax striatellus	1	2	1	1	-	2	2	-	1	-	1	1	1	-	1	1	2	1	-	1	1	-	21
Halyomorpha halys	1	2	1	1	-	1	1	-	1	-	1	1	1	-	1	1	1	1	-	1	1	-	18
Rhopalosiphum maidis	1	2	1	1	-	1	1	-	1	-	1	1	1	-	1	1	1	-	-	-	2	-	17
Apis mellifera	1	1	-	1	-	-	-	-	1	-	1	1	1	-	1	1	1	1	-	1	1	-	14
Tribolium castaneum	1	2	1	1	-	1	2	-	1	-	1	1	1	-	1	1	1	1	-	1	1	-	18
Limnephilus lunatus	1	1	1	1	-	1	1	-	1	1	1	-	1	-	1	-	1	2	-	1	1	-	16
Papilio xuthus	1	3	1	1	-	1	2	-	1	5	3	1	1	-	1	1	1	1	-	1	1	1	26
Ctenocephalides felis	2	2	5	1	-	5	3	-	2	-	1	1	-	-	1	2	3	2	1	-	-	-	31
Drosophila melanogaster	1	2	1	1	-	1	3	-	1	-	1	1	-	-	1	1	1	1	-	1	2	-	19
	A	B	C	D	E	F	G	H	J1	J2/3	K	L1	L2	M	N1/4	N2/3	O	P	Q1	Q2	T	Orphan	Total

Fig. 2 Summary of the Fox gene complement of the panarthropod species used in the analysis. The top row gives the name of the Fox gene sub-family; the numbers in the boxes give the count of duplicated genes in a given sub-family per species; a dash (-) indicates that no clear-cut homologs of this particular sub-family could be identified in the genome of the respective species. The panarthropod Fox gene sequences were included in three separate phylogenetic analyses (see text and Fig. S1 to Fig. S3). Dark green boxes indicate that the respective genes were assigned to this sub-family in all three of these analyses. Light green boxes indicate that the respective genes were assigned to this sub-family in two out of the three analyses. All remaining genes that could not be confidently placed in one of the sub-families in the three phylogenetic analyses are given in the column "Orphan" and the total number of Fox genes per species is given in the column "Total"

Interestingly, FoxT appears to be a male-specific gene in the brown plant hopper *Nilaparvata lugens* where it is predominantly (if not exclusively) expressed in male nymphs and adult males (Lin et al. 2021). It is thus likely that FoxT is involved in motile cilia development of sperm.

The Fox genes FoxI, FoxR, and FoxS have evolved in the deuterostome/vertebrate lineage (Tu et al. 2006; Wotton et al. 2008; Yu et al. 2008a; b) and are therefore not present in panarthropods, as expected. However, FoxE,

FoxH, FoxM, and FoxQ1 are more widely distributed in the metazoans (see Fig. S1) and therefore their lack in (most) arthropods/panarthropods indicates specific events of gene loss. FoxE has apparently been lost already in the panarthropod ancestor, because FoxE is also absent in the onychophoran *Euperipatoides kanangrensis* and the tardigrade *Ramazzottius varieornatus*. FoxH is absent from arthropods, but this seems to be an arthropod-specific

event of gene loss, because a clear ortholog of FoxH is present in *Euperipatoides kanangrensis*.

Another Fox gene absent from arthropods is FoxM, and this Fox gene is apparently also absent from tardigrades and onychophorans. However, there is one Fox gene in *Euperipatoides kanangrensis*, namely *c200914_g1_i1*, that we could not confidently place in a sub-family, and this gene is actually included in FoxM in the phylogenetic analysis that also includes non-panarthropods. In the panarthropod-only analyses, *c200914_g1_i1* is included with FoxJ2/3 when only the forkhead domain is used, but it is placed at the base of FoxO when the entire sequence is used. The latter placement also indicates that *c200914_g1_i1* could actually be a representative of FoxM, because genuine FoxM genes are closely related to FoxO and FoxP (see Fig. S1). More information about the Fox gene complement of additional onychophorans is required to establish whether FoxM is present in non-arthropod panarthropods.

Finally, an intriguing case of gene loss is represented by FoxQ1. This gene is absent from all investigated arthropods, except for the bark scorpion *Centruroides sculpturatus*. The identity of the scorpion gene as a genuine member of FoxQ1 is confirmed not only by the phylogenetic sequence analysis, but also by the clustering of this gene with FoxF, FoxC, and FoxL1 (see below). Thus, at least in the chelicerates, the presence of FoxQ1 appears to represent the ancestral state, and FoxQ1 is also present in the basally branching panarthropods. The loss of FoxQ1 therefore appears to be specific for the myriapods, crustaceans, and insects and might represent an apomorphy for the Mandibulata.

Fox genes in the arthropod model species *Drosophila melanogaster*

The Fox gene complement of the main arthropod model system, the fly *Drosophila melanogaster*, has been analyzed previously (Lee and Frasch 2004), and FlyBase lists 19 genes with a Fork head domain (Thurmond et al. 2019). Most of these genes are readily assigned to one of the Fox gene sub-families, but some genes proved to be difficult to classify in previous studies (Mazet et al. 2003; Lee and Frasch, 2004). In our analysis, we were able to assign all 19 Fox genes of *Drosophila melanogaster* to sub-family. The assignment of most of the Fox genes is non-controversial and is consistent with previous analyses: *fork head* (*fkh*) is assigned to FoxA, *fd96Ca* and *fd96Cb* are assigned to FoxB, *crocodile* (*croc*) is assigned to FoxC, *fd59A* is assigned to FoxD, *biniou* (*bin*) is assigned to FoxF, and *sloppy paired 1* (*slp1*) and *sloppy paired 2* (*slp2*) are assigned to FoxG. The gene *fd19B* was previously unassigned to a sub-family, but previous analyses have already demonstrated a close relationship of *fd19B* with the two *slp* genes (e.g.,

Lee and Frasch 2004; Pascual-Carreras et al. 2021), and our analysis confirms this assignment of *fd19B* to FoxG. The assignment of the *Drosophila melanogaster* genes *FoxK*, *FoxL1*, *foxo*, and *FoxP* to the sub-families of the same name was also confirmed in our analysis. In addition, the assignment of *jumeau* (*jumu*) and *Checkpoint suppressor 1-like* (*CHES-1-like*) to FoxN1/4 and FoxN2/3, respectively, was also confirmed. The gene *fd102C* was previously tentatively assigned to FoxQ2 (Lee and Frasch 2004) and we confirm this assignment in our present analysis. The assignment of the gene *CG32006* was previously unclear; it was not assigned to a sub-family by previous studies (e.g., Mazet et al. 2003; Vij et al. 2012), or was assigned to FoxM (Pascual-Carreras et al. 2021). In our analysis, it is firmly placed in the FoxJ1 sub-family, which is notable for its highly conserved role in the formation of motile cilia (Vij et al. 2012; Stubbs et al. 2008; Yu et al. 2008a; b). Our finding of a FoxJ1 homolog in *Drosophila melanogaster* is therefore surprising, because *Drosophila melanogaster* lacks motile cilia in all somatic cells, except for bipolar neurons, and was previously believed to have no FoxJ1 homolog (Vij et al. 2012). However, although our phylogenetic analysis suggests that *CG32006* is a homolog of FoxJ1, the sequence of *CG32006* is also rather diverged from the other members of this Fox sub-family, indicating that *CG32006* might have lost its conserved role in motile cilia formation in *Drosophila melanogaster*. Interestingly, another Fox gene with previously unclear assignment, *fd3F*, appears to take over the role of FoxJ1 in the bipolar chordotonal neurons (Newton et al. 2012). The *fd3F* gene, however, does not group with the FoxJ1 sub-family (as would be expected from its function in motile cilia formation), but is a representative of the newly recognized Fox gene sub-family, FoxT (Lin et al. 2021), that is present throughout the Panarthropoda, and in *Drosophila melanogaster* includes a second previously unassigned Fox gene, *Circadianly Regulated Gene* (*Crg-1*) (Lin et al. 2021).

Clustering of Fox genes in the genome

We have also analyzed the location of the Fox genes in the genomes of those panarthropod species for which sufficient linkage information is available. It has been reported previously that the genes FoxQ1, FoxF, FoxC, and FoxL1 are often clustered together in the genome (Shimeld et al. 2010a; b; Mazet et al. 2006). We find partial conservation of this cluster in many panarthropod species (Fig. 3, Supplementary Fig. S4). Interestingly, we find a complete Q1-F–C-L1 cluster in the scorpion *Centruroides sculpturatus*, and an almost complete cluster (lacking only FoxC) in the tardigrade *Ramazzottius varieornatus*, indicating
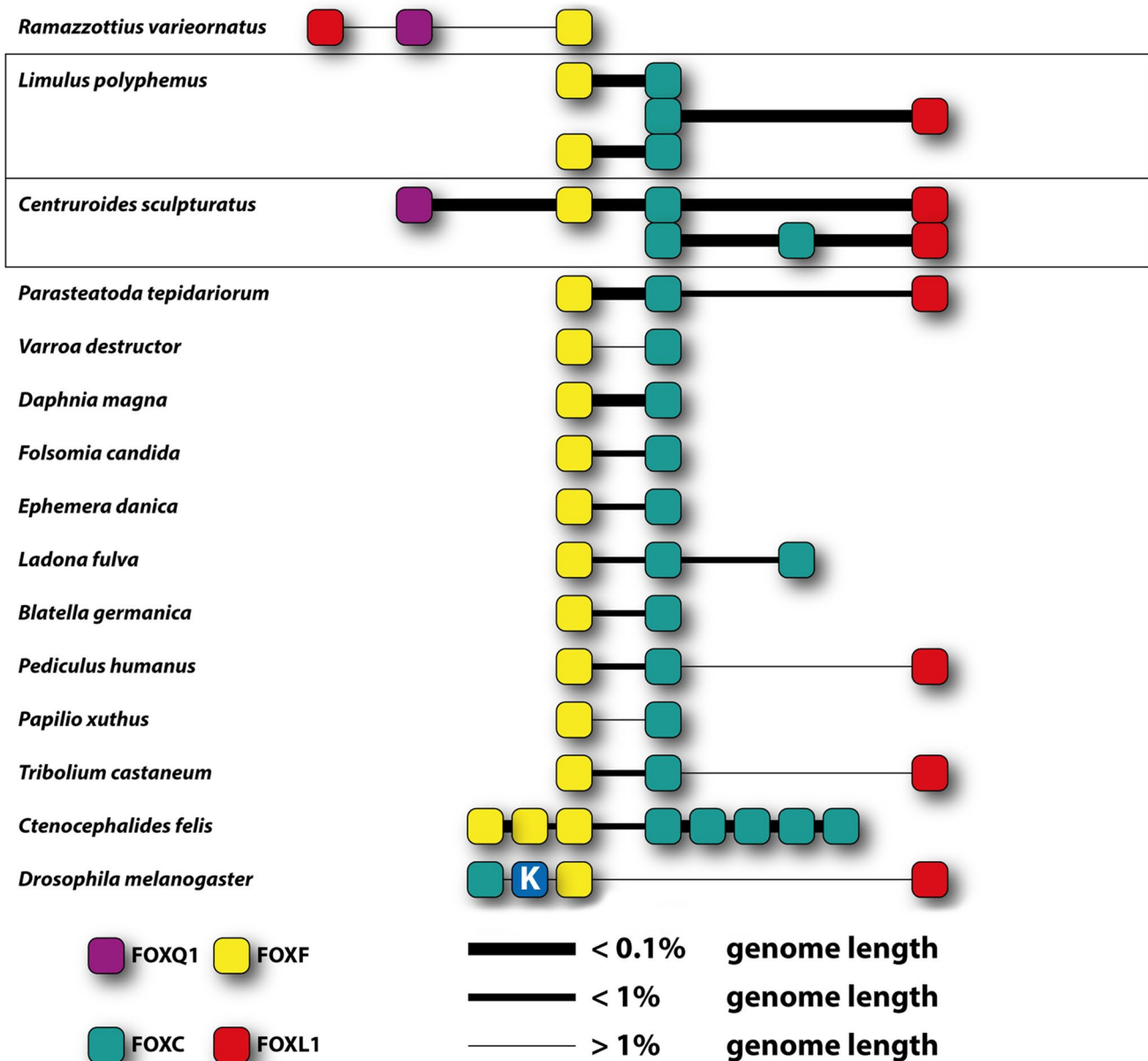
Fig. 3 Summary of the chromosomal clustering of the Fox genes FoxQ1 (purple), FoxF (yellow), FoxC (green), and FoxL1 (red) in the genome of panarthropods. The distance (relative to the genome size) between the genes is indicated by the width of the line connecting the genes, with thick lines indicating close proximity (see legend in the figure). Boxes denote duplicated clusters within one species. Note that the cluster in *Drosophila melanogaster* also contains FoxK. See Fig. S4 for details about accession numbers and chromosomal distances

that this cluster was present in the last common ancestor of all arthropod groups and also in the panarthropod ancestor. In insects and crustaceans, FoxF and FoxC are always clustered and FoxL1 is still within the cluster in some species, e.g., the louse *Pediculus humanus*, the beetle *Tribolium castaneum*, and the fly *Drosophila melanogaster*, but FoxQ1 has been lost entirely not only from this cluster, but also from the genome.

Methods

Selected species and genome/transcriptome sequence resources

For this analysis, we selected high-quality genomes from species at important systematic positions. We used genomes and their translated protein sequences (official gene sets) from publicly available resources (Poelchau et al. 2015; Evans et al. 2013; Agarwala et al. 2018; Howe et al. 2020), except for the two transcriptomes of the onychophoran *Euperipatoides kanangrensis* (Janssen and Budd 2013) and the millipede *Glomeris marginata* (Janssen and Posnien 2014) (see Supplementary Table S1 for details on sources and Supplementary Table S2 for accession numbers).

Identification of forkhead domain–containing genes.

To identify forkhead domain–containing genes for further analysis, we used the hmmscan function of the HMMER3 software (version 2.3.1) (Eddy 2011) package to scan the downloaded gene sets against the PfamA database (release Pfam30.0) (El-Gebali et al. 2019) for conserved forkhead domains. We set a score of 20 as a cut-off for inclusion in the further analysis, since the lowest scoring forkhead gene, which was previously published, is *FoxR1* from the mouse *Mus musculus* with a score of 23.4. We identified isoforms of genes via the corresponding gff annotation files and chose the sequence with the highest forkhead domain score at each genomic locus. In case of multiple sequences on the different scaffolds/chromosomes of a species, we noted their respective positions and distances relative to one another, in order to see clustered forkhead genes. Information on accession numbers, genomic location, distances, and forkhead domain scores of the selected sequences are listed in Supplementary Table S2.

Sequence alignment and phylogenetic analysis

From the identified genes, we extracted the forkhead domains as predicted by hmmscan and used the sequences of the domains and the full-length protein sequences for the alignments, using ClustalOmega (Sievers and Higgins 2014) with standard settings. We inferred phylogenetic trees on the alignments using RAxML (version 8.2.11) (Stamatakis 2014; Stamatakis 2015). Since the forkhead domain genes show an extremely conserved domain, and the mapping of bootstrap trees on a maximum-likelihood tree resulted in very low support values (not shown), we decided to summarize the bootstrap trees (-b, 1,000 trees each) with the Majority Rule Consensus function (Wilkinson 1996; Stamatakis and Izquierdo-Carrasco 2011) of RAxML (-J MR). The

alignment of the forkhead domain of all arthropod, onychophoran, and tardigrade sequences used in the analysis is provided in Supplementary Figure S5.

Declarations

References

Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E et al (2018) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 46:D8-13

Bemm F, Burleigh L, Forster F, Schmucki R, Ebeling M, Janzen CJ, Dandekar T, Schill RO, Certa U, Schultz J (2017) Draft genome

of the eutardigrade Milnesium tardigradum sheds light on ecdyso-zoan evolution. bioRxiv. https://doi.org/10.1101/122309

Cadigan KM, Grossniklaus U, Gehring WJ (1994) Localized expression of sloppy paired protein maintains the polarity of Drosophila parasegments. Genes Dev 8:899–913

Carlsson P, Mahlapuu M (2002) Forkhead transcription factors: key players in development and metabolism. Dev Biol 250:1–23

Driscoll TP, Verhoeve VI, Gillespie JJ, Johnston JS, Guillotte ML, Rennoll-Bankert KE, Rahman MS, Hagen D, Elsik CG, Macaluso KR, Azad AF (2020) A chromosome-level assembly of the cat flea genome uncovers rampant gene duplication and genome size plasticity. BMC Biol 18:70. https://doi.org/10.1186/s12915-020-00802-7

Eddy SR (2011) Accelerated Profile HMM Searches. Pearson WR, editor. PLoS Comput Biol 7:e1002195

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC et al (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432

Evans JD, Brown SJ, Hackett KJJ, Robinson G, Richards S, Lawson D et al (2013) The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. J Hered 104:595–600

Fritzenwanker JH, Gerhart J, Freeman RM, Lowe CJ (2014) The Fox/Forkhead transcription factor family of the hemichordate Saccoglossus kowalevskii. EvoDevo 5:17

Garcia-Fernàndez J (2005) The genesis and evolution of homeobox gene clusters. Nat Rev Genet 6:881–892

Guijarro-Clarke C, Holland PWH, Paps J (2020) Widespread patterns of gene loss in the evolution of the animal kingdom. Nat Ecol Evol 4(4):519–523. https://doi.org/10.1038/s41559-020-1129-2

Hannenhalli S, Kaestner KH (2009) The evolution of Fox genes and their role in development and disease. Nat Rev Genet 10:233–240

Harper A, Baudouin Gonzalez L, Schönauer A, Janssen R, Seiter M, Holzem M, Arif S, McGregor AP, Sumner-Rooney L. Widespread retention of ohnologs in key developmental gene families following whole genome duplication in arachnopulmonates. bioRxiv https://doi.org/10.1101/2020.07.10.177725

Hope IA, Mounsey A, Bauer P, Aslam S (2003) The forkhead gene family of Caenorhabditis elegans. Gene 304:43–55

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J et al (2020) Ensembl Genomes 2020-enabling non-vertebrate genomic research. Nucleic Acids Res 48:D689–D695

Janssen R, Budd GE (2013) Deciphering the onychophoran 'segmentation gene cascade': gene expression reveals limited involvement of pair rule gene orthologs in segmentation, but a highly conserved segment polarity gene network. Dev Biol 382:224–234

Janssen R, Posnien N (2014) Identification and embryonic expression of Wnt2, Wnt4, Wnt5 and Wnt9 in the millipede Glomeris marginata (Myriapoda: Diplopoda). Gene Expr Patterns 14:55–61

Kaestner KH, Knöchel W, Martínez DE (2000) Unified nomenclature for the winged helix / forkhead transcription factors. Genes Dev 14:142–146

Katoh M, Katoh M (2004) Human FOX gene family. Int J Oncol 25:1495–1500. https://doi.org/10.3892/ijo.25.5.1495

Kaufmann E, Knöchel W (1996) Five years on the wings of fork head. Mech Dev 57:3–20

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649

Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland PW, Chu KH, Hui JH (2016) Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. Heredity (edinb) 116(2):190–199. https://doi.org/10.1038/hdy.2015.89

Lai E, Prezioso VR, Smith E, Litvin O, Costa RH, Darnell JE (1990) HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. Genes Dev 4:1427–1436. https://doi.org/10.1101/gad.4.8.1427

Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM (2008) Genesis and expansion of metazoan transcription factor gene classes. Mol Biol Evol 25:980–996

Lee H-H, Frasch M (2004) Survey of forkhead domain encoding genes in the Drosophila genome: classification and embryonic expression patterns. Dev Dyn 229:357–366

Lemons D, McGinnis W (2006) Genomic evolution of Hox gene clusters. Science 313:1918–1922

Li C, Tucker PW (1993) DNA-binding properties and secondary structural model of the hepatocyte nuclear factor 3/fork head domain. Proc Natl Acad Sci U S A 90:11583–11587

Lin HY, Zhu CQ, Zhang HH, Shen ZC, Zhang CX, Ye YX (2021) The genetic network of Forkhead gene family in development of Brown Planthoppers. Biology (Basel) 10(9):867. https://doi.org/10.3390/biology10090867

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Magie CR, Pang K, Martindale MQ (2005) Genomic inventory and expression of Sox and Fox genes in the cnidarian Nematostella vectensis. Dev Genes Evol 215:618–630

Mazet F, Yu JK, Liberles DA, Holland LZ, Shimeld SM (2003) Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. Gene 316:79–89

Mazet F, Amemiya CT, Shimeld SM (2006) An ancient Fox gene cluster in bilaterian animals. Curr Biol 16:R314–R316

Monteiro AS, Ferrier DEK (2006) Hox genes are not always colinear. Int J Biol Sci 2:95–103

Newton FG, zur Lage PI, Karak S, Moore DJ, Göpfert MC, Jarman AP (2012) Forkhead transcription factor Fd3F cooperates with Rfx to regulate a gene expression program for mechanosensory cilia specialization. Dev Cell 22:1221–33

Pascual-Carreras E, Herrera-Úbeda C, Rosselló M, Coronel-Córdoba P, Garcia-Fernàndez J, Saló E, Adell T (2021) Analysis of Fox genes in Schmidtea mediterranea reveals new families and a conserved role of Smed-foxO in controlling cell death. Sci Rep 11(1):2947. https://doi.org/10.1038/s41598-020-80627-0

Poelchau M, Childers C, Moore G, Tsavatapalli V, Evans J, Lee C-Y, Lin H, Lin J-W, Hackett K (2015) The i5k Workspace@NAL - enabling genomic data access, visualization and curation of arthropod genomes. Nucleic Acids Res 43:D714-719. https://doi.org/10.1093/nar/gku983

Pohl BS, Knöchel W (2005) Of Fox and Frogs: Fox (fork head/winged helix) transcription factors in Xenopus development. Gene 344:21–32

Shimeld SM, Boyle MJ, Brunet T, Luke GN, Seaver EC (2010a) Clustered Fox genes in lophotrochozoans and the evolution of the bilaterian Fox gene cluster. Dev Biol 340(2):234–248. https://doi.org/10.1016/j.ydbio.2010.01.015

Shimeld SM, Degnan B, Luke GN (2010b) Evolutionary genomics of the Fox genes: origin of gene families and the ancestry of gene clusters. Genomics 95(5):256–260. https://doi.org/10.1016/j.ygeno.2009.08.002

Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T, Buffry AD, Chao H, Dinh H, Doddapaneni H, Dugan S, Eibner C, Extavour CG, Funch P, Garb J, Gonzalez LB, Gonzalez VL, Griffiths-Jones S, Han Y, Hayashi C, Hilbrant M, Hughes DST, Janssen R, Lee SL, Maeso I, Murali SC, Muzny DM, Nunes da Fonseca R, Paese CLB, Qu J, Ronshaugen M, Schomburg C, Schönauer A, Stollewerk A, Torres-Oliva M, Turetzek N, Vanthournout B, Werren JH, Wolff C, Worley KC, Bucher G, Gibbs RA, Coddington J, Oda H, Stanke M, Ayoub NA, Prpic NM, Flot

JF, Posnien N, Richards S, McGregor AP (2017) The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol 15(1):62. https://doi.org/10.1186/s12915-017-0399-x

Sievers F, Higgins DG (2014) Clustal Omega. Curr Protoc Bioinformatics 48(3.13.1):16

Stamatakis A (2015) Using RAxML to Infer Phylogenies. Curr Protoc Bioinformatics 51(6.14.1.):6.14.14

Stamatakis A, Izquierdo-Carrasco F (2011) Result verification, code verification and computation of support values in phylogenetics. Brief Bioinform 12:270–279

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Stubbs JL, Oishi I, Izpisúa Belmonte JC, Kintner C (2008) The forkhead protein Foxj1 specifies node-like cilia in Xenopus and zebrafish embryos. Nat Genet 40:1454–1460

Suga H, Chen Z, de Mendoza A, Sebé-Pedrós A, Brown MW, Kramer E et al (2013) The Capsaspora genome reveals a complex unicellular prehistory of animals. Nat Commun 4:2325

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ et al (2019) FlyBase 2.0: the next generation. Nucleic Acids Res 47:D759-65

Tu Q, Brown CT, Davidson EH, Oliveri P (2006) Sea urchin Forkhead gene family: phylogeny and embryonic expression. Dev Biol 300:49–62

Vij S, Rink JC, Ho HK, Babu D, Eitel M, Narasimhan V et al (2012) Evolutionarily ancient association of the FoxJ1 transcription factor with the motile ciliogenic program. PLoS Genet 8:e1003019

Weigel D, Jürgens G, Küttner F, Seifert E, Jäckle H (1989) The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. Cell. 57(4):645–658. https://doi.org/10.1016/0092-8674(89)90133-5

Wilkinson M (1996) Majority-rule reduced consensus trees and their use in bootstrapping. Mol Biol Evol 13(3):437–444. https://doi.org/10.1093/oxfordjournals.molbev.a025604

Wotton KR, Mazet F, Shimeld SM (2008) Expression of FoxC, FoxF, FoxL1, and FoxQ1 genes in the dogfish Scyliorhinus canicula defines ancient and derived roles for fox genes in vertebrate development. Dev Dyn 237:1590–1603

Yang M, Xu F, Liu J, Que H, Li L, Zhang G (2014) Phylogeny of forkhead genes in three spiralians and their expression in Pacific oyster Crassostrea gigas. Chinese J Oceanol Limnol 32:1207–1223

Yoshida Y, Koutsovoulos G, Laetsch DR, Stevens L, Kumar S, Horikawa DD, Ishino K, Komine S, Kunieda T, Tomita M, Blaxter M, Arakawa K (2017) Comparative genomics of the tardigrades Hypsibius dujardini and Ramazzottius varieornatus. PLoS Biol. 15(7):e2002266

Yu J-K, Mazet F, Chen Y-T, Huang S-W, Jung K-C, Shimeld SM (2008a) The Fox genes of Branchiostoma floridae. Dev Genes Evol 218:629–638

Yu X, Ng CP, Habacher H, Roy S (2008b) Foxj1 transcription factors are master regulators of the motile ciliogenic program. Nat Genet 40:1445–1453