Article
Discoveries

Evolutionary history of major chemosensory gene families across Panarthropoda

Joel Vizueta[1#], Paula Escuer[1#], Cristina Frías-López[1], Sara Guirao-Rico[2], Lars Hering[3], Georg Mayer[3], Julio Rozas[1], Alejandro Sánchez-Gracia[1]

[1] Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain

[2] Institute of Evolutionary Biology (CSIC-UPF), Passeig de la Barceloneta 37-49, 08003 Barcelona, Spain

[3] Department of Zoology, Institute of Biology, University of Kassel, Heinrich-Plett-Str. 40, D-34132 Kassel, Germany

# These authors contributed equally to this work.

Corresponding authors: Julio Rozas and Alejandro Sánchez-Gracia. E-mail: jrozas@ub.edu, elsanchez@ub.edu. Phone: (+34) 934021495, (+34) 934035304

Data deposition: The raw sequence data generated for this work have been deposited at the Sequence Read Archive (SRA) under Bioproject PRJNA607887. Additional data, including the *de novo* assembly and annotation of the *E. rowelli* transcriptome, and results generated in this study have been deposited in *Figshare* (https://doi.org/10.6084/m9.figshare.12369638.v1).

Abstract

Chemosensory perception is a fundamental biological process of particular relevance in basic and applied arthropod research. However, apart from insects, there is little knowledge of specific molecules involved in this system, which is restricted to a few taxa with uneven phylogenetic sampling across lineages. From an evolutionary perspective, onychophorans (velvet worms) and tardigrades (water bears) are of special interest since they represent the closest living relatives of arthropods, altogether comprising the Panarthropoda. To get insights into the evolutionary origin and diversification of the chemosensory gene repertoire in panarthropods, we sequenced the antenna- and head-specific transcriptomes of the velvet worm *Euperipatoides rowelli* and analyzed members of all major chemosensory families in representative genomes of onychophorans, tardigrades and arthropods. Our results suggest that the NPC2 gene family was the only family encoding soluble proteins in the panarthropod ancestor and that onychophorans might have lost many arthropod-like chemoreceptors, including the highly conserved IR25a receptor of protostomes. On the other hand, the eutardigrade genomes lack genes encoding the DEG-ENaC and CD36-SNMP proteins, the chemosensory members of which have been retained in arthropods; these losses might be related to lineage-specific adaptive strategies of tardigrades to survive extreme environmental conditions. Although the results of this study need to be further substantiated by an increased taxon sampling, our findings shed light on the diversification of chemosensory gene families in Panarthropoda and contribute to a better understanding of the evolution of animal chemical senses.

Keywords: Onychophora, Tardigrada, Chemosensory-related proteins, Antenna-specific transcriptome, Comparative genomics, BITACORA

Introduction

Major animal lineages have evolved independently strikingly similar olfactory pathways (Ache and Young 2005). Indeed, in both mammals and insects, a group of small soluble proteins is responsible for the peripheral detection and solubilization of the odorant compounds (but see Sun et al. 2018 for a broader perspective of these molecules). These proteins are secreted into the aqueous space such as the olfactory mucosa in vertebrates and the sensillar lymph in insects, which is in direct contact with the external environment (Pelosi 1994; Tegoni et al. 2000; Leal 2013). Odorants activate the highly tuned transmembrane receptors located in the dendrites of olfactory neurons, triggering electrical signals that are initially processed in intermediate brain structures (e.g., the olfactory bulb in vertebrates and olfactory glomeruli in insects) and subsequently integrated in higher brain centers (Pelosi 1996; Sánchez-Gracia et al. 2009).

Although soluble proteins and membrane receptors are encoded by large multigene families varying from tens to thousands of copies per genome in both mammals and insects, their evolutionary origin is completely different. While the soluble proteins of insects (mostly represented by odorant-binding, OBP, and chemosensory proteins, CSP; Pelosi et al. 2014) are small, globular alpha-helix-rich proteins, those of

mammals are much larger and with a typical beta barrel domain (belonging to the lipocalin family, mammalian OBP; Tegoni et al. 2000). Non-homology of mammalian and insect olfactory receptors with their characteristic seven-transmembrane domains is also evident from their inverted membrane topologies and different signal transduction mechanisms. In insects, the chemoreceptor superfamily, composed of olfactory (*Or*) and gustatory (*Gr*) receptor families, and the ionotropic receptor (*Ir*) subfamily (a group of highly divergent members of the ionotropic glutamate receptor superfamily, *iGluR*, involved in smell and taste) are ligand-gated ion channels (Joseph and Carlson 2015). Conversely, mammalian olfactory (OR) and taste (T1R and T2R) receptors are G protein-coupled receptors (GPCRs) that activate the second messengers indirectly through gating the corresponding ion channels (Wicher 2012). Furthermore, in mammals, salty and sour stimuli are known to be sensed by amiloride-sensitive ion channels (ENaC; Ben-Shahar 2011), a gene family that has been reported to play a role in insect pheromone perception (Lu et al. 2012), besides being involved in salt and water reception taste and osmotic stress responses (Liu et al. 2003; Chen et al. 2010). Finally, another family related to the human fatty acid transporter CD36 (Vogt et al. 2009), the sensory neuron membrane protein family (SNMP) has been also associated with chemosensory neurons in insects.

All this knowledge, however, is based only on a few invertebrate lineages, with data completely missing from many other important bilaterian clades (Eyun et al. 2017; Vizueta et al. 2018). Among these unexplored taxa, Onychophora (velvet worms) and Tardigrada (water bears) are especially relevant since they represent the closest living relatives of arthropods, with which they have been united in the so-called Panarthropoda (Nielsen 1995; Giribet & Edgecombe 2017). Onychophorans and tardigrades are, thus, key for understanding the evolutionary changes that have taken place in the arthropod lineage. It remains unknown, for instance, whether the chemosensory gene repertories found in arthropods were the result of specific adaptations to the extraordinarily range of environments they inhabit (both aquatic and terrestrial) or whether they were already present in the last common ancestor of Panarthropoda. In other words, to what extent do onychophoran and tardigrade genomes encode members of the arthropod chemosensory families?

Onychophorans most likely originated from an aquatic ancestor over 500 million years ago (Rota-Stabelli et al. 2013), although the ~200 extant species of this group are exclusively terrestrial (Oliveira et al. 2012; Murienne et al. 2014; Oliveira et al. 2016). Velvet worms are elongated, soft-bodied invertebrates that inhabit tropical and temperate forests of the southern hemisphere and around the equator. One remarkable feature of velvet worms is the high phenotypic and anatomic conservation with respect to their Cambrian ancestors (lobopodians), emerging as an important outgroup and excellent model for evolutionary studies of arthropods (e.g., Mayer et al. 2010; Ou et al. 2012; Pauli et al. 2016; Janssen 2017; Martin et al. 2017; Petersen et al. 2019). In onychophorans, the main chemosensory perception structures are located on the antennae (fig. 1A), although the lip papillae surrounding the mouth might also have sensory cells responding to chemical stimuli (Storch & Ruhberg 1977; Storch & Ruhberg 1993). The antennae and the oral lips are innervated by differentiated groups of cell bodies located in different brain regions, suggesting that these structures might have some chemosensory specialization (Martin & Mayer 2014; Martin et al. 2017).

However, only the antennae are associated with the olfactory lobes, which are situated in the protocerebrum (Schürmann 1995; Mayer et al. 2010).

Tardigrades, or water bears, are represented by approximately 1,300 described microscopic species that inhabit marine and semi-terrestrial environments and feed on algae or plant and animal cell fluids (Degma et al. 2020). These animals are renowned for their miniaturized body and ability to survive extreme environmental conditions (Clegg 2002; Horikawa et al. 2013; Smith et al. 2016; Gross et al. 2019). Unlike onychophorans and arthropods, they do not possess modified limbs with a clear chemosensory function, which is likely performed by internal structures covered with a cuticle of variable permeability (e.g., Mayer et al. 2013; Møbjerg et al. 2018). The phylogenetic relationships between arthropods, onychophorans and tardigrades and even the validity of Panarthropoda as a clade are still under debate, although the first two are consistently recovered as sister groups in most molecular phylogenetic analyses (Laumer et al. 2019).

Here, we present a comprehensive comparative genomics analysis across members of the chemosensory gene families of the three major subgroups of Panarthropoda. Our aim is to shed light on the origin and evolution of molecular components of the chemosensory system in these invertebrates and, more specifically, to determine which molecules (or gene families) are responsible for chemoreception in onychophorans and tardigrades and to clarify their evolutionary relationship to those characterized in arthropods. For the analyses, we obtained the specific transcriptomes from the antennae, the head and the rest of the body of the velvet worm *Euperipatoides rowelli.* We integrated these transcriptomic data with information obtained from publicly available genomic data of this onychophoran species (i5K Consortium 2013; Thomas et al. 2020) and two tardigrades, *Hypsibius exemplaris* (formerly referred to as "*H. dujardini*") and *Ramazzottius varieornatus* (Hashimoto et al. 2016; Koutsovoulos et al. 2016; Yoshida et al. 2017), and with transcriptomic and genomic data from arthropods.

Our results uncovered striking differences in the chemosensory repertoires of panarthropods, including the absence of some key families (which do not only encode chemosensory genes) in specific lineages, and allow a more precise delimitation of their origin. These findings highlight the need for extending molecular studies to taxa that have not received much attention in order to better understand the emergence of major genetic innovations and the diversification of animals.

Results

Novel, mostly complete onychophoran reference gene set

The publicly available draft genome of *E. rowelli* is highly fragmented and largely incomplete; only 43.9 % and 47.3 % of genes conserved in Eukaryota (Eu) and Metazoa (Mt) (Based on BUSCO gene collection; ran under the "genome" mode; Seppey et al. 2019), are complete, respectively, whereas 30.4 % of Eu and 23.3 % of Mt genes are missing. On the other hand, the two genome assemblies of tardigrades show good continuity and completeness statistics. These assemblies contain a high proportion of complete genes, ranging from 85 % to 95 %, and only a few fragmented

or missing genes (BUSCO, supplementary table S1, Supplementary Material online). Unlike the genome draft, our deep transcriptome sequencing data from *E. rowelli* (i.e., 60 to 80 million reads per RNAseq experiment) allowed to obtain a mostly complete reference gene set (Table 1). The final consensus transcriptome of this species consists of 1,072,091 non-redundant transcripts. Although this huge number would indicate that the transcriptome is highly fragmented, several lines of evidence suggest the opposite. On the one hand, the sequencing library was prepared using RiboZero instead of the classical poly-A approach, hence our transcriptome contains all RNAs (after ribosomal RNA depletion), including short and large non-coding RNA transcripts. Indeed, only 8.9 % of our consensus transcripts encode putative proteins, the rest being short noncoding sequences (supplementary table S2, Supplementary Material online). On the other hand, we identified all CEG members, most of them being complete (supplementary table S3, Supplementary Material online), and 99.0 % and 99.1 % complete BUSCO Eu and Mt genes (using the "transcriptome" mode), respectively; the remaining 1 % of BUSCO genes are also present but fragmented (supplementary table S1, Supplementary Material online).

Antennal and head-specific transcriptomes of *E. rowelli*

The consensus transcriptome of *E. rowelli* includes a total of 191,116 candidate protein-coding sequences (encoding 245,070 putative peptides), 95,433 of which are functionally annotated. This number, although still quite high, is similar or even lower to those obtained in the other currently available transcriptomes of this and other onychophoran species (Hering et al. 2012; Mapalo et al. 2020). We found 39,128 genes (20.5 %) with detectable expression in the three anatomical compartments. About 8 % of the protein-coding transcripts are expressed exclusively in the antenna (*ANT*) (supplementary fig. S1, Supplementary Material online), which is in agreement with the lower number of cells and molecular functions expected in these appendages. Conversely, almost 11 % of transcripts are expressed exclusively in the head (*HEAD*). Finally, we found 4,615 (~2.4%) transcripts expressed in these two compartments but not in the rest of the body (*REST*). The differential expression analysis (based on RSEM and DESeq2; Li and Dewey 2011; Love et al. 2014) revealed that 9,129 putative protein-coding transcripts are significantly overexpressed in *ANT*, 351 in *HEAD*, 352 in *REST*, and 6,722 in *HEAD+REST*. As expected, we found among the transcripts overexpressed in *ANT* and *HEAD* several GO terms that are enriched in biological functions associated with the response to chemical and external stimuli (supplementary fig. S2, Supplementary Material online).

The chemosensory gene repertoire in Panarthropoda

We identified 440 sequences encoding putative members of the major arthropod chemosensory families in the onychophoran transcriptome (and genome draft) and the two tardigrade genomes (86 in *E. rowelli,* 266 in *H. exemplaris* and 88 in *R. varieornatus*) (supplementary table S4, Supplementary Material online). Although most of the chemosensory genes found in tardigrades (352 out of 354) had annotated structural features in the general feature format files (GFF), many of them lacked a fitting functional annotation. Using BITACORA (Vizueta et al. 2020), we were able to annotate (and in some cases curate) as chemosensory genes 310 GFF features

previously labeled as hypothetical proteins, and to identify new candidate sequences (two novel genes, one in each species).

Chemoreceptors

The *Gr* family is the largest chemosensory gene family in tardigrades. We identified 192 sequences encoding GR-like proteins (the minimum number of protein-coding sequences that can be unequivocally attributed to different gene family copies, S_{MIN}, was 190, 162 of them encoding complete proteins) and 49 (S_{MIN} = 47, 46 complete) in *H. exemplaris* and *R. varieornatus*, respectively (fig. 1B; supplementary table S4, Supplementary Material online). In contrast, we only found three transcripts encoding putative members of this family in *E. rowelli*. One of these copies encoded a complete GR-like member with the protein domain characteristic of this family (7TM chemoreceptor; PF08395). The other two transcripts are short sequences with some similarity to the transmembrane domain of some arthropod GRs. Noticeably, one of them is expressed exclusively in *ANT* (supplementary table S5, Supplementary Material online) whereas the other one might be a pseudogene or an incorrect transcript due to assembly artifacts or sequencing errors. Both this and all the results bellow obtained from our compartmentalized transcriptome data were qualitatively reproduced when the other three transcriptomic sources of *E. rowelli* were used as the subject of our searches (supplementary table S6, Supplementary Material online).

As with arthropod copies of the same family in previously reported gene trees (Eyun et al. 2017; Vizueta et al. 2018), the newly identified tardigrade and onychophoran GR-like sequences form lineage-specific clades in the Panarthropoda tree (fig. 2, supplementary fig. S3, Supplementary Material online). The presence of two phylogenetically unrelated tardigrade-specific clades, and three onychophoran GRs interspersed with other arthropod copies, would suggest that this family underwent an expansion in the ancestor of panarthropods, followed by a second more recent burst in a tardigrade subclade containing *H. exemplaris* with the loss of most of its members in the onychophoran lineage.

The IR/iGluR gene family is the second largest chemosensory family in the three species surveyed, with 47, 26 and 22 IR/iGluR encoding sequences in *H. exemplaris*, *R. varieornatus* and *E. rowelli*, respectively (19, 13 and 12 of them are complete; in this case, we calculated S_{MIN} only for the whole family since the copies estimated from partial fragments could not be unambiguously assigned to one of the two subfamilies; see supplementary table S5, Supplementary Material online, for further details). The phylogenetic tree of the ligand-gated ion channel domains (LCD) of these receptors show a similar picture to that of the GRs, with the predominance of lineage-specific clades. According to the phylogenetic and OrthoFinder results (supplementary table S7, Supplementary Material online), *H. exemplaris* encodes seven Kainate, two AMPA and five NMDAR receptors, whereas the *R. varieornatus* iGluR repertoire is composed of eight Kainate, and five NMDAR receptors, with no AMPA homolog found in this species. In addition, we identified a candidate homolog of the co-receptor IR25a in both tardigrade species. Based on the phylogenetic relationships of the LCD sequences, tardigrades would encode 27 (*H. exemplaris*) and 11 (*R. varieornatus*) divergent IR proteins, thus predicting a chemosensory function of this family in this animal group. In

the case of *E. rowelli*, however, we only found significant evidence for the presence of iGluR members (10 Kainate, one AMPA and nine NMDAR receptors). Specifically, we identified two antennal expressed sequences encoding partial fragments of an IR/GluR protein that are phylogenetically related to some arthropod divergent IRs; nevertheless, the poor node support and the very short length of the aligned region preclude us from drawing firm conclusions about their subfamily identity (fig. 3, supplementary fig. S4 and S5, Supplementary Material online). In fact, the remarkable absence of expression (but also of the signal of a gene in the genome draft) of an IR25a homolog in *E. rowelli* could point to a complete loss of this subfamily of ancient chemoreceptors in Onychophora. Tardigrades would also lack some highly conserved members of this subfamily occurring across arthropods, such as IR8a, IR93a and IR76b, suggesting important changes in the chemosensory role played by this subfamily also in water bears.

Other candidate chemoreceptors and related chemosensory genes

We identified 48 DEG-ENaC sequences in the transcriptome of *E. rowelli*. Although only 10 of them encoded complete receptors, we estimated a S_{MIN} = 25 in this species (Tables S5 and S7), a value which is similar to that found in other arthropods (fig. 4). Surprisingly, the two tardigrades do not encode any DEG-ENaC members, suggesting a complete loss of the family. The phylogenetic tree of the DEG-ENaC family in Panarthropoda is also characterized by the presence of large lineage-specific clades, pointing to a similar mode of evolution as for the other surveyed receptors. Interestingly, many of the members of this family are expressed in *ANT* and/or *HEAD* (19 out of 46 transcripts) of *E. rowelli*, being the family with the greatest number of copies expressed in the chemosensory structures of this species.

Our analysis also uncovered two transcripts encoding CD36-SNMP proteins in *E. rowelli*, a family phylogenetically related to the SNMPs of arthropods but missing in tardigrades. These transcripts are specific or differentially expressed in *ANT* (supplementary fig. S5 and table S5, Supplementary Material online). We also detected the expression of other genes in *ANT* and *HEAD* of *E. rowelli* that have been related directly or indirectly with arthropod chemosensory activity. For instance, we found 20 antenna-specific copies of the GPCR family 3 of receptors (out of 91 characterized in the whole transcriptome) and homologs of the ODR4-like and Pinocchio proteins with differential expression in the *ANT* compartment (supplementary table S5, Supplementary Material online).

Noticeably, the NPC2 is the only family encoding soluble proteins in tardigrades and onychophorans, which fully lack members of the OBP-like and CSP families. We identified 9, 7, and 11 complete *Npc2* genes in the genomes of *H. exemplaris* and *R. varieornatus* and the transcriptome of *E. rowelli*, respectively (supplementary table S4, Supplementary Material online). These family sizes represent a considerable increase in the number of copies with respect to non-panarthropod invertebrates, in which this family typically consists of a single gene (Pelosi et al. 2014). These results suggest an expansion of the NPC2 family in the last common ancestor of Panarthropoda (fig. 5). However, after the expansion, this family shows the lowest turnover rate among the surveyed chemosensory families. It is remarkable that eight NPC2 members are

differentially or specifically expressed in the *ANT* compartment of *E. rowelli* (supplementary table S5, Supplementary Material online), indicating a hypothetical chemosensory role of this family in onychophorans.

Discussion

Evidence suggests that arthropods, onychophorans and tardigrades colonized the land independently after their initial split from an aquatic ancestor (Rota-Stabelli et al. 2013). Similar processes occurred in the three major arthropod groups including pancrustaceans, myriapods, and chelicerates, which originated from an aquatic ancestor 550-450 mya (Lozano-Fernandez et al. 2016). These terrestrialization events might have impacted many aspects of chemosensory perception in these animals. Nonetheless, the extensive comparative genomics analyses in arthropods have revealed a very similar qualitative chemosensory gene composition in all lineages (i.e., we found the same families in most of them, but see Brand et al. 2018 and Vizueta et al. 2018 for two exceptions), suggesting the presence of these proteins in the last common ancestor of Arthropoda. The analysis of representative species across bilaterians revealed that GRs and IRs involved in arthropod chemoreception might have originated through the co-option of ancient gustatory receptor-like (*Gr*-like) and ionotropic glutamate receptor (*iGlur*) genes, respectively (Croset et al. 2010; Krishnan et al. 2014; Robertson 2019). GR-like proteins might have already been present in the last common ancestor of metazoans, as they have been identified in many animal lineages (Robertson 2015; Eyun et al. 2017), and its ancestral function is still under debate (Robertson 2019). Similarly, the origin of the chemosensory IRs, including the co-receptor IR25a and other divergent sequences that evolved independently in different lineages, has been dated back to the protostome ancestor (Croset et al. 2010; Eyun et al. 2017). Although functional evidence of the participation of these proteins in chemoreception comes from studies of insects, various tissue-specific transcriptomes from crustaceans, myriapods and spiders have confirmed the specific or preferential expression of GR and divergent IR genes in the chemosensory structures of crustaceans (Kozma et al. 2020), spiders (Vizueta et al. 2017) and centipedes (Frías-López C, unpublished data).

Intriguingly, we found that *E. rowelli* encodes an exceptionally low number of GR-like members, the lowest reported from panarthropods, and we did not find any trace of the highly conserved co-receptor IR25a. These results are likely not caused by a lack of sensitivity since the sequencing depth of our transcriptomes should be sufficient for detecting lowly expressed genes even in the antenna, as we performed nine independent RNAseq experiments, enriched in compartment-specific transcripts and each yielding between 60 and 80 million reads. Furthermore, we have corroborated all these findings in the full body transcriptomes from six onychophoran species (including representatives of Peripatidae and Peripatopsidae), in addition to *E. rowelli* (supplementary table S6, Supplementary Material online). In fact, the marginal GR-like repertoire size detected in all these species is similar to that observed in *C. elegans* (and other major nematode clades; see supplementary table S6, Supplementary Material online), in which a non-chemosensory role has been established. Similarly, the absence of an IR25a homolog and the doubtful presence of divergent IRs in these transcriptomes would also certainly question the role of this family in velvet worm

chemoreception. Still, these results have to be approached with caution due to high fragmentation of the surveyed transcriptomes and the unavailability of well-assembled, complete genome sequences. Besides, the picture is further complicated by the observation that at least one of the two putatively functional onychophoran GR-like copies is expressed in the antenna of *E. rowelli*, precluding us from drawing a firm conclusion. Thus, it remains uncertain whether these few members of the GR-like lineage have, or ever had, a chemosensory function in onychophorans.

In light of these remarkable absences, some members of the DEG-ENaC or other receptor families, such as the GPCR family 3 or the TRP channels, might have a chemosensory function in *E. rowelli*. In fact, we detected in the antenna of this species the expression of a candidate homolog of the *C. elegans odr-4* gene, which is required for localizing a subset of odorant GPCRs in the cilia of olfactory neurons of this nematode (Dwyer et al. 1998). In *C. elegans*, the olfactory receptors are synthesized in the endoplasmic reticulum of the olfactory neurons, trafficked to the cell surface membrane and transported to the tip of the olfactory cilium, where they bind to odorants. Interestingly, the chemoreceptors of onychophorans, which are situated on the antennal tip and covered with a specialized thin cuticle (fig. 1A), also contain receptor cells with branched cilia (Storch & Ruhberg 1977), suggesting that the onychophoran *odr-4* homolog might be expressed in these cells. TRP channels are highly conserved non-voltage gated, cation channels with a role in insect thermosensation and mechanosensation (Venkatachalam and Montell 2007) that have been attributed to gustation and repellency (Fowler and Montell 2013). We have conducted a prospective search for members of this family in the tardigrades *H. exemplaris* and *R. varieornatus* and the onychophoran *E. rowelli*, detecting a noticeable number of gene copies. From the 67 good-quality annotated onychophoran copies, two are antenna specific, and 11 show differential expression in this compartment (supplementary table S8, Supplementary Material online); at least three of these TRP candidates show remote sequence similarity with members of the TRPA (1) and TRPM (2) subfamilies, which are involved in nociception in insects and taste and cold perception in mammals (Matsuura et al. 2009; Kang et al. 2010: Kwon et al. 2010).

On the other hand, we have found that the two tardigrade genomes could have completely lost the DEG-ENaC family, a group of metazoan-specific membrane proteins that play a role in salt taste, mechanoreception and chemoreception, among other functions (Chen et al. 2010; Ben-Shahar 2011; Lu et al. 2012), and are present in other panarthropods, including in the antenna and the head of the onychophoran *E. rowelli*. It is largely known that tardigrades are organisms extraordinarily resistant to extreme conditions, with unique features among metazoans such as surviving in space, enduring very high pressures and radiation, or surviving extreme temperatures or prolonged desiccation (Møbjerg et al. 2011; Rebecchi et al. 2011; Fernandez et al. 2016; Hashimoto et al. 2016; Hering et al. 2016; Tsujimoto et al. 2016). Recent comparative genomics and transcriptomics studies in these animals have uncovered frequent losses and expansions in stress-related gene pathways, although affecting independent genes in different lineages (Yoshida et al. 2017; Kamilari et al. 2019). Interestingly, in the fruit fly *D. melanogaster*, some DEG-ENaC proteins are involved in maintaining osmotic and intestinal stem cell homeostasis (Kim et al. 2017), regulating

the neuronal response to heat stress (Zheng et al. 2014) or are the target of mutants with lethal desiccation phenotypes of larvae (Johnson and Carder 2012). Moreover, the loss of function of one member of this family extends lifespan and health span, increases internal water stores due to the loss of the ability to sense external water, and exhibits significantly increased survivorship under desiccating conditions (Waterson et al. 2014). In fact, additional searches in six publicly available transcriptomes of different tardigrade species have confirmed the absence of this family in eutardigrades but not in heterotardigrades (supplementary table S6, Supplementary Material online), suggesting that the loss of DEG-ENaC family could be part of a lineage-specific adaptation of eutardigrades to survive in extreme environments. The putative loss of the CD36/SNMP family in eutardigrades is another interesting finding, as members of this conserved family play important sensory, digestive, and immune system roles in *D. melanogaster* (Nichols and Vogt 2008; Vogt et al. 2009).

Our study also revealed that NPC2 members are the only soluble proteins present in tardigrades and onychophorans, some of which are specifically or differentially expressed in the antenna of the *E. rowelli*. This result points to this family as the only panarthropod candidate to perform functions like those documented from arthropod soluble proteins. Overall, our findings, when integrated with previous studies on several arthropod lineages (Eyun et al. 2017; Vizueta et al. 2018), point to at least three evolutionarily independent co-options from ancestral, non-chemosensory soluble protein families, to further participate in chemoreception (fig. 1B), namely in the last common ancestors of (i) Panarthropoda (NPC2; members of this family are involved in the metabolism of cholesterol in *C. elegans*; Sym et al. 2000), (ii) Arthropoda (OBP); and (iii) Mandibulata (CSP). These staggered co-options might have served to progressively adapt peripheral chemoreception to the new chemical world.

Finally, it is worth noting that, if the true phylogeny of Panarthropoda was different from that considered here, the origin and evolutionary history of some of these families would be quite different. If we consider, for example, the phylogenetic hypothesis placing Tardigrada as sister to Nematoda (e.g., Yoshida et al. 2017; Arakawa 2018; Laumer et al. 2019), onychophorans would have lost ionotropic receptors and would have never had chemosensory GR-like proteins. In this case, the insect-type gustatory receptors would have appeared in two (or more) independent GR-like expansions in tardigrades and arthropods. Alternatively, GR-like genes would have been present in the last common ancestor of Panarthropoda but lost in velvet worms and nematodes. Nonetheless, it is worth noting that many of our conclusions are based on the lack of evidence in similarity-based searches, many of them in transcriptomic data and, therefore, must be considered with caution. Further broader taxonomic studies including complete genome assemblies, currently unavailable, and supported by functional evidence, will be needed to confirm the striking absences found in this study and to determine their actual biological meaning.

Taken together, our findings shed light on the diversification of members of the chemosensory gene families across Panarthropoda, including hypothetized origin of some of the surveyed families (fig. 1B). We have found considerable differences in the chemosensory repertoires of panarthropods, including striking absences in specific

lineages, which vindicates the importance of conducting evolutionary genomics studies on the closest arthropod relatives, such as onychophorans and tardigrades. Paradoxically, these clades have not received much attention since the beginning of the genomics era, although they might be crucial for understanding the emergence and diversification of major evolutionary innovations in arthropods.

Materials and Methods

Specimens

Specimens of *Euperipatoides rowelli*, Reid, 1996 (Onychophora, Peripatopsidae) were obtained from decaying logs in the Tallaganda State Forest (New South Wales, Australia; 35°28'S, 149°32'E, 954 m) in October 2011 and January 2013. They were collected under the permit numbers SL100159 and SL101720 issued by the National Parks & Wildlife Service New South Wales and exported under the permit numbers PWSP104061 and PWSP208163 provided by the Department of Sustainability, Environment, Water, Population and Communities. The collected specimens were maintained in the laboratory as described previously (Baer and Mayer, 2012).

Genome data

The genome sequences, annotations and predicted proteins of two tardigrade species, *Hypsibius exemplaris* (v3.5.1, Ensembl Tardigrades Genomes) (Koutsovoulos et al. 2016) and *Ramazzottius varieornatus* (Rv101, Ensembl Tardigrades Genomes) (Hashimoto et al. 2016; Yoshida et al. 2017), and the draft assembly of the onychophoran *E. rowelli*, sequenced as part of the i5K initiative (i5K Consortium 2013; Thomas et al. 2020), were retrieved from http://ensembl.tardigrades.org and https://www.hgsc.bcm.edu/arthropods/velvet-worm-genome-project), respectively. Note that *H. exemplaris* was commonly referred to as "*Hypsibius dujardini*" before its formal description by Gąsiorek et al. (2018).

Transcriptome data

Samples

We used four different sources of transcriptome data of *E. rowelli*. The first was obtained in our tissue-specific transcriptome sequencing experiment of three juvenile individuals (representing three biological replicates). The other three consisted in the raw data of two whole individual RNA-seq experiments (one female [ER9] and one male [ER10]) retrieved from Baylor i5K Initiative Pilot Project [HGSC] (accession numbers: SRX973445 and SRX973444, for the female and male, respectively), and the transcriptome assembly of the *E. rowelli* sample used in Hering et al. (2012).

RNA extraction, library preparation and sequencing

We generated new transcriptomics data from three *E. rowelli* juvenile individuals (supplementary table S9). This species does not show sexual dimorphism with respect to the structure of antennae and chemoreceptors, and juveniles are active hunters

shortly after birth. For each individual, we built three separate RNAseq libraries: the antenna (*ANT*; ensuring that the cut was below the antennal rings with chemoreceptors), the head (*HEAD;* butting behind the slime papillae, and the rest of the body (*REST*), henceforth referred to as anatomical compartments (fig. 1A). All dissections were performed after snap-freezing individuals in liquid nitrogen, which were starved for one week in the laboratory.

The small amount of tissue (and therefore of total RNA) contained in the antennae of a single individual led us to consider a specific extraction protocol specially designed for small amounts of starting material. For *ANT*, we used the PicoPure RNA Isolation Kit (Arcturus, Applied Biosystems, USA) and TRIzol reagent (Invitrogen, Waltham, MA), especially designed to consistently recover high-quality total RNA from fewer cells. In the case of *HEAD* and *REST*, where the amount of tissue was not a limiting factor, we used the RNeasy Mini kit (Qiagen, Venlo, Netherlands) and TRIzol reagent (Invitrogen). In addition, *ANT* RNA was amplified with RiboAmp HS PLUS Kit (Arcturus) to obtain the necessary amount for sequencing (two amplification rounds). We determined the amount and integrity of RNA using a Qubit Fluorometer (Life Technologies, Grand Island, NY) and an Agilent 2100 Bioanalyzer (CCiTUB, Barcelona, Spain), respectively. All library preparation steps and RNA sequencing were carried out in Macrogen Inc., Seoul, South Korea. Briefly, ribosomal RNA was depleted with Ribo-Zero Kit and fragmented into small pieces. Double-stranded cDNA was synthesized with random hexamer (N6) primers (Illumina) and Illumina PE adapters were ligated to the ends of adenylated cDNA fragments. The nine transcriptomes (e.g., three anatomical compartments in three biological replicates) were sequenced independently in the HiSeq 4000 system (100 bp paired end reads) according to the manufacturer's instructions (Illumina, San Diego, CA).

Data pre-processing and transcriptome assembly

We used NGSQCToolkit (Patel and Jain 2012) to filter low quality reads (raw reads with more than 30 % of bases with quality scores <20) from raw data. Filtered reads were further corrected for sequencing errors with the program SEECER v_0.1.3 (Le et al. 2013). We generated a consensus transcriptome from the entire collection of reads of all individuals and anatomical compartments, which was used as the reference for differential expression analyses. We assembled these consensus reference transcripts using Bridger (*k*-mer size = 31; Chang et al. 2015). All contigs with contaminant sequences, i.e. those matching the UniVec vector database, the genomes of *Escherichia coli, Pseudomonas aeruginosa, Staphylococcus aureus, Saccharomyces cerevisiae* and *Homo sapiens,* were removed from this reference using the software Seqclean (https://sourceforge.net/projects/seqclean/). Finally, clean contigs were clustered into putative transcripts defined in the assembly (analogous to the Trinity gene-isoform nomenclature).

Functional annotation of the *E. rowelli* transcriptome

We carried out exhaustive BLAST searches (*E*-value = 10^{-5}; using the assembled reference transcripts as a query) against NCBI-nr, Swiss-Prot and an updated version of the ArthropodDB database (see Vizueta et al. 2020 for further details); for this study

we included in the latter database the predicted proteins and functional annotations of the tardigrade *H. exemplaris* and the nematode *Caenorhabditis elegans* (Consortium* 1998; Yoshida et al. 2017). Coding sequences (CDS) and their conceptual translations were inferred from TransDecoder results (Haas et al. 2013). In addition, we searched the predicted proteins for specific domain signatures with InterProScan (Jones et al. 2014), and signal peptides and transmembrane domains were predicted using SignalP and TMHMM, respectively (Krogh et al. 2001; Petersen et al. 2011).

Gene ontology (GO) terms (Ashburner et al. 2000) were inherited from the results of the BLAST and InterProScan searches (we used the top five positive hits with an *E*-value $<10^{-5}$). We also identified the KEGG enzymes and pathways (Kanehisa and Goto 2000), CEG members (Core Eukaryotic Genes; Parra et al. 2007, 2009), and Metazoa and Eukaryota conserved genes included in BUSCO v3.1.0 (Seppey et al. 2019).

Identification and annotation of chemosensory families

We used the bioinformatics pipeline BITACORA (Vizueta, et al. 2020; 2020b) to identify and annotate the members of the major arthropod chemosensory gene families (CS) in the surveyed transcriptomes and genomes. We first built a database for each of the focal gene families, hereinafter the olfactory (OR), gustatory (GR), ionotropic (iGLuR/IR/*Ir*) and epithelial sodium channel (DEG/ENaC) receptor families, the genes encoding the chemosensory (CSP), odorant-binding (OBP; this family also included the *Obp-like* family recently identified in Vizueta et al. 2017) and Niemann-Pick C2 (NPC2) soluble protein families, and the genes encoding the sensory neuron membrane protein (SNMP) family. Each database included the protein sequences of these families from different arthropod lineages, obtained from the literature (Croset et al. 2010; Colbourne et al. 2011; Vieira and Rozas 2011; Chipman et al. 2014; Robertson 2015; Gulia-Nuss et al. 2016; Vizueta et al. 2018), and was used to construct an HMM profile for each family. In addition, the repertoire of *Drosophila melanogaster*, *Daphnia pulex*, *Strigamia maritima* and *Ixodes scapularis w*as selected as representative of each major arthropod lineage for fig. 1 and to build Panarthropoda gene family trees. We also included in our searches representative members of the chemosensory families from other organisms, including odorant-binding proteins and olfactory and taste receptors of vertebrates (InterPro signatures IPR002448, IPR000725 and IPR007960, respectively; see supplementary table S1B in Frías-López et al. 2015), and *C. elegans* serpentine receptors, known to be involved in nematode chemoreception (Vidal et al. 2018). In each family specific search, we ran two iterative rounds of BITACORA. Specifically, we used the *full mode* on tardigrade genomes (taking advantage of existing GFF annotations), the *genome mode* on *E. rowelli* genomic draft, where no structural annotation is available, and the *protein mode* on the predicted peptides from the *E. rowelli* compartmentalized transcriptome.

All sequences identified in our searches as possible members of one of the focal CS families were classified in different categories based on the structural and functional criteria applied in Vizueta et al. (2018). Briefly, all coding sequences with premature stop codons were classified as non-functional or erroneous copies; this category would include putative pseudogenes, genes with sequencing errors or assembly artifacts. Among the remaining proteins, we distinguished between complete (>80 % of the

559 average length of the family protein domain) and incomplete copies; in addition, and
560 only for the *Gr* and *iGluR/Ir* families, we required complete copies to contain a
561 minimum of five of the seven transmembrane domains (predicted with the software
562 TMHMM version 2.0c; Krogh et al. 2001, and Phobius version 1.01; Käll et al. 2004) or
563 the presence of the ligand-gated ion channel domain (LCD; Pfam identifier PF00060; a
564 domain present in all subfamilies; Croset et al. 2010), respectively. Finally, we
565 estimated the minimum number of different copies of a family, S_{MIN}, as in Vizueta et al.
566 (2018) (this value can be interpreted as an estimate of the actual number of family
567 copies in this species). All proteins identified in this study are provided in the
568 supplementary material.

569 **Expression profiling in *E. rowelli***

570 We mapped the pre-processed reads of each individual and anatomical compartment
571 back to the consensus reference transcriptome using Bowtie2 version 2.2.3 (set as
572 default; Langmead and Salzberg 2012). We used RSEM 1.2.19 software to obtain read
573 counts and TMM-normalized FPKM (Li and Dewey 2011). For the analysis, we
574 considered that a gene is expressed when the FPKM value is higher than 0.01, a
575 reasonable cut-off given the low expression levels reported for other arthropod
576 chemosensory genes (Zhang et al. 2014). The differential gene expression analysis
577 across anatomical compartments was conducted with DESeq2 (Love et al. 2014)
578 considering the three sequenced individuals (per anatomical compartment) as
579 biological replicates and adjusting the *P*-values for the false discovery rate (FDR;
580 Benjamini and Hochberg 1995).

581 **Phylogenetic analyses**

582 We built a multiple sequence alignment (MSA) per each focal CS family using MAFFT
583 ('--auto' option; Katoh and Standley 2013) and used IQ-TREE version 1.6.5 to estimate
584 the best fit substitution models and gene family trees (Nguyen et al. 2015). Node
585 support was estimated from 1,000 ultrafast bootstrap replicates (Hoang et al. 2018).
586 Tree images were drawn using the iTOL web server (Letunic and Bork 2007, 2019).
587 We also assessed the orthologous relationships of some of the surveyed
588 chemosensory gene family members using OrthoFinder v2.2.7 with default options
589 (Emms and Kelly 2015, 2019).

590 **GO enrichment**

591 We used R and GOstats to carry out a GO enrichment analysis (Falcon and Gentleman
592 2007), and REVIGO to generate a graphical representation of the results (Supek et al.
593 2011). We also used Blast2GO suite (Conesa et al. 2005; Götz et al. 2008) to identify
594 KEGG pathways enriched in the list of candidates (Kanehisa and Goto 2000).

595 Acknowledgments

596

597 We are thankful to members of the Mayer laboratory for their support with animal
598 husbandry. We gratefully acknowledge Dave M. Rowell, Ivo de Sena Oliveira, Sandra

599 Treffkorn, Franziska Anni Franke and Michael Gerth for their assistance with collecting
600 the specimens and Noel N. Tait for his help with permits. Ivo de Sena Oliveira and
601 Christine Martin kindly provided images of *E. rowelli* for fig. 1A. The staffs of the
602 National Parks & Wildlife Service New South Wales (Australia) and the Department of
603 Sustainability, Environment, Water, Population and Communities (Australia) are
604 gratefully acknowledged for providing the collection and export permits. AS-G is a
605 Serra Húnter Fellow. This work was supported by the Ministerio de Economía y
606 Competitividad of Spain (CGL2013-45211, CGL2016-75255) and the Comissió
607 Interdepartamental de Recerca I Innovació Tecnològica of Catalonia, Spain
608 (2017SGR1287). JV and PE were supported by an FPI grant (Ministerio de Economía
609 y Competitividad of Spain, BES-2014-068437 and BES-2017-081740, respectively).
610 GM received support from the German Research Foundation (DFG: MA 4147/10-1).

611 References

612 Ache BW, Young JM. 2005. Olfaction: Diverse Species, Conserved Principles. Neuron
613 48:417–430.
614 Arakawa, K. 2018. The complete mitochondrial genome of *Echiniscus testudo*
615 (Heterotardigrada: Echiniscidae). Mitochondrial DNA Part B 3:810–811.
616 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K,
617 Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology.
618 The Gene Ontology Consortium. Nat. Genet. 25:25–29.
619 Baer A, Mayer G. 2012. Comparative anatomy of slime glands in Onychophora (velvet
620 worms). J. Morphol. 273:1079–1088.
621 Ben-Shahar Y. 2011. Sensory Functions for Degenerin/Epithelial Sodium Channels
622 (DEG/ENaC). Adv. Genet. 76:1–26.
623 Benjamini YH, Hochberg Y. 1995. Controlling the False Discovery Rate - A Practical
624 And Powerful Approach To Multiple Testing. J. R. Stat. Soc. 57:289–300.
625 Bhatla N, Horvitz HR. 2015. Light and Hydrogen Peroxide Inhibit *C.elegans* Feeding
626 through Gustatory Receptor Orthologs and Pharyngeal Neurons. Neuron 85:804–
627 818.
628 Brand P, Robertson HM, Lin W, Pothula R, Klingeman WE, Jurat-Fuentes JL, Johnson
629 BR. 2018. The origin of the odorant receptor gene family in insects. Elife 7.
630 Chang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D, Cramer CL, Huang X. 2015. Bridger: a
631 new framework for de novo transcriptome assembly using RNA-seq data.
632 Genome Biol. 16:1–10.
633 Chen Z, Wang Q, Wang Z. 2010. The amiloride-sensitive epithelial Na+ channel
634 PPK28 is essential for *Drosophila* gustatory water reception. J. Neurosci.
635 30:6247–6252.
636 Chipman AD, Ferrier DEK, Brena C, Qu J, Hughes DST, Schröder R, Torres-Oliva M,
637 Znassi N, Jiang H, Almeida FC, et al. 2014. The first myriapod genome sequence
638 reveals conservative arthropod gene content and genome organisation in the
639 centipede *Strigamia maritima*. PLoS Biol. 12:e1002005.
640 Clegg JS. 2001. Cryptobiosis - A peculiar state of biological organization. Comp
641 Biochem Physiol B Biochem Mol Biol.128:613-24
642 Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S,
643 Aerts A, Arnold GJ, Basu MK, et al. 2011. The ecoresponsive genome of *Daphnia*
644 *pulex*. Science 331:555–561.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21:3674-6.
C. elegans Sequencing Consortium.1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282:2012–2018.
Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R. 2010. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet. 6:e1001064.
Degma P, Bertolani R, Guidetti R. 2020. Actual checklist of Tardigrada species. http://www.tardigrada.modena.unimo.it /miscellanea/Actual%20checklist%20of%20Tardigrada.pdf, pp. 48. Accessed 29-02-2020
Dwyer ND, Troemel ER, Sengupta P, Bargmann CI. 1998. Odorant receptor localization to olfactory cilia is mediated by ODR-4, a novel membrane-associated protein. Cell 93:455–466.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20:238.
Eyun S, Soh HY, Posavi M, Munro JB, Hughes DST, Murali SC, Qu J, Dugan S, Lee SL, Chao H, et al. 2017. Evolutionary History of Chemosensory-Related Gene Families across the Arthropoda. Mol. Biol. Evol. 34:1838–1862.
Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. Bioinformatics 23:257–258.
Fernandez, C, Vasanthan, T, Kissoon, N, Karam, G, Duquette, N, Seymour, C, and Stone, JR. 2016. Radiation tolerance and bystander effects in the eutardigrade species *Hypsibius dujardini* (Parachela: Hypsibiidae). Zool. J. Linn. Soc. 178:919–923.
Fowler MA, Montell C. 2013. Drosophila TRP channels and animal behavior. Life Sciences 92:394–403.
Frías-López C, Almeida FC, Guirao-Rico S, Vizueta J, Sánchez-Gracia A, Arnedo MA, Rozas J. 2015. Comparative analysis of tissue-specific transcriptomes in the funnel-web spider *Macrothele calpeiana* (Araneae, Hexathelidae). PeerJ 3:e1064.
Gąsiorek P, Stec D, Morek W , Michalczyk L. 2018. An integrative redescription of *Hypsibius dujardini* (Doyère, 1840), the nominal taxon for *Hypsibioidea* (Tardigrada: Eutardigrada). Zootaxa 4415:45–75.
Giribet G, Edgecombe GD. 2017. Current understanding of Ecdysozoa and its internal phylogenetic relationships. Integr. Comp. Biol. 57:455–466.
Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36:3420–3435.
Gross V, Treffkorn S, Reichelt J, Epple L, Lüter C, Mayer G. 2019. Miniaturization of tardigrades (water bears): Morphological and genomic perspectives. Arthropod. Struct. Dev. 48:12–19.
Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, Sattelle DB, de la Fuente J, Ribeiro JM, Megy K, et al. 2016. Genomic insights

693 into the *Ixodes scapularis* tick vector of Lyme disease. Nat. Commun. 7:10507.
694 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB,
695 Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction
696 from RNA-seq using the Trinity platform for reference generation and analysis.
697 Nat. Protoc. 8:1494–1512.
698 Hashimoto T, Horikawa DD, Saito Y, Kuwahara H, Kozuka-Hata H, Shin-I T, Minakuchi
699 Y, Ohishi K, Motoyama A, Aizu T, et al. 2016. Extremotolerant tardigrade genome
700 and improved radiotolerance of human cultured cells by tardigrade-unique protein.
701 Nat. Commun. 7.
702 Hering L, Henze MJ, Kohler M, Kelber A, Bleidorn C, Leschke M, Nickel B, Meyer M,
703 Kircher M, Sunnucks P, Mayer G. 2012. Opsins in Onychophora (Velvet Worms)
704 suggest a single origin and subsequent diversification of visual pigments in
705 arthropods, Mol. Biol. Evol. 29: 3451–3458,
706 Hering, L, Bouameur, JE, Reichelt, J, Magin, TM, Mayer G. 2016. Novel origin of lamin-
707 derived cytoplasmic intermediate filaments in tardigrades. eLife 5:e11117.
708 Hildebrand JG, Shepherd GM. 1997. Mechanisms of olfactory
709 discrimination:converging evidence for common principles across phyla. Annu.
710 Rev. Neurosci. 20:595–631.
711 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2:
712 Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522.
713 Horikawa DD, Cumbers J, Sakakibara I, Rogoff D, Leuko S, Harnoto R, Arakawa K,
714 Katayama T, Kunieda T, Toyoda A, et al. 2013. Analysis of DNA repair and
715 protection in the tardigrade *Ramazzottius varieornatus* and *Hypsibius dujardini*
716 after exposure to UVC radiation. PLoS One 8(6):e64793.
717 i5K Consortium (2013). The i5K initiative: advancing arthropod genomics for
718 knowledge, human health, agriculture, and the environment. J. Hered. 104:595–
719 600.
720 Janssen R. 2017. Comparative analysis of gene expression patterns in the arthropod
721 labrum and the onychophoran frontal appendages, and its implications for the
722 arthropod head problem. EvoDevo 8:1.
723 Johnson WA, Carder JW. 2012. *Drosophila* nociceptors mediate larval aversion to dry
724 surface environments utilizing both the painless TRP channel and the DEG/ENaC
725 subunit, PPK1. PLoS One 7:32878.
726 Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J,
727 Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function
728 classification. Bioinformatics 30:1236–1240.
729 Joseph RM, Carlson JR. 2015. *Drosophila* Chemoreceptors: A Molecular interface
730 between the chemical world and the brain. Trends Genet. 31:683–695.
731 Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and
732 signal peptide prediction method. J. Mol. Biol. 338:1027–1036.
733 Kamilari M, Jørgensen A, Schiøtt M, Møbjerg N. 2019. Comparative transcriptomics
734 suggest unique molecular adaptations within tardigrade lineages. BMC Genomics
735 20:607.
736 Kang K, Pulver SR, Panzano VC, Chang EC, Griffith LC, Theobald DL, Garrity PA.
737 2010. Analysis of Drosophila TRPA1 reveals an ancient origin for human chemical
738 nociception. Nature 464: 597–600
739 Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic
740 Acids Res. 28:27–30.

741 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
742 improvements in performance and usability. Mol. Biol. Evol. 30:772–780.
743 Kim K, Hung RJ, Perrimon N. 2017. miR-263a regulates ENaC to maintain osmotic and
744 intestinal stem cell homeostasis in *Drosophila*. Dev. Cell 40:23–36.
745 Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H,
746 Thomas F, Aboobaker AA, Blaxter M. 2016. No evidence for extensive horizontal
747 gene transfer in the genome of the tardigrade *Hypsibius dujardini*. Proc. Natl.
748 Acad. Sci. U. S. A. 113:5053–5058.
749 Kozma MT, Ngo-Vu H, Wong YY, Shukla NS, Pawar SD, Senatore A, Schmidt M,
750 Derby CD. 2010. Comparison of transcriptomes from two chemosensory organs in
751 four decapod crustaceans reveals hundreds of candidate chemoreceptor proteins.
752 PloS One 15: e0230266
753 Krishnan A, Almén MS, Fredriksson R, Schiöth HB. 2014. Insights into the origin of
754 nematode chemosensory GPCRs: putative orthologs of the srw family are found
755 across several phyla of protostomes. PLoS One 9:e93048.
756 Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane
757 protein topology with a hidden Markov model: application to complete genomes. J.
758 Mol. Biol. 305:567–580.
759 Kwon Y, Kim SH, Ronderos DS, Lee Y, Akitake B, Woodward OM, Guggino WB, Smith
760 DP, Montell C. 2010. *Drosophila* TRPA1 channel is required to avoid the naturally
761 occurring insect repellent citronellal. Curr Biol. 20:1672–1678.
762 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat.
763 Methods 9:357–359.
764 Laumer CE, Fernández R, Lemer S, Combosch D, Kocot KM, Riesgo A, Andrade SCS,
765 Sterrer W, Sørensen MV, Giribet G. 2019. Revisiting metazoan phylogeny with
766 genomic sampling of all phyla. Proc Biol Sci. 286:20191941
767 Le H-S, Schulz MH, McCauley BM, Hinman VF, Bar-Joseph Z. 2013. Probabilistic
768 error correction for RNA sequencing. Nucleic Acids Res. 41:e109.
769 Leal WS. 2013. Odorant Reception in Insects: Roles of Receptors, Binding Proteins,
770 and Degrading Enzymes. Annu. Rev. Entomol. 58:373–391.
771 Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic
772 tree display and annotation. Bioinformatics 23:127–128.
773 Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
774 developments. Nucleic Acids Res. 47:256-259.
775 Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data
776 with or without a reference genome. BMC Bioinformatics 12:323.
777 Li S, Picimbon JF, Ji S, Kan Y, Chuanling Q, Zhou JJ, Pelosi P. 2008. Multiple
778 functions of an odorant-binding protein in the mosquito *Aedes aegypti*. Biochem.
779 Biophys. Res. Commun. 372:464–468.
780 Liu L, Leonard AS, Motto DG, Feller MA, Price MP, Johnson WA, Welsh MJ. 2003.
781 Contribution of *Drosophila* DEG/ENaC genes to salt taste. Neuron 39:133–146.
782 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
783 dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.
784 Lozano-Fernandez J, Carton R, Tanner AR, Puttick MN, Blaxter M, Vinther J, Olesen J,
785 Giribet G, Edgecombe GD, Pisani D. 2016. A molecular palaeobiological
786 exploration of arthropod terrestrialization. Philos. Trans. R. Soc. B Biol. Sci. 371.
787 Lu B, LaMora A, Sun Y, Welsh MJ, Ben-Shahar Y. 2012. ppk23-dependent
788 chemosensory functions contribute to courtship behavior in *Drosophila*

789 *melanogaster*. PLoS Genet. 8:e1002587.
790 Mapalo MA, Arakawa K, Baker CM, Persson DK, Mirano-Bascos D, Giribet G. 2020.
791 The unique antimicrobial recognition and signaling pathways in tardigrades with a
792 comparison across Ecdysozoa. G3 (Bethesda). 10:1137-1148.
793 Martin C. Mayer G. 2014. Neuronal tracing of oral nerves in a velvet worm –
794 Implications for the evolution of the ecdysozoan brain. Frontiers in Neuroanatomy
795 8:7.
796 Martin C, Gross V, Hering L, Tepper B, Jahn H, Oliveira IS, Stevenson PA. Mayer G.
797 2017. The nervous and visual systems of onychophorans and tardigrades:
798 learning about arthropod evolution from their closest relatives. J. Comp. Physiol.
799 A. Neuroethol. Sens. Neural Behav Physiol. 203:565-590.
800 Matsuura H, Sokabe T, Kohno K, Tominaga M, Kadowaki T. 2009. Evolutionary
801 conservation and changes in insect TRP channels. BMC Evol Biol. 9:228.
802 Mayer G, Kauschke S, Rüdiger J. Stevenson PA. 2013. Neural markers reveal a one-
803 segmented head in tardigrades (water bears). PLoS One 8:e59090.
804 Mayer G, Whitington PM, Sunnucks P. Pflüger HJ. 2010. A revision of brain
805 composition in Onychophora (velvet worms) suggests that the tritocerebrum
806 evolved in arthropods. BMC Evol. Biol. 10:255.
807 Møbjerg N, Halberg KA, Jørgensen A, Persson D, Bjørn M, Ramløv H, Kristensen RM.
808 2011. Survival in extreme environments — on the current knowledge of
809 adaptations in tardigrades. Acta Physiol. 202:409–420.
810 Møbjerg, N., Jørgensen, A., Kristensen, R.M. & Neves, R.C. (2018) Morphology and
811 Functional Anatomy. In: R.O. Schill (ed.) Water Bears: The Biology of
812 Tardigrades. Springer Nature Switzerland AG, Cham, Switzerland, pp. 57–94.
813 Murienne J, Daniels SR, Buckley TR, Mayer G, Giribet G. 2013. A living fossil tale of
814 Pangean biogeography. Proc. Biol. Sci. 281:20132648.
815 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and
816 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol.
817 Biol. Evol. 32:268-274
818 Nichols Z, Vogt RG. 2008. The SNMP/CD36 gene family in Diptera, Hymenoptera and
819 Coleoptera: *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles gambiae*,
820 *Aedes aegypti*, *Apis mellifera*, and *Tribolium castaneum*. Insect Biochem. Mol.
821 Biol. 38:398–415.
822 Nielsen C. 1995. Animal Evolution: Interrelationships of the Living Phyla. Oxford
823 University Press, Oxford
824 Oliveira IS, Read VM. Mayer G. 2012. A world checklist of Onychophora (velvet
825 worms), with notes on nomenclature and status of names. ZooKeys 211:1–70.
826 Oliveira IS, Bai M, Jahn H, Gross V, Martin C, Hammel JU, Zhang W. Mayer G. 2016.
827 Earliest onychophoran in amber reveals Gondwanan migration patterns. Curr.
828 Biol. 26:2594–2601.
829 Ou Q, Shu D, Mayer G. 2012. Cambrian lobopodians and extant onychophorans
830 provide new insights into early cephalization in Panarthropoda. Nat. Commun.
831 3:1261.
832 Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core
833 genes in eukaryotic genomes. Bioinformatics 23:1061–1067.
834 Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009. Assessing the gene space in draft
835 genomes. Nucleic Acids Res. 37:289–297.
836 Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation

837 sequencing data. PLoS One 7:e30619.
838 Pauli T, Vedder L, Dowling D, Petersen M, Meusemann, K, Donath A, Peters RS,
839 Podsiadlowski L, Mayer C, Liu S, Zhou X, Heger P, Wiehe T, Hering L, Mayer G,
840 Misof B, Niehuis O. 2016. Transcriptomic data from panarthropods shed new light
841 on the evolution of insulator binding proteins in insects. BMC Genomics 17:861.
842 Pelosi P. 1994. Odorant-binding proteins. Crit. Rev. Biochem. Mol. Biol. 29:199–228.
843 Pelosi P. 1996. Perireceptor events in olfaction. J. Neurobiol. 30:3–19.
844 Pelosi P, Iovinella I, Felicioli A, Dani FR. 2014. Soluble proteins of chemical
845 communication: an overview across arthropods. Front. Physiol. 5:320.
846 Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating
847 signal peptides from transmembrane regions. Nat. Methods 8:785–786.
848 Petersen M, Armisen D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Misof B.
849 2019. Diversity and evolution of the transposable element repertoire in arthropods
850 with particular reference to insects. BMC Evol. Biol. 19:11.
851 Rebecchi L, Altiero T, Cesari M, Bertolani R, Rizzo AM, Corsetto PA, Guidetti R. 2011.
852 Resistance of the anhydrobiotic eutardigrade *Paramacrobiotus richtersi* to space
853 flight (LIFE–TARSE mission on FOTON-M3). J. Zoolog. Syst. Evol. Res. 49:98–
854 103
855 Robertson HM. 2015. The insect chemoreceptor superfamily is ancient in animals.
856 Chem. Senses 40:609–614.
857 Robertson HM. 2019. Molecular evolution of the major arthropod chemoreceptor gene
858 families. Annu. Rev. Entomol. 64:227–242.
859 Rota-Stabelli O, Daley AC, Pisani D. 2013. Molecular timetrees reveal a Cambrian
860 colonization of land and a new scenario for ecdysozoan evolution. Curr. Biol.
861 23:392–398.
862 Saina M, Busengdal H, Sinigaglia C, Petrone L, Oliveri P, Rentzsch F, Benton R. 2015.
863 A cnidarian homologue of an insect gustatory receptor functions in developmental
864 body patterning. Nat. Commun. 6:6243.
865 Sánchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major
866 chemosensory gene families in insects. Heredity 103:208–216.
867 Schürmann FW. 1995. Common and special features of the nervous system of
868 Onychophora: A comparison with Arthropoda, Annelida and some other
869 invertebrates. In: Breidbach O., Kutsch W. (eds) The Nervous Systems of
870 Invertebrates: An Evolutionary and Comparative Approach. Experientia
871 Supplementum, vol 72. Birkhäuser Basel. p. 139-158.
872 Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing genome assembly and
873 annotation completeness. Methods Mol. Biol. 1962: :227-245.
874 Smith FW, Boothby TC, Giovannini I, Rebecchi L, Jockusch EL, Goldstein B. 2016. The
875 compact body plan of tardigrades evolved by the loss of a large body region. Curr.
876 Biol. 26:224-229.
877 Storch V, Ruhberg H. 1977. Fine structure of the sensilla of Peripatopsis moseleyi
878 (Onychophora). Cell Tissue Res. 177:539-553.
879 Storch V, Ruhberg H. 1993. Onychophora. In: Harrison FW, Rice ME, editors.
880 Microscopic Anatomy of Invertebrates. New York (NY): Wiley-Liss. p. 11–56.
881 Sun JS, Xiao S, Carlson JR. 2018. The diverse small proteins called odorant-binding
882 proteins. Open Biol. 8:180208.
883 Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes
884 long lists of gene ontology terms. PLoS One 6:e21800.

Sym M, Basson M, Johnson C. 2000. A model for niemann-pick type C disease in the nematode *Caenorhabditis elegans*. Curr Biol. 10:527–530.
Tegoni M, Pelosi P, Vincent F, Spinelli S, Campanacci V, Grolli S, Ramoni R, Cambillau C. 2000. Mammalian odorant binding proteins. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1482:229–240.
Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, Anstead CA, Ayoub NA, Batterham P, Bellair M, et al. 2020. Gene content evolution in the arthropods. Genome Biol. 21:15.
Tsujimoto M, Imura S, Kanda H. 2016. Rocovery and reproduction of an Antarctic tardigrade retrieved from a moss sample frozen for over 30 years. Cryobiology 72:78–81.
Venkatachalam K, Montell C. 2007. TRP Channels TRP: transient receptor potential. Annu Rev. Biochem. 76:387–417.
Vidal B, Aghayeva U, Sun H, Wang C, Glenwinkel L, Bayer EA, Hobert O. 2018. An atlas of *Caenorhabditis elegans* chemoreceptor expression. PLoS Biol. 16:e2004218.
Vieira FG, Rozas J. 2011. Comparative genomics of the odorant-binding and chemosensory protein gene families across the arthropoda: Origin and evolutionary history of the chemosensory system. Genome Biol. Evol. 3:476–490.
Vizueta J, Frías-López C, Macías-Hernández N, Arnedo MA, Sánchez-Gracia A, Rozas J. 2017. Evolution of chemosensory gene families in arthropods: Insight from the first inclusive comparative transcriptome analysis across spider appendages. Genome Biol. Evol. 9:178–196.
Vizueta J, Macías-Hernández N, Arnedo MA, Rozas J, Sánchez-Gracia A. 2019. Chance and predictability in evolution: The genomic basis of convergent dietary specializations in an adaptive radiation. Mol. Ecol. 28:4028–4045.
Vizueta J, Rozas J, Sánchez-Gracia A. 2018. Comparative genomics reveals thousands of novel chemosensory genes and massive changes in chemoreceptor repertories across chelicerates. Genome Biol. Evol. 10:1221–1236.
Vizueta J, Sánchez-Gracia A, Rozas J. 2020. BITACORA: A comprehensive tool for the identification and annotation of gene families in genome assemblies. Mol Ecol Resour. 10.1111/1755-0998.13202.
Vizueta J, Escuer P, Sánchez-Gracia A, Rozas J. 2020. Genome mining and sequence analysis of chemosensory soluble proteins in arthropods. Methods Enzymol. In press.
Vogt RG, Miller NE, Litvack R, Fandino RA, Sparks J, Staples J, Friedman R, Dickens JC. 2009. The insect SNMP gene family. Insect Biochem. Mol. Biol. 39:448–456.
Waterson MJ, Chung BY, Harvanek ZM, Ostojic I, Alcedo J, Pletcher SD. 2014. Water sensor ppk28 modulates Drosophila lifespan and physiology through AKH signaling. Proc. Natl. Acad. Sci. U. S. A. 111:8137–8142.
Wicher D. 2012. Functional and evolutionary aspects of chemoreceptors. Front. Cell. Neurosci. 6:48.
Xiu C, Xiao Y, Zhang S, Bao H, Liu Z, Zhang Y. 2019. Niemann-Pick proteins type C2 are identified as olfactory related genes of *Pardosa pseudoannulata* by transcriptome and expression profile analysis. Comp. Biochem. Physiol. Part D Genomics Proteomics 29:320–329.
Yoshida Y, Koutsovoulos G, Laetsch DR, Stevens L, Kumar S, Horikawa DD, Ishino K, Komine S, Kunieda T, Tomita M, et al. 2017. Comparative genomics of the

933 tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. PLOS Biol.
934 15:e2002266.
935 Zhang Y, Zheng Y, Li D, Fan Y. 2014. Transcriptomics and identification of the
936 chemoreceptor superfamily of the pupal parasitoid of the oriental fruit fly,
937 *Spalangia endius* Walker (Hymenoptera: Pteromalidae). PLoS One 9:e87800.
938 Zheng X, Valakh V, DiAntonio A, Ben-Shahar Y. 2014. Natural antisense transcripts
939 regulate the neuronal stress response and excitability. Elife 2014.

Tables

Table 1. Summary of the transcriptomic data newly generated for this study and the functional annotation statistics

	ANT	*HEAD*	*REST*	Total	Covered by reads[b]	Protein-coding
Assembled contigs	313,898	640,096	538,450	1,212,132	865,014	245,070
Unique sequences (transcripts)	246,146	541,258	448,126	1,072,091	742,596	191,116
Average length of transcripts (nt)	611	495	527	427	473	490
Longest transcript (nt)	56,010	56,010	56,010	56,010	56,010	55,107
CEG sequences	450	438	440	458	453	458
Sequences with GO annotation	22,514	43,734	41,035	69,901	55,577	69,901
Sequences with functional annotation[a]	29,156	59,247	54,911	95,433	75,060	95,433

[a] based on Interpro and BLAST searches (include annotations without GO)
[b] transcripts with mapped reads CEG, cluster of essential genes

Figure legends

Figure 1. Chemosensory structures in the onychophoran *E. rowelli* and summary of major findings. (A) Anterior end of a specimen with anatomical compartments indicated by dotted lines. Insets illustrate scanning electron micrographs of mouth surrounded by lip papillae (asterisks in left micrograph; scale bar: 300 µm) and putative chemosensory organs situated on antennae (asterisks in right micrograph; scale bar: 20 µm). Images provided by Ivo de Sena Oliveira and Christine Martin. Note that chemosensory related genes are expressed in the anatomical compartments with expected chemosensory function. Numbers refer to those genes specifically or differentially expressed in antenna (*ANT*) and head (*HEAD*). (B) Minimum estimates of gene family sizes (S_{MIN}) in the genomes from nine major ecdysozoan lineages (numbers for iGluR and IR subfamilies correspond to complete copies; see Results). Solid and empty colored boxes in the phylogeny indicate gains and losses of particular gene families, respectively. Purple and light-brown shadings denote membrane receptors and soluble proteins, respectively. [†]Three very short sequences encoding parts of the iGluR/IR ligand-gated ion channel domain (PF00060) that, although they are phylogenetically related to IRs, could not be unambiguously assigned to this subfamily (supplementary fig. S4). [#]One complete GR receptor and two sequences resulting from partial BLAST hits. *Values obtained after new BITACORA searches in these genomes.

Figure 2. Maximum likelihood phylogenetic tree of GR family in panarthropods. We excluded all partial proteins and putative pseudogenes and artifacts from the analysis. The color code for species is the same as in fig. 1B. Nodes with bootstrap support values >90 % are shown as solid circles. Scale bar represents one amino acid

substitution per site. See supplementary fig. S3 for a gene tree including all identified sequences.

Figure 3. Maximum likelihood phylogenetic tree of iGluR/IR ligand-gated ion channel domains (PF00060) in panarthropods. Only complete domains were used for this analysis. The color code for species is the same as in fig. 1B. Nodes with bootstrap support values >90 % are shown as solid circles. Scale bar represents one amino acid substitution per site. See supplementary fig. S4 for a gene tree including all identified sequences

Figure 4. Maximum likelihood phylogenetic tree of DEG-ENaC family in panarthropods. The color code for species is the same as in fig. 1B. Boxes in the outer circle indicate the genes specifically (first layer) or differentially (outer layer) expressed in *ANT* (purple), *HEAD* (green), or both (orange). Nodes with bootstrap support values >90 % are shown as solid circles. Scale bar represents one amino acid substitution per site.

Figure 5. Maximum likelihood phylogenetic tree of NPC2 family in panarthropods. Boxes in the outer circle indicate the genes specifically (first layer) or differentially (outer layer) expressed in *ANT* (purple), *HEAD* (green), or both (orange). Nodes with bootstrap support values >90 % are shown as solid circles. Scale bar represents one amino acid substitution per site.

Supplementary Data

Supplementary tables

Supplementary table S1. BUSCO analysis of the *H. exemplaris, R. varieornatus* and *E. rowelli* genomes and *E. rowelli* transcriptome.

Supplementary table S2. Functional annotation and transcript lengths in *E. rowelli.*

Supplementary table S3. Distribution of coverage lengths in CEG BLASTX results.

Supplementary table S4. Summary of sequences identified in the *H. exemplaris, R. varieornatus* genomes and *E. rowelli* and transcriptome.

Supplementary table S5. Gene expression profiles of candidate chemosensory genes in different anatomical compartments of *E. rowelli.*

Supplementary table S6. Summary of chemosensory gene family members identified in whole body transcriptomes of onychophoran and tardigrades and in the genome of representative nematodes.

Supplementary table S7. OrthoFinder analysis of IR/iGluR sequences.

Supplementary table S8. Gene expression profiles of TRP genes in different anatomical compartments of *E. rowelli.*

Supplementary table S9. Summary of the RNA-Seq samples and data.

Supplementary figure legends

Supplementary figure S1. Venn diagram showing the number of protein-coding transcripts across *E. rowelli* anatomical compartments. Differentially expressed genes are shown in brackets.

Supplementary figure S2. Tree maps with the results of the GO enrichment analysis of differentially expressed genes generated with REVIGO. A) and B) Molecular function and Biological process in the *ANT* compartment. C) and D) Molecular function and Biological process in the *HEAD* compartment.

Supplementary figure S3. Maximum likelihood phylogenetic tree of the Panarthropoda GR family. All identified sequences are included. The color code for species is the same as in fig. 1B. Nodes with bootstrap support values >90 % are shown as solid circles. Scale bar represents one amino acid substitution per site.

Supplementary figure S4. Maximum likelihood phylogenetic tree of Panarthropoda iGluR/IR ligand-gated ion channel domain (PF00060). The tree includes both complete and partial domains. The color code for species is the same as in fig. 1B. Boxes in the outer circle indicate the genes specifically or differentially expressed in *ANT* (purple), *HEAD* (green), or both (orange). Nodes with bootstrap support values >90 % are shown as solid circles. Scale bar represents one amino acid substitution per site.

Supplementary figure S5. Maximum likelihood phylogenetic tree of the Panarthropoda CD36-SNMP family. The color code for species is the same as in fig. 1B. Boxes in the outer circle indicate the genes specifically or differentially expressed in *ANT* (purple) or in ANT+HEAD (orange). Nodes with bootstrap support values >90 % are shown as solid circles. Scale bar represents one amino acid substitution per site.