Duplications and Positive Selection Drive the Evolution of Parasitism-Associated Gene Families in the Nematode *Strongyloides papillosus*

Praveen Baskaran[†], Tegegn G. Jaleta[†], Adrian Streit, and Christian Rödelsperger*

Department for Evolutionary Biology, Max-Planck-Institute for Developmental Biology, Tübingen, Germany

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: christian.roedelsperger@tuebingen.mpg.de.

Abstract

Gene duplication is a major mechanism playing a role in the evolution of phenotypic complexity and in the generation of novel traits. By comparing parasitic and nonparasitic nematodes, a recent study found that the evolution of parasitism in Strongyloididae is associated with a large expansion in the Astacin and CAP gene families.

To gain novel insights into the developmental processes in the sheep parasite *Strongyloides papillosus*, we sequenced transcriptomes of different developmental stages and sexes. Overall, we found that the majority of genes are developmentally regulated and have one-to-one orthologs in the diverged *S. ratti* genome. Together with the finding of similar expression profiles between *S. papillosus* and *S. ratti*, these results indicate a strong evolutionary constraint acting against change at sequence and expression levels. However, the comparison between parasitic and free-living females demonstrates a quite divergent pattern that is mostly due to the previously mentioned expansion in the Astacin and CAP gene families. More detailed phylogenetic analysis of both gene families shows that most members date back to single expansion events early in the *Strongyloides* lineage and have undergone subfunctionalization resulting in clusters that are highly expressed either in infective larvae or in parasitic females. Finally, we found increased evidence for positive selection in both gene families relative to the genome-wide expectation.

In summary, our study reveals first insights into the developmental transcriptomes of *S. papillosus* and provides a detailed analysis of sequence and expression evolution in parasitism-associated gene families.

Key words: adaptation, nematode, transcriptome, development, stage-specific expression, genome evolution.

Introduction

Gene duplication is one of the major mechanisms involved in the formation of new genes, leading to phenotypic complexity and novel traits (Ohno 1970; Zhang 2003). Well-characterized cases of evolution by gene duplication include the duplication of opsin genes resulting in the formation of color sensitive retina pigments (Dulai et al. 1999), optimization of oxygen binding affinity of human beta-globin duplicates (Hurles 2004) and duplication of immunoglobulin genes in the vertebrate adaptive immune system (Zhang 2003).

Genome sequencing of multiple nematode species has laid the foundation for studying various aspects of genome evolution, including gene duplication in one of the most species-rich animal phyla (Rödelsperger et al. 2013). This includes nematode-specific questions such as those associated with the evolution of mating systems (Denver et al. 2011) and parasitism (Dieterich and Sommer 2009). In their groundbreaking work on the phylogeny of nematodes, Blaxter et al. (1998) suggested that parasitism evolved multiple times independently in the phylum Nematoda. This pattern still holds true even after deeper taxon sampling of the nematode phylogeny (van Megen et al. 2009). Consequently, to gain further insight into the evolution of parasitism at the genomic level, multiple independent studies focusing on different nematode clades and incorporating parasitic species, as well as nonparasitic outgroups, are needed. However, only with the ability to generate numerous high-quality genome assemblies from next

generation sequencing data, it became feasible to investigate the genomic basis of parasitism at high phylogenetic resolution. A recent study in the Strongyloididae family of nematodes identified the CAP and Astacin gene families as primary candidates for genes associated with parasitism (Hunt et al. 2016). These two families showed extreme duplication patterns that coincided with the emergence of parasitism, and they continued throughout speciation events that ultimately resulted in different parasitic species infecting a wide range of vertebrate hosts including humans. Even though the expansion of CAP and Astacin gene families cannot be proven to be the causative event leading to parasitism in Strongyloididae, previous experimental studies have given hints on what role these families may have during infection. CAP proteins are cystine rich secretary proteins, functioning as immunomodulatory molecules in parasitic nematodes (Moyle et al. 1994). Because of their inhibitory effect on neutrophil and platelet activity, these proteins are considered to be a potential vaccine candidate against hookworms (Moyle et al. 1994; Del et al. 2003; Hunt et al. 2016). On the other hand, Astacins are metallopeptidases found in a variety of different species (Park et al. 2010). Functional roles for Astacin proteases in parasitic nematodes include host tissue penetration by infective L3s (Williamson et al. 2006), cuticle formation and ecdysis (Gamble et al. 1996; Stepek et al. 2010, 2011, 2015), and tissue digestion, penetration and migration (Gallego et al. 2005).

The identification of candidate genes associated with parasitism is of utmost importance for the development of potential treatments, especially in a family of parasitic nematodes such as Strongyloididae. The fact that gene duplications have previously been reported to be associated with parasitism in plants, arthropods, and trematodes (Yang et al. 2015; Clarke et al. 2015; Cantacessi et al. 2012) further supports the nature of CAPs and Astacins as strong parasitism-associated candidate genes.

In this study, we focus on the nematode *Strongyloides papillosus*, a common gastrointestinal parasite of domestic and wild ruminants (Jäger et al. 2005). *S. papillosus* has a unique and complex life cycle that consist of parasitic and free-living generations (Streit 2017; Viney and Lok 2015). The adult parasites are only females and live in the mucosa of the small intestine where they lay embryonated eggs that pass with host feces. The parasitic female reproduces parthenogenetically (clonal reproduction). The progenies of parasitic females have two developmental choices: either to develop directly into an all-female infective third stage (iL3) (direct or asexual cycle) or to develop into facultative free-living males and females (indirect or sexual cycle). The free-living generations mate and undergo sexual reproduction in the environment and all their progeny develop into iL3. The life cycle of *S. papillosus* is illustrated in figure 1A. The alternating life cycles between free-living and parasitic generations provide a unique opportunity to apply available genetic and molecular tools for

a better understanding of the basic biology and evolution of parasitism in *Strongyloides* (Streit 2017). Hunt et al. (2016) reported draft genome sequences for four species of *Strongyloides* and transcriptomic comparisons between parasitic and free-living stages for *S. ratti* and *S. stercoralis* but not *S. papillosus*. To complement the work by Hunt et al. (2016), we have sequenced and characterized ten transcriptomes that were sampled throughout the development of *S. papillosus*. Comparative transcriptomic analyses indicate a high degree of conservation in developmental regulation across *Strongyloides* species. Consistent with the results of Hunt et al. (2016), we found that the CAP and Astacin gene families are strongly enriched among genes that are differentially expressed between parasitic and nonparasitic stages. We undertook a detailed analysis of the evolutionary history of both gene families and correlated this with patterns of differential expression. Altogether, these results will improve our knowledge of evolutionary forces that facilitate the acquisition of parasitism in animal parasitic nematodes.

Materials and Methods

Animals and Parasites Husbandry

All experiments were performed on the *S. papillosus* isolate LIN, originally isolated from sheep and maintained in rabbits (Eberhardt et al. 2007). Pathogen-free female rabbits (New Zealand White) purchased from Charles River Laboratories (Sulzfeld, Germany) were subcutaneously inoculated with about 2000 infective larvae (iL3) in order to cultivate different developmental stages of *S. papillosus*. Ten days postinfection (p.i) the feces were collected overnight, mixed with saw dust and cultured at 25 °C as described previously (Eberhardt et al. 2007). Worms of different developmental stages were isolated after the times specified below using the Baermann technique (Eberhardt et al. 2007). Stages 1 and 2 larvae (parasitic L1/L2) derived from parasitic adult females were isolated after 8 h of culturing. Free-living adult males and females were collected after 28 h of culturing. L1/L2, which were progenies of free-living generations were collected after 48 h of culturing. To isolate iL3, the culture dishes were placed in larger dishes with water (Eberhardt et al. 2007) and after 7 days of culture the iL3s that had crawled into the water were collected. These larvae were a mixture of iL3 from the direct and the indirect cycles. The isolate LIN, when kept in rabbits, has such a strong tendency towards the indirect cycle (Eberhardt et al. 2007) that the majority of the iL3 were the progeny of free-living animals. At 22 days after infection, the rabbits were sacrificed to harvest the parasitic adults. The small intestines were dissected from the rabbits, its contents removed, slit longitudinally and carefully washed with PBS. The cleaned intestinal tissues were spread inside down over wire mesh placed on top of a Baermann funnel, filled with phosphate-buffered saline (PBS) and incubated at 37 °C for
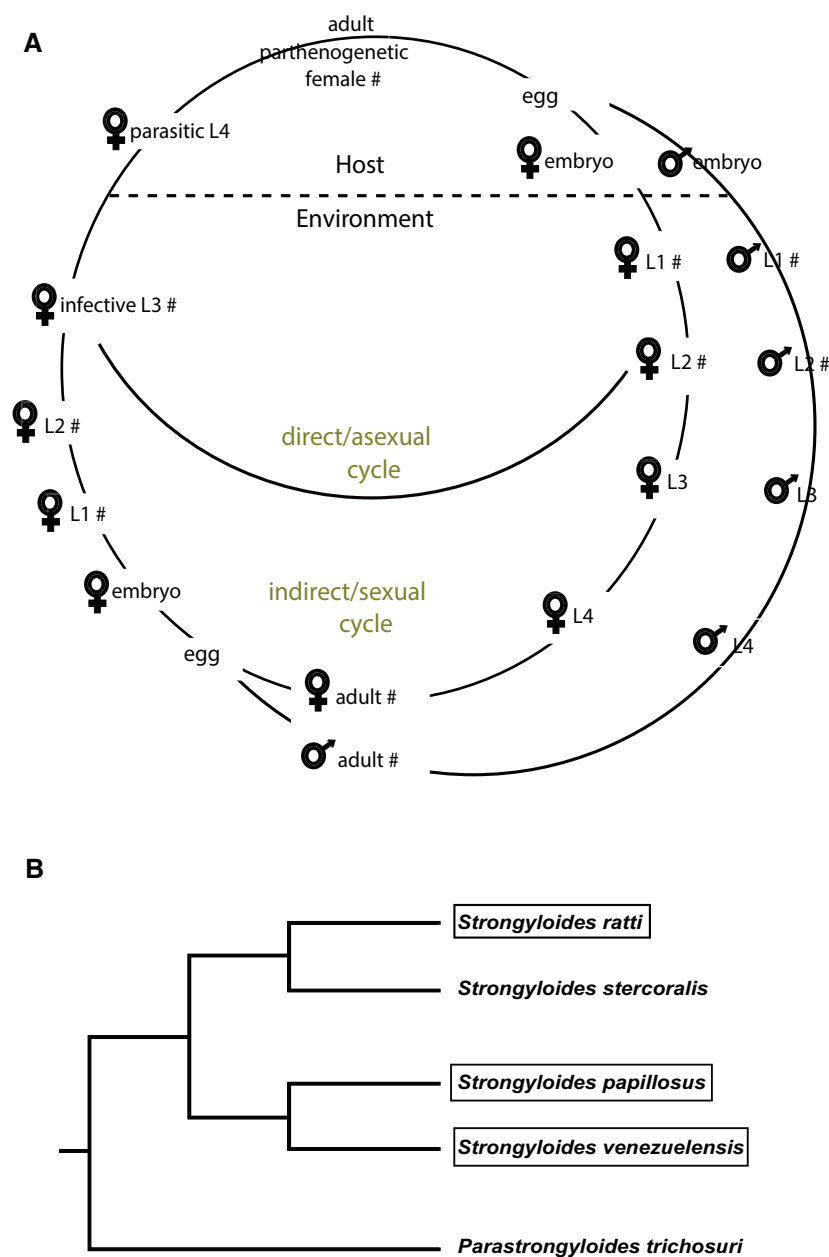
Fig. 1.—The life cycle of *S. papillosus*. (*A*) The life cycle of *S. papillosus* consists of free-living and parasitic stages. Parasitic worms are female only. Within the host, parasitic females reproduce clonally. Their female offspring can either develop directly into infective larvae (iL3) or develop into facultative free-living females that reproduce sexually with males (indirect cycle). Their progeny develops into iL3. Stages for which transcriptomes were sequenced are labeled with #. (*B*) The schematic tree shows the phylogenetic relationships between different *Strongyloides* species (Hunt et al. 2016). The species that are subject of this study are highlighted by a box.

3 h. The worms were then handpicked from the sediment of the Baermann funnel, washed with PBS, counted and stored at −80 °C in TRIzol reagent (Ambion Inc., USA).

Total RNA Extraction, RNA-Seq Libraries Preparations and cDNA Synthesis

Frozen worms were homogenized by multiple freezes and thawed with subsequent vigorous vortexing. Total RNA was isolated with the standard TRIzol extraction method following the manufacturer's instructions (Ambion Inc., USA). The integrity of the RNA was verified by gel electrophoresis and the concentration was measured by a Qubit fluorimeter (Invitrogen Life technologies, USA). RNA-seq libraries were prepared using TruSeq RNA library preparation kit v2 (Illumina Inc., USA) according to manufacturer's instructions of 1μg of total RNA in each sample. The concentration of each

Table 1

Sample Information Table

Developmental Stages	Number of Biological Replicates	Sample Name Replicate 1	Sample Name Replicate 2	ENA Sample id Replicate 1	ENA Sample id Replicate 2
Parasitic females	1	Parasitic female		ERS1214240	
Free living Females	2	Free living female rep1	Free living female rep2	ERS1214234	ERS1214235
Free living males	2	Free living male rep1	Free living male rep2	ERS1214236	ERS1214237
Free living L1/L2 (Free-living mother derived L1/L2)	2	Free living L1/L2 rep1	Free living L1/L2 rep2	ERS1214232	ERS1214233
Infective larvae L3	2	iL3 rep1	iL3 rep2	ERS1214238	ERS1214239
Parasite L1/L2 (Parasitic mother derived L1/L2)	1	Parasitic L1/L2		ERS1214241	

NOTE.—General information about the sequenced developmental stages, number of biological replicates, sample names used in the main text and European nucleotide archive (ENA) accession id.

library was measured by Qubit and Bioanalyzer (Agilent Technologies, USA) and then normalized to 5 nM. The samples were sequenced as 150 bp paired ends in one multiplexed lane using HiSeq2000 platform (Illumina Inc., USA) in an in-house genome facility.

Analysis of Sequencing Data

Sequencing reads were preprocessed using Trimmomatic (Bolger et al. 2014) to remove adapters and trim low quality reads. The preprocessed reads were aligned to the *S. papillo-sus* reference genome (PRJEB525) using Tophat version 2 (Trapnell et al. 2009), followed by expression level estimation using Cufflinks2 (Trapnell et al. 2012). Differentially expressed genes were identified for all possible pairwise comparisons of samples using Cuffdiff (Trapnell et al. 2012). Genes with $\log_2$ fold change greater than 1 or less than −1 and with false discovery rate (FDR) corrected P value <0.05 were treated as differentially expressed. For differential expression analysis by Cuffdiff, available biological replicates (table 1) were grouped together.

Orthologous Clustering and Protein Domain Annotation

In order to determine the homology relationship of *S. papillo-sus* genes, we downloaded *S. ratti* (PRJEB125) and *S. vene-zuelensis* (PRJEB530) gene annotations from WormBase (WBPS4) and classified genes into different homology classes using orthoMCL (Li 2003).

Protein domains of *S. papillosus* protein sequences were annotated using the hmmsearch program in combination with the PFAM domain database obtained from HMMER package (Version 3.0). Domain enrichment analysis of differentially expressed genes was done in R and Fisher's exact test was used to test for statistical significance (FDR-corrected P value <0.05).

Phylogenetic Analysis

Multiple sequence alignments were generated using Clustal omega (Sievers et al. 2014). Individual sequences showing only spurious alignments were removed using trimAl (Capella-Gutiérrez et al. 2009). Prottest (Darriba et al. 2011) was used to identify the best substitution model for tree re-construction. Phylogenetic trees were generated with the help of the Phangorn R package (Schliep 2010). Interactive tree of life web server (Letunic and Bork 2011) was used to display the differential gene expression information of all comparisons for Astacin and CAP genes with the phylogenetic tree. We used CODEML program from the PAML v4.8 (Yang 2007) package to estimate the sequence evolutionary rates and to test for positive selection. Evidence for positive selection was assessed using the site models m7, m8 and m8a. All the models assume different ω values between sites. Alignment sites are divided into ten (m7) or 11 (m8, m8a) classes. Model m7 assumes that all classes of sites are either neutrally evolving or negatively selected (ω ≤ 1). Models m8 and m8a add one explicit class of sites with ω > 1 (m8) and ω = 1 (m8a). We compared the model m8 versus m7 and m8a, respectively, and defined as evidence for positive selection if both the m7 and m8a models could be rejected over m8 in a likelihood ratio test with P value <0.01.

Identification of Conserved Developmental Regulation

In order to combine expression data with phylogenetic data and test for conservation of developmental regulation, we arbitrarily chose two large gene families, G-protein coupled receptors (GPCRs) and Neurotransmitter-gated ion channel (NGIC), for closer investigation. Both families are significantly enriched in various pairwise comparisons of different developmental stages (supplementary figs. S1 and S2, Supplemental Material online). In order to compare our *S. papillosus* data

with expression data from *S. ratti*, we obtained differentially expressed genes in the corresponding stages for *S. ratti* from Hunt et al. (2016) and reconstructed NGIC (fig. 3) and GPCRs (supplementary fig. S4, Supplemental Material online) gene trees using sequences from *S. ratti* and *S. papillosus*. Here, we restricted the pairwise comparison to three stages: free-living female, parasitic female and iL3 stage, because the data for *S. ratti* (Hunt et al. 2016) were available only for these stages. Phylogenetic analysis of these two gene families shows that one-to-one orthologs from *S. papillosus* and *S. ratti* are grouped together in both gene trees, supporting the correctness of our approach for assigning orthologous classes.

Results

The Majority of *S. Papillosus* Genes are Developmentally Regulated

In order to get insights into the developmental processes in *S. papillosus*, we sequenced ten *S. papillosus* transcriptomes that were sampled throughout the development of male and female nematodes (table 1). Principal component analysis of the estimated expression levels indicates that 55% of variation can be explained by the first two principal components (fig. 2A). The ten transcriptomes group into four clusters: young larvae (L1/L2), infective larvae (iL3), adult males and adult females (including parasitic and free-living). This indicates that variation between different developmental stages and sexes is considerably larger than variation between biological replicates (fig. 2A). Therefore, we conclude that the sequenced transcriptomes robustly capture gene expression profiles during the development of *S. papillosus*. By performing differential gene expression analysis in a pairwise manner, we found that the number of differentially expressed genes vary widely from 0.2% (parasitic L1/L2 vs. free-living L1/L2) to 45% (iL3 vs. parasitic female stage) (fig. 2B). Overall, 73% of *S. papillosus* genes are found as significantly differentially expressed (FDR-corrected *P* value <0.05) in at least one comparison. When focusing on the comparison between free-living and parasitic females, we found that only 10% of genes were differentially expressed and 4.4% (917) of genes are up-regulated in parasitic females. This relatively small set of infection-associated genes is consistent with previous data from Hunt *et al.* (2016), which posited that 909 and 1188 genes were up-regulated in parasitic females in *S. ratti* and *S. stercoralis* respectively.

High Degree of Conservation in Developmental Transcriptomes across *Strongyloides* Species

As we have previously observed a strong trend of ancient duplications to have increased gene dosage of developmentally regulated genes in the nematode *Pristionchus pacificus* (Baskaran et al. 2015), we wanted to test whether this also holds true in the comparison between *S. ratti* and *S. papillosus*. Thus, we first classified the differentially expressed genes

into four different classes (Li 2003) based on their orthology relationships with *S. ratti*: one-to-one orthologs (N = 9302 genes) with one gene per species, 6428 *S. papillosus* genes with many-to-X relationships (many-to-many, many-to-one, many-to-zero), *S. papillosus* genes with one-to-many relationship in *S. ratti* (N = 82), and *S. papillosus* singletons (N = 2483) without any closely related gene either in *S. papillosus* or in *S. ratti*. Next, we tested whether genes that are identified as being differentially expressed in particular comparisons tend to be a result of duplication events since the separation from the *S. ratti* lineage. In contrast with previous results from the comparison between *P. pacificus* and *C. elegans*, which represents a much larger evolutionary distance (Baskaran et al. 2015), we found that in *Strongyloides*, the majority of gene sets identified by pairwise comparisons shows a significant enrichment of one-to-one orthologs (FDR-corrected *P* value <0.05, Fisher's exact-test, fig. 2C). One notable exception is the genes that are differentially expressed between free-living females and parasitic females. This set is significantly enriched in genes that have undergone duplications in the *S. papillosus* lineage (many-to-X).

Given the substantial divergence between *S. papillosus* and *S. ratti*, (fig. 2E, inset) and widespread evidence for negative selection ($\omega < 1$, fig. 2E), we further tested whether the conservation is also reflected at the level of expression. Visual inspection of phylogenetic trees (fig. 3 and supplementary fig. S3, Supplemental Material online) in combination with expression data for two large gene families, GPCRs and NGIC, shows that most genes in both gene families are indeed one-to-one orthologs and that expression profiles are indeed very similar. This implies that in addition to the high level of sequence conservation, as shown by the enrichment of one-to-one orthologs and a strong signature of negative selection ($\omega < 1$), the expression profiles between the two species are also highly conserved.

Astacin and CAP Gene Families Are Overrepresented among Genes with High Expression in Parasitic Stages

In order to further characterize the identified gene sets, we performed protein domain enrichment analysis (see Materials and Methods). In total, we found 221 gene families that are significantly enriched in at least two comparisons. Protein kinases (PF00069), HSP20 (PF00011), Motile sperm domain containing proteins (PF00635), Collagens (PF01391), different subfamilies of GPCRs, and a few other families are significantly enriched in more than five comparisons (supplementary figs. S1 and S2, Supplemental Material online). In agreement with the finding of Hunt et al. (2016) based on data from *S. ratti* and *S. stercoralis*, we find the Astacin (PF01400) and CAP (PF00188) gene families to be the most significantly enriched among genes up-regulated in parasitic females in comparison with free-living females ($P < 10^{-20}$, Fisher's exact-test, fig. 2D). Apart from the enrichment of members of these two gene
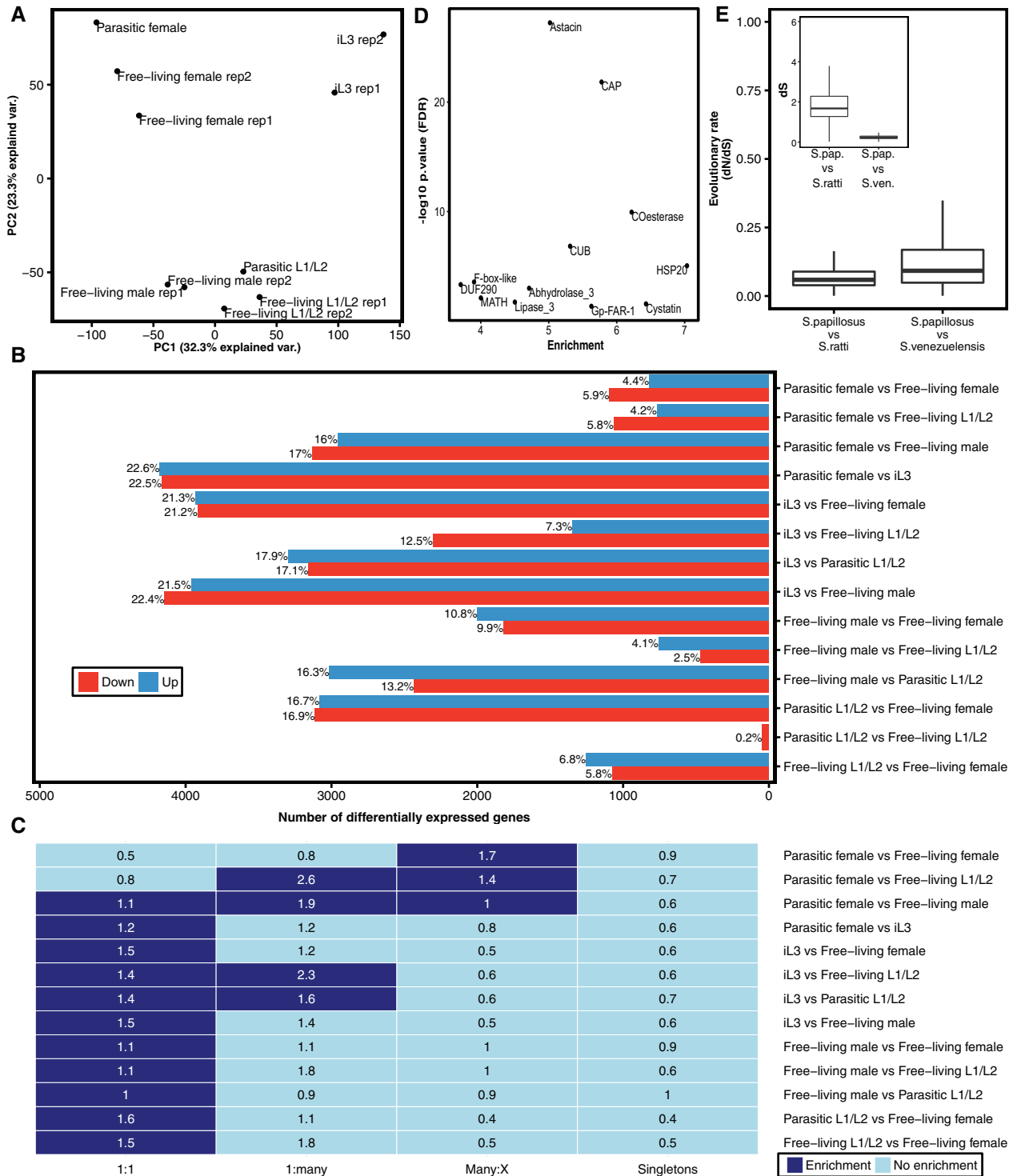
Fig. 2.—Comparison of *S. papillosus* developmental transcriptomes. (*A*) Principal component analysis of expression values shows a subdivision of transcriptomes into distinct clusters that are defined by developmental stage and sex. (*B*) Numbers of significantly differentially expressed genes (FDR-corrected *P* value <0.05) for each comparison. (*C*) Overrepresentation analysis of distinct orthology classes reveals that most comparisons are enriched in one-to-one orthologs. Dark blue color indicates significant enrichment (FDR-corrected *P* value <0.05 and enrichment value >1). The number inside each cell represents the enrichment value. (*D*) Enrichment of protein domains (PFAM) among genes up-regulated in parasitic females in comparison with free-living females (FDR-corrected *P* value <0.05 and enrichment value >1). (*E*) Box plot representation of the evolutionary rates of one-to-one orthologous genes as measured in dN/dS (ω) across two different species comparisons. Both comparisons show that one-to-one orthologous genes are under strong purifying selection. The inlay shows age of one-to-one orthologous genes as measured in dS (synonymous substitutions rate). Median dS of *S. papillosus–S. venezuelensis* one-to-one orthologs is < 1, indicating that number of synonymous substitutions is not saturated.
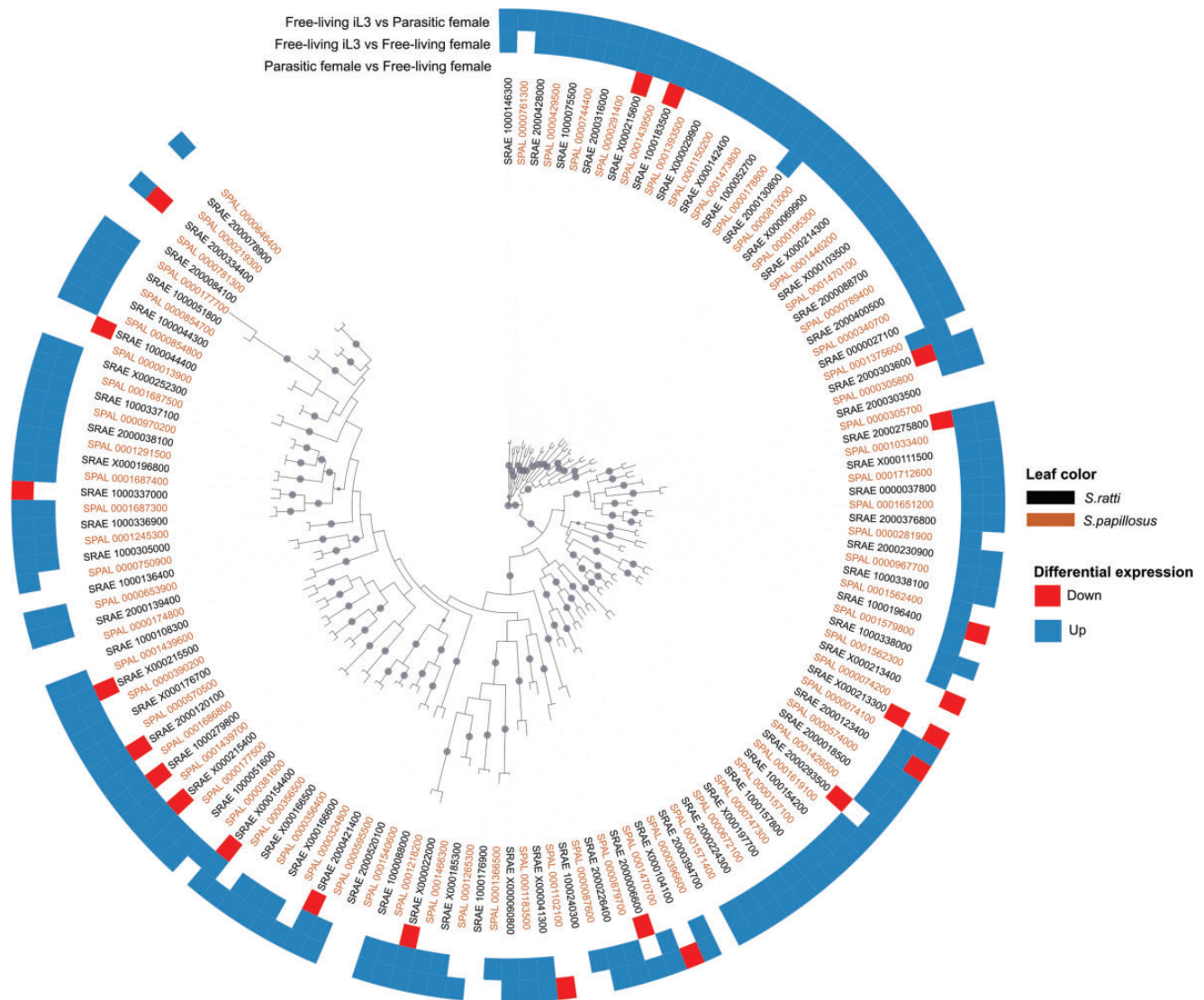
FIG. 3.—Gene expression and sequence evolution of NGIC gene family. Phylogenetic tree of NGIC genes from *S. papillosus* (Brown) and *S. ratti* (black). Heat map around the phylogenetic tree indicates differential expression pattern of each gene in three different comparisons. In the Heatmap, up-regulation, down-regulation and no change are shown in blue, red and white, respectively. This tree shows that a majority of one-to-one orthologs in NGIC gene family have similar expression patterns in *S. papillosus* and *S. ratti*. Dots indicate internal nodes with 100/100 bootstrap support.

families in the parasitic female stage, we also found a considerable number of Astacin and CAP genes up-regulated in iL3.

Individual Expansion Events in the *Strongyloides* Lineage Gave Rise to Most Astacins and CAP Genes

To complement the previous findings of Hunt et al. (2016) with a more detailed analysis of the evolutionary history of CAP and Astacin gene families, we reconstructed phylogenetic trees for both families and integrated the newly acquired *S. papillosus* expression data (figs. 4 and 5, respectively). Both phylogenies show that most genes fall into *S. papillosus* specific subtrees. Similarly, *C. elegans* paralogs also cluster in

lineage-specific subtrees, indicating that there is a trend for duplication of these families even in nonparasitic nematode lineages. Interestingly, the few *S. papillosus* Astacin genes that appear to have one-to-one orthologs in *C. elegans* (*nas-36*, *dpy-31*, *nas-37*, *nas-33*, *nas-30*, *toh-1*, *hch-1* in fig. 5) show quite unique expression profiles, suggesting that these profiles are tightly regulated and are highly dosage sensitive (fig. 5). Although the bootstrap support for a single expansion of CAP genes giving rise to almost all extant members of this family is weak (52/100), there is one branch with a support of 96/100 replicates that contains 97 (59%) of CAP genes (fig. 4). For Astacins, we indeed found one highly supported branch that gave rise to 193 (97%) of all genes in this gene family (fig. 5).
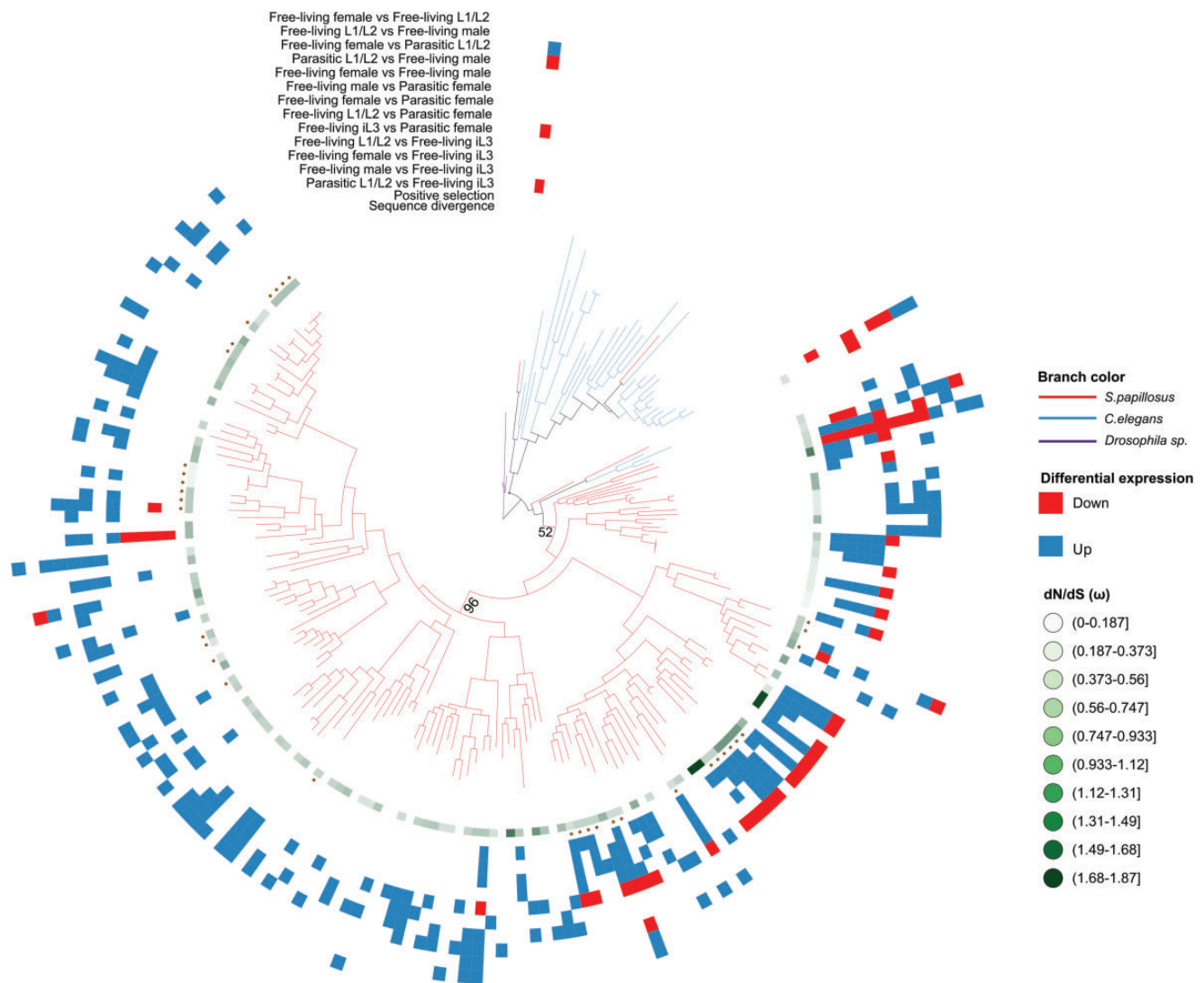
Fig. 4.—Gene duplication in CAP gene family. Phylogenetic tree reconstructed using CAP genes amino acid sequence from *S. papillosus* (red branches), *C. elegans* (blue branches) and *Drosophila Sp.* (purple branches). The Tree shows lineage-specific expansion of CAP genes in the *Strongyloides* lineage. Due to poor visibility, only bootstrap values that are relevant for our analysis are shown. Bootstrap support for all branches of the tree is shown in supplementary figure S5, Supplemental Material online. The color patterns and symbols for each *S. papillosus* gene were generated by testing sequence evolution and differential expression. The green gradient indicates the sequence evolutionary rate as measured in dN/dS (ω). The star symbol indicates the evidence for positive selection in the gene sets. The heatmap of blue, red and white shows the differential expression pattern for each *S. papillosus* gene based on the pairwise comparison of all sequenced developmental stages. Up-regulation, down-regulation and no change in gene expression are shown in blue, red and white, respectively.

Two Major Patterns Dominate Expression Profiles of Astacins and CAP Genes

Visual inspection of the differential expression patterns across both families reflects the above-mentioned finding that most genes either show high expression in parasitic females or in iL3. These two patterns dominate the general picture of expression profiles and comparison with the gene tree reveals a strong phylogenetic signature in the expression profiles, that is, most iL3 specific genes cluster in a specific subtree while genes with high expression in parasitic females cluster in a different subtree (see supplementary figs. S4 and S5, Supplemental Material online for phylogenetic trees with all bootstrap support values included). Even though we are far from understanding the roles of Astacins and CAP genes in *Strongyloides* parasites, our analysis strongly indicates that mutually exclusive groups of Astacins and CAP genes play a role in the stage where parasites infect the host and in the stage where adult parasites proliferate within the host. This clearly demonstrates the mechanism of subfunctionalization in both gene families.
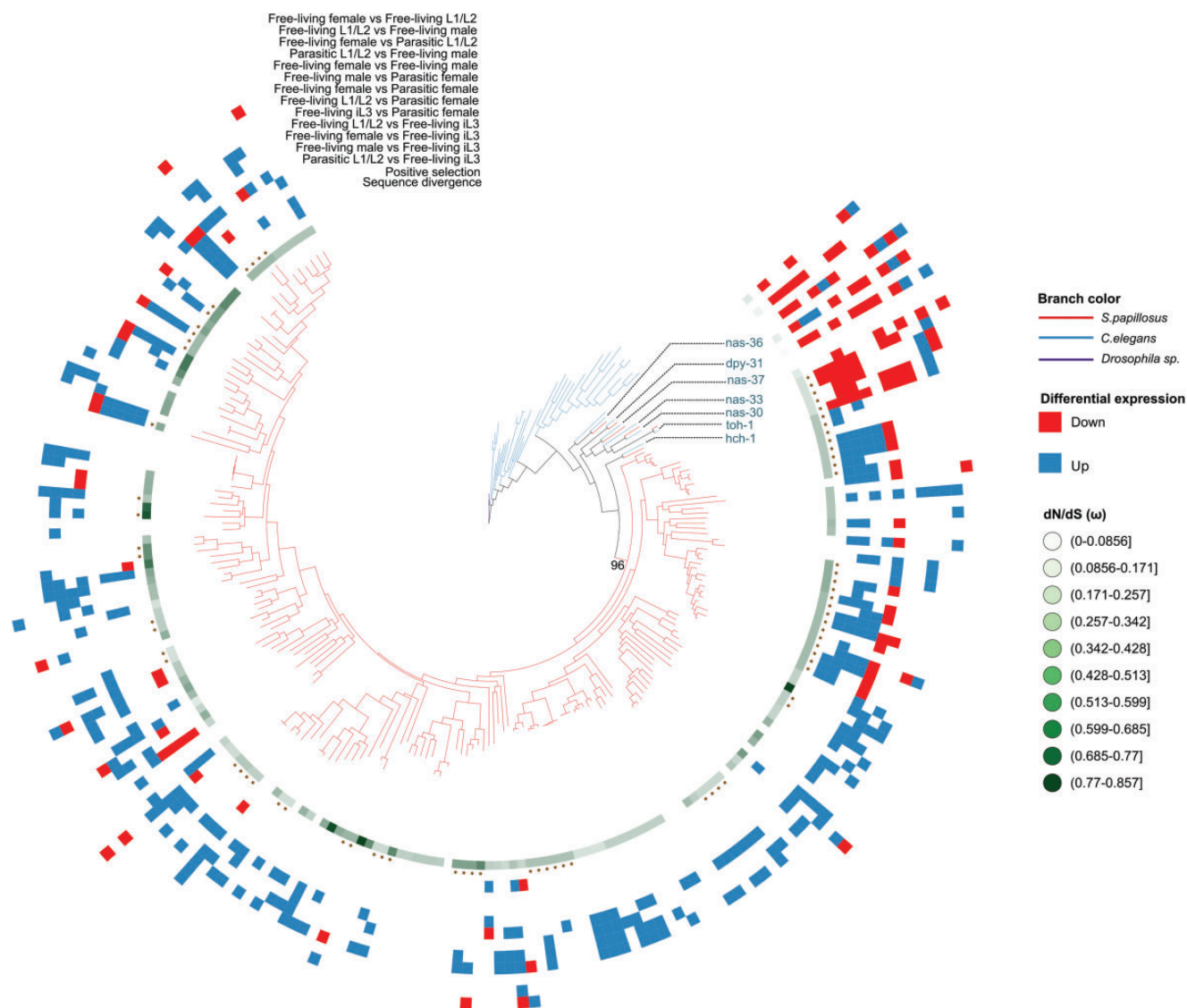
FIG. 5.—Gene duplication in Astacin gene family. Phylogenetic tree of Astacins from *S. papillosus* (red branches), *C. elegans* (blue branches) and *Drosophila Sp.* (purple branches). The tree shows lineage-specific expansion of CAP genes in the *Strongyloides* lineage. A complete tree with labels indicating bootstrap support for all branches of the tree is shown in supplementary figure S6, Supplemental Material online. The color patterns and symbols for each S. *papillosus* gene were generated by testing sequence evolution and differential expression. The green gradient indicates the sequence evolutionary rate as measured in dN/dS (ω). The star symbol indicates the evidence for positive selection in the gene sets. The heatmap of blue, red and white shows the differential expression pattern for each *S. papillosus* gene based on the pairwise comparison of all sequenced developmental stages. Up-regulation, down-regulation and no change in gene expression are shown in blue, red, and white, respectively.

Increased Evidence of Positive Selection in CAP and Astacin Gene Families

Finally, we wanted to test whether the adaptation to novel hosts is paralleled by increased evidence for positive selection. As the divergence between *S. papillosus* and *S. ratti* indicates that synonymous substitutions are already saturated (more than one substitution is expected per synonymous site, fig. 2E inset), we identified orthology clusters with the more

closely related *S. venezuelensis* (Hunt et al. 2016) and screened all orthologous clusters with more than two genes for positive selection. More precisely, we performed a likelihood ratio test to decide whether a model including a number of sites with $\omega > 1$ explains the alignments better than models with either only negative or neutral selection (see Materials and Methods). Overall, we found 28 (67%) and 14 (58%) of orthologous clusters in Astacin and CAP families, respectively, with evidence for positive selection. This represents a strong

enrichment when compared with all other clusters tested ($N = 943$), of which only 21% showed evidence for positive selection (FDR-corrected P value <0.05, Fisher's exact test).

Discussion

The expansion of Astacin and CAP gene families in Strongyloididae has been recently identified as the most extreme genomic feature distinguishing parasitic and nonparasitic lineages within this clade of nematodes (Hunt et al. 2016). In this study, we complemented previous work by sequencing and analyzing the developmental transcriptome of the parasitic nematode S. papillosus. This first characterization of developmental transcriptomes from S. papillosus showed that the majority of S. papillosus genes are developmentally regulated, and, in addition, most of them have one-to-one orthologs in S. ratti. Furthermore, the new transcriptomic data allowed us to gain some insights into the evolution of gene expression in the Strongyloides lineage, which has not been the focus of the previous study by Hunt et al. (2016). Our comparative transcriptomic analysis shows remarkably similar expression between orthologous gene pairs (fig. 3). Given the substantial divergence between the species, this indicates a high degree of evolutionary constraint acting at sequence and expression level. This high degree of conservation is in contrast to our previous study comparing the developmental transcriptome of P. pacificus with C. elegans, in which most developmentally regulated genes are derived from ancient gene duplications (Baskaran et al. 2015). In our opinion, this disagreement is most likely due to the vastly different evolutionary distances between the two comparisons. Consistent with the observation that different timescales will reveal quite distinct evolutionary patterns, we found that even in P. pacificus there is wide-spread evidence for selection against copy number variation when comparing different strains of the same species (Baskaran and Rödelsperger 2015). This microevolutionary picture of P. pacificus is much more similar to our current analysis comparing developmental transcriptomes of S. papillosus and S. ratti.

In light of the generally high evolutionary constraint acting on sequence and expression levels, the divergent patterns in the comparison between free-living and parasitic females can be speculated to represent signatures of adaptive events in response to discrete and continuous changes in the environment, such as host switches and adaptation of the host immune response. This is further supported by the increased evidence for positive selection in Astacins and CAP genes, which are most significantly enriched in the comparison between free-living and parasitic females. The finding of increased evidence for positive selection in these two gene families also distinguishes our work from the study by Hunt et al. (2016), which mainly used the gain of Astacins and CAP genes in parasitic species in comparison to the nonparasitic outgroup together with the enrichment of these gene families

among genes that are upregulated in parasitic relative to nonparasitic females as argument to advertise the patterns observed in these two families as "genomic basis of parasitism."

Together with initial evidence from experimental evolution studies supporting a strong adaptive potential of gene duplications in nematodes (Farslow et al. 2015), our results suggest that despite its generally conservative nature, evolution can be, under certain circumstances, extremely fast. If gene duplications have an adaptive potential, for example, increase the ability of Strongyloides nematodes to infect their host, it is much more likely that this will increase the probability of their retention. Thus, it could be possible that the strongly divergent patterns which dominate comparisons of distantly related species, such as between P. pacificus and C. elegans (Baskaran et al. 2015) or between S. stercoralis and C. elegans (Stoltzfus et al. 2012), reflect the evolutionary history of multiple adaptive events. The alternative scenario that most of these changes can be explained by neutral events (e.g., genetic drift and demographic effects) appears less likely, as the overwhelming picture of several recent studies of genome evolution in nematodes seems to be that selection generally acts against change in protein-coding regions (Rödelsperger et al. 2014, Baskaran and Rödelsperger 2015, this study). We would further like to point out that recent studies of phenotypic evolution indicate towards much stronger selection than what was traditionally assumed by population genetic models (see Messer et al. [2016] for review). The overall evolutionary slower rates of phenotypic evolution at larger timescales suggest that selection on certain phenotypes may often fluctuate in direction. In the case of Strongyloides parasites, fluctuating direction of selection could consist of different host haplotypes encoding for immune response genes that may change in the frequency within the population. If individual gene duplicates would yield different capabilities to overcome specific host immune systems, loss of certain host immune response haplotypes would make certain gene duplicates obsolete and could also explain the higher rate of gene loss in Astacins and CAP genes as observed by Hunt et al. (2016). Thus, despite the extremely limited knowledge about the actual function of individual members of these two families, our study strongly supports the role of Astacin and CAP gene families as prime candidates for gaining mechanistic insights into the infection process and to study its variation across different hosts.

Ethics Statement

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All animal experiments were conducted in accordance with national and international animal welfare guidelines and legislation. The rabbits were kept in an in-house animal facility, that is regularly inspected by the local authorities

(Veterinäramt Tübingen). Permits were granted by the Regierungspräsidium Tübingen (AZ35/9185.82-5).

Data Availability

Raw sequencing data were submitted to the European Nucleotide archive.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Baskaran P, Rödelsperger C. 2015. Microevolution of duplications and deletions and their impact on gene expression in the nematode *Pristionchus pacificus*. PLoS One 10:e0131136.

Baskaran P, et al. 2015. Ancient gene duplications have shaped developmental stage-specific expression in *Pristionchus pacificus*. BMC Evol Biol. 15:doi: 10.1186/s12862-015-0466-2.

Blaxter ML, et al. 1998. A molecular evolutionary framework for the phylum Nematoda. Nature 392:71–75.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Cantacessi C, et al. 2012. Insights into SCP/TAPS proteins of liver flukes based on large-scale bioinformatic analyses of sequence datasets. PLoS One 7:e31164.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Clarke TH, Garb JE, Hayashi CY, Arensburger P, Ayoub NA. 2015. Spider transcriptomes identify ancient large-scale gene duplication event potentially important in silk gland evolution. Genome Biol Evol. 7:1856–1870.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Del VA, Jones BF, Harrison LM, Chadderdon RC, Cappello M. 2003. Isolation and molecular cloning of a secreted hookworm platelet inhibitor from adult *Ancylostoma caninum*. Mol Biochem Parasitol. 129:167–177.

Denver DR, Clark KA, Raboin MJ. 2011. Reproductive mode evolution in nematodes: insights from molecular phylogenies and recently discovered species. Mol Phylogenet Evol. 61:584–592.

Dieterich C, Sommer RJ. 2009. How to become a parasite—lessons from the genomes of nematodes. Trends Genet. 25:203–209.

Dulai KS, von Dornum M, Mollon JD, Hunt DM. 1999. The evolution of trichromatic color vision by opsin gene duplication in new world and old world primates. Genome Res. 9:629–638.

Eberhardt A, Mayer W, Streit A. 2007. The free-living generation of the nematode *Strongyloides papillosus* undergoes sexual reproduction. Int J Parasitol. 37:989–1000.

Farslow JC, et al. 2015. Rapid Increase in frequency of gene copy-number variants during experimental evolution in *Caenorhabditis elegans*. BMC Genomics 16:1044.

Gallego SG, et al. 2005. Identification of an astacin-like metallo-proteinase transcript from the infective larvae of *Strongyloides stercoralis*. Parasitol Int. 54:123–133.

Gamble HR, Fetterer RH, Mansfield LS. 1996. Developmentally regulated zinc metalloproteinases from third- and fourth-stage larvae of the ovine nematode *Haemonchus contortus*. J Parasitol. 82:197–202.

Good RT, et al. 2014. The molecular evolution of cytochrome P450 genes within and between drosophila species. Genome Biol Evol. 6:1118–1134.

Huerta-Cepas J, Gabaldón T. 2011. Assigning duplication events to relative temporal scales in genome-wide studies. Bioinformatics 27:38–45.

Hunt VL, et al. 2016. The genomic basis of parasitism in the *Strongyloides clade* of nematodes. Nat Genet. 48:299–307.

Hurles M. 2004. Gene duplication: the genomic trade in spare parts. PLoS Biol. 2:E206.

Jäger M, et al. 2005. Endoparasites in calves of beef cattle herds: management systems dependent and genetic influences. Vet Parasitol. 131:173–191.

Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39:W475–W478.

Li L. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.

Messer PW, Ellner SP, Hairston NG. 2016. Can population genetics adapt to rapid evolution?. Trends Genet. 32:408–418.

Moyle M, et al. 1994. A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. J Biol Chem. 269:10008–10015.

Nagayasu E, et al. 2013. Transcriptomic analysis of four developmental stages of *Strongyloides venezuelensis*. Parasitol Int. 62:57–65.

Ohno S. 1970. Evolution by gene duplication. Berlin: Springer Verlag. doi: 10.1007/978-3-642-86659-3.

Park J-O, et al. 2010. Characterization of the astacin family of metalloproteases in *C. elegans*. BMC Dev Biol. 10:1–13.

Rödelsperger C, Neher RA, Weller AM, Eberhardt G, Witte H, Mayer WE, Dieterich C, Sommer RJ. 2014. Characterization of genetic diversity in the nematode Pristionchus pacificus from population-scale resequencing data. Genetics 196:1153–1165.

Rödelsperger C, Streit A, Sommer RJ. 2013. Structure, function and evolution of the nematode genome. In: eLS. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9780470015902.a0024603.

Schliep KP. 2010. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593.

Sievers F, et al. 2014. Fast scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:539.

Stepek G, McCormack G, Birnie AJ, Page AP. 2011. The astacin metalloprotease moulting enzyme NAS-36 is required for normal cuticle ecdysis in free-living and parasitic nematodes. Parasitology 138:237–248.

Stepek G, McCormack G, Page AP. 2010. Collagen processing and cuticle formation is catalysed by the astacin metalloprotease DPY-31 in free-living and parasitic nematodes. Int J. Parasitol. 40:533–542.

Stepek G, McCormack G, Winter AD, Page AP. 2015. A highly conserved, inhibitable astacin metalloprotease from *Teladorsagia circumcincta* is required for cuticle formation and nematode development. Int J Parasitol. 45:345–355.

Streit A. 2017. Genetics: modes of reproduction and genetic analysis. Parasitology Mar 22: 1–11. doi: 10.1017/S0031182016000342.

Streit A. 2008. Reproduction in *Strongyloides* Nematoda: a life between sex and parthenogenesis. Parasitology 135:285–294.

Stoltzfus JD, Minot S, Berriman M, Nolan TJ, Lok JB. 2012. RNAseq analysis of the parasitic nematode *Strongyloides stercoralis* reveals

divergent regulation of canonical dauer pathways. PLoS Negl Trop Dis. 6:e1854.

Thompson FJ, Barker GLA, Hughes L, Viney ME. 2008. Genes important in the parasitic life of the nematode *Strongyloides ratti*. Mol Biochem Parasitol. 158:112–119.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Trapnell C, et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 7:562–578.

van Megen H, et al. 2009. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology 11:927–950.

Viney ME, Lok JB. 2015. The biology of *Strongyloides* spp. WormBook. pp.1–17. doi: 10.1895/wormbook.1.141.2.

Williamson AL, et al. 2006. *Ancylostoma caninum* MTP-1, an astacin-like metalloprotease secreted by infective hookworm larvae, is involved in tissue migration. Infect Immun. 74:961–967.

Yang Z, Wafula EK, Honaas LA, Zhang H, Das M, Fernandez-Aparicio M, Huang K, Bandaranayake PC, Wu B, Der JP, et al. 2015. Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evol. 32:767–790.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Zhang J. 2003. Evolution by gene duplication: an update. Trends Ecol Evol. 18:292–298.

Associate editor: Davide Pisani