Molecular Characterization of Chaetae Formation in Annelida and Other Lophotrochozoa

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

(Dr. rer. nat.)

im Fachbereich Biologie, Chemie und Pharmazie

der Freien Universität Berlin

vorgelegt von

Anne-Christin Zakrzewski

aus Potsdam

November 2011

Gutachter

Erster Gutachter:	Prof. Dr. Klaus Hausmann Institut für Biologie Fachbereich Biologie, Chemie, Pharmazie Freie Universität Berlin
Zweiter Gutachter:	Prof. Dr. Thomas Bartolomaeus Institut für Evolutionsbiologie und Ökologie Mathematisch-Naturwissenschaftliche Fakultät Rheinische Friedrich-Wilhelms-Universität Bonn
Datum der Disputation:	16.12.2011

Erklärung

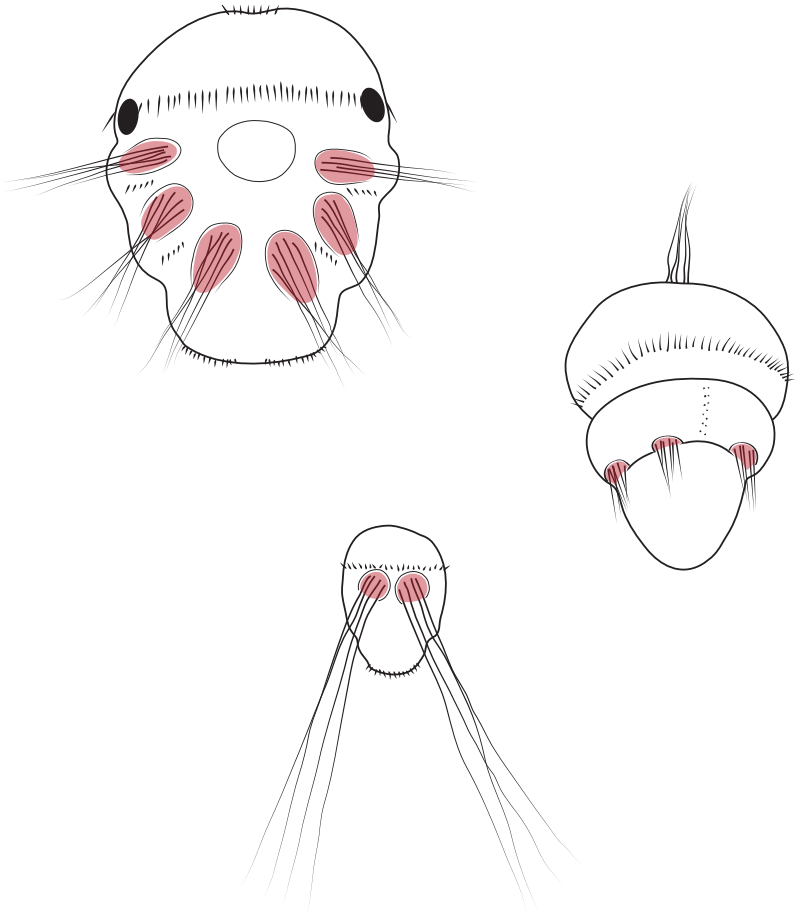
Hiermit erkläre ich, dass ich die vorliegende Arbeit ohne die unzulässige Hilfe Dritter und ohne die Verwendung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Berlin, den 05. November 2011

Anne-Christin Zakrzewski

Molecular Characterization of Chaetae Formation in Annelida and Other Lophotrochozoa



Anne-Christin Zakrzewski

Danksagung

Mein herzlicher Dank geht an Herrn Prof. Dr. Klaus Hausmann, der die abschließende Betreuung dieser Dissertation ermöglichte und dankenswerterweise die Kommisionsleitung des Promotionsverfahrens übernommen hat.
Ganz besonders bedanken möchte ich mich bei Herrn Prof. Dr. Thomas Bartolomaeus für jegliche Unterstützung, die freundliche Aufnahme in seine Arbeitsgruppe und die Betreuung dieser Dissertation.
Mein besonderer Dank geht ebenfalls an Dr. Harald Hausen für seine fachliche Betreuung und die inhaltliche Unterstützung bei etwaigen Stipendienanträgen, die dieser Arbeit zugrunde liegen.

Ich danke für die Unterstützung meiner Forschung durch ein EMBO-Forschungsstipendium für meinen Gastaufenthalt am MFPL in Wien, sowie durch ein Elsa-Neumann-Stipendium. In diessem Rahmen möchte ich auch dem ASSEMBLE-Projekt danken, das die Arbeit in Frankreich ermöglichte. Mein besonderer Dank gilt auch den Angestellten der schwedischen Forschungsstationen „Lovén Centre Kristineberg“ (hier sei besonders Matthias Obst gedankt) und „Lovén Centre Tjärnö“ (insbesondere Lars Hagström und Bertil Rex).

Für die Betreuung während meiner Zeit am MFPL in Wien möchte ich mich besonders herzlich bei Dr. Florian Raible und Dr. Kristin Tessmar-Raible bedanken, die diese Monate sehr fruchtbar machten. Die molekulare Ausgangslage dieser Arbeit wurde durch Prof. Dr. Detlev Arendt und seiner Arbeitsgruppe am EMBL Heidelberg geschaffen und ich danke insbesondere für den Zugriff auf die PEPD-Datenbank und die damit einhergehenden Klone. In diesem Kontext möchte ich auch Dr. Stephan Schneider für die zur Verfügungstellung eines Klones danken.

Dr. Christoph Bleidorn gilt mein besonderer Dank, da er Conrad Helm und Anne Weigert mit ins Projekt brachte und wir eine wunderbare „Myzostomiden-Zeit“ hatten. DANKE!

Da diese Arbeit mit vielen „Standortwechseln“ verbunden war, haben mich viele wunderbare Menschen in dieser Zeit begleitet. Ich kann aus Platzgründen nicht jeden einzelnen nennen. Mein Dank gilt besonders Claudia Lohs, Juliane Zantke und Benjamin Backfisch für eine tolle Zeit am MFPL. Julia Leven, Tobias Kaller und Björn Quast möchte ich für ihre besondere Hilfbereitschaft in Berlin und/oder Bonn danken. Waltraud Brackwehr (Freie Universität Berlin) machte so manche unmöglich anmutenden Dinge möglich – Danke dafür! Für eine gute Zeit an meinem derzeitigen Standort „in der Nähe des Polarkreises“ (Sars Centre Bergen, Norwegen) danke ich vor allem Prof. Dr. Daniel Chourrout, Philipp Sell, Michael Tessmann und Matteo Ugolini.

Esther Ullrich-Lüter kann ich gar nicht genug danken, da ich nicht weiß wo ich anfangen soll. Ich bin einfach froh, daß die vielen Kilometer zwischen uns nichts an unserer Freundschaft ändern konnten. Untrennbar damit verbunden ist mein Dank an Carsten Lüter und seine Begeisterung für Brachiopoden.

Danken möchte ich nicht zuletzt meiner Familie, die mich für diesen Weg vorbereitet, mir immerwährend beigestanden und mich in allen Entscheidungen gestützt haben. Danke für eure Liebe! Auch meinen Schwiegereltern möchte ich an dieser Stelle für ihre tatkräftige Unterstützung danken.

Zuletzt möchte ich mich bei Alexander bedanken, der nun nicht nur eine Diplomarbeit, sondern auch eine Doktorarbeit und das damit einhergehende „Nomadentum“ zusammen mit mir überstanden hat. Danke für deine Geduld und Liebe.

Table of Contents

Index of Figures

Index of Tables

1 Introduction

Many aspects of the phylogenetic interrelationships of the major bilaterian taxa are still a matter of debate, but recent molecular investigations consistently recovered the clades Deuterostomia, Ecdysozoa and Lophotrochozoa. The Lophotrochozoa hypothesis (established by Halanych et al. 1995 and rephrased by Halanych 2004) comprises all traditional Spiralia (except Arthropoda and sometimes Acoelomorpha; e.g., Hejnol et al. 2009), but additionally includes the former radialian taxa of Lophophorata (Brachiopoda, Phoronida, Bryozoa). Several molecular analyses (comprising single gene, 18S + 28S rDNA and multigene analyses) revealed the clade Lophotrochozoa, but differ in terms of a definition *sensu lato* (Petersen & Eernisse 2001, Halanych 2004, Passamaneck & Halanych 2006, Dunn et al. 2008, Giribet 2008, Hejnol et al. 2009, Paps et al. 2009a, 2009b, Nesnidal et al. 2010) or *sensu stricto* (Helmkampf et al. 2008, Nesnidal et al. 2010) by the inclusion or exclusion of a weakly supported clade "Platyzoa" (comprising Platyhelminthes, Gastrotricha, Gnathostomulida, Rotifera and sometimes Myzostomida and Acoelomorpha; e.g., Dunn et al. 2008). The relationships within Lophotrochozoa remain unclear, as the mentioned analyses yielded highly incongruent topologies. In addition to this, no unambiguous shared morphological characters have been identified so far clearly supporting Lophotrochozoa, and evolution of body morphology within the Lophotrochozoa is far from being understood due to the high morphological diversity of this group. In the context of this controversy, the evolutionary origin and phylogenetic relevance of β-chitinous chaetae has been repeatedly discussed (e.g., Eeckhaut et al. 2000, Peterson & Eernisse 2001) as these complex structures are well known from annelids, and similar structures are also present in some other lophotrochozoan taxa.

1.1 Chaetae – characteristic structures in annelids

Many studies have been investigating the morphology and structure of various types of annelid chaetae (e.g. uncini and compound chaetae) in addition to their characteristic arrangement (Bouligand 1966 and 1967, Gustus & Cloney 1973, Specht 1988, Hausen & Bartolomaeus 1998, Bartolomaeus 2002), making it a well-studied structure in annelids that is important for species identification (Bhaud 1978, Fauchald 1977, Thomassin & Picard 1972) and can furthermore be used for phylogenetic considerations (Meyer & Bartolomaeus 1996, Bartolomaeus & Meyer 1997, Bartolomaeus 1998, Hausen 2005). It is known that chaetae emerge from epidermal follicles that in turn consist of one basal chaetoblast and several laterally surrounding follicle cells (Bouligand 1967). Many annelids do not only possess single chaetae, but bundles of them (each emerging from an epidermal follicle), which are united within the so-called chaetal sac. In general, chaetae grow by basal apposition of new material (i.e., consisting of inorganic hardening constituents and cross-linking proteins) that is added to the base of the growing chaeta, where the chaetoblast is located (Bouligand 1967; O'Clair & Cloney 1974). The chaetoblast itself bears a prominent apical microvilli surface, which is suggested to determine the form and shape of the growing chaeta, as cross-sections of chaetae revealed that the chaeta is anchored within the microvilli surface. When the microvilli are withdrawn due to chaetae elongation, typical "honey comb" patterns (Hausen 2005) remain in cross-sections of chaetae, indicating that the hollow channels correspond to the former position of the microvilli.

1.2 Chaetae and chaetae-like structures in other taxa

It has long been known that chaetae occur not only in annelids as traditionally defined, but also in the Echiura (Orrhage 1971), Siboglinidae (George & Southward 1973, Orrhage 1973, Schroeder 1984, Specht 1988) and Myzostomida (Rouse & Fauchald 1997, Eeckhaut et al. 2000 and 2003). Some recent studies group these three taxa with or within the annelids, making up the Annelida *sensu lato* (Hejnol et al. 2009, Struck et al. 2011), although other recent studies place myzostomids near or within a clade of Platyhelminthes and other acoel organisms ("Platyzoa"; Eeckhaut et al. 2000, Halanych 2004, Dunn et al. 2008). Additionally, larval and adult brachiopods are known to possess annelid-like chaetae (Gustus & Cloney 1972, Storch & Welsch 1972, Orrhage 1973, Lüter 2000a), although within Brachiopoda, e.g., *Argyrotheca cistellula* and thecideoid brachiopods lack chaetal structures (D'Hondt & Franzen 2001). Similar structures are also common to other taxa, namely molluscs – sensory hairs of the polyplacophorans (Leise & Cloney 1982) and the Kölliker's organ of juvenile octopods (Brocco et al. 1974) – as well as the gizzard teeth in a subgroup of the bryozoans (Gordon 1975). Ultrastructural data suggest that some of these chaetae-like structures show potentially similar patterns to the annelids, even on the cellular level. The Kölliker's organ, for example, emerges from an epidermal follicle that consists of one basal cell and a few lateral cells surrounding a chaetae-like chitinous structure (Brocco et al. 1974). Especially the potential homology of brachiopod chaetae has been discussed most controversially (Schepotieff 1903, Gupta & Little 1970, Gustus & Cloney 1972, Orrhage 1973, George & Southward 1973), since these are strikingly similar to annelid chaetae.

1.3 Ancient or multiple independent origins of annelid and brachiopod chaetae?

Despite their similarities, even within the most recent morphological studies (Lüter & Bartolomaeus 1997, Lüter 2000a), annelid and brachiopod chaetae have each been interpreted as structures with an independent origin. On the one hand, other brachiopod characters (e.g., coelom formation via enterocoely, modified radial cleavage patterns, lophophore structure) are supposed to be more complex and numerous, therefore outweighing this single complex feature and placing the Brachiopoda within the Lophophorata (= "Tentaculata") at the base of the Radialia (Hyman 1959, Valentine 1973, Ax 1989, Eernisse et al. 1992, Lüter & Bartolomaeus 1997, Lüter 2000b, Nielsen 2001). On the other hand, annelids are placed within the Articulata based on characters such as segmentation, coelomic/mesodermal compartments and metanephridia that they share with the Arthropoda (Ax 1999, Nielsen 2001, Scholtz 2002).

In contrast to this, all molecular-based analyses (e.g., Dunn et al. 2008, Helmkampf et al. 2008, Hejnol et al. 2009, Nesnidal et al. 2010) suggest that brachiopods are much closer related to annelids, as both are grouped within the protostomian clade Lophotrochozoa (Halanych et al. 1995, De Rosa 2001, Passamaneck & Halanych 2006). Although the inner relationships of the Lophotrochozoa

are controversially discussed, some analyses (Dunn et al. 2008, Helmkampf et al. 2008, Paps et al. 2009a and 2009b, Nesnidal et al. 2010) indicate that Annelida and Brachiopoda share a common ancestor together with molluscs, nemerteans and phoronids. However, since several of these taxa lack chaetae, much more comparative information about the complex process of chaetogenesis is needed to infer whether chaetae evolved independently in annelids and brachiopods or if they are homologous and subsequently were reduced several times.

1.4 Molecular tools for unravelling cell type evolution

Over the years, many studies have been investigating chaetae, including their formation, mainly based on histological and ultrastructural investigations (George & Southward 1973, Orrhage 1973, O'Clair & Cloney 1974, Specht 1988, Lüter 2000a, Lanterbecq et al. 2008), indicating that new levels of comparison are necessary to overcome difficulties in homologizing chaetae. Accordingly, the comparison of cells or even cell types involved in chaetogenesis might be of special interest, as these cells are already well-characterized on the ultrastructural level and provide therefore a good base for "molecular fingerprinting" of cell types (Arendt 2005). This strategy promises to detect homology by combining classical morphological features with molecular data and is therefore a good tool for further comparisons. However, chaetae and their formation processes have hardly been studied on the molecular level. Recent studies of genes that are not specific to chaetae formation, such as *engrailed*, *Hox*, *Notch*, *Delta* and *hes2* in the annelid *Capitella teleta* (Seaver & Kaneshige 2006, Fröbius et al. 2008, Thamm & Seaver 2008), consider that some of their studied expression patterns partly overlap with those of chaetal sac regions. Consequently, the question arises whether it is possible to identify specific molecular markers for integrative characterization of cell types involved in chaetogenesis.

1.5 Molecular and structural characteristics of chitin synthesis

While the present study deals with chitinous chaetae, the process of chitin synthesis is obviously of interest. Next to cellulose, chitin is highly abundant in nature and is present in nearly all traditional spiralian taxa (Jeuniaux 1982) – including annelids, molluscs (Peters 1972, Beedham & Trueman 1968, Haszprunar 1992) and arthropods (Rudall 1955) – as well as taxa like cnidarians, Rotifera, Cycloneuralia, entoprocts and ectoprocts, as well as diatoms (Blackwell et al. 1967) and Fungi (Foster 1949, Kent & Whitehouse 1955). Chitin itself is a linear polymer that consists of N-acetyl-glucosamine subunits. Due to a process catalyzed by a membrane-bound enzyme – the chitin synthase – each subunit is polymerized with the elongating chain of the chitin polymer (Merzendorfer 2006). In fungi and arthropods, it is known that chitin mainly contributes to the mechanical strength of the fungal cell walls and the arthropod exoskeleton. However, not all chitinous structures consist of the same variants of chitin. X-ray diffraction methods revealed three different crystalline forms:

α-, β- and γ-chitin (Lotmar & Picken 1950, Rudall & Kenchington 1973, Kramer & Koga 1986). The three variants mainly differ in their arrangement of chains. Whereas the α-type (e.g., specific to the arthropod cuticle) is characterized by an antiparallel orientation of the N-acetyl-glucosamine chains that allows tight packing and thus a higher stability (Giraud-Guille & Bouligand 1986), the β-form (mainly found in annelids and brachiopods) shows an arrangement of chains orientated in a parallel manner, which contributes to more flexibility (as required for chaetae) and soft chitinous structures (required for peritrophic matrices and the peritrophic matrix-formed cocoons in insects or the cuticle of the pedicle in brachiopods; Rudall 1955 and 1963, Rudall & Kenchington 1973). As arthropods are characterized by a prominent α-chitinous exoskeleton, as well as α-chitinous trachea and β-chitinous peritrophic matrices lining the gut epithelium, it is thus not surprising that chitin synthesis is best-studied in this group, in particular in insects (Merzendorfer & Zimoch 2003, Merzendorfer 2006). For example, because the exoskeleton surrounds the complete body, insect chitin is secreted by epidermal cells onto the epidermal surface of the entire organism. In contrast, formation of β-chitinous chaetae in annelids requires a distinct and local chitin synthesis and chitin secretion (i.e., by the few cells within a chaetal follicle), as chaetae are filamentous, thin structures (exhibiting a variety of different morphological types) and are therefore suggested to require a fine-tuned regulation (Hausen 2005). Bouligand (1967) and O'Clair & Cloney (1974) proposed a cellular model of chaetogenesis via interaction of a chaetoblast and several follicle cells, which has been confirmed by many investigators over the years. Nevertheless, many gaps remain in understanding especially the molecular details of the chaetae formation process, such as the localization of chitin synthesis as well as what kind of mechanism is producing such a diversity of chaetal structures. In this context, the chitin synthase group of enzymes is of special interest, as in the well-studied Fungi and Ecdysozoa (especially insects and nematodes) at least two different chitin synthases are present within an organism and perform chitin polymerization with a temporal and spatial differentiation (Roncero 2002, Zhang et al. 2005, Merzendorfer 2006). For example, the insect chitin synthase 1 is involved in ectodermal cuticle formation (Merzendorfer 2006), whereas the other (chitin synthase 2) forms peritrophic matrices of the endoderm (Ibrahim et al. 2000). As within Lophotrochozoa, chitin synthases have so far only been found in bivalve molluscs (Weiss et al. 2006, Suzuki et al. 2007), the reconstruction of the evolution of chitin synthases within Metazoa has been restricted to the α-chitin-bearing Ecdysozoa (Zhu et al. 2002).

1.6 Suitable organisms for studying chaetogenesis

Platynereis dumerilii is a typical nereid "polychaete" with a well-studied development (Hauenschild & Fischer 1969, Fischer et al. 2010) and is easy to keep as well as breed in high quantities, which in turn defines it as a good model organism. Even more important, transcriptome and low-coverage genome data of *P. dumerilii* were recently generated, which, in combination with EST libraries and a random *in situ*-hybridization screening of D. Arendt and colleagues (EMBL Heidelberg, Germany), provided an ideal start point for studying chaetogenesis on a molecular level. Additionally, *Capitella*

teleta was selected for this study, as it is a "polychaete" featuring a direct development, different types of chaetae and a completely sequenced genome. In completing the taxon sampling, a representative of the chaetae-bearing, annelid-like myzostomids was included, namely *Myzostoma cirriferum*, because transcriptome data is available from a collaborator (C. Bleidorn, University of Leipzig, Germany). Finally, as the annelid-like brachiopod chaetae have been discussed most controversially, *Macandrevia cranium* was added to the studied organisms. It is a typical articulate brachiopod which, in addition to the general adult chaetae that surround the complete mantle epithelium, possesses two bundle pairs of larval chaetae (*contra* D'Hondt & Franzen 2001). In the process of this research, our lab (H. Hausen, Sars Center Bergen, Norway) together with our collaborators (Michael Kube, MPI for Molecular Genetics Berlin, Germany) generated *M. cranium* 454 transcriptome sequences.

1.7 Aims of the present study

The aims of this thesis were to unravel novel aspects of chaetae formation by combining structural and molecular methods for the first time. Various gene expression studies based on a comprehensive set of marker genes using different methods of *in situ*-hybridizations were performed. These screenings comprised four different developmental stages of *P. dumerilii* and one stage in each of the three additional study organisms, *C. teleta*, *M. cirriferum* and *M. cranium*. To enable a subsequent cell type characterization via these specific molecular markers, the expression studies were complemented by in-depth structural and immunhistochemical analyses of the cellular components of a typical annelid chaetal sac. Furthermore, orthology or paralogy of the molecular markers was determined via evolutionary analyses of their sequences together with previously published sequences of other organisms. And finally, the amino acid sequences of these novel markers were studied in the light of potential structural characteristics indicating their possible role during chaetogenesis.

These strategies gave me the unique opportunity to assess homology of chaetogenesis cell type markers among the two sampled annelids, the myzostomid and the brachiopod, to gain new insights into the evolution of lophotrochozoan chitinous chaetae.

2 Material & Methods

2.1 Animal supply and culture (incl. breeding protocols)

Platynereis dumerilii **(Audouin & Milne-Edwards, 1833) (Annelida: Nereididae)**

All specimens were obtained from our own *P. dumerilii* culture (initially at FU Berlin, then at University Bonn and at Sars Centre Bergen) which is maintained following the procedure described by Dorresteijn (1990) and Fischer & Dorresteijn (2004). All developmental stages (including the adults) were kept at an ambient temperature of 18°C. As *P. dumerilii* is sensitive to moonlight, the culture was held under a strict light cycle similar to the natural moon cycle – one week of artificial moon followed by three weeks without any moonlight. The animals matured 7-14 days after the moon was switched off. For fertilization, the mature males and females were picked from their aquaria, washed in natural sea water (NSW) and transferred to a glass bowl with NSW. After successful fertilization, the batches were kept in an incubator at 18°C until they developed to the desired developmental stage. One day after fertilization, the embryos were de-jellied, dead embryos discarded and the batches refilled with fresh NSW.

Capitella teleta **Blake, Grassle & Eckelbarger, 2009 (Annelida: Capitellidae)**

Embryos of *C. teleta* were received from our in-house culture at Sars Centre Bergen (Norway). The stem colony originated from a culture of the Grassle Lab (Rutgers University, New Jersey/USA) and was maintained in the laboratory at 18°C according to the culture methods developed by Grassle and Grassle (1976) and Seaver et al. (2005). As in *P. dumerilii*, *C. teleta* was also held under a strict light cycle, which was synchronous with the natural moon cycle. In our culture the animals were kept in plastic boxes (10 x 10 x 6 cm^3) with NSW. They live and feed on a substrate that is slightly anoxic (grain size ≤ 500 µm; mud flats). The sediment was changed every second week. During this process the substrate was checked for brooding tubes filled with larvae which were then analyzed under a dissecting microscope. After opening of the tubes, the larval stages were selected and sorted for different developmental stages, according to Seaver et al. (2005).

Myzostoma cirriferum **Leuckart, 1836 (Myzostomida)**

In collaboration with C. Bleidorn (University of Leipzig, Germany), the different developmental stages of *M. cirriferum* were collected in Roscoff in December 2009 and spring 2011. Like all representatives of the taxon *Myzostomida*, *M. cirriferum* is associated with echinoderms (Grygier 2000). In this special case, *M. cirriferum* lives ectocommensal on the crinoid *Antedon bifida* which was collected by scuba diving at Morgat (Brittany, France). To harvest the myzostomids from their hosts' surface, several crinoids were rinsed in 10% $MgCl_2$ (in NSW). This procedure anaesthetized the myzostomids which then fall off their hosts. After quick washes in NSW, both myzostomids and crinoids recovered from the $MgCl_2$ treatment. As myzostomids are hermaphrodites, all individuals were separated into 6-well plates with NFSW and cultured at room temperature. After a few hours,

these animals laid batches of fertilized eggs. To prevent bacterial growth, the water of the cultures was changed several times per day. For stages older than larvae, free living young stages were picked from their hosts and fixed directly. The most critical stages to obtain were juveniles and young males which remain within the pinnules of the crinoid arms. To get these stages, it was necessary to screen for localized pinnular deformations (Eeckhaut & Jangoux 1993) and subsequent dissection.

***Macandrevia cranium* (Müller, 1776) (Brachiopoda: Rhynchonelliformea)**

Adult specimens of *Macandrevia cranium* were collected by dredging in October 2007 and 2008 at a depth of 100-110 m near Tjärnö (Sweden; 58° 59.170 N, 011° 05.102 E to 58° 59.200 N, 011° 04.935 E). To ensure constant ambient temperature, the animals were maintained in cold rooms with running seawater at 7-8°C. Embryos were obtained by a strict protocol of artificial fertilization and a constant ambient temperature of 7-8°C. Gametes were collected using a pipette with a disposable tip. The eggs were transferred to meshes (40 μm) and maintained in running seawater for 8 h, whereas the sperms were re-suspended in tubes with natural filtered seawater (NFSW) until motility was observed under the microscope. Fertilization occurred in 6-well plates filled with NSFW. Ten minutes after addition of sperms, the fertilized eggs were washed several times by transferring the meshes to new wells containing NFSW. Subsequent to the washing steps, the plates were connected to flowing seawater for at least 24 hours. When the developing embryos started swimming in the water column, they were transferred to nets containing 6-well plates filled with NFSW. The entire NFSW was changed once a day and the embryos or developing larvae were sampled in time intervals of 3 h until 183 hours past fertilization (hpf).

Embryo fixation

In all breeding experiments, stages were fixed for different methods such as ultrastructural studies, immunohistochemistry, *in-situ* hybridization and RNA extraction. For protocols, see chapters 2.3 Immunhistochemistry, 2.4.2 Whole-mount *in-situ* hybridization and 2.5.1 Transmission electron microscopy.

2.2. Molecular cloning – general techniques

2.2.1 Extraction of total RNA

TRIzol (Invitrogen, 15596018) was used to isolate total RNA from *P. dumerilii* (of 48h-stages) and *C. teleta* embryos of mixed stages (mainly stages 5-8). Extraction of total RNA from *M. cranium* and *M. cirriferum* was done using the RNeasy Mini Kit (protocol: Purification of Total RNA from Animal Cells Using Spin Technology; Qiagen, 74104). The concentration of isolated RNA was determined by spectrophotometry (NanoDrop, Peqlab). Quality and integrity of total RNA was examined using

gel electrophoresis (2 µl total RNA plus 3 µl 2x RNA Loading Dye [NEB, B0363S]; 1% agarose gel in 1x TAE). Finally, the isolated RNA was stored at -80°C.

2.2.2 cDNA library synthesis

The reverse transcription of total RNA into cDNA took place using the SuperScript II Reverse Transcriptase Kit (Invitrogen, 18064-022), the Transcriptor High Fidelity cDNA Synthesis Kit (Roche, 05081955001) or the SMART™ RACE cDNA Amplification-Kit (Clontech, 634914) and the SMARTer™ RACE cDNA Amplification-Kit (Clontech, 634923), respectively. Depending on further experiment strategies, the first-strand cDNA synthesis was processed using either random hexamers, gene-specific primers or oligo(dT)s. cDNA-libraries were generated for *P. dumerilii* (48-75 hpf stages), *C. teleta* (stages 5-8), *M. cranium* (mixed stages) and *M. cirriferum* (juveniles and adults).

2.2.3 Specific PCR

In terms of gene candidates, whose BLAST searches against the *P. dumerilii* transcriptome yielded hits containing more sequence information up- and downstream the existing clone fragment, PCR strategies based on specific primers were used. This was the case for elongation of the chitin synthase and myosin light chain kinase of *P. dumerilii*. Several different kits (HotStarTaq Plus DNA Polymerase, Qiagen 203601; Phusion High-Fidelity DNA Polymerase, Finnzymes F-530S; Go Taq Flexi DNA Polymerase, Promega M8305) were used to perform PCRs, but the final concentration of the used components for each PCR reaction were similar (Tab. 1).

Tab. 1: General PCR reaction mixture.

components	final concentration
10x buffer	1x
dNTPs (10 mM each)	200 µM each
$MgCl_2$ (25 mM)	1.5-2.5 mM
primer (10 µM)	0.2-2 µM
DNA Polymerase	1-1.25 U/50 µl
template	X µl
Milli-Q	X µl

The alignment for transcriptomic sequences with data from the *Platynereis* EST library (D. Arendt et al., unpublished data) was processed using the 'ClustalW Multiple Alignment' or 'Cap contig assembly program' function of BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). The primer binding region was chosen by eye or via the public software Primer3 (Rozen and Skaletsky, 2000; http://frodo.wi.mit.edu/primer3/) and the melting temperature was calculated with the public software NetPrimer (http://www.premierbiosoft.com/netprimer/index.html).

The specific PCR programs for the different kits are listed in below. The resulting PCR products were assessed under ultraviolet light after gel electrophoresis and subsequent staining with ethidium bromide (EtBr).

Hot Star Taq program

1.) 95°C – 5 min, 2.) 94°C – 45 sec, 3.) X°C – 45 sec, 4.) 72°C – Y min, 5.) repeat steps 2.-4. 34 times, 6.) 72°C – 10 min, 7.) store at 4°C ∞

Phusion program

1.) 98°C – 2 min, 2.) 98°C – 10 sec, 3.) X°C – 1 min, 4.) 72°C – X min, 5.) repeat steps 2.-4. 32 times, 6.) 72°C – 10 min, 7.) store at 4°C ∞

Go Taq Flexi program

1.) 94°C – 5 min, 2.) 94°C – 30 sec, 3.) X°C – 30 sec, 4.) 72°C – X min, 5.) repeat steps 2.-4. 34 times, 6.) 72°C – 10 min, 7.) store at 4°C ∞

2.2.4 Degenerated primed PCR

The degenerated PCR strategy was used to amplify the chitin synthase genes in *P. dumerilii*. Based on amino acid (AA) sequence alignments of known protostome chitin synthase genes degenerated, primers (i.e., exhibiting variable nucleotide sites; the listed abbreviations for ambiguous nucleotides follow the IUPAC nomenclature of nucleic acids) in combination with specific primers were designed in conserved sequence regions (see Appendix Tab. 1). The alignment of those sequences was compiled by MA FFT (E-INS-i; Katoh & Toh 2008) and ClustalX (http://www.clustal.org/clustal2/). The software Oligo6 (Molecular Biology Insights Inc., Cascade, USA) and a standard codon-usage table was used to generate the primers and calculate the specific melting temperatures.

All PCRs were performed with cDNA templates. The mainly used cDNA libraries were RACE-cDNA enriched for 3' and 5' elongation, respectively. But also cDNA synthesized with gene-specific primers was applied. To perform the PCR, the HotStarTaq Plus DNA Polymerase (Qiagen, 203601) was used. The PCR reaction (50 µl) was set up as follows (Tab. 2).

Tab. 2: PCR reaction mixture used for degenerated PCR.

Components	volumes
10x buffer	5 µl
dNTPs (10 mM each)	1 µl
Hot Star Taq	0.25 µl
degenerated primer (100 µM)	1.5 µl
specific primer (5 µM)	2.5 µl
Milli-Q water	36.75 µl
Template	3 µl

The following PCR program is the basic program which was used to amplify the genes of interest. Minor or major variations in the protocol were done regarding the annealing temperature (depending on the different primers), number of cycles and elongation times (depending on expected fragment size) (Tab. 3).

Tab. 3: PCR program used for degenerated PCR.

initial denaturation	95°C	5/15 min	
denaturation	95°C	1 min	
annealing	X°C (-5°C)	2 min	6 cycles
elongation	72°C	3-4 min	
denaturation	94°C	1 min	
annealing	X°C	2 min	36 cycles
elongation	72°C	4 min	
elongation	72°C	10 min	
stored at	4-10°C		

In addition to the degenerated PCR, a nested PCR was conducted to enrich the expected fragment in the PCR product. For this, 1 µl of the first PCR was diluted in 2 µl of Milli-Q generating the template of the nested PCR. The PCR reaction was set up as described above. The program to run the reaction was identical to the first PCR but lacked the first six cycles. Resulting products were analyzed by running an agarose gel (Sigma, A6013) with the appropriate agarose concentration (0.8-2% agarose in 1x TAE) and stained with EtBr or SybrSafe. All bands with the expected size were excised and ligated into vectors for cloning. In the case of no bands because of weak amplifications, the gel was prepared for a southern blot (see chapter 2.2.7 High stringency Southern Blots and Radioactive Hybridization).

2.2.5 TAIL-PCR

According to Liu & Whittier (1995), TAIL-PCR is an abbreviation of thermal asymmetric interlaced PCR that uses three nested specific primers and one degenerated primer which is designed to bind

arbitrarily to sequences. The principle is to specifically amplify fragments of unknown sequence that flank known sequences. In this study, the TAIL-PCR approach was used to amplify upstream and/or downstream regions of genes of interest out of BAC DNA. Therefore the specific primers were selected by eye based on the criteria of a length of approximately 30 nucleotides and a melting temperature of around 70°C. The melting temperature was calculated with the online available NetPrimer software. The degenerated primers (see Appendix Tab. 2) were generated by Florian Raible (MFPL Vienna, Austria). In three sequential nested reactions, each of the eight degenerated primers was used in combination with the specific primers.

The reaction mixtures for each TAIL-PCR reaction are as follows:

1st TAIL reaction:

1 µl BAC DNA or 3 µl cDNA; 1 µl degenerated primer 1-8 (100 µM); 0.8 µl specific primer for 1st TAIL reaction (5 µM); 1 µl dNTPs (10 mM each); 4 µl 5x buffer HF (Finnzymes); 0.25 µl Phusion Polymerase (Finnzymes); add Milli-Q up to 20 µl total reaction volume.

2nd TAIL reaction:

1 µl 1:50 dilution of the 1st TAIL reaction; 1 µl degenerated primer 1-8 (100 µM); 0.8 µl specific primer for 2nd TAIL reaction (5 µM); 1 µl dNTPs (10 mM); 4 µl 5x buffer HF (Finnzymes); 0.25 µl Phusion Polymerase (Finnzymes); add Milli-Q up to 20 µl total reaction volume.

3rd TAIL reaction:

2.5 µl 1:50 dilution of the 2nd TAIL reaction; 2.5 µl degenerated primer 1-8 (100 µM); 2 µl specific primer for 3rd TAIL reaction (5 µM); 2.5 µl dNTPs (10 mM); 10 µl 5x buffer HF (Finnzymes); 0.5 µl Phusion Polymerase (Finnzymes); add Milli-Q up to 50 µl total reaction volume.

The programs for each TAIL reaction are given in the following:

1st TAIL PCR:

1.) 98°C – 30 sec, 2.) 94°C – 1 min, 3.) 95°C – 1 min, 4.) 94°C – 1 min, 5.) 65°C – 1 min, 6.) 72°C – 45 sec, 7.) repeat steps 4.-6. 5 times, 8.) 94°C – 1 min, 9.) 25°C – 3 min, 10.) 72°C – 3 min, 11.) 94°C – 30 sec, 12.) 64°C – 1 min, 13.) 72°C – 3 min, 14.) 94°C – 30 sec, 15.) 64°C – 1 min, 16.) 72°C – 3 min, 17.) 94°C – 30 sec, 18.) 44°C – 1 min, 19.) 72°C – 3 min, 20.) repeat steps 11.-19. 14 times, 21.) 72°C – 5 min, 22.) store at 10°C

2nd TAIL PCR:

1.) 98°C – 10 sec, 2.) 94°C – 30 sec, 3.) 64°C – 1 min, 4.) 72°C – 3 min, 5.) 94°C – 30 sec, 6.) 64°C

– 1 min, 7.) 72°C – 3 min, 8.) 94°C – 30 sec, 9.) 44°C – 1 min, 10.) 72°C – 3 min, 11.) repeat steps 2.-11. 13 times, 12.) 72°C – 5 min, 13.) store at 10°C

3rd TAIL PCR:

1.) 98°C – 10 sec, 2.) 94°C – 30 sec, 3.) 64°C – 1 min, 4.) 72°C – 3 min, 5.) 94°C – 30 sec, 6.) 64°C – 1 min, 7.) 72°C – 3 min, 8.) 94°C – 30 sec, 9.) 44°C – 1 min, 10.) 72°C – 3 min, 11.) repeat steps 2.-11. 11 times, 12.) 72°C – 5 min, 13.) store at 10°C

5 µl of the three or at least the last two TAIL-PCR reactions were analyzed on a 0.8-2% agarose gel. When a band was visible, the respective fragment was cloned.

2.2.6 RACE-PCR

RACE primers and their binding region within the known sequence were chosen by eye. Melting temperatures and secondary structures were checked using the public software NetPrimer. RACE-PCR was performed using one specific primer in combination with the universal primer (UPM) and a cDNA template that was created with the SMART™ RACE/SMARTer™ RACE kit (Clontech). All PCR products were analyzed by gel electrophoresis. Subsequently, the gels were southern blotted and hybridized with high stringent conditions (see 2.2.7) using the hitherto cloned fragment of the gene as a probe. Bands of samples with a positive hybridization signal were excised from an agarose gel and directly subcloned (for detailed information see chapter 2.2.9 Cloning). For some genes, the Advantage® 2 PCR Kit (Clontech, 639206) was used after RACE-PCR (instead of southern blotting and radioactive hybridization) to select for the correct fragment. The protocol was identical to the user manual.

2.2.7 High stringency Southern Blots and Radioactive Hybridization

After gel electrophoresis and visualization under ultraviolet light, the gel was placed in a denaturing solution (1.5 M NaCl, 0.5 M NaOH) and incubated for 30 min at room temperature. The gel was southern blotted overnight using a nylon blotting membrane with a pore size of 0.45 µm (Peqlab, 39-2010). The membrane was rinsed with 0.2x SSC/0.1% SDS and pre-hybridized in RapidHyb Buffer (Amersham Life Science) for 30 min at 65°C. Parallel to pre-hybridization, the probe was denatured in a thermo block at 100°C for 5 min and labeled with the RadPrime DNA Labeling System Kit (Invitrogen, 18428-011) according to manufacturer's instructions. The radioactive nucleotide that was used was P^{32}dCTP. Before hybridization, the radioactive labeled probe was purified using the illustra ProbeQuant™ G-50 Micro Columns (GE Healthcare, 28-9034-08). Half of the total amount of the labeled probe (denatured at 95°C for 5 min) was added to the pre-hybridized blotting membrane and hybridized for 2.5 h at 65°C. Afterwards, the membrane

was rinsed once shortly and two times for 30 min in 0.2x SSC/0.1% SDS at 65°C. The blotting membrane was exposed with an intensifier screen at -80°C until a signal was detected.

2.2.8 Gel electrophoresis and gel purification

A suitable amount of the PCR product (in general, 5 µl) was mixed with 6x DNA Loading Dye (NEB, B7021S) and analyzed on a 0.8-2% agarose/1x TAE gel (Sigma, A6013). The gel was run in a *kuro*Gel miniplus 10 (VWR) and *kuro*Gel midiplus 15 electrophoresis unit (VWR), respectively. After electrophoresis, the gel was stained in an EtBr/1x TAE (Sigma, 46067) solution for 15-30 min. Alternatively, SYBR® Safe (Invitrogen, S33102) was added to the hardening gel. The DNA of the excised bands was isolated using the MinElute Gel Extraction Kit (Qiagen) following the user manual. The resulting eluate was used for ligation into an appropriate cloning vector.

2.2.9 Cloning

Depending on the used polymerase, desired fragments were cloned using the pGEM®-T Easy Vector System (Promega, A1360) or CloneJET™ PCR Cloning Kit containing the pJET1.2blunt vector (Fermentas, K1231). For recipes of the ligation mixture, see manuals. Transformation was done using One Shot® TOP10 Chemically Competent *E. coli* cells (Invitrogen, C4040-03). A water bath (VWR, Grant GD100) at 40°C was used to heat shock the competent cells. On an LB-agar plate (containing 50 µg/ml of Ampicillin), the cell ligation mixture was spread out and placed into an incubator at 37°C overnight. At least 4 colonies were picked and set up for overnight culture in LB medium (containing 50 µg/ml of Ampicillin) at 37°C in a shaking incubator. Colony PCR or restriction enzyme digest was used to check if the cloning was positive and if the transformed plasmid had the correct insert size.

2.2.10 Colony PCR

The standard primer pairs M13FW, M13REV and Sp6, T7 were used to check for an insert. For this, either half of the colony or 1 µl of the overnight culture was applied as template. The colony PCR reaction was set up as follows:

5 µl 5X Green GoTaq® Flexi Buffer; 5 µl MgCl2; 0.5 µl dNTPs (10mM each); 1 µl FW primer (10 µM); 1 µl REV primer (10 µM); 0.15 µl GoTaq® Flexi DNA Polymerase (Promega); add Milli-Q up to 25 µl total reaction volume.

Colony PCR program:

1.) 95°C – 5min, 2.) 94°C – 30sec, 3.) X°C – 30sec, 4.) 72°C – Xmin, 5.) repeat steps 2.-4. 34 times, 6.) 72°C – 10min, 7.) stored at 4°C ∞

The elongation time of the PCR program was adapted to the expected insert size (plus the length of the multiple cloning site). 5 µl of each PCR product were analyzed using gel electrophoresis.

2.2.11 Plasmid preparation

The plasmid preparation was performed using the Miniprep kit (Qiagen, 27104) or a non-commercial technique. Therefore, 1.5 ml overnight culture was centrifuged for 5 min at 5000 rpm and the supernatant subsequently removed. In the following, three different buffers (P 1-3) were added with an equal amount of 200 µl each. First, alkaline lysis solution I (P 1: 50 mM glucose, 25 mM Tris-Cl pH 8.0, 10 mM EDTA pH 8.0) was added to resuspend the bacterial pellet. Additionally, alkaline lysis solution II (P 2: 0.2 M NaOH, 1% SDS) was added, followed by alkaline lysis solution III (P 3: 5 M potassium acetate, glacial acetic acid). Mixing was done by inverting the tubes 4-6 times between each buffer step. Finally the solution was spun for 10 min at 13,300 rpm and the supernatant decanted to another tube. To precipitate the DNA, 450 µl isopropanol was added, the solution mixed by pipetting and subsequently centrifuged for 10 min at 13,300 rpm; the supernatant was removed. Before resuspending the DNA, the pellet was rinsed with 1 ml of 70% ethanol and spun for 3 min at 13,300 rpm. The remaining pelleted DNA was air dried and then resuspended in 30 µl of Milli-Q. The final DNA concentration of the plasmid was determined by spectrophotometry using NanoDrop (PeqLab).

2.2.12 Restriction enzyme analysis

As an alternative to colony PCR, the restriction enzyme analysis was used to check for the size of the insert. Depending on the vector, a two-time cutter (*Mlu*I) or two one-time cutting enzymes (*Bam*HI, *Eco*RI, *Kpn*I, *Not*I, *Sal*I) were selected for restriction digest. If it was necessary to check which clones were more likely identical, a polycutter like *Hinf*I (NEB) was used. 1 µg of plasmid was used as template. The recipe was as follows: 1.5 µl plasmid; 1 µl enzyme buffer; 0.5 µl enzyme (NEB or Fermentas); add Milli-Q up to 10 µl total reaction volume and incubate for 1.5-2 h at 37°C. The digest was analyzed via gel electrophoresis.

2.2.13 Sequencing

A suitable amount of plasmid prep was sent for sequencing. Sequencing was initially done by the company Agowa (Berlin, Germany) and later by the company Seqlab (Bergen, Norway). For the latter, one sequencing reaction had to be set up and run before sending it to the service. For protocol information see http://seqlab.uib.no/Doc/protocolBigDye.

2.2.14 Bioinformatics

The obtained sequences were downloaded from the server (Agowa and Seqlab, respectively) and cleaned from the multiple cloning site fragments of the vector using BLAST VecScreen (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html). To compare the resulting inserted sequences with genes listed in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/), the insert was queried in the BLAST program. DNA sequences were aligned with the 'Cap contig assembly program' function of BioEdit and checked by eye for ambiguities. Protein sequences of species of interest were obtained from the NCBI database and aligned using MAFFT (E-INS-i) or ClustalX. Domain prediction and further functional characterization were performed using the ExPASy Prosite (Sigrist et al. 2010) and SMART database (http://smart.embl-heidelberg.de/), whereas the prediction of signal peptides was done with the SignaIP 3.0 server (Dyrløv Bendtsen et al. 2004). Analyses for the prediction of transmembrane domains were run using the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). All these selected alignments were used to calculate a bootstrapped ML (maximum likelihood) evolutionary tree using the RAxML program (Stamatakis et al. 2008). All ambiguous positions or long autapomorphic insertions in the alignment were excluded using the BioEdit software. The alignments can be found at the end of the Appendix and the used GenBank or UniProt accessions are listed there as well (Appendix, Tab. A5).

2.3 Immunhistochemistry (IHC)

Two slightly different fixations were used to fix the embryos. The differences in processing were confined to buffer constitution and storage of the embryos. Half of the larvae were fixed for 2 h at room temperature (RT) in 4% Paraformaldehyde (PFA) buffered in phosphate-buffered saline (PBS, pH 7.4) plus 0.1% Tween 20 (PTW) and afterwards stored in 100% methanol (MeOH) at -20°C. This fixation is mainly used for *in-situ* hybridization (ISH) techniques and is listed here, because IHC stainings were also done after ISH. Another way of fixing animals for IHC was obtained by 4% PFA, buffered in phosphate-buffered saline (PBS, pH 7.2) without Tween, for 1 h at 4°C. The fixed material was stored in storage buffer (0.05 M PB/0.3 M NaCl/0.05% NaN_3) at 4°C. For co-localization purposes, the chitinous chaetae were stained with WGA (wheat germ agglutinin) Alexa Fluor 633 (Invitrogen, W21404) in concentrations of 0.1-0.5 µg/µl. The incubation time varied between 15 min to 24 h. The staining was stopped via several washes in PTW. For antibody staining, e.g., after ISH, the embryos were incubated in blocking buffer (0.05 M PB/0.3 M NaCl/0.25% BSA/0.1% Tween/0.05% NaN_3) at RT for 2 h and afterwards rinsed in PBT washing buffer (0.05 M PB/0.3 M NaCl/0.1% Tween). The primary antibody was diluted in incubation buffer (0.05 M PB/0.3 M NaCl/0.1% Tween/0.05% NaN_3) and incubated at 4°C overnight. The incubation was stopped with three washes in PBT at 4°C followed by three washing steps for 2 h at RT. The second antibody was incubated and stopped as described for the primary antibody. Finally, the embryos were rinsed three times for 1 h in PBT at RT and transferred into

storage buffer (0.05 M PB/0.3 M NaCl/0.1% Tween/0.25% DABCO/0.05% NaN_3) or directly into the embedding medium (90% glycerol/1x PBS/0.25% DABCO). The staining was analyzed using confocal laser scanning microscopy (CLSM) (Leica TSC SP2).

2.4 *In-situ* hybridization (ISH)

2.4.1 Probe generation

The selection for the *P. dumerilii* genes was based on a random *in-situ* hybridization screening (D. Arendt lab, unpublished data, EMBL Heidelberg, Germany), whose resulting gene expression patterns were deposited in the *Platynereis* Expression Pattern Database (PEPD base). From this, the candidates exhibiting a gene expression pattern restricted to the chaetal sac region were selected. Afterwards, all selected *P. dumerilii* genes were picked from the existing EST library (pCMV SPORT 6 vector, Invitrogen) in the lab of D. Arendt (EMBL Heidelberg, Germany). All genes for *C. teleta*, *M. cranium* and *M. cirriferum* were amplified from cDNA and cloned into the pGEM-T Easy or pJET1.2blunt vector. For probe generation, the plasmids were linearized using one time cutting enzymes according to their specific cloning vector. After purification of the linearization (Nucleotide Removal Kit, Qiagen), the transcription reaction was prepared. For this, the following components were added in the following order: 1 µg linearized plasmid (max. 12.5 µl), 2 µl 0.1 M DTT (Promega), 1.3 µl NTPs (NEB), 0.7 µl Digoxigenin-UTP (DIG-UTP) (Roche, 11209256910), 0.5 µl RNase inhibitor, 2 µl 10x transcription buffer (Roche), 1 µl RNA-Polymerase (Roche, T7: 10881767001, Sp6: 10810274001) and finally filled up to 20 µl total volume with Milli-Q. The transcription reaction was incubated at 37°C for 4 h in a thermal cycler (VWR, Applied Biosystems Veriti Thermal Cycler). To eliminate the remaining DNA, each sample was incubated with 1 µl DNaseI (NEB) at 37°C for 15 min. All transcription reactions were purified using the RNeasy Mini Kit (protocol: RNA Cleanup; Qiagen, 74104) and subsequently checked on a 1% gel (1% agarose (Sigma) in 1x TAE, stained with ethidiumbromide). All positive probes were diluted in 75 µl Hyb-Mix (50% formamide, 5x standard saline citrate (SSC), 50 µg/ml heparin, 0.1% Tween 20, 5 mg/ml torula RNA) for RNA stabilization and afterwards stored at -20°C.

2.4.2 Whole-mount *in-situ* hybridization (WMISH)

The generally used protocol was adapted after Jékely and Arendt (2007). Larval material for ISH was fixed in 4% PFA in phosphate-buffered saline (PBS, pH 7.4) plus 0.1% Tween 20 (PTW) for 2 h at room temperature while shaking. The fixation was stopped by three washes in 100% MeOH for 30 min each and then stored at -20°C for at least one night. In the first step of *in-situ* hybridization (ISH), the embryos were rehydrated in a 75%/50%/25% MeOH/PTW series, 2 min each, rinsed twice in PTW for 5 min and digested in 100 µg/ml (in the case of *P. dumerilii*) and

10 µg/ml proteinase K (for *M. cranium* and *M. cirriferum*), respectively. In the case of *P. dumerilii*, the incubation time for the digests was adapted according to each developmental stage (45 sec for 40-48 h, 1 min for >48-72 h-stage, 2 min for older worm-shaped juveniles). In consequence of the lower concentration of proteinase K in *M. cranium* and *M. cirriferum*, the digestion time was increased to 2 min for all processed larval stages. To stop the digestion, embryos were rinsed twice in 2 mg/ml glycine/PTW and post-fixed afterwards for 20 min in 4% paraformaldehyde/PTW (pH 7.4). Following fixation, the embryos were washed five times for 5 min in PTW and afterwards pre-hybridized in Hyb-Mix (50% formamide, 5x standard saline citrate [SSC], 50 µg/ml heparin, 0.1% Tween 20, 5 mg/ml torula RNA [Sigma]) at 65°C for 1 h. For probe hybridization, a DIG-11-UTP-labeled RNA probe (final probe dilution of 5-30 µl/200 µl Hyb-Mix) was denatured in Hyb-Mix at 80°C for 10 min and added to the pre-hybridized embryos. Probe hybridization was carried out overnight at 65°C in a water bath. To wash out the probe, embryos were washed twice in 50% formamide/2x SSC/0.1% Tween 20 for 20 min, once in 2x SSC/0.1% Tween 20 for 15 min and twice in 0.2x SSC/0.1% Tween 20 for 20 min. After two quick exchanges of PTW, the embryos were blocked in 5% sheep serum in PTW for 1 h at room temperature and incubated in 1:2000 anti-DIG-antibody (Anti-Digoxigenin-AP, Roche 11093274910) in 2.5% sheep serum/PTW at 4°C overnight. The antibody incubation was stopped by washing the embryos six times in PTW for 10 min each, followed by two quick washes in staining buffer (100 mM Tris-Cl, pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$, 0.1% Tween 20) to equilibrate the embryos. Finally, the embryos were stained with 337.5 µg/ml NBT (Roche, 11383213001) and 175 µg/ml BCIP (Roche, 11383221001) diluted in staining buffer. After several hours the staining was stopped by three washes in PTW. For microscopy, the embryos were mounted in glycerol and analyzed using Nomarski microscopy (Zeiss) and CLSM (Leica CLSM TCS SPE, software LAS AF 1.6.1).

2.4.3 Double fluorescent *in-situ* hybridization (double FISH)

In general, the protocol was similar to the above described standard WMISH protocol, but in double fluorescent *in-situ* hybridizations (double FISH) one probe was labeled with Digoxigenin (DIG) and a second was labeled with Fluorescein (FLUO). The detection of both probes was done separately using the TSA Plus Fluorescence System (Perkin Elmer, NEL741, NEL752001KT). The protocol procedure was performed following Tessmar-Raible et al. (2005) and using anti-DIG- and anti-FLUO-PODs (Anti-Digoxigenin-POD, Roche 11207733910; Anti-Fluorescein-POD, Roche 11426346910). The analysis of the gene expression patterns was done using the CLSM (Leica).

2.4.4 Detection of both fluorescence- and DIG-labeled probes

The general protocol was followed as described above, but according to Tessmar-Raible et al. (2005), it is important to first detect the DIG-labeled probe using the standard WMISH protocol

(2.4.2). In addition, the FLUO-labeled probe was detected using the protocol for FISH (2.4.3). In the case of two probes that were known to differ in their strength of staining, the weaker one was labeled with DIG-UTPs to be detected via NBT/BCIP while the stronger one was combined with FLUO-UTPs for the FISH method. Detailed information can be found in Tessmar-Raible et al. (2005).

2.5 Structural investigations

2.5.1 Transmission electron microscopy (TEM)

Larval stages of *P. dumerilii* (48 hpf) and *M. cranium* were fixed in 2.5% glutaraldehyde (in 0.05 M PB / 0.3 M NaCl; pH 7.2; a few crystals of Rutheniumred) at 4°C for 1 h, rinsed several times in PBS (0.05 M PB / 0.3 M NaCl, pH 7.2) and subsequently stored in storage buffer (0.05 M PB / 0.3 M NaCl / 0.05% NaN_3) at 4°C. Postfixation occurred in 1% OsO_4 in PBS for 20 min. Afterwards the samples were dehydrated in an ascending acetone series and embedded in plastic resin (Araldit M [Fluka 10951]; Araldit M Hardener 964 [Fluka 10953]; BDMA [Benzyldimethylamin, Agar Scientific Ltd. R1062]) via several Propylenoxid-Araldit-mixtures. For hardening, the resin was put in an oven at 60°C for 2 days. Serial sections were cut with a diamond knife (Diatome 45°MT 4526) by using a Reichert ultra-microtome S and automatically stained with Uranyl Acetate (30 min, 2%) and Lead Citrate after Reynolds (1963) at room temperature for 25 min in a Phoenix Ultrastainer (Phoenix staining technologies, Berlin). Analysis of the serial sections was carried out using a Philips CM120 BioTWIN equipped with digital imaging plates (Ditabis).

2.5.2 Scanning electron microscopy (SEM)

All stages were fixed and post-fixed following the protocol for TEM studies. Subsequently, the larvae were rehydrated in a graded ethanol series and critical point dried with a Balzer CPD 030 Critical Point Dryer. Subsequently, the samples were sputtered with gold or gold/palladium in a Balzer 70 Union SCD 040 or a Low Voltage Cool Sputter Coater EMITECH K 550, respectively. Specimens were examined with a Fei Quanta 200 scanning electron microscope (SEM).

2.6 Image processing

Analysis and merging of image stacks was done using the programs ImageJ (v1.44i), Adobe Photoshop CS5 and Adobe Illustrator CS5. Alignments of the ultrastructural sections for further 3D reconstructions was performed using Adobe Illustrator CS5, whereas the final 3D model was designed using 3D Studio Max.

3 Results

3.1 Chaetogenesis in *Platynereis dumerilii*

Chitinous chaetae are very prominent in annelids and exhibit a variety of shapes, e.g., the compound type of chaetae in *P. dumerilii* (Fig. 1 A+B). The presence of such bizarre structures that are identical within and between conspecific organisms, attracts wide interest in the formation of chaetae (chaetogenesis), as a strict genetic control can be suggested to guide the formation process. Here, chaetae are studied on different levels, combining data from ultrastructure, immunohistochemistry and gene expression, to reveal new insights into chaetogenesis.

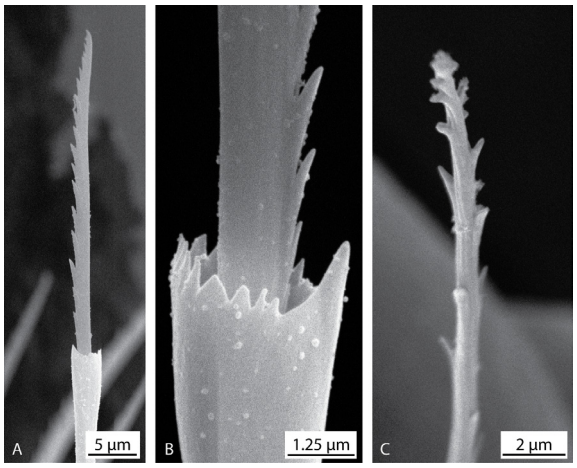


Fig. 1: Different types of chaetae – *Platynereis dumerilii* vs. *Macandrevia cranium*. A+B – Compound-type annelid chaeta of a 77 hpf old *P. dumerilii* larva. C – Larval brachiopod chaeta (lacking a coating layer) of *M. cranium* (132 hpf).

For that purpose I first screened an EST library of *Platynereis dumerilii* (D. Arendt, unpublished data) searching for genes expressed in the chaetal sac region and investigated their expression in different developmental stages. Some of these genes have previously been shown to be expressed in 48-hpf stages (for these 7 candidate genes of my initial gene expression studies see the diploma thesis of Zakrzewski [2007]: "Strukturelle und molekulare Charakterisierung von Chaetoblasten bei *Platynereis dumerilii* [Polychaeta]", translation: "Structural and molecular characterization of chaetoblasts in *Platynereis dumerilii* [Polychaeta]"). Altogether, in the present study I finally localized 15 candidate genes to be exclusively expressed in chaetae forming cells in several developmental stages. For the purpose of bioinformatic analyses I first sequenced the complete insert of all 15 clones, yielding a fragment size range between 0.5 kb and 2.5 kb. BLAST searches performed against GenBank's amino acid sequence database revealed that some of the candidate genes are orthologous to already published genes, whereas many of the genes seem to be completely unknown yet. Nonetheless, all these genes and their expression specificity are of special interest for they represent excellent molecular markers in the process of chaetae formation. As the low number of significant BLAST results may be due to the relative short sequence lengths, Perl script-based computational studies against a 454-transcriptome data library of *P. dumerilii* were performed to elongate the existing sequences. The general idea of this strategy was to BLAST search the existing query sequence against the transcriptome and check for sequence hits that were able to elongate the existing sequence. If a sequence could be prolonged, the new sequence was always used for another round of BLAST searches and elongation. This iterating process was done until no further prolonging hits were found. As a result, the computational approach yielded longer sequences

for four of the fifteen candidate genes. Upcoming transcriptome and genome data of *P. dumerilii* were assumed to yield another valuable source for sequence elongation, but the sequence coverage was still too low and assembly was not yet sufficient for these approaches. As in the context of this study the functional sequence characterization is of additional importance, analyses of the amino acid sequence were performed for each gene. Therefore, the longest open reading frame was selected for each of the genes and the corresponding protein sequence was used as a query in different databases. For domain prediction and further functional characterization, database searches were run using SMART and ExPASy, as well as the SignalP 3.0 Server and the TMHMM Server v. 2.0. Additional further analyses were performed for two of the candidate genes.

3.2 Candidate genes for chaetae formation in *P. dumerilii*

3.2.1 Tissue specificity and functional characterization

To check the specificity and activity of the selected candidate genes, I screened their expression in several developmental stages, which will be illustrated in the following. As the internal morphology of *P. dumerilii* differs between the developmental stage of a larva and an already worm-shaped juvenile, a short schematic overview (Fig. 2) of the larval ontogeny is shown to ease the interpretation of the further light microscopic pictures. For the ISH screen I chose three different developmental larval stages – 40 hpf, 48 hpf, 72 hpf – and a juvenile stage with at least ten distinct segments. In 40-hpf (Fig. 2 A) and 48-hpf stages (Fig. 2 B), the larva is more or less spherical, whereas in 72 hpf (Fig. 2 C) it exhibits a pear shape. While in the latter stage chaetae are visible from the outside (Fig. 2 C), in both early stages, chaetae do not protrude the epidermal layer. Nevertheless, even in these two early stages chaetae development had already started, as light microscopic studies show. In the earliest stage, two chaetal sacs are differentiated, but only the chaetal sacs of the first segment already possess chaetae (Fig. 2 A). In 48-hpf stages, the chaetae are more prominent (even in all segments) and close to "piercing" through the body wall (Fig. 2 B).

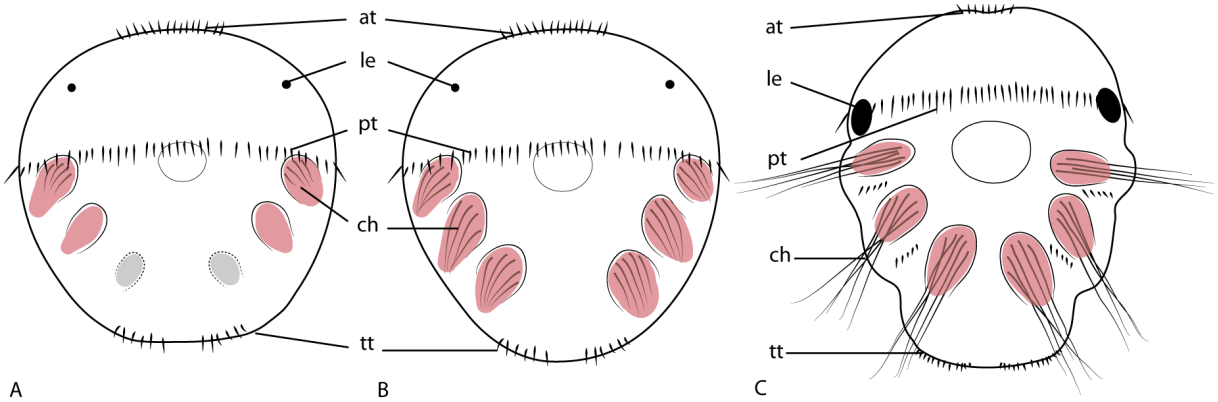


Fig. 2: Schematic drawing of *Platynereis dumerilii* larvae. A – 40-hpf stage with short chaetae already developed within the chaetal sacs of the first setiger, whereas the anlage of the second setiger lacks chaetae. A third posterior located setiger is not yet structurally characterized. B – 48-hpf stage. C – 72-hpf stage. at – apical tuft; ch – chaetae; le – larval eyes; pt – prototroch; red – chaetal sacs, consisting of follicle cells and more basal located chaetoblasts; tt – telotroch.

Pd_CS1

During ISH studies it became evident that the gene *Pd_CS1* is highly expressed in the chaetal sac cells, especially in early stages of chaetogenesis, and that it is restricted solely to this specific region (Fig. 3). Already in 40-hpf stages, both the first setiger as well as the anlage for the second setiger feature strong expression patterns (Fig. 3 A). As *Pd_CS1* is also expressed in the second, still chaetae-lacking segment, this gene might be one of the first-activated genes being involved in the process of chaetae formation. Even in 48-hpf embryos, each of the three setigers features strong ISH signals (Fig. 3 B), which are proximally located within the chaetal sacs. Stainings in 72-hpf old larvae exhibit a pattern that is similar to 40- and 48-hpf stages (Fig. 3 C). The ISH signals are restricted to the proximal part of the chaetal sacs, indicating that the cellular localization of gene expression presumably is restricted to the basal cells. Light microscopic investigations of older, juvenile stages (Fig. 3 D) using a higher magnification (600-1000x) reveal gene expression patterns at the base of the acicula (Fig. 3 E-H), internal chitinous support structures of the parapodia that are formed in a way comparable to the hitherto characterized larval chaetae. As previously depicted for other genes, posterior (and thus younger) segments present stronger expression patterns than the more anterior ones. As this gene seems to have a broad expression range, I was interested in its molecular and functional characterization. The original clone sequence of *Pd_CS1* had a length of 1318 bp, whereas the mentioned computational approaches were not able to elongate the fragment. However, as this gene appeared to be strongly expressed during chaetae formation, further experiments were performed to reveal more sequence information (PCR-based approach; see chapter 3.4.1). BLAST searches showed that *Pd_CS1* is highly similar to known chitin synthases and database searches for functional characterization revealed that the amino acid sequence of *Pd_CS1* possesses 12 transmembrane domains and a chitin synthase domain. Thus, *Pd_CS1* clearly is a chitin synthase and of special interest in the context of formation of chitinous chaetae. The evolution of chitin synthases will be analysed later (see chapter 3.4).

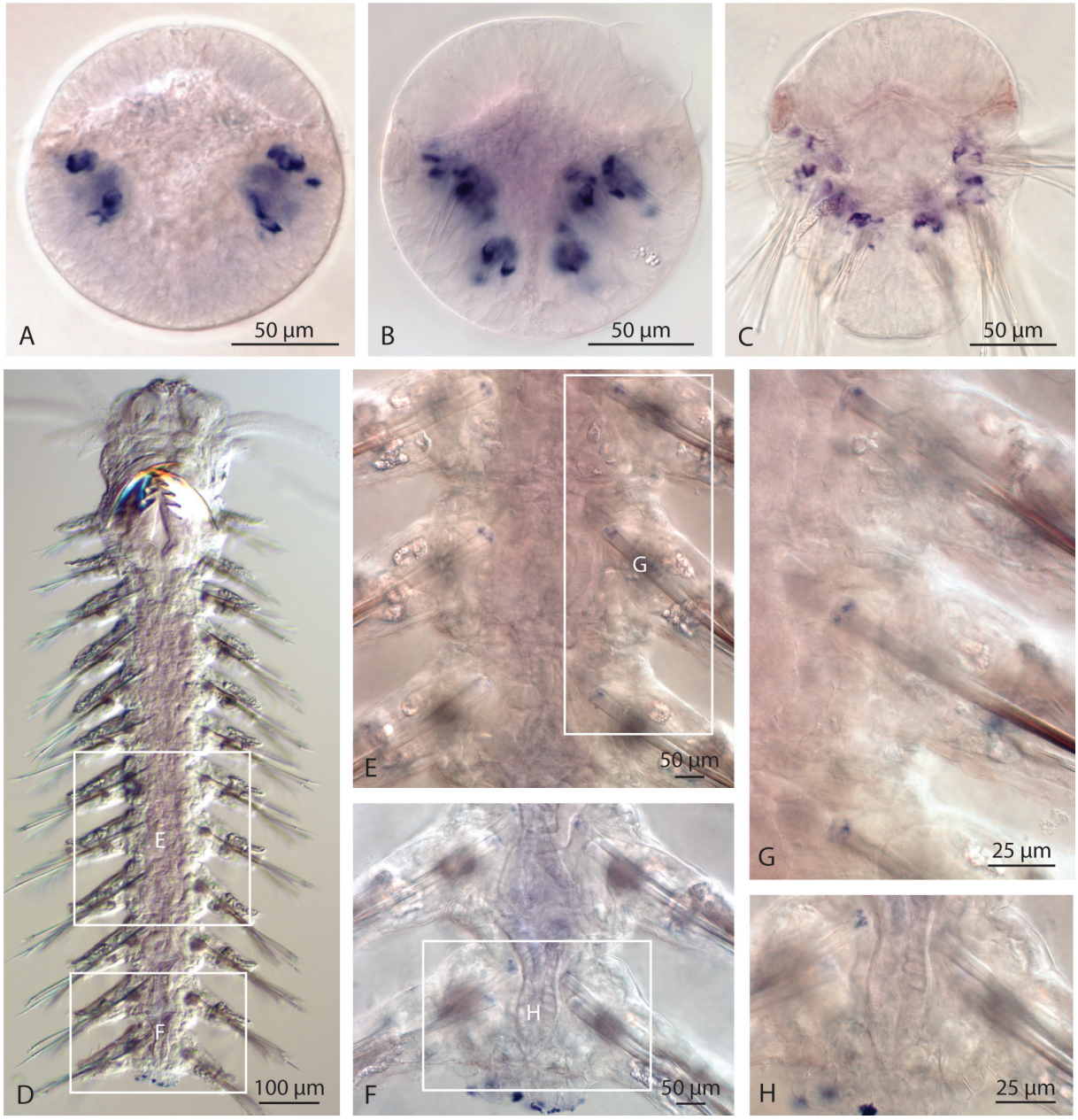


Fig. 3: ISH with *Pd_CS1* in *Platynereis dumerilii*. A – In the 40-hpf stage, gene expression is present in the first setiger and the anlage of the second setiger. As the staining indicated to be strong, it was stopped after 3 h. B – Expression in 48 hpf old larvae is present in all three setigers and their chaetal sacs. The staining, that was stopped after 2.5 h, is restricted to the proximal-most cell layers. C – In 72 hpf old larvae, each chaetal sac of each setiger exhibits a prominent staining, that was stopped after 2.5 h of incubation. D-H – ISH signals in a juvenile stage (staining stopped after 3 h). E+F – Close-up of the middle (E) and posterior segments with evidence for gene expression. G+H – The detailed view of single aciculae within the parapodia reveals a faint staining at the proximal part of the chaeta.

Pd_MLCK1

ISH with *Pd_MLCK1* turned out to be difficult, because the NBT/BCIP stainings developed weakly and are difficult to discern from the background signal (Fig. 4). Nevertheless, first ISH signals are even present in 40-hpf old embryos, where first setiger and the anlage of the second one express a weak, but distinct pattern (Fig. 4 A). In 48-hpf stages, the staining in both setigers is enhanced,

resulting in stronger ISH signals located in the proximal part of the chaetal sacs (Fig. 4 B). In the youngest setiger, the gene expression patterns show a faint staining comparable with the situation in 40-hpf stages described above. After 72 h of development, a transition of gene expression can be observed, as stainings are present in the second and third setiger, but absent in the first and oldest one (Fig. 4 C). In contrast to the investigated younger stages, ISH signals are also present in epidermal cell layers at the posterior rim of the first and second chaetae bearing segment. It is not clear whether these are specific stainings or artefacts. Nevertheless, these putatively chaetogenesis-

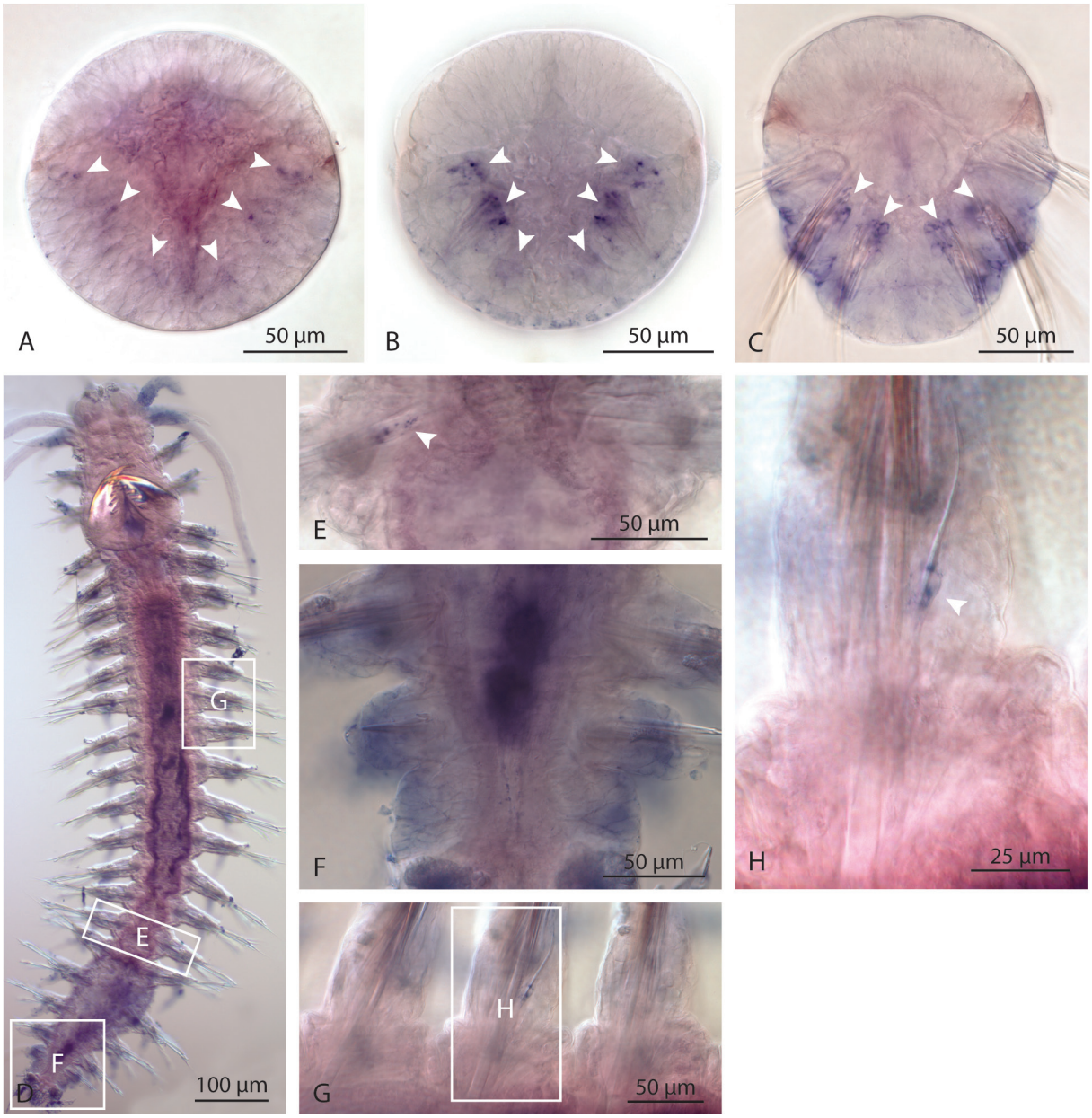


Fig. 4: ISH with *Pd_MLCK1* in *Platynereis dumerilii*. Specific ISH signals are indicated by arrows. A – ISH studies in 40-hpf-stages (staining stopped after 3 h) reveal an expression not only restricted to the first setiger, but also to both anlagen of the following setiger, where a structural localization of the chaetal sacs is not yet possible. B – In 48-hpf stages, ISH pattern are present in first anterior-most setigers. The staining, that is restricted to the proximal part of the chaetal sacs only, was stopped after 3 h. C – In 72-hpf stages the staining gets blurry. It had to be stopped after 1 d to avoid strong background. Despite, a NBT/BCIP precipitation is present surrounding the proximal part of the chaetal sacs. D-H – ISH signals are present in several setigers of the juvenile stage. E-H – Detailed view of selected setigers (E+F) and their parapodia (G+H) showing chaetae-specific gene expression patterns (staining stopped after 3 h).

unrelated patterns are solely present in 72-hpf stages, for investigated older stages (Fig. 4 D) lack signals in the corresponding regions, but show the typical expression patterns along the chaetae (Fig. 4 E-H). The sequencing analysis of the *Pd_MLCK1* clone yielded a fragment size of 2188 bp, whereas further elongation approaches obtained 2787 bp of total nucleotide sequence information. BLAST searches showed that *Pd_MLCK1* is highly similar to myosin light-chain kinases (MLCKs). Further characterization of *Pd_MLCK1* is subject of a separate chapter (see chapter 3.5).

Pd_CAML

The analysis of the ISH signals showed that *Pd_CAML* is expressed in all screened stages (Fig. 5), although the NBT/BCIP staining appears to be weak (Fig. 5 A-C). First evidence for gene expression was obtained in 40-hpf stages, where only the first setiger shows an ISH signal restricted to the chaetal sac region (Fig. 5 A), whereas no expression pattern can be detected in the second one. The staining within the chaetal sacs of the first setiger is weak and difficult to discern from the background (Fig. 5 A). Twelve hours later (48 hpf), the ISH signals are not only restricted to the first but also second chaetae-bearing segment (Fig. 5 B). This "step-by-step" activation of *Pd_CAML* is completed in 72-hpf stages, as all setigers express the gene, indicated by stainings in the chaetal sacs of all segments (Fig. 5 C). In comparison with younger stages the chaetal sac cells of 72-hpf show a stronger staining, although it is still difficult to localize the expression on the cellular level. Likewise, ISH analysis of juvenile stages revealed that some setigers exhibit ISH signals, but with low signal intensity (Fig. 5 D-J). In the context of molecular characterization, computational approaches were not capable to obtain more nucleotide sequence information than the existing 1059 bp. Nonetheless, database searches indicate that *Pd_CAML* is a calmodulin-like (CAM-like) protein. Alignment analysis with representatives of vertebrate and invertebrate calmodulins showed that there is sequence similarity to "standard" calmodulins (these exhibit hardly any sequence differences between vertebrates and invertebrates), with highest similarity in the C-terminal region. Additional support for a common evolutionary origin with calmodulins is yielded by the existence of four EF-hand domains that are suggested to be calcium-binding motifs. The evolution of such EF-hand proteins will be analysed later (see chapter 3.6).

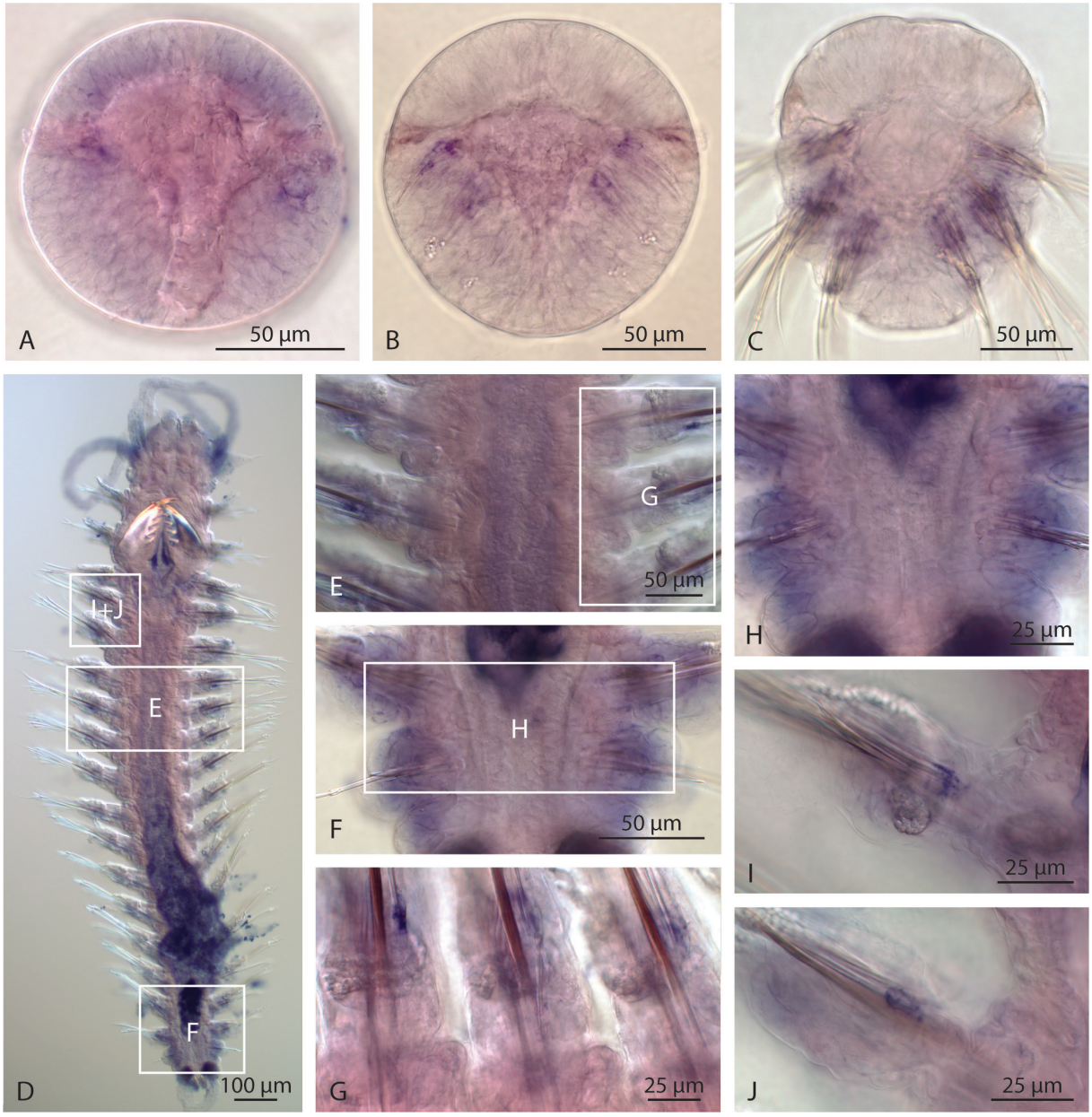


Fig. 5: ISH with *Pd_CAML* in *Platynereis dumerilii*. A – In 40-hpf stages, a faint expression is present in the chaetal sac region of the first developing setiger. B – Only first and second setiger of 48 hpf old larvae exhibit a chaetae-specific pattern that is more restricted to the proximal part of the chaetal sacs. C – All three setigers of 72-hpf stages show an expression in the chaetal sacs that can be distinguished from the remaining tissue. D-J – Specific, but weak expression is present in nearly all segments along the body length of juveniles. E/F/H – Close-up of selected setigers with specific ISH signals. F+H – Detailed view of the posterior-most segment (pygidium) showing extremely weak ISH signals. G/I/J – Details of the parapodia expressing *Pd_CAML*.

Pd_fcmg1

Investigation of the expression of *Pd_fcmg1* show that this gene is not expressed in 40-hpf stages (Fig. 6 A), whereas in 48 hpf (Fig. 6 B), 72 hpf (Fig. 6C) and juveniles (Fig. 6 D-G), gene expression can be detected. Additional ISHs on 42-hpf stages (data not shown) show the same lack of expression as stated for 40-hpf stages. Therefore *Pd_fcmg1* is suggested to act as one of the later activated genes in the process of chaetogenesis. The analysis of all data obtained by ISH

experiments show the strongest local gene expression in 48-hpf stages. Here all chaetal sacs of all three segments feature strong gene expression. Closer microscopic analysis revealed that, if present, in all stages the gene is expressed by cells guiding the proximal region of a single chaeta. Sequencing and additional elongation approaches of the clone of *Pd_fcmg1* obtained a nucleotide sequence of about 1427 bp length. BLAST searches against public sequence databases revealed no clear orthologs. Database searches for domain prediction and further functional characterization revealed evidence for a signal peptide.

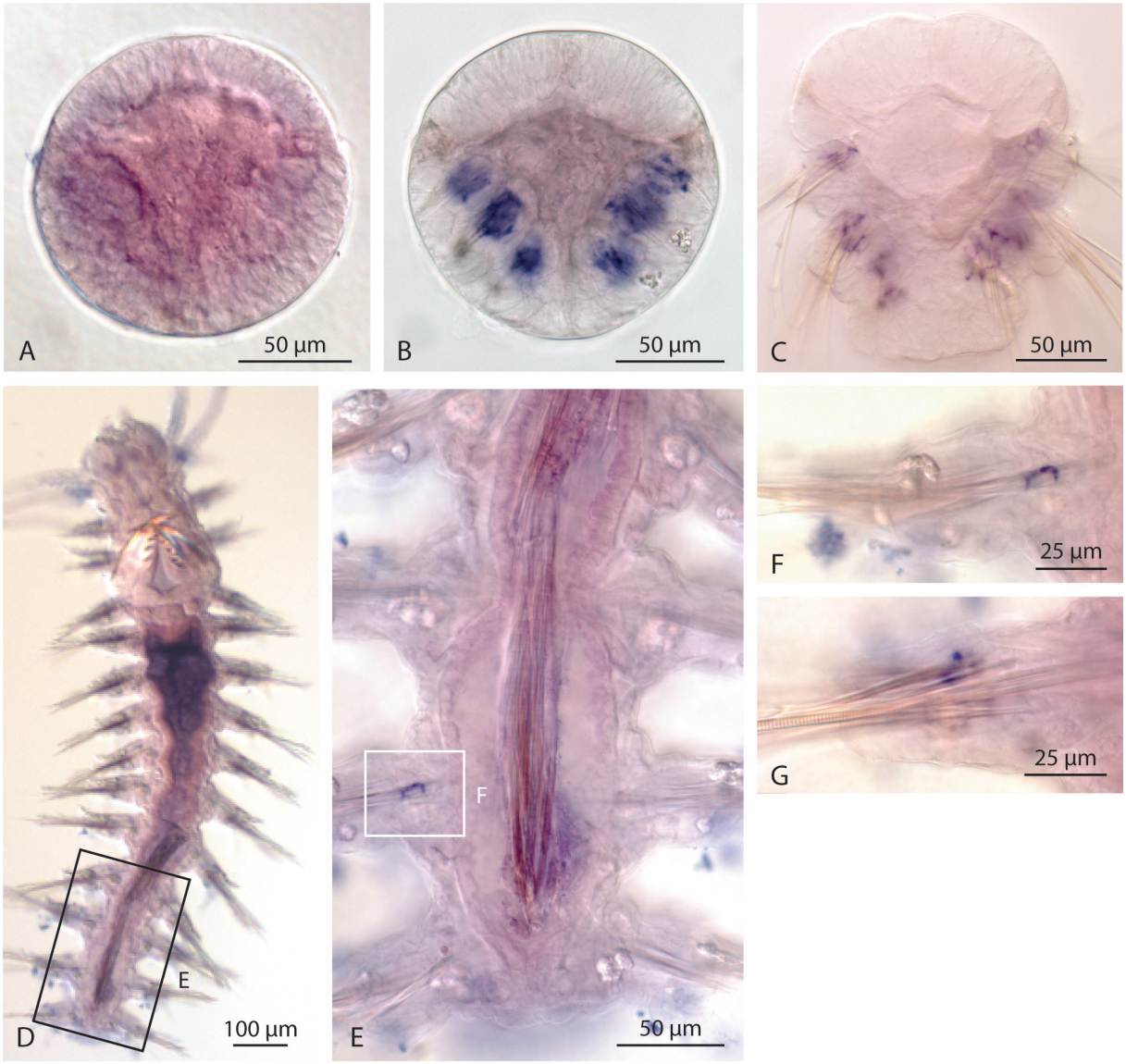


Fig. 6: ISH with *Pd_fcmg1* in *Platynereis dumerilii*. A – 40-hpf stage showing no expression of *Pd_fcmg1*. B-G – Blue-violet staining indicates gene expression. B – 48 hpf old larval stage (dorsal view) with ISH signals restricted to the proximal cell layers of all chaetal sacs. C – All setigers show proximally chaetal sac-specific ISH signals in 72-hpf stage (ventral view). D-G – In juveniles (ventral view), *Pd_fcmg1* is mainly expressed in the posterior-most segments. E – Last four segments of the posterior end of a juvenile. One expression signal is highlighted. F+G – Detailed view of one parapodium showing an ISH signal more at the proximal part of a chaeta.

Pd_fcmg2

All investigated stages older than 36 hpf, which were hybridized with an antisense probe of *Pd_fcmg2* show a more or less intense gene expression restricted exclusively to the chaetal sac region (Fig. 7). While in 36-hpf stages no gene expression can be detected (data not shown), in 40-hpf stages both setigers (i.e., segments exhibiting chaetae) feature distinct expression patterns (Fig. 7 A). Despite that there is clear evidence for the activity of this gene in this stage, it seems to be only weakly expressed, as the staining had to be applied over a duration of 1.5 days (d). In 42-hpf stages (data not shown), gene expression is detected after 1 d, but the stainings are still faint. In contrast, 48 hpf old larvae (Fig. 7 B) exhibit strong staining patterns (staining stopped after 2.5 h), suggesting that at this time of chaetae formation, *Pd_csmg2* is highly expressed in chaetal sacs. When comparing older stages (Fig. 7 C-I) with the 48-hpf stage, the signal strength of the ISH appears to have decreased during further development (72-hpf stage stopped after 4 h; juvenile stage after 3 h). This decrease in gene expression may be due to the fact that each chaetal sac features already prominent cheatae, which may not grow much further (Fig. 7 C-I). Accordingly, in juveniles only the chaetal sacs of the youngest segments (the posterior ones), where many new chaetae are formed, exhibit detectable ISH staining (Fig. 7 D-I). Although *Pd_fcmg2* is partly strongly expressed, the cellular localization of *Pd_fcmg2* expression is difficult. Mainly the ISH data of older stages indicate that this gene is a marker exclusively expressed in proximally situated chaetal sac cells, but not the basal most ones (Fig. 7 B, C, F). On the molecular level, *Pd_fcmg2* represents a clone with 2554 bp nucleotide sequence information. However, based on this fragment size, no similarity to already published sequences was found. Nevertheless, investigations on amino acid level revealed a PAN domain. This kind of domain has significant functional versatility, as they fulfil diverse biological functions by mediating protein-protein or protein-carbohydrate interactions. Additionally, it became evident that the amino acid sequence of *Pd_fcmg2* contains a signal peptide, which is of importance in the context of intercellular trafficking. BLAST searches yielded strong similarities of the C-terminal region of *Pd_fcmg2* to certain EST sequences of other annelids (*Lumbricus rubellus*, *C. teleta* and *M. cirriferum*) including eight conserved cysteine sites.

Pd_cbmg3

Pd_cbmg3 features relatively strong expressions in all screened stages (Fig. 8). Even in stages fixed at 40 hpf, a very prominent expression (staining stopped after 3 h) is present solely within the chaetal sacs of the first setiger and its anlage in the second segment (Fig. 8 A), although only the first setiger exhibits visible chaetal structures (Fig. 8 A detail). ISH on 36-hpf stages (data not shown) reveal that *Pd_cbmg3* is not expressed at this time point, similar to the situation in the presumptive anlage of the third setiger of 40-hpf stages. In 48-hpf stages, where three distinguishable segments are established, all of them feature strong gene expression (Fig. 8 B). The analysis of 72-hpf stages, however, reveals a transition in gene expression, because merely setigers two and three exhibit ISH

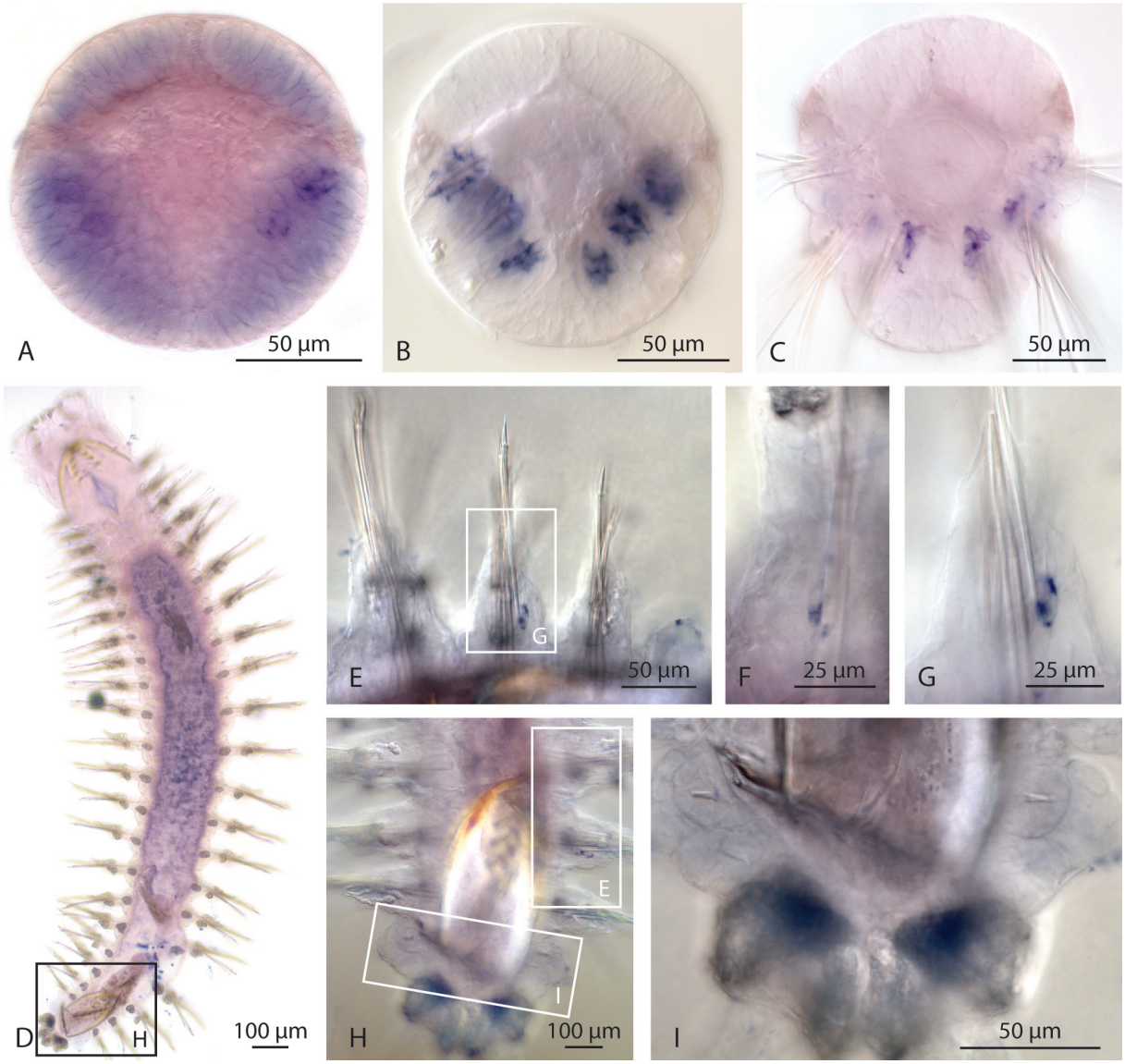


Fig. 7: ISH with *Pd_fcmg2* in *Platynereis dumerilii*. A – The ISH signal in a 40 hpf old larva is restricted to the chaetal sacs of the first setiger and the anlage of the second setiger. The faint blue color in the background is caused by unspecific binding. B – In the 48-hpf stage, the majority of the chaetal sacs cells yielded a prominent ISH staining, showing no expression solely within the distally located cells. C – All three setigers show an expression (not all in one focus plane) exclusively restricted to the chaetal sacs (ventral view). D-I – Expression in juveniles of *P. dumerilii* (ventral view). D – Specific expression signals are mainly present in the posterior-most segments (highlighted). E-G – Detailed view of ISH signal in selected parapodia, showing cellular expression. H – Posterior segments including the pygidium (youngest segment is highlighted). I – Youngest segment with the anlage of chaetal sacs that show the first tips of chaetae.

stainings (Fig. 8 C). This shift of expression to the more posterior (and younger) segments is also present in the juvenile stage (Fig. 8 D-J), where only the last seven segments show gene expression. Concerning the localization of gene expression for *Pd_cbmg3*, light microscopic investigations reveal that, in all screened stages, this gene is exclusively expressed in the proximal-most cell layer. Concerning the molecular characterization of that gene, I obtained the full insert information of about 893 nucleotides, whose length was elongated to 992 bp due to computational tools against transcriptome data. However, BLAST searches revealed no similarity to any specific known genes, but to uncharacterized EST sequences of other annelids (*Alvinella pompejana*, *Lanice conchilega*, *C. teleta* and *M. cirriferum*) that share ten conserved cysteine sites within the region of sequence similarity.

Additionally, database searches for functional characterization indicate that *Pd_cbmg3* contains a signal peptide, which is known to act in cellular trafficking. Additionally, algorithms used for the prediction of (secondary) protein structure identified at least one transmembrane domain.

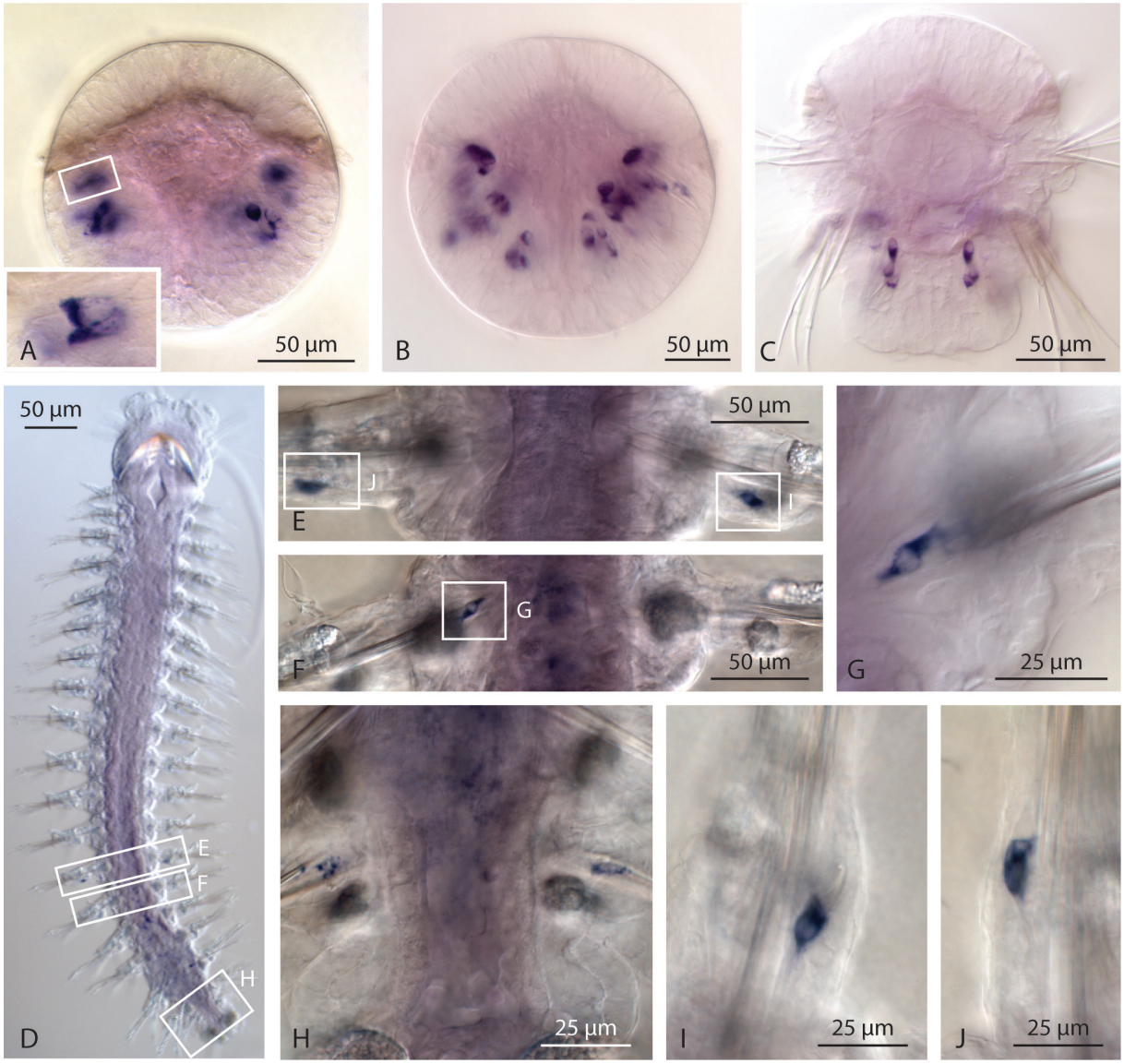


Fig. 8: ISH with *Pd_cbmg3* in *Platynereis dumerilii*. A – Strong chaetal sac-specific expression in 40-hpf stages. Staining was stopped after 3 h. B – In 48 hpf old larvae, prominent ISH signals are present in the proximal-most part of the chaetal sacs of all three setigers. The NBT/BCIP incubation was stopped after 2.5 h staining. C – Ventral view of a 72-hpf stage with expression solely restricted to the chaetal sacs of setigers two and three (second setiger out of focus). Here, the staining was stopped after 4 h of incubation. D – ISH pattern in a juvenile of *P. dumerilii* (staining stopped after 3 h). E/F/H – Detailed view of ISH signals per segment. G/I/J – Detailed view of expression of *Pd_cbmg3* within single parapodia along a young chaeta.

Pd_fcmg4

The expression analysis of *Pd_fcmg4* suggests that this gene is not expressed in 36-hpf stages (data not shown), 40-hpf stages (Fig. 9 A) and the juvenile stage (Fig. 9 D-H), whereas very strong ISH signals are obtained in 48-hpf stages (stopped after 2.5 h of staining; Fig. 9 B) as well as in 72-

hpf stages (stopped after 1.5 h; Fig. 9 C). Thus, concerning chaetae formation, *Pd_fcmg4* can be assumed to act as one of the later activated genes. However, it seems likely that a detectable amount of transcript is present only within a narrow time frame, as ISH signals are absent in all segments of the juvenile stage (Fig. 9 D-H). Nevertheless, the ISH investigations show that in stages expressing *Pd_fcmg4*, the staining is always restricted to cells of the middle and upper layers of the chaetal sac (Fig. 9 B+C). As I am interested in the products of the chaetal genes, I performed domain predictions and further functional characterization studies using the 784 bp sequence information of the *Pd_fcmg4* clone as template. Elongation approaches based on 454-transcriptome data yielded no longer sequences and BLAST searches against public sequence databases revealed no clear orthologs. Within the EST library of *M. cirriferum*, one sequence was found, which in addition to overall sequence similarity share conserved cysteine sites. Furthermore, investigations on amino acid level suggest that *Pd_fcmg4* contains a signal peptide, which affects protein targeting and post-processing.

Fig. 9: ISH with *Pd_fcmg4* in *Platynereis dumerilii*. A – In 40-hpf stages, the specific candidate gene is not yet expressed. B – All three setigers of 48 hpf old larvae show a strong expression, exclusively in the chaetal sacs. C – In 72-hpf stages, setigers one to three show a strong ISH signal (noto- and neuropodium are not in the same focus plane and therefore some ISH signals are out of focus) that is restricted to the proximal part of the chaetal sacs only. D-H – ISH studies in juveniles of *P. dumerilii* yielded no evidence for specific gene expression in the chaetae region.

Pd_fcmg5

I performed ISH for four developmental stages, whose analysis revealed that ISH signals are present in all stages (Fig. 10 A-J). ISH studies using a mixed batch of 40-hpf and 40.5-hpf embryos showed that half of the embryos exhibited an ISH signal, whereas the other half did not. These observations suggest that a detectable expression of *Pd_fcmg5* is reached during the time between 40 and 40.5 h of development post fertilization. Additionally, in those embryos that show an ISH signal (Fig. 10 A) the expression is stronger in setiger one and nearly absent in the less developed and chaetae-lacking anlage of the second setiger. 48-hpf stages (Fig. 10 B) show the same situation,

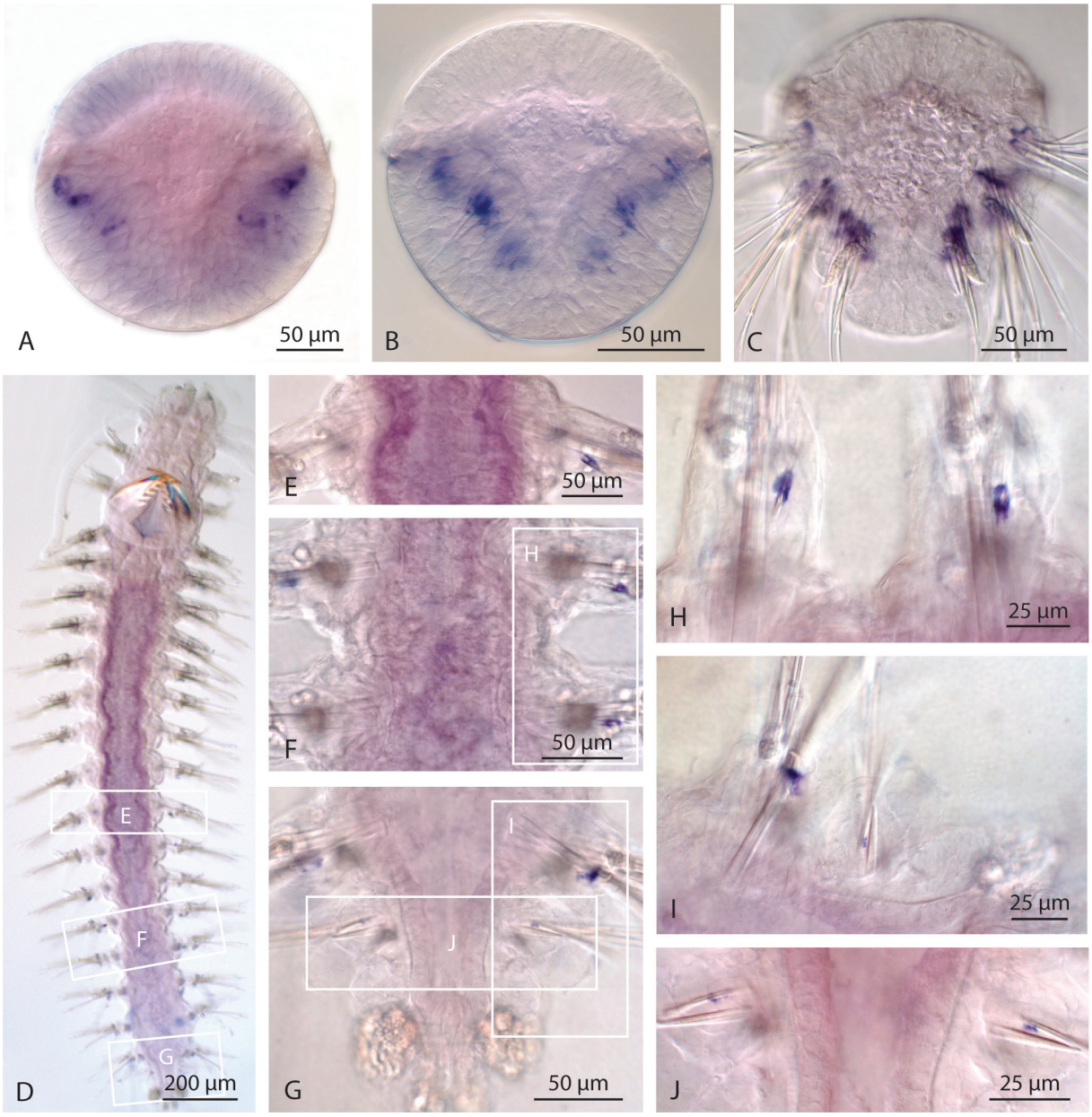


Fig. 10: ISH with *Pd_fcmg5* in *Platynereis dumerilii*. A – 40/40.5 hpf old larval stage, showing ISH signals restricted to the chaetal sacs of the first setiger and the anlage of the second (chaetae-lacking) one. B+C – In 48-hpf (B) and 72-hpf stages (C), the gene *Pd_fcmg5* is expressed in all three setigers, where it marks the mid-region of each chaetal sac. D-J – Gene expression patterns in segments of the juvenile stage. E-G – Close-up of setigers exhibiting ISH signals of the mid- and posterior body region. H-J – Selected single parapodia that show young, growing chaetae and an expression surrounding part of the chaeta. J – Youngest setiger with weak but present expression of *Pd_fcmg5* in the newly developed chaetae.

as strong ISH signals occur in the first and second setigers, whereas weak ones can be observed in the third and youngest one. Even in 72-hpf stages (Fig. 10 C) this shift of signal intensity is evident, but in the opposite way. Here, *Pd_fcmg5* exhibits a strong expression in setigers two and three, but a weak staining in the first and oldest one. The investigations of this ISH show that in chaetal sacs with young and growing chaetae, *Pd_fcmg5* is expressed, as even in the juvenile stage (Fig. 10 D-J) each setiger with short setae along the complete body length (Fig. 10 H-I) – and especially the youngest posterior-most segment (Fig. 10 J) – exhibits a staining. In conclusion, the ISH investigations have revealed that *Pd_fcmg5* consistently appears to be a strong marker of the chaetal sac mid region. Certainly it would be very interesting to know more about the functional aspect of that gene. General BLAST searches against public protein and nucleotide data bases, based on the existing clone sequence (1057 bp length), yielded no significant similarity hits. BLAST searches against EST databases provided hits in two other annelids (*Ridgeia piscesae* and *M. cirriferum*) that possess twelve conserved cysteine positions (see Appendix, Fig. A1). Additionally, further investigations on amino acid level indicate that this gene contains at least one transmembrane domain.

Pd_csmg10

All larval stages that were hybridized with an antisense probe of *Pd_csmg10* show a predominantly strong gene expression, which is exclusively restricted to the chaetal sac region (Fig. 11). While the same probe concentration was used for all screened stages, the first stainings appeared in juvenile stages already after 1.5 h (Fig. 11 D-I) and the last stainings were stopped after 3 h. The ISH signals are mainly present in the mid body region, but completely absent in the last six segments (Fig. 11 D). High resolution analysis (600-1000x) shows an unambiguous cellular localization (Fig. 11 E-I) that is primarily restricted to young chaetae stages, although no signal is visible in the chaetal sacs of the youngest posterior-most segment. During the ISH analysis of the early developmental stages, it became evident that the expression of *Pd_csmg10* is strongest in 48- (Fig. 11 B) and 72-hpf stages (Fig. 11 C), as the developing staining was stopped after 4 and 3 h, respectively. In both cases, the ISH signals are exclusively restricted to cells in the proximal part of the chaetal sac. In contrast, *Pd_csmg10* expression appears to be less prominent in the youngest stage (40 hpf), because the NBT/BCIP staining had to be applied over a period of 1.5 d (Fig. 11 A). The sequencing analysis of the *Pd_csmg10* clone yielded a gene fragment length of about 1070 bp. Approaches with the aim to elongate the existing information via searches within a 454-transcriptome library revealed no further sequence information. BLAST searches against public sequence databases revealed no clear homologs. Additionally, database searches for functional characterization revealed no evidence for a signal peptide within the amino acid sequence, though it seemingly has at least one transmembrane domain.

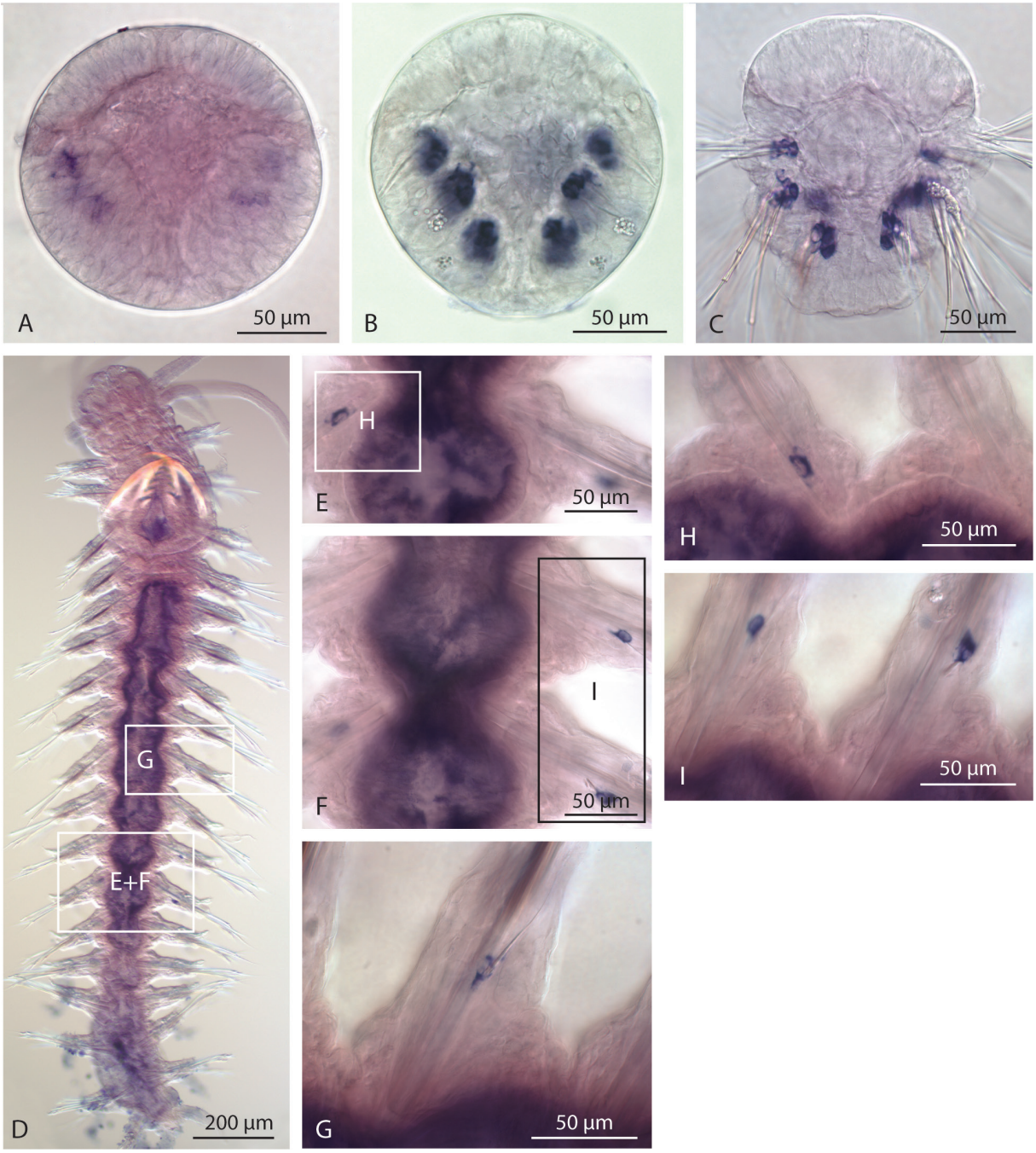


Fig. 11: ISH with *Pd_csmg10* in *Platynereis dumerilii*. The analysis shows that in all stages the gene expression is present and restricted to the chaetae-specific tissue only. A – In 40-hpf stages, the staining to detect the expression developed over a period of 1.5 d. Both, first setiger and the anlage of its second setiger show a faint, but present ISH signal. B+C – The developing NBT/BCIP-staining was stopped after 4 h (B – 48-hpf stage) and 3 h (C – 72-hpf stage) of incubation, respectively. All three setigers exhibit an intense expression in the proximal part of the chaetal sac. Because each setiger consists of a noto- and neuropodium, in 72-hpf stages, the expression is not always located in the same focal plane. D-I – In juveniles of *P. dumerilii*, the gene expression pattern developed after 3 h of NBT/BCIP incubation and shows a broad presence in nearly all setigers, but complete absence in the last six segments. E-I – Detailed view of selected setigers (E+F) and their parapodia (G-I) showing a strong expression at the proximal part of a chaeta.

Pd_fcmg11

ISH investigations with probes against *Pd_fcmg11* show that this gene is expressed in early developmental stages (Fig. 12 A-C), but not in older ones (Fig. 12 D). The latter stages were stained over a period of 1.5 d with several exchanges of the staining solution, but no signal was obtained (Fig. 12 D-G). However, first stainings are present in 40-h old larvae (Fig. 12 A), whereas 36-hpf stages exhibit no ISH signals (data not shown). In 40-hpf stages, the chaetal sacs of the first setiger and the anlage of the second one feature more or less intense gene expression. Even stronger ISH signals are present in 48-hpf stages, where the chaetal sacs of all three segments feature an intense staining (Fig. 12 B). Here, the expression is restricted to the more proximally situated cells of the chaetal sac. In 72-hpf stages I found the same situation of intense expression patterns being present in all three setigers (Fig. 12 C). Computational approaches using public database searches or

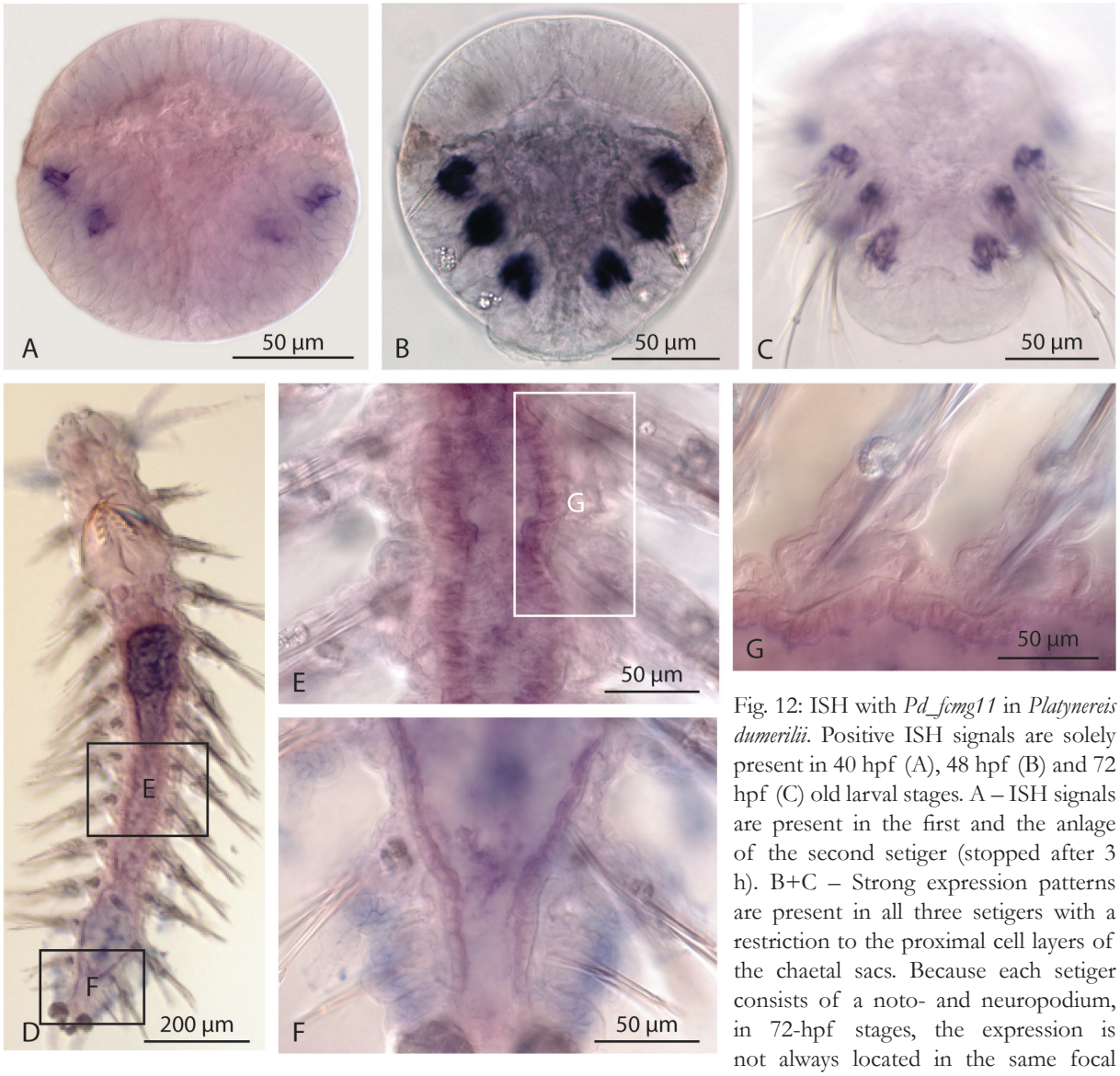


Fig. 12: ISH with *Pd_fcmg11* in *Platynereis dumerilii*. Positive ISH signals are solely present in 40 hpf (A), 48 hpf (B) and 72 hpf (C) old larval stages. A – ISH signals are present in the first and the anlage of the second setiger (stopped after 3 h). B+C – Strong expression patterns are present in all three setigers with a restriction to the proximal cell layers of the chaetal sacs. Because each setiger consists of a noto- and neuropodium, in 72-hpf stages, the expression is not always located in the same focal plane. D-G – Despite an NBT/BCIP incubation of about 1.5 d, no indication for an expression of *Pd_fcmg11* in setigers of the juvenile stage was found.

the *P. dumerilii* genome/transcriptome database yielded neither similarity to already published genes nor further sequence information to elongate the known 801 bp sequence of *Pd_fcmg11*. BLAST searches against public sequence (EST) databases revealed one similar sequence in *M. cirriferum* sharing twelve conserved cysteine sites (see Appendix, Fig. A1). Database searches for functional characterization obtained evidence that *Pd_fcmg11* contains a signal peptide – a signal sequence that guides proteins from the cytosol to certain organelles and to the cell membrane, respectively.

Pd_fcmg13

Based on ISH investigations, it became apparent that *Pd_fcmg13* is a highly expressed gene in invariably all developmental stages (Fig. 13). Despite its strong expression, *Pd_fcmg13* is exclusively present in chaetae-specific tissues; the chaetal sacs. As the investigations show, the gene is inactive in 36-hpf stages (data not shown), but first ISH signals are present in the first setiger and the anlage of the second one of 40-hpf stages (Fig. 13 A). Accordingly, there is a very small time frame, in which the gene expression is probably started. In 48-hpf stages, all chaetal sacs feature very strong ISH signals (Fig. 13 B) that developed within 30 min of NBT/BCIP incubation. Compared with 72-hpf stages, where the staining was present within 2 h of incubation, the strength of gene expression seems to peak in 48 h old larvae. The analysis of 72 h old larvae shows that the second and third setigers feature stronger ISH signals than the first one, suggesting a possible initiation of down-regulation of *Pd_fcmg13* in the oldest setiger. Because of the intense stainings in the early developmental stages, the cellular localization of these signals turned out to be difficult, but it can be suggested that this gene is expressed in the proximal chaetal sac cells, as well indicated by ISH studies in juvenile stages. In these stages, all setigers (Fig. 13 D) feature gene expression restricted to a few single cells (Fig. 13 E-M), as revealed using high magnification light microscopic analyses. Each setiger exhibiting young and growing chaetae shows an ISH signal located around the proximal region of the specific chaeta (Fig. 13 F, K, M). Concerning the functional characterization, similarity search yielded no hits to already published genes. As the reason was thought to be the short (1060 bp) length of sequence information, PERL script-based approaches were performed against the 454-transcriptome data to obtain more nucleotide sequence information. These studies were not able to yield further sequence information for *Pd_fcmg13*. Additional database searches to analyse the protein structure obtained no evidence that the amino acid sequence of *Pd_fcmg13* contains a motif for a signal peptide or at least an anchor protein, but domain searches revealed at least one transmembrane domain.

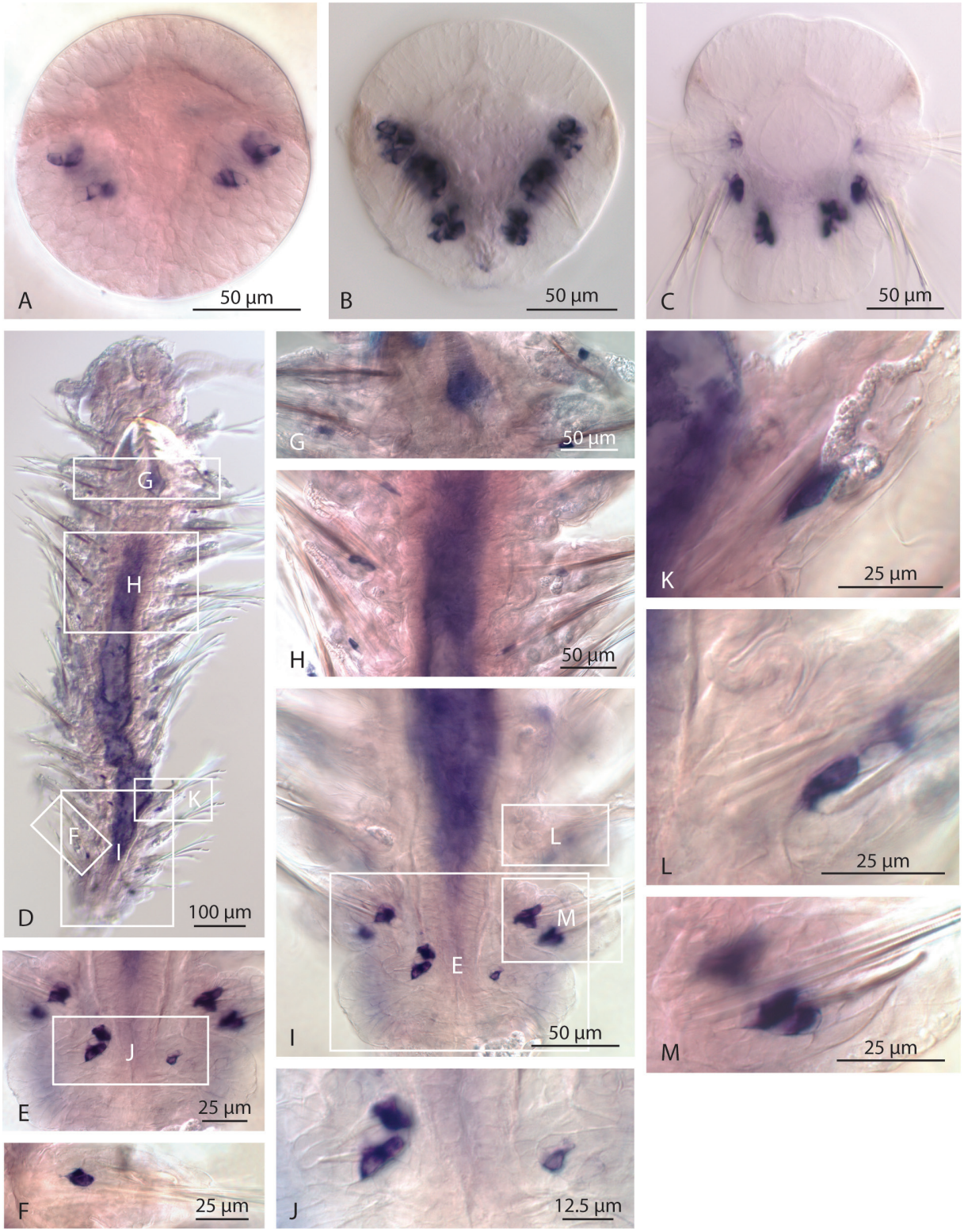


Fig. 13: ISH with *Pd_fcmg13* in *Platynereis dumerilii*. The analysis shows that gene expression is present in all screened stages and that they are exclusively restricted to the chaetal sacs and their proximal cell layers. A – In the 40-hpf stage, the expression was present after 2 h of NBT/BCIP incubation. B+C – The staining was stopped after 30 min (in 48-hpf stages) and 2 h (72-hpf stages) of incubation, as all chaetal sacs of the three setigers presented strong stainings. D-M – Juvenile stage. The ISH signals developed within 4 h of NBT/BCIP incubation and are present in all setigers along the body. Detailed views of selected setigers (E/G/H/I/J) and their parapodia (F/K/L/M) reveal expression being restricted to the proximal chaetal sac cells.

Pd_csmg15

ISH studies show that *Pd_csmg15* is not expressed in the 40-hpf stage (Fig. 14 A), as no NBT/BCIP precipitation is detectable. The first ISH signals are present in 48 h old embryos (Fig. 14 B), where all three setigers feature a strong staining, although the third and youngest setiger exhibits a staining that is weaker than the anterior ones, whose staining intensities indicate a high gene expression level. A similar expression pattern was obtained in ISH studies of 72-hpf stages, showing that *Pd_csmg15* is expressed in all three setigers (Fig. 14 C). But in comparison with the 48-hpf stages, it is more difficult to distinguish the staining from the background signal. Nevertheless, light microscopic studies of 48 h and 72 h old larvae indicate that the gene expression of *Pd_csmg15* is restricted to the middle and proximal cells of the chaetal sacs. Further ISH investigations in juvenile stages show that this gene is solely expressed in 48- and 72-hpf stages, as no setiger of the juvenile worm, not even the youngest posterior-most one, features a specific expression pattern (Fig. 14 D-F). Also for this gene I am interested in the molecular and functional characterization. The clone insert length of *Pd_csmg15* is 1369 bp, whereas further computational approaches for sequence elongation yielded no additional information. Additionally, BLAST search results using the existing 1.3 kb insert showed no similarity to already published genes. Database searches for functional characterization yielded no evidence that *Pd_csmg15* contains structural information that would predict a signal peptide motif, though domain predictions discovered at least one transmembrane domain.

Pd_fcmg18

Analyses of the ISH studies performed with *Pd_fcmg18* show that the first setiger in 40-hpf stages and the anlage of the second setiger exhibit ISH signals in the chaetal sac region (Fig. 15 A), but these are difficult to distinguish from the background signal. In 48-hpf stages all three setigers feature an NBT/BCIP staining (Fig. 15 B) that surrounds the chaetal sac cells along each chaeta. The intense oval signals outside the chaetal sacs within the distal part of the second and third setigers, are restricted to the glands of these segments, which are known to bind unspecifically to many kinds of RNA probes (F. Raible, pers. communication). ISH studies on 72-hpf stages revealed that *Pd_fcmg18* is not expressed at this stage (Fig. 15 C), as no visible, specific NBT/BCIP precipitation was obtained. ISH investigations in juvenile stages indicate that *Pd_fcmg18* is also not expressed in chaetal sacs of developmental stages older than 48 h (Fig. 15 D). The only precipitate that is present within the posterior-most segments appears to be restricted to glands solely (Fig. 15 E+F). The sequencing analysis of the *Pd_fcmg18* clone yielded a fragment size of 798 bp, whereas further elongation approaches obtained a total of 912 bp nucleotide sequence information. BLAST searches against public sequence databases revealed no clear homologs. Database searches for the functional characterization of this gene yielded neither an evidence for a signal peptide nor an anchor protein. Additionally, even domain searches revealed no specific hits.

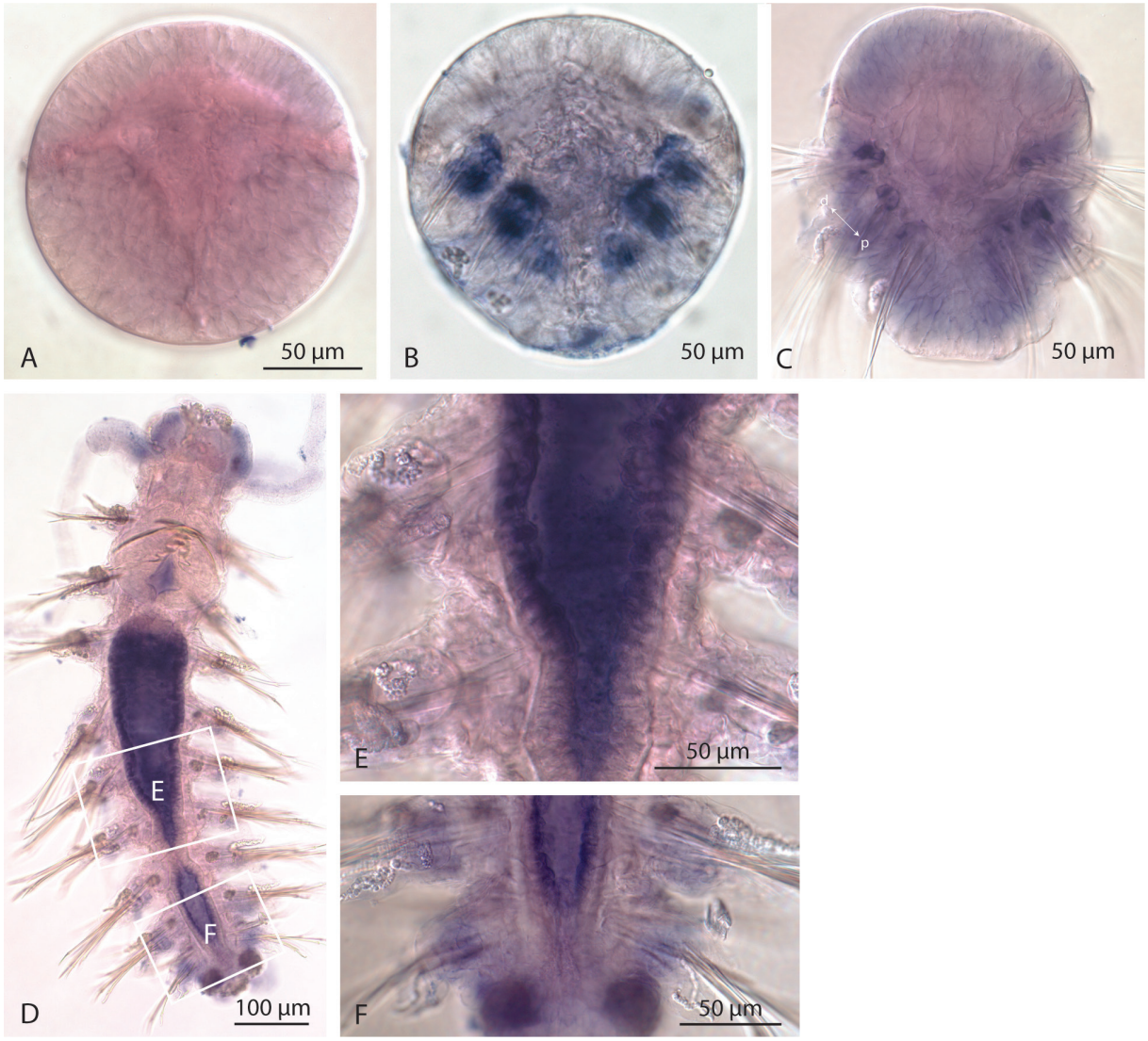


Fig. 14: ISH with *Pd_csmg15* in *Platynereis dumerilii*. The analysis indicates a narrow time range of gene activity that is only active in 48 h (B) and 72 h (C) old larvae. A – No evidence for the expression of *Pd_csmg15* in the 40-hpf stage. B+C – Presence of gene expression restricted to the middle and proximal part of the chaetal sacs. Staining was stopped after 1.5 d (in 48-hpf stage) and 4 h (72-hpf stage), respectively. D – The NBT/BCIP incubation caused a lot of background, whereas the staining had to be stopped after 2 h. However, no ISH signal was detected in any of the setigers.

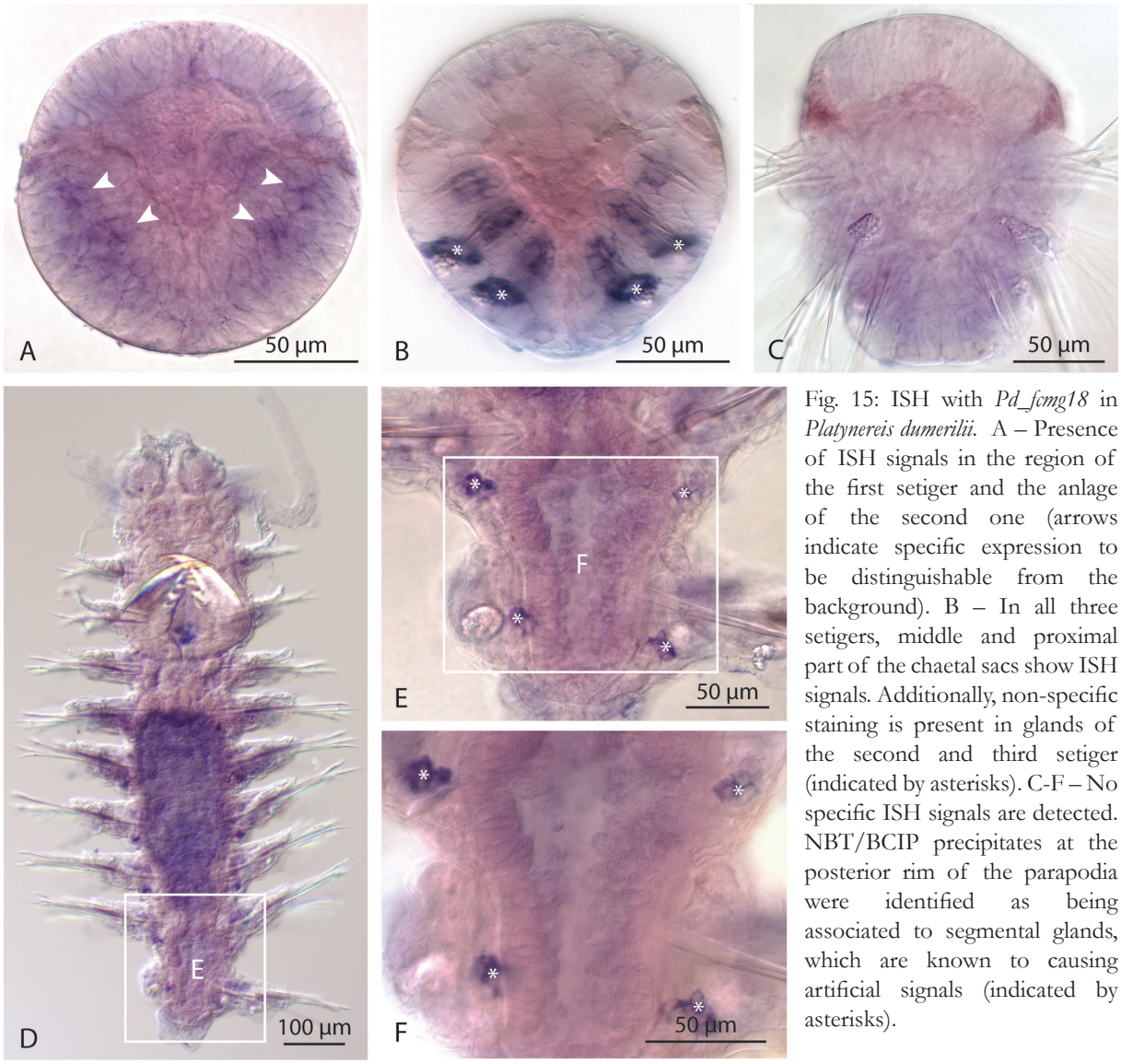


Fig. 15: ISH with *Pd_fcmg18* in *Platynereis dumerilii*. A – Presence of ISH signals in the region of the first setiger and the anlage of the second one (arrows indicate specific expression to be distinguishable from the background). B – In all three setigers, middle and proximal part of the chaetal sacs show ISH signals. Additionally, non-specific staining is present in glands of the second and third setiger (indicated by asterisks). C-F – No specific ISH signals are detected. NBT/BCIP precipitates at the posterior rim of the parapodia were identified as being associated to segmental glands, which are known to causing artificial signals (indicated by asterisks).

Pd_SCA2L

Due to ISH studies, I observed that *Pd_SCA2L* is already expressed in 40-hpf stages (Fig. 16 A), but absent in younger 36 h old embryos (data not shown). A closer analysis of 40-hpf stages reveals that the ISH signal is restricted to the first setiger and the anlage of the second setiger, although a possible third segment (anlage of the third setiger, that is not yet fully differentiated) can already be suggested, as indicated due to a faint NBT/BCIP staining at the posterior end of the embryo. In 48-hpf stages, all three chaetae-bearing segments feature ISH signals (Fig. 16 B) that are present only in the chaetal sac area. In contrast to the above described genes, it is apparent that the ISH signals of *Pd_SCA2L* are more restricted to the distally located cell layers within the chaetal sac structure. Investigations on 72-hpf stages show a comparable ISH pattern as described for 48 h old embryos. However, whereas in the latter stage all setiger feature more or less similar signal intensities, in 72-hpf stages only the second and third setiger exhibit a strong staining (Fig. 16 C). The first setiger however presents a faint ISH signal that is shifted more to the distal-most cell layer. While studying older stages (Fig. 16 D), faint ISH signals were exclusively

obtained within the last three posterior segments (Fig. 16 F-H), whereas the more anterior ones show no expression (Fig. 16 E). In the posterior segments, the staining outlines presumably single cells along the length of the chaeta (Fig. 16 G+H). BLAST searches against public EST databases yielded sequence similarity to an unknown protein from other annelids (*Perinereis nuntia*, *Alvinella pompejana*, *C. teleta*) and moderate sequence to unknown proteins of other taxa (*Crassostrea gigas*, *Schistosoma mansoni*, insects, echinoderms, *Saccoglossus kowalevskii*) as well as a SCA2-like protein from vertebrates. Domain prediction analysis revealed that *Pd_SCA2L* contains structural information about an N-terminal signal peptide motif that is predicted to be involved in intercellular trafficking and at least one transmembrane domain at the C-terminal end. Further structural characterization was conducted by domain analysis of the annotated vertebrate SCA2 proteins and subsequent search for shared features in SCA2-like proteins (see chapter 3.7).

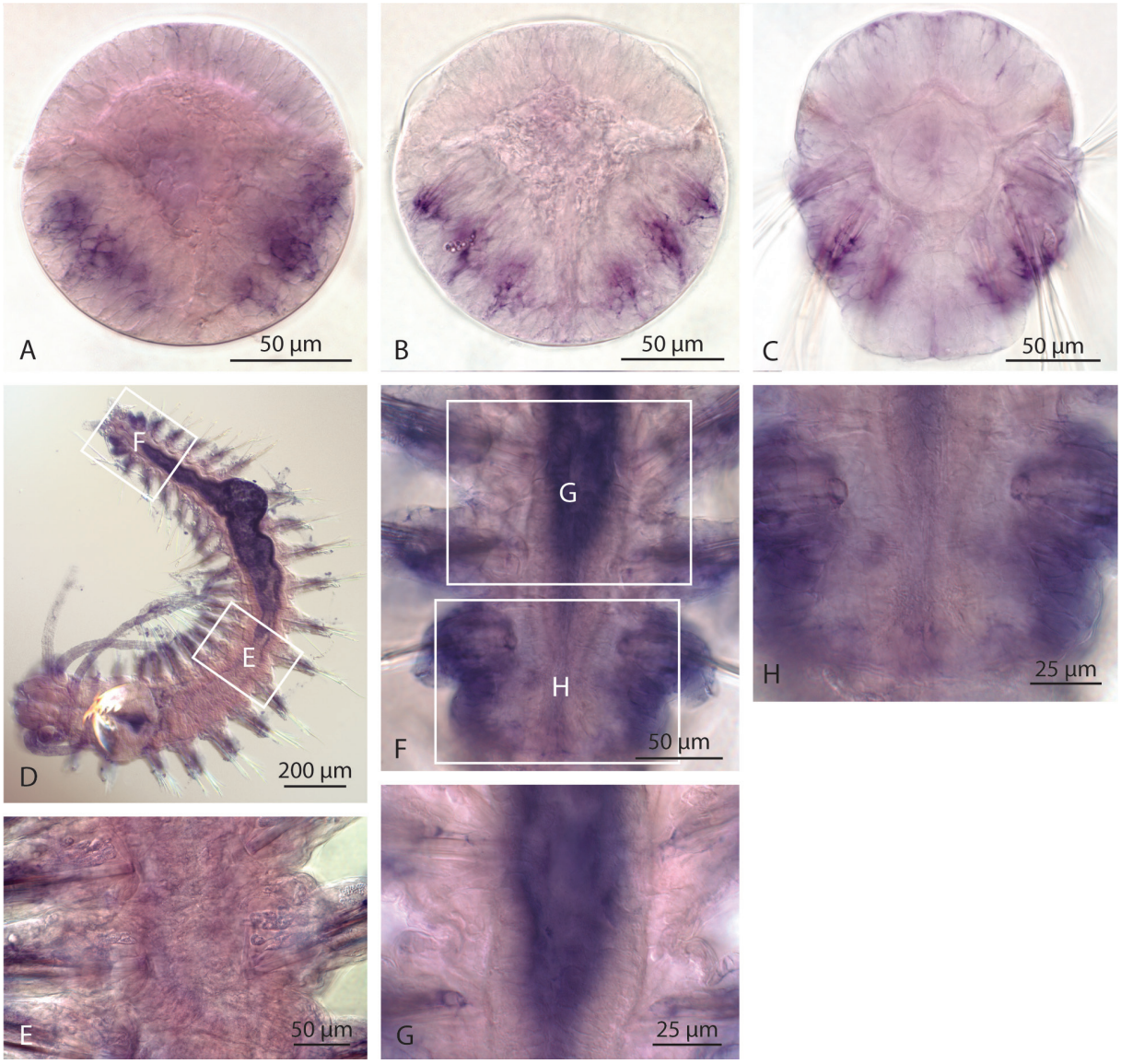


Fig. 16: ISH with *Pd_SCA2L* in *Platynereis dumerilii*. Specific signals are present in all screened stages. A – In 40 h larvae, *Pd_SCA2L* is not only expressed in the chaetal sacs of first setiger and the anlage of the second one, but also in the presumptive anlage of the third setiger. B+C – All three setigers and their corresponding chaetal sacs show distally located ISH signals. However, the staining of the first setiger in 72-hpf stages is decreased and only the distal-most cells are stained. D-H – In juvenile stages, expression is solely present within the last posterior segments. F-H – Detailed view of weak, proximally ISH signals.

Pd_csmg21

ISH studies to analyze the gene expression of *Pd_csmg21* show that well-detectable signals are present in the differentiated first setiger and the anlage of the second one in 40-hpf stages (Fig. 17 A). In contrast, ISH signals are completely absent in investigated younger stages (36 hpf old embryos; data not shown). As in 40 hpf old embryos, chaetae are not yet fully developed, the evaluation of the cellular localization of the gene expression is difficult. 48-hpf stages reveal a better possibility to estimate whether a gene is presumably expressed in chaetoblasts or in follicle cells. Here, all three chaetae-bearing segments feature strong ISH signals exclusively restricted to the chaetal sac region (Fig. 17 B), although the staining of the third and youngest setiger is weaker than the two. The ISH signal covers half of the cells of each chaetal sac and suggests a

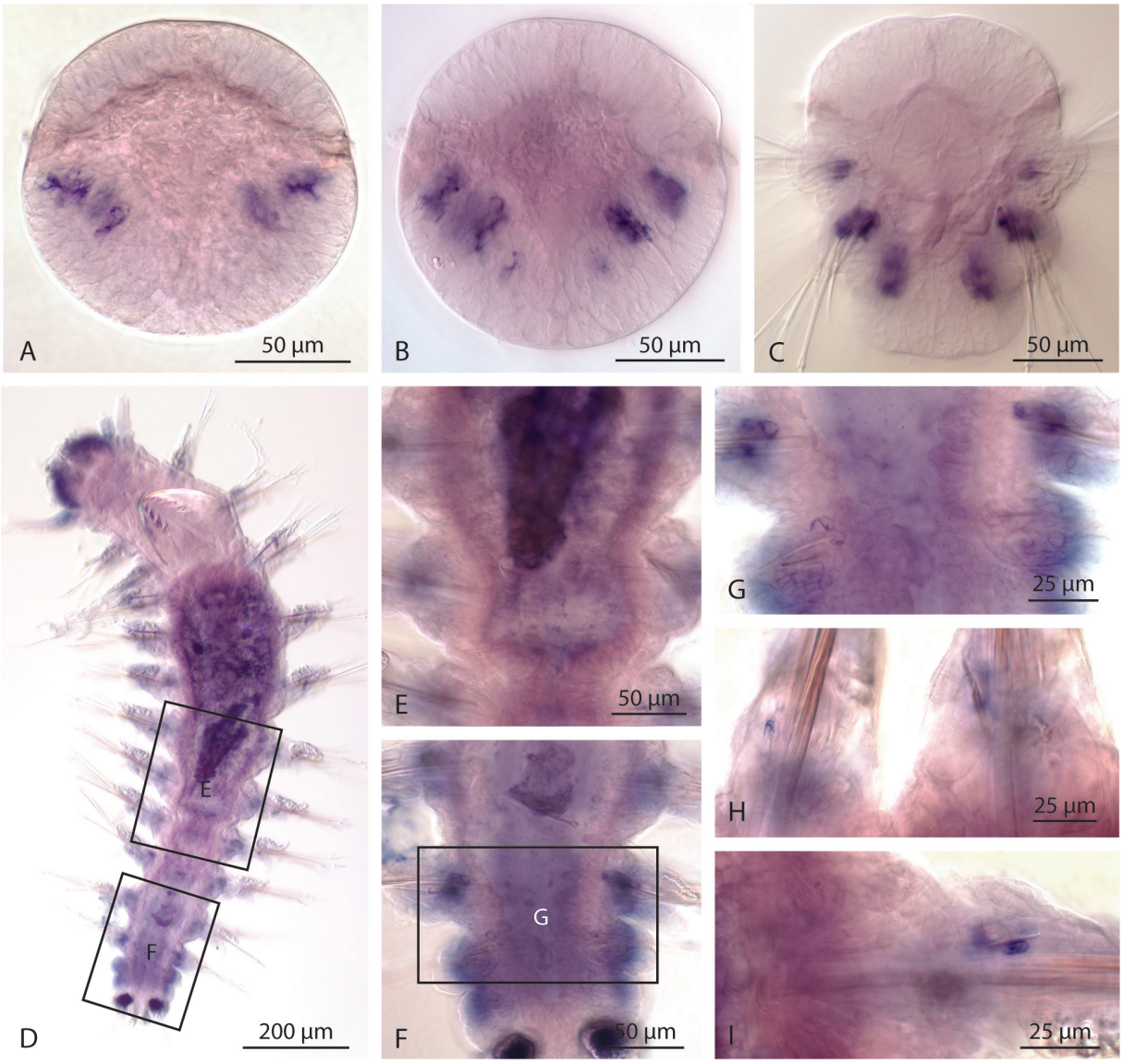


Fig. 17: ISH with *Pd_csmg21* in *Platynereis dumerilii*. Specific signals are present in all screened stages. A – In 40-hpf stages, prominent gene expression patterns are present in the chaetal sacs of the first setiger and the anlage of the chaetal sacs of the second setiger. B+C – All three setigers and their corresponding chaetal sacs show an expression restricted to the middle and proximal region. D-I – ISH studies in juvenile stages show that the expression is solely present within the last posterior segments.

cellular localization within the more proximally located cell layers. In 72 h old embryos, a similar pattern is evident, namely strong ISH signals in all three setigers (Fig. 17 C). Nevertheless, a shift in gene expression is apparent, as the chaetal sacs of the first and oldest setiger exhibit weaker stainings compared with the situation in 48-hpf stages. Light microscopic investigations of juveniles show that ISH signals are present only in the posterior most segments (Fig. 17 D-I), as only here young chaetal stages are surrounded by an NBT/BCIP signal also restricted to their proximal region (Fig. 17 G-I). Sequencing of the existing clone insert revealed an 1175 bp fragment, whose length could not be elongated with the applied approaches. Nevertheless, investigations on amino acid level revealed that *Pd_csmg21* contains a signal anchor protein. These proteins are uncleaved signal proteins that target the protein to the membrane of the endoplasmatic reticulum. Additional database searches for domain predictions discovered seven transmembrane domains, which confirms the suggestion of *Pd_csmg21* being a membrane located protein.

3.3 Cellular localization of gene expression patterns

3.3.1 3D TEM analysis of chaetal sacs

While screening all the different genes, it became evident that the gene expression for many genes tends to be cell-specific. However, the analyses using light microscopy revealed that there must be a very compact cellular arrangement within each chaetal sac that makes it difficult to decide in which cell type an ISH signal is located. Generally, annelid chaetae emerge from an epidermal follicle (Fig. 18) that consists of follicle cells and one basal chaetoblast (Bouligand 1967, Specht, 1988, Hausen 2005). Previous investigations on ultrastructural level (Fig. 19 A) discovered that the follicle cells not only surround the developing chaeta by simply being located on top of each other. For the position of follicle cell one and two it became apparent that they both encompass the chaeta on the same level as the chaeta arises from the apical microvilli surface of the chaetoblast. Ultrastructural

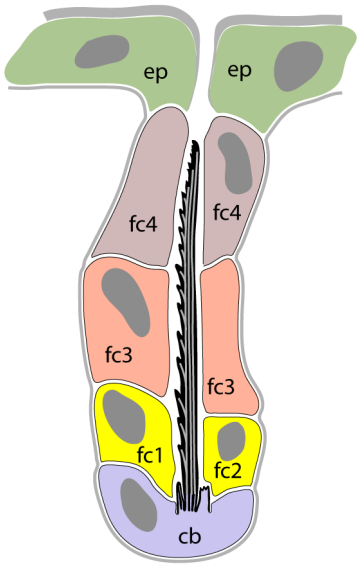


Fig. 18: Schematic drawing of a single chaetal follicle that gives rise to one developing chaeta. One follicle consists of one basally located chaetoblast followed by additional layers of follicle cells (follicle cells one to four). The first layer of follicle cells that is distally located of the chaetoblast is characterized by the position of both follicle cells within one level, whereas in all following follicle cell layers, a single cell encompasses the growing chaeta. cb – chaetoblast; ep – epidermal cell; fc – follicle cell (one to four).

investigations revealed that both cells, follicle cell one and two, comprehend the same conspicuous vesicles. Follicle cell three is distally located of follicle cells one and two, followed by follicle cell four and finally epidermal cells.

A look into several developmental stages of *P. dumerilii* already gives a suggestion about the upcoming problem, as each chaetal sac consists not only of one follicle but several ones (Fig. 19 A), resulting in bundles of chaetae per chaetal sac. The situation becomes even more complex by young chaetae, as formation of new chaeta generally starts by invagination of a new follicle from the epidermal surface, which then continuously sinks down (Specht, 1988, Bartolomaeus 1998, Hausen 2005). To get more information about the cellular arrangement within a chaetal sac, I did ultrastructural investigations (TEM analysis) on 48-hpf stages of *P. dumerilii*. At that stage, chaetogenesis seems to be very active as light microscopy reveals chaetal sacs with prominently developed chaetae that are close to protruding the body surface. For exact analysis of a chaetal sac, an araldite-embedded *P. dumerilii* larva was orientated into a lateral position to cut cross-sections of a chaetal sac (Fig. 19 A). These sections were processed for TEM analysis, one selected chaetal sac was photographed and the resulting image stack was aligned (Fig. 19 B). Redrawing of all nuclei, chaetae and extracellular matrix (ECM) using the software 3D Studio Max resulted in a 3D model of a chaetal sac with all its information about the cellular arrangement within this structure (Fig. 19 C). This model shows that follicle cells present the majority of chaetal sac cells and are arranged in layers along the chaetae (Fig. 19 D). Nevertheless, the information about the position of chaetoblasts in relation to that of the chaetae is the most important one. This model indicates that in 48-hpf larva of *P. dumerilli* a chaetoblast can be localized by the relative position to its corresponding chaeta, simply by following the length of one chaeta down to its base. The nucleus localized underneath the chaetal basis belongs to the chaetoblast of that chaeta. Additionally it became evident that at the base of the chaetal sac, only chaetoblasts are adjacent to the extracellular matrix. Chaetoblasts are followed by a proximal recessed layer of follicle cell one and two, distally followed by follicle cell three and additionally four. It became apparent that some chaetoblasts are not located at the base of the chaetal sac but embedded among follicle cells. These chaetoblasts belong to follicles of young developing chaetae.

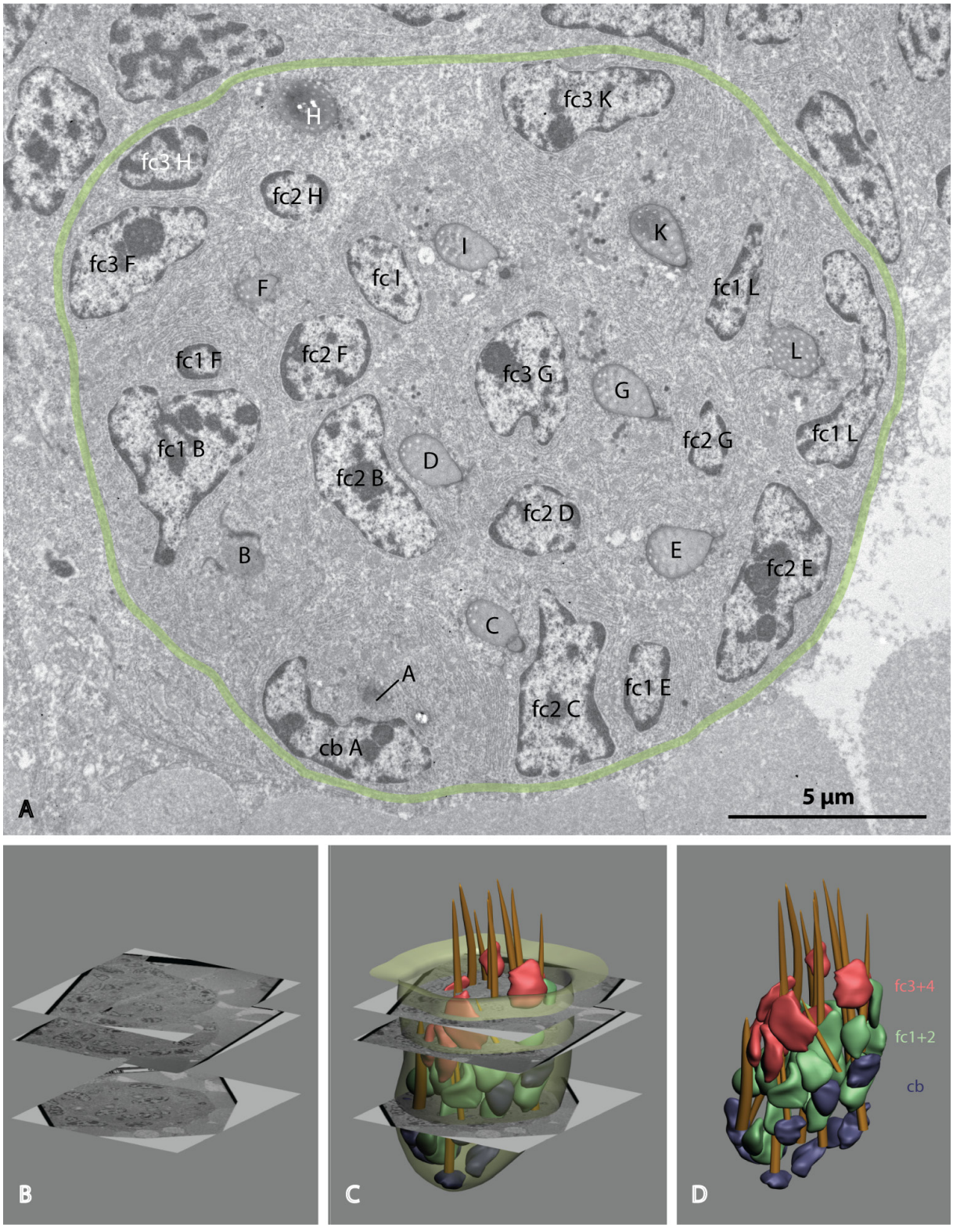


Fig. 19: 3D TEM analysis of chaetal sacs in 48 h old larvae of *Platynereis dumerilii*. A – Cross section of a chaetal sac (surrounded by extra cellular matrix) with several chaetae (A to L) and nuclei of the corresponding follicle cells (fc) and chaetoblasts (cb). B – Alignment of chaetal sac cross-sections (TEM). C – 3D reconstruction of the chaetal sac and its cellular arrangement including the z-stack alignment of cross-sections (from A). D - 3D reconstruction of the cellular arrangement within a chaetal sac based on the ultrastructural information of the cross-sections. Three different cellular layers are distinguishable. The most proximal located cells, mainly colocalized by the base of a chaeta, can be specified as chaetoblasts. The following layer consists of the follicle cells 1 and 2 that surround the chaetal shaft from both sides. The next layer is characterized by the follicle cells 3 (followed by follicle cell layer 4 that is not visualized here) located in the mid to distal region of the chaetal sac. cb/blue – chaetoblast, fc – follicle cell, brown – chaetae, semi-transparent green – extracellular matrix (ECM) surrounding the chaetal sac.

3.3.2 Immunhistochemical 3D-imaging of chaetal sacs

Having the information about the cellular arrangement within a chaetal sac based on 3D TEM analysis, new strategies can be developed using immunohistochemistry to create a comparable 3D dataset of a chaetal sac. As a result of the previous analysis, chaetae have to be used to localize the position of the chaetoblasts. Therefore I developed a protocol to detect chaetae in combination with a fluorescent dye or an antibody. As chaetae consist of β-chitin, Wheat Germ Agglutinin (WGA) conjugated to a fluorescent label suggests to be the best candidate to detect chaetae. Experiments using different concentration levels to obtain ideal results showed that WGA stains not only the N-acetyl glucosamine subunits of distal, therefore more compact chitin material-containing chaetal region, but also the proximal-most at the base of a chaeta. Additionally, the combination of WGA staining (Fig. 20 A) with a Propidium Iodide-induced nuclei staining (Fig. 20 B) revealed data similar to that one of the 3D-TEM-analysis (Fig. 20 C). Therefore, the results were processed using the Amira software in combination with volume rendering technique for chaetae visualization and surface rendering, to outline the nuclei.

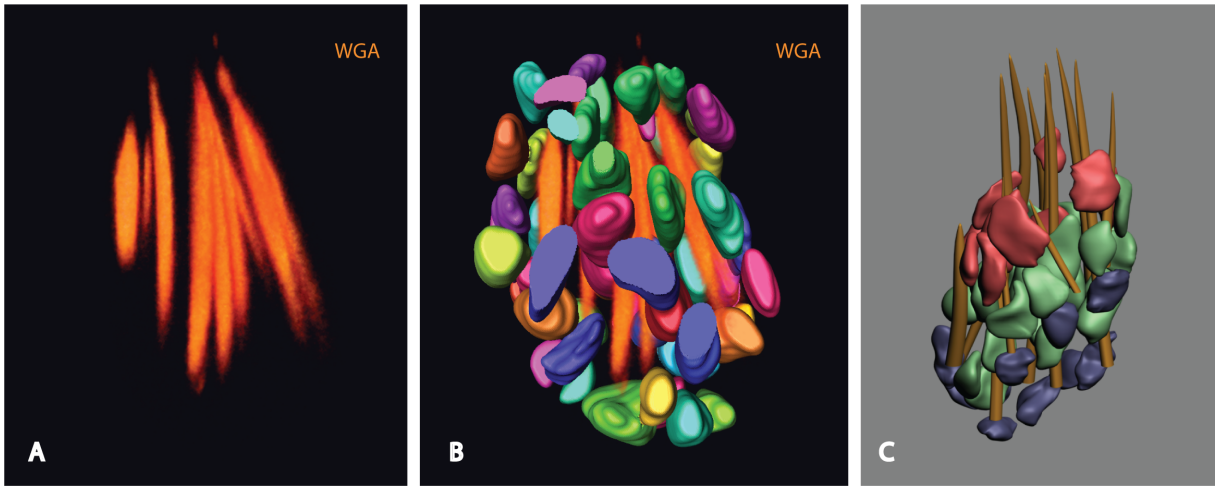


Fig. 20: 3D-imaging of chaetal sacs based on immunhistochemical data. A – Detection of chitinous chaetae by conjugated Wheat Germ Agglutinin (WGA). B – Co-localization of chaetae (WGA) and nuclei (Propidium Iodid) showing a comparable picture as known from the 3D-TEM-analysis model (C). orange – chaetae, stained with WGA; multicolored – nuclei, stained with PI.

3.3.3 Application of FISH for cellular localization

Finally, having the information of the cellular arrangement, which facilitates the identification of the different chaetal sac cells by 3D immunohistochemistry, new ISHs were started. To gain the necessary high resolution, a combination of Fluorescent *in-situ* hybridization (FISH), Propidium Iodid nuclei staining and WGA chaetae staining was performed and recorded with confocal microscopy. In some cases the ISH signal was instead visualized by NBT/BCIP precipitation and detected either as reflection signal or in transmission mode. Further, double ISHs were used for colocalization analyses.

Pd_cbmg3

New investigations on *Pd_cbmg3* showed a similar expression pattern compared to the previous WMISH (Fig. 8). The ISH signal suggests an expression in the proximal cells of the chaetal sacs in all three setigers (Fig. 21 A). However, only confocal microscopy allows unambiguous cellular localization of gene expression. A detailed view shows that a very prominent ISH staining is present in single cells covering the proximal-most part within the chaetal sac (Fig. 21 B). Using WGA-stained chaetae to localize the nuclei of the chaetoblasts it becomes evident that these are enclosed by the ISH signal with *Pd_cbmg3* probes (Fig. 21 B+C). Thus, *Pd_cbmg3* is identified to be a chaetoblast marker. As the signal is very prominent and has good cellular resolution (Fig. 21 C), it became apparent that this gene would give a good candidate for use in colocalization studies.

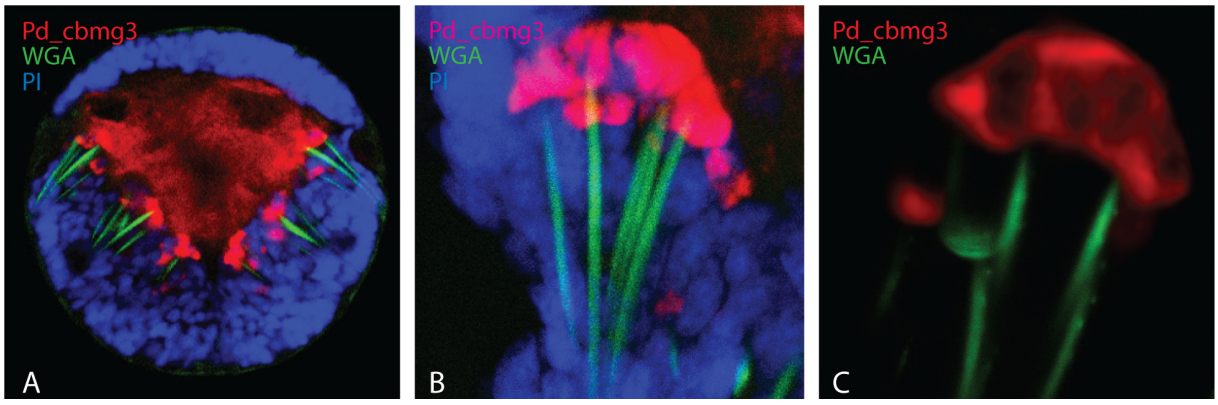


Fig. 21: FISH with *Pd_cbmg3* in 48-hpf stages of *Platynereis dumerilii*. A – Ventral larval view with gene expression present in all three setigers. B/C – Detailed view (z-projections) of one selected chaetal sac. Using the WGA staining of the chaetae to localize their chaetoblasts reveals that the expression of *Pd_cbmg3* colocalizes with that of the chaetoblast cells. C – Higher magnification analysis shows that this genes expression has a clear cellular resolution. red – ISH signal; green – chaetae, stained with Wheat Germ Agglutinin (WGA); blue – nuclei, stained with Propidium Iodid (PI).

Pd_CS1

Studies using the FISH technique for *Pd_CS1* show that in 48-hpf stages, clear and strong expression patterns are present in all three setiger (Fig. 22 A) as described above (Fig. 2). But still the cellular localization is difficult. Higher magnification of one chaetal sac of the second setiger indicates expression of *Pd_CS1* in cells at the base of each chaeta, characterized as chaetoblasts (Fig. 22 B). Therefore, double FISHs for *Pd_CS1* were performed to test if there is a coexpression with the highly specific chaetoblast marker gene *Pd_cbmg3*. Already the total view of a larva in lower magnification indicates the strong local correlation between the two gene expressions (Fig. 22 C). Detailed analysis of the third left setiger's chaetal sac present an exact overlap of the ISH signals of *Pd_CS1* (Fig. 22 D) with *Pd_cbmg3* (Fig. 22 E). Additional studies, using different tyramids to label the specific probes revealed the same exact colocalization for both genes (Fig. 22 F). With these experiments it became evident that firstly, *Pd_cbmg3* works perfectly well as colocalization marker in double FISHs and secondly, the chitin synthase 1 of *P. dumerilii* is expressed solely within the chaetoblast cells. The latter fact is most interesting because it is the first evidence in annelids about where the chitin synthesis takes place.

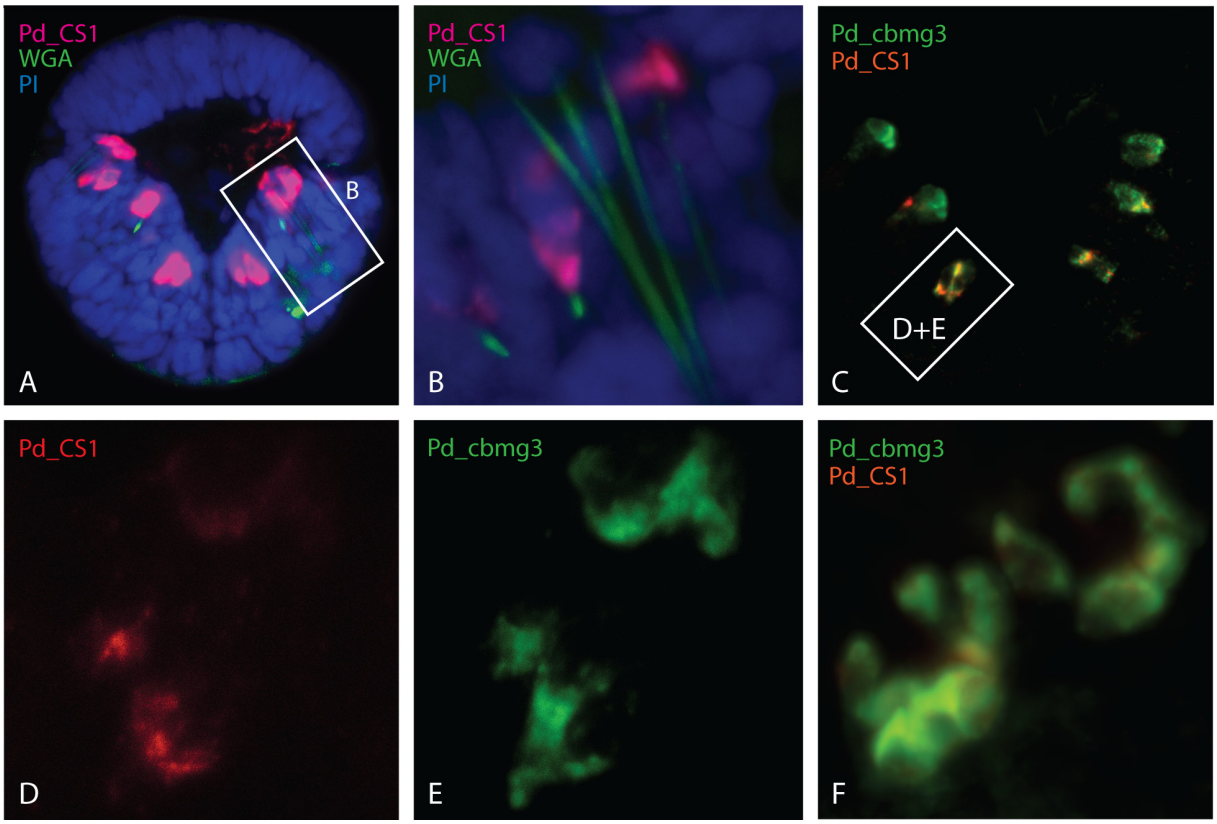


Fig. 22: FISH and double FISH with *Pd_CS1* in 48-hpf stages of *Platynereis dumerilii*. A+B – FISH experiments. A – Ventral view presenting ISH signals in the proximal parts of all chaetal sacs of the three setigers. B – Detailed view (z-projection) of one chaetal sac of the right second setiger. Colocalization of chaetoblast cells using WGA staining indicates an expression in chaetoblast cells. C-F – Double FISH experiments using the (Fluorescein-detected) *Pd_cbmg3* as a specific chaetoblast marker to test for colocalization with (Cyanine5-detected) *Pd_CS1*. C – Ventral view (z-projection), showing both genes expressed within the three setigers restricted to the same cells. D+E – Selected single chaetal sac presents expression of *Pd_CS1* (D) and *Pd_cbmg3* (E), respectively. F – Z-projection of two selected chaetal sacs that definitely reveal a co-expression of both genes within the same cells – the chaetoblast. red/green – ISH signal; long green structures – chaetae, stained with Wheat Germ Agglutinin (WGA); blue – nuclei, stained with Propidium Iodid (PI).

Pd_fcmg5

Analyses of *Pd_fcmg5* expression show that, in comparison with general WMISHs, the use of FISH technique enables a more reliable interpretation about the cellular localization of this gene. Overview images already indicate an expression of *Pd_fcmg5* that is restricted to follicle cells (Fig. 23 A). A closer analysis of the second (left) setiger shows that there is no ISH signal in the proximal-most cells at the base of the chaetae, whereas upper cell layers of the distal chaetal sac region are prominently stained (Fig. 23 B). Interpretation of this cellular restriction suggests that *Pd_fcmg5* is expressed in follicle cells only. By comparison with the data of the TEM analysis, gene expression can be more precisely localized as being most likely expressed in the follicle cells three and four. 3D reconstruction of the same confocal stack data indicate the same suggestion of *Pd_fcmg5* expression being restricted to the distal-most two follicle cells (Fig. 23 C). This characterization of *Pd_fcmg5* and the high signal quality commends this gene as a potential colocalization marker for follicle cells in addition to *Pd_cbmg3*.

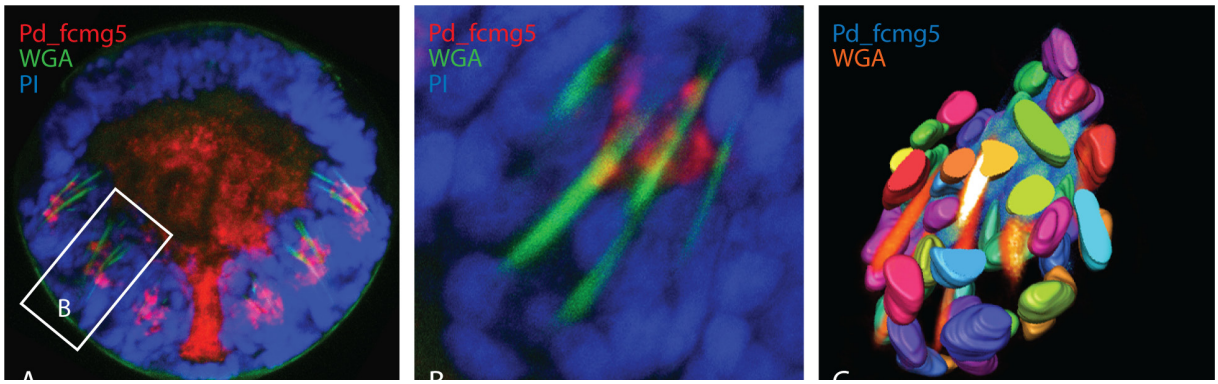


Fig. 23: FISH with *Pd_fcmg5* in 48-hpf stages of *Platynereis dumerilii*. A – The ventral view shows that ISH signals are present in all developed chaetal sacs with a restriction more to the distal part of the chaetal sac. B – Detailed view (z-projection) of the left chaetal sac of the second setiger. No gene expression is present in the chaetoblast cells at the base of the chaetae, whereas distally located follicle cells (presumably follicle cell three and four) present a staining. C – 3D reconstruction of the chaetal sac stack-data of the second setiger. red – ISH signal; green/orange – chaetae, stained with Wheat Germ Agglutinin (WGA); blue/multicolored – nuclei, stained with Propidium Iodid (PI).

Pd_csmg10

Compared to the above described genes, *Pd_csmg10* occurs to show a different expression pattern, as it seems that this gene is expressed in two different cell layers. A closer analysis of selected chaetal sac data shows on the one hand a prominent expression within the chaetoblast cells at the base of WGA-stained chaetae (Fig. 24 A). On the other hand, there is evidence for ISH signals present in the first follicle cells located apically of the chaetoblasts (Fig. 24 B). 3D reconstructions from image z-stacks support this interpretation (Fig. 24 C).

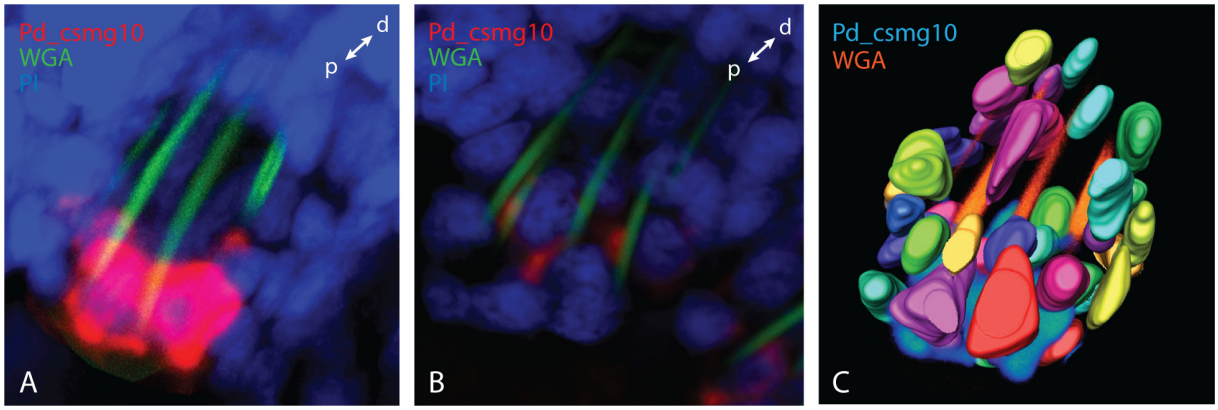


Fig. 24: FISH with *Pd_csmg10* in 48-hpf stages of *Platynereis dumerilii*. The study indicates that this gene is expressed both in follicle cells and chaetoblasts. A/B – Z-projections of different subsets of image layers of the same chaetal sac. A – ISH signals are located in cells directly at the base of each visualized chaeta. Additionally there is evidence for an expression within the following follicle cell layer, too. B – Detailed analysis of the relative position of chaetae to that of the nuclei reveals that *Pd_csmg10* is also expressed within the proximal-most follicle cells. C – 3D reconstruction of the whole z-stack of A + B. blue/multicolored – nuclei, stained with Propidium Iodid (PI); d – distal; green/orange – chaetae, stained with Wheat Germ Agglutinin (WGA); p – proximal; red – ISH signal.

Pd_fcmg2

The gene expression of *Pd_fcmg2* in 48-hpf stages shows a similar pattern as described above (Fig. 7). The chaetal sacs of all three setigers show an ISH signal that is proximally located (Fig. 25 A; not

all setigers in the same focal plane). A closer analysis of the third chaetal sac (left-sided) reveals that expression of *Pd_fcmg2* is restricted to follicle cells only (Fig. 25 B). ISH signal is present in follicle cells one and two as well as in follicle cell three.

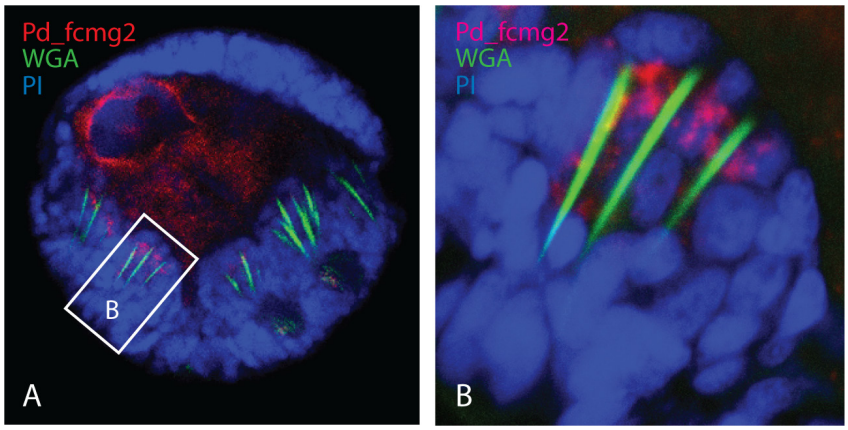


Fig. 25: FISH with *Pd_fcmg2* in 48-hpf stages of *Platynereis dumerilii*. A – Overview (not all setigers are located within the same focal plane), showing that *Pd_fcmg2* is expressed in all chaetal sacs of all three setigers. B – The detailed analysis of one selected chaetal sac reveals that the expression is present in follicle cells only, but in several layers (presumably follicle cells one to three). blue – nuclei, stained with Propidium Iodid (PI); green – chaetae, stained with Wheat Germ Agglutinin (WGA); red – ISH signal.

Pd_fcmg4

FISH studies in 48-hpf stages showed (Fig. 26 A), that transcripts of *Pd_fcmg4* are present in all yet developed chaetal sacs. At this stage, their expression is more prominent in the middle region of each chaetal sac. By comparing these findings with the TEM data, it is likely that the expression is restricted to follicle cells. Accordingly, a detailed analysis of one selected chaetal sac shows that no basal cell presents an ISH signal, whereas the cells (follicle cells) aligned at half of the length of the chaetae exhibit a strong and prominent expression (Fig. 26 B+C). In contrast to *Pd_fcmg2*, the expression of this gene seems to be restricted to a low number of cells, suggesting a presence in follicle cell three only, for colocalization of the expression with the position of chaetae indicate that the staining is not overlapping with the position of follicle cells one and two.

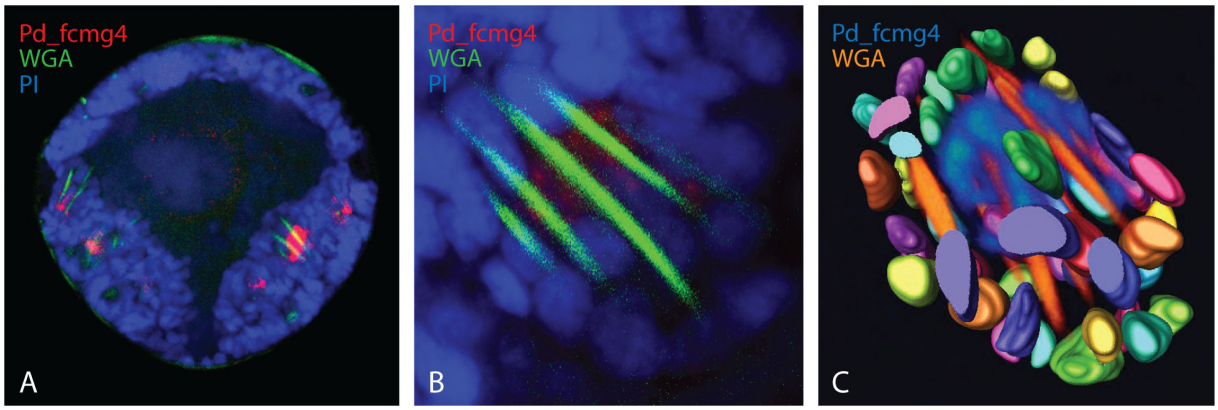


Fig. 26: FISH with *Pd_fcmg4* in 48-hpf stages of *Platynereis dumerilii*. A – Overview that shows the presence of *Pd_fcmg4* expression in each chaetal sac of the setigers. The ISH signals are localized more in the mid-region of each chaetal sac. B – Z-projection of one selected chaetal sac, presenting an expression in the middle follicle cell layer. C – 3D reconstruction of the z-stack data used in B. blue/multicolored – nuclei, stained with Propidium Iodid (PI); green/orange – chaetae, stained with Wheat Germ Agglutinin (WGA); red – ISH signal.

Pd_CAML

The first results of the FISH studies showed that *Pd_CAML* is present in all three setigers in 48-hpf stages (Fig. 27 A). Each chaetal sac seemed to exhibit a relatively broad staining (caused by relatively high unspecific background staining), making it difficult to specify the cell that expresses the gene. The analysis focusing on just one single setiger at higher magnification indicates an expression of *Pd_ CAML* within several layers of follicle cells (Fig. 27 B). However, a clear suggestion about whether or not the gene is additionally expressed in the chaetoblasts cannot be made from these data. Therefore, further ISHs were performed in combination with the chaetoblast-specific gene *Pd_CS*. If there is an overlap in both ISH signals, *Pd_ CAML* would be suggested to be additionally expressed in chaetoblasts. By a closer analysis of the double FISH experiments it became evident that there is no co-expression of *Pd_ CAML* with the chaetoblast marker, wherefore this gene is expressed in follicle cells only (Fig. 27 C).

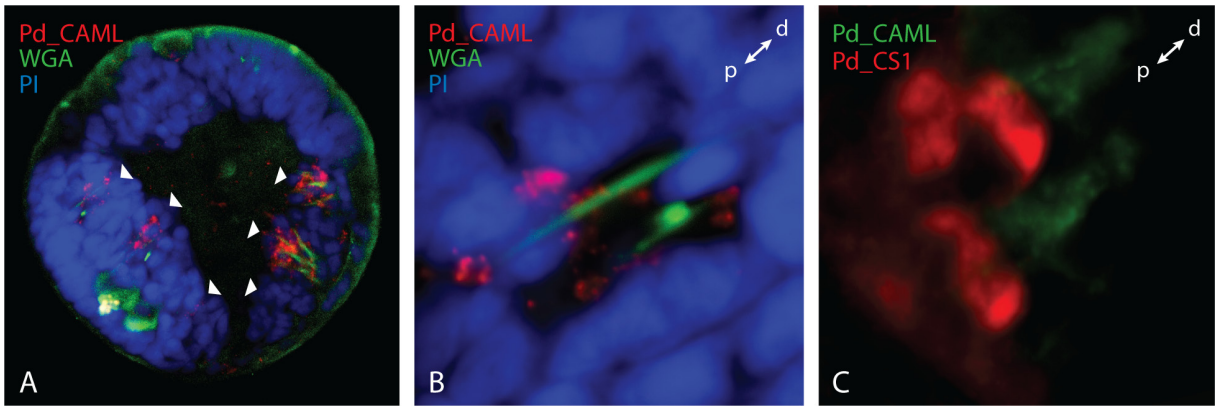


Fig. 27: FISH with *Pd_CAML* in 48-hpf stages of *Platynereis dumerilii*. A – Ventral view (not all setigers are aligned within the same focal plane), presents expression patterns in all chaetal sacs. B/C – A closer analysis indicates that *Pd_CAML* is expressed in follicle cells only, with a restriction to the mid to proximal situated follicle cell layers within the chaetal sac (z-projections). C – Double FISH, using the chaetoblast-specific gene *Pd_CS1* reveals that there is no evidence for a co-expression of both genes. blue – nuclei, stained with Propidium Iodid (PI); d – distal; green – chaetae, stained with Wheat Germ Agglutinin (WGA); p – proximal; red/green – ISH signal.

Pd_fcmg11

Due to the analysis of gene expression with *Pd_fcmg11* it became evident that this gene is expressed in follicle cells only, as the ISH signals are located in the mid to distal chaetal sac region, whereas no expression patterns are present in the proximal chaetal sac area at the base of each chaeta (Fig. 28).

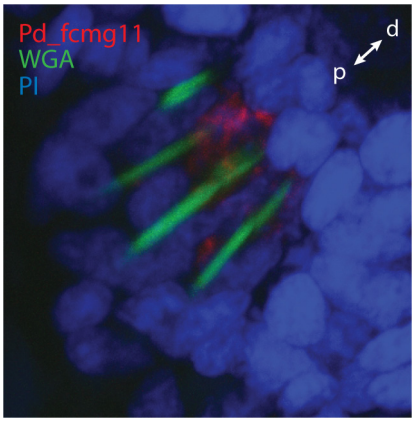


Fig. 28: FISH with *Pd_fcmg11* in 48-hpf stages of *Platynereis dumerilii*. Z-projection of one selected chaetal sac that gives evidence about the expression of *Pd_fcmg11* solely restricted to follicle cells, for there is no ISH signal at the base of any chaeta. blue – nuclei, stained with Propidium Iodid (PI); d – distal; green – chaetae, stained with Wheat Germ Agglutinin (WGA); p – proximal; red – ISH signal.

Pd_fcmg13

All investigated larvae showed prominent ISH signals in all chaetal sacs of the three setiger (Fig. 29 A). However, to obtain information about the cellular localization of the genes expression, a detailed analysis on the resolution level of single chaetal sacs is necessary. These investigations revealed that expression of *Pd_fcmg13* is exclusively restricted to follicle cells of the proximal-most layers (Fig. 29 B), as the marked cells are not localized at the base of the chaetae, which pass through the ISH signal. In comparing the data with information obtained by ultrastructural investigations the expressing cells can be identified as follicle cells one and two.

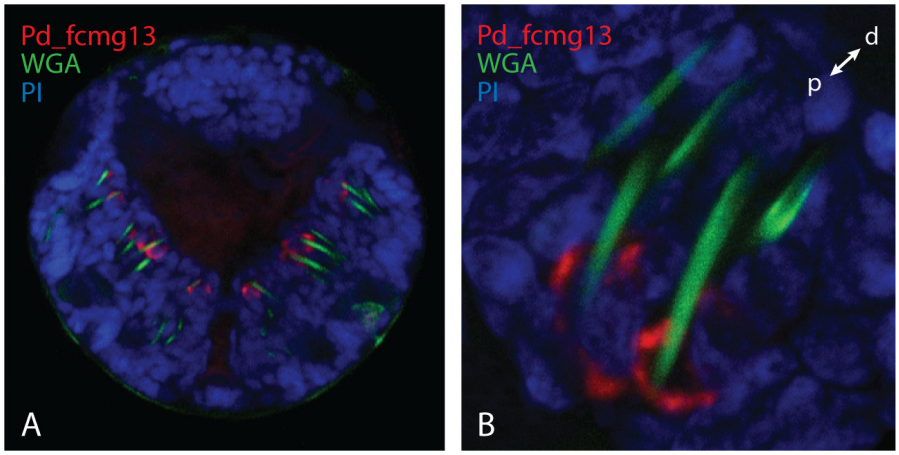


Fig. 29: FISH with *Pd_fcmg13* in 48-hpf stages of *Platynereis dumerilii*. A – Ventral view, presenting expression patterns in all chaetal sacs. B – Detailed view of one single chaetal sac indicating that *Pd_fcmg13* is expressed in follicle cells only, with a restriction to the more proximally situated cells within the chaetal sac (z-projection). blue – nuclei, stained with Propidiumiodid (PI); d – distal; green – chaetae, stained with Wheat Germ Agglutinin (WGA); p – proximal; red – ISH signal.

Pd_MLCK1

The analysis of FISH with *Pd_MLCK1* proved to be difficult, as in all experiments a strong, potentially artificial background staining developed, making it difficult to differentiate between signal and noise. Nevertheless, weak staining is visible in the proximal part of the chaetal sacs (Fig. 30 A+B). Additional double FISH studies also were not successful. For that reason, co-localization analysis was performed based on FISH combined with standard ISH to finally obtain information on which cells are expressing *Pd_MLCK1*. In these experiments, at first *Pd_MLCK1* was detected via NBT/BCIP precipitation method, whereas the colocalization marker *Pd_CS1* was fluorescently labeled and detected with the Tyramid technique (Perkin Elmer) afterwards. Finally, these experiments revealed that *Pd_MLCK1* is exclusively expressed in follicle cells. Detailed high magnification analysis of the data shows that the *Pd_MLCK1* expressing cells are arranged directly on top of the apical surface of the chaetoblasts indicating the position of the follicle cell one and two (Fig. 30 C).

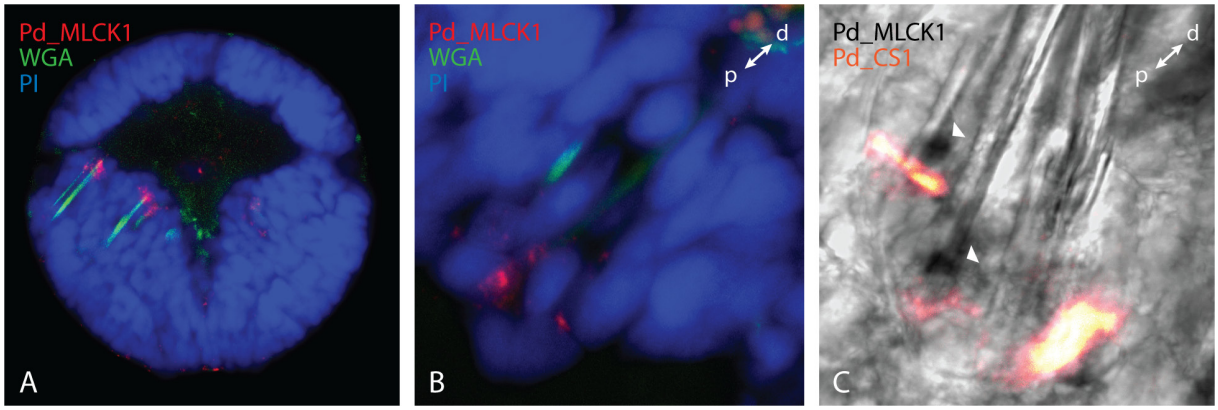


Fig. 30: FISH with *Pd_MLCK1* in 48- and 72-hpf stages of *Platynereis dumerilii*. A – In 48-hpf stages, expression patterns are present in proximal regions of the chaetal sac (not all setigers in the same focal plane). B – Analysis on a single chaetal sac level gives first hints on gene expression within the proximal follicle cells. C – Detailed view of a chaetal sac in 72-hpf stages. arrows – showing the position of NBT/BCIP precipitation (black spots) in C; blue – nuclei, stained with Propidiumiodid (PI); d – distal; green – chaetae, stained with Wheat Germ Agglutinin (WGA); p – proximal; red – ISH signal.

With the help of these detection methods, ranging from single fluorescent detection (FISH) over approaches of combined fluorescent and NBT/BCIP detection to double FISHs, cellular localization of gene expression was possible for a wide range of genes. In total, expression of ten of fifteen candidate genes was clearly assignable to specific subsets of the chaetal sac cells.

3.4 Evolution of chitin synthases

3.4.1 Identification of chitin synthases in *P. dumerilii*

Of special interest in the context of the formation of chitinous chaetae is the candidate gene *Pd_CS1*, as BLASTx searches suggest strong similarity to the chitin synthase family. For further sequence analyses it was necessary to obtain as much sequence information as possible.

Therefore primers for RACE-PCR were designed to amplify the up- and downstream regions of the chitin synthase gene (see Appendix, Tab. A4). As the gel electrophoresis analysis of the PCR amplifications never detected any visible products, the gel was southern blotted and radioactive high stringency hybridizations performed with a chitin synthase specific probe. The hybridization revealed bands only for the nested PCRs, where a 0.7- and 0.8-kb band was detected, but after sequencing it turned out that the contigs contained a large putatively intronic region.

For that reason, a new strategy was developed. Alignments of other possible closely related chitin synthases from *C. teleta* as well as published mollusc chitin synthases showed that these genes obtain a highly conserved chitin synthase domain and some of them an additional presumably conserved myosin motor domain at the five prime end (Weiss et al. 2006). Thus, the new strategy was to use this information to connect the sequences of both domains using degenerated PCR, instead of RACE-PCR. With the help of K. Tessmar-Raible and F. Raible (MFPL Vienna, Austria) I designed degenerated and specific primers (including primers for two nested reactions) for the corresponding regions (see Appendix, Tab. A1). However, again no appropriate products were achieved, even after high stringency hybridization. As the previous experiments with the first chitin synthase clone revealed that PCR amplifications based on its sequence information never obtained any usable product or contained potential intron sequences, I received a second candidate clone with the help of Stephan Schneider (Iowa State University, USA). Sequence alignment with other chitin synthases revealed that this clone sequence is located downstream of my initial chitin synthase fragment. Therefore primers were designed to amplify the region between both fragments, but again without success.

Finally, local BLAST searches against a 454-library of the *P. dumerilii* transcriptome revealed two contigs upstream of the new chitin synthase fragment. An alignment analysis with other chitin synthases suggested that the new contigs and the clone sequence of *P. dumerilii* show sequence similarity to specific conserved regions (of the known full length sequence of *Atrina rigida*) – one approximately 300 amino acids (aa) behind the possible myosin motor domain, whereas the others aligned to the core region of the chitin synthase domain and another downstream region, respectively. Subsequently, primers were designed that are located in the conserved regions to amplify the unknown regions between the up- and downstream fragments. With this strategy, finally 4.8 kb of continuous nucleotide sequence information were obtained. The BLASTp search of the new 4.8 kb hit revealed a prediction similar to the chitin synthase of the molluscs *A. rigida*, *Pinctada fucata* and *Mytilus galloprovincialis* (high e-value of 0.0).

In parallel to the cloning strategy, the arising genome and transcriptome data of *P. dumerilii* were continuously checked on the presence of chitin synthase sequences. By this, two additional chitin synthase gene candidates were uncovered and one of them could be successfully amplified via PCR. Nonetheless, the presence of a myosin motor domain could not yet be validated for any of the three candidate genes. Therefore, a new PCR strategy was developed, trying to use specific myosin motor domain forward primers in combination with reverse primers, specific to the five prime region of each of the three chitin synthase genes. To obtain the presumptive myosin motor domain of *P. dumerilii*, I BLAST searched the myosin motor domain of *P. fucata* against the *P. dumerilii* genome and transcriptome. The best hit was BLAST searched against GenBank and yielded high support for similarity to the myosin motor domains of the above mentioned molluscs. In addition to the above mentioned primers, primers were designed to obtain the myosin motor domain exclusively. However, no PCR approach yielded a positive band neither for the myosin motor domain itself nor the hypothesized fusion of myosin motor domain with one of each of the chitin synthase candidate genes.

In summary, I was able to extend the formerly 1318 bp chitin synthase fragment to a 4.8 kb sequence, which is almost half of the expected sequence information (11-12 kb of the full length gene). Even though information about the potential 5' region (and therefore the presence of a myosin motor domain) is still absent, the current sequence information is continuously translatable including the complete catalytic chitin synthase domain and extending to the 3' end of the open reading frame.

3.4.2 Molecular evolution of chitin synthases

Chitin synthases are not only common in organisms like annelids and the well investigated arthropods that both feature prominent chitinous structures, such as β-chitin chaetae and α-chitin cuticles, respectively. From GenBank and JGI, additional sequences and gene predictions were obtained of taxa like Fungi, nematods, Cnidaria, Porifera, molluscs and chordates. For several species multiple chitin synthase sequences were found. This includes species like *Lottia gigantea* with nine chitin synthases (*Logi_CS1* to *Logi_CS9*), *C. teleta* with four (*Cate_CS1* to *Cate_CS4*) and *P. dumerilii* with three (*Pd_CS1* to *Pd_CS3*) potential chitin synthases. Two chitin synthases were found in *Manduca sexta* (*Mase_CS1 +Mase_CS2*), *Brugia malayi* (*Brma_CS1 + Brma_CS2*), *Caenorhabditis elegans* (*Cael_CS1 + Cael_CS2*), *Amphimedon queenslandica* (*Amqu_CS1 + Amqu_CS2*; annotated as hypothetical protein prediction) and *Nematostella vectensis* (*Neve_CS1 + Neve_CS2*; annotated as hypothetical protein prediction), respectively. Nonetheless, in the remaining investigated species a single candidate was found, namely in *Myzostoma cirriferum* (*Myci_CS*), *Macandrevia cranium* (*Macr_CS*), *A. rigida* (*Atri_CS*), *P. fucata* (*Pifu_CS*), *M. galloprovincialis* (*Myga_CS*), *Dirofilaria immitis* (*Diim_CS*), *Meloidogyne artiellia* (*Mear_CS*) and *Hydra magnipapillata* (*Hyma_CS*). In terms of chordates, I found enzymes whose sequences were highly similar to this of chitin synthases. This involves *Ciona intestinalis* (*Ciin_CS*); annotated as hypothetical protein prediction), *Branchiostoma floridae* (*Brfl_CS*; annotated as hypothetical protein prediction), *Danio rerio*

(*Dare_CS*; annotated as hypothetical protein prediction) and *Xenopus tropicalis* (*Xetr_CS*; annotated as hypothetical protein prediction), respectively. As in fungi, chitin synthases appeared to be quite abundant (Roncero 2002) I conducted an initial evolutionary analysis of the six fungal chitin synthase "classes" *sensu* Roncero (2002; see Appendix, Fig. A2) and then selected one representative of Basidiomycota (*Lentinula edodes*; *Leed_CS1*) as well as one Ascomycota representative (*Colletotrichum graminicola*; *Cogr_CSc*). All respective amino acid sequences were aligned using the MAFFT server (version 6, E-INS-i). The resulting alignments were manually edited to exclude ambiguous alignment positions and postions with many gaps and used for a maximum likelihood (ML) analysis (RAxML, Stamatakis et al. 2008). As the amino acid sequences of the two chitin synthases of *H. robusta* (*Hero_CS1* + *Hero_CS2*) proved to be too short, they were excluded from the main analysis. However, to get an idea about the presumable position within the tree, a second tree analysis was performed including these data.

The resulting tree of the main analysis (without *H. robusta*; Fig. 31) was rooted to the two representatives of Basidiomycota and Ascomycota, revealing that all metazoan chitin synthases group together within one clade that is supported by 100% of the bootstrap (BS) replicates. Within that metazoan clade, major chitin synthase groups can be identified. The first branch represents a clade of chitin synthases of Cnidaria and Porifera, which obtained 100% bootstrap support and depict a sister group relationship to all chitin synthases of Bilateria (77% BS support). The basal branching pattern of bilaterian chitin synthases (e.g., the monophyly and position of chordate chitin synthases) has low bootstrap values, but several major subgroups are well-supported. The topology suggests (but with low bootstrap support of 23%) that Nematoda B chitin synthases (100% BS support) are sister to a clade (74% BS support) containing all other known chitin synthases of Ecdysozoa (Nematoda A + Arthropoda; with 99% BS support) and one clade comprising a certain subset of all sampled lophotrochozoan chitin synthases (80% BS support). It is apparent that this clade contains only lophotrochozoan chitin synthases (Mollusca D + Annelida D) that lack a myosin motor domain (MMD). On the other hand, all lophotrochozoan chitin synthases possessing a myosin motor domain cluster within a clade that is well-supported by 95% of the bootstrap replicates and contains all fragmentary chitin synthases from *P. dumerillii*, *M. cranium* and *M. cirriferum*. These MMD-bearing lophotrochozoan sequences assemble three moderately to strongly supported groups (clade A: Mollusca A + Annelida A + Brachiopoda; clade B: Mollusca B + Annelida B; clade C: Mollusca C + Annelida C). It is striking that none of the three groups contains only mollusc or annelid chitin synthases, but in every clade genes of both – molluscs and annelids – consistently group together (once even with a brachiopod chitin synthase and once with a myzostomid chitin synthase). The earliest-branching clade (100% BS support) consists of one of the *L. gigantea* chitin synthases (Mollusca group C) and the Annelida group C chitin synthases (*Pd_CS1* and *Cate_CS2*). On the other hand, there is moderate support (clade A + clade B; 68% BS support) for a close relationship of the two remaining groups of the sampled lophotrochozoan chitin synthases that are each well supported (clade A with 81% BS support; clade B with 91% BS support). Within clade A, the chitin synthases of *A. rigida*, *P. fucata* and *M. galloprovincialis*, as well as three chitin synthases of *L. gigantea* represent one well supported group (Mollusca group A; 97% BS support) that groups

together with the chitin synthase clade of *C. teleta* (*Cate_CS1*) and *M. cranium* (Annelida group A + Brachiopoda; 57% BS support). Within clade B, a mollusc chitin synthase of *L. gigantea* (Mollusca group B) branches as sister to the chitin synthases of Annelida group B (100% BS support). This clade suggests strong support (100% BS replicates) that the chitin synthase of *M. cirriferum* groups together with two chitin synthases of *P. dumerilii* (*Pd_CS2* and *Pd_CS3*; 100% BS support).

The second chitin synthase tree (see Appendix, Fig. A3) includes *H. robusta* and revealed similar results concerning the branch support and tree topology. In terms of *H. robusta*, both chitin synthases are closely related to each other (95% BS support) and indicated to be sister group to the non-MMD-bearing chitin synthases of *C. teleta* (*Cate_CS3* + *Cate_CS4*). However, this relationship is only supported by a weak BS value (33% BS support).

Based on this chitin synthase tree, the common origin of all lophotrochozoan MMD-bearing chitin synthases indicates that a linkage of the MMD with a chitin synthase is homologous in all these Lophotrochozoa. Additional evidence for monophyly of all MMD-bearing chitin synthases is given by the presence of a "bridge" region within the chitin synthase domain. This region is characterized by an insertion of around 27 to 62 amino acids into the otherwise strictly conserved chitin synthase domain (Fig. 32). This CS domain alignment suggests that this region is absent in all groups lacking an MMD (all representatives of Nematoda group B, Nematoda group A + Arthropoda and Mollusca D + Annelida D, respectively), but present in all lophotrochozoan MMD-bearing chitin synthases. For some of the chitin synthases nested within this clade (*Pd_CS1*, *Pd_CS2*, *Pd_CS3*, *Myci_CS* and *Cate_CS2*) no MMD-related sequence has yet been obtained, but they all feature this "bridge" region within the chitin synthase domain. Furthermore, since all other metazoan chitin synthases lack MMDs, it can be suggested that presence of an MMD-linkage evolved once within the ancestor of the sampled lophotrochozoans and independently from the situation in Fungi. This is further supported by a phylogenetic analysis (Fig. 33) of MMDs in comparison with representatives of different myosin gene subclasses (*sensu* Odronitz & Kollmar 2007), where the lophotrochozoan MMDs are monophyletic (97% BS support; including *Miye_Myo*, a previously unclassified myosin from *Mizuhopecten yessoensis*) and cluster within a clade of class III and related myosins. Additionally, these lophotrochozoan MMDs are only distantly related to the MMDs of fungal chitin synthases, as the latter belong to class XVII myosins (Odronitz & Kollmar 2007).

Domain prediction analyses (Fig. 34) using SMART search revealed that metazoan chitin synthases generally possess seven to eleven transmembrane domains upstream of the chitin synthase domain (but downstream of the MMD, if present) and two to five transmembrane domains behind. In contrast, chitin synthases of fungi bear only two to three transmembrane domains between their MMD and chitin synthase domain and none downstream of the CS domain. Some predictions include additional domains in a few sequences, such as an IQ domain in *Macr_CS* + *Logi_CS5*. As most additional domain predictions were restricted to single sequences (such as a SCOP domain in *Pd_CS1*, two repeat regions in *Brfl_CS*, a SAM domain in *Amqu_CS2* and Cytb5 + DEK_C domains in *Cogr_CS*), they were not considered in the comparative structural analysis (Fig. 34).

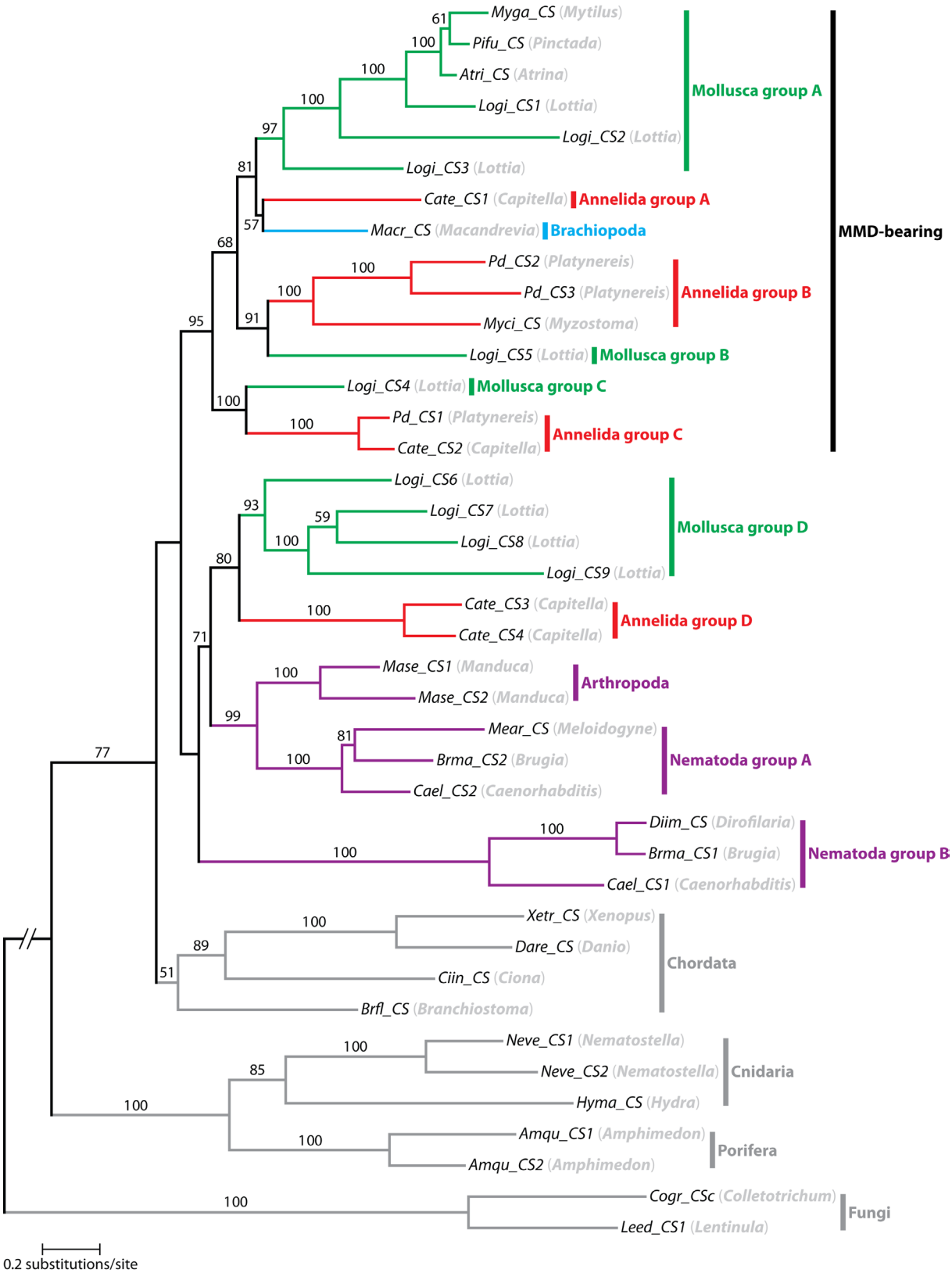


Fig. 31: Evolutionary tree of metazoan chitin synthase domains. The ML phylogram (RAxML, WAGF model) is derived from an alignment (969 amino acid positions) of 40 sequences and bootstrap values above 50% are shown. Chitin synthases of all sampled lophotrochozoan taxa are closely related. The analysis revealed two major clades, of which one is characterized by all lophotrochozoan chitin synthases (CS) that are linked to a myosin motor domain (MMD), whereas the other clade comprises all representatives of the sampled lophotrochozoans and ecdysozoans that lack a MMD. See Appendix for the sequence alignment and for an analysis including two *Helobdella robusta* chitin synthases (Appendix, Fig. A3).

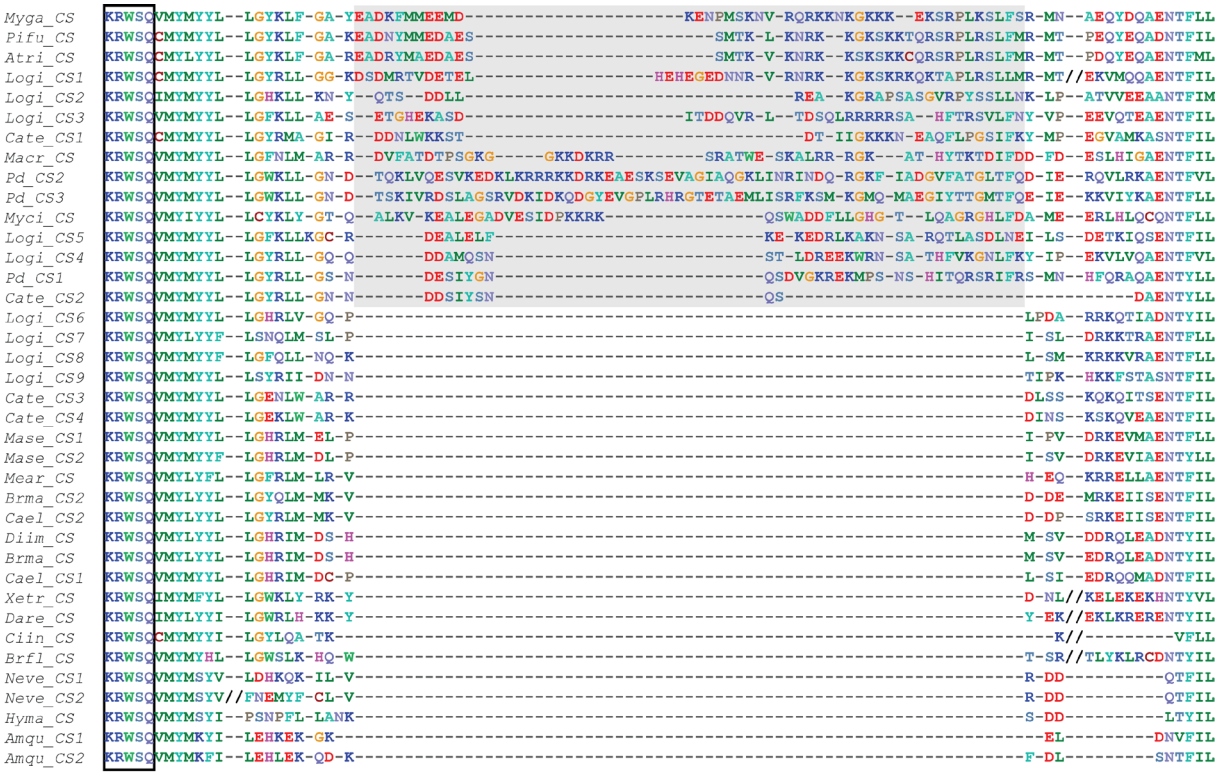


Fig. 32: Alignment of the "bridge" region within the chitin synthase domain of chitin synthases. The "bridge" region (highlighted in grey) starts 15 amino acid positions after the conserved "KRWSQ"-motif (boxed region) and seems to be specific to all myosin motor domain-bearing chitin synthases of all sampled lophotrochozoans.

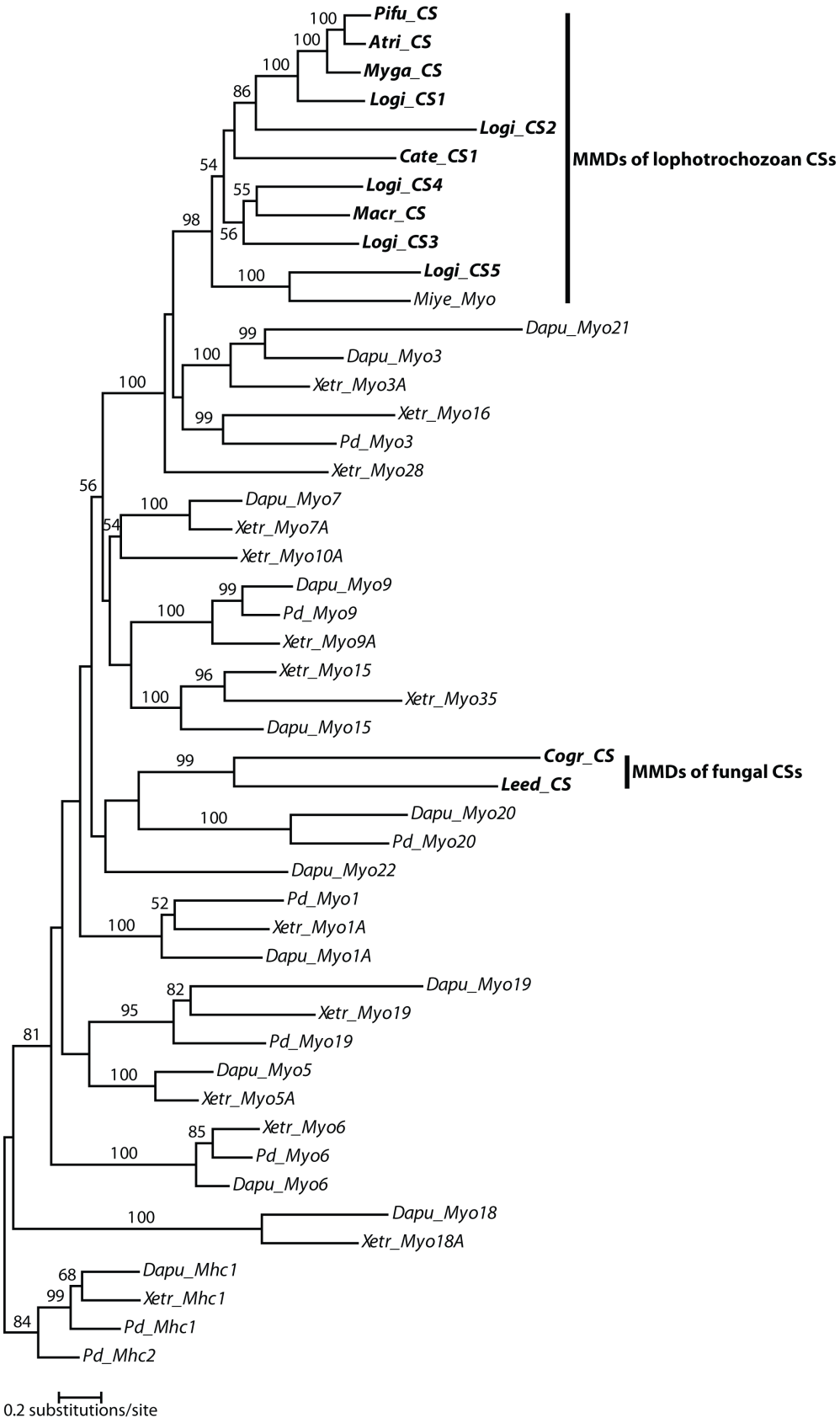


Fig. 33: Evolutionary tree of chitin synthase-linked myosin motor domains (MMDs) and other myosins. The ML phylogram (RAxML, RTREVF model) is derived from an alignment (665 amino acid positions) of 48 sequences and bootstrap values above 50% are shown. The analysis includes myosin hits from the *P. dumerilii* transcriptome and representatives of the myosin classes *sensu* Odronitz & Kollmar (2007) of *Xenopus tropicalis* (*Xetr*), *Daphnia pulex* (*Dapu*) and *Mizuhopecten yessoensis* (*Miye*). See Appendix for the sequence alignment.

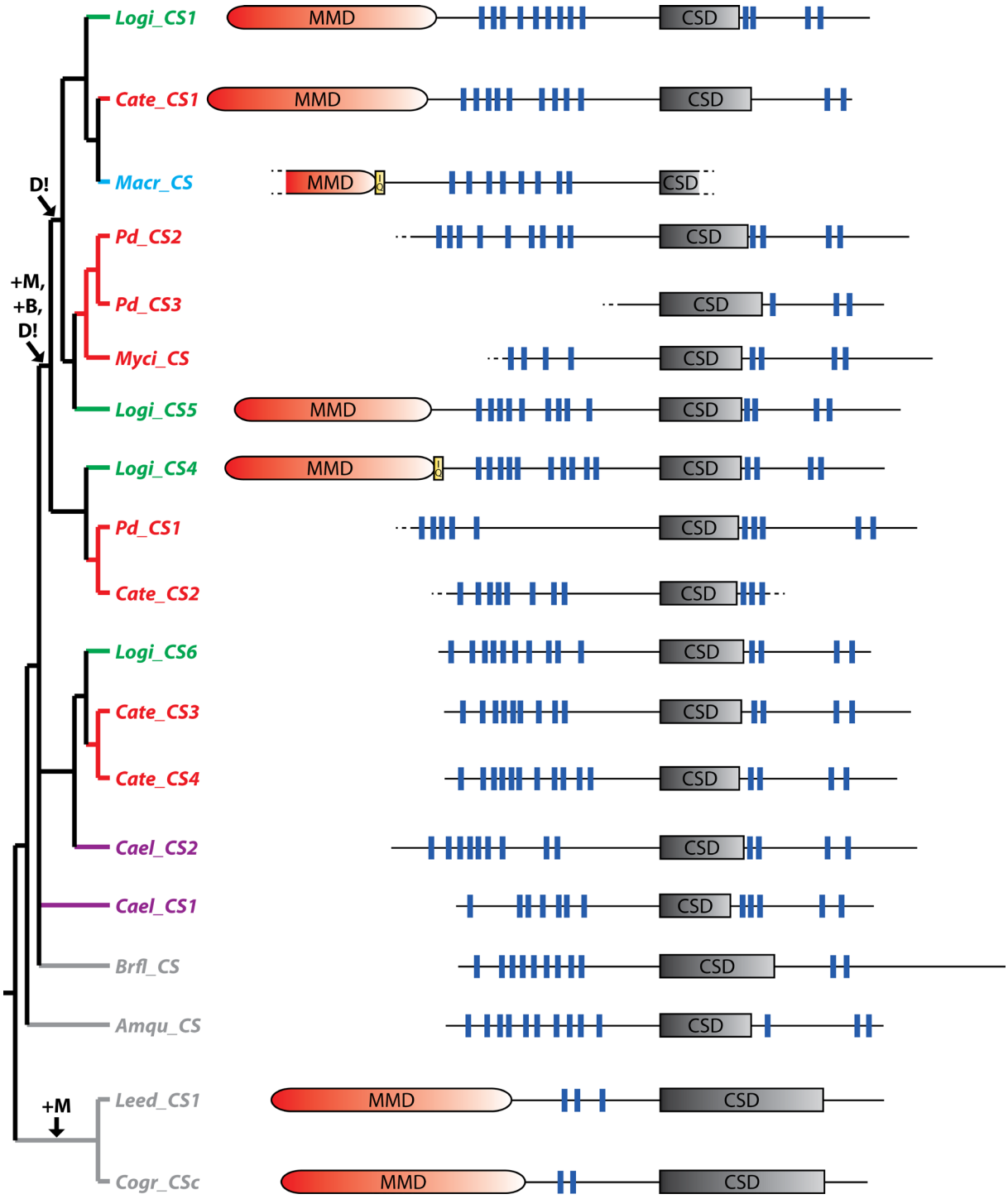


Fig. 34: Domain structure of metazoan chitin synthases in the light of a simplified tree topology of Fig. 31. The annotation of putative domains was conducted using the SMART search server. Branches with less than 50% bootstrap support were collapsed. Putative evolutionary events are mapped on the tree. Green branches are mollusk-specific, red branches are annelid-specific, blue branches are brachiopod-specific, violet branches ecdysozoan-specific and grey branches are of non-protostomian taxa. blue (vertical) bar – transmembrane domain; +B – insertion of bridge-motif; CSD – chitin synthase domain; D! – gene duplication; IQ – IQ motif; +M – linkage of chitin synthase with myosin motor domain; MMD – myosin motor domain.

3.5 Myosin light chain kinases in *P. dumerilii*

In the context of myosin motor domain-bearing chitin synthases another candidate gene, namely a myosin light chain kinase (MLCK), appears to be interesting for further investigations, as this gene might act as a regulator in a possible interaction of the myosin motor domain of the chitin synthase with actin filaments, as known from the musculature system (Kamm & Stull 2001). Therefore, investigations were performed to study the molecular and structural aspects of that gene.

First BLASTx searches indicated that *Pd_MLCK1* is similar to a MLCK of *Strongylocentrotus purpuratus* (e-value: 1e-109). As the initial *Pd_MLCK1* sequence fragment of *P. dumerilii* lacked a catalytic domain that is typical for functional smooth muscle myosin light chain kinases (smMLCK), the prediction was still vague. Consequently, the idea was to check if the existing *Pd_MLCK1* sequence information can be extended to the downstream-located catalytic domain using PCR. Therefore, protein sequences of vertebrate smMLCK were selected using the GenBank sequence database (MYLK_SHEEP, accession number: O02827; MYLK_RABIT, accession number: P29294; MYLK_BOVIN, accession number: Q28824), whose sequences were analysed with the SMART search database to extract the core sequence of the catalytic domain. The analysis revealed two different conserved catalytic domain sequences that were used in a BLASTp search against the *C. teleta* transcriptome and genome data to obtain the *C. teleta* MLCK (JGI accession number: 148032). After extraction of the so identified catalytic domain region in *C. teleta*, this specific sequence was used for a BLASTx analysis against the recently generated *P. dumerilii* transcriptome data. With this BLAST method I received a contig sequence that contained information highly identical to the catalytic domain of *C. teleta*. A reBLAST of the contig sequence against the GenBank database revealed a prediction for a MLCK similar to that one of *Saccoglossus kowalevskii* (e-value of 1e-47), whereas the SMART analysis helped to find the region of the catalytic kinase domain. With the sequence information about a catalytic domain in *P. dumerilii*, forward primers were placed into the downstream region of the "original" *Pd_MLCK1* sequence and the reverse primer into the upstream region of the catalytic domain (see Appendix, Tab. A3). Using this PCR approach, I amplified a product showing that *Pd_MLCK1* has a catalytic kinase domain and thus resembles a functional MLCK. In combination with *P. dumerilii* transcriptome data, a 6042 bp sequence was obtained, whose translated amino acid sequence includes eleven Ig domains, an Fn domain and a catalytic protein-kinase domain (Fig. 35). Additionally, a triple repeat motif "DFRxxL" typical for an acting binding site (Kamm & Stull 2001) was found twice, first in the amino acids 440-501 and second at position 1428-1489, as this appears to be part of a large duplicated region within *Pd_MLCK1*.

BLAST searches in EST and transciptome databases identified additional MLCKs (Fig. 35) with a kinase domain (*Cate_MLCK*, *Logi_MLCK*), as well as short fragments of Ig domains from other representatives of Lophotrochozoa (*Hero_MLCK*, *Myci_MLCK*, *Macr_MLCK*). In a recent update of the *P. dumerilii* transcriptome data, I could also detect a second fragment of a MLCK (*Pd_MLCK2*) that contains three Ig domains. Evolutionary analysis (see Appendix, Fig. A4) indicates that both MLCKs of *P. dumerilii* are closely related to each other and are together potentially orthologous to the other lophotrochozoan MLCKs.

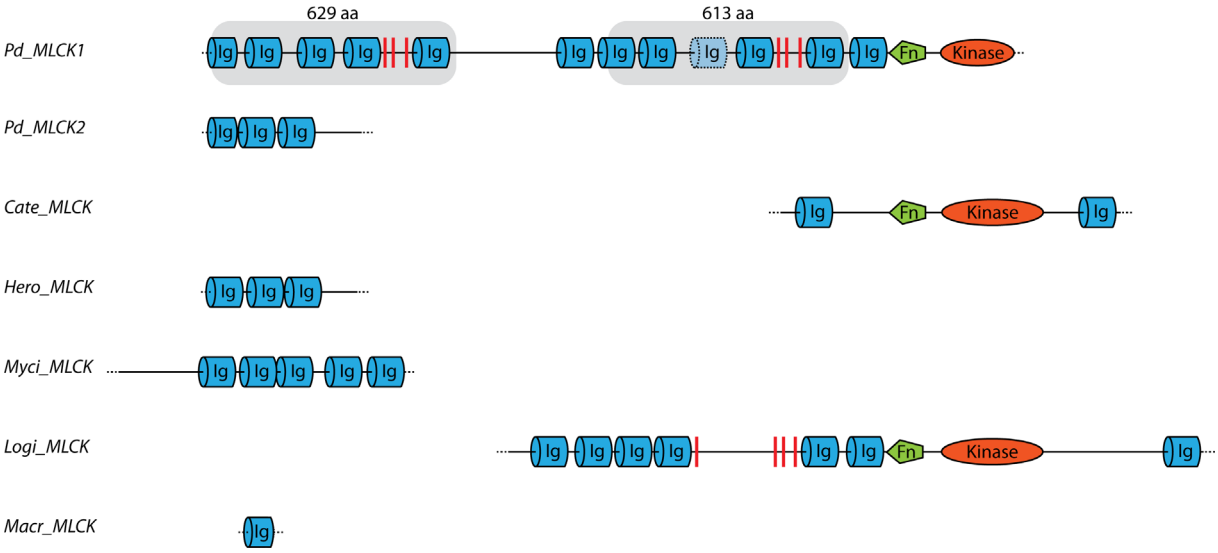


Fig. 35: Domain structure of lophotrochozoan MLCKs. The annotation of putative domains was conducted using the SMART search server. The two MLCKs of *P. dumerilii* (*Pd_MLCK1* and the potential second MLCK *Pd_MLCK2*) are in comparison with MLCKs of *C. teleta* (*Cate_MLCK*) and *Lottia gigantea* (*Logi_MLCK*), as well as potential fragments of *Helobdella robusta* (*Hero_MLCK*), *M. cirriferum* (*Myci_MLCK*) and *M. cranium* (*Macr_MLCK*). blue – Ig domains; green – Fibronectin type-III domain (Fn domain); orange – Protein kinase domain; red vertical lines – actin-binding motifs ("DFRxxL"). Boxed regions appear to be largely identical to each other (97% nucleotide sequence similarity after exclusion of indels).

3.6 A calmodulin-like protein in *P. dumerilii*

BLAST searches with *Pd_CAML* indicated that this gene is similar to the calmodulin proteins. However, a closer analysis of the amino acid sequence information revealed that it is different from "standard" calmodulins (Fig. 36). For evolutionary analysis several other sequences were obtained from GenBank, which likewise are similar to, but do not belong to the "standard" calmodulins. Especially one sequence of *Haliotis* is of interest, as this is a so-called calmodulin-like, potential calcium-regulatory protein (Nikapitiya et al. 2010). The topology of the phylogenetic analysis of EF-hand proteins (Fig. 37) groups *Pd_CAML* within a clade of other annelid calmodulin-like proteins (85% BS support), namely of *Alvinella pompejana* and *Lanice conchilega*. These annelid calmodulin-like proteins group together with the *Haliotis* calmodulin-like proteins (*Hads_CAML1*, *Hads_CAML2*, *Hadv_CAML1 and Hadv_CAML2*), but the BS support is only 37%. As I could obtain "standard" calmodulins from almost all metazoan clades, it became apparent that these calmodulins feature high amino acid sequence conservation (see very short branch lengths within the phylogenetic tree). In contrast to this, all calmodulin-like sequences (also from *M. cirriferum* and *Monezia expansa*) are more variable and differ to a certain degree from the "standard" calmodulins.

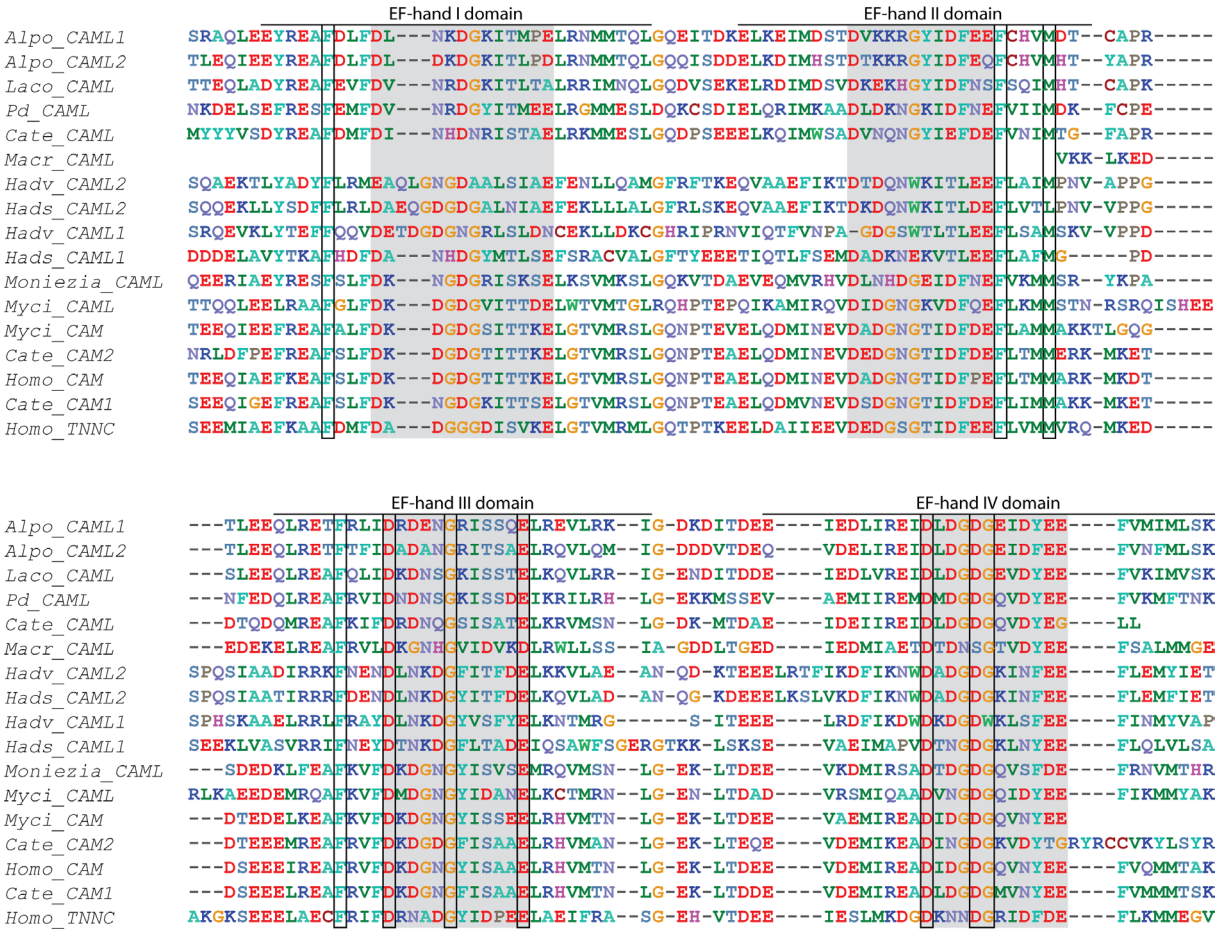


Fig. 36: Alignment of the EF-hand domain region of calmodulins (CAM), calmodulin-like (CAML) proteins and troponin C (TNNC). EF-hand motifs are highlighted in grey, whereas conserved amino acid postions are boxed.

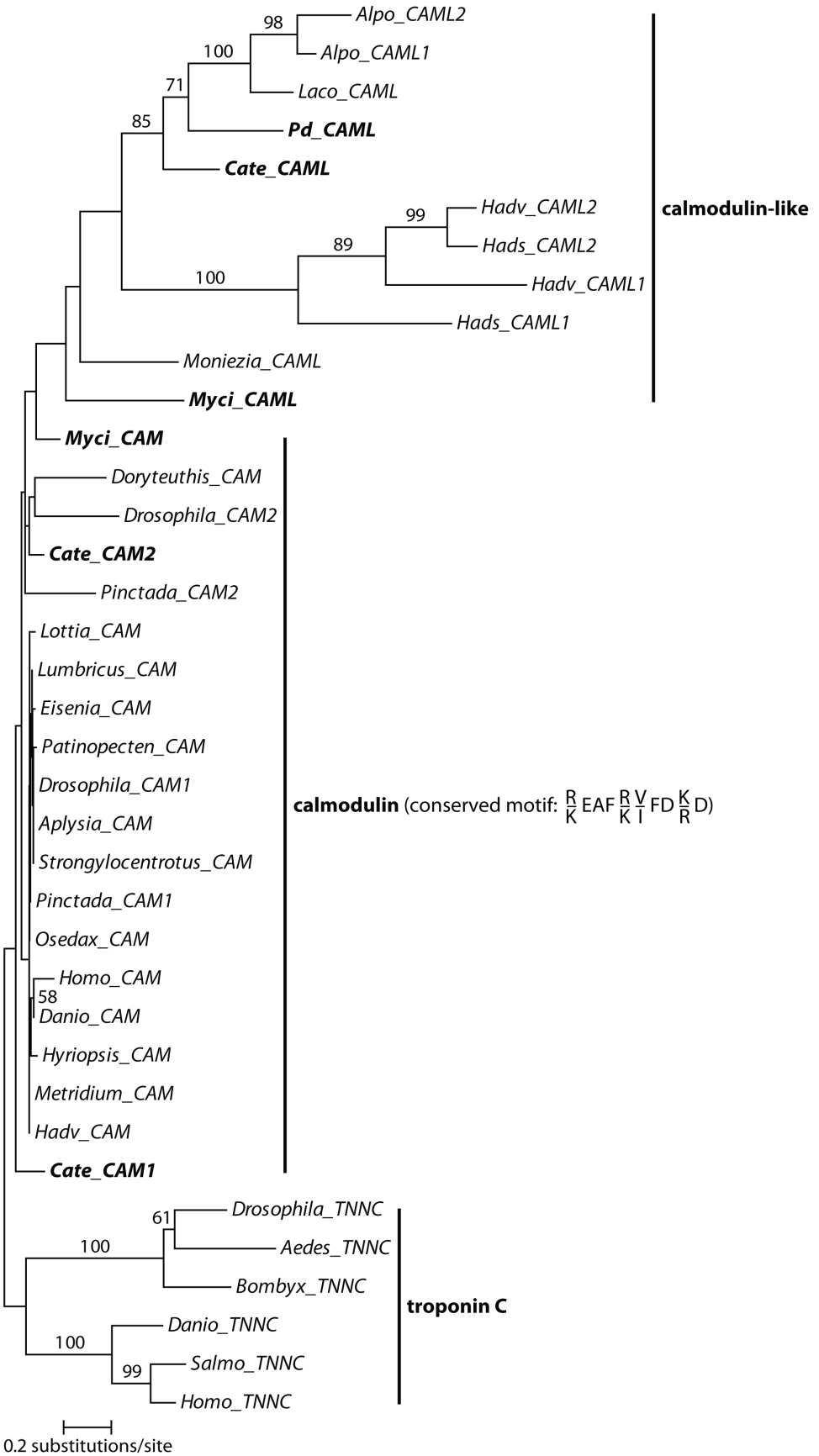


Fig. 37: Evolutionary tree of the EF-hand proteins CAM and CAML. The ML tree (RAxML, RTREVF model) is derived from an alignment (169 amino acid positions) of 37 calmodulin and calmodulin-like sequences including troponin C as an outgroup. Bootstrap values above 50% are shown. Study organisms are highlighted in bold. As the fragment of *M. cranium* (*Macr_CAML*) was too short, it was excluded from the analysis. See Appendix for the sequence alignment.

3.7 A stem cell antigen 2-like protein in *P. dumerilii*

As previously stated, *Pd_SCA2L* contains an N-terminal signal peptide motif as well as a C-terminal transmembrane domain, and shows moderate sequence similarity to a vertebrate stem cell antigen 2-like (SCA2L) protein (*Notophthalmus viridescens*, GenBank accession number: GO931319). The annotation of this protein indicated that this gene is similar to the stem cell antigen 2 (SCA2) known from the chicken (*Gallus gallus*), which is characterized by a LU domain (bearing ten conserved cysteine sites at specific positions). Sequence alignment (Fig. 38) of the *Gallus_SCA2* LU domain with *Pd_SCA2L* and all sequences found via BLAST search (exhibiting high to moderate similarity to *Pd_SCA2L*) corroborate that these ten cysteine sites are highly conserved. Notably, even the remaining sequence of the *Gallus_SCA2* LU domain aligns well to all SCA2-like proteins and suggests that this domain feature is – although not predicted by SMART search in most invertebrate taxa – common to SCA2 and the SCA2-like proteins studied here. Additionally, an evolutionary analysis was conducted, but yielded a tree with low resolution (see Appendix, Fig. A5).

Fig. 38: Alignment of the LU domain and adjacent regions of SCA2 and SCA2-like proteins (SCA2L). Conserved cysteine positions are boxed and the boundary of the LU domain is marked by a horizontal line.

3.8 Chaetae-specific genes in other organisms

After analysis of chaetae-specific gene expression in *P. dumerilii*, I searched for similar genes in other chaetae-bearing organisms. One potential candidate species was *Capitella teleta*, as this genome is sequenced. A second candidate was the myzostomid *Myzostoma cirriferum*, of which I could screen transcriptome sequence data provided by C. Bleidorn (University of Leipzig, Germany). Finally the brachiopod *Macandrevia cranium* was selected for further investigations, as own transcriptome data are under construction. Computational approaches revealed several potential orthologs for all selected species.

3.8.1 *Capitella teleta*

BLAST searches against the *C. teleta* genome revealed four chitin synthases (*Cate_CS1* to *Cate_CS4*). Based on the phylogenetic analysis of chitin synthases (Fig. 31), two of the four paralogs are each orthologous to a chitin synthase of another study organism (e.g., *Cate_CS1* to *Macr_CS*). Interestingly, one chitin synthase (*Cate_CS2*) revealed to be orthologous to *Pd_CS1*, which has been shown to be expressed in the chaetoblasts of *P. dumerilii*. In terms of the ambiguity whether or not *Pd_CS1* possesses a myosin motor domain, *Cate_CS2* is of special interest, as this lacks a myosin motor head and it became interesting to prove whether this gene likewise is chaetoblast-specific. PCR yielded an appropriate gene fragment for *Cate_CS2*. A 1.2 kb *in-situ* hybridization probe was transcribed and used for further gene expression studies. Finally, the ISH studies with *Cate_CS2* revealed that the gene expression is indeed chaetae-specific. Closer analysis of the chaetal sacs indicates that the expression is restricted solely to the chaetoblasts, as each ISH signal is located directly at the proximal base of a single chaeta (Fig. 39 B+C). This result shows that the ortholog in *C. teleta* to that of *Pd_CS1* is expressed in the same cell type as described for *Pd_CS1* in *P. dumerilii*. Further potential orthologs were also found for other chaetal sac specific genes of *P. dumerilli*, namely *Pd_fcmg2*, *Pd_cbmg3*, *Pd_CAML*, *Pd_MLCK1* and *Pd_SCA2L*.

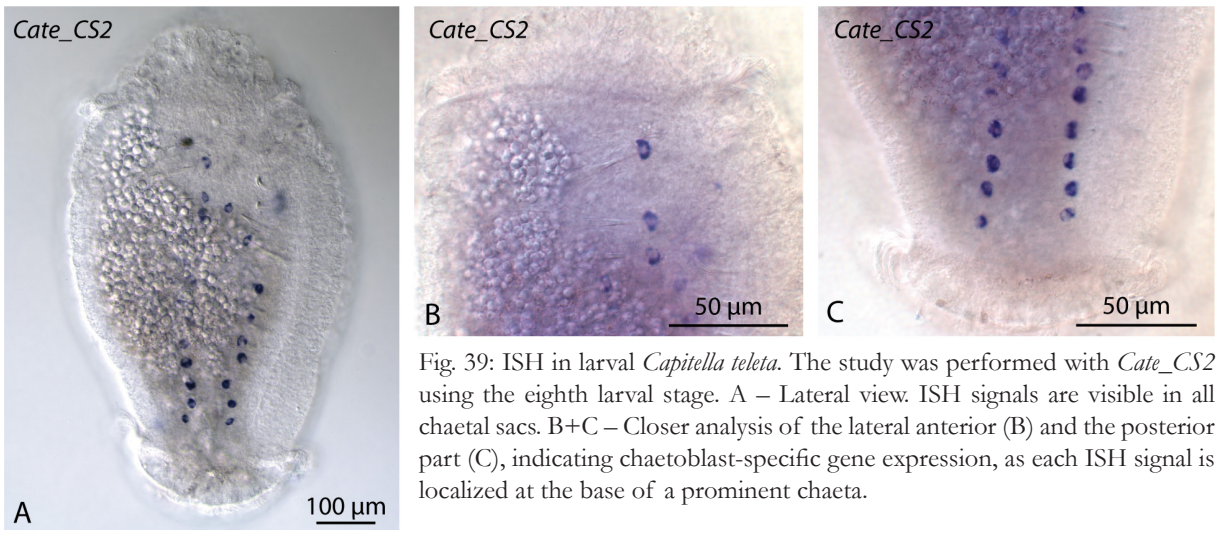


Fig. 39: ISH in larval *Capitella teleta*. The study was performed with *Cate_CS2* using the eighth larval stage. A – Lateral view. ISH signals are visible in all chaetal sacs. B+C – Closer analysis of the lateral anterior (B) and the posterior part (C), indicating chaetoblast-specific gene expression, as each ISH signal is localized at the base of a prominent chaeta.

3.8.2 *Myzostoma cirriferum*

All studies on *M. cirriferum* were performed in collaboration with C. Helm and C. Bleidorn (University of Leipzig, Germany). To identify orthologs of *P. dumerilii* genes, the chaetae-specific *P. dumerilii* clone sequences were BLAST searched (tBLASTx) against the *M. cirriferum* transcriptome library, while in parallel a BLASTn analysis was run against the *P. dumerilii* genome data (to check if the respective sequence is present in the hitherto assembled genomic sequence ressources of *P. dumerilii*). In a second step, the most significant *M. cirriferum* hits were reBLASTed against the genomic data of *P. dumerilii*. Gene orthology was assumed, if the best hit of the *M. cirriferum* reBLAST was the same as the corresponding one hit by the *P. dumerilii* clone sequences itself. Due to this search strategy potential orthologous genes have been found for nine chaetae-specific genes of *P. dumerilii* (*Pd_fcmg2*, *Pd_cbmg3*, *Pd_fcmg4*, *Pd_fcmg5*, *Pd_CAML*, *Pd_CS1*, *Pd_MLCK1*, *Pd_fcmg11*, *Pd_SCA2L*).

Two genes were selected for further ISH analysis to investigate if these orthologs are expressed in the same tissue as known from *P. dumerilii*. However, first of all an ISH protocol had to established, as these results are the first gene expression data ever obtained for myzostomids. This was done in collaboration with C. Helm. As the morphology of myzostomid larvae (Fig. 40 D+E) differ

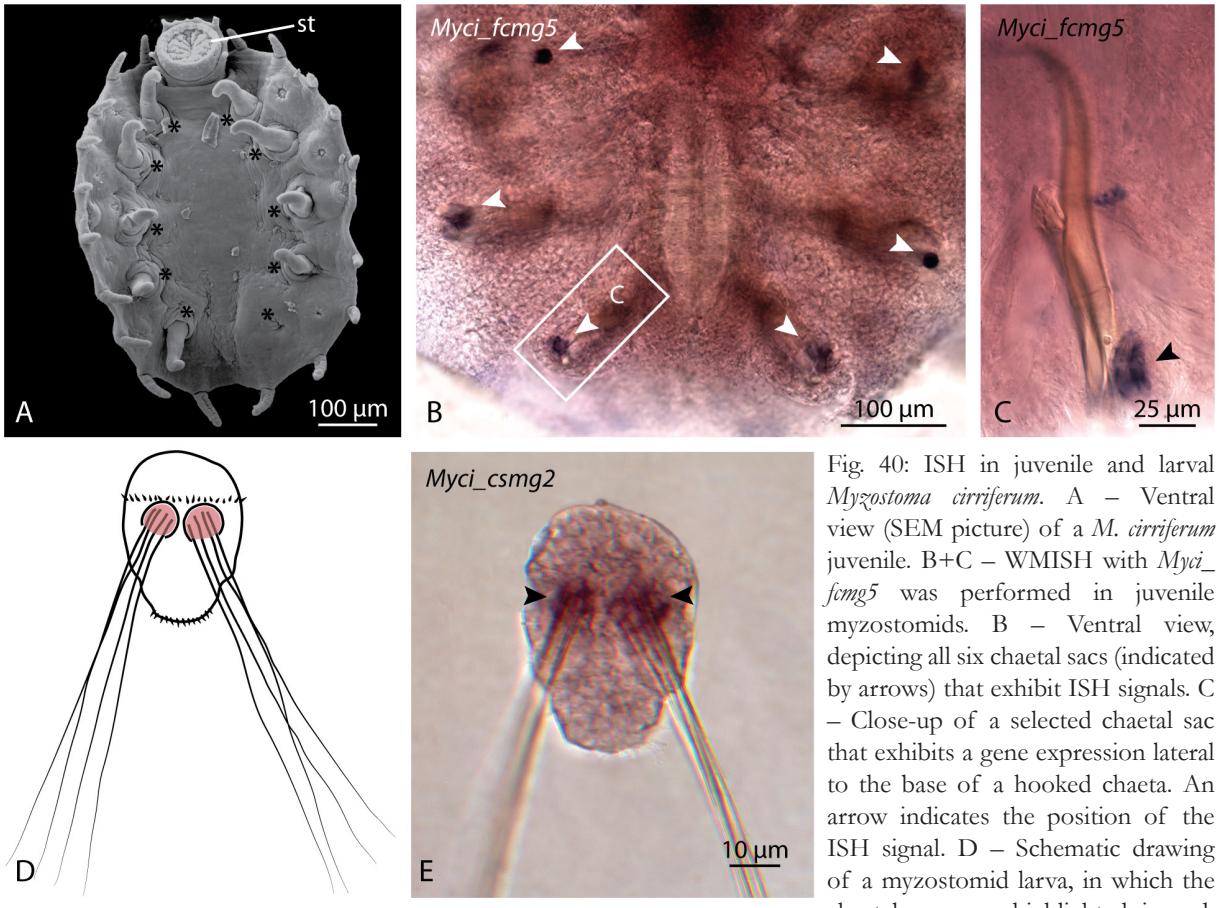


Fig. 40: ISH in juvenile and larval *Myzostoma cirriferum*. A – Ventral view (SEM picture) of a *M. cirriferum* juvenile. B+C – WMISH with *Myci_fcmg5* was performed in juvenile myzostomids. B – Ventral view, depicting all six chaetal sacs (indicated by arrows) that exhibit ISH signals. C – Close-up of a selected chaetal sac that exhibits a gene expression lateral to the base of a hooked chaeta. An arrow indicates the position of the ISH signal. D – Schematic drawing of a myzostomid larva, in which the chaetal sacs are highlighted in red. E – WMISH with *Myci_csmg2* in a myzostomid metatrochophore stage. Arrows mark the chaetal sac position showing ISH signals.

strongly from that of adults (Fig. 40 A) and juveniles (Fig. 40 B), ISHs were performed with juveniles and larvae (if possible). Expression studies with *Myci_fcmg5* (365 nt probe) show that this gene is expressed in a distinct circular pattern, which is visible in each of the chaetal sac regions in 3-segmented juveniles (Fig. 40 B). The shape and size of this pattern suggests that this signal is restricted to a single cell (Fig. 40 C). Notably, the expression seems to be restricted to a proximal cell within the chaetal sac, which does not lies underneath, but laterally of the base of a hooked chaeta. ISH studies with *Myci_csmg2* (178 nt probe) were performed with larval material of the metatrochophora stage (Fig. 40 E). At this stage, the ISH signal is restricted exclusively to the chaetal sacs, but unambiguous cellular localization of the signal is not possible.

3.8.3 *Macandrevia cranium*

Bioinformatic analysis of transcriptome sequences of *M. cranium* were performed based on the same BLAST strategy as described for *M. cirriferum*. This revealed three presumptive orthologs to the genes *Pd_CS1*, *Pd_MLCK1*, and *Pd_CAML* of *P. dumerilii*, of which two were selected for further ISH investigations. In terms of *Macr_MLCK*, BLAST searches against Genbank indicate similarities to known myosin light chain kinases. Due to the evolutionary analyses of myosin light chain kinases (see Appendix, Fig. A4) it became evident that *Macr*_MLCK is orthologous to *Pd_MLCK1*, as both *P. dumerilii* MLCKs cluster within one clade, defining all other myosin light chain kinases as orthologous to *Pd_MLCK1* and *Pd_MLCK2*. SMART domain predictions yielded at least one Ig domain that is a typical domain for MLCKs. The absence of further domain predictions *Macr*_MLCK is most probably due to the short length of the sequence fragment.

ISH studies with *Macr_MLCK* (0.9 kb probe) were performed with larval stages that already possess chaetae (Fig. 41 A+B) and show that the expression is restricted solely to chaetal sacs (Fig. 41 C). For chaetal sacs of larval brachiopods consists only of chaetoblasts underneath epidermal cells (Lüter 2000, 2001), the ISH signals can be allocated to the chaetoblasts of the larval chaetae, as the signal is located at the base of a chaeta (Fig. 41 D). The second candidate gene shows a high sequence similarity to a chitin synthase, whereas the evolutionary analysis of chitin synthases (Fig. 31) supports this fact and suggests an orthology of *Macr_CS* to *Cate_CS1* of C. *teleta*. Additional ISHs with a 3.2 kb probe of *Macr_CS* indicate that this gene likewise is chaetae-specific, as its expression is restricted to proximal-most chaetal sac regions where the chaetoblasts are located (Fig. 41 E+F). The coexpression of these two genes suggests a potential interactive role of the MLCK with that of the chitin synthase within the chaetoblast.

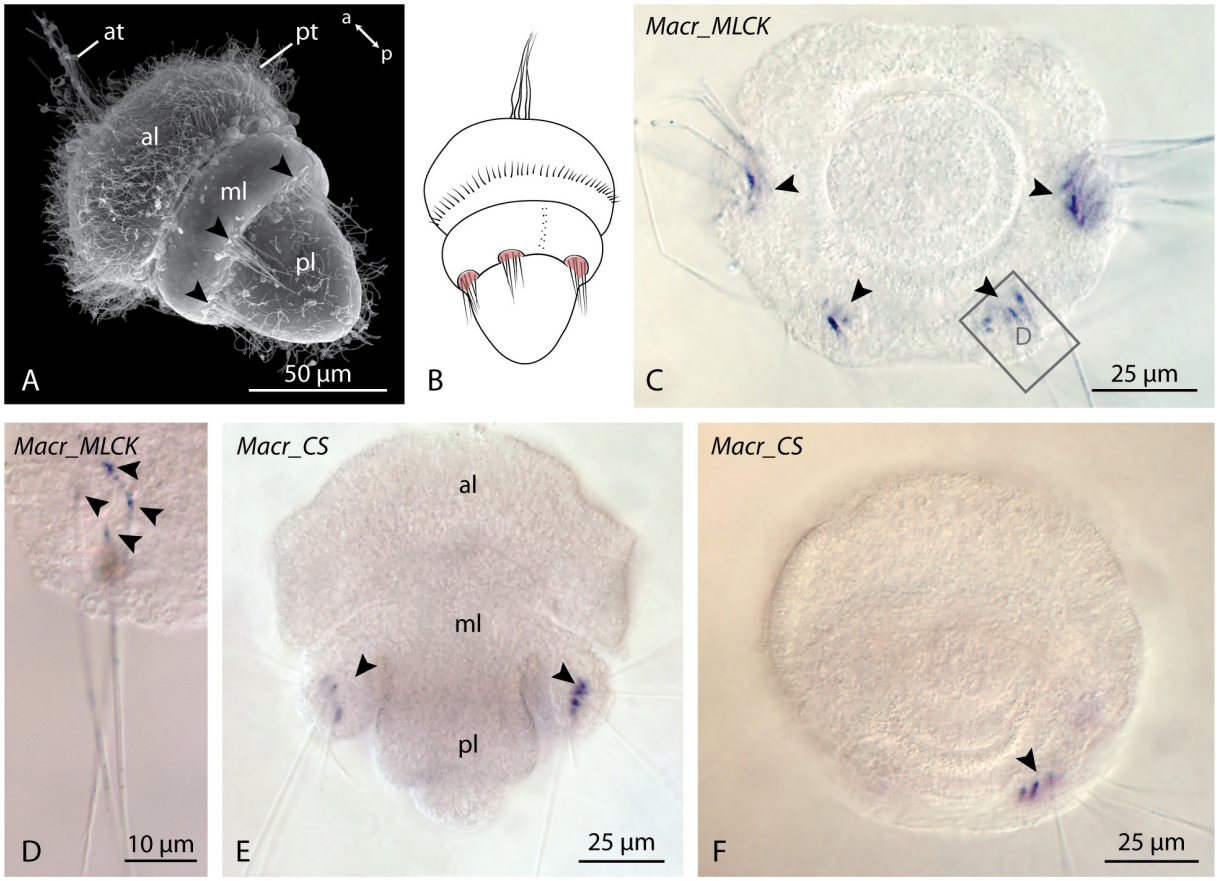


Fig. 41: ISH in larval *Macandrevia cranium*. A – Dorso-lateral view (SEM picture) of a *M. cranium* larva (114 hpf). B – Schematic drawing of an *M. cranium* larva (chaetal sacs are highlighted in red). C-F – WMISH in larval stages (132 hpf) with *Macr_MLCK* (C+D) and *Macr_CS* (E+F). C – Posterior view onto the pedicle and mantel lobe, depicting all four chaetal sacs that exhibit ISH signals. D – Detailed, lateral view of a selected chaetal sac. Arrows indicate the basal position of the ISH signals. E – Ventral view of a brachiopod larva. Arrows mark the chaetal sac position showing ISH signals. F – Posterior view. Not all four chaetal sacs are located within the focal plane. Gene expression (indicated by an arrow) is restricted to the posterior-most part of the chaetal sac.

4 Discussion

This study combines structural and molecular methods to gain new data on chaetae formation in annelids and other chaetae-bearing taxa. It becomes evident that the cells involved in chaetogenesis can be characterized by specific sets of molecular markers. The results provide new levels of comparison and thus insights into the evolution of chaetogenesis-specific genes, functional aspects of chaetae formation, evolution of cell types involved in chaetogenesis and of chaetae in general.

4.1 Structural and molecular characterization of cell types involved in *Platynereis dumerilii* chaetogenesis

According to Arendt (2003), a cell type is a number of cells that express the same orthologous genes. This cell type-specific set of genes is defined as "molecular fingerprint" (Arendt 2005). Molecular fingerprinting is widely used in interpreting cell type evolution, especially the evolution of photoreceptor cells (Arendt 2003, Arendt et al. 2004, Suga et al. 2008, Ullrich-Lüter et al. 2011). In terms of chaetogenesis, such a comparison of "molecular fingerprints" was not yet possible due to the lack of molecular data, but might be valuable in deciding pro or contra a homology of chaetae forming cells, possibly even between distantly related taxa. As the shape of chaetae is species-specific in annelids, Hausen (2005) suggests that chaetogenesis must be under strict genetic control, specifying how a certain chaeta is formed, which makes this process an ideal candidate to study cell type evolution. Based on a random ISH screen in *Platynereis dumerilii* done by D. Arendt and colleagues (unpublished data; EMBL Heidelberg, Germany), I obtained a set of fifteen genes that appeared to be part of the chaetae formation, as they were shown to be expressed in the chaetal sacs. Previous studies suggested that several cells (one chaetoblast and several follicle cells) are involved in the formation of a single chaeta (Bouligand 1967, O'Clair & Cloney 1974), thus indicating a large amount of cells within a chaetal sac, which is the sum of all chaetal follicles of one bundle of chaetae. The ultrastructural analyses performed in this study support this notion, pointing towards the upcoming problem to identify single cells. Only the combination of ISH with immunhistochemistry (using WGA in FISH for co-localization purposes) and especially the double FISH (based on cell-specific markers) made it possible to localize the expression of ten of the fifteen candidate genes on a cellular level. As a result, two candidate genes show an expression restricted to chaetoblasts (*Pd_cbmg3*, *Pd_CS1*), in contrast to nine genes that are only expressed in follicle cells (*Pd_fcmg2*, *Pd_fcmg4*, *Pd_fcmg5*, *Pd_CAML*, *Pd_fcmg11*, *Pd_fcmg13*, *Pd_MLCK1*). These results reveal that at least the chaetoblast is characterized by an own set of genes (*Pd_cbmg3*, *Pd_CS1*), but also by its particular position at the base of a chaeta and, due to its specific morphology of an apical microvilli surface, it can be interpreted as an own cell type. Furthermore, there is evidence that a second cell type can be defined by *Pd_fcmg13* that is specifically expressed in follicle cells one and two which are located apically of the chaetoblast. Results of the ultrastructural investigations support this view, as these two cells differ from the remaining follicle cells, as they oppose each other and together surround the base of the growing chaeta; this is in contrast to the other follicle cells, which singly enwrap the chaetae. Finally, both follicle cells three and four may belong to a

third cell type, characterized by the expression of *Pd_fcmg5*, but as it was difficult to unambiguously relate the expression signals of several genes to either follicle cell three or follicle cell four, it remains unclear whether both cells are of the same or of two different cell types. Interestingly, one gene (*Pd_csmg10*) was found that seems to represent an intermediate situation, as its expression is exhibited by chaetoblasts and proximal-most follicle cells, i.e., cell type I and II, respectively. This may hint to common ancestry. All cells of a chaetal follicle are of epidermal origin (Bouligand 1967), no matter if they become a follicle cell or a chaetoblast. It is thus conceivable that there was once a chaetae-bearing ancestor in which the chaetal sac was not characterized by different cell types.

As the above data unveiled the presence of specific molecular fingerprints for the different cells of a chaetal follicle, it is important to characterize the sequence of such marker genes in terms of sequence similarity to known genes of other model organisms and in terms of sequence features (e.g., domains, binding sites, motifs and duplications). Since this current study presents the first molecular and gene expression data about chaetogenesis, I did not expect to find much information on their structure and potential function, given the fact that on one hand the *P. dumerilii* genomic and transcriptomic resources are still under construction, and on the other hand all established and well-investigated model organisms, like *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus* and *Danio rerio*, do not possess chaetae and are therefore unlikely to exhibit the relevant genes. In addition, most of the annotations of the upcoming lophtrochozoan data are likewise based on investigations performed on model organisms of Ecdysozoa and Deuterostomia (Tessmar-Raible & Arendt 2003).

Nevertheless, in the present study I was able to identify fifteen molecular candidate genes (Tab. 4) that are highly specific to chaetal sacs, of which only one (*Pd_fcmg18*) could not be further characterized, as investigations yielded neither similarity to other organisms nor special structural sequence characteristics.

Even though the genes *Pd_csmg10*, *Pd_fcmg13*, *Pd_csmg15* and *Pd_csmg21* did not yield similarity to any sequenced gene, they can be further characterized by a structural domain – a transmembrane domain –, whereas *Pd_csmg21* additionally contains a signal anchor motif and *Pd_fcmg1* an N-terminal signal peptide.

Better to characterize are the five genes *Pd_fcmg2*, *Pd_cbmg3*, *Pd_fcmg4*, *Pd_fcmg5* and *Pd_fcmg11*, as their sequences show similarities to other organisms; strikingly Annelida. However, as these hits are not yet annotated, the characterization is mainly based on the fact that they all contain several conserved cysteine sites, an N-terminal signal peptide motif (not in *Pd_fcmg5*) and transmembrane domains in *Pd_cbmg3* and *Pd_fcmg5*. In addition, *Pd_fcmg2* contains a PAN domain that is suggested to be of functional importance as these domains act in diverse biological functions, such as mediating protein-protein or protein-carbohydrate interactions.

Interestingly, two of the fifteen genes (*Pd_CAML* and *Pd_SCA2L*) showed sequence similarity not only to closely related organisms (e.g., annelids), but also to major metazoan taxa, including

Lophotrochozoa. As the hits comprise even vertebrates, there is a better chance to find already annotated genes, as for example *Pd_CAML*, which is suggested to be similar to calmodulin proteins. Four EF-hand motifs have been predicted for the amino acid sequence of *Pd_CAML*, which are typical features of calmodulins that act in processes such as cell signaling in general (Lewit-Bentley & Réty 2000) and the regulation of calcium flux for shell formation (Rousseau et al. 2003, Li et al. 2004, Nikapitiya et al. 2010), in particular. Even though *Pd_CAML* shows homology to calmodulins, its amino acid sequence was not as highly conserved compared to the usual calmodulins. In contrast, the amino acid sequence of *Pd_CAML* is closely related to other annelid calmodulin-like proteins, as well as to calcium-regulatory proteins in *Haliotis* (Nikapitiya et al. 2010), than to the slightly shorter (148-149 amino acid length) and highly conserved "standard" calmodulins. On the other hand, *Pd_SCA2L* was characterized as being similar to the stem cell antigen 2 (SCA2) in *Gallus gallus*. This protein is member of the Ly-6 family, a group of small cysteine-rich (LU domain) cell surface proteins (Classon & Coverdale 1994) that are suggested to mediate cell-cell adhesion (Bamezai & Rock 1995). Even though domain predictions lack information about the LU domain (containing ten conserved cysteine sites), it was found that there is high sequence similarity in terms of the position of the cysteine sites. Additionally, *Pd_SCA2L* contains an N-terminal signal peptide motif and a transmembrane domain.

Finally, two of the fifteen investigated *P. dumerilii* genes can be found in all studied organisms (comprising *Capitella teleta*, *Myzostoma cirriferum*, *Macandrevia cranium*), as well as in other Lophotrochozoa (comprising molluscs and annelids) and "non-lophotrochozoans". Due to the high number of similar sequences, I was able to perform structural characterizations as well as evolutionary analyses to reveal information about the gene evolution of *Pd_CS1* and *Pd_MLCK1*. *Pd_MLCK1* shows sequence similarity to a wide range of taxa, including all studied organisms, other Lophotrochozoans as well as representatives of Ecdysozoa, Echinodermata and Vertebrata. The similarity includes even structural characteristics, like Ig domains, a kinase domain and an Fn domain, which are typical features of myosin light chain kinases. The last remaining gene (*Pd_CS*) is even more interesting, as it is highly similar to a chitin synthase, additionally characterized by several transmembrane domains and most importantly by a chitin synthase domain. In the context of chaetogenesis, which is the main focus of this study, it is apparent that this gene is highly interesting, as annelid chaetae are (β-) chitinous structures. Previous investigators suggested that the chitin synthesis must be located in the chaetoblast (Bouligand 1967, O'Clair & Cloney 1974), but until now no evidence was presented for such a hypothesis. This current study for the first time gives evidence that the chitin synthesis indeed takes place in the chaetoblast, as data obtained by double FISH show a complete overlap of the *Pd_CS1* ISH signal with that of the chaetoblast marker *Pd_cbmg3*. As *Pd_CS1* and *Pd_MLCK1* are suggested to be of special interest in terms of chaetae formation, their data will be discussed in separate chapters (see 4.3 and 4.4).

In terms of chaetae formation, several of the above stated structural characteristics might be very important, in addition to the better characterized genes (*Pd_CAML*, *Pd_CS1*, *Pd_MLCK1*, *Pd_SCA2L*), as for example a protein that contains a PAN domain is conceivable to act in tanning

processes at the base of the growing chaeta, where new material is added and has to be cross-linked. Genes that are characterized by signal peptide motifs might also be of interest, as these motifs are targeting signals that enable cellular trafficking to correctly place a protein extra- or intracellularly (Blobel & Sabatini 1971). Interestingly, all of the signal peptide-containing genes (*Pd*_cfmg2, *Pd*_cbmg3, *Pd*_fcmg4, Pd_*fcmg5*, *Pd_fcmg11* and *Pd_SCA2L*) and *Pd_fcmg5* show a high conservation in terms of cysteine sites, which are suggested to play an important role in many structural proteins (e.g., crosslinking of proteins) due to their reactive thiol group (Bulaj et al. 1998). On the other hand, transmembrane helices typically are able to span an entire biological membrane, which makes transmembrane proteins ideal candidates to act in or at cell membranes. In terms of one single gene candidate (*Pd_csmg21*), a signal-anchor motif was predicted next to the presence of a transmembrane domain. Signal-anchors are transmembrane proteins belonging to the type II membrane proteins, which possess a single transmembrane domain, and are assumed to act in protein translocation, especially in interaction with the endoplasmatic reticulum (Sakaguchi et al. 1992). Nonetheless, even though at present the function of all the genes is unknown, the domain prediction revealed several structural and functional elements that indicate the potential these genes might have for future studies.

Tab. 4: Structural characteristics of the fifteen *Platynereis dumerilii* chaetal sac markers and their homologs in other organisms.

***P. dumerilii* chaeto-genesis marker**	**homologous sequences in other organisms**					**structural sequence characteristics**			
	C. teleta	***M. cirriferum***	***M. cranium***	**other Lophotrochozoa**	**"non-Lophotrochozoa"**	**SP**	**TMD**	**CCS**	**other structures**
Pd_fcmg1						X			
Pd_fcmg2	X	X		*Lumbricus rubellus*		X		X	PAN domain
Pd_cbmg3	X	X		*Alvinella pompejana, Lanice conchilega*		X	X	X	
Pd_fcmg4		X				X		X	
Pd_fcmg5		X		*Ridgeia piscesae*			X	X	
Pd_CAML	X	X	X	CAML in *Alvinella pompejana, Lanice conchilega, Haliotis* spp., *Moniezia expansa;* CAM in Lophotrochozoa	CAM in Cnidaria, Ecdysozoa, Echinodermata and Vertebrata				EF-hand motifs
Pd_CS1	X	X	X	*Helobdella robusta, Atrina rigida, Pinctada fucata, Mytilus galloprovincialis, Lottia gigantea*	Porifera, Cnidaria, Ecdysozoa, Chordata		X		chitin synthase domain
Pd_MLCK1	X	X	X	*Helobdella robusta, Lottia gigantea*	Ecdysozoa, Echinodermata, Vertebrata				Ig domains, Fn domain, kinase domain, actin-binding motifs
Pd_csmg10							X		
Pd_fcmg11		X				X		X	
Pd_fcmg13							X		
Pd_csmg15							X		
Pd_fcmg18									
Pd_SCA2L	X	X		*Perinereis nuntia, Alvinella pompejana, Crassostrea gigas, Schistosoma mansoni*	SCA2L in Insecta, Echinodermata, Hemichordata, Vertebrata; SCA2 in *Gallus gallus*	X	X	X	LU domain
Pd_csmg21							X		signal anchor

Abbreviations: SP – signal peptide, TMD – transmembrane domain, CCS – conserved cysteine sites

4.2 Temporal dynamics in gene activity during chaetogenesis

This study also presents dynamics in gene activity, as different developmental stages (ranging from 40-hpf larvae to worm stage) for some of the genes showed that not all of them are active at the same time. Rather, several of the investigated genes seem to work successively. During my ISH screening, I found that, for example, the genes *Pd_fcmg11*, *Pd_fcmg13*, *Pd_SCA2L* as well as *Pd_csmg21* are the first genes that are detectable. For all of them, gene expression was not present in screened 36-hpf stages but strong ISH signals were yielded in experiments on 40 hpf old larvae. Also the chitin synthase *Pd_CS1* belongs to these „early genes". Additionally, all investigated 40-hpf stages were characterized by strong gene expressions, which indicate that during the earliest steps of chaetogenesis these genes must be highly abundant, especially considering that they were activated within the first four hours of chaetal follicle development. On the other hand, some genes are suggested to be activated subsequently, as their gene expression is only present in screened 48-hpf and 72-hpf stages (*Pd_fcmg4*, *Pd_csmg15*). Additionally, some evidence was obtained (indicated by ISH signal intensity) suggesting a possible up- and down-regulation of gene activity. While using the same probe concentration per candidate gene and corresponding stage, some screened stages exhibit increased or decreased signal intensities per setiger when compared to younger or older stages (*Pd_fcmg1*, *Pd_fcmg2*, *Pd_cbmg3*, *Pd_fcmg5*, *Pd_fcmg13*, *Pd_SCA2L* and *Pd_csmg21*). For example, *Pd_csmg21* exhibits strong gene expression even in 40-hpf stages, when the second setiger is not completely differentiated yet. In 48-hpf stages, expression is still mainly restricted to the setiger one and two, though weak ISH signals can be detected in the third setiger (the "youngest", but yet well-developed one). When finally this state is compared with the situation in 72-hpf stages, it becomes apparent that the intensity of the ISH signal of the first setiger has decreased, whereas its intensity stayed constant in the second and is even increased within the third one (if compared to 48 hpf old larvae). Additionally, there are some genes whose gene expression appears to be strong in nearly all screened stages (*Pd_CS1*, *Pd_csmg10*, *Pd_fcmg11*, *Pd_fcmg13*, *Pd_SCA2L* and *Pd_csmg21*). This group seems to resemble candidate genes which have to be expressed during all stages of chaetogenesis. Finally, further studies have to be performed to yield information about how these genes interact to orchestrate chaetae formation.

4.3 Evolution of chitin synthases and their possible function

The data of the present study provide comprehensive insights into the diversity of chitin synthases, whose previous investigations mainly focused on model organisms in Nematoda (e.g., Veronico et al. 2001), Arthropoda (e.g., Zimoch & Merzendorfer 2002, Hogenkamp et al. 2005, Merzendorfer 2006) and Fungi (e.g., Roncero 2002, Ruiz-Herrera & Ortiz-Castellanos 2010). I found chitin synthases in other metazoan taxa, including the first evidence for the presence of chitin synthases in annelids, namely three in *P. dumerilii* (*Pd_CS1* to *Pd_CS3*), four in *C. teleta* (*Cate_CS1* to *Cate_CS4*), two in *Helobdella robusta* (*Hero_CS1* + *Hero_CS2*) and one in *M. cirriferum* (*Myci_CS*). In addition,

even low-coverage transcriptome sequencing of *M. cranium* revealed the first known brachiopod chitin synthase. Although recent studies in molluscs described one chitin synthase in *Atrina rigida*, an ortholog in *Mytilus galloprovincialis* (Weiss et al. 2006) and one in *Pinctada fucata* (Suzuki et al. 2007), I found nine chitin synthases in total in the genome data of *Lottia gigantea*. Notably, this number is even higher than the large number of chitin synthases in *Aspergillus fumigatus*, namely seven (Roncero 2002). Furthermore, in computational searches I found chitin synthases for three groups, namely Porifera (*Amphimedon queenslandica*), Cnidaria (*Nematostella vectensis*, *Hydra magnipapillata*) and Chordata (*Ciona intestinalis*, *Branchiostoma floridae*, *Danio rerio* and *Xenopus tropicalis*), although until now, these animals have been suggested to not contain chitin at all (except for a few cnidarians: hydrozoan polyps, e.g., *H. magnipapillata*, and the pneumatophores of Siphonophora colonies; Rudall 1955).

The evolutionary analyses of the most conserved region of chitin synthases (CS domain) revealed that, when using all "classes" (*sensu* Roncero 2002) of fungal chitin synthases as an outgroup, the above mentioned metazoan chitin synthases have one common ancestor and are therefore monophyletic. Interestingly and in congruence with previous studies that sampled only nematodes and insects for Metazoa (Roncero 2002, Ruiz-Herrera et al. 2002), the fungal chitin synthases of "class" IV and "class" V (MMD-bearing) represent the sister group to this clade. Within the large group of metazoan chitin synthases, the many chitin synthases of molluscs and of annelids each form four distinct groups. It is worth noting that in all four cases, each group of annelid chitin synthases clusters together with one of the four molluscan groups (one group including the chitin synthase of *M. cranium* and one including the *M. cirriferum* chitin synthase). As it was stated for fungi ("class" V), an MMD was also found in the majority of the sampled lophotrochozoan groups, but with a restriction to three (group A to C) of the four clades. In contrast, the broad taxon sampling puts the MMD-lacking chitin synthases (group D) into close relationship to a part of the two chitin synthases known from Ecdysozoa (whereas another nematode chitin synthase group emerges from a basal bilaterian polytomy). As a consequence, the hitherto assumed duplication event (Zhu et al. 2002) that led to the two ecdysozoan chitin synthases did not happen in the last common ancestor of Ecdysozoa, but must have occurred at least within the last common ancestor of Protostomia or even Bilateria. In contrast to this, all three remaining groups (group A + B + C, each comprising only lophotrochozoan taxa) are monophyletic as a whole and group A + B are sister to group C. Consequently, in the ancestor of the sampled lophotrochozoans, a first duplication led to the ancestor of group C and of the groups A + B, followed by a second duplication (of A + B) that yielded the ancestors of group A and group B. The last common ancestor of annelids, myzostomids, brachiopods and molluscs thus possessed one chitin synthase without an MMD and three MMD-bearing chitin synthases.

As shown above, the minimum set of chitin synthases within a chitin-secreting organism appears to be at least two paralogous proteins, but their origins within the diversity of chitin synthases is caused by several independent duplication events of the chitin synthase gene (e.g., in the fungal lineage, in the ecdysozoan lineage, in the lophotrochozoan lineage). Successful duplication events,

leading to increased chitin synthase diversity within an organism after the retention of both daughter genes (paralogs), provide the potential for recruiting each of the paralogous genes for specialized or even different functions (Roncero 2002). This fact is at least supported by functional investigations in insects (Zhu et al. 2002, Zimoch & Merzendorfer 2002, Hogenkamp et al. 2005, Merzendorfer 2006), in which two chitin synthase genes are present in two different cell types, but each achieving a different function. One of the genes (e.g., *Mase_CS1*) is active in epidermal cells having a function in cuticle formation (molting processes), whereas the other one (e.g., *Mase_CS2*) is specific to endodermal cells of the midgut (for formation of peritrophic matrices; Ibrahim et al. 2000). These findings are additionally supported by studies in Nematoda, as here one chitin synthase gene (e.g., *Cael_CS1,* Nematoda group B) is involved in oogenesis and eggshell formation, as well as in embryonic development (Harris et al. 2000), whereas the gene expression studies of the second gene (e.g., *Cael_CS2,* Nematoda group A) reveal that it is exclusively expressed within the pharynx (Zhang et al. 2005). Since in both cases the investigated chitin synthase genes exhibited cell type specificity in addition to specific temporal activity during development, such a situation can be assumed for the different chitin synthases present in the ISH-screened organisms of this study. As I was able to localize the expression of two orthologous chitin synthases of the annelid group C (*Pd_CS1* + *Cate_CS2*), it is now clear that they are restricted to the chaetoblast. As annelid chaetae, in contrast to larval brachiopod chaetae (Lüter 2000a and 2001), possess a coating layer (Ebling 1945) and protrude from a chaetal sac with follicle cells, it is conceivable that at least one of the remaining chitin synthases in *P. dumerilii* (*Pd_CS2*, *Pd_CS3*) and *C. teleta* (*Cate_CS1*, *Cate_CS3*, *Cate_CS4*) is localized in the follicle cells to serve for the coating function during chaetogenesis. Localization of the expression of a paralog of this chitin synthase group in *M. cranium* (*Macr_CS*, sister to annelid group A, i.e., *Cate_CS1*) revealed that this gene is expressed in the same cell type, which is the only cell that constitutes a larval brachiopod chaetal sac (Lüter 2000a and 2001). It seems likely that once a higher coverage of the *M. cranium* transcriptome has been achieved, more chitin synthases will be found and that they may also comprise orthologs of *Pd_CS1* and *Cate_CS2*.

All here ISH-investigated chitin synthases (*P. dumerilii*, *C. teleta*, *M. cranium*) belong to the MMD-bearing clade, even though, as not all are represented by full-length sequences, not all of them contain information about their N-terminal region. Nevertheless, the topology of the chitin synthase domain tree argues for their position within the MMD-bearing clade and is additionally supported by the presence of the "bridge" region within the chitin synthase domain. This implies that full-length sequence information may either uncover MMDs in these chitin synthases or show that the MMDs were lost secondarily, then raising questions about functional aspects of the MMDs.

The results of this study corroborate that the MMD-linkage of the chitin synthases of Fungi ("class" V) and all sampled Lophotrochozoa evolved twice independently, since in evolutionary analyses of chitin synthase domains both clades are distantly related. Additionally, the evolutionary analysis of myosin motor heads revealed that the MMD of lophotrochozoans is closely related to "class" III myosins, which group far away from the fungal "class" XVII MMDs (Odronitz & Kollmar 2007). As such duplication and translocation events, leading to a successful linkage of

a myosin motor head with a chitin synthase, have to maintain their functionality, which includes shared promoters and shared splicing, they tend to be rare (Cope et al. 1996). In Fungi, in which such an event apparently occurred once (Roncero 2002), MMDs linked to a chitin synthase ("class" V) became functionally very important. In general, MMDs of fungi "class" V chitin synthases interact with the actin cytoskeleton in hyphal tips, while deletion or point mutation experiments in the MMD yielded loss of chitin synthase functionality (Horiuchi et al. 1999, Steinberg 2000, Takeshita et al. 2005). In addition, MMD-linked chitin synthases of parasitic Fungi are necessary for their pathogenic potential, as they conduct cell wall modification of the hyphal tip after entering the epidermis of the host (Madrid et al. 2002, Liu et al. 2004, Weber et al. 2006, Werner et al. 2007). Further evidence for the importance of such an MMD in interacting with elements of the cytoskeleton was also found in a study on the bivalve mollusc *Atrina rigida* (Weiss et al. 2006), which was the hitherto first record of an invertebrate chitin synthase containing an MMD. In their study, Weiss et al. (2006) suggest that the MMD might likewise interact with cytoskeletal components and thus participate in mollusc shell formation. Having in mind that the chaetoblast in annelids is characterized by a prominent microvilli pattern at its apical surface (Bouligand 1967, O'Clair & Cloney 1974), an adapted model may be useful to explain the fine-tuned process of chaetae formation. Microvilli bear distinct F-actin filaments that interact in a filamentous network, thus, it can be hypothesized that in chaetogenesis the chitin synthase myosin motor head binds to the actin filament network, in a way comparable to the above mentioned situation in Fungi and molluscs. This may facilitate effective recruitment of the enzyme to the microvillar surface of the chaetoblast. In turn, this indicates that the chitin synthase (being a transmembrane protein) is not only integrated into any membrane, but specifically into the membrane of the microvilli of the chaetoblast. Here, the enzyme catalyses the chemical reaction, in which UDP-N-acetyl-D-glucosamine (UDP-GlcNAc) is utilized as an activated sugar donor to produce the growing chitin polymer (Glaser & Brown 1957). As I was able to show, *M. cranium* possesses an MMD-linked chitin synthase (*Macr_CS*) whose expression is restricted to the chaetoblast and therefore suggests a potentially interactive role during chaetogenesis between the MMD and the apical actin-filament cytoskeleton.

In addition to the four study organisms, this analysis shows that there are chitin synthases in non-chaetae-bearing lophotrochozoans that are similar to the cell type-specific chitin synthases involved in annelid and brachiopod chaetogenesis, i.e., Mollusca group A is closely related to the brachiopod chaetoblast-specific chitin synthase (*Macr_CS*) and Mollusca group C is sister to the two annelid chaetoblast-specific chitin synthases (*Pd_CS1* and *Cate_CS2*). Beedham & Trueman (1968) and Haszprunar (1992), respectively, assumed chitin being present in several clades of Mollusca, thus suggesting that the ancestral state of molluscs might have been a cuticle without calcified parts, but consisting mainly of proteins, mucopolysaccharides and chitin (Peters 1972). Gene expression studies of Weiss et al. (2006) showed that the chitin synthase of *M. galloprovincialis* (*Myga_CS*, grouped within Mollusca group A in the present study) is expressed in shell-associated tissue. There are no other records of chitin synthases in *M. galloprovincialis* that could have been tested,

thus leaving the question about the presumptive function of the other mollusc chitin synthases unresolved. Notably, four of the nine chitin synthases of *L. gigantea* belong to Mollusca group D and neither feature an MMD nor are related to the MMD-bearing clade, but are more closely related to a part of the ecdysozoan chitin synthases. So far, group D chitin synthases of molluscs and annelids have neither been studied on a functional level nor via gene expression analyses. However, in this context *H. robusta* might be interesting, as this annelid belongs to the hirudinean subtaxon of Clitellata (Struck et al. 2011) and thus has secondarily lost chaetae. Both chitin synthases of *H. robusta* group in close relationship to the primarily MMD-lacking chitin synthases (group D) of annelids and molluscs (more specifically, close to *Cate_CS3* + *Cate_CS4* of *C. teleta*), and thus do not belong to the set of genes (*Pd_CS1* und *Cate_CS2* und *Macr_CS*) for which chaetoblast-specific gene expression is proven.

The investigations presented here revealed that to possess more than one chitin synthase is not untypical among chitin-secreting organisms, which was even shown for the relatively low-coverage genomic and transcriptomic sequence data of *P. dumerilii*. This certainly raises further questions about possible functional aspects and the presumptive complexity of the mechanism of chaetae formation itself. As I was able to pinpoint the expression of the brachiopod chitin synthase *Macr_CS* to the chaetoblast, as well as the expressions of the annelid chitin synthases *Pd_CS1* and *Cate_CS2*, it appears that at least two of the chitin synthase groups (A and C) of the sampled chaetae-bearing Lophotrochozoa are involved in chaetogenesis. Although one can expect that a chitin synthase involved in the formation of the coating layer (i.e., potentially indicated by an expression within follicle cells) has yet to be identified, the remaining annelid chitin synthase groups (B and D) do not necessarily all have to be involved in chaetogenesis. Rather, they could secrete another chitinous structure, namely the peritrophic membrane of the gizzard (Peters 1968, Peters & Walldorf 1986) that appears to be widespread among annelids (e.g., Nereidae, Terebellidae, Lumbricidae; Peters 1968). Thus, the high diversity of chitin synthases might be the result of specialisations of paralogs in terms of temporal and spatial restrictions of chitin secretion (Roncero 2002).

4.4 Possible interactive role of MLCK

The follicle cell marker *Pd_MLCK1* shows strong sequence and structure similarities to a myosin light chain kinase (MLCK). MLCKs are enzymes that phosphorylate sites on the N terminus of the regulatory light chain of myosin (generally "class" II myosins). Investigations of Stull et al. (1998) revealed that vertebrates possess two MLCK genes, of which one encodes the skeletal muscle MLCK, whereas the other generates the MLCK of the smooth muscle type. Studies on the smooth muscle MLCK gene show that, in contrast to the skeletal muscle MLCK gene, the former gives rise to three transcript variants regulated by different promoters (Birukov et al. 1998, Smith et al. 1998), which are thereby expressed in a cell-specific manner. The basic structural and functional elements of a smooth muscle MLCK is a catalytic kinase domain, one fibronectin (Fn) module as well as

three immunoglobin (Ig) modules, and three tandemly repeated actin-binding motifs (3x DFRxxL) at the N terminus, but their functions for MLCK activity are not completely understood (Kamm & Stull 2001). The second smooth muscle MLCK variant contains an N-terminal extension of six Ig modules and three further actin-binding motifs in addition to the basic subset of domains. It is hypothesized that this N-terminal extension increases the sensitivity to actin-containing filaments (Kudryashov et al. 1999, Yang et al. 2006). In the present study, the structure of the initially found MLCK of *P. dumerilii* (*Pd_MLCK1*) resembles the structural and functional elements of a long type smooth muscle MLCK. Such an assessment was not possible for most of the other MLCKs (including the paralog *Pd_MLCK2* and the ortholog *Macr_CS*) that I found via screening of recently updated databases, since they each comprise short fragments without an Fn domain and a kinase domain. The basic subset of MLCK-typical domains (i.e., Ig, Fn and kinase domains) was found in *C. teleta* (*Cate_MLCK*) and *L. gigantea* (*Logi_MLCK*), for which gene expressions were not studied.

In the process of chaetogenesis, the question remains why an MLCK might be important and what role it may play within the mechanism of chaetae formation. As described by Kamm & Stull (2001), it is suggested that MLCKs are enzymes that bind both to microfilaments and myosin filaments. For that purpose they use the DFRxxL motif to target F-actin, whereas the C-terminal Ig modules bind to the light chain of myosins. This results in a tied position of the MLCK in which phosphorylation of the regulatory myosin light chain is possible. Thinking about an enzyme whose ability is to interact with both F-actin and myosin, ideas about a mechanistic model are conceivable. Especially for long-type MLCKs this might be an interesting hypothesis, since the increased number of Ig domains and actin-binding motifs is suggested to be important for cytoskeleton organization (Yang et al. 2006), as for example utilized in the coordination of the apical filament in cells. As in chaetogenesis the chaetoblast is suggested to play an important role (Bouligand 1967, Hausen 2005) and considering that its apical surface is characterized by prominent microvilli which contain a high amount of F-actin, the DFRxxL motifs of this enzyme can be suggested to interact here. On the other hand, thinking about the myosin-binding Ig modules of this enzyme, it is important to remember the myosin motor head structure of most of the lophotrochozoan chitin synthases. As it has these two domains (for F-actin and myosin, respectively), such an MLCK is conceivable to act as a regulator in the hypothetical interaction of the myosin motor domain of the chitin synthase with the F-actin-containing microvilli of the chaetoblast. This would be a simple model to explain chitin synthesis dynamics at the apical surface of the chaetoblast. However, in *P. dumerilii* it was found that both enzymes, namely *Pd_MLCK1* and *Pd_CS1*, are expressed in different cell types (*Pd_MLCK1* in follicle cells and *Pd_CS1* in chaetoblasts, respectively), which would not support such a hypothesis. Additionally, the hypothesized myosin motor head could not be detected for *Pd_CS1* in this study. Nonetheless, in *M. cranium*, where a presumable MLCK ortholog to *P. dumerilii* was found, gene expression studies revealed that both genes (*Macr_CS* and *Macr_MLCK*) are expressed within the same cell, namely the chaetoblast. This supports the hypothesis of an interactive role of MLCK enzymes in chaetoblast-specific processes during chaetogenesis. In addition, my above described investigations on *P. dumerilii* genome and transcriptome data yielded evidence for a second MLCK

(*Pd_MLCK2*), and the possibility should not be ruled out that differentially regulated transcript variants may be present (as observed for vertebrate MLCKs by Birukov et al. 1998 and Smith et al. 1998). Further experiments will show whether *Pd_MLCK2* or another transcript variant of *Pd_MLCK1* is expressed in the chaetoblast instead of the follicle cells. Moreover, future studies will show whether one of the not yet further studied chitin synthases of *P. dumerilii* possesses an MMD that is necessary for MLCK interaction and is expressed in follicle cells.

4.5 Evolution of chaetae and phylogenetic implications

4.5.1 Homology of chaetae

The investigated set of chaetae-specific molecular markers in *P. dumerilii*, which was established during this study, was finally used for a comparative approach in other representatives of lophotrochozoans, to test whether or not these genes are solely annelid-specific or, beyond that, also specific to other chaetae-bearing lophotrochozoans. The sampled groups comprised three species, of which one was the sedentary annelid *C. teleta*, whereas the others, the myzostomid *M. cirriferum* and the brachiopod *M. cranium*, were investigated more intensively.

Capitella teleta

One may ask why amongst the study organisms, two species that are accepted to be annelids were selected. *C. teleta* was chosen for one specific aim. The evolutionary tree analysis revealed that *Cate_CS2* is a clear ortholog of *Pd_CS1*, for which the presence of a myosin motor head could neither be validated nor disapproved. However, sequence analysis of *Cate_CS2* yielded information that this gene likewise lacks this specific region. Nevertheless, as the myosin motor-lacking chitin synthase of *P. dumerilii* is expressed within the chaetoblast, it was interesting to check whether this gene (*Cate_CS2*) is expressed in the same cell type, which was indeed the case. The presence of these orthologous genes in the chaetoblasts of the two annelids clearly indicates the common origin of the cells, but also suggests that the function of *Pd_CS1* in the chaetoblasts indeed may not depend on an MMD.

Myzostoma cirriferum

Several studies of the last years deal with the phylogenetic position of the Myzostomida (based on morphological data: Haszprunar 1996; molecular data: Eeckhaut et al. 2000, Littlewood et al. 2001, Giribet et al. 2004, Bleidorn et al. 2007 and 2009; combined analyses: Zrzavý et al. 2001 and 2009). Even in recent studies, myzostomids tend to have an unstable position (probably due to their very long branches; see, for instance, the trees of Dunn et al. 2008 and Struck et al. 2011),

as they cluster, for example, either near "Platyzoa" (Eeckhaut et al. 2000, Halanych 2004), within Bryozoa (Passamaneck & Halanych 2006), within "Platyzoa" (Dunn et al. 2008) or within Annelida (Hejnol et al. 2009) and were often excluded in analyses of reduced taxon samplings. A variety of other molecular data sets consistently place Myzostomida at the base of or within Annelida; this includes evidence based on mitochondrial gene order as well as mitochondrial sequence analyses (Bleidorn et al. 2007 and 2009) or phylogenomic analyses using a denser taxon sampling within annelids (Hejnol et al. 2009, Struck et al. 2011). Thus, in including the potentially basal-branching annelid (Struck et al. 2011) *M. cirriferum* in the taxon sampling (in addition to *C. teleta* as a sedentary and *P. dumerilii* as an errant "polychaete") for gene expression studies with the new chaetae-specific molecular markers, a new level of comparison has been established that is independent from previous molecular or morphological levels. The same was possible for *M. cranium*, a representative of brachiopods where the high degree of morphological similarity of their chaetae to annelid chaetae has long been most controversially discussed (e.g., Storch & Welsch 1972, Orrhage 1973, Gustus & Cloney 1972, Ax 1989, Conway Morris 1993, Lüter & Bartolomaeus 1997, Lüter 2000a and 2001).

The computational approach revealed several candidates for *M. cirriferum* and *M. cranium* that resemble genes that are orthologous to *P. dumerilii* and *C. teleta*. It is striking that more genes have been isolated for the myzostomid (nine candidate genes) than for the brachiopod (three candidate genes), but this might partly be due to the quality and coverage of the transcriptome data, which is much higher in *M. cirriferum* (Christoph Bleidorn, pers. communication, Florian Raible, pers. communication). It is also conceivable that this is due to the proposed close phylogenetic relationship of annelids and myzostomids (Bleidorn et al. 2007, Bleidorn et al. 2009, Hejnol et al. 2009, Struck et al. 2011). Orthology assessment via evolutionary sequence analysis was possible for four markers (*Pd_CS1*, *Pd_MLCK1*, *Pd_CAML*, *Pd_SCA2L*), as a large number of homologous sequences from other organisms, including well-studied model organisms within Ecdysozoa and Vertebrata, were identified. For five other markers (*Pd_fcmg2*, *Pd_cbmg3*, *Pd_fcmg4*, *Pd_fcmg5*, *Pd_fcmg11*), potential orthologs were only found in *M. cirriferum* and a few other annelids, making an evolutionary sequence analysis unfeasible. Still, as for all investigated species only one gene copy was found (i.e., a moderate or good BLAST hit confirmed via reBLAST) and that was furthermore well-alignable with other sequences of the same marker, each sequence can be assumed to be not only homologous but potentially orthologous to the corresponding *P. dumerilii* chaetogenesis marker. This is to the exception of chitin synthase and myosin light chain kinase, where I have shown that paralogs exist. Nevertheless, the two paralogs of the MLCK gene in *P. dumerilii* appeared to be orthologous to the other found MLCKs and in the case of the chitin synthase paralogs, *Pd_CS1* is orthologous to *Cate_CS2*, *Myci_CS* is orthologous to *Pd_CS2* + *Pd_CS3* and *Macr_CS* is orthologous to *Cate_CS1*.

In terms of gene expression studies in *M. cirriferum* (performed in collaboration with C. Helm, University of Leipzig, Germany), this study reports first evidence for chaetal sac specificity of presumable orthologs in myzostomids. Both sampled markers (*Myci_fcmg5*, *Myci_csmg2*) are not

only found to be expressed within chaetal sac-specific cells, but even within the same cell types known from *P. dumerilii*. This was the case for both screened developmental stages, no matter if larval or adult material was used for ISH. Thus these two markers provide new evidence for the homology of chaetae in Annelida and Myzostomida. The remaining seven marker candidates that are potentially orthologous to the *P. dumerilii* chaetogenesis markers promise to further substantiate this finding.

Macandrevia cranium

Comparable results were yielded with ISH studies performed with larval *M. cranium*. Both screened genes (*Macr_CS*, *Macr_MLCK*) are exclusively expressed in chaetal follicles and are, as their sequences were identified from the transcriptome using sequence similarity, therefore homologous to those of *P. dumerilii*. As in the evolutionary analysis *Macr_CS* was grouped within group A of the MMD-bearing chitin synthases, it became apparent that this gene is a paralog of *Pd_CS1*, which is part of group C of the MMD-bearing chitin synthases. In the case of *Macr_MLCK*, the evolutionary analysis shows that *Macr_MLCK* is orthologous to the two *P. dumerilii* MLCKs.

As ultrastructural investigations of Lüter (2000a, 2001 and 2007) revealed that, in contrast to annelids, chaetal sacs of larval brachiopods do not contain follicle cells, it became interesting to check where ISH signals of genes homologous to *P. dumerilii* might be located, especially those that are known to be expressed in follicle cells. Interestingly, my studies revealed that both screened genes (*Macr_CS*, *Macr_MLCK*) exhibit an ISH signal within the chaetoblast, in difference to *P. dumerilii* where *Pd_MLCK1* is expressed in follicle cells only. This indicates that in brachiopod larval chaetae all genes and processes involved in chaetogenesis are restricted to a single cell and cell type – the chaetoblast. This situation found in the larval brachiopod may well reflect the ancestral state and it seems likely that the annelid and adult brachiopod follicle cells originate from chaetoblasts. This is further corroborated by the coexpression of *Pd_csmg10* in the chaetoblast and the two proximal follicle cells of *P. dumerilii*.

Since with this study there is now a set of chaetal sac-specific molecular markers available in *P. dumerilii*, to which potential orthologs (in addition to *M. cirriferum*) are present in *M. cranium* (note that even *Macr_CS*, though a paralog of *Pd_CS1*, is an ortholog of the annelid *Cate_CS1*), and whose first ISH experiments revealed that these genes are expressed in the same cell types, possible consequences for interpreting chaetae evolution in general and the position of Brachiopoda within the metazoan tree of life in particular have to be discussed. As most molecular analyses (e.g., Dunn et al. 2008, Helmkampf et al. 2008, Hejnol et al. 2009, Nesnidal et al. 2010) suggest that the last common ancestor of annelids and brachiopods lived during early lophotrochozoan radiation, which dates back at least 600 Million years ago (Peterson et al. 2008, Blair 2009), it is worth noting that in the present work at least three marker candidates with a clear sequence similarity to *P. dumerilii* have been found for *M. cranium*. Even more important is the evidence obtained by the

evolutionary analyses of chitin synthases performed in this study. These investigations revealed not only that *Macr_CS* is a paralog of *Pd_CS1*, but that it belongs to the MMD-bearing chitin synthases (featuring also a "bridge" region within the CS domain) and groups within a clade of chitin synthase genes (group A) from Annelida, Mollusca and Brachiopoda. Additional prediction analysis for the amino acid sequence of *Macr_CS* yielded structural evidence for the presence of an MMD (phylogenetically related to the MMD of annelids and molluscs) linked to the chitin synthase. Altogether, these new characters (chitin synthase with an MMD and a "bridge" region) indicate that, at least from the perspective of chitin synthase gene structure and sequence, brachiopods share derived characters with the other sampled lophotrochozoans and are thus more similar to them than to other protostomians or deuterostomians.

Given that, as stated above, the candidate genes found in *M. cranium* are indicated to be homologous and their gene expressions (already performed for *Macr_CS* and *Macr_MLCK*) are exclusively localized within comparable cell types as described for the annelid *P. dumerilii*, it is parsimonious to assume that such genes, their distinct expression and their potential role within chaetogenesis evolved just once. In that case, brachiopod chaetae would be structures homologous to that of annelids and would have been present in the last common ancestor of these two taxa. If the position of brachiopoda is considered to be in a protostomian clade named Lophotrochozoa (e.g., Halanych et al. 1995), fewer independent losses of chaetae would have to be assumed in comparison to the traditional position at the base of Radialia within Lophophorata (e.g., Hyman 1959).

4.5.2 Evolutionary origin of chaetae within Lophotrochozoa

As during the last years, all molecular studies have exclusively favored the Lophotrochozoa hypothesis (Halanych et al. 1995) and did not support the traditional Lophophorata ("Tentaculata"), I will focus in the following on Lophotrochozoa only. The Lophotrochozoa were first described by Halanych et al. (1995) and comprise a clade consisting of Annelida, Mollusca, Brachiopoda, Phoronida and Bryozoa, thus including the lophophorate taxa. However, the study of Halanych et al. (1995) was performed based on an incomplete taxon sampling and thus required further investigations. With his review in 2004, Halanych redefined Lophotrochozoa as a clade that comprises the taxa Annelida, Mollusca, Brachiopoda, Phoronida, Nemertea, Entoprocta, Ectoprocta, Platyhelminthes, Gastrotricha, Rotifera and Gnathostomulida. Some other recent phylogenomic analyses have additionally placed either Acoelomorpha (Petersen & Eernisse 2001, Dunn et al. 2008) or Chaetognatha (Giribet 2008, Hejnol et al. 2009) within Lophotrochozoa. Until now, the interrelationships of the lophotrochozoan taxa remain unresolved, leaving a big polytomy at the base of the clade, although some authors suggest that "Platyzoa" (comprising Platyhelminthes, Gnathostomulida, Rotifera and Gastrotricha) are sister to a clade termed Lophotrochozoa *sensu stricto* (Giribet 2008, Helmkampf et al. 2008, Nesnidal et al. 2010). Nevertheless, large-scale phylogenomic studies (Dunn et al. 2008, Hejnol et al. 2009) found contradictory topologies. In

addition, no molecular study has so far reported an exclusive sister group relationship between Annelida and Brachiopoda. If such a scenario would be hypothesized, one has to assume that chaetae evolved once in the last common ancestor of Brachiopoda and Annelida, and was lost at least twice within the group of annelids (i.e., in the ancestor of Sipunculida and Hirudinea, respectively). However, the least inclusive taxon (Fig. 42) within Lophotrochozoa that is supported by the tree topology of multiple independent analyses (but with weak to moderate support) comprises a clade of Annelida + Brachiopoda + Mollusca + Nemertea + Phoronida (ABMNP; Dunn et al. 2008, Helmkampf et al. 2008, Paps et al. 2009a and 2009b, Nesnidal et al. 2010). The inner topology of that ABMNP clade remains highly ambiguous, as some studies suggest Mollusca to be the earliest branch (Dunn et al. 2008, Helmkampf et al. 2008), whereas others place Annelida as sister group of the remaining ABMNP subtaxa (Nesnidal et al. 2010; also Paps et al. 2009a and 2009b if nemerteans are not considered, as well as Hejnol et al. 2009 if phoronids are not considered). Even though the phylogenetic position of myzostomids is highly unstable in the few phylogenetic analyses that include them, their affinity to "platyzoan" taxa (Fig. 42 A; Eeckhaut et al. 2000, Halanych 2004, Passamaneck & Halanych 2006, Dunn et al. 2008) is potentially due to long branch attraction (Bleidorn et al. 2009). In addition to the homology of annelid and myzostomid chaetae shown in the present study, the body of independent molecular evidence (mitochondrial gene order, mitochondrial sequence analysis, phylogenomic analyses) supports an annelid affinity (Bleidorn et al. 2007 and 2009, Hejnol et al. 2009, Struck et al. 2011) and it thus seems plausible that Myzostomida at least belong to the ABMNP clade, if not even to the Annelida *sensu lato*. With the given lack of phylogenetic resolution within the ABMNP clade, suggesting the emergence of chaetae in the last common ancestor would assume several secondary reductions on the lineages leading to the chaetae-lacking taxa, namely one loss in Phoronida, one in Nemertea and at least one in Mollusca (or even more if similar structures in polyplacophorans and octopods are chaetae-derived and therefore homologous).

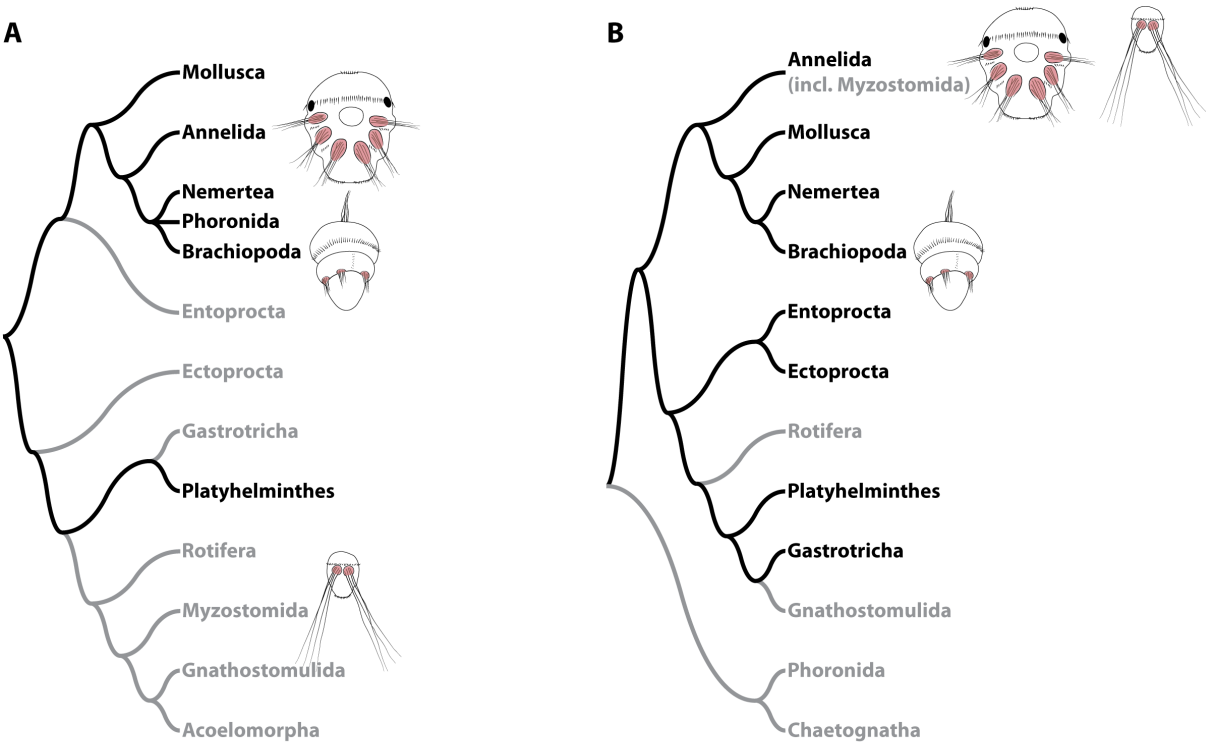
The insights gained with this study provide new characters which are relevant for hypotheses on the evolution of chaetae. The data set and especially the character of an MMD-linked chitin synthase indicate that chaetae can be assumed for the last common ancestor of at least all sampled lophotrochozoans, namely Annelida, Brachiopoda and Mollusca (all belonging to the above mentioned ABMNP clade that remains internally unresolved). Assuming chaetae for an ancestor that is even shared by molluscs, one has to argue about the ancestral function of MMD-bearing chitin synthases, since molluscs do not possess chaetae (except for potentially chaetae-like structures in polyplacophorans and juvenile octopods). Investigations done by Weiss et al. (2006) and Schönitzer & Weiss (2007) revealed that the chitin synthase (*Atri_CS* and *Myga_CS*, both are group A chitin synthases) plays an essential role in nacre formation of molluscs. They hypothesized the MMD to be especially important for interaction with cytoskeletal elements as they enable controlled mineral deposition in shell biogenesis. Considering that for the brachiopod representative of chitin synthase group A (*Macr_CS*) I could show a spatial and temporal connection with chaetogenesis, it seems plausible that, if molluscs secondarily lost chaetae, their group A chitin synthase changed its

function from chitin secretion in chaetogenesis to secretion of a chitin matrix in shell formation. Assuming chaetae for the ground pattern of molluscs is even supported by a few fossils. Scheltema & Ivanov (2002) and Scheltema et al. (2003) published studies about *Wiwaxia* suggesting that this fossil shows some aplacophoran characters. Conway Morris & Peel (1995) and Butterfield (2006) however suggest that, as *Wiwaxia* is assumed to bear microvilli-secreted chaetae, it should be interpreted as a stem-group annelid. In contrast, Caron et al. (2006) propose that *Wiwaxia* and *Odontogriphus* are early representatives of molluscs indicated by the presence of a radula. Additionally, there is another fossil that seems to unite characteristics of molluscs as well as chaetae, namely the mollusc *Pelagiella* (J. Vinther, personal communication) that bears prominent chaetae similar to those found in annelids. This unique character combination of chaetal structures and a shell-covered body constitutes a possible ancestral condition in the context of the evolution of chaetae and shells (i.e., their chitinous parts) within Lophotrochozoa, although this largely dependents on the phylogenetic placement of *Pelagiella*, which is not yet clear. However, considering the recently published study of Kocot et al. (2011) about deep molluscan phylogeny, it has to be assumed that, if chaetae were present within the ground pattern of molluscs, then they must have been lost at least three times independently. This would have to be assumed once within the Aculifera (in the ancestor of the Aplacophora), as polyplacophorans feature chaetae-like sensory hairs (Leise and Cloney 1982), and at least twice within the Conchifera (within Cephalopoda and in the ancestor of the remaining the Conchifera). The assumption of at least two reductions of chaetae within Conchifera is based on studies of Brocco et al. (1974) who describe a chaetae-like structure – the Kölliker's Organ – for octopods (though only present in late embryonic and juvenile stages).

The presence of homologous chaetae as a shared character in the last common ancestor of Brachiopoda and Annelida has also been discussed based on the fossil record. Conway Morris & Peel (1995) studied halkieriid fossils and hypothesized that this group plays an important role in interpreting annelid, brachiopod and mollusc evolution. As, e.g., *Halkieria* features both characters of molluscs (shells) and annelids (sclerites interpreted as chaetae), it was interpreted as a stem-group brachiopod (Conway Morris 1998) or as a mollusc (Conway Morris & Peel 1990, Peel 1991, Vinther & Nielsen 2005). A recent study of Vinther et al. (2008) even suggests an affinity of annelids to brachiopods indicated by the shell plates of annelid-associated fossil machaeridians. In the light of a robust phylogenetic framework for Lophotrochozoa that has yet to be established by future large-scale molecular analyses (e.g., via whole-genome comparisons of rare genomic changes, Rokas & Holland 2000), the placement of these fossils might shed more light on the evolution of ancient characters, such as chaetae and shells.

To conclude, I was able to establish a comprehensive set of molecular markers for cell type characterization in *P. dumerilii* chaetal sacs. This enabled comparative analyses of homologous marker genes and their cellular localization in *C. teleta*, *M. cirriferum* and *M. cranium*. The data obtained in the present study constitute the first molecular evidence that annelid, myzostomid and brachiopod chaetae are homologous. Further corroboration can be expected from expression analyses of additional marker candidates of this study and of the still ongoing sequencing projects

of the *M. cirriferum* and *M. cranium* transcriptomes as well as the *P. dumerilii* genome. Especially future expression analyses of chitin synthases appear to be of strong interest, as this study has revealed that the sampled lophotrochozoans possess an unexpected diversity of MMD-linked chitin synthases, of which two groups appear to be involved in chaetogenesis. Based on these insights, the last common ancestor of the sampled lophotrochozoans featured three MMD-linked chitin synthases that were involved in formation of chaetae and potentially other chitinous structures, such as shells. Therefore, the lingulid brachiopods are of great importance, as they are suggested to be basal-branching brachiopods that contain (apart from brachiopod chaetae) a chitinophosphatic shell (Williams et al. 1994) and represent an extant brachiopod taxon that comprises the oldest fossils (reaching into the Lower Cambrium; Lehmann & Hillmer 1997). As chitin synthesis also plays an important role in nacre formation of molluscs (Weiss et al. 2006), future studies might even reveal that the shells of molluscs and brachiopods are homologous. This, of course, will need extensive comparative studies on the complex process of shell formation. However, if evidence for shell homology will be discovered, it can be assumed that the last common ancestor of at least a part of the lophotrochozoans (probably the ABMNP clade) did not only possess chaetae (made of β-chitin) and three MMD-linked chitin synthases, but also a chitin-containing shell. It still remains to be solved if the polyplacophoran sensory hairs and the octopod Kölliker's Organ are relics of chaetae, but if they are indeed homologous structures, this would be further evidence that chitinous chaetae are an ancient feature of at least a part of the lophotrochozoan radiation.



C

phylogenetic study	data type	Lophotrochozoa *sensu lato*	Lophotrochozoa *sensu stricto*	Annelida + Mollusca + Nemertea + Phoronida + Brachiopoda	Annelida + Nemertea + Phoronida + Brachiopoda	Mollusca + Nemertea + Phoronida + Brachiopoda
Petersen & Eernisse 2001	18S rDNA	+ (+ Acoelomorpha)	–	–	–	–
Halanych 2004	review	+	–	–	–	–
Passamaneck & Halanych 2006	18S + 28S rDNA	+	–	–	–	–
Dunn et al. 2008 (tree A)	**150 genes**	+ (+ Acoelomorpha)	–	+	+	–
Giribet 2008	review	+ (+ Chaetognatha)	+	–	–	–
Helmkampf et al. 2008	7 genes	+	+	+	+	–
Hejnol et al. 2009 (tree B)	**1487 genes**	+ (+ Chaetognatha)	–	–	–	–
Paps et al. 2009a	18S + 28S rDNA	+	–	+	–	–
Paps et al. 2009b	13 genes	+	–	+	–	–
Nesnidal et al. 2010	78 genes	+	+	+	–	+

Fig. 42: Origin of chaetae in the lophotrochozoan tree of life. Drawings of larval stages of three of the four study organisms (*Platynereis dumerilli*, *Myzostoma cirriferum*, *Macandrevia cranium*) have been placed on the topologies of the hitherto largest phylogenomic analyses. A – Lophotrochozoan part of the tree of Dunn et al. (2008). B - Lophotrochozoan part of the tree of Hejnol et al. (2009). C – Summary of the phylogenetic support for Lophotrochozoa *sensu lato*, Lophotrochozoa *sensu stricto* (i.e., *sensu lato* without "Platyzoa") and potential internal relationships containing the chaetae-bearing groups Annelida and Brachiopoda. Note that for such an internal resolution, Myzostomida were not considered, as only four of the named studies sampled this taxon, placing them either near "Platyzoa" (Halanych 2004), within Bryozoa (Passamaneck & Halanych 2006), within "Platyzoa" (Dunn et al. 2008) or within Annelida (Hejnol et al. 2009).

5 References

Ax, P. (1989). Basic phylogenetic systematization of the Metazoa. In *The hierarchy of life*. (B. Fernholm, K. B., H. Jörnvall, ed.), pp. 229–245. Elsevier, Amsterdam.

Ax, P. (1999). Annelida. In *Das System der Metazoa II - Ein Lehrbuch der phylogenetischen Systematik*, pp. 59-62. Gustav Fischer Verlag, Stuttgart, Jena, Lübeck, Ulm.

Arendt, D. (2003). Evolution of eyes and photoreceptor cell types. *International Journal of Developmental Biology* **47**, 563-572.

Arendt, D., Tessmar-Raible, K., Snyman, H., Dorresteijn, A. W. & Wittbrodt, J. (2004). Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science* **306**, 869.

Arendt, D. (2005). Genes and homology in nervous system evolution: comparing gene functions, expression patterns, and cell type molecular fingerprints. *Theory in Biosciences* **124**, 185-197.

Audouin, J. V., & Milne Edwards, A. (1833). Recherches pour servir a l'histoire du littoral de la France. In *Annelides*, vol. 2, pp. 1-290. Paris.

Bamezai, A. & Rock, K. L. (1995). Overexpressed Ly-6A.2 mediates cell-cell adhesion by binding a ligand expressed on lymphoid cells. *Proceedings of the National Academy of Sciences* **92**, 4294.

Bartolomaeus, T. & Meyer, K. (1997). Morphogenesis and phylogenetic significance of hooked setae in Arenicolidae (Polychaeta, Annelida). *Invertebrate Biology* **116**, 227-242.

Bartolomaeus, T. (1998). Chaetogenesis in polychaetous Annelida - significance for annelid systematics ant the position of the Pogonophora. *Zoology* **100**: 348-364.

Bartolomaeus, T. (2002). Structure and formation of thoracic and abdominal uncini in Fabricia stellaris (Müller, 1774)-implication for the evolution of Sabellida (Annelida). *Zoologischer Anzeiger-A Journal of Comparative Zoology* **241**, 1-17.

Beedham, G. & Trueman, E. (1968). The cuticle of the Aplacophora and its evolutionary significance in the Mollusca. *Journal of Zoology* **154**, 443-451.

Bhaud, M. (1978). Morphological variations of the modified setae of chaetopterids during ontogenesis. *Ophelia* **17**, 199-206.

Birukov, K. G., Schavocky, J. P., Shirinsky, V. P., Chibalina, M. V., Van Eldik, L. J. & Watterson, D. M. (1998). Organization of the genetic locus for chicken myosin light chain kinase is complex: multiple proteins are encoded and exhibit differential expression and localization. *Journal of cellular biochemistry* **70**, 402-413.

Blackwell, J., Parker, K. D. & Rudall, K. M. (1967). Chitin fibres of the diatoms *Thalassiosira fluviatilis* and *Cyclotella cryptica*. *Journal of Molecular Biology* **28**, 383-385.

Blair, J. (2009). Animals (Metazoa). In *The timetree of life* (Hedges, S. B. K., S., ed.), pp. 223-230. Oxford University Press, Oxford.

Blake, J. A., Grassle, J. P. & Eckelbarger, K. J. (2009). Capitella teleta, a new species designation for the opportunistic and experimental Capitella sp. I, with a review of the literature for confirmed records. *Zoosymposia* **2**, 25-53.

Bleidorn, C., Eeckhaut, I., Podsiadlowski, L., Schult, N., McHugh, D., Halanych, K. M., Milinkovitch, M. C. & Tiedemann, R. (2007). Mitochondrial genome and nuclear sequence data support Myzostomida as part of the annelid radiation. *Molecular Biology and Evolution* **24**, 1690.

Bleidorn, C., Podsiadlowski, L., Zhong, M., Eeckhaut, I., Hartmann, S., Halanych, K. & Tiedemann, R. (2009). On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evolutionary Biology* **9**, 150.

Blobel, G. & Sabatini, D. D. (1971). Ribosome-membrane interaction in eukaryotic cells. *Biomembranes* **2**, 193-195.

Bouligand, Y. (1966). Etude ultrastructurale des soies et des cellules associees chez une annelide polychete *Haplosyllis depressa*. *C. R. Acad. Sci. Paris* **263D**, 378-381.

Bouligand, Y. (1967). Les soies et les cellules associees chez deux Annelides Polychetes. *Zeitschrift für Zellforschung* **79**: 332-363.

Brocco, S. L., O'Clair, R. M. & Cloney, R. A. (1974). Cephalopod integument: the ultrastructure of Kölliker's organs and their relationship to setae. *Cell and Tissue Research* **151**, 293-308.

Bulaj, G., Kortemme, T. & Goldenberg, D. P. (1998). Ionization-reactivity relationships for cysteine thiols in polypeptides. *Biochemistry* **37**, 8965-8972.

Butterfield, N. J. (2006). Hooking some stem group "worms": fossil lophotrochozoans in the Burgess Shale. *BioEssays* **28**, 1161-1166.

Caron, J. B., Scheltema, A. H., Schander, C. & Rudkin, D. (2006). A soft-bodied mollusc with radula from the Middle Cambrian Burgess Shale. *Nature* **442**, 159-163.

Classon, B. J. & Coverdale, L. (1994). Mouse stem cell antigen Sca-2 is a member of the Ly-6 family of cell surface proteins. *Proceedings of the National Academy of Sciences* **91**, 5296.

Conway Morris, S. & Peel, J. S. (1990). Articulated halkieriids from the Lower Cambrian of north Greenland. *Nature* **345**, 802-805.

Conway Morris, S. (1993). The fossil record and the early evolution of the Metazoa. *Nature* **361**, 219-225.

Conway Morris, S. C. & Peel, J. S. (1995). Articulated halkieriids from the Lower Cambrian of North Greenland and their role in early protostome evolution. *Philosophical Transactions: Biological Sciences* **347**, 305-358.

Conway Morris, S. (1998). *The Crucible of Creation*. Oxford University Press. Oxford.

Cope, M. J. T. V., Whisstock, J., Rayment, I. & Kendrick-Jones, J. (1996). Conservation within the myosin motor domain: implications for structure and function. *Structure* **4**, 969-987.

De Rosa, R. (2001). Molecular data indicate the protostome affinity of brachiopods. *Systematic Biology* **50**, 848.

d'Hondt, J. L. & Franzen, A. (2001). Observations on embryological and larval stages of *Macandrevia cranium* (Müller, 1776) (Brachiopoda, Articulata). *Invertebrate Reproduction and Development* **40**, 153-162.

Dorresteijn, A W C (1990). Quantitative analysis of cellular differentiation during early embryogenesis of *Platynereis dumerilii*. *Development Genes and Evolution* 199(1): 14-30.

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sorensen, M. V., Haddock, S. H. D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M., Wheeler, W. C., Martindale, M. Q. & Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745-749.

Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783-795.

Ebling, F. J. (1945). Formation and nature of the opercular chaetae of *Sabellaria alveolata*. *Quarterly Journal of Microscopical Science* **85**, 153-176.

Eeckhaut, I., Jangoux, M. (1993). Life cycle and mode of infestation of *Myzostoma cirriferum* (Annelida), a symbiotic myzostomid of the comatulid crinoid *Antedon bifida* (Echinodermata). *Diseases of aquatic Organisms* 15(3): 207-217.

Eeckhaut, I., McHugh, D., Mardulyn, P., Tiedemann, R., Monteyne, D., Jangoux, M. & Milinkovitch, M. C. (2000). Myzostomida: a link between trochozoans and flatworms? *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**, 1383.

Eeckhaut, I., Fievez, L. & Muller, M. (2003). Larval development of Myzostoma cirriferum (Myzostomida). *J Morphol* **258**, 269 - 283.

Eernisse, D. J., Albert, J. S. & Anderson, F. E. (1992). Annelida and Arthropoda are not sister taxa: a phylogenetic analysis of spiralian metazoan morphology. *Systematic Biology* **41**, 305.

Fauchald, K. (1977). The polychaete worms, definitions and keys to the orders, families and genera. Natural History Museum of Los Angeles County, *Science Series* **28**.

Fischer, A. & Dorresteijn, A. (2004). The polychaete *Platynereis dumerilii* (Annelida): a laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. *BioEssays* **26**, 314-325.

Fischer, A. H. L., Henrich, T. & Arendt, D. (2010). The normal development of *Platynereis dumerilii* (Nereididae, Annelida). *Frontiers in Zoology* **7**, 31.

Foster, J. W. (1949). Chemical activities of fungi. Academic Press, New York, N. Y.

Fröbius, A. C., Matus, D. Q. & Seaver, E. C. (2008). Genomic organization and expression demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan *Capitella* sp. I. *PLoS One* **3**, e4004.

George, J. D. & Southward, E. C. (1973). A comparative study of the setae of Pogonophora and polychaetous Annelida. *Journal of the Marine Biological Association of the United Kingdom* **53**, 403-424.

Giraud-Guille, M. & Bouligand, Y. (1986). Chitin-protein molecular organization in arthropod. In *Chitin in nature and technology* (Muzzarelli, R., Jeuniaux, C. & Gooday, G. W., eds.), pp. 29-35. Plenum, New York.

Giribet, G., Sørensen, M. V., Funch, P., Kristensen, R. M. & Sterrer, W. (2004). Investigations into the phylogenetic position of Micrognathozoa using four molecular loci. *Cladistics* **20**, 1-13.

Giribet, G. (2008). Assembling the lophotrochozoan (= spiralian) tree of life. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 1513.

Glaser, L. & Brown, D. H. (1957). The synthesis of chitin in cell-free extracts of Neurospora crassa. *Journal of Biological Chemistry* **228**, 729.

Gordon, D. P. (1975). The resemblance of bryozoan gizzard teeth to "annelid-like" setae. *Acta Zoologica* **56**, 283-289.

Grassle, J. and J. F. Grassle (1976). Sibling species in the marine pollution indicator *Capitella* (Polychaeta). *Science* 192: 567-569

Grygier, MJ (2000) Class myzostomida. Polychaetes and allies: the southern synthesis. Fauna of Australia 4: 297-329.

Gupta, B. & Little, C. (1970). Studies on Pogonophora. 4. Fine structure of the cuticle and epidermis. *Tissue and Cell* **2**, 637-696.

Gustus, R. M. & Cloney, R. A. (1972). Ultrastructural similarities between setae of brachiopods and polychaetes. *Acta Zoologica* **53**, 229-233.

Gustus, R. M. & Cloney, R. A. (1973). Ultrastructure of the larval compound setae of the polychaete Nereis vexillosa Grube. *Journal of Morphology* **140**, 355-366.

Halanych, K. M., Bacheller, J. D., Aguinaldo, A., Liva, S. M., Hillis, D. M. & Lake, J. A. (1995). Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science* **267**, 1641.

Halanych, K. M. (2004). The New View of Animal Phylogeny. *Annual Review of Ecology, Evolution, and Systematics* **35**, 229-256.

Harris, M. T., Lai, K., Arnold, K., Martinez, H. F., Specht, C. A. & Fuhrman, J. A. (2000). Chitin synthase in the filarial parasite, *Brugia malayi*. *Molecular and Biochemical Parasitology* **111**, 351-362.

Haszprunar, G. (1992). The first molluscs small animals. *Italian Journal of Zoology* **59**, 1-16.

Haszprunar, G. (1996). The Mollusca: coelomate turbellarians or mesenchymate annelids. In *Origin and evolutionary radiation of the Mollusca* (Taylor, J. D., ed.), Oxford University Press, Oxford, pp. 3–28.

Hauenschild, C. & Fischer, A. (1969). *Platynereis dumerilii*, Gustav Fischer Verlag, Stuttgart.

Hausen, H. & Bartolomaeus, T. (1998). Setal structure and chaetogenesis in Scolelepis squamata and Malacoceros fuliginosus (Spionidae, Annelida). *Acta Zoologica* **79**, 149-161.

Hausen, H. (2005). Chaetae and chaetogenesis in polychaetes (Annelida). *Hydrobiologia* **535**, 37-52.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W. E. G., Seaver, E., Wheeler, W. C., Martindale, M. Q., Giribet, G. & Dunn, C. W. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings of the Royal Society B: Biological Sciences* **276**, 4261-4270.

Helmkampf, M., Bruchhaus, I. & Hausdorf, B. (2008). Phylogenomic analyses of lophophorates (brachiopods, phoronids and bryozoans) confirm the Lophotrochozoa concept. *Proceedings of the Royal Society B: Biological Sciences* **275**, 1927.

Hogenkamp, D. G., Arakane, Y., Zimoch, L., Merzendorfer, H., Kramer, K. J., Beeman, R. W., Kanost, M. R., Specht, C. A. & Muthukrishnan, S. (2005). Chitin synthase genes in Manduca sexta: characterization of a gut-specific transcript and differential tissue expression of alternately spliced mRNAs during development. *Insect biochemistry and molecular biology* **35**, 529-540.

Horiuchi, H., Fujiwara, M., Yamashita, S., Ohta, A. & Takagi, M. (1999). Proliferation of intrahyphal hyphae caused by disruption of csmA, which encodes a class V chitin synthase with a myosin motor-like domain in *Aspergillus nidulans*. *Journal of bacteriology* **181**, 3721.

Hyman, L. H. (1959). *The invertebrates. Smaller coelomic Groups*, Vol. 5, McGraw-Hill, New York.

Ibrahim, G. H., Smartt, C. T., Kiley, L. M. & Christensen, B. M. (2000). Cloning and characterization of a chitin synthase cDNA from the mosquito Aedes aegypti. *Insect biochemistry and molecular biology* **30**, 1213-1222.

Jékely G, Arendt D (2007) Cellular resolution expression profiling using confocal detection of NBT/BCIP precipitate by reflection microscopy. *Biotechniques* 42: 751-755.

Jeuniaux, C. (1982). La chitine dans le regne animal. *Bull. Soc. Zool. France* **107**, 363-386.

Kamm, K. E. & Stull, J. T. (2001). Dedicated myosin light chain kinases with diverse cellular functions. *Journal of Biological Chemistry* **276**, 4527.

Katoh, K. & Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in bioinformatics* **9**, 286.

Kent, P. W., Whitehous, M. W. (1955). Biochemistry of the aminosugars. Academic Press, New York, N. Y.

Kocot, K. M., Cannon, J. T., Todt, C., Citarella, M. R., Kohn, A. B., Meyer, A., Santos, S. R., Schander, C., Moroz, L. L., Lieb, B. & Halanych, K. M. (2011). Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452-456.

Kramer, K. J. & Koga, D. (1986). Insect chitin: physical state, synthesis, degradation and metabolic regulation. *Insect Biochem.* 16, pp. 851–877.

Kudryashov, D. S., Chibalina, M. V., Birukov, K. G., Lukas, T. J., Sellers, J. R., Van Eldik, L. J., Watterson, D. M. & Shirinsky, V. P. (1999). Unique sequence of a high molecular weight myosin light chain kinase is involved in interaction with actin cytoskeleton. *FEBS Letters* **463**, 67-71.

Lanterbecq, D., Bleidorn, C., Michel, S. & Eeckhaut, I. (2008). Locomotion and fine structure of parapodia in Myzostoma cirriferum (Myzostomida). *Zoomorphology* **127**, 59-68.

Lehmann, U. & Hillmer, G. (1997). *Wirbellose Tiere der Vorzeit: Leitfaden der systematischen Paläontologie der Invertebraten*, 4., Enke, Stuttgart.

Leise, E. M. & Cloney, R. A. (1982). Chiton integument: ultrastructure of the sensory hairs of *Mopalia muscosa* (Mollusca: Polyplacophora). *Cell and Tissue Research* **223**, 43-59.

Leuckart, F. S. (1836). In Beziehung auf der Haarstern (*Comatula*) und *Pentacrinus europaeus*, so wie auf das Schmarotzerthier auf *Comatula*. *Notiz. Gebiete Natur- und Heilk. Gesammelt Mitgetheilt Froriep* **59**, 129-131.

Lewit-Bentley, A. & Réty, S. (2000). EF-hand calcium-binding proteins. *Current Opinion in Structural Biology* **10**, 637-643.

Li, S., Xie, L., Zhang, C., Zhang, Y., Gu, M. & Zhang, R. (2004). Cloning and expression of a pivotal calcium metabolism regulator: calmodulin involved in shell formation from pearl oyster (Pinctada fucata). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **138**, 235-243.

Littlewood, D. T. J., Olson, P. D., Telford, M. J., Herniou, E. A. & Riutort, M. (2001). Elongation factor 1-alpha sequences alone do not assist in resolving the position of the Acoela within the Metazoa. *Molecular Biology and Evolution* **18**, 437.

Liu, Y.-G., Whittier, R. F. (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25**(3): 674-681.

Liu, H., Kauffman, S., Becker, J. M. & Szaniszlo, P. J. (2004). *Wangiella* (*Exophiala*) *dermatitidis* WdChs5p, a class V chitin synthase, is essential for sustained cell growth at temperature of infection. *Eukaryotic Cell* **3**, 40-51.

Lotmar, W., Picken, L. E. R. (1950). A new crystallographic modification of chitin and its distribution. *Cellular and Molecular Life Sciences* **6**, 58-59.

Lüter, C. & Bartolomaeus, T. (1997). The phylogenetic position of Brachiopoda—a comparison of morphological and molecular data. *Zoologica Scripta* **26**, 245-253.

Lüter, C. (2000a). Ultrastructure of larval and adult setae of Brachiopoda. *Zool. Anz.* 239(1): 75-90.

Lüter, C. (2000b). The origin of the coelom in Brachiopoda and its phylogenetic significance. *Zoomorphology* **120**, 15-28.

Lüter, C. (2001) Brachiopod larval setae - a key to the phylum's ancestral life cycle? In: Brunton, C.H.C., Cocks, L.R.M. & Long, S.L. (eds.) Brachiopods - Past and Present. Taylor and Francis, London, pp. 46-55.

Lüter, C. (2007). Anatomy. In: Selden, P.A. (Ed), Treatise on Invertebrate Paleontology, part H: Brachiopoda (revised), Vol. 6. The Geological Society of America and University of Kansas, Boulder and Lawrence, pp. 2321-2355.

Madrid, M. P., Di Pietro, A. & Roncero, M. I. G. (2003). Class V chitin synthase determines pathogenesis in the vascular wilt fungus *Fusarium oxysporum* and mediates resistance to plant defence compounds. *Molecular Microbiology* **47**, 257-266.

Merzendorfer, H. & Zimoch, L. (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of experimental biology* **206**, 4393.

Merzendorfer, H. (2006). Insect chitin synthases: a review. Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology **176**, 1-15.

Meyer, K. & Bartolomaeus, T. (1996). Ultrastructure and formation of the hooked setae in *Owenia fusiformis* delle Chiaje, 1842: implications for annelid phylogeny. *Canadian journal of zoology* **74**, 2143-2153.

Müller, O. F. (1776). *Zoologiae Danicae prodromus: seu Animalium Daniae et Norvegiae indigenarum characteres, nomina, et synonyma imprimis popularium*, typis Hallageriis.

Nesnidal, M. P., Helmkampf, M., Bruchhaus, I. & Hausdorf, B. (2010). Compositional heterogeneity and phylogenomic inference of metazoan relationships. *Molecular Biology and Evolution* **27**, 2095.

Nielsen, C. (2001). *Animal Evolution*, (2nd edition), 563 pp., Oxford University Press, USA, Oxford.

Nikapitiya, C., De Zoysa, M., Whang, I., Kim, S. J., Choi, C. Y., Lee, J. S. & Lee, J. (2010). Characterization and expression analysis of EF hand domain-containing calcium-regulatory gene from disk abalone: Calcium homeostasis and its role in immunity. *Fish & Shellfish Immunology* **29**, 334-342.

O'Clair, R. M. & Cloney, R. A. (1974). Patterns of morphogenesis mediated by dynamic microvilli: chaetogenesis in Nereis vexillosa. *Cell and Tissue Research* **151**, 141-157.

Odronitz, F. & Kollmar, M. (2007). Drawing the tree of eukaryotic life based on the analysis of 2,269 manually annotated myosins from 328 species. *Genome Biology* **8**, R196.

Orrhage, L. (1971). Light and electron microscope studies of some annelid setae. *Acta Zoologica* **52**, 157-169.

Orrhage, L. (1973). Light and electron microscope studies of some brachiopod and pogonophoran setae. *Zoomorphology* **74**, 253-270.

Paps, J., Baguñà, J. & Riutort, M. (2009a). Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. *Proceedings of the Royal Society B: Biological Sciences* **276**, 1245.

Paps, J., Baguñà, J. & Riutort, M. (2009b). Bilaterian phylogeny: a broad sampling of 13 nuclear genes provides a new lophotrochozoa phylogeny and supports a paraphyletic basal acoelomorpha. *Molecular Biology and Evolution* **26**, 2397.

Passamaneck, Y. & Halanych, K. M. (2006). Lophotrochozoan phylogeny assessed with LSU and SSU data: evidence of lophophorate polyphyly. *Molecular Phylogenetics and Evolution* **40**, 20-28.

Peel, J. (1991). Functional morphology, evolution and systematics of early Paleozoic univalve molluscs. *Bulletin Grønlands Geologiske Undersøgelse* **161**, 1-116.

Peters, W. (1968). Vorkommen, zusammensetzung und feinstruktur peritrophischer membranen im tierreich. *Zoomorphology* **62**, 9-57.

Peters, W. (1972). Occurrence of chitin in mollusca. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry **41**, 541-550.

Peters, W. & Walldorf, V. (1986). Endodermal secretion of chitin in the 'cuticle' of the earthworm gizzard. *Tissue and Cell* **18**, 361-374.

Peterson, K. J. & Eernisse, D. J. (2001). Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. *Evolution and Development* **3**, 170-205.

Peterson, K. J., Cotton, J. A., Gehling, J. G. & Pisani, D. (2008). The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 1435.

Reynolds, E. S. (1963). The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *The Journal of Cell Biology* **17**, 208.

Rokas, A. & Holland, P. W. H. (2000). Rare genomic changes as a tool for phylogenetics. *Trends in Ecology & Evolution* **15**, 454-459.

Roncero, C. (2002). The genetic complexity of chitin synthesis in fungi. *Current genetics* **41**, 367-378.

Rouse, G. W. & Fauchald, K. (1997). Cladistics and polychaetes. *Zoologica Scripta* **26**, 139-204.

Rousseau, M., Plouguerne, E., Wan, G., Wan, R., Lopez, E. & Fouchereau-Peron, M. (2003). Biomineralisation markers during a phase of active growth in Pinctada margaritifera. *Comparative Biochemistry and Physiology-Part A: Molecular & Integrative Physiology* **135**, 271-278.

Rozen, S. & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-386.

Rudall, K. M. (1955). The distribution of collagen and chitin. In "Fibrous Proteins and their Biological Significance". *Symp. Soc. Exp. Biol*, **9** (1955), pp. 49–71.

Rudall K.M. (1963). The chitin/protein complexes of insect cuticles. In *Advances in Insect Physiology* (J.W.L. Beament, J. E. T. & Wigglesworth, V. B., eds.), Vol. Volume 1, pp. 257-313. Academic Press.

Rudall, K. M. & Kenchington, W. (1973). The chitin system. *Biological Reviews* **48**, 597-633.

Ruiz-Herrera, J., Gonzalez-Prieto, J. M. & Ruiz-Medrano, R. (2002). Evolution and phylogenetic relationships of chitin synthases from yeasts and fungi. *FEMS Yeast Research* 1, 247-256.

Ruiz-Herrera, J. & Ortiz-Castellanos, L. (2010). Analysis of the phylogenetic relationships and evolution of the cell walls from yeasts and fungi. *FEMS Yeast Research* **10**, 225-243.

Sakaguchi, M., Tomiyoshi, R., Kuroiwa, T., Mihara, K. & Omura, T. (1992). Functions of signal and signal-anchor sequences are determined by the balance between the hydrophobic segment and the N-terminal charge. *Proceedings of the National Academy of Sciences* **89**, 16-19.

Scheltema, A. H. & Ivanov, D. L. (2002). An aplacophoran postlarva with iterated dorsal groups of spicules and skeletal similarities to Paleozoic fossils. *Invertebrate Biology* **121**, 1-10.

Scheltema, A. H., Kerth, K. & Kuzirian, A. M. (2003). Original molluscan radula: comparisons among Aplacophora, Polyplacophora, Gastropoda, and the Cambrian fossil Wiwaxia corrugata. *Journal of Morphology* **257**, 219-245.

Schepotieff, A. (1903). Untersuchungen über den feineren Bau der Borsten einiger Chätopoden und Brachiopoden. *Zeitschrift für Wissenschaftliche Zoologie* **74**, 656-710.

Scholtz, G. (2002). The Articulata hypothesis - or what is a segment? *Organisms Diversity & Evolution* **2**, 197-215.

Schönitzer, V. & Weiss, I. (2007). The structure of mollusc larval shells formed in the presence of the chitin synthase inhibitor Nikkomycin Z. *BMC Structural Biology* **7**, 71.

Schroeder, P. C. (1984). Annelid Chaetae. In *Biology of the Integument* (Breiter-Hahn, J., Matoltsy, A. G. & Richards, K. S., eds.), Vol. 1, pp. 297-309. Springer Verlag, Berlin.

Seaver, E. C., Thamm, K. and Hill, S. D. (2005), Growth patterns during segmentation in the two polychaete annelids,

Capitella sp. I and *Hydroides elegans*: comparisons at distinct life history stages. *Evolution & Development*, 7: 312–326.

Seaver, E. C. & Kaneshige, L. M. (2006). Expression of 'segmentation' genes during larval and juvenile development in the polychaetes *Capitella* sp. I and *H. elegans*. *Developmental biology* **289**, 179-194.

Sigrist, C. J. A., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A. & Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research* **38**, D161.

Smith, A. F., Bigsby, R. M., Word, R. & Herring, B. P. (1998). A 310-bp minimal promoter mediates smooth muscle cell-specific expression of telokin. *American Journal of Physiology-Cell Physiology* **274**, C1188.

Specht, A. (1988). Chaetae. In *The Ultrastructure of Polychaeta* (Westheide, W. & Hermans, C. O., eds.), Vol. 4, pp. 45-59. Gustav Fischer, New York.

Stamatakis A, Hoover P, Rougemont J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol.* 75:758–771.

Steinberg, G. (2000). The cellular roles of molecular motors in fungi. *Trends in microbiology* **8**, 162-168.

Storch, V. & Welsch, U. (1972). Über Bau und Entstehung der Mantelrandstacheln von *Lingula unguis* L. (Brachiopoda). *Zeitschrift fur Wissenschaftliche Zoologie* **183**, 181-189.

Struck, T. H., Paul, C., Hill, N., Hartmann, S., Hosel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G. & Bleidorn, C. (2011). Phylogenomic analyses unravel annelid evolution. *Nature* **471**, 95-98.

Stull, T., Lin, P., Krueger, J., Trewhella, J. & Zhi, G. (1998). Myosin light chain kinase: functional domains and structural motifs. *Acta physiologica scandinavica* **164**, 471-482.

Suga, H., Schmid, V. & Gehring, W. J. (2008). Evolution and functional diversity of jellyfish opsins. *Current Biology* **18**, 51-55.

Suzuki, M., Sakuda, S. & Nagasawa, H. (2007). Identification of chitin in the prismatic layer of the shell and a chitin synthase gene from the Japanese pearl oyster, Pinctada fucata. *Bioscience, biotechnology, and biochemistry* **71**, 1735-1744.

Takeshita, N., Ohta, A. & Horiuchi, H. (2005). CsmA, a Class V chitin synthase with a myosin motor-like domain, is localized through direct interaction with the actin cytoskeleton in *Aspergillus nidulans*. *Molecular Biology of the Cell* **16**, 1961-1970.

Tessmar-Raible, K. & Arendt, D. (2003). Emerging systems: between vertebrates and arthropods, the Lophotrochozoa. *Current opinion in genetics & development* **13**, 331-340.

Tessmar-Raible, K., Steinmetz, P. R. H., Snyman, H., Hassel, M. & Arendt, D. (2005). Fluorescent two-color whole mount in situ hybridization in *Platynereis dumerilii* (Polychaeta, Annelida), an emerging marine molecular model for evolution and development. *Biotechniques* **39**, 460.

Thamm, K. & Seaver, E. C. (2008). Notch signaling during larval and juvenile development in the polychaete annelid *Capitella* sp. I. *Developmental biology* **320**, 304-318.

Thomassin, B. & Picard, C. (1972). Etude de la microstructure des soies de polychetes Capitellidae et Oweniidae au microscope electronique a balayage: un critere systematique precis. *Marine Biology* **12**, 229-236.

Ullrich-Lüter, E. M., Dupont, S., Arboleda, E., Hausen, H. & Arnone, M. I. (2011). Unique system of photoreceptors in sea urchin tube feet. *Proceedings of the National Academy of Sciences* **108**, 8367.

Valentine, J. W. (1973). Coelomate superphyla. *Systematic zoology* **22**, 97-102.

Veronico, P., Gray, L., Jones, J., Bazzicalupo, P., Arbucci, S., Cortese, M., Di Vito, M. & De Giorgi, C. (2001). Nematode chitin synthases: gene structure, expression and function in *Caenorhabditis elegans* and the plant parasitic nematode *Meloidogyne artiellia*. *Molecular Genetics and Genomics* **266**, 28-34.

Vinther, J. & Nielsen, C. (2005). The early Cambrian *Halkieria* is a mollusc. *Zoologica Scripta* **34**, 81-89.

Vinther, J., Van Roy, P. & Briggs, D. E. G. (2008). Machaeridians are Palaeozoic armoured annelids. *Nature* **451**, 185-188.

Weber, I., Aßmann, D., Thines, E. & Steinberg, G. (2006). Polar localizing class V myosin chitin synthases are essential during early plant infection in the plant pathogenic fungus *Ustilago maydis*. *The Plant Cell Online* **18**, 225-242.

Weiss, I. M., Schönitzer, V., Eichner, N. & Sumper, M. (2006). The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. *FEBS Letters* **580**, 1846-1852.

Werner, S., Sugui, J. A., Steinberg, G. & Deising, H. B. (2007). A chitin synthase with a myosin-like motor domain is essential for hyphal growth, appressorium differentiation, and pathogenicity of the maize anthracnose fungus *Colletotrichum graminicola*. *Molecular Plant-Microbe Interactions* **20**, 1555-1567.

Williams, A., Cusack, M. & Mackay, S. (1994). Collagenous chitinophosphatic shell of the brachiopod *Lingula*. *Philosophical Transactions: Biological Sciences*, 223-266.

Yang, C. X., Chen, H. Q., Chen, C., Yu, W. P., Zhang, W. C., Peng, Y. J., He, W. Q., Wei, D. M., Gao, X. & Zhu, M. S. (2006). Microfilament-binding properties of N-terminal extension of the isoform of smooth muscle long myosin light chain kinase. *Cell research* **16**, 367-376.

Zhang, Y., Foster, J. M., Nelson, L. S., Ma, D. & Carlow, C. K. S. (2005). The chitin synthase genes chs-1 and chs-2 are essential for *C. elegans* development and responsible for chitin deposition in the eggshell and pharynx, respectively. *Developmental biology* **285**, 330-339.

Zhu, Y. C., Specht, C. A., Dittmer, N. T., Muthukrishnan, S., Kanost, M. R. & Kramer, K. J. (2002). Sequence of a cDNA and expression of the gene encoding a putative epidermal chitin synthase of *Manduca sexta*. *Insect biochemistry and molecular biology* **32**, 1497-1506.

Zimoch, L. & Merzendorfer, H. (2002). Immunolocalization of chitin synthase in the tobacco hornworm. *Cell and Tissue Research* **308**, 287-297.

Zrzavý, J., Hypša, V. & Tietz, D. F. (2001). Myzostomida are not annelids: molecular and morphological support for a clade of animals with anterior sperm flagella. *Cladistics* **17**, 170-198.

Zrzavý, J., Riha, P., Pialek, L. & Janouskovec, J. (2009). Phylogeny of Annelida (Lophotrochozoa): total-evidence analysis of morphology and six genes. *BMC Evolutionary Biology* **9**, 189.

6 Appendix

Fig. A1: Alignment of the conserved cysteine sites of *Pd_fcmg5*, *Pd_fcmg11* and respective orthologs. Conserved cysteine sites are boxed, other conserved sites are highlighted in grey.

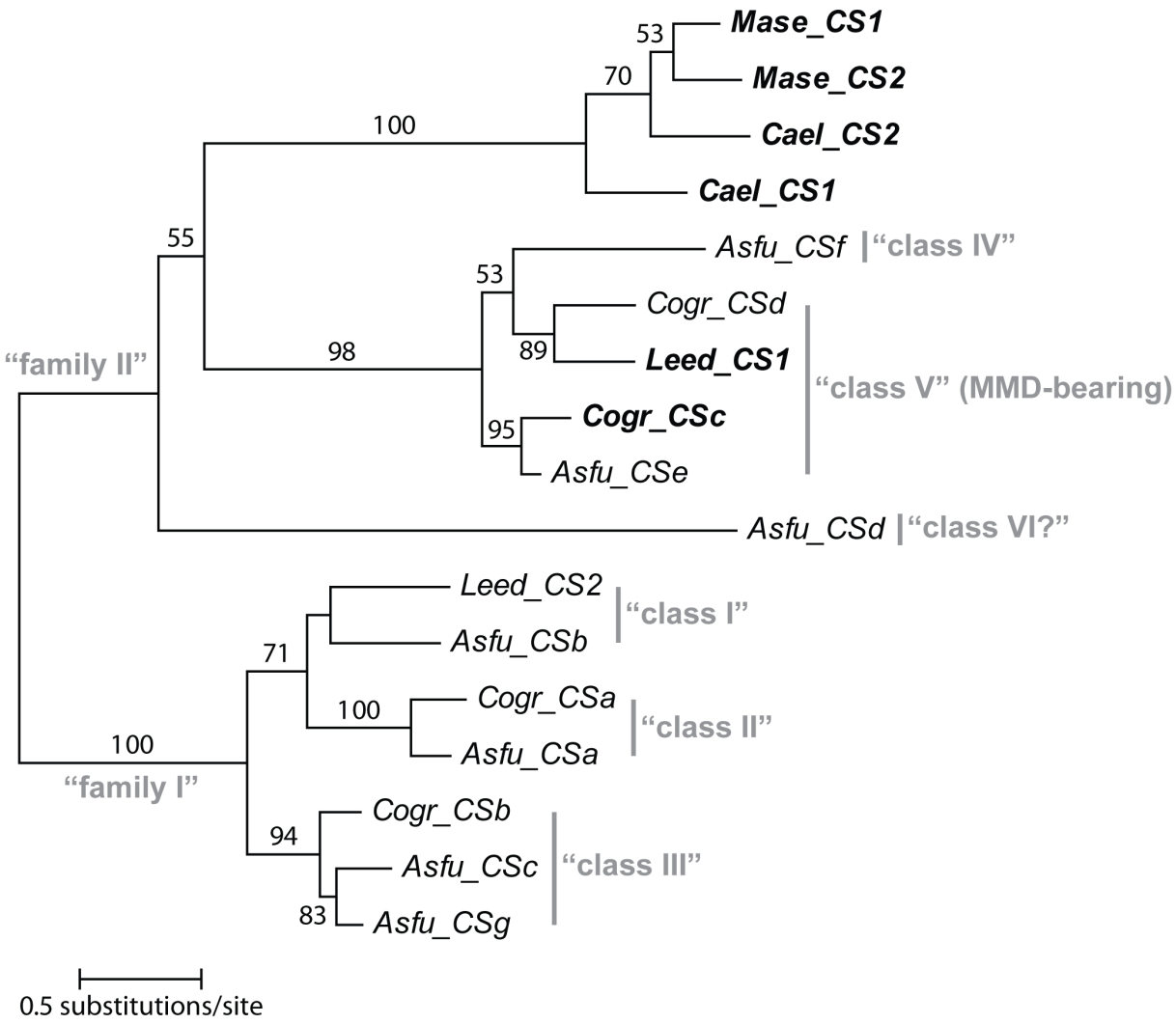


Fig. A2: Evolutionary tree of fungal chitin synthase domains. The ML phylogram (RAxML, JTT model) is derived from an alignment (449 amino acid positions) of 17 sequences, including representatives of the six fungal chitin synthase "classes" *sensu* Roncero (2002). Bootstrap values above 50% are shown. Chitin synthases that were used in the phylogenetic tree analyses of Fig. 31 and Fig. A3 are marked in bold. See Appendix for the sequence alignment.

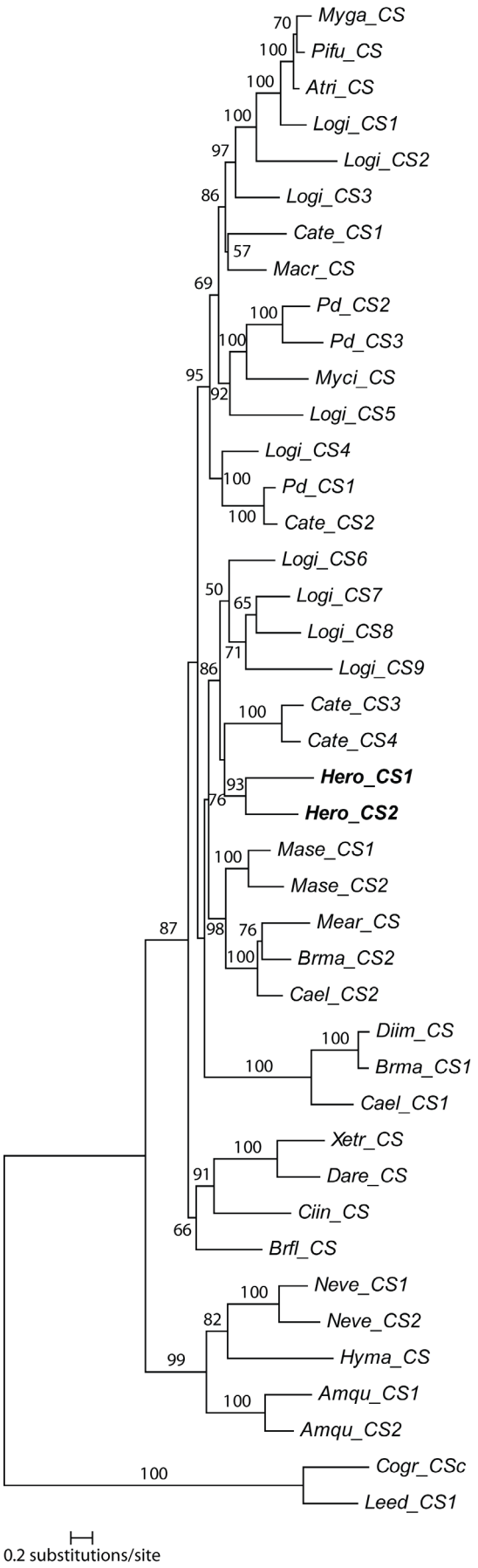
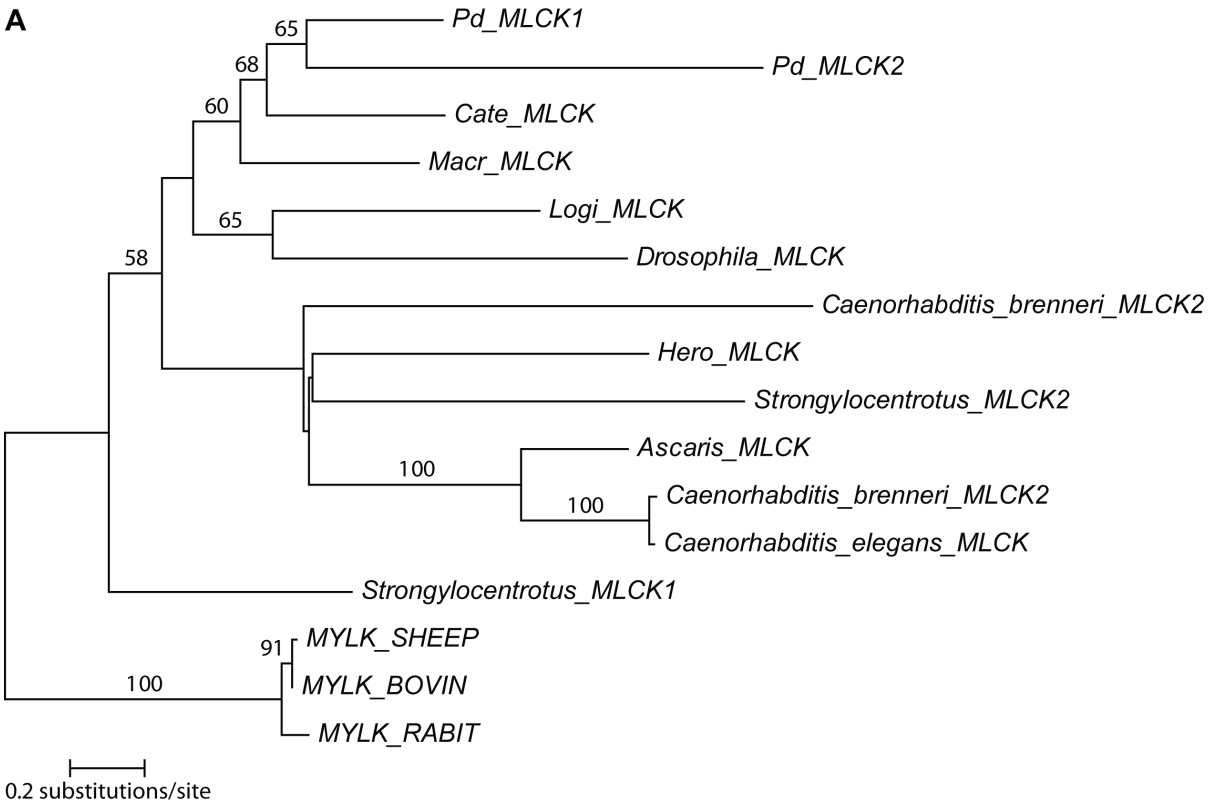


Fig. A3: Evolutionary tree of metazoan chitin synthase domains including *Helobdella robusta*. The ML phylogram (RAxML, WAGF model) is derived from an alignment (969 amino acid positions) of 42 sequences and bootstrap values above 50% are shown. See Appendix for the sequence alignment.



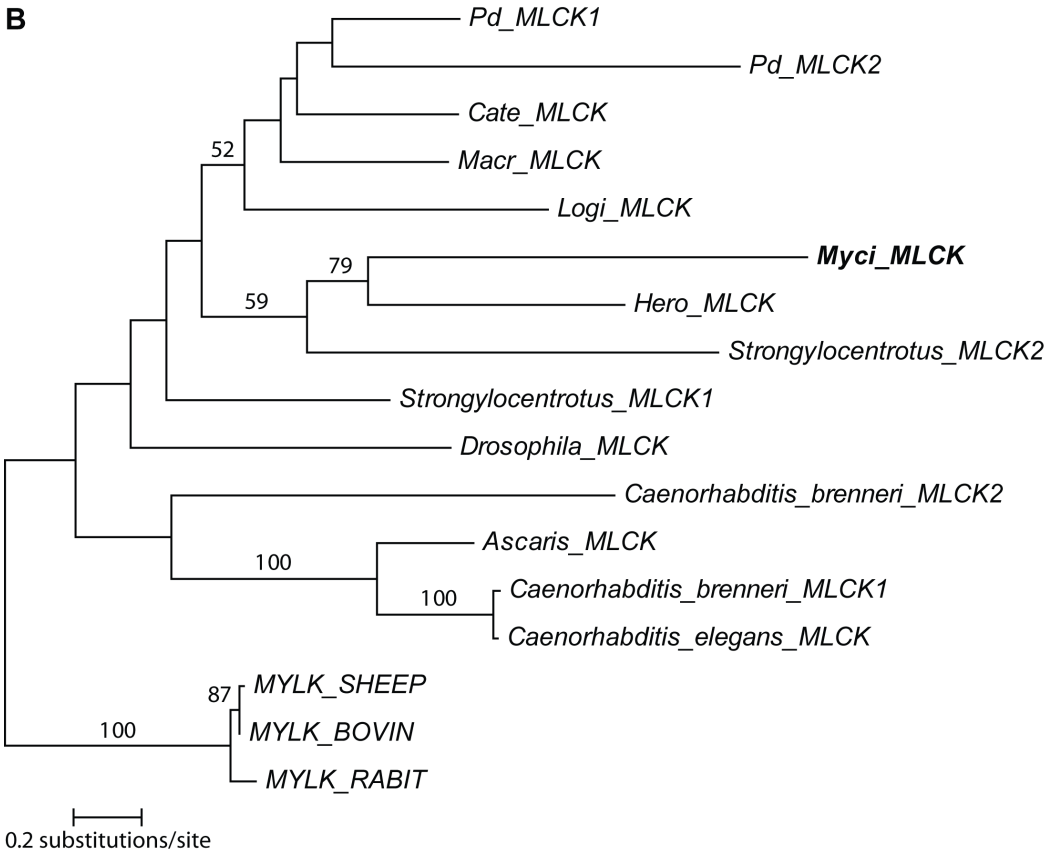


Fig. A4: Evolutionary trees of lophotrochozoan MLCKs. A – tree excluding *Myci_MLCK*. B – tree including *Myci_MLCK* (bold). The ML phylograms (RAxML, RTREVF model) are derived from an alignment (1077 amino acid positions) of 16 or 17 sequences and bootstrap values above 50% are shown. Note that the low resolution is probably due to the very short sequence fragments from some taxa. See Appendix for the sequence alignment.

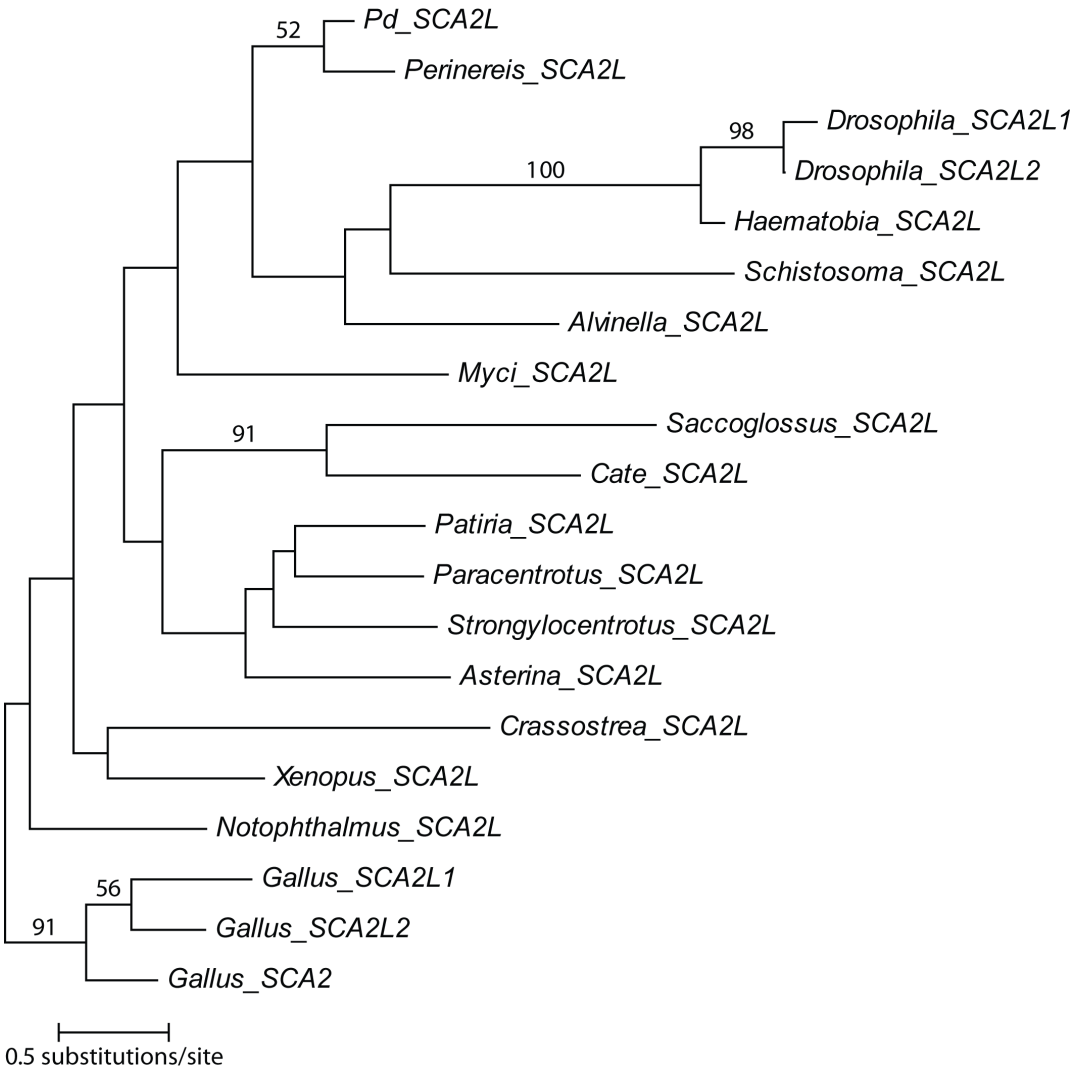


Fig. A5: Evolutionary tree of metazoan SCA2L proteins. The ML phylogram (RAxML, VTF model) is derived from an alignment (146 amino acid positions) of 20 sequences and bootstrap values above 50% are shown. Note that the low resolution is probably due to the short total alignment length. See Appendix for the sequence alignment.

Tab. A1: List of nested degenerate and specific primers used for degenerate PCR to achieve the MMD in *Pd_CS1*. Each degenerate primer was combined with each specific one.

primer ID	primer sequence (5'-3')
CS_deg_U1	TTYYANACXTAYATHGGXGA
CS_deg_U2	GGNGARWSNGGXDSXGGXAARACXGA
CS_deg_U3	MGNGAYGCNYTXGCXAARGMXYTXTA
CS_deg_U4	GGNTTYYTXGARAARAAYMGXGA
CS_deg_lo1	RTYTCNCKXGTXCCCCAXSWXAC
CS_Pdu_spec_U1	ACATTACTGCTGCAACAAGGATAC
CS_Pdu_spec_U2	ATGTGTATTGTGCAGTCCGGGATG
CS_Pdu_spec_lo1	TGACTCCATCTCTTCTTGTGTC
CS_Pdu_spec_lo2	ACGTCCGCCGTAGGGTGTGGG
CS5k_lo1	TGCTTCAGTCAGAAGGAC
CS5k_lo2	AATATGACACCAAAGGACGG
CS5k_lo3	GTTGCCAAAGATAGACTTG

Tab. A2: TAIL-PCR primers.

primer ID	primer sequence (5'-3')
TailN-1	NGTCGASWGANAWGTT
TailN-2	GTNCGASWCANAWGTT
TailN-3	WGTGNAGWANCANAGA
TailN-4	NGTCGASWGANAWCTT
TailN-5	GTNCGASWCANAWCAA
TailN-6	WGTGNAGWANCANTCT
TailN-7	NGTCGASWGANAWAGA
TailN-8	WGTGNAGWANCANGTT

Tab. A3: Fusion-PCR primers to amplify *Pd_MLCK1* with catalytic domain as well as the core region exclusively.

forward primer	primer sequence (5'-3')	reverse primer	primer sequence (5'-3')
Pd_MLCK_FUS_up1	GAAGAGAGGAAGAAGGATTTGG	Pd_MLCK_FUS_lo1	CGCCCTGAAGACTTCTCAATAC
Pd_MLCK_FUS_up2	GAAAGAAGGCTTGCGATGG	Pd_MLCK_FUS_lo2	CAGACGCTGCCGAATTTG
Pd_MLCK_cat_dom_up1	CAAATTCGGCAGCGTCTG	Pd_MLCK_cat_dom_lo1	CAGTATGTAACAAATAACTCCGATGC
Pd_MLCK_cat_dom_up2	GCCTATGTATTGAGAAGTCTTCAGGG	Pd_MLCK_cat_dom_lo2	GCTCCACATATCCGTAGCGTATG

Tab. A4: RACE PCR primers (including nested primers) to amplify upstream region of *Pd_CS1*.

primer ID	primer sequence (5'-3')	PCR template
Cs_primer lo1_5RACE	GAAGTAGCAACAGCGCCTGAGGTTC	5' RACE cDNA (mix of 48+50 hpf embryos)
Cs_primer lo2_5RACE	GCGAGCAAAAATGTGTTGTCAGACAG	5' RACE cDNA (mix of 48+50 hpf embryos)
Cs_primer lo3_5RACE	CTTGATGTCCTTTGTCTTTGACCACAGG	5' RACE cDNA (mix of 48+50 hpf embryos)

Tab. A5: List of all used GenBank and UniProt accessions.

homolog	organism	abbreviation	accession number
Pd_fcmg2	*Lumbricus rubellus*		CF810041
Pd_cbmg3	*Alvinella pompejana*		GO190881
	Lanice conchilega		FR763479
Pd_fcmg5	*Ridgeia piscesae*	*Ridgeia_fcmg5L*	EV802866
Pd_CAML	*Aedes aegypti*	*Aedes_TNNC*	XP_001662251
	Alvinella pompejana	*Alpo_CAML1*	GO174136
	Alvinella pompejana	*Alpo_CAML2*	FP548682
	Aplysia californica	*Aplysia_CAM*	NP_001191509
	Bombyx mori	*Bombyx_TNNC*	NP_001037594
	Danio rerio	*Danio_CAM*	BC045298
	Danio rerio	*Danio_TNNC*	NP_852475
	Doryteuthis pealeii	*Doryteuthis_CAM*	JK337034
	Drosophila melanogaster	*Drosophila_CAM1*	AAF58542
	Drosophila melanogaster	*Drosophila_CAM2*	CAA53630
	Drosophila melanogaster	*Drosophila_TNNC*	CAA53628
	Eisenia fetida	*Eisenia_CAM*	ABC48922
	Haliotis discus	*Hads_CAML1*	ABO26649
	Haliotis discus	*Hads_CAML2*	ADB02986
	Haliotis diversicolor	*Hadv_CAM*	ABU43070
	Haliotis diversicolor	*Hadv_CAML1*	ABY87375
	Haliotis diversicolor	*Hadv_CAML2*	ABY87354
	Homo sapiens	*Homo_CAM*	BG426626
	Homo sapiens	*Homo_TNNC*	NP_003270
	Hyriopsis cumingii	*Hyriopsis_CAM*	ADT61781
	Lanice conchilega	*Laco_CAML*	FR761557
	Lottia gigantea	*Lottia_CAM*	FC631001
	Lumbricus rubellus	*Lumbricus_CAM*	CAC14791
	Metridium senile	*Metridium_CAM*	Q95NR9
	Moniezia expansa	*Moniezia_CAML*	FE942139
	Osedax rubiplumus	*Osedax_CAM*	ACO55636
	Patinopecten sp.	*Patinopecten_CAM*	P02595
	Pinctada fucata	*Pinctada_CAM1*	AAQ20043
	Pinctada fucata	*Pinctada_CAM2*	AAV73912
	Salmo salar	*Salmo_TNNC*	ACH70760
	Strongylocentrotus intermedius	*Strongylocentrotus_CAM*	AB081568

Tab. A5: Continued.

homolog	organism	abbreviation	accession number
Pd_CS1	*Amphimedon queenslandica*	*Amqu_CS1*	XP_003389565
	Amphimedon queenslandica	*Amqu_CS2*	XP_003385441
	Aspergillus fumigatus	*Asfu_Csa*	EDP53850
	Aspergillus fumigatus	*Asfu_CSb*	EDP47283
	Aspergillus fumigatus	*Asfu_CSc*	EDP50922
	Aspergillus fumigatus	*Asfu_CSd*	EDP56496
	Aspergillus fumigatus	*Asfu_Cse*	EAL93639
	Aspergillus fumigatus	*Asfu_CSf*	EDP48750
	Aspergillus fumigatus	*Asfu_CSg*	EDP52317
	Atrina rigida	*Atri_CS*	Q288C6
	Branchiostoma floridae	*Brfl_CS*	XP_002602987
	Brugia malayi	*Brma_CS1*	Q9GQC3
	Brugia malayi	*Brma_CS2*	Q4VW80
	Caenorhabditis elegans	*Cael_CS1*	AAX62732
	Caenorhabditis elegans	*Cael_CS2*	O17368
	Colletotrichum graminicola	*Cogr_Csa*	AAL23717
	Colletotrichum graminicola	*Cogr_CSb*	AAL23718
	Colletotrichum graminicola	*Cogr_CSc*	AAL23719
	Colletotrichum graminicola	*Cogr_CSd*	AAL55424
	Danio rerio	*Dare_CS*	XP_691594
	Dirofilaria immitis	*Diim_CS*	Q9GQ90
	Hydra magnipapillata	*Hyma_CS*	XP_002162504
	Lentinula edodes	*Leed_CS1*	BAF41220
	Lentinula edodes	*Leed_CS2*	BAJ08815
	Manduca sexta	*Mase_CS1*	Q8T4U2
	Manduca sexta	*Mase_CS2*	AAX20091
	Meloidogyne artiellia	*Mear_CS*	Q8T5G8
	Mytilus galloprovincialis	*Myga_CS*	A5HKN1
	Nematostella vectensis	*Neve_CS1*	XP_001637059
	Nematostella vectensis	*Neve_CS2*	XP_001633545
	Pinctada fucata	*Pifu_CS*	A7BIC0
	Xenopus tropicalis	*Xetr_CS*	XP_002942397
Pd_MLCK1	*Ascaris suum*	*Ascaris_MLCK*	ADY40407
	Bos taurus	*MYLK_BOVIN*	Q28824
	Caenorhabditis brenneri	*Caenorhabditis_brenneri_MLCK1*	EGT59399
	Caenorhabditis brenneri	*Caenorhabditis_brenneri_MLCK2*	EGT41856
	Caenorhabditis elegans	*Caenorhabditis_elegans_MLCK*	CAA88976
	Drosophila melanogaster	*Drosophila_MLCK*	O01653
	Oryctolagus cuniculus	*MYLK_RABIT*	P29294
	Ovis aries	*MYLK_SHEEP*	O02827
	Strongylocentrotus purpuratus	*Strongylocentrotus_MLCK1*	XP_001195424
	Strongylocentrotus purpuratus	*Strongylocentrotus_MLCK2*	XP_001196303

Tab. A5: Continued.

homolog	organism	abbreviation	accession number
Pd_SCA2L	*Alvinella pompejana*	*Alvinella_CAM2L*	GO106308
	Asterina pectinifera	*Asterina_CAM2L*	DB387660
	Crassostrea gigas	*Crassostrea_CAM2L*	AM864987
	Drosophila melanogaster	*Drosophila_CAM2L1*	AW941054
	Drosophila melanogaster	*Drosophila_CAM2L2*	EB490223
	Gallus gallus	*Gallus_CAM2*	L34554
	Gallus gallus	*Gallus_CAM2L1*	XM_418412
	Gallus gallus	*Gallus_CAM2L2*	XM_418413
	Haematobia irritans	*Haematobia_CAM2L*	FD460073
	Notophthalmus viridescens	*Notophthalmus_CAM2L*	GO931319
	Paracentrotus lividus	*Paracentrotus_CAM2L*	AM570596
	Patiria miniata	*Patiria_CAM2L*	EX452553
	Perinereis nuntia	*Perinereis_CAM2L*	GT630674
	Saccoglossus kowalevskii	*Saccoglossus_CAM2L*	FF609909
	Schistosoma mansoni	*Schistosoma_CAM2L*	AM046898
	Strongylocentrotus purpuratus	*Strongylocentrotus_CAM2L*	CD293714
	Xenopus laevis	*Xenopus_CAM2L*	CB197007

Alignment for evolutionary analysis of CAM homologs

```
Alpo_CAML2              AMNDFTLEQIEEYREAFDLF DL---DK-DGKITLPDLRNM MTQLGQQISDDELKDIMHST DTKKRGYIDFEQFCHVMHTY -AP---------RTLEEQLR
Alpo_CAML1              YWNTFSRAQLEEYREAFDLF DL---NK-DGKITMPELRNM MTQLGQEITDKELKEIMDST DVKKRGYIDFEEFCHVMDTC -AP---------RTLEEQLR
Laco_CAML               KMATFTTEQLADYREAFEVF DV---NR-DGKITLTALRRI MNQLGQDVSEKELRDIMDSV DKEKHGYIDFNSFSQIMHTC -AP---------KSLEEQLR
Pd_CAML                 -TTRFNKDELSEFRESFEMF DV---NR-DGYITMEELRGM MESLDQKCSDIELQRIMKAA DLDKNGKIDFNEFVIIMDKF -CP---------ENFEDQLR
Cate_CAML               ----------DYREAFDMF DI---NH-DNRISTAELRKM MESLGQDPSEEELKQIMWSA DVNQNGYIEFDEFVNIMTGF -AP---------RDTQDQMR
Hadv_CAML2              -MAHLSQAEKTLYADYFLRM EAQLGNG-DAALSIAEFENL LQAMGFRFTKEQVAAEFIKT DTDQNWKITLEEFLAIMPNV -APP-----GSPQSIAADIR
Hads_CAML2              -MSHLSQQEKLLYSDFFLRL DAEQGDG-DGALNIAEFEKL LLALGFRLSKEQVAAEFIKT DKDQNWKITLDEFLVTLPNV -VPP-----GSPQSIAATIR
Hadv_CAML1              -MATLSRQEVKLYTEFFQQV DETDGDG-NGRLSLDNCEKL LDKCGHRIPRNVIQTFVNPA -GDGSWTLTLEEFLSAMSKV -VPP-----DSPHSKAAELR
Hads_CAML1              -MNHFDDDELAVYTKAFHDF DA---NH-DGYMTLSEFSRA CVALGFTYEEETIQTLFSEM DADKNEKVTLEEFLA----F -MGP-----DSEEKLVASVR
Moniezia_CAML           GQQEEQEERIAEYRESFSLF DK---NG-DGRISKSELKSV MKSLGQKVTDAEVEQMVRHV DLNHDGEIDFNEFVKMMSRY ----------KPASDEDKLF
Myci_CAML               LKTEFTTQQLEELRAAFGLF DK---DG-DGVITTDELWTV MTGLRQHPTEPQIKAMIRQV DIDGNGKVDFQEFLKMMSTN -RSRQISHEERLKAEEDEMR
Myci_CAM                MVEKLTEEQIEEFREAFALF DK---DG-DGSITTKELGTV MRSLGQNPTEVELQDMINEV DADGNGTIDFDEFLAMMAKK TLGQ--------GDTEDELK
Doryteuthis_CAM         MAKELSEKQIAEIKDAFDMF DI---DG-DGQITSKELRSV MKSLGRTPSDAELEEMIREV DTDGNGTIEYAEFVEMMAKQ -MGP--------TDPEKEMR
Drosophila_CAM2         -MSELTEEQIAEFKDAFVQF DK---EG-TGKIATRELGTL MRTLGQNPTEAELQDLIAEA ENNNNGQLNFTEFCGIMAKQ -MRE--------TDTEEEMR
Cate_CAM2               -----------EFREAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DEDGNGTIDFDEFLTMMERK -MKE--------TDTEEEMR
Pinctada_CAM2           MAEDLTEEQIAEFREAFHLF DK---DG-SGSISAEELGTV MRSLGQNPNEQELQDLVEEI DTDGNGEVDFNEFLAMMAKK -MKD--------TDSEEEIR
Lottia_CAM              ----LTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Lumbricus_CAM           MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Eisenia_CAM             -------------------- -------------------- ----------SELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Patinopecten_CAM        MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGDGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Drosophila_CAM1         MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Aplysia_CAM             MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Strongylocentrotus_CAM  MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Pinctada_CAM1           MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Osedax_CAM              ---------------FSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Homo_CAM                MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Danio_CAM               MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Hyriopsis_CAM           MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMAKK -LKD--------RDSEEEIR
Metridium_CAM           MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Hadv_CAM                MADQLTEEQIAEFKEAFSLF DK---DG-DGTITTKELGTV MRSLGQNPTEAELQDMINEV DADGNGTIDFPEFLTMMARK -MKD--------TDSEEEIR
Cate_CAM1               -----------EFREAFSLF DK---NG-DGKITTSELGTV MRSLGQNPTEAELQDMVNEV DSDGNGTIDFDEFLIMMAKK -MKE--------TDSEEELR
Drosophila_TNNC         MSDELTKEQTALLRNAFNAF DP---EK-NGYINTAMVGTI LSMLGHQLDDATLADIIAEV DEDGSGQIEFEEFTTLAARF -LVE-----EDAERMMAELK
Aedes_TNNC              -MMEDEEQRLIIMRKAFQMF DT---TK-CGFIDTVKISTI LNTLGQQFDEGELQDLIDEE DPEETGRVNFDQFANIASNF -LIE----DEDSEAIQQELK
Bombyx_TNNC             WADELPPEQIAVLRKAFDGF DH---NR-SGSIPCDFVADI LRMMGQPFNKKILEELIEEV DADKSGRLEFPEFVTLAAKF -IVE-----EDAEAMQKELR
Danio_TNNC              AAEQLTDEQKNEFRAAFDIF VQ---DAEDGCISTKELGKV MRMLGQNPTPEELQEMIDEV DEDGSGAVDFEEFLVMMVRC -MKD-----DSKGRPEEELA
Salmo_TNNC              ARSFLSEEMLNEFKAAFDMF DT---DG-GGDISTKELGQV MRMLGQNPTRQELDEIIEEV DEDGSGTIDFEEFLVMMVRL -LKE-----DQAGKSEEELA
Homo_TNNC               ARSYLSEEMIAEFKAAFDMF DA---DG-GGDISVKELGTV MRMLGQTPTKEELDAIIEEV DEDGSGTIDFEEFLVMMVRQ -MKE-----DAKGKSEEELA

Alpo_CAML2              ETFTFIDADANGRITSAELR QVL--QMIGDDDVTDEQVD- ---ELIREIDLDGDGEIDFE EFVNFMLSK
Alpo_CAML1              ETFRLIDRDENGRISSQELR EVL--RKIGDKDITDEEIE- ---DLIREIDLDGDGEIDYE EFVMIMLSK
Laco_CAML               EAFQLIDKDNSGKISSTELK QVL--RRIGENDITDDEIE- ---DLVREIDLDGDGEVDYE EFVKIMVSK
Pd_CAML                 EAFRVIDNDNSGKISSDEIK RIL--RHLGEKKMSSEVAE- ---MIIREMDMDGDGQVDYE EFVKMFTNK
Cate_CAML               EAFKIFDRDNQGSISATELK RVM--SNLG-DKMTDAEID- ---EIIREIDLDGDGQVDYE GLL------
Hadv_CAML2              RKFNENDLNKDGFITFDELK KVL--AEAN-QDKTEEELRT FIKDFIKNWDADGDGKINFE EFLEMYIET
Hads_CAML2              RRFDENDLNKDGYITFDELK QVL--ADAN-QGKDEEELKS LVKDFIKNWDADGDGKINFE EFLEMFIET
Hadv_CAML1              RLFRAYDLNKDGYVSFYELK NTM--R----GSITEEELR- ---DFIKDWDKDGDWKLSFE EFINMYVAP
Hads_CAML1              RIFNEYDTNKDGFLTADEIQ SAWFSGERGTKKLSKSEVA- ---EIMAPVDTNGDGKLNYE EFLQLVLSA
Moniezia_CAML           EAFKVFDKDGNGYISVSEMR QVM--SNLG-EKLTDEEVK- ---DMIRSADTDGDGQVSFD EFRNVMTHR
Myci_CAML               QAFKVFDMDGNGYIDANELK CTM--RNLG-ENLTDADVR- ---SMIQAADVNGDGQIDYE EFIKMMYAK
Myci_CAM                EAFKVFDKDGNGYISSEELR HVM--TNLG-EKLTDEEVA- ---EMIREADIDGDGQVNYE E--------
Doryteuthis_CAM         EAFRVFDKDGNGLITAAELR QVM--ANFSDEKLTSEEIS- ---EMIREADIDGDGMVNYE EFVKMMTPK
Drosophila_CAM2         EAFKIFDRDGDGFISPAEIR FVM--INLG-EKVTDEEID- ---EMIREADFDGDGMINYE EFVWMISQK
Cate_CAM2               EAFRVFDKDGDGFISAAELR HVM--ANLG-EKLTEQEVD- ---EMIKEADINGDGKVDYE RGVVSVMYN
Pinctada_CAM2           EAFRVFDRDDKGFITASELK HIM--TTLG-EKMDDEEAE- ---EMVAAADIDGDGEINYE EFVKMISMK
Lottia_CAM              EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIIRG----------- ---------
Lumbricus_CAM           EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVTMMMSK
Eisenia_CAM             EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREA----------- ---------
Patinopecten_CAM        EAFRVFDKDGDGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVTMMTSK
Drosophila_CAM1         EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVTMMTSK
Aplysia_CAM             EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVTMMTSK
Strongylocentrotus_CAM  EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVTMMTSK
Pinctada_CAM1           EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVKMMMSK
Osedax_CAM              EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDG------- ---------
Homo_CAM                EAFRVFDKDGNGYISAAELR HVM--TKLG-EKLTDEEVA- ---ETDQGSRSRW------- ---------
Danio_CAM               EAFRVFDKDGNGYISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVQMMTAK
Hyriopsis_CAM           EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVQMMTSK
Metridium_CAM           EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVKMMTSK
Hadv_CAM                EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDEEVD- ---EMIREADIDGDGQVNYE EFVKMMTSK
Cate_CAM1               EAFRVFDKDGNGFISAAELR HVM--TNLG-EKLTDDEVD- ---EMIREADLDGDGMVNYE EFVMMMTSK
Drosophila_TNNC         EAFRLYDKEGNGYITTGVLR EIL--RELD-DKLTNDDLD- ---MMIEEIDSDGSGTVDFD EFMEVMTGG
Aedes_TNNC              EAFRMYDREGNGYITTSTLK EIL--AALD-DKLTNDDLD- ---GMINEIDADGSGTVDFD EFMEMMTGE
Bombyx_TNNC             EAFRLYDKEGNGYIPTSSLR EIL--RELD-EQLTDDELD- ---GLIQEIDTDGSGTVDFD EFMEMMTGE
Danio_TNNC              ELFRMFDKNADGYIDLDELK LML--EATG-EAITEDDIE- ---ELMRDGDKNNDGKIDYD EFLEFMKGV
Salmo_TNNC              ECFRVFDKNADGYIDREEFA III--RSTG-EQISEEEID- ---ELLKDGDKNADGMLDFD EFLKMMENV
Homo_TNNC               ECFRIFDRNADGYIDPEELA EIF--RASG-EHVTDEEIE- ---SLMKDGDKNNDGRIDFD EFLKMMEGV
```

Alignment for evolutionary analysis of fungal CS homologs

```
Mase_CS1                NTFLLTLDGDIDFQPHAVRL LIDLMK----KNKNLGAACG RIHPVGS--------GPMVW YQLFEYAIGHWLQKATEHMI GCVLCSPGCFSLFRGKALMD
Mase_CS2                NTYLLALDGDIDFKPSAVTL LVDLMK----KDKNLGAACG RIHPVGS--------GFMAW YQMFEYAIGHWLQKATEHMI GCVLCSPGCFSLFRGKALMD
Cael_CS2                NTFILTLDGDVDFTPSSVYL LVDLMK----KNRRLGAACG RIHPRGD--------GAMVW YQKFEYAIGHWLQKATEHMI GCVMCSPGCFSLFRAYALMD
Cael_CS1                NTFILAIDGDSKFEPDALLR LLHLMN----AKSDIGCACG RIHPIGN--------GIMVW YQKFEYAIAHWFQKAAEHVF GCVLCAPGCFSLFRASALMD
Asfu_CSf                YEVVLMVDADTKVFPDSLTH MISAMV----KDPDVMGLCG ETKIANKTD------SWVTM IQVFEYFISHHQSKAFESVF GGVTCLPGCFCMYRIKAPKG
Cogr_CSd                YEYLMMVDADTCVREDSLNR LVSACA----NDAKIAGICG ETGLQNDDK------SWWTM IQVYEYFISHNLAKAFESLF GSVTCLPGCFTMYRLRTVDK
Leed_CS1                YEYIFTIDADTTVTPESLNR LVASAA----IDSNIIGICG ETKVANEDT------SWWTM IQVYEYYISHHLTKAFESLF GSVTCLPGCFSLYRIRTADK
Cogr_CSc                YEFMFQIDADTVVAPDSATR MVSAFL----YDTRLIAVCG ETALTNAKA------SFITM IQVYEYWISHNLTKAFESLF GSVTCLPGCFSMYRIRAAES
Asfu_CSe                YEFILQVDADTVVAPDSATR MVAAFL----NDTRLIGVCG ETALTNAKN------SAVTM IQVYEYYISHNLTKAFESLF GSVTCLPGCFTMYRIRSAET
```

```
Asfu_CSd        NLFVLFIDSDCILDRVCLQN FMYDMELKPGSKHDMLAMTG VITSTTDRG------SLLTL LQDMEYVHGQLFERSVESSC GAVTCLPGALTMLRFSAFRK
Leed_CS2        PNVCILLDVGTKPTGTSIYE LWKCFD----KHSNVGGACG EICVDTGRGCSLLLTSPLAA SQNFEYKMSNILDKPLESVF GYISVLPGAFSAYRYAALQN
Asfu_CSb        PNICILLDVGTKPEPTALYH LWKAFD----QDSNVAGAAG EIKAGKGKNMMGLL-NPLVA SQNFEYKMSNILDKPLESVF GYITVLPGALSAYRFFALQN
Cogr_CSa        PNICVLIDAGTKPGGNSIYH LWKAFD----LEPMCAGACG EIKAMLGTGGKHLL-NPLVA TQNFEYKMSNILDKPLESAF GFISVLPGAFSAYRYVALQN
Asfu_CSa        PNICVLLDAGTKPGRDSIYQ LWKAFD----VEPMCGGACG EIKVMLSHG-KKLL-NPLVA GQNFEYKLSNILDKPLESAF GFISVLPGAFSAYRYVALQN
Cogr_CSb        PEVCILIDAGTKPSPRSLLA LWEGFY----NDKDLGGACG EIHAMLGKGGKKLL-NPLVA VQNFEYKISNILDKPLESSF GYVSVLPGAFSAYRFRAIMG
Asfu_CSc        PEICILLDAGTKPGSKSLLA LWEAFY----NDKTLGGACG EIHAMLGRGWRNVL-NPLVA AQNFEYKISNILDKPLESAF GYVSVLPGAFSAYRYRAIMG
Asfu_CSg        PEVCILLDAGTKPGPKSLLS LWEAFY----NDKDLGGACG EIHAMLGKGWKNLI-NPLVA AQNFEYKISNILDKPLESSF GYVSVLPGAFSAYRFRAIMG

Mase_CS1        -----------DNVMKKYTL RSD-------------EARH YVQ-YDQGEDRWLCTLLL-- -QRGYRVEYSAASDAYTHCP EGFSEFYNQRRRWVPSTVAN
Mase_CS2        -----------DNVMKKYTL TSN-------------EARH YVQ-YDQGEDRWLCTLLL-- -QRGYRVEYSAASDAYTHCP ERFDEFFNQRRRWVPSTMAN
Cael_CS2        -----------DNVARRYAL KSE-------------EPKH FIQ-YDQGEDRWLCTLLL-- -QRGYRVEYCAASDAQTFAP EGFNEFFNQRRRWIPSTIFN
Cael_CS1        -----------DNIMHKYTK TAS-------------EPRH YVQ-YDQGEDRWLSTLLL-- -KQGYRIEYAAASDAETYAP EGFEEFFNQRRRWTPSSIAN
Asfu_CSf        GQNYWVPILANPDVVEHYSE NVV-------------DTLH KKNLLLLGEDRYLSTLML-R TFPKRKQIFVPQAVCKTVVP DKFMVLLSQRRRWINSTVHN
Cogr_CSd        GK----PLIIADGVIRDYAV CDV-------------DTLH KKNLLSLGEDRYLTTLMT-K YFPSMKYKFIPDAYCQTAAP ESWSVLLSQRRRWINSTIHN
Leed_CS1        GR----PIIISNRVIDEYAE PNV-------------DTLH KKNLFSLGEDRYLTTLLM-K HFPTFKTKFNPDAVAHTVAP ESTRVLFSQRRRWINSTVHN
Cogr_CSc        GK----PLFVSREVVEAYAT IRV-------------DTLH MKNLLHLGEDRYLTTLLL-K FHNKYKTKYTNRAHAWTIAP DSWAVFLSQRRRWINSTVHN
Asfu_CSe        AK----PLFVSKEVVDAYAE IRV-------------DTLH MKNLLHLGEDRYLTTLLL-K HHSKYKTKYISSAKAWTIAP ESWTVFLSQRRRWINSTVHN
Asfu_CSd        --------------MAKYYF ADKAEQ---------CEDFF DYGKCHLGEDRWLTHLFMVG ARKRYQIQMCAGAFCKTEAV QTFSSLLKQRRRWFLGFITN
Leed_CS2        GPDGKGP-------LASYFK GETMHG-----GGPTGAGLF ERN-MYLAEDRILCFEIVTK KREGWVLKYVKSAKASTDVP ASVPEFISQRRRWLNGSLFA
Asfu_CSb        DADGNGP-------LNQYFK GETLHG--------KDADVF TAN-MYLAEDRILCWELVAK REERWVLKFVKSAVGETDVP DTIPEFISQRRRWLNGAFFA
Cogr_CSa        DKNGQGP-------LEKYFA GEKLEG--------AGAGIF TSN-MYLAEDRILCFELVTK RNCHWILQYVKSATGETDVP DTVTELVLQRRRWLNGSFFA
Asfu_CSa        DKNGQGP-------LERYFL GEKMHG--------ANAGIF TAN-MYLAEDRILCFEIVTK RNCRWLLQYVKSSTGETDVP DRMAEFILQRRRWLNGSFFA
Cogr_CSb        R-----P-------LEQYFH GDHTLSKILGKKGIDGMNIF KKN-MFLAEDRILCFELVAK AGQKWHLSYIKAAKGETDVP EGAAEFISQRRRWLNGSFAA
Asfu_CSc        R-----P-------LEQYFH GDHTLSKRLGKKGIEGMNIF KKN-MFLAEDRILCFELVAK AGYKWHLTYVKASKGETDVP EAAPEYISQRRRWLNGSFAA
Asfu_CSg        R-----P-------LEQYFH GDHTLSKQLGKKGIEGMNIF KKN-MFLAEDRILCFELVAK AGSKWHLTYVKASKAETDVP EGAPEFISQRRRWLNGSFAA

Mase_CS1        IMDLLVDCKHTIKINDNIST PYI-----AYQMMLMGGTIL GPGTIFLMLVGAFVAAFRID ------------------NW TSFEYNLYPIAIFVFVCFTM
Mase_CS2        IFDLLADSKRTVQVNDNIST LYI-----VYQCMLMMGTIL GPGTIFLMMIGAINAITGMS ------------------NM HALLFNLVPVLTFLVVCMTC
Cael_CS2        IMDLLKDYRNVVRVNESISI WYI-----IYQLVMLISSIL GPGTIFVMIIGAISISFSID ------------------TL ISLVIVSIPVVVFIVVCLTA
Cael_CS1        TVDLLMDYKRASENNDAISY AYI-----AYQFLVIFFSML GPAIIFTMLVFAQVAAFELR ------------------GS DVMLYNGIPIGFFIVLCFTT
Asfu_CSf        LMELVLVRDLC--------- ---------------------- ---------------------- ---------------------- -------------GTFCFSM
Cogr_CSd        LFELMKLKEMC--------- ---------------------- ---------------------- ---------------------- -------------GFCCFSM
Leed_CS1        LCELVMLPELI--------- ---------------------- ---------------------- ---------------------- -------------GFCCFSM
Cogr_CSc        LMELIPMSQLC--------- ---------------------- ---------------------- ---------------------- -------------GFCCFSM
Asfu_CSe        LIELIPMQQLC--------- ---------------------- ---------------------- ---------------------- -------------GFCCFSM
Asfu_CSd        EVCMLTDVRL---------- ---------------------- ---------------------- ---------------------- ----WKRYPL-----------
Leed_CS2        SIHSTVFWFRIWTSGQNFFR KIVLTIEFIYNAVQLLFTWT SLANFYLAFYFLVSSATSDS TNDPFK------FMAEGAGK DVFEVVLKVYIALIFVVLVC
Asfu_CSb        AVYSLVNVKQLWKTDHSVPR KILLQIEAFYQFLNLLFTYF GLANFYLAFFFIAGSLSDER ID----------PFGHNMGK YIFFVLRYACILVMCLQFII
Cogr_CSa        GIYAIAHFYEFFRSDHSMLR KLMFFVEFVFNTINLIYAWF AIGNFFLVFKILTTSLGDDN -------------LLGRTGE ILGVVFTWVYGIALITCFVL
Asfu_CSa        AVYAIAHFYQIWRSDHSFMR KFMLLVEFIYQTINMIFAWF NIGNFFLVFHILTTYLGDAE -------------LLGTTGK VLGVVFEWLYLATLVTCFVL
Cogr_CSb        SLYSLMHFGRMYKSGHNLVR MIFFHIQLVYNILQVLFTWF SLGSYYLTTTVIMDLVGNPV VSEDPKLARHGWPFGDTATH IFNALLKYLYLAFVILQFIL
Asfu_CSc        SLYSIMHFGRIYKSGHSFVR MFFLHIQMIYNCCQLIMTWF SLASYWLTSSVIMDLVGTPS SHNKYK----AWPFGNDASP IVNFFVKYGYLLVLMLQFVL
Asfu_CSg        GIYSLMHFGRMYKSGHNIVR MFFLHIQMLYNIFSTVLTWF SLASYWLTTTVIMDLVGTPS DNNGNK----AFPFGKTATP IINTIVKYVYLGFLLLQFIL

Mase_CS1        K-------SEYQLLVAQILS TAYAMIMMAVIVGTALQLGE DG------------------ -----IGSPSAIFLISLSSS FFIAACLHPQEFWCIVP--G
Mase_CS2        K-------SETQLMLANLIT CFYAMVMMFVIVSIVLQISQ DG------------------ -----WLAPSSMFTAATFGI FFVTAALHPQEIICLLY--I
Cael_CS2        K-------PEHQLICAQTIG AIFAMLMTAVVVGTSLQLQK DG------------------ -----LLSPHSMFTVAVATS FLTAAILHPLEFTCIIP--G
Cael_CS1        E-------SNIQLIYAKYMS IAYAFVMLAVLVATSSQIVL ET------------------ -----VLAPTSLFIVTMVGI FFFAACLHPKEFTNIIH--G
Asfu_CSf        Q-------F---VVFIELVG TLVLPAAIAFTFYVVIISII KK------------------ ----PVQIIPLVLLALILGL PGVLIVVTAHRLVYVLW--M
Cogr_CSd        R-------F---VVFIDLFG TIILPATTIYLGYMIYLAAS ST------------------ --GQFP-IISIIMLAAVYGL Q-ALIFILKRQWQHIGW--M
Leed_CS1        R-------I---FVFVDLLS TMILP---STLVYLVIVVSL GK------------------ --YRFP-LIAIIMIAAGYGL Q-VLIFILKREFMLIGW--M
Cogr_CSc        R-------F---IVFVDLLS TIVQPVVVAYIIYLIVMVVQ NP------------------ --TVVP-VAAFVMLGAIYGL Q-AIIFILRRKWEMIGW--M
Asfu_CSe        R-------F---VVFVDLLS TVIQPVTLAYIIYLIYWLVK DT------------------ --STIP-YTSLILLAAIYGL Q-ALIFIIRRKWEMVGW--M
Asfu_CSd        ------------LCLVRFMQ NTIR--TTALLFFIIALSLI TT------------------ --SSSINDLPVGFIAISLGL NYVLMFYLGAKLKRYKAWLF
Leed_CS2        SLGNRPQGSKWTYILAIVLF GLCNVVTLYCAGYTVYLAVP HTVA-------GWEDFPALV K-STPALQDIVISLAATYGL YFIGSFMHFEPWHMFTS--F
Asfu_CSb        SMGNRPQGAKKLYMSGIIVY SIIMVYTAFCALYLVVLELM AKAGVGKK---------ELA V-SDSLFINIVVSLLSTVGL YFYSSFLYLDPWHMFTS--S
Cogr_CSa        AMGNRPAGSGPYYITMVYFW AFIMIYLLFAAVFIAVKAII ADVH------DSNGFNITDL F-KNPVFYTLIISVMSTYGI WLIASLLMFDPWHMITS--F
Asfu_CSa        SLGNRPGGSNKFYMTMVYFW IGIMIYLTFAAIFVTVKSIQ KEVA-------DNSFSVGQL F-SNSQFFSIFVSLGSTYVM WLLASLIFLDPWHMFTS--F
Cogr_CSb        ALGNRPKGSKYTYIASFVVF GLIQSYILILSMYLVVQAFQ TPLSQQISLDSG-KDFVQSF FGGTNAAGVILVALVTIYGL NFIASFMYLDPWHMFTS--F
Asfu_CSc        ALGNRPKGTKLAYTMSFLWF SLVQFYVLILSFYLVANAFM GGM-IDFDFDQGVGNFLSSF F-SSTGGGIVLIALVSTYGI YIVASILYMDPWHILTS--S
Asfu_CSg        ALGNRPKGSKFSYLASFVVF GIIQVYVVIDALYLVVRAFS GSAPMDFTTDQGVGEFLKSF F-SSSGAGIIIIALAATFGL YFVASFMYLDPWHMFTS--F

Mase_CS1        IIYLLSIPSMYLLLILYSII NLNVVSWGTREVQTKKTKKE IELH----I
Mase_CS2        SIYYITIPSMYMLLIIYSLC NLNNVSWGTREVAQKKTAKE MELQ----I
Cael_CS2        TIYFLAIPCMYMLLPIYSVC NMHTVSWGTREDPRPTEKNT LAKKTPGNK
Cael_CS1        VVFFLMIPSTYVFLTLYSLI NLNVITWGTREAVAKATGQK TKKAPMEQF
Asfu_CSf        FIYLLSLPIWNFVLPTYAYW KFDDFSWGDTRKTAGEQDKG HE---AGEG
Cogr_CSd        IIYIMAFPIYSFALPIYSFW NQDNFSWGNTRIVIGEK-GN KQLVAVDDE
Leed_CS1        VVYLLSYPIYSLFLPVYSFW CMDEFGWGNTRLVIGEG-KD KKVIVNDDD
Cogr_CSc        ILYIVAIPVFSFALPLYSFW HMDDFNWGNTRVIAGEA--G KKIVISDEG
Asfu_CSe        IVYLLALPVFSLALPLYSFW HMDDFTWGNTRIITGEK--G RKVVISDEG
Asfu_CSd        PLMFILNPFFNWLYMVYGIL TAGQRTWGGPRADAATADE- --------H
Leed_CS2        IQYLFFLPSYVNILMMYAMC NLHDVTWGTKGDNGAAKDLG GAKKVKGEG
Asfu_CSb        AQYFALLPSYICTLQVYAFC NTHDVTWGTKGDNTINNDLG AARIINGTT
Cogr_CSa        VQYMLLTPTYTNILNVYAFC NTHDISWGTKGDDKAES-LP TVSTKDGSG
Asfu_CSa        IQYMLLTPTYINVLNIYAFC NTHDITWGTKGDDKAEK-LP SANMKPG-G
Cogr_CSb        PHYLVLMSTYINILMVYAFN NWHDVSWGTKGSDKTEA-LP SAQVSKGEK
Asfu_CSc        WAYFLGMTTSINILMVYAFC NWHDVSWGTKGSDKADA-LP SAQTKKADG
Asfu_CSg        PAYMCVQSSYINILNVYAFS NWHDVSWGTKGSDKADA-LP SAKTTKDEG
```

Alignment for evolutionary analysis of metazoan CS homologs

```
Myga_CS         LYLFFIVMILGCTVASKMSL LLITSGISKDENARGENIVL LVICLCAPIGWNWLNSFMKI LFGGKTWPSMKTFFVLLLFE GLQTFGMCLLLFRVLPSTDF
Pifu_CS         LYFFFFIMILGCAVTSKMSL LLLTSGINKDTESRGEHMVG LLFCMCGPMLWNWLMAFMKI LFGGKEWPSFTTFAILLLLE CLQTFGLCLLLFRILPSTDF
Atri_CS         LYVFFVIMILGSIVASKMSL LLLTSGINKDENSRGEHLVT LLFCMCGPMLWNWFMAFMRI LFGGKEWPSMKTFLILLLFE NVQTFGMCLLLFRVLPSTDF
Logi_CS1        LYLFVVIVILASTVASRISL LLMTSGIKGDEKSKGESSVV LLFCLCSPILWSWLNAVMKI LFGGKEWPSLKMFIIMLLLE MMQIVGMSLLLFRVLPGTDF
Logi_CS2        MYIFLLVFVCGSCSVSKLSM YWISKSVIVTSKHRNKLFIQ LIWCITVPMIYKWCLTMFKA LFATSNKPSKKAIFVLTVVD VFQTCGRCVMILLVMPYLSP
Logi_CS3        CYQALFITVLGSAVVSKLSL LTMTSGIELGNNDANPTHTM LLISVCFPYFCWIFAYTFKS LFGSTGWPSITMTIITLILE LLHSFGLCLLLFRILPKVDI
Cate_CS1        VYALLFLVLLASLVVSKGSL ILMASSLKIIIGRRDTMVTL LVTCMCVPHVLTFIVHVFKG LFGSQPWPSIVEFIQILVLE SIQTAAMCALVFIVLPELDI
```

```
Macr_CS          MYVFLFVMVLGTAIVSKVGL LLLTSGSLMHTA-----LTL LVIVICIPYALTFITHFGKA LFGSQPWPSWRTTTCILSVE LLHSFGLSLLVFRILPSLDM
Pd_CS2           MYFVIFLLVLGSATISQLCH LMLTAGVGINSDRWKRAVLQ LLAIICVPRVLTFIFSLVKS WFGHMPWPTTRLFLFVMLIE TLQTTGETLLVIRVLPRFGT
Pd_CS3           -------------------- -------------------- -------------------- -------------------- --------------------
Myci_CS          -------------------- -------------------- -------------------- -------------------- --------------------
Logi_CS5         LYLLFFCIVLGCCVCQKLSL IITVQNINPTNADANNYIKN LVLAICIPTSLNILISTSKL LFGHFKCPSVITFILAIFCE VIHAFGLCLLVYGVLPYCNT
Logi_CS4         AYLCIFICVLTSALASKGSL LLMTHAIGNVKQNRDRWVYM VIAVMCLPYLVSFMECLFKV MFGNYPRPSIGNLSWILMIE SIHSLGLCLFMFRVLPCLDT
Pd_CS1           -----------------VSL LIMTTGAGDLRKERDLIIAM LAFAICIPNLFIFAESLLKS IFGNKAWPSFGVIFVVLLTE AAHTFGLCLLVYRVLPQFDG
Cate_CS2         VYLIFFAIIICCAFASKVSL LAMTAGVGDLAKERDLIIIM LTLAMCVPNLFVFLESLLKS VFGNQKWPTFGTIFVVLLTE TAHCFGLCLFLFRVLPQFDV
Logi_CS6         VGIFLTLAVLMSTVLSKLSL LLITSNIYNNMTVEIRWIWA LFACVVIPYVFNFGRSMWRM CFKSTRNPTWGVLLAVLLFE SIHSTGICIFTFHLLPHIDS
Logi_CS7         ASLFLFTTILGTAIISKLSF LFMTYHINPIQKLDVKWLWS LFLSMCLPYAFIFVGNVWVN CFKTTRSITLLPFITVFVTE SLHSVGLCLLVFYVIPSFDP
Logi_CS8         LCTFFFLLVLGTSTLSKLSF IFITFYINPGTGIDIKWIWS LILILCAQYLFTFLHSLWKM IFKKTGSIRCIPFLVSMVTE TLHSIGLCTFIFYVLPSFDP
Logi_CS9         LCIIFFLIVLATSLITKLTF IIIAYNINIPSNTDIRWIWS LVLIIIAPYIFNLAKCLHTQ CFKKAKKMKWKYLPFTLIIE TIHSIGLGIFVYIASSSGDA
Cate_CS3         FIGMISLIVLACLVTSKISL FFITSGMNPHFKNNIRWIWA LHLSMCVPYMLCMVKSIWKV TFKTKTSPDWGTLAWVLVVE SVHSCGLFFLAFLGLPKLDV
Cate_CS4         FILLLATLVLGCLTSSKVSL LFITASMNPFFKNNVRWIWS LHMAMCVPYVVCCLKSLWKV TFKMKAAPKGSTVLWVLLVE SLHTCGLFMFAFLGLPRLDA
Hero_CS1         -------------------- -------------------- -------------------- -------------------- --------------------
Hero_CS2         -------------------- -------------------- -------------------- -------------------- --------------------
Mase_CS1         AYLVVFVVVLGSGVLAKGTM LFMTSLLKKDRRERVAWMWA VLAAFAIPELGTLIRSIRIC FFKSSKRPSFVQFMVVFIAE SLHTIGMGLLFFKILPELDV
Mase_CS2         AYILVFCAVLGSAVIAKGTL LFITSQLKKDRQERITWLWA ALIVFGVPELGVFLRSVRIC FFKTAKKPTKTQFIIAFITE TLQAIGIAALVLIILPELDA
Mear_CS          TFVALFVTTLGSSILAKLSL LVMAAGLLISPKNAAKWAWA LLLAICIPELLCFARSLHRS LFRKVRGPSFLQFLLVFTVE SVHAFGLGALVFAIMPRGMV
Brma_CS2         CFCVLFILTLLSAIISKATF LLMTSAIGTIKMHVVKWIWA VYISLCIPELQCFIGSIHRA FFRNVKKPTFAQFIIVFVVE TLNAIGVSILAFKILPNLDA
Cael_CS2         TFLVLFLLTLGSAVVAKSTF ILMTSAIKLKNAHVVKWVWA TLLALSAPEALCFVRSMHRT MFRNVKRPTFIQFVFVLIIE TFHSIGVGILVFRIFPDLDA
Diim_CS          MFLVSHLLLIFGAATSKLIV ILLATNIARSSEVTISVIFA LWLIQNIPDVMAIIQSLYRI Y--KYPKERKCATLLLFVME CCRSIGIATLFFHVFPSLDP
Brma_CS1         IFVVSHLFLIIGAAISKLVV ILLATNIELKSSVTVSVVCA LWLIQNIPDVMAIIQSLYRI Y--KYPKERKGAALTLFMME CSRSMGMAILFFHVFPSLDP
Cael_CS1         LFALCNIVLTLGSVFSKLIV LIMATNIRRTSTTTAGIYLS LLLIQCFPDTINLIRSGIDM WKGQCGQ----LVKSVVVLE SLRAIGLAVLSFHVFPQLDL
Xetr_CS          VWALAGLLVLASGFFSKVSL LFVVAMTN----NKPFGTLI IGIMLVIPSILTFIKSMWKL IFTHSPSPHKCDVFTVCIIE ALVALGSAMLVVVAMPHFNI
Dare_CS          ------MIVFILALLSKTSF LLLITFGN----QKPTALLA IGFVLVGPSFLLLLKGTWKF IFKNSTMPSKKTCLWVLCVE ILVALGAAVLTIVAMPHFDI
Ciin_CS          SYLIFGGLLFGGALISKACL MILVNSAK----IRAWSFIV ILLMLLLPEVLTMMYSLYRI IMKAEPDMDMNTVIWGGFWE FLHAVGEGILFFKVLPYLDA
Brfl_CS          SYVLFFHAVLGSAIGSKMAL IILASNTMHHGNERAKYMLM LMGTILVPEVMTVLRNIIRI TFSKMKGPSLGVLWKVCLIE ALSTSGICLFVLRVLPRSNF
Neve_CS1         -------------------- -------------------- -------------------- -------------------- --------------------
Neve_CS2         -------------------- -------------------- -------------------- -------------------- --------------------
Hyma_CS          -------------------- -------------------- -------------------- -------------------- --------------------
Amqu_CS1         -------------------- -------------------- -------------------- -------------------- --------------------
Amqu_CS2         VRILLVASVLTCLVLSKLTI IKILADLH----KAANLYWM LFFIVMIPNLISWIRTVLSG LLSKSPRPKLTALLGCIAIA FLEAVAECFLFFYILITFPP
Cogr_CSc         --------TVMQATISSKPT R--APSVMRKTTEQGASGQF LASLDNVTKSVADPSTNSYF VFCLK--PNDRRIANQFDSK CVRTQVQTFGIAEISQRLSA
Leed_CS1         --------IVVQAQVSSRPL RTLTPFVSNFTGPKPVSGHC LAFVRTT------------- -------PAHQILLTNADVE YI-----------------A

Myga_CS          FRGLVITFAVCQIPSLLKVI PNLSISEIVAVVMNIGAFLV QVSAIPFFTIGEEGGGEWEL PVSLVLISIGWWENYVSGEW TVIPFKQWRSILQDVRETSY
Pifu_CS          FRGLVISFAVCQIPSLLKVI PNPSISEVVVIIMNIAAFLV QVSSIPFFTVGEPNNCEWEL PLAIILLSIGWWENYVSGEW TVIPFKQWRSILQDVRETSY
Atri_CS          FRGLITTFAICQVPSLLKVI PNPSVSEIVAIIMNISAFLV QVSAIPFFSIGEDSTCEWEL PVSLILLSIGWWENYVSGEW TVIPFKQWRSILQDVRETSY
Logi_CS1         FRGITVTFTIFQIPGFLSLI IACSLWDYLKMFCSLLAFVC QAAALFFFTISDFSSLKWEL PVAAILISVGWIENFISGEW SLPSFRSWRCEMQDVKETSF
Logi_CS2         IETTLCTLSVGLLPAILNIH ---RRRFIYVIFVSFLVFLV QLGSPFAVIFTGQLRLGALF CLVCFCVTVTWWENFAFGKL KCFDLNEVRKNLRRVRTTGY
Logi_CS3         FGALIILSSVSAIPSILQMF TMPYKKKVLGSILHSVAAFF QVGCIVMTIFLGRGSVSWEL PVSLFLISLGYWENFMDGDI AILPFISWIKTLHVVRQRLY
Cate_CS1         VRGVLFLSATSLIPAICKVH LSDGGWRKLWYIGNASAALV QLAAIVALSCGGLMRRWWVI PLALVFVSLTWWENFVEKNL KIIPLKSWKVRLHTLKQKGC
Macr_CS          IRGLMVMVTVCTVPGFMSIC QASSCRTNCGRMFRCIAPLL QLSASVIMIMTGHPAMTWEI PVALILISLGWWENFADRDR ETLKLKSWRTLLHGVREKSM
Pd_CS2           IMSLCMLFSVGLVPSMLKMI LDSPAVKVMDKFFTFIAFLA QGAGIGFPILVLDAWMYWEI PTAMVLTSLKWWENFVHRNR GPFKLARIKQQLFACRHKLT
Pd_CS3           -------------------- -------------------- -------------------- -------------------- --------------------
Myci_CS          -------------------- -------------------- ------------SSVFLWEA PVALFLISLTWWENFVDFSN TN-FVADLKKITQSCKVKMS
Logi_CS5         ILMVLLLSAVAIIPSILLTL ---RVCFALSKMFDGLSILA QGSVIAVMLLWFFDFRIIEI CMGFVCVSVGWWENFLDDRF LCNRLKNFRYELQSGRHYTV
Logi_CS4         VRSLLLMNALCSVPAILRLF ---ARRKITTVILDFLSVCM QLSAFVIVMVSMLSDMSWEI PLALVLISVNWWENFVDKDF RLIPVKDYKDTLHKVRVKSY
Pd_CS1           VRATLLMCTSCMIPAILKLI MTKGDRGPLSIIIDIIAVLM QGSMVFLITHFLELVDIVHV VSAVSLISFRYWENFIDRDI GAIPIQSFKMSLRVGRCKTY
Cate_CS2         VRALLLMCATSIVPAFFKLV FTKGNRGPLSVIIDIFALLM QCSVFFLMTRYSDLIRLLEI AASLFLISLRYWENYIDRDV GILEVQAFKQTMRNGRCKTY
Logi_CS6         LRGLVLMLGVAPGISCLKLF QEEGGRSIGVIALDVIALII QASALLLWTIKNDTDENWSI PISLFLISFGWWENYINRFT CLRHLLKLKKSIRRMRTKVY
Logi_CS7         LIGCLMVMSVGSVPGIVRIF KSPWLPNIVKRILDFLVLAV HIVVIVLWIIRASIALTSLI PISLFLCSLNNWDNFVKGD- --SFLVDLKTQIRNQRVKLY
Logi_CS8         LIGCMLTWNVASIPGLLKIF NDLKTKTYVRRFVDTVMVTL HLGCIATWTYNSNIYVTILL PISLFLISISWWDNFVTTSK ENGCMSSVKDSLEECKIRTT
Logi_CS9         LYTLYMMQGVSLVPAILKCL RAGKSTSGTSMIVACLSAII HGGFLVIIGVKSASNSQWLF PVALLLISVHWWENFVPRTG EKGSLERIANKIR--HTKLE
Cate_CS3         FRGMLTLCAIGCVPSLLRAL -TLSKTSKKERILDALSAIV QFAGMLAWMLKQ-VIDFWPI PVGVILISVVWWENFIDKYH GQKRLTNHRRKVQKSQVKIS
Cate_CS4         FRGILTMCALGFVPSLLRAC RPGLKTGKTERGLDILSAVI QIGGIIAWLVKR-AIDYWPI AIGMVLISVVWWENFIDRYQ EDIRLENHKRDIQKAHVKLN
Hero_CS1         -------------------- -------------------- -------------------- -------------------- --------------------
Hero_CS2         -------------------- -------------------- -------------------- -------------------- --------------------
Mase_CS1         IKGAMVTNCLCIIPAILGLL NSRDSKRFVRVIVDMAAIVA QVTGFFVWPLLENKPVLWLI PVSALCISLGWWENYVTRQS PIKSLARLKDGLNFSRYYTY
Mase_CS2         VKGAMLMNATCAIPALLNIF DRMDSKFSIKLILDVLAISA QATAFVVWPLMERTPVLWTI PVACVLVSLGFWENFVDTYN KSTVLQELRDNLKRTRYYTQ
Mear_CS          ITMLQLGNSLCLIPSLLLPL SRSRSRWPLLLLLDGSAILA QSSA-AIWRGSIPLERFGFV FLCTSLISIAWWQNFVHPHS FTRFFAHYAAKLRECRSKTF
Brma_CS2         ITGAMLTNFACFVPSILLLL RRPNRLAILLMAIDTLCIIV QI--IALWTIPNLQKFKFIL PFCLILISLGWWQNFVHIES ALRALARFAVTLNERRSKTY
Cael_CS2         VTAAQLTNAMCFVPAILSVI SRKPNKSLLLVIIDFAAIAA QSSGFWAMFLPNLQKHLVAI PVSLTLISLAWWQNFVHRDS VVRTLAKFAQRLSERRSKTY
Diim_CS          LHCLILSVSTVFMPLYFKIK CSFDPIWTLKNRLKLLVDSC SYSILFLIIFSSPFFKYIAL PLSLILSSFGFWDIWIDESH SYRYLYQLKYGMRKMSSMTR
Brma_CS1         LYCLILSVYVVFMPLYLKIK ---NPIWSLKKRLKLLLHSC SYSVLFLVVFASPFFKYIAL PLGLILSSFGFWDLWINEGH CMHYLYQLKYGTRKMTSMTR
Cael_CS1         ARCLVLSACFPLVAVLQRSL SAARTGRSFRNRLGRCFVAI PHVIMFLVLMSSKFTAIIAL PIGVICTSAGFWESWIDTTH SFDELYRLKYAVRKMNTTTK
Xetr_CS          ISNLMILSSVYTIPSAVQII --NPNWKMSIPIISLVLLII GHILFCMGYLLELQIYVSIA IVATILVSLTWFENLLLLLQ KT-TFHHV-EPEHDTRNVLY
Dare_CS          VTNVMILNSVSILSAIFQVI ADCRKRFILLPLLSIVFIVL GYVLFVVNYLAFLLIPILLA IFGTICVSVNWWENYSALFT IL-SLNEIPKDIAKSSNAVN
Ciin_CS          TVALLMLPLVGIIPNCIHIQ RRPKQRRIRDLVLAVVALVL QLCGLLVCFACFLNYTYIWA FSGIVLTSCRYWENFVTGEF IFIKAQTLQNKLDATRFKTS
Brfl_CS          IFSLLLMNGVCFVPAILKIH ---SGRREIMLFLDVISWLM QTSALCLFPAQPYNGPTWEI TLSLFLISFRWWETFAGADI VFVPLYKLKVGLSEVRCKTY
Neve_CS1         -------------------- -------------------- -------------------- -------------------- --------------------
Neve_CS2         -------------------- -------------------- -------------------- -------------------- --------------------
Hyma_CS          -------------------- -------------------- -------------------- -------------------- --------------------
Amqu_CS1         -------------------- -------------------- -------------------- -------------------- --------------------
Amqu_CS2         ALSILLMCGVYSTQSLVESF GSLPRYRCLIALILVAGFLC QTSGLASITVFLDLLGLVLV IPCLVVLSFTWSGRLQKRTF APKTGQA-KELNTIARWKAS
Cogr_CSc         DFSLFLPFGEFLTPLMYGEK REQGNEKSLEEVEEFKDSPS RKRWVFMVYVMT-----WFV PDILIRIRVAWRE------- --------------------
Leed_CS1         DFDLAQFCGRYVTPNAFRGQ LVV--KDAPDSVEEVPSTRS RRFWLGTVWACT-----WFI PSFLLTVRLAWRE------- --------------------

Myga_CS          VLVGPLKIGLCIFLSRFLTN DTGEFNATTSDFSEVGVSYS LMFIQLGSGIICTYLAGLAC KLHMQKAAFALPLTLAPPLT LVVVFLQCSYQFLPAGGWFC
Pifu_CS          FLIGPCKIGLCVLLSRLLTN NTSEFNATTSKFSEVSVSYS LMFIQIGSGIICTYLAGLAC KLHMQRMAFALPLTLAPPLS LAVIYFQCEYHFLPAGGWFC
Atri_CS          FLIAPFKIGLAVLMARLLTN NTGEFNATTSQYSEVGVAYS LMFIQIGSGIICSYLAGLAC KLHMQRTAFALPLTLAPPVS LLFVYMQCEYNFLPSGGWFC
Logi_CS1         FLIGPLKIGMAVLMAYLLV- DTSNDDPDSNDAIAHFTTYS LMYLHIGSGIVCTYLAGMAC KLHMQKTAFALPLLLSPPIS LMCIYLQCQYMFIPSGNWMC
Logi_CS2         IYSAPIQIVLSGLFMHFLLD NEDSFNYYWKSFSEHFKEFG LMYIQMISTNMCVYCSGIAC RLHMQQLGFALPIVIV-PIV FTALGSLCFP-----VEWYT
Logi_CS3         TFVGVWKIGCSFLFAMLLVD DTSQSKANMTEIEDHAEQFG PLYLHLVSAWICSYVSGLAC KLCMQLFGFSLPLSLVTPVS ATVIILQCRYNFIPEYFWIC
Cate_CS1         IISSLWNAGVIIAFPFIVFP DPNQNKRAKSDIL---MQFS VPLAQGLASLFAFLSGSLAC KLCMQQVAYSLPLLLSAPMT VVVLVLQCQYGFIP-YFWFC
Macr_CS          VLVSLWKIGWTIGFAYLLLW DRVSFNMTTSGLPNIWKDIA PAIAQIIASLLCYYFGRLAC KLCMQRFCFSIPLLLSTPVT FAVLFCQCRFGFLNFYLWFC
Pd_CS2           AITSAWSIFVCIAFYYILEA ASLQNTNNNNNMTYYILSVV PGMLQIAFACVCYYFCTLAS KLMMQKISFVAPLYLSIAAS ATIIILQCHGILPPFFTWLC
Pd_CS3           -------------------- -------------------- -------------------- -------------------- --------------------
Myci_CS          LIKNCMTIVVTIAFVLSLEA LTVRLNETDEQFEIANISSV PVVVQILVSFALFYFGTLAC KLKMQTVSFALPIFLSLPVT VLLIQLQCEGHIPHILNWIC
Logi_CS5         IPASVAKLLVVLFMTYAYFP ------SIIDDILEGEIILV PVIMLHLASIVAFYIGGLAC KLDMQIISFSLSLVLVTPLT IVAIVLDCEYKSISK----C
Logi_CS4         LIANLWKVGLTFGFAFLLVP NYGLNETVCSRIPRNWFTYM PLLLQVCASGLCYYFARLAC KLCMQRFSFAIPLSLATPVA VASIITVCYLRPEKTLMWAC
Pd_CS1           IFASLWKIGLTLAFAYILIP NYYNYDYQEEPVVKVLFRFL PLIIQAVSGAICYYFARVAC KLCMQGFGFSLPLTLVTPAT ASIFCYLCYLQGWTRGFWKC
Cate_CS2         IFASLWKIALTLAFAYVLVP DAYDFDSNSTNIIKILYRFL PLIIQSASGAICYYFSRLAC KLCMQGFSLMLPLTLLTPAT VSIFCYLCHLDGWTRGHWPC
Logi_CS6         TLVSLWKIILTLGIMTVVVS RPFNNPSECPRLVVLSNPFW VMLFQVFACLLCYSFAKSAC KILLQVGSFALPLVLSMPIM IGTLISSCESWKGTSIYWNC
```

```
Logi_CS7          LVISLWKILISFVFLLVLFG VTASNCSIFGELNHYFFPFL VALVNIVSSVFCYKLSKAAC KILVQVACFTIPLSLAVPVT FALLILSYSSSNANLIQWVP
Logi_CS8          FITSAWKIFLTIGLPALCFS YYAKDCSVFTNIEHNMLPLL VACLNILSSVICYKTGKSAC KIMAQVPCFTVPVMLSGPLC VAIIISTYTFPEEITLPWV-
Logi_CS9          FCGMLFKLIITFLVPIVIFA PDRSDCSLFGNVTITYVPLI LAGVNILSSVIGFKVAKTAC VIRSQIACFTIPLILVTPLS FVIILLSYKQHFIQSVAWPV
Cate_CS3          IFVSLWKMVFSTGLLLALYS PGTGFNALFDQNNIVYDWLY VFAAQSVPGIICKYLFTLAV RGSMQVPCIAVPMLLSTPIS LGVSIWACYDCNSAWLFWNC
Cate_CS4          IYVSLWKLIFSTALLLAFFG TGVGWDALFDAKNMDYDWLI LFAAQAAPGLLCKYFFTMAV RAAIQIPAITIPMVLSTPVA LAVSIWACYDCSNTWLFWNC
Hero_CS1          ---------- ---------- ---------- ---------- ----------
Hero_CS2          ---------- ---------- ---------- ---------- ----------
Mase_CS1          RFMSIWKIFVFMMCNLFFIW LFGQHNIVVEEVQAYNSAVY VMLIQIFAAYICYIFGKFAC KILIQGFSYAFPINLVIPLV VNFLIAACGIRNGDTLFFES
Mase_CS2          RVLSVWKIIVFMACILISLH MFGERQYVVNEVLIWTSALW VALIQVGAAYFCFGSGKFAC KILIQNFSFTLALTLVGPVA INLLIAFCGMRNADPLFYEI
Mear_CS           VVLSPWKCLIFTFCMFQIYD DRVE----LNQFTTANDALW LVFVQAGSVLLCQLCAKFAC KVVMQRVGLALPVVLSIPFG ILFLAYSCRQKATNPLFWQC
Brma_CS2          VLISIWKCIVYLLMVFLVLT NYVELNQYTTSFD----ALW IVLAQAIAVILTYHTSKFAC KVMLQRTCFALPVVLSVPAT VAVLLAMCRKRYIDALFWRC
Cael_CS2          AFVSLWKICIYVVCCFLIYS NYVE----RNQLTMAYDALW LVIFQFGAVFVCYHSSKFAC KVMMQRMGFALPMALSVPVT VLLLSTNCRMRQKDSLFWQC
Diim_CS           FMVSMLRILTSTAVVWIVKF RYLPFKIVKRSLILYTLAMW IIFLN----FFLRFISRLLS TMSMRVLAALHPLLIIVPLP MIIFIGTCFIKPLCALRWHC
Brma_CS1          FLISALRILTSSAIVWIVKL RYVPFQAVKTSLILYTLAFW IIFLN----FVLRFISRLLS AMSMRVLAVLHPLLIVVPLL TMIFIGMCFITPACSLRWHC
Cael_CS1          LIVSLMRIVCTVSVLVSAVY INDHKKLNSSHFVLLLLATG IIVLH----FVMRGISRFLA ALDLHPFSFVHPLSIAPLIA YGYVRYACQSPTCSILHWVC
Xetr_CS           ICSSVVRIAVTAAVVGSCVP LSRHDWTQLKSIADLQLLLI LLAIQAVTSATCHWFGVIAC KMHYVKRGFAGPLIMTTPAV CIALFIVFIDQYRVAITIFC
Dare_CS           IISSLTRILITSAVIGVYVA LIGDGWKSVKIVF-ETVVIA LVVIQTLSSALCRWFVVVAC KMHALRRSFVMPMYLASIVV LAVFLSPLVITFPVSISYLV
Ciin_CS           FIFSLMKCGVLIGFCFLTFV VTARYNKIISLLSIIARLIY VFVTQGCCSFAAYFFGRVAC LTLIQREAYAFPTVLVTPVI AAFSAALAVYRQAGGNFLSF
Brfl_CS           IYAGITNVLIIALFTFLILA ADIKLYNIFEPVKYFVQISL LFFVQLVTSATCFLFSRLAC MTCVHRVGFALPLCLATPAT IVMFIIDCQWGNVFRNIWEC
Neve_CS1          ---------- ---------- ---------- ---------- ----------
Neve_CS2          ---------- ---------- ---------- ---------- ----------
Hyma_CS           ---------- ---------- ---------- ---------- ----------
Amqu_CS1          ---------- -MANYNLFSHQLYWSNTYIA IFATHVITSLLAYVLSFLSC YMTLHVFGLALPVLFATPLS IAAYIALINY-----HLFYY
Amqu_CS2          FIATSVRLLTYPFLSYCIFK FIPKGESIFYAFNWSGHYIS VFAVHVVASFLAYALSRLAC YMTLHVFGLALPVLLATPIS LCIYFISNSPASIEGNLFPF
Cogr_CSc          ---------- ----------KLAINMM IWLFCLFAAFFIVIFPKLIC PQLKNYAGKDISALFPIQVS ALCQGKDGSVDPHVTLDYRN
Leed_CS1          ---------- ----------KVTICFL ILLLNGIVIFYIVIFGRLLC PVLEALAGLDLTYYFPVPLV LGC----GNLGPTSTMKLSF

Myga_CS           PELYSLLIPLICAVLLWLSY SITVSHIWFPQCERMAKIEK LFITPHFETIFPDFTLTMKR RRNDKEIKITGVGDDTYCEE IYSVNNRIPFVYVCATMWHE
Pifu_CS           PELMELLVPLICALLLWLSY CITVSHIWFPQSERMAKIEK LFITPHFEGIFPDFTLGLRR RRNDKEVKLTGVGEDTYMDD VYNGSNITPQVYVCATMWHE
Atri_CS           PEIISLLIPLICAVLLWLSY SITVSHIWFPQSERMAKIEK LFITPHTDGILPDFTLSLRR RRNDKEVKLTGVGEDTYMDD IYSSSGVTPQVYACATMWHE
Logi_CS1          SDINKMIIPLGCAGALWISY CLIVSHVWFPQCERMAKTEK LFVTPHYDGIFPDFNLTMRR RRIDQEVRRTR--------- KLTEDYIIPTIYACATMWHE
Logi_CS2          IPGQADWITIGLTIGVYLSY LFLTRHIWQPKAERMAKIEK LFSFGHSDPVFPDVSLALKR RQDKAENIEYK--------- KKKQTFKQPMVYICATMWHE
Logi_CS3          PTDELVIYHYILLGISWLSQ IIINAHIWFPQSDRMSKVER LFVSPTRC-VLTDQSLLMRR RRWEKGTFLYQYGDDEEFYE MSQRNESVPKIYACATMWHE
Cate_CS1          PEDSTSIWQYSLFVLWWVSQ MMIAGHIWRPRQNNMDRTEK LFVRPMYNAALIEQSLLLNR RRRDDEIRCSENDYGYVDDD VPASKEYVPMIYACATMWHE
Macr_CS           PEGQLLWWHYGCFGVWWISQ MLITAHIWFPHAERLAKSEK LFVLPFYDGVLIEQSLLLNR RRNEKNLISNSGSISEETLY KSSDDRVIPRIYACATMWHE
Pd_CS2            DYGVLSAWAIVAIAAWVVSL LYLAVHVWRFNNEKMAMRER LFVTPFYAGPLIVPSMLMNR RSLDQELRLDLANIDTRHKK TDEDEKPVPKIYACATMWHE
Pd_CS3            ---------- ---------- ---------- ---------- ----------
Myci_CS           FV--QSWWHYVAAFVWFMSF FWICRHIFKSKLNKLAKTES VYTMQYYSGVYAPSSLALNR KRNDKEVLTTGTGVVRGNPR NAAEQRVVPMIYLCATMWHE
Logi_CS5          TELSGPSTDLVVGVAWSLSL MGIPRHFWFPRQERLPKHER MFVNPLYCVVLTTESFLLNR RRHMRRVLKTVNDNESYPKD DQPQNAEVPMVYACATMWHE
Logi_CS4          SEKKWLIWHLGLGGLWWASQ LWITRYIWTPKAPRLAFTER LFILPLYCGALIEQSLLLNR RRNDKERILKESMDSNTDAM RKQREQIVPKIYVCATMWHE
Pd_CS1            SEASSFHWQMGCAALWWLSQ LWINNHIWFPLSERLAKVER LFVLPQYCGILLEQSLMLNR RRNDWEEIPKLSGSIIEGEK ARIKEGVNNRIYVCATMWHE
Cate_CS2          PEENSFHWQIGCAGLWWLSQ LWISNHMWFSKSERLAKVER LFVLPQYCGILLEQGIMLNR RRNDFEDVPQPSGSIIEGER ARLQEGIHNRIYVCATMWHE
Logi_CS6          HATLLADYYIVIALAWWLSF MWVAVHIWFPRVERLVQTER LFVQTMYCGICLEQSLILNR RRDDKDRDKQRDEETVQQTR GSLRTDVTPMLYVCATMWHE
Logi_CS7          IIGLLQQYWFVIGFAGYLSF LYIVGHNWHPRTRRLLATDK LFTRSLYCGILLEQSMMLTR TRDEQVEDPWEKKVDFLMDT SKIRTDSTPFIYICATMWHE
Logi_CS8          HIHFTLKWWIPIGLLGYITM LYIVGHTFTPRSERLARTET LFVKPLFCGALVEQSILLNR RRDEEDNREYKDQIVDSEDI PDKKKDTIPLIYICATMWHE
Logi_CS9          LRVYSTNLYLPIMFVEFINL LITVDYLWKPGTEKTASTAM LFVKPLYCGILFDQSLMLNR SSDDDIDRKKQTMGNSSRTT SSHKEKVIPKIYACATLWRE
Cate_CS3          FSVDVLLYRWGVVVAFWVAE MWTTRYLWFPKVERLAKTGR IFTMPRYCSLMLEQSLALNR RRNDKEFEVVDQANMKVKIN VKESRKVVPKIYCCATMWHE
Cate_CS4          FSTDVLIYRWGVILLFWLAE LWTTRYLWYPKVERLAKTER IFTHPFYCSLTLGQSLAFNR RRNDEEFELLGDMNLDTKVN VKPPSKMNPMIYCCATMWHE
Hero_CS1          ---------- ---------- ---------- ---------- ----------
Hero_CS2          ---------- ---------- ---------- ---------- -NMMKDKIPFVYLCATMWHE
Mase_CS1          PPSDFISRQMAWILLWLLSQ TWITIHIWTPKAERLASTEK LYVLPMYNGLLIDQSMALNR KRDDQKDVKTEHTDNTSSPK AVKSSDSITRIYACATMWHE
Mase_CS2          PPREYVGHEMAWVLLWLISQ AWIVFHTWQPRCERLSATDK LFAKPWYIGPLIDQSLLLNR TKDLDNDCQVEGSD--AIVK DIKPFDSITRIQVCATMWHE
Mear_CS           PTRPFHWQRFLLWLCWWLSQ CWITIHLWLPRQERLAKSEK LFVLGYIDAPFPEHSIALDR RRDDKIQIRSEDIDTEEEAE LPSSADSVCKIYVCATMWHE
Brma_CS2          NTSEFLLSPLTQIFAILMVQ LWITLHLWRPRHERLAKTAK LFILPYNSALFTDQCLALNR RCDDKTKIRTEKVNETMIKE MASSADAVTKIYVCATMWHE
Cael_CS2          NGASMSLADFWIWLCWLASQ FWITIHLWNPKHERLAKSEK LFILPYYIGAFVDQSLAFNR RRDDKAKIKAEDLEFDAEDD SASSADAITKIYACATMWHE
Diim_CS           LAWKVSQKSLCFGVIWLIAY LWAYQHAKAP---KFDPDDD VFITSILNGLMVEQSLIVYR VSLNKKKINDDEVNAGDIRI TNDEIDKTVTLYICATMWHE
Brma_CS1          LA---N--SLCF-VLWLIIY LWAYQHVKAP---KFSADDD VFITTLLNGLTIEQSLVVYR VSHNKKEIDDNAEVDDDLRI TNDEIDKTVTLYICATMWHE
Cael_CS1          DQWFQSARGIYICLIWLLVG C---YRGWRLVRQRFDTNEE IIMPPVCNGLCIEQSLVVFQ HSLNRQEKTMLSDENDELRI RNDEVDRVSTVYGCATMWHE
Xetr_CS           QNSMCNIVMLFISIAWYLGF VLCTLYVWKLKVQRIERTTQ LFVRRLYEAAYIDQSMLLNT RYKLESKEKFE-------SK ENQSSADKVMVYLCATMWHE
Dare_CS           GNTLLGLVCLGCSLCWWLGL VLSTVYIWFLKIHRIERTKD LFVQRMYEGAFLEQSLMLNT RFEIRKKIKEK---------- --KHEKETIRVFLCATMWHE
Ciin_CS           VGKPFYWTGIVCVLLFWLSL ILFTVFIWNSKHDRLQRTEV IFTKALYGGPLIDQVLMLNR RQFESNYNKNQSRRRSRRLT MTANHKFTPQLFLCATMWHE
Brfl_CS           DV---SWREITIGVLWWLSQ LWISSHIWFRSSHRFAPTSK LFVRPMYCGGLVEQCLMLNR KSKDENEARTG--SNVKIVT DVVAAPVNSLIYLCATMWHE
Neve_CS1          ---------- ---------- ---------- ---------- ----------
Neve_CS2          ---------- ---------- ---------- ---------- ----------MYHE
Hyma_CS           ---------- ---------- ---------- ---------- ----------
Amqu_CS1          QD---FGYMILTSLLLWFSQ LISNIKLFTLKNAILASEEN IFVRPYYNSILLEQSLAMNR DVSCKSNVQFN--------- ---GNRQMCTMYVCSTMYRE
Amqu_CS2          QTAPLVHYAFGFSFLLWVGQ TVAGVNLFRFKNAILASDED MFVRPYYNSALLEQFLAINR VVSHKPQVMTD--------- ---GRRESCRVFVCSTMFRE
Cogr_CSc          TNRYFDWYFEQMT---MLKN LLKGAVGYSPQYVKTLANKQ QNILG---GRVYDMTQYLAG GNKMQAKAGESVDTRNARCQ FAEYLVLVVSIILCSILYTE
Leed_CS1          KNGEFDWYTATFQ---PVIN QHIGTLVHSWDEIQSYASDE DIIWGVYNNNLYDLTDYFYT QDQ--------TDFRTARCQ APNITLIVVAAILMSTMYTE

Myga_CS           TRQEMTQLLKSLFRLDYVHC ASKLAQDKFRIKDPDFFDLE MHIIFDDAFELDESVKYIPN GFVRLLYDCMEDAARSVVKG PVILSSPEKVPTPYGGKLIW
Pifu_CS           TRQEMTQLLKSLFRLDYVHC ASKLAQEKFRINDPDFFNLE LHVIFDDAFELDEKVKYIPN SFVKQLVECMEDAARSVVKG PISLLPPEKVATPYGGKLIW
Atri_CS           TRQEMTQLLKSLFRLDYVHC ASRLAQEKFRIHDPDYYDLE LHIIFDDAFELDDKVKYVPN SFVRQLIECMEDAARSVVKG PISIQPPEKIPTPYGGRLVW
Logi_CS1          TRQEMTQLLKSLYRLDYVQC ASRLAQQQFRIRDPDYYNFE IHIIFDDAFEIDDDFKYVPN GFVQQLIECMEDAARSVVKG PIIIAPPVKIPTPYGGRLVF
Logi_CS2          IQQEMLQLLKSLFRLDFDIC KRYLAQKKFETNIEDLYTAE IHIIFDDAFETDEKTQRMMN GYVRQLIQVMPMACSSVSKG SVVLKPVKKITTPYGGRLEW
Logi_CS3          TRNEMTQLLKSIFRMDIDHS ARFLAQKYYQIHDPDYYEFE AHIFFDDAMALDANNTLLPN TFVRDLVDCLDEALSSVHER QISLGPPKRTPTPYGGRFTW
Cate_CS1          TKNEMTQLLKSIFRMDKDQS KRRNLRADPTFKDPDYYEFE THIMFDDAMDLNDDDEFIAN EFVHQLVSVIDHAAR----- -----RPVKTPTPYGGRLTW
Macr_CS           TRNEMTQLLKSIFRMDFDQC ARRNAQEVFKVRDPDYYEFE THILFDDAMELNEQDEMVPN EFVNQFIEVIDNAASSVHES VVTLGSPEKLSTPYGGRLVW
Pd_CS2            TEQEMKQLLKSIFRMDMDQC ARRIALKKFGVGKSDYYEFE TQIFFDDAWQENANKIFSPN TRVKMFLSVINQAASAVYET EVELQPPVKYPTAYGGRLVW
Pd_CS3            ---------- ---------- ---------- ----------VYET EVELHPPEKVPTPYGGQLVW
Myci_CS           TELEMKQLLKSIFRLDSDQN ARRLAEEKLRIIDPDYYEFE THIFFDDAMELDEEDRFVPN SYVRKLIDVVDEAASSVMKV AWKLPPPQIFPTPYGGRLIW
Logi_CS5          TRTEMVQLMKSIHRVDAEQC MRRTVE-QLSGPDPDFFNYE AHIFFDDIMEQNDDEEWQVN DFVALLVECMEEAISSVLKT YLKPHPPIKISTPYGGQLVY
Logi_CS4          TENEMIQILKSLFRLDIDQS ARRNAQDYFSVQDPDFYEFE ANIFVDDAMEVNEKGEWMPN TFVTQLVDLVDEAASAIHES EIKLMAPVKIPTPYGGRLVW
Pd_CS1            TANEMVQVLKSVMRMDVDQA ARSNAQEYFGVKDPDYYEFE AHVFIDDAMEYNKQTDRICN RFVRQLVDVIELAATSIHET QLVLAPPIKTPTPYGGRLTW
Cate_CS2          TANEMVQVLKSIMRMDEDQA ARCNAQEYFGVKDPDYYEFE AHIFIDDGMDFNKDKGPACN SFVKQLVDVMELAATSIHET QLVLAPPIKTPTSYGGRLTW
Logi_CS6          TDNEMMQILTSIFRLDLDQC TRKHAYLFFDVMDPDYYEFE AHIFFDDAFESHEDEEYQTN AFVKQLVSVVERAASAVHRV QVKMESPAKYPTPYGGRLEW
Logi_CS7          TENEMIQVLRSLFRLDEDQS ARRKAVKAFEI-DVEYYEFE AHIFFDDAFEFHEEEDYDVN DFVKMLTRVIGVAANTQHRT SMNIPPPTKIPTPYGGRLLW
Logi_CS8          TEQEMINMLKSIFRLDSDQS ARKLAQQFFDLTDPDYYEFQ GHIFFDDAFQSHSDDDYMVN DFVKQLVRTMQKAANDVHGA SMVIKPPKRILTPYGGRLEW
Logi_CS9          NEKEMTQLLKSIFRMDLDQF DKSFAKRISDGPKPHTYEFE AHIFFDDAFKEDSK-KYTVN RYVLQLLKVIEMASRAVYID SVELSKPNKYLTPYGGRFEW
Cate_CS3          NEAEMMQIMKSCFWLDKDQH ARRMVKDEFGLKDDDYYEFD IHIWFDDAFELHMVGSYSVN KFVKQFCQVIQDAAFAVHRV RINLDKPEKIVTPYGGQFKY
Cate_CS4          TEDEMVQILKSCFRLDKDQH ARRMAKDVFRVKVKDYYEFE IHILFDDAFEPHLEGNFEVN QFVKQFCRVIQDAGLAVHNV RLSMYVPEKVETPYGGQLKY
Hero_CS1          ---------- ---------- ----------VTDPDYYEFE ANIFFDNAMNNHDDPDYTAN DYVILLVEVVREAAKGVHDV STRLDDPTKYPTPYGGRLEW
Hero_CS2          GRSEMFCMLKSIFKLDIDQC ARKNAQTFFKFKDPDYYDFE GHIFFDDAFIFFKNNVYEIN SYVRDFIDTIQDAATAAHGI TPNLQEPIIYQTPYGGRLEY
Mase_CS1          TKDEMMEFLKSILRLDEDQC ARRVAQKYLRVVDPDYYEFE THIYLDDAFEISDHSDSQVN RFVKLLVDTIDEAASNVHQT NIRIRPPKKYPAPYGGRLTW
Mase_CS2          TNEEMIEFLKSIFRLDEDQS ARRVAQKYLGIVDPDYYELE CHIFMDDAFEISDHSDSQVN RFVKCLVDAVDEAASEVHLT NVRLRPPKKYPTPYGGKLIW
```

```
Mear_CS             SALEMGCMLKSIFRLDKDQC ARRNAQRYLKVVDPDYYVFE AHIYIDDAFELDENGNPHPN KFVHQLLEKMDEAAS----T KLQLRTPRICVTSYGGRISY
Brma_CS2            TALEMTCMLKSIFRLDEDQC ARRNAQKYLKIIDPDYYEFE AHIFFDDAFEINEYGEPVIN KFVQQLIEKIDEAASALHQT QMKLKSPKKISTPYGGRLSF
Cael_CS2            TGVEMTCMLKSLFRMDEDQC ARRNAQKYLKVIDPDYYEFE AHIFFDDAYDVNEYGEPEIN KFVKQIVNVIDQAASAVHQT QMRLKPPKKAKTPYGGKLTY
Diim_CS             TRIEMMQMLKSIMKLDKEHS LIMAKR---RRSDDIRYRLE AHIFFDDAWEDQAECGRTPN AFFQTFFHLLLELTQNENTD EKSDKDRVLVNTAYGGRLVV
Brma_CS1            TRIEMMQMLKSIFRLDKEHS LMLIER---RRSDDIKYRLE AHIFFDDVWEDQMECGRAPN GFFQTFFHLLLELTQNESVN ESAENDRVLVNTAYGGRLVV
Cael_CS1            TETEMRQVLRSILKLDVDHA TRMNNKK----ANELRYRLE GHIFFDDAWEDVEEDKRQPN EYFNMFFDLLNEMTGERLNE EGKMETRILVNTPYGGRLVV
Xetr_CS             TFDEMLKILTSLFRLDKYKS NKK----------DTLNLE AHIFFDDAFTETQTGKRHVN VYVEYLISAFEEVHRDVFNN ISGCTQQKIMMTPYGGRLCY
Dare_CS             TYDEMMKIIISMFRMDKYRP KTEQK---------SDVEFE FHIYVDDAFKDVKE-ERHTN EYVEILVDVIKEVYLSVFKN QQPLPSQKIISTPFGGRLEY
Ciin_CS             TEQEMTQILTSLMRLDKDQG SRRENS----QKDPDYYNFQ IHVMFDDAMETPKGSDIKTN KFVQQLCTLIPLAANKVDGT GYDLSPPVIYPTPYGGRLEF
Brfl_CS             TEEEMVQILKSICRMDLDQA SQTSVP----DTQRDNYRCE THILFDDAMDLRDDGKLVVN DFVRQLCKLVPTAANAVRDD QVHVLDPTLTPTPYGGRLVW
Neve_CS1            -------------------- -------------------- -------------------- -------------------- ---MKACVKVKTPYGMQLRW
Neve_CS2            ADYEMEQLLNSLHDVDRYRR K------------ARRHIE SHVFFDGCIKSDE-----LN SYILQLVSLVESTLR----- -IPLNKCTKMKTPYGMQLRW
Hyma_CS             -------------------- -------------------- ----------------DVVK VYALQLLSLLKETMG----- -IDPGKACKLETPYGLQLKW
Amqu_CS1            NETEMKQMLISIKRLSDLYK KEEEKQSFI-----KADQYE SHIFFDGGCRENQ-----LY PFAIQFLSLIPDTLD----- -VKLKDAEKWKTPYGYRFQW
Amqu_CS2            NAKEMRQMLRSIRRISKFFR KDQNI---------FKYQYE SHIFFDGGCREDD-----LS QFAIQLMSLLEDTLG----- -IKLRNATKQKTPYGYRFMW
Cogr_CSc            DEESLRRAIDSAARMHYDDK RKLLV---------------- --VICDGMIIG---QDKPTP RIVLDILGVSDTVDPEPLES LGEGQKQHNMGKVYSG--LY
Leed_CS1            GEKSLRRTIDSLAALNYDDK RKLIF---------------- --LICDGNITG---SDRTTP RIVQDILGVDPKLDPEPLKS VGEGSRALNYGKVYSG--LY

Myga_CS             TMPGHTKLHVHMKDKNKMRH RKRWSQVMYMYYLLGYKLFA YEADKFMMEDQAENTFLLTL DGDVDFKPDAVKLLIDRMKK NRKVGAVCGRIHPIGSGPMV
Pifu_CS             TMPGHTKLVVHMKDKNKMRH RKRWSQCMYMYYLLGYKLFA KEADNYMMEEQADNTFILTL DGDVDFKPDSVKLLIDRMKK NKKVGAVCGRIHPIGSGPMV
Atri_CS             TMPGHTKLNVHMKDKNKMRH RKRWSQCMYLYYLLGYKLFA READRYMAEEQAENTFMLTL DGDVDFRPDSVKLLIDRMKK NKKVGAVCGRIHPIGSGPMV
Logi_CS1            TMPGQTRMIVHIKNKNMIRH RKRWSQCMYMYYLLGYRLLG KDSDMRTVDQQAENTFILAL DGDVDFKPESVKLLIDRMKK NKKVGAVCGRIHPIGSGPMV
Logi_CS2            IMPGGTKMEVHLKDKNKIRH KKRWSQIMYMYYLLGHKLLN YQTSDDLLREEAANTFIMAL DGDVDFGPESVRLLIDRMKK NRDVAAVCGRIHPIGQGPMV
Logi_CS3            DLPGATSLVVHLKDKHKIRH KKRWSQVMYMYYLLGFKLLE SETGHEKASTEAENTFILTL DGDVDFKPDAVRLLVDRMKK NKKVGAACGRIHPIGSGPMV
Cate_CS1            TLPGGTMLIAHLKDKVKIRH KKRWSQCMYMYYLLGYRMAI RDDNLWKKSMKASNTFILAL DGDVDFKPDAVRLLVDQMKA NRKVASSCGRIHPIGSGPLI
Macr_CS             MLPGGTKLNVHLKDKMKIRH RKRWSQVMYMYYLLGFNLMR RDVFATDTPIGAENTFILTL DGDVDFKPDAVRLLVDRLKK NKKVGAACGRIHPTGQGPMV
Pd_CS2              QLPGGNNLVVHMKDKNKIRH KKRWSQVMYMYYLLGWKLLN DTQKLVQESRKAENTFVLAL DGDVDFRPEAVLLLVDRLKK NPKTGAACGRIHPIGSGPMV
Pd_CS3              NLPGGNKMVIHLKDKNKIRH KKRWSQVMYMYYLLGWKLLN DTSKIVRDSYKAENTFLLVL DGDVDFKPEAVVLLIDRLKR NPKTGGACGRIHPIGSGPVV
Myci_CS             TLHGKNQIIVHIKDKNKIRH KKRWSQVMYIYYLLCYKLYT QALKVKEALLQCQNTFLLAL DGDVDFKPEAVILLVDRMKK NPKTGAACGRIHPIGSGPMV
Logi_CS5            KMPGGNLLYIHLKDKNKIRH RKRWSQVMYMYYLLGFKLLC RDEALELFKIQSENTFILAL DGDTDFSPVSVQILLDRMKK NDKAAAACGRIHPIGSGPIV
Logi_CS4            TLPGGNLLIAHIKDRLKIRH KKRWSQVMYMYYLLGYRLLQ QDDAMQSNSVQAENTFVLAL DGDVDFKPHAVQLLVDRMRK SKKVGAACGRIHPIGSGPMI
Pd_CS1              TLPGGNAFYVHIKDKNKIRH KKRWSQVMYMYYLLGYRLLS NDESIYGNQAQAENTYLLTL DGDIDFKAESVTLLVDRMKK NKKVGAACGRIHPIGTGPMV
Cate_CS2            TLPGGNLLYLHIKDKNKIRT KKRWSQVMYMYYLLGYRLLN NDDSIYSNQSDAENTYLLTL DGDIDFKAESVTLLVDRMKK NKKVGAACGRIHPIGTGPMV
Logi_CS6            HLPGQNKLICHMKDKAKIRL RKRWSQVMYMYYLLGHRLVQ PLPDARRKQTIADNTYILTL DGDVDFKPSAVQLLVDRMKK NDKVGAACGRIHPIGTGPMV
Logi_CS7            TMPGGNTLTAHLKDKTKIRH RKRWSQVMYLYYFLSNQLMS LPISLDRKKTRAENTFLLAL DGDVDFQPEALRMLVDRMKK NPKVGAACGRIHPIGSGPMV
Logi_CS8            DLPGGNKIIAHIKDKAKIRH RKRWSQVMYMYYFLGFQLLQ KL-SMKRKKVRAENTFLLAL DGDVDFQPMAVQILVDRMKT NPNLGAACGRIHPIGSGPMV
Logi_CS9            TMPGDNLLVVHFKDKKKIRV KKRWSQVMYMYYLLSYRIIN NTIPKHKKFSTASNTFILAL DGDIDFYPDAVHLLLDRMRK DGQLGAAAGRIHPKGSGPMV
Cate_CS3            TFPGGNKMYVHLKDKMLIRH KKRWSQVMYMYYLLGENLWR RDLSSKQKQITSENTFILAL DGDVDFQPDAVRMLLDLMKR DLNVAATCGRIHPIGSGPMV
Cate_CS4            TFPGGNKMIVHLKDKLKIRH RKRWSQVMYMYYLLGEKLWR KDINSKSKQVEAENTFILAL DGDVDFRPEAVRLLLDLMNR DLSVAAACGRIHPIGSGPMV
Hero_CS1            TLPGENKLVVHLKDSAKIRH KKRWSQVMYMYYFLGYRLCD ETLTEDRKMTRSNHTFLLSL DGDVDFEPMALIFLIDLMKR EEKVGAVCGRIHPIGSGPLV
Hero_CS2            ILPGENKLIVHLKDSAKIRH RKRWSQVMYMYYFLGNNLxx xxxxxxRKEIRKENTFLLAL DGDIDFEPTALLLLIDLMKR DEMVGAACGRIHPTGTGPMV
Mase_CS1            VLPGKTKMICHLKDKAKIRH RKRWSQVMYMYYLLGHRLME LPIPVDRKEVMAENTFLLTL DGDIDFQPHAVRLLIDLMKK NKNLGAACGRIHPVGSGPMV
Mase_CS2            TMPGKNKLICHLKDKSKIRH RKRWSQVMYMYYFLGHRLMD LPISVDRKEVIAENTYLLAL DGDIDFKPSAVTLLVDLMKK DKNLGAACGRIHPVGSGFMA
Mear_CS             VLPWRNRLSIHLKNKLLIRQ RKRWSQVMYLYFLLGFRLML RVHEQKRRELLAENTFILTL DGDVDFQPECVHLLVDLMRK NRRLGAACGRIHPRGSGLMV
Brma_CS2            IMPGKNRLTIHLKDKQKIRV RKRWSQVMYLYYLLGYQLMK VDDEM-RKEIISENTFILTL DGDVDFSPQCVHLLIDLMKK DRRLGAACGRIHPRGTGLMI
Cael_CS2            IMPGKNKLFIHLKDNQKIRH KKRWSQVMYLYYLLGYRLMM KVDDPSRKEIISENTFILTL DGDVDFTPSSVYLLVDLMKK NRRLGAACGRIHPRGDGAMV
Diim_CS             RLPAGTLLFVHLKDKQLVRH KKRWSQVMYLYYLLGHRIMD SHMSVDDRQLEADNTYILAI DGDSKFEPSSVMKLLRLMNM KNEIGCACGRIHPIGEGIMI
Brma_CS1            RLPAGTLLFVHLKDKQLVRH KKRWSQVMYLYYLLGHRIMD SHMSVEDRQLEADNTYILAI DGDSKFEPSAVMKLLRLMNA KNEIGCACGRIHPIGEGIMI
Cael_CS1            KLPSGTLLFVHLKDKKMIRH KKRWSQVMYMYYLLGHRIMD CPLSIEDRQQMADNTFILAI DGDSKFEPDALLRLLHLMNA KSDIGCACGRIHPIGNGIMV
Xetr_CS             ILPHGNLLYVHLKDKQRIRH KKRWSQIMYMFYLLGWKLYR KYDNLAEKAKEKHNTYVLAL DGDTDFQPSSLLLLVDRLKR YPGVGAACGRIHPTGMGPMV
Dare_CS             TLPKGNVMIVNLKDKKLIRH KKRWSQIMYLYYILGWRLHK KYYEKFEAGRERENTYILAL DGDTDFQPSAVMLLIDRLKL YPDVGAACGRIHPTGTGPMV
Ciin_CS             ILPHGNLMVVHLKDKAKIRH RKRWSQCMYMYYILGYLQ-- ----ATKKAYKLDDVFLLAL DGDVDFQPHAVHLLLDRMKR NPDVGAACGRIHPTGSGPVV
Brfl_CS             RLPGGHKMVAHLKDKMRIRH RKRWSQVMYMYHLLGWSLKH QWTSRIKAELRCDNTYILAL DGDVDFQPSAVQLLVDLMRR NPQVGAACGRIHPIGSGPLI
Neve_CS1            KLPGGMPFTVHLKDNTKVKN KKRWSQVMYMSYVLDHKQ-- ------KILVRDDQTFILTT DADVNFKHDSMEALIDYMLQ DETIGAVCGRTHPMGKGPLV
Neve_CS2            KLPGGMPFTIHLKDNNKVKN KKRWSQVMYMSYVFNEMY-- ------FCLVRDDQTFILTT DADVNFKHESMEALIDYMLQ DEKVGAVCGRTHPMGKGPLV
Hyma_CS             RLPNGMPFIIHLKDTVKVKS KKRWSQVMYMSYIPSNPF-- -----LLANKSDDLTYILAT DADVMFTPDDVMALMDFMSR NAKVGAVCGRTHPLGSGPLV
Amqu_CS1            KIGTSMPFVVHFKDNLKVRN KKRWSQVMYMKYILEHKE-- -------KGKELDNVFILTT DGDVDFKAESAVALLDMLAR DPQVGAVCSRTHPLGYGPIY
Amqu_CS2            LIDNCMPFMIHLKDNHKVRN KKRWSQVMYMKFILEHLE-- -----KQDKFDLSNTFILTT DADIDFKAESAVVLLDMLAR DPQVGAVCSRTHPLGSGPIY
Cogr_CSc            EVQGHVPFLVVVKEVSRPGN RKRDSQMIIMRFLPMSPLLY HQIRNIIGVNPTFYEFMFQI DADTVVAPDSATRMVSAFLY DTRLIAVCGETALTNAKASF
Leed_CS1            EFEGHVPYMVVVKERSKPGN RKRDSQILLMHYLPMSPLLY HQMRNVIGIDPAFYEYIFTI DADTTVTPESLNRLVASAAI DSNIIGICGETKVANEDTSW

Myga_CS             WYQEFEYAVGHWLQKAAEHV FGCVLCCPGCFSLFRGSAVM DDNVLKMYTTPPTEARHYIQ FEQGEDRWLCTLMLQQGHRI DYCAGSDALTFAPETFNEFF
Pifu_CS             WYQQFEYAVGHWLQKAAEHV FGCVLCCPGCFSLFRGSAVM DDNVMKMYTTKPTEARHYIQ FEQGEDRWLCTLMLQQGHRI DYCAGADALTFAPETFNEFF
Atri_CS             WYQQFEYAVGHWLQKAAEHV FGCVLCCPGCFSLFRGSAIM DDNVLKMYTTKPTEARHYIQ FEQGEDRWLCTLMLQQGHRI DYCAGADALTFAPETFNEFF
Logi_CS1            WYQQFEYAVGHWLQKAAEHV FGCVLCCPGCFSLFRASALM DDNVMKMYTTKPTEARHYIQ FEQGEDRWLCTLLLQQGHRI DYCAGADALTFAPETFNEFF
Logi_CS2            WYQQFEYSVGHWLQKATENV LGCVLCCPGCFSLFRGLSLM DDNVMRLYAIDATEAKHFIQ YEQGEDRWLCTLLLQQGYKI DYCAGSDAFTYAPETFHDFY
Logi_CS3            WYQQFEYAIGHWLQKATEHV FGCVLCAPGCFSLFRGSALM DDNVARLYTTRATEAGEYVQ YDQGEDRWLSTLLLQQGHRI DYCAAADALTHAPETFSEFF
Cate_CS1            WYQEFEYAIGHWLQKAAEHV YGSVLCSPGCFSLFRGSCLM DDNVMRTYCSQADEPSEVVQ FDQGEDRWLCTLLLQQGYRI TYCAASDALTYAPESFKEFY
Macr_CS             WYQKFEYA------------ -------------------- -------------------- -------------------- --------------------
Pd_CS2              WYQIFEYAIGHWMQKASEHV FGCVLCSPGCFSLFRGSTLM DTNVMARYTTKSTEASHFVQ YDQGEDRWLCTLIIQQGYRV DYCAGADALTFAPEGFREFF
Pd_CS3              WYQIFEYAIGHWMQKATEDV FGCVLCSPGCFSLFRGSTLM DTNIMCRYTTKSTEPVHFLQ YDQGEDRWLSTLVVQQGYRI DYCAGADALTYAPVAFNEFF
Myci_CS             WYQVFEYAVGHWLQKAAEHV LGCVLCSPGCFSLFRGSAIL ANNVGAMYTHIAEQPAQFLM YDQGEDRWLCTLLLKEGYRV DYSAASDALTYCPETFNEFF
Logi_CS5            WFQKFEYAVGHWLQKATEHV FGCVLCSPGCFSLFRASALM DQNVMRKYTILPSDPGHYLQ YDQGEDRWLCTLLLQQGYRV DYAAAADAVTYAPEGFNEFF
Logi_CS4            WYQIFEYAVGHWLQKAAEHM LGCVLCSPGCFSLFRGSALM DDNVMRTYATRSSEARHYLQ YDQGEDRWLCTLLLQQGYRV EYCAASDAMTHAPERFKEFF
Pd_CS1              WYQKFEYAIGHWLQKAAEHM LGCVLCSPGCFSLFRGSAIM DDNVMRTYATKSTEARHYVQ YDQGEDRWLCTLLLQQGYKV EYCAAADASTYAPETFREFY
Cate_CS2            WYQKFEYAIGHWLQKAAEHM LGCVLCSPGCFSLFRGSAIM DDNVMRTYATKSTEARHYVQ YDQGEDRWLCTLLLQQGYKV EYCAAADASTYAPETFHEFY
Logi_CS6            WYQKFEYAVSHWLQKATEHM IGCVMCSPGCFSLFRGSAIM DDNVMRKYATLSTEAAHYVQ YDQGEDRWLCTLMLQQGYRV EYCAAADAYTYCPEGFNEFF
Logi_CS7            WYQKFEYAVSHWLQKATEHI IGCVLCSPGCFSLFRSSALM DDNVMKKYTTPPTEPRHYVQ YDQGEDRWLCTLLLQQGYRV EYCAASDSYTFAPEGFYEFF
Logi_CS8            WYQKFEYAVSHWLAKATEHM LGCVLCSPGCFSLFRGSALM DDNVMKKYTTPPTEARHYVQ YDQGEDRWLCTLLLQQGYKV EYAAASDSYTFAPEGFYEFY
Logi_CS9            WYQKFEYSVGHWLQKATEHK LGCVLCSPGCFSLFRGSALL QHNVINKYATIPRQAHHYIQ FDMGEDRWLCTLLLQAGKRI EYCAAADASTFCPEGFTEFY
Cate_CS3            WYQKFEYAVGHWMSKSTEHV LGCVLCSPGCFSLFRGSALM DDNVMKTYTRRPTEGIHYVQ YDQGEDRWLCTLLLQQGFKV EYSAAADALTYCPEGFGEFF
Cate_CS4            WYQKFEYAIGHWLSKSTEHV LGCVLCSPGCFSLFRGSALM DDNVMKTYTRRPTEAIHYVQ YDQGEDRWLCTLLLQAGYKV EYSAASDALTYCPVGFGEFF
Hero_CS1            WFQEFEYAIGHWLQKATEHV MGCILCTPGCFSLFRGSAIM DENVMRSYTMLAVDPLQHLQ YDQGEDNWLCTLLLQQGYKV EYCAASDSYTYCPEGLTEFF
Hero_CS2            WYQQFEYALGHWMQKAAEHV FGCVLCSPGCFSLFRGSALM SANVMAMYTTMSTEPLHYIQ YDQGEDRWLCTLLLQQGYRV EYCAASDSYTHAPEGLHEFF
Mase_CS1            WYQLFEYAIGHWLQKATEHM IGCVLCSPGCFSLFRGKALM DDNVMKKYTLRSDEARHYVQ YDQGEDRWLCTLLLQRGYRV EYSAASDAYTHCPEGFSEFY
Mase_CS2            WYQMFEYAIGHWLQKATEHM IGCVLCSPGCFSLFRGKALM DDNVMKKYTLTSNEARHYVQ YDQGEDRWLCTLLLQRGYRV EYSAASDAYTHCPERFDEFF
Mear_CS             WYQKFEYAVGHWLQKATEHM IGCVLCSPGCFSLFRSSALI DDNVARKYATKSEKPFHYVQ YDQGEDRWLCTLLLQRGYRV EYCAASDALTFAPEGFSEFF
Brma_CS2            WYQKFEYAIGHWLQKATEHM IGCVLCSPGCFSLFRSLALM DDNVTRRYASKAENPVDFIQ YDQGEDRWLCTLLLQRGYRV EYCAASDALTYAPEGFNEFF
Cael_CS2            WYQKFEYAIGHWLQKATEHM IGCVMCSPGCFSLFRAYALM DDNVARRYALKSEEPKHFIQ YDQGEDRWLCTLLLQRGYRV EYCAASDAQTFAPEGFNEFF
Diim_CS             WYQKFEYAISHWFQKATEHI FGCVLCAPGCFSLFRASALM DDNIMHKYTKIACEARHFVQ YDQGEDRWLSTLLLKQGYRI EYAAFADAETYAPESFHEFF
Brma_CS1            WYQKFEYAISHWFQKAAEHV FGCVLCAPGCFSLFRASALM DDNVMHKYTKTASEARHFVQ YDQGEDRWLSTLLLKQGYRI EYAAVADAETYAPEGFHEFF
Cael_CS1            WYQKFEYAIAHWFQKAAEHV FGCVLCAPGCFSLFRASALM DDNIMHKYTKTASEPRHYVQ YDQGEDRWLSTLLLKQGYRI EYAAASDAETYAPEGFEEFF
Xetr_CS             WYQKFEYAVGHWLQKSSEHV FGCVLCSPGCFSLFRASALM DDNVLKKYSTKAAEAAHYVQ YDQGEDRWLCTLLLQQGWRV EYNAASDAYTNAPQEFVEFY
Dare_CS             WYQKFEYAVGHWLQKTAEHV LGCVLCSPGCFSLFRGAALM DDNVMKRYTTRASEASHYVQ YDQGEDRWLCTLLLQQGWRV EYNAASDAYTNAPQDFKEFY
Ciin_CS             WFQKFEYAVGHWLQKTAEHV LGCVLCSPGCFSMFRGSAIV DVNVAKMYTTKATEALHYVQ WDQGEDRWLCTLLLKQGWRI EYSAASDSYTFAPEEFKEFY
```

```
Brfl_CS          WYQMFEYAIGHWFQKTAEHV LGCVLCSPGCFSLFRATAIM DDNVMHKYTTKPTEAKHFVQ YDQGEDRWLCTLMLQQGWRV EYSAASDAFTYAPEGF----
Neve_CS1         WYQIFDYAIGHWFMKVANHM FGSVLCCPGCFSLYRCQAVR D--ILPTYSTTVNEAFEFLT KDMGEDRWMCTLMVQAGWRL MYCAAAEDVTYCPMNFDEFY
Neve_CS2         WYQIFDYAIGHWFQKVANHM LGSVLCCPGCFSLYRCKAVK D--ILPEYSTNVNEASEFLT KDMGEDRWMCTLMVQAGWRL LYCAAAEDVTYCPTTFDEFY
Hyma_CS          WYQIFDYAIGHWLLKVADHV FGTVLCSPGCFSVYRAAAVR D--ILPHYATGVNCAMDFLM KDMGEDRWFCTLMIEAGWKL EYCATAENSTYCPDNFDEFF
Amqu_CS1         WYQIFDYAIGHWLQKSTEHI LGSVLCCPGCFSMFRCSALQ Q--CLETFSSEVEDAMEFLT KDMGEDRWLCTLLIEKGWRL EYCAISNNKTYCPMTPDDLY
Amqu_CS2         WYQVFDYAIGHWLQKSAEHV LGSVLCCPGCFSMFRCSALK Q--CLPTYSTEVSTALEFLM KDMGEDRWLCTLLVENSWRL EFCAISNDKTYCPLSFDEFY
Cogr_CSc         IIQVYEYWISHNLTKAFESL FGSVTCLPGCFSMYRIRAFV SREVVEAYATIRVDTLHMKN LHLGEDRYLTTLLLKNKYKT KYTNRAHAWTIAPDSWAVFL
Leed_CS1         WIQVYEYYISHHLTKAFESL FGSVTCLPGCFSLYRIRTII SNRVIDEYAEPNVDTLHKKN FSLGEDRYLTTLLMKPTFKT KFNPDAVAHTVAPESTRVLF

Myga_CS          NQRRRWSPSTLANMMDLLAS WRDTVRINDNISRPYMLYQF VLMASTIIAPSTVILMITGS YHSVFKLSIFESYLLSLLPV VIYLAICLTMKSDIQILAAA
Pifu_CS          NQRRRWSPSTLANMMDLLSS WRDTVRINDNISRPFVLYQF VLMASTILGPSTVILMITGS YHSVLNLSIWQSYLLSLLPV VSYLVICMTMKSNHQITAAA
Atri_CS          NQRRRWSPSTLANMMDLLSS WRDTVRINDNISRPYVLYQF VLMASTILGPSTIILMITGS YHSVLGLNIWQSYLLSILPV MVYLAICMTMKSDHQIFAAA
Logi_CS1         NQRRRWSPSTLANMMDLLSS WRETCRINDNISRPYILYQF ILMASTILAPSTVVLMITGS YHSVFGIDMWWSYVISVIPV AIYIVICMTQKNNLQITAAA
Logi_CS2         IQRRRWSPSTMANVMDLLGS WKTTIKMNNNISTPFMFYQF ILMASSILAPGTVSFMIAGS YTAVLNFNAWQSYLLSVLPI LAFILLCLKAKTETQVAVAA
Logi_CS3         NQRRRWGPSTLANIIDLLGE WKNTVRINDNISTLYTIYQI TLLASTVLGPATVLLMMAGA YTVVFKTTILQSYTLAIVPV IFYIVLCMYSKAETQLTVGA
Cate_CS1         NQRRRWLPSTMANLIDLMSS TKPTVHKNDNVSYLFMLYQF ILFASTTLGPGTILLMIAGS YNAILGTDLWDSHILAIGPV IYFVIICFTTHPDTQVIVAQ
Macr_CS          -------------------- -------------------- -------------------- -------------------- --------------------
Pd_CS2           NQRRRWVTSTLANQMDLMSS YKTTVYMNDNISTPFMMYQF LLTVSTVLGPSTIILAIASA FKEILVIHVFWSYFLAILPI AFFILICHVRKEKEQLVVAG
Pd_CS3           NQRRRWITSTLANVLDLLAS YRTTVYINDNISLLYMFYQL LLTVATLLGPSTIIIAIASS LRDILVIHEGWALAITLVPT IGFIAVSLVCKEKTQLICAA
Myci_CS          NQRRRWAPSTMMNVMDLLSD YKTTVANNDNISNFYILYQT LLMASTILGPSTFVLAQALA FADIFRFELWVAYIVSLAPI VFYIAICFKTKTATQLYVGQ
Logi_CS5         NQRRRWMPSTMANIMDLLGD YKNTIRNNNNISMIYIIYQG ALMLSTVLGPATVLMMISGA FKDIFELDILLAYALSLIPA AAFVVICLVCKSKTQLIVAV
Logi_CS4         NQRRRWIPSTLANIMDLLQS YRTTVKINDNISYLYMAYQG LLMLSTILGPATVLLMMAGA FNAVIRTSLWQSYLLAVGPS VLYLIVCFIAKSETQLNVAA
Pd_CS1           NQRRRWIPSTLANMMDLLSD YHRTVLVNDNISYLYMMYQA VVMGASLLAPATVILMVAGA IHVVVGGALYWLWLTSIGAA VFYMIVCFKCSPDRQITIAG
Cate_CS2         NQRRRWIPSTLANMMDLMSD YHRTVLVNDNISYLYMMYQA VVMGASLLAPATVILMVAGA VHVVIGGALYWLWLVSIGAA IFYMLVCFKAKPETQITVAG
Logi_CS6         NQRRRWSPSTMANIMDFLAN WRGMIRKNENISCLYVMYQL LLFFSSIVTPGTTFLLIVGA INMAFTLDLLDALLANLAPI GLFMVVCFVAKPEWQLRMAA
Logi_CS7         NQRRRWTPSTMANILDLLNN WKSVTRKNEDISCLYIMYQM GLMASSIVTPGTIFLLIVGA ITTAFELPLFGSLILNIIPV GIFIIMCFTTKTSTQLAYAA
Logi_CS8         NQRRRWTPSTMANILDLLTT WNTTTKNNDDISWLYMFYQA SLMVSSIITPGTIFLLVLGA LRTAYDIPLYGALVINLVPV VFFVFLCLKAPSKYQLAYAG
Logi_CS9         NQRRRWTPSTMANIIDLLSS GNSLRNNNPNISVLYIMYQV FLFVSSILTPGTIFLLILGA FNTAYTVPLYVALILNLLPV VVFIILCYSATENIQLGYAA
Cate_CS3         NQRRRWTPSTFANILDLLSS AGRTVQNNQNISTLYMIYQA SLFVSSILGPSTIFLVIVNS LNIAINIGKTWSFVLVIIPM VLYILICLLAKPDVQIMTSA
Cate_CS4         NQRRRWTPSTFANILDLLST AGATAKKNQNISTIYMIYQI SLFVSSVLGPATIFLLIVNS LTIALTIHPAGSLMIVLIPM TLFIIICLVAKPDVQIAMAS
Hero_CS1         NQRKRWMPSKMTNMLDLLRS WKHTTYINENISFMYLFYQA AIQVASLLSPGIVFMVTVGA LNIS---------------- --------------------
Hero_CS2         KQRRRWMVSTMANILDLLVS WKLTTKKNENISSLYIFYQM MLFGVSIVGPGTIFLLISGA VNVA---------------- --------------------
Mase_CS1         NQRRRWVPSTVANIMDLLVD CKHTIKINDNISTPYIAYQM MLMGGTILGPGTIFLMLVGA FVAAFRIDNWTSFEYNLYPI AIFVFVCFTMKSEYQLLVAQ
Mase_CS2         NQRRRWVPSTMANIFDLLAD SKRTVQVNDNISTLYIVYQC MLMMGTILGPGTIFLMMIGA INAITGMSNMHALLFNLVPV LTFLVVCMTCKSETQLMLAN
Mear_CS          NQRRRWIPSTMANVIDLLRD YRNVVRVNDSVSIWYIAYQL VMLFSSVLGPGTIFLMIVGA ISISFNIDTRLALLIVTTPV LCFCVCCLTCGTETQLLLAQ
Brma_CS2         NQRRRWIPSTLANIIDLLQD YKNVINVNESISIWYILYQC IMLISSILGPGTIFLMVVGA LSISFNIDTSLSLFVVTLPV ALFCFMCFISDSEKQLMAAQ
Cael_CS2         NQRRRWIPSTIFNIMDLLKD YRNVVRVNESISIWYIIYQL VMLISSILGPGTIFVMIIGA ISISFSIDTLISLVIVSIPV VVFIVVCLTAKPEHQLICAQ
Diim_CS          NQRRRWTPSSIANTIDLLTD YKRVCQNNNSISKMYILYQS MVISFSLLSPSIVFTMLVYA QASTFEVESAKMLLYNSLPI FTFIMLCFFANSNYQIIFAK
Brma_CS1         NQRRRWTPSSIANTVDLLAD YKRVCQSNNSISRMYILYQS MVIGFSLLSPSIVFTMLVYA QVSTFEAESDKMLLYNSLPV LTFIGLCFIADSNSQIIFAK
Cael_CS1         NQRRRWTPSSIANTVDLLMD YKRASENNDAISYAYIAYQF LVIFFSMLGPAIIFTMLVFA QVAAFELRGSDVMLYNGIPI GFFIVLCFTTESNIQLIYAK
Xetr_CS          NQRRRWGPSTMANTLDLLHT GVQTAKKNPSLSLLYILYQT LFMGASILSPATVCLMIAGA FSFVFGLGTNLSLFLAVLPP AIYILICFVTKPNTQITVAA
Dare_CS          NQRRRWGPSTMANTIDLLGS GQLTAERNSSISKLYILYQV LSMGASILGPATVCLMISGG FVFVFKMSENDALILAIVPP AIYLILCFKLKSDTQLTIAA
Ciin_CS          NQRRRWGPSTMANIFDILMD AKLTVKNNIYISWGYISYQI GLMVSSILGPATVVMIVQGA YQYVFHWSSAVALVVSLIPV IIFIILCYTTSADTQVAVAG
Brfl_CS          -------------------- ------------------KM AMMISSVLGPATVCMMIAGA LTYAFKWDNVTAMAVSVAPP IGFIILCFIVKPDTQINVAA
Neve_CS1         KQRRRWMPSTLANLALLISE WRVTLGNNEHISVPFIIFQL FLLVSTIIGPSTVILVIVGG MVYGVGANENTSVILLSLIT LTYGLICLFMSQDFQLKTAK
Neve_CS2         NQRRRWMPSTLANLILLISE WRLVVTNNDYISVPFIIFQL ILLLATVIGPSTVILVITGG MNYGIATNEVTTVTIMSLIT LAFGLICLFTSQSFQLLIAK
Hyma_CS          KQRRRWGPSTLANSLVVINE QKRLRENNNAINLFFIAYQG VLLISSVIGPSTVLMVVAGG LEYAYTVNMVVVLILLIVVT VLYTLVCLYCSQDTQLKCSK
Amqu_CS1         KQRRRWIGSTIANIGLMIKK AKKITKNNESINWTFIIYQL IAFFSTTPSPATVILVIASG LTASFGINSTAVIVVTGLVA VCYGIVCVITSVKTQINVAK
Amqu_CS2         KQRRRWIPSTIANLWLLISK AGKVTRRNESVTWAFILYQI IVIISTIISPATVILIISSG LSTSFNFNPTAVIVLIGLVA VIYGFICIFTSQKTQLDIAK
Cogr_CSc         SQRRRWINSTVHNLMELIPM SQLC---------------- -------------------- -------------------- ---GFCCFSMRF---IVFVD
Leed_CS1         SQRRRWINSTVHNLCELVML PELI---------------- -------------------- -------------------- ---GFCCFSMRI---FVFVD

Myga_CS          VITALYAVIMMIATVGTVIS IVTENFGSPNVVFLTGLTTI FLIAGILHPQEFFCLVYGAL YFMVVPSTFILLTIFYLCNL NNVSWGTRETPKKLTKEEEE
Pifu_CS          VITALYSVIMMIATVGTIVS IVTENFGSPNVVFLSGLVII FVIAGILHPQEFFCLVYGVL YFMTVPSTFILLTIYYLANL NNVSWGTRETPKKLTKEEEE
Atri_CS          VITAIYSVVMMIATVGTIIS IVTENFGSPNVVFLSGLVII FVIAGILHPQEFFCLIYGVL YFMTVPSTFVLLTVYYLCNL NNVSWGTRETPKKLTKEEEE
Logi_CS1         ALTAMYTVVMMIATVGTFIS IATESFNSPNVVFLSGLAII FVVAGLLHPQELFCLIYGTL YFMVVPSTFILLTVYYLCNL NNVSWGTRETPKKLTPEEEA
Logi_CS2         VLSSVYAIVMVIVSVGTVIN IVQEDFYTPNVLFLVGLSAI FVITALIHPKELMCLIHGIL YYLTVPSTFVFLTVFFLCNL NIVSWGTREAPKKTDDDD-E
Logi_CS3         IMSAIYAVTMTVVLVGTVGT AVEGSITSPNVIFMIMLAAV FFIAALMHPEEFTCVIPGVL YFICIPAGYLVLTTYFLCNL NIVSWGTREVPVRKTKEDIE
Cate_CS1         VLSALYAVVMMIVMVATVVN SFEESILSPNVLFFLCLAGF YTLAAILHPQECACIIHGLV YFLCIPSGFLLLPVYSLCNL NVVSWGTREIPKRKSKEELE
Macr_CS          -------------------- -------------------- -------------------- -------------------- --------------------
Pd_CS2           ILSAVYAGIMVVVLVGILKS FRNESILTPSLLFIAILALC FIIGGLLHPYEFFCLAPGVL YYLCIPAGYLILIIYSFSNV NVVSWGTRETPKPASQVQEP
Pd_CS3           VLSLLFAIAMVVVFVGILKV MVLESILAPSAAFVLCITFC FLLGGLLHPYEFSCLAPGLI YFLCIPAGYLVLTIYSFSNM HVVSWGTRETPKPVDPGENE
Myci_CS          VLSVAYACVMIVIIVGVING FAESEVTNPTLILFIALSSV FVIAGYIHPYEFWCLPCGLI YYLCIPAGYVVLTIYSITNL NVVSWGTREVAGRKTKQEIE
Logi_CS5         FLSTYYAFVMVVVLAGTLKN AVTESPWHPSVIFLLALVII FIITGLLHVYEFYCLPWGAL YFLCIPSGYLLLVIFSLCNL DNISWGTREVAKRKTKAEQE
Logi_CS4         ILSAIYSLLMMAVVVGTLVQ IAEDTIVSPNAVFLLMLLAI FIIAACFHPQEFICLLPGAL YFLCIPSGYVLLMLYAMCNL HVVSWGTRETAHLAHLKPLP
Pd_CS1           YMSTFYAILMLAVTVGIVVQ TSQDSWTSPNALFIMVFIGI FMLAGLLHPEEFMCLIPGVL YFLCIPSGFLLLFIYSMINM NIVSWGTREIPKLEDLNTVK
Cate_CS2         YMSTAYAILMLAVSVGIVVQ TAQDSWTSPNAMFIMVITGI FLLAGLLHPEEFMNLIPGIL YFLCIPSGFLLLFIYAMINM KNVSWGTREM----------
Logi_CS6         ILSGVYALVMMLVVVGLALS VKAEGLCSPATIFMIFLIAV FMIAAILHPMEFLNIFYGVL YVLAIPSMSMLLMIYSMCNL HVVSWGTRETATAVPLNAKN
Logi_CS7         VLSVIYSLVMMLVVVGLIKQ AADNGFCSVSTIFLVAVCGL FFITSLLHPQEFTCILHGFL YFLSIPSMSMLLMIYSLGNL HVVSWGTRESPKPVTSETGQ
Logi_CS8         ILSIFYSLLMMLVLVGLVRQ AAEYGFCSVTTIFLTTVSSI FVVSAFLHPQEFTCILHGFL YFLTIPSMSMLLVIYSLANL HNVSWGTREVKQAVPPPNKQ
Logi_CS9         VMSTIYSLVMMIVIVGLIKQ IVESEFCSVTAIFFLGVVGI FVISAMFHPREILNIFHGFM YFLTIPSSSMLLIIYALGNL NNGSWGTRVDTTVPARTEIP
Cate_CS3         VLSTLMAVVQIVVLVGIMRQ VASDSVCHPNTMLLIFMTGS FVIGGLMHPQEATTLVHGFI YYLAIPTMYLLLIVYSICNM NVVSWGTREEVKTSTEKEEE
Cate_CS4         VLSTLMAIVQVVVLVGIMRQ VAGDGLCHPNTMLLIFMVGT FIIAALLHPQECTIIFHGFI YYLAIPTMYLLLIIYSICNL NVVSWGTREVPKTELEKENE
Hero_CS1         -------------------- -------------------- -------------------- -------------------- --------------------
Hero_CS2         -------------------- -------------------- -------------------- -------------------- --------------------
Mase_CS1         ILSTAYAMIMMAVIVGTALQ LGEDGIGSPSAIFLISLSSS FFIAACLHPQEFWCIVPGII YLLSIPSMYLLLILYSIINL NVVSWGTREVQTKKTKKEIE
Mase_CS2         LITCFYAMVMMFVIVSIVLQ ISQDGWLAPSSMFTAATFGI FFVTAALHPQEIICLLYISI YYITIPSMYMLLIIYSLCNL NNVSWGTREVAQKKTAKEME
Mear_CS          VIGALFAMLMTAVIVGTSLQ IQKDGLLSPHSIFLFTVLGS WSFSALLHPLEFGCLLPCGL YFLAIPCMYMLLPVYSLCNL NTVSWGTRENASVSSSSTGQ
Brma_CS2         IIGALFAMLMTAVIVGTVIQ MQKDGIMSPHSIFLLFVIGS FFTAAILHPLEFTCITPGIL YFLAIPCMYMLLPIYSLCNL NNIAWGTRENPKVEGNDPI-
Cael_CS2         TIGAIFAMLMTAVVVGTSLQ LQKDGLLSPHSMFTVAVATS FLTAAILHPLEFTCIIPGTI YFLAIPCMYMLLPIYSVCNM HTVSWGTREDPRPTEKNTLA
Diim_CS          FISVLYGFIMLAVAMATINQ IILETAISPTPIFVLCMIGI FILAAILHPQEFHNIIYGPI FFLMIPCTYIFMALYALINL NIINWGTREAASIAAGKHAS
Brma_CS1         LISVIYGFVMLAVAIATTNQ IILETALSPTPIFVLCMIGI FVLAAILHPQEFHNIFYGLI FFLMIPCTYIFMSLYALINL NIINWGTREAASVAIGNNAS
Cael_CS1         YMSIAYAFVMLAVLVATSSQ IVLETVLAPTSLFIVTMVGI FFFAACLHPKEFTNIIHGVV FFLMIPSTYVFLTLYSLINL NVITWGTREAVAKATGQKTK
Xetr_CS          FLSVGYAFLMTATFLSIIGG IVNNTILTPTGIFLVSMGLI YIITALLHPKEFSLIIYGLL YIICVPSGYLLLTIYSLVNM HIVSWGTRESAAPKTEKKED
Dare_CS          IMSIMYAFLMTGSILSIIGD LVQDTFLTPSGLFLIGMVVL YLITAGLHPKEFSLIIYGLL YFICIPSGYLLLAIYSIVNM NNVSWGTRESGGQAKTSAVN
Ciin_CS          ILTILYALVMMAVIVGVVGA MATNSILDPNNIFMIMLVGL YAFTGLVHPQELMCLIHGIL YYLCVPSAFIFLMVYSLTNM NNVSWGTREVKTAAPSGDDA
Brfl_CS          IMSAMYACVMMAVVVGIAGQ ITVEGPLSPVAIFTFLMCGL FLTAAIIHPQEFWNIACGAL YLLCVPSGYLLLIIYSLCNM NIVSWGTREVKKKGGTKKAA
Neve_CS1         FLTFVFAVIMILVTIGVAAQ IANELEMLPSTLYLGGLAGI FMVAALMHPTEAYCLINGLW YLLCLPSGYLLLTIYSVVNM TDRSWGNEHDLADFESSREL
Neve_CS2         LLTFAFSIVMLMVAVGVAVQ ISKDLEVLPASVSTLYLGGI FVLAALLHPKEAYCLLNGIF YLLCLPSGYLLLTVYSIVNI TDRSWGKSLLPFSAPTDNPN
Hyma_CS          LLTLMFAIIMSIAIVGIIRQ IVDDVAFNPSTWYMFTISAI FILTAILHGSEAPYVFHGVW YLLCLPSGYLLLTIYSVCNL TDRSWGTREVKSQSQVGDKS
Amqu_CS1         GLTAAFAMVMAIVIVGMIMD KTKELLVTESTVYLALFVII FLTTAILHFSEGFALIHSLW YLLGLPAGYLFLLIFSTANL NDRSWGTREAKISLEKPFQL
Amqu_CS2         ILTFIFSIVMAAVVVGIIKE TVQDILIAESTIYISLFAIV FVTAAILHFFEGFALFQCLW YLLGLPSGYLLLLIYSTANL NDRRWGTREAADKSEQS---
Cogr_CSc         LLSTIVQPVVVAYIIYLIVM VVQNPTVVP-VAAFVMLGAI YGLQAIIFRRKWEMIGWMIL YIVAIPVFSFALPLYSFWHM DDFNWGNTRVIAGEA-GKKI
Leed_CS1         LLSTMILP---STLVYLVIV VSLGKYRFP-LIAIIMIAAG YGLQVLIFKREFMLIGWMVV YLLSYPIYSLFLPVYSFWCM DEFGWGNTRLVIGEGKDKKV
```

```
Myga_CS             RD----LIKNF---NGTVRH LSRNEWKFMKYLHPEDKQHK EKIKQDLITLKIAITLQFFS MFMHRWGTH
Pifu_CS             RE----VIRSL---NGPVSY IDQDEWRFMKYLHPEDQQHK QKIKEDLISLKIAISLQFLS MFMHRWGTH
Atri_CS             RD----LLKNI---NGPVSH IEHDEWNFMKYLHPEDQQHK QKIKEDLICLKIAITLQFLS MFIHRWGTH
Logi_CS1            RD----AIKTL---NGPVEY LPESEWQYLKYLHPEDKAHK QKIKEDLVLIRLALGLQFVS MFLHRWGTH
Logi_CS2            NQ----MESGL---DGIEGE LSENEWQKLKFLLPENKERQ EQIASDLKSIRALILTQFIA MLIHRWGTL
Logi_CS3            EE----LIKEI---DGPMIF LDAREWEQLKYLYPEDKEHQ EKIASDLKSLRCILFIQFVG MFVHRWGTH
Cate_CS1            SI----LLKEV---DGKVEY LKKKEWKQAIHLFPRDKQKE EEMSMALKSMRIILFAQFLA SLSHRWGTH
Macr_CS             -------------------- -------------------- -------------------- ---------
Pd_CS2              IE----LLQQL---DGPISY LPRDEWDEFKYVKPDSGAAK ARLTAQLLELRMLLILQLVG MIIHRWGTQ
Pd_CS3              VE----LLREL---EGDICH LCPDEWEDFKYVKPKNLQKQ QELKENLVSLRMLIALQLIG MIVHRWHTQ
Myci_CS             IEM--------KALEGIVDR LDKDEFRGLRYLKPKDVWKE KQDGKALIELRLILAVQFVC MLIHRWGTQ
Logi_CS5            DA----LLLEI---KGEVRA LDSDEWKDLKYLYPKDPEKE QQVKTSLQELRIVLLLQFIG MLIHRYGTH
Logi_CS4            RQ----VMEKL---GGEVQY LPVREWQRCKYLHPKDKNHE AKVSVDLKSMRFILLLQFAA MLRHRYGTH
Pd_CS1              RSIVRQ-------LRGPVRY LDNNEWQKLKYLYPLDKDHE AKIKTDLRTLRLIILIQFIS MLFHRYGTH
Cate_CS2            -------------------- -------------------- -------------------- ---------
Logi_CS6            GV----ILDKL---EGRVRY LSPEEWKELKYLFPNNVDHQ KKIKVDLKELRVSLLIQFIA MFFHRMSTH
Logi_CS7            NQ----VMDRL---NGNVEL LDPSEWKELTYLFPKDADHE KKMQGDLLELRVLLLIQFFA MMFHRISTH
Logi_CS8            -------------------- -------------------- ----------L--------- --------H
Logi_CS9            ------ILEKL---RGKSRS VQEDEWNNLPYLAPKGTQAF SAATSSFHPAKIHLGISFFV LLFKKMNS-
Cate_CS3            ERVHEE-------LRGKIGE LDDGEWKELKYLYPKNKIEE ERVTDQLKDLRIIVAIQFIC MFGHRLMTQ
Cate_CS4            KDMQVE-------LRGECKE LHYDEWKDLKYLFPADKIEQ DRAAMQLKELRIIVGIQFFC MFGHRLMTQ
Hero_CS1            -------------------- -------------------- -------------------- ---------
Hero_CS2            -------------------- -------------------- -------------------- ---------
Mase_CS1            LH----IASTL---RGEVDF LSPPEWKDLKYLYPEDKDEK ARISRDLKELRLILVIQFTA MLFHRFGTH
Mase_CS2            LQ----IANSI---KGEVDF LTTAEWKDLVYLRPENKEEQ ERIKTDLKNLRFVMLIQFVA MLSHRSYTH
Mear_CS             FSGKRELDENFPFRRFEPIV LDTEEWREMKYLRPPNSNEQ ARIQRGLNELRIILFIQFLA MLFHRFGTK
Brma_CS2            NE----MMDKL---GAEKDC LDPDEWNDVKYLTPHNQLEQ KNIRDGLMLLRSILIIQFIG MLLHRFGTH
Cael_CS2            KKTPGNKVDRLALKRAEREY LEPEEWNDVRYLSPMDGKDM DRLRAGLIAIRSILVIQFLA MLCHRFGTL
Diim_CS             DTELHNLLQRLFFKIIDFAR LTSSDWNGMQYLKPITEEET RQDVANLASLRIILVVQTCG MFAHRINTL
Brma_CS1            QEIQKK--------VCSRGR LNAAEWNGMQYLKPFTEEEM RRDVASLVSLRIILVVQTCG MFAHRITTH
Cael_CS1            KAPMEQFIDRVDIVKKGFRL ISCREWNELAYLKPTTPAEM KAVAEGLASLRIILFVQTLG MLLHRLNTM
Xetr_CS             NKKVKY-----NVL--QKED LPEEEWEDVSYLEPEDKHKQ DEIERDLKSLRMLLILQFLG LIYHRIYTV
Dare_CS             SLKKKF-----SLRSESACE LSEEEWKEMKYLHPEKPEQL AIIANDLRELRALLIIQFIA MLYHRIYTQ
Ciin_CS             NVKALESRKTESLEDGIIEY LDNDEWKYMNYLYPENKDEK ARIKENLRDLRIILLIQFTC MLMHRWSTF
Brfl_CS             SGPQPKALDKVVFASCPVQN LDEGEWKSLKYLHPEDKQDQ ERIAGELKSLRVVLVIQFVS MVFHRFSTI
Neve_CS1            LAPGERNGNSLTIDLEVDNS DDGAEWDGLACLKPAGFDGN TELKDKLEDLRTIILFQFIA MLAHRLTTA
Neve_CS2            HL----DDDYLQIKKEAKVV ADGVSWEGLISLEPAAFSTS SELKEKLTSLRVIIIVQFLT MVCHRLATA
Hyma_CS             IIDQIK-------------- ----KWCCIRVGQPNEKGNK SNPYTDSNMSSTISVIDEND VLIHDR--R
Amqu_CS1            KDIKNWLEEQR--EIEQSTG LEEDDWEELSTLRPEHLTKI TELRNNLKQLRILIMVQFFA MIFHRVVTV
Amqu_CS2            ------CKEGMLTNDHPNKP LDDDEWNEMTTLRPDHLTQA SQLRSDLKYLRILISIQFLG MIVHRVVTV
Cogr_CSc            VISDEG-----NSQFGGSQF FSPEKKQELRFGVPAKRQHI NSATEALLSGQ--------- --------L
Leed_CS1            IVNDDD-----NPFAGSSQ- -NPNK-----LLCP--DDFY SCVVEPL--EE--------- --------P
```

Alignment for evolutionary analysis of MMD homologs

```
Pifu_CS             DDLSHLEVLDENTIVQAIKS RYNKEKYYTYIGDILVAVNP CKHLTLFDDQYHVEYENLIT RSKKSPHLFWIADHAYRELC GTGRNQCILVSGESGAGKTE
Atri_CS             DDLSDLEVLDENTIVQALRT RFNKEKFYTYIGDILVAVNP CKPLNLFDLKYHGEYENLVS RTQKAPHLFWTADNAYRSLC ETGRNQCILVSGESGAGKTE
Myga_CS             EDLSELEILDETSIVQTLRC RFNKDIFYTYIGDILVAINP CKPLHLFDEKNHHDYKNLTV RSQRPPHLFWVADQAFRAMQ DTKRNQCILVSGESGAGKTE
Logi_CS1            DDLCQLEILDENNILQTLKK RYERDRFYTYIGDILVAVNP CKPTNIFDERNHASYSSVSV RSNNPPHLFWVADQCYRNML DSGMSQCILVSGESGAGKTE
Logi_CS2            DDISQLETIDNEILLDCIKQ RFEQNSFYQH---------- ------------EIYTNIRD RSSQPAHIFWAADFAYRAL- KDGNGQCLVVTGDSGAGKTE
Cate_CS1            -------------------- ------FQTYIGDVLLAVNP FKQIDVYSKKNHQDFTRSHK RSDNAPHIFWVAAEALDRLR ETGMSQSILINGESGAGKTE
Logi_CS4            DDLALLSILDEKTIVENLKQ RYIKNEFYTYIGDILLFINP FKTVDIYGKSDHKQYTNVSV RSNLKPHIFWIADHTHLRML SSHKSQNIVVSGESGSGKTE
Macr_CS             -------------------- -------------------- -------------------- -------------------- --------------------
Logi_CS3            SDLSQLSFLDENVILNELKK RYKEEQYYTYLGDVLVALNP NKSIPIFDDKHHEEYSNSST DSNRTPHIFWIAQQAYNRLL ATCQNQCILVSGESGAGKTE
Logi_CS5            GDFSRLSYLNEEILLSSLKS RYKKDKIYSKCGDILIAINP NRPLPIYDAEHHQRFRWQNA STQQEPHVFATAALAFQRLV ETKTDQVILVSGESGAGKTE
Miye_Myo            DDLSQLGNLDNATIKRTLQS RYAKDKIYTYCGDILIAVNP FKDLPIFGKKQHEEYHWKTL QRMPPPHVFNMAARAYRRIH ETRTNQVILLEGNSGAGKTE
Dapu_Myo21          DDLASLEVVDEEKVLLTLGN RYSRGQFQTYAGTALLIVTP QSDQSLVYEKYHERYMYK-A RSDEEPHVFAVVDRAWQDML HHQEHQNIILSGDRNSGKSF
Dapu_Myo3           DDLASLENFSIDSIVDQLEQ RYSNSVIYTYIGDILVAVNP FTELSIYSDREVFMYRNR-A RSDNPPHVFAMADAAYHAML HQRRSQCIVISGESGAGKTV
Xetr_Myo3A          DDLATLEALDENTVTDQLQK RYNRDQIYTYVGDILVAVNP FRPMDLYSSQYSQMYIGA-K RTDNPPHIFAVADLAYQSMV TYNSDQCIVISGESGAGKTE
Xetr_Myo16          DDLATLSELTDSSLLYEIQK RFNNNQIYTYIGNILLLVNP YKNLPIYSALVTQLYNNNSG RSSLPPHIFACSERAYHMLF QERRSQCFILSGESGSGKTE
Pd_Myo3             DDLSKLPEVTTDRLLVELNE RYNRDRIYTLMGDILIALNP FKPLPIYGTSVSQQYHDESQ QKSLAPHVYSIAEQSHRALL SQQISQCCVISGESGAGKTE
Xetr_Myo28          EDAALLAELTENKLLETLRD RYRDHRIYTYVGDILIAVNP FRDLPLYSPEISDIFSSR-P LSSLPPHIFAVADRAYNSLQ GNPRNQCVVISGDSGAGKTE
Dapu_Myo7           QDMISLGDLHEAGILRNLHI RYNESIIYTYTGSILVAVNP YQILPIYTAEQIKMYRER-K IGELPPHIFAIGDNCYSQMK RFRQDQCIVISGESGAGKTE
Xetr_Myo7A          EDMIRLGDLNEAGILRNLLI RYREHLIYTYTGSILVAVNP YQLLPIYTPDQIRLYTNR-K IGEMPPHIFAIADNCYFNMQ RNNKDQCCIISGESGAGKTE
Xetr_Myo10A         EDMAALSDLHEGSIMYNLFQ RYQQNKIYTYIGSILASVNP YKMIALYDCNAVELYSRH-H MGEIPPHIFAVASECYRCLW KRHDNQCVLISGESGAGKTE
Dapu_Myo9           PDLCQLPDLNEETLMDNLRA RFQAGHIYTYVGSILIAVNP FKFYPIYNPKYVKLYQNR-R LGELPPHIFAVADAAYNAMM RDRKPQCIVISGESGSGKTE
Pd_Myo9             -DLCNLPDLNENTLLKSLES RFDKGNIYTYVGSILISLNP FKFFPIYNPKYVKLYQN-RR LGDLPPHIFAIADAAYHTML HKKQNQCIVISGESGSGKTE
Xetr_Myo9A          EDLCGLPDLSERALLETLRT RFRQEKIYTYVGTILVAVNP FRFLPIYNPKYVKMYDNR-R LGDREPHIYAVADAAYHAML RRQRNQCIVISGETGSGKTQ
Xetr_Myo15          HDMTQLEDLQEASVLGNLRK RFEREVIYTYIGSILLSMNP YKMLNIYGTDHVLKYEGK-A LGENPPHLFAIANVAYTKLM DAKNNQCIIISGESGSGKTE
Xetr_Myo35          SDLCFLKDVCESAVLLCLKR RFHRNETYTHVGHMLLSMNP GKALNIHSPDMAQVYWDG-E SVEKPPHIFAVAEDAYSLSQ TSDVAPNILLSGHSGSGKSE
Dapu_Myo15          EDMISISDLNEASLLWNLKV RYDGQHIYSYTGNILVSVNP YKMFDIYGLDMVKKYENQ-I LGTLPPHLFAIGSAAYDRMT SRDQPQVVIISGESGAGKTE
Leed_CS1            --PYATIYPTDDAILAVLQA RFRADLPYTRAGATFVAINP YKTLASVNNASAKEYEERSK DTSLPPHIYDLAARMYLVMR RRNESQALLARGITGSGKTS
Cogr_CSc            -------SQSDTQLTAHLAS RFHVSLPTARLSSHLISLNT YTS----------------K DGSAMGGAEDIADRAWIRLG HRSENQAVVFLGESGSGKST
Dapu_Myo20          DDLIHLAPLTEDAVLKTLQA RFYADEFFTNIGPILLAFNP YNEVG----NALTLSSTRN- -IVQKPQLVRVVQEAVRQQS ETGYPQAIILSGCSGSGKTY
Pd_Myo20            -------------------- -------------------- -------------------- -------------------- --------------------
Dapu_Myo22          SDLTLISNIDERGINLTLST RYKRDEIYTCAGTILIAVNP YKDLPIYGLDYVERYQKKHE LVNREPHVYLLTESAFTSFQ HKGVNQSVVISGESGAGKTE
Pd_Myo1             EDHT-----NIDAFLENLRK RFQENLIYTYIGPVLVSVNP YKQIDIYNQNVSESYRN-VN FYELPPHIYAIGDAAYRSVK NESRDQCVLISGESGAGKTE
Xetr_Myo1A          GDMILLESLSEESLLVNLKE RFKSDEIYTYIGNVVISVNP YKSLPIYSPEKVEEYRNC-N LYQLKPHIYAVADDAYQSLR DRDRDQCILITGESGAGKTE
Dapu_Myo1A          GDFVLLDEISPDSFIHNLRI RFNGGKFYTYIGEVCVSVNP YRTLNIYGPDYVSQYKGR-E IFERPPHIFAIGDAAYKTMK RNGKDTCIVISGESGSGKTE
Dapu_Myo19          EDLTHLKKLDSSTVLQCLQS RYKQKQFYTWAGPTLVALNP LKDVPLYSKRMMHLFAHE-S FTYKTPHIFALGQRTLRHLQ LGKTHQAIVVNGESGAGKTC
Xetr_Myo19          DDLTKVNPVTTATVLKCLQA RYSAGVFYTNAGCTVVAVNP FRPVCLYSSEVMKEYHAASN PQGCKPHIFTVAEQAYKNVQ SQIVNQSIIVSGESGAGKTW
Pd_Myo19            EDLTKLSPLSTPAVLHCLQQ RYTEDEIYTRAGETLVAINP FKVIPLYDEERMEKYHNHLE SERRKPHVYETAELAFRAMV NENHNQSIIVSGESGAGKTE
Dapu_Myo5           NDLTSLSYLHEPAVLHTLKV RFNYNAIYTYCGIVLVAINP YQELSIYSQDTVLAYRNRNQ YGSLDPHIFAVAEEAFTKME RESQDQSIIVSGESGAGKTV
Xetr_Myo5A          NDLTALSYLHEPAVLHNLKV RFDSKLIYTYCGIVLVAINP YEQLPIYGSDIINAYSGQ-N MGDMDPHIFAVAEEAYKQMA RDERNQSIIVSGESGAGKTV
Xetr_Myo6           EDNCSLMYLNEATLLHNIQV RYSKDKIYTYVANILIAVNP YFDVPLYTSEQIKQYHGK-S LGTLPPHVYAIADKAFRDMK VLKMSQSIIVSGESGAGKTE
Pd_Myo6             DDNCALMYLNEATLLNNLRI RYKKDAIYTYVANILLAINP YKELPLYSKQTIASYTG-KS LGTMAPHIFAIADKAFRDMK MLKESQSIIVSGESGAGKTE
Dapu_Myo6           DDNCSLMFLNEATFLNNVAL RYKRDQIYTYVANILIAVNP YFEIKLYSKDAIKKYQGK-S LGTLPPHVFAVVDKAFRDMR VLKQSQSIVVSGESGAGKTE
Dapu_Myo18          EDLAQLRHLNEASLLHTLRC RYAASLPHTYAGQSLVVINP VTPLAVYSERMIQLFRGC-K AEDMPPHIYSSAQSAYRSML ATRCDQSLIFVGRSGSGKSH
Xetr_Myo18A         EDLASLVYLNESSVLHTLRQ RYGANLIHTHAGTSMVIMNP MSSPSIYSEKVMHMFKGS-R REDMSPHIYGVAQAAYWNML MTRQDQSIVLLGSSGSGKTT
Dapu_Mhc1           EDMADLTYLNDAAVLHNLRQ RYYHKLIYTYSGLFCVAINP YKRFPIYTQRVIKMYIGK-R RNEVPPHIFCISDGAYMDML TNHENQSMLITGESGAGKTE
Xetr_Mhc1           EDMAMMTHLNEASVLYNLKE RYAAWMIYTYSGLFCATVNP YKWLPVYNPEVVAGYRGK-K RMEAPPHIFSLSDNAYQAML TDRENQSVLITGESGAGKTV
```

```
Pd_Mhc1          EDMANLTYLNEASILHNLRT RYYNSLIYTYSGLFCVAINP FRRLPIYGMNVVLKFKG-KK RNEVPPHLFCIADTAYSNML MDRENQSMLITGESGAGKTE
Pd_Mhc2          EDMAELTCLNEASVLHNLKE RYYSGLIYTYSGLFCVVVNP YQRLPIYTEEVIEQYKG-KK RHDVPPHVFAIADTAYRSML QDREDQTILCTGESGAGKTE

Pifu_CS          STKYMIRHLMKISPSDDTRL LDKIVQINPLLEAFGNAATV MNGNSSRFGKFIELHYTEHG KLLGAKIDDYILEKSRVVHR TCGEKNFHVFYALFAGMSRE
Atri_CS          STKYMIQHLMKISPSDDSLL LDKIVQINPLLEAFGNAATF MNKNSSRFGKFIELHYSEYG QLLGAKIDDYILEKSRVVHR SNGEKNFHVFYALFAGMSRE
Myga_CS          STKYMIQHLMKLSPSDDESL LDKIVQINPLLEAFGNASTV MNKNSSRFGKFIELQYAEDG SLLGAKIDDYILEKSRVVHR SPGEKNFHVFYALFAGMSRD
Logi_CS1         SSKYIIRHLMHVSSSEDRQY LDKIVQVNPLLEAFGNASTV VNNNSSRFGKFVELYYKDNG EILGAKFHDYILEKSRVVHR CRGEKNFHIFYALLAGMSRE
Logi_CS2         STKLILKHLIHLN-KSPSDL IAKLEKINQVLEFFGNAETS LNENSSRFGKLVEISYTDNN TIEFAKIEGYILEKFRVTEP NENEKNFHAFYALFAGMSDC
Cate_CS1         STKLVISHLAYRA-AHDKDL ARKVTQVNPLLEAFANAQTP VNTNSSRFGKYIQLDFNPSN QITGGKVYDYLLEKSRVVQR ATTDSNFHVFYYLVAGTSTE
Logi_CS4         SVKHMISHVTYKS-ESIHFL DVKINEVNPLLEAFGNAQTL MNDNSSRFGKNLELKFTSKG KLAGAVLTDYLLEKSRVVQH GPGERNFHIFYYLFAGMTKD
Macr_CS          -------------------- -------------------- -------------------- -------------------- --------------------
Logi_CS3         STKLAISHLSYMCSTDSNNL QEKIIQINPLLESFGNAVTP MNDNSSRFGKLIELHFTEDG HLSGAKIFDYLLEKSRVVHH GSGEQNFHFLYYLFDGLTED
Logi_CS5         TTKFIIKHIVHLCGNAQSDL HKKILNINPLMEAFGNARTN LNGNSSRFGKFIQINFDAEG GIQGAVVRDYLLEKSRVVHQ PNGEGNFHIFYAFIAGVDPS
Miye_Myo         STKYMVKHLVSLCPKETGDL HERIVKINPLLEAFGNAKTT MNDNSSRFAKYLEMSFATNG QVTGAIVRDYLLEKSRVVDQ MDKEGNFHIFYCLFAGAPVT
Dapu_Myo21       NFTQIVRHLCYMAKG-NQTV ADRIEQLPLILDCFGNAATE GNANSSRHLRYLEMKFTDSG KLSGAIINVFMLEKWRLSET SKSNRNYHIFYYVYYGLKLE
Dapu_Myo3        TANYLLKQLVALGKAPNRDL EDKILMMNPIMEAFGNARTG INDNSSRFGKYLDVTFTSAG MVSGARLSVYLLEHSRVVQQ AANEKNFHIFYYLCDGLAAD
Xetr_Myo3A       GAHLLVQQLTVLGKANNRTL QEKILQVNNLVEAFGNACTI INENSSRFGKYLEMKFTSNG TVVAAQISEYLLEKSRVVHQ ALGEKNFHIFYYIYAGLSEK
Xetr_Myo16       SCKHIVKHLAFRCGSSKNSL YSKIKHVNCILEAFGHARTS INNSSSRFINFLELQFCEKK QTLSG---------------- ----------YFMMDGLSAE
Pd_Myo3          TCKMIVEHLNVIAHSLEANL SKKINQVNPLLEAFGNAQTV KNSNSSRFGKYLELVFSSSG EILGGKVNEYLLESSRIVFH GPQETSFHIFYLLYAGLTXX
Xetr_Myo28       STKLLLKHLVRRSRG-NLQL GQQILQVNPLLEAFGNAQTV MNPNSSRFGKYIQLRF-RDG AVRGAKINEYLLEKSRVSHQ DPGEKNFHIFYYMLGGIPDE
Dapu_Myo7        STKLILQYLAAISGK-HSWI EQQILEANPILEAFGNAKTV RNDNSSRFGKYIDISFNKHG TIEGANIEQYLLEKSRIVAQ NPGERNYHIFYCMLAGMTKE
Xetr_Myo7A       STKLILQFLAAISGQ-HSWI EQQVLEANPILEAFGNAKTI RNDNSSRFGKYIDIHFNKKG AIEGAKIEQYLLEKSRVCRQ AQDERNYHIFYCMLKGMSPE
Xetr_Myo10A      STKLILKFLSAMSQHSTASV EQAILESSPIMEAFGNAKTV YNNNSSRFGKFIQLNICQKG HIQGGRIVDYLLEKNRVVRQ NPEERNYHIFYALLEGAGKE
Dapu_Myo9        STNFLLHHLTALSQKGGCGV EQTILSAGPVLEAFGNAKTS HNNNSSRFGKFIQVNYRENG QVHGALVQKYLLEKTRICFQ AKNERNYHVFYHLLVGSSAQ
Pd_Myo9          STNLLLHHLSALS-QRGVGI ENTILGGGPVLEAFGNAKTV HNNNSSRFGKFIQVNYKENG ---------------------- ----------------------
Xetr_Myo9A       STNFLIHHLTALSQKGASGV EQIILGAGPVLEAFGNAKTY QNNNSSRFGKFIQVNYHETG TVRGAYVEKYLLEKSRLVYQ EQQERNYHVFYYLLAGASEE
Xetr_Myo15       ATKLVLRYLVAVNQR--RGV TSKILEATPLLESFGNAKTV RNDNSSRFGKFIEIYL-EEG VICGAITSQYLLEKSRIVFQ AKNERNYHIFYEMLAGLPSQ
Xetr_Myo35       AVKLMALYLVTLSTR----Q EEKILQLEKVLESFGNAKTV FNDNSTRFGQVLQLYL-HSG AVVGGMVSPFLLEKSRVVFQ AQGERGFHVFYELLRGLPAH
Dapu_Myo15       ATKLIMQYLAAVNNSSSSLV TEQILEASPLLESFGNATTV RNDNSSRFGKYTQIFF-RNG AIAGAKITEYLLEKSRIVTH AADERNYHVFYELLKGLRPE
Leed_CS1         NLRLFTNQILRLSSSKEAKI AEQVKALGIVLESFGNSKTL MNPNASRHSRYTELHFNERG RIAGAKVLTFGLDKSRLNRL TQEERSFHVFYQFIAGCTSA
Cogr_CSc         IRSHLLTAFLG---KSSTPL STKVSLAAYVFDTLTTTKTA TTPTASKAGLFYELQYTTNP VLIGAKLLDHRLERSRITDI PTGERNFHILYYLMAGTSAA
Dapu_Myo20       ASMLLLRQLFDVAGGGETDA FKHLAAAFTVLRSLGSAKTQ ANSESSRIGHFIEVQV-TDG ALYRTKIHCYFLDQTRVVRQ LAHEKNYHIFYQMLAGLTVE
Pd_Myo20         -------------------- -------------------- -------------------- -------------------- --------------------
Dapu_Myo22       TAKFVLNYLCSVTSN-STWV QHQIIEANTILEAFGNAKTV RNDNSSRFGKFMQVCFDAQV KISGCIIQDYLLELSRITVQ SKGERNYHVFYQLLAAAQHD
Pd_Myo1          ASKKILQYIANTSTHQHDHI KDRLLQSNXXXEESGKAI-- -------------------- -------------------- --------------------
Xetr_Myo1A       ASKLVMSYVAAVCGK-VNQV KEQLLQSNPVLEAFGNAKTI RNDNSSRFGKYMDIEFDFKG DPMGGVISNYLLEKSRVVKN VKGERNFHIFYQLLAGGGTE
Dapu_Myo1A       ASKIIMRYLAAVTNLSIERV KNILLQSNAILETFGNAKTN RNDNSSRFGKYMDIHFDFKG DPIGGHINNYLLEKSRVIVQ QAGERNFHSFYQLLFGAPDG
Dapu_Myo19       NALYLLSFLAAMETNNSSNI EVLLCQSNPLLEALGNAQTL KNENSSRFGKLIRLVYLGEG QLVSAHIETYLLESTRVARQ AAGERNFHIFYQMLAGLSDS
Xetr_Myo19       TSRCLMKFYATVSASRVERI ESRVLDSNPVMEAFGNACTL RNHNSSRFGKYIQLQLNRTQ QITGASIQTYLLEKTRVAHQ APLERNFHVFIYQVVKGASRH
Pd_Myo19         SASHLMRYLVTISGEQDERI EQRILISNPILESFGNASTL RNHNSSRFGKFIELHYTEHG KLLGAKIDDYI--------- ----------------------
Dapu_Myo5        SAKYAMRYFATAGGSAETQV ERKVLASSPIMEAIGNAKTT RNDNSSRFGKYIELGFNKDY HIQGAGMRTYLLEKSRVVFQ SAEERNYHIFYQLCAASSLP
Xetr_Myo5A       SAKYAMRYFATVSGSAETNV EERVLASNPIMEAIGNAKTT RNDNSSRFGKYIEIGFDKRY RILGAHMRTYLLEKSRVVFQ AEEERNYHIFYQLCASASLP
Xetr_Myo6        NTKFVLRYLTESYGS-GQDI DERIVEANPLLEAFGNAKTV RNNNSSRFGKFVEIHFNEKN SVVGGFVSHYLLEKSRICVQ GQDERNYHIFYRLCAGAPED
Pd_Myo6          GTKHILKYLTEVWGSHAGPI QQRIVESNPLLEAFGNAKTV RNNNSSRFGKFIEIHFDAKY LVTGGYISHYLLEKSRICKQ GRDERNYHIFYRMCAGAPEK
Dapu_Myo6        STKYILRYLCESGGLSAGPI EQQILDANPLLEAFGNAKTT RNNNSSRFGKFIEIHFTSKY TIAGGFISHYLLEKSRIVTQ SKEERNYHFFYQLCAGASEK
Dapu_Myo18       NFRNSLHYLCLAAGCPKIIT VEKLNSICLLLEAFGNCRTA LNTNATRFTNIFSLDFDQSG QIAAASIQVLMPERERVTRR PDGQPSFNVFYELVAGATGE
Xetr_Myo18A      SCQHLIQYMATISGSIKILS AEKWQAVYTILEAFGNSSTS LNGNATRFSQIISLDFDQTG QVASASIQTMLLEKLRVARR PANEASFNVFYYLLAGADNT
Dapu_Mhc1        NTKKVIQYFAQIAKDTGGNL EDQIVQTNPVLEAFGNAKTT RNDNSSRFGKFIRIHFGNSG KLAGADIETYLLEKARVISQ QALERSYHIFYQIMSGKLPT
Xetr_Mhc1        NTKRVIQYFATIAALGQGNL EDQIIQANPLLEAFGNAKTV RNDNSSRFGKFIRIHFGTTG KLSSADIETYLLEKSRVTFQ LSAERSYHIFYQILTNKKPE
Pd_Mhc1          NTKKVISYFAQVAASSAGTL EDQIVQANPVLEAYGNAKTT RNNNSSRFGKFIRIHFGPSG KIAGCDIETYLLEKSRVTYQ QPGERNYHIFYWLLSDAFPD
Pd_Mhc2          NTKKVIQYLAYVAASLRSXL EQQLLQANPILEAFGNAKTV KNDNSSRFGKFIRINFDSSG FISGANIETYLLEKSRSVRQ APNERAFHIFYQM-------

Pifu_CS          RLLYYFLEDPDCHRIMRDEE MQCGVFSDSEEYEHYRKMFT DLTTIMIQIGFSEEHIQVIF LILAAILHLANIVFMPIETT DG---VTVADEYPLHAVAKL
Atri_CS          RLLYYFLEDPDCHRIMREDD VQRGVFQDAEEYQHYKSMFS DLTVIMSHIGFSDEHIQVIF LILAAILHLANIVFMPIDST DG---VSVADEYPLHAVAKL
Myga_CS          RLLYYFLEDPDCHRIMRDDD QSSSVFKDNEEYEYYHQMFT NLTKIMAEIGFSEEHISVIF LVLAAVLHLANIQFVTCEET DG---VTVADEYPLHAVAKL
Logi_CS1         KLLYYFLEDPNHYRILKDAN TNQPVFGSKNQLLYYKQMYD NLTNIMYQIGFTEEDLTFIH MTISAILNLANIEFIPNLDA EG---VTIADDYPLKAVCNL
Logi_CS2         ELRELELEELSEYRILTTFN KV--------EEERSYERKYN EMINILDEVGLNKETRDQIH SILAAILHLTNIEFDEDPSG AIGF-LDVKALESIDRFGRM
Cate_CS1         EKERFCLQDADAFKILRREG AS-RVFRNDYQIKENAEYFE EMRQTMRNIGLEREEILFTL ASVAAILNLTNIEFEETGSS DGR--LGIVNLSFLENVSRL
Logi_CS4         ELRNYSLGHPEDFRILRNDN RNRPVFKSTRDLENSRLMFH KQLNVMKLVGFTDWEISMVF TVLSAVLHVTNIVYIKDEKT EK---ARFADLKNLEIVAVL
Macr_CS          -------------------- -------------------- -------------------- -------------------- --------------------
Logi_CS3         KKKYYYLEDIEQFRVLQNGR RN-----KSEQVERNKQKFH QQIDIMKMIGFSEEDISSIL TLLSAILNVTNLVFVDDTEV DG---VNIENEFHLRVATSF
Logi_CS5         LKRKLLLDK-KSYRITKN-- -------IVGDLDYHRNQYN DYLKIMRLIKLPKEDIEAVQ IILAAILHITEIRFEEDGDS QG---VVVSNEDEMINAASL
Miye_Myo         VLKNLHLKDARTYRIVKGNE EL------LTRTEFYRAMYQ EQIEVLKSINLEQEDIDIIH TILAAILLITQVEFLEPDDP NEP--MKIKDTTFVENVADL
Dapu_Myo21       GLEKYFLDGQTNMRFLPI-- -----GDSDAEISYYLSGYK KLKKYLRFWDMEGDEQSFIF HTIVAVLLIGQIQFGD---- -RNGESYITNPDILNKVSTL
Dapu_Myo3        NKADYFLDSRRTHRYVTLP- -----PLTSLERQIYAEKFQ SVRKGFDQLGFASDELDAIY SVMAAVIHIGDLDLVPGPGT DNTDRCKIANLPQANIVAKL
Xetr_Myo3A       KKAHYKLPDIKPPRYLQNEH IKM--VQDFMNKPFYKSQFE LIEQCFKVIGFSMEELGSVY SILAAILNVGNIDFTSVASE HQIDKSHINNPTVLENAASL
Xetr_Myo16       EKYSLYLKSLSAHRYLNQNV PED--TISLSASTQNREKLD NLKQALIALGYNNLEIENLF VILSAILHIGDIRFTA---L TDAETAFVSDLQLIEQVAGM
Pd_Myo3          XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXX XXX--XXXXXXXXXXXXX
Xetr_Myo28       EKQVYGLLQPSLYRYIGS-- -----KWEEATPSHWVESYQ RVCNAMRMVGFQEQEELDLK VILSGVLSLGNVVFEP--QE SGGVGVCPTAMGWLKAAAGQ
Dapu_Myo7        DKQKFDLQDASQYKYLTGG- -----NSTTCQGRNDANEFA EIRSAMKVLLFTEPEISDIL RVLAALLHLGNVSFKG-VVI SNMDASEIPDPSNAIRVSKL
Xetr_Myo7A       QKKKLSLGQASDYNYLCMG- -----KCTTCDGRDDSKEYA NIRSAMKVLMFTDTENWEIS RLLAAILHMGNLRYEA-RMY DNLDACEVVYSTSLTTAATL
Xetr_Myo10A      ERDAFYLLQAEKYHYLNQS- -----GCILDETISDHETFQ EVKAAMEVMKFSREEVREIF RLLAGILHLGNMEFIT---- --AGGAQISSKTALGRASEL
Dapu_Myo9        EKQALHLLPPEAYRYLNQS- -----DYHNLDEGDECYELS RLKQSMEFVGFCSETQRRIF SVLSAVLHIGNLEFQPKSTY NHDESVTVKNTAVVATIAQL
Pd_Myo9          -------------------- -------------------- -------------------- -------------------- --------------------
Xetr_Myo9A       EKAEFHLEHPESYHYLNQRT RKPQRDSLGGEGEDLRHDFE RLQLAMEMVGFMPATRRQIV SLLSAILHLGNIQYKR--KT YRDESIDISNPETLNIVSEL
Xetr_Myo15       QKQMFYLQDAETYYYLNQG- -----GNCDIPAKSDADDFR RLLNAMESLSFTGEDQDSIF RILSSILHLGNVYFEK-YET ESQEIASVVSASEIRVVAEL
Xetr_Myo35       EKQRLYLQEPETYYYLNQG- -----RACDVRGKNDKQDFV LLGKCLRAIGLSFTELTAIW AVLSAILQLGNICFSS-CEK DSFGLAAAFSATEVRIVASL
Dapu_Myo15       EKERYGLTAADNYFYLNQG- -----GHCESATKHDGQDFK TLLSAMQILGFNAEEQDVIF RILASVLHLGNVFFHRKALK HGQEGVEMGSDAEVRWVSHL
Leed_CS1         ERDTLNNIEDPSDYALLASSG TY---RLPAGPSSDDSVAMG DLRAAMRTLGFKPKATAAIF SLLIAILLLGNLEFGEGDFH TVS--AHVTNVEVLDHVSRL
Cogr_CSc         EKSHLGFETPKRWKYLGHPT QL--------KVGINDAEGFQ VFKNALKKLEFPRSEIAEIC QILASILHIGQLEFETTSDT QATGDTAVKNKDVLGIVAAF
Dapu_Myo20       ERARLSLEGVDNLRYLNR-- -----GDTRCNQAEDARRFA NWKACLGVLGIP---FLDVV RVLAAVLLLGNLQFHE---T NGGGGQCETHEAELKAVAGL
Pd_Myo20         -------------------- -------------------- -------------------- -------------------- -----VEVQGNNEIKSVSSL
Dapu_Myo22       KESNLMLKSWNHYEYVNQS- -----GCASINGVDDSKKFD GLRLALNVLQVTTEDTESIF SVVAGILWLGNLHFKD--ID GEKCELLAEDEEIVAIIGKL
Pd_Myo1          -------------------- -------------------- -------------------- -------------------- -------VHGEDHVNNMAEL
Xetr_Myo1A       LLGKLKLVPCSKYNYLNN-- -----DPSSLNGMGDNENFQ IVKNAMEIIGLSDAEVTSIL EIVAVILKLGNVEIKGQFQA NGMAACYITDTKGIKEIGAL
Dapu_Myo1A       EAAKLQLKRPTFYHFLRQG- -----GVAKVDSISDRTDYK NVSSAFRALNFDGPSVDTIW KIVAAVLHLGNAEFKS---- -EGDKTRLANSATLSTVASL
Dapu_Myo19       LKTELELKDVPIWEIASERD CN----------AVDEAQFH ATVAALEKLSL-EDTANQLW KLLAALLHLGNIKFTVTQET -----WTTNQRHHLESAARL
Xetr_Myo19       EREEWNLPEKANFSWLPNY- ----------ENNLEEDDFE VTKDAMLHLGIDQTTQNNIF KILSGLLHLGNIQFSDSVDE SQPCEPLNYTQEFASVAASL
Pd_Myo19         -------------------- -------------------- -------------------- -------------------- --------------------
Dapu_Myo5        EMAYLQLQHQDHFSYTRKG- -----NCPTIDGVDDLAEFQ ETRRALTLLGFSEDQQADMF RVFAGLLHLGNVTIVD---A DHEGSNIPKTDTYLASFCSL
Xetr_Myo5A       EFKMLRLGTANDFHYTKQG- -----GSPVIEGVDDQREMK NTRQACTLLGIGDSYQMGIF RILAAILHLGNVEFKS---R DSDSCLVPPKHAPLNIFCDL
Xetr_Myo6        IRQKLFLCSPDTFRYLNRGC TRYFAGALKDPLLDDHGDFN RMCIAMKKIGLDDTEKLDLF RVVAGVLHLGNIDFEEAGST SGGCTLRKKSSESLQCCAKL
Pd_Myo6          LREQLRLGSPDSFNYLRYGC TQYFGGSLVDIQLDDVNDFN RMDSAMSKMGLPEDEKMAVY TVVAGVLHLGNIEFEEETEG SKGSKVTKQSEGSLVIAAKL
Dapu_Myo6        LKGQLQLRSPDHFHYLSQGC TQYFLGSLNDPLLDDANDFK QTDKALSNLGMSDDEKLAIY SAVAAILHLGNITFEDAEDA RGGCQIAKSSHEALRNASAL
Dapu_Myo18       LRRYLQLDSVSGGGPNAEPC AYMTPLQSAEDRQKAAVSYA RLVAALESLGVAETDSRCIW SILAAIYHLGVAGAMR---G VSNNKIVFARPAAAQRAAIL
```

```
Xetr_Myo18A         LRTELHFNH------LAENN VFGIISAKAEEKQKAAQLFA KLQAAMKILGISSEEQKAFW LILGAIYHLGSAGATK---- --AGRKQFARHEWAQKAAYL
Dapu_Mhc1           LKADCCLVDIYQYNFVSQ-- -----GKITIPSMDDSEEMA LTDEAFEILGMG-EQRPEIW KITAAVMHFGTMKFKQ---R GREEQADPDGTQEGENVAKM
Xetr_Mhc1           IVEMLLLTTPYDYPSISQ-- -----GELVVKSINDEEELM ATDSAIDVLGFNQEEKMGIY KMTGAVMHHGNLKFKQ---K QREEQAEPDSTEVADKIAYL
Pd_Mhc1             YHDKLVVTPPGLFFFINQGC LT-------VDGMDDPEEMK LTDDAFAVLGFTTEEKMGLY KCTCAILHMGEMKFKQRPRE EQ---AEADGTAEAEKVSFL
Pd_Mhc2             -------------------- -------------------- -------------------- -------------------- --------------------

Pifu_CS             LGIEDELTEALISNVSNIKG EKIQSWKSLREANDSRDALA KDLYSRLFGWIVGQVNRNIW PQRRGPSIGLLDMSGFENFR LNGFDQFLINISNEKIQQYF
Atri_CS             LGIEDELTEALISNVNTIKG EKIQSWKSLREANDSRDALA KDLYSRLFGWIVGQINRNIW GKRRGSSIGLLDMSGFENFK VNGFDQLLINISNEKLQQYF
Myga_CS             LGIEDELTEALISNVNYIKG ERIQSWKNFRDANNSRDALA KDLYSRLFGWIVGQINRKIW TAKRGPSISLLDLSGFENFT NNSFDQFFINASNERLQQYF
Logi_CS1            LRISDTFMDALVSQVNIIRG ESFRCPKNEREANNSRDALC KELYARLFGWIVGQINRNLS SDSDGPSIGLLDMSGFENFE RNSFDQFLINVSNERLQQHF
Logi_CS2            LNLEEENIRRLLLSTEIRDE KRL---KTLSDAEDGRDALA RELYSCLFYGIIDVINSILA NNHKKKSIRILDISGFENLK NNSFEQLLINTVNENLQQYF
Cate_CS1            LQVDQDLERALLCNEVHIQG EEVLRPRSNADASDSRDALA KALYHRLFGWLVSRINDHLR PTSRSNFIGILDLSGFENFE ENRFEQLLINAANERMQFYF
Logi_CS4            LQIDIVLSDVLISSANYTRG ERLVIHRTVEQAAESRDALA KALYSRLFGWIVRQINTRLK PDKSGPSLGILDMPGFENLK INSLEQLCINVANEQMQTYF
Macr_CS             -------------------- -------------------- -------------------- -------------------- --------------------
Logi_CS3            LGIEIELTTALISSTIFVRG ERVETLKKKHQAEDCRDTIA KILYSRLFGWIVGQINVHLK PKTRNSTVSILDLAGFECLP MNTFDQLCINAANERLQQYF
Logi_CS5            LGVDPDFGKSLVLEDIVLPD ETITTFKSLEQAEDGRDALA KSLYERLFGWMIQNINVSLH TQQRMGSLGILDICGFEIMR NNGLEQLCINLVNEQLQSFM
Miye_Myo            LNVSYELGHALIATKQTYVG ETLVKRKSMYQAIDSRDAFA KALYERIFGWIVRQINLNLH PSKGSTSIGILDIAGFERLE INSMEQMCINLINERLQSFT
Dapu_Myo21          LNVDSALGWALCNSSKFSNG NLEVEPNTCSESAQARDALA RGLYSRLFDWLVNTLNLNMC LGREKHSMGIVDPPGFDSLR RNSFEQLISNITNEQLSYHY
Dapu_Myo3           LGVEPQVIEALTSVSVVTRG ETITRHNDLDAACATREAMA KGLYARLFDWVVQQINRHLS FGREPLSVGLLDIFGFENHG RNSFEQLCINIANEQIQYFF
Xetr_Myo3A          LCIQPDLKDALTSHCVVTRG ETIIQQNTVEKAADVRDAMS KALYGRLFSWIVNRINALLK PEKAGFNIGILDIFGFENFK KNSFEQLCINIANEQIQFYF
Xetr_Myo16          LQVSPDLSAALTTEVQYFKG DMITRRHSVDTAEFYRDLLA KSLYCRLFSFLVNNINYYLQ QQSPPLTVGILDIFGFEDFQ KNSFEQLCINLTNEKIHQHI
Pd_Myo3             XXXXXXXXXXXXXXXXXXXX XXXXXXRSLTQAADARDALA KTLYRRLFSWIVNGINQLIQ NVKSDLCLGIFDIFGFENFD KNSFEQLCINLTNEQLQFYF
Xetr_Myo28          FGVQEELLSCLTCTLSLTRG ESIRRLHSQQQAEDSRDSIA RGVYSRLFGWIVCKINELLA GDMELQEIGILDIFGFENFS VNRFEQLCINLANEQLQNFF
Dapu_Myo7           LGVDPRMVDALTTKTIFAQG DSVVSRLSKVQAVDVRDAFA KGIYGRLFIWIVTKLNCAIR KSDDTSSIGVLDIFGFENFS INSFEQFCINYANENLQQFF
Xetr_Myo7A          LEVGPQLMNCLTSRTIITRG ETVSTPLSMEQALDVRDAFV KGIYGRLFVWIVDKINAAIY RKAARRSIGLLDIFGFENFT VNSFEQLCINFANENLQQFF
Xetr_Myo10A         LGLDSEMAEALTQRSMILRG EEISTPLSVEQAVDSRDSVA MALYSQCFAWVIMKINSRIK GKDDFKSIGILDIFGFENFE VNRFEQFNINYANEKLQEYF
Dapu_Myo9           LRVKEELHQALISKRARASG ETLVINYRMPEAIAARDAMA KCLYGSLFDWIVMQLNHALL AKKRAYSIGVLDIFGFEDFG MNSFEQFCINYANEHLQYYF
Pd_Myo9             -------------------- -------------------- -------------------- -------------------- --------------------
Xetr_Myo9A          LKVREELLDALITRKTVTVG EKLILPYRLPEAITVRDSLA KSLYSALFDWIVFRINHALL NNSQTLSIGVLDIFGFEDYG TNSFEQFCINYANERLQHYF
Xetr_Myo15          LQISPELQKAITYKVTETMR EKILTPLTVDSAVDARDAIA KILYSLLFSWLTDRINKLVY PKQDALSIAILDIYGFEDLS FNSFEQLCINYANEYLQFFF
Xetr_Myo35          LQISPELQRAITQRVTVTSY DRVLSPLSVEASIDVRDAIS QTLYSLLFNWLLEKGNSWLH PQKMDKTLGIVDVCGFENLG VNSLEQLCRNYANEHIQHYS
Dapu_Myo15          LQLRTDIVAALTTKTTEARN EKLHTPLNIDQALDARDAIA KALYSSLFGWLVNRVNSIVN KGERSTSINILDIFGFEDFP ENSFEQLCINYANENLQFYF
Leed_CS1            LGVSSELSQALTNKTSYVRK ELYTALLNAQQSALQRDQLV RDLYAIMFAFVVETANPQGA PSSPHSQIVLLDQPGYQAYG QNGFDEFCINFADEMLQSYF
Cogr_CSc            LGVGTQLQTTLSYKTKMIHK ERVTIMLDPVGARSHANELA RTLYSLLVAYIVESINQRIC APEVANTISIIDFPGFAAAT HSALDQLLNNAACEALYNLT
Dapu_Myo20          LGVTPALLRGLTMRTHNVRG QLVKSICDANMSGGARDALA KALYCRTVATIVRRANSLKR LSSTDGFIGILDMFGSRIVS DSGNVLLTVNYIRDQMQHFF
Pd_Myo20            LGVSPVLYRALTTRTKRSRG QVFKSLADASTANYNKDCLA RALYSRTVYAIVRRANSCKR PTTANSFVGILDMFGYEDNK ENELEQLCINLCAEKLQQYY
Dapu_Myo22          LGLSSELKEVLILRQINVRG NVTEIPLRLHEARENRHAMA RVLYSRTFSWLLHHINRCTS PGETNRFLGILDIFGFENFE QNSFEQLCINYTNEKLHRFF
Pd_Myo1             LGCSSTLLSALQNRSIHAEQ QKVTTPLSRDQAVYARDALA KGIYERLFSWLVQKINKSLA CSKKKTVLGLLDIYGFEIFE SNSFEQFCINYCNEKLQQLF
Xetr_Myo1A          IGLDSSLENAFSVRTVEAKQ EKVVTTLNQSQGIYARDALA KNLYDRLFSWLVQRINESIK VSEGRKVMGVLDIYGFEILE VNSFEQFIINYCNEKLQQIF
Dapu_Myo1A          LCVDPNLNRALCHRVIAANG ELMEKGHTTSEAVYGRDAFA KALYDRLFSWIVERVNSAIE VQRPGTVIGVLDIYGFEIFD INSFEQFCINYCNEKLQQLF
Dapu_Myo19          LGLNSNLFQVFTVRNFQASP AVVRPCSSQNECAARRDTLI KLLYRMMFDTVLESVNEKLQ HTNTPKYLCILDLYGFESFQ GNSLEQLCINYANERLQYYF
Xetr_Myo19          LKIPVSLLERLSIRTITAGK QQVKKPCRKSECDTRRDCLA KTIYARLFEWLVTVINENIC AEAWNNFIGLLDVYGFESFP ENNLEQLCINYANEKLQQHF
Pd_Myo19            -------------------- -------------------- -------------------- -------------------- --------------------
Dapu_Myo5           MGLDVALRKWLCFRQIVSMK EVFTKPMTKAEASFARDALA KHIYSLLFQKIVTMINKSLA SSSPHRFIGVLDIYGFETFE WNSFEQFCINYANEKLQQQF
Xetr_Myo5A          MGVEYEMSHWLCHRKLVTAA ETYIKPISRLQATNARDALA KHIYAFLFNWIVCHVNKALH SSTQNSFIGVLDIYGFETFE INSFEQFCINYANEKLQQQF
Xetr_Myo6           LGLDQDLQVSLTSRVMLTKG TVIKVPLKVQQAENARDALA KAIYSHLFDHVVNRVNQCFP FERSSFFIGVLDIAGFEYFE HNSFEQFCINYCNEKLQQFF
Pd_Myo6             LGMDKDLHEALVSRVMQAKG TVIKVPLKVGEASNARDALA KSVYSKLFDYIVKRVNDSLP FTTSKNYIGVLDIAGFEYFQ ENSFEQFCINYCNEKLQQFF
Dapu_Myo6           IGVDPELHHALLSRVMQPKG TVIMVPLKVFEASNARDALA KAIYSRLFDYIVHRINQSIP FQSSMHYVGVLDIAGFEYFT VNSFEQFCINYCNEKLQQFF
Dapu_Myo18          LGIGSLLAQLAFYATPVQDG AKLSMADKHLDPSEALEGFA QGLYVEAFSALVALINRSIS STATISSIVAVDTPGFQNPT SATFQDLCHNYTQERLQLLF
Xetr_Myo18A         LGCSLELSSSIFKHQPKTDG SQGDGSGTKLSALDCLEGMA SGLYSELFTLLISLLNRALK SSQSLCSIMIVDTPGFQNPR GATFEELCHNYAQERLQALF
Dapu_Mhc1           MGVDGPLYMNFLKPRIKVGN EFVTQGRNVNQVVYSIGAMA KAIFDRLFKWLVKRVNETLE TGQRVTFIGVLDIAGFEIFD YNGFEQLCINFTNEKLQQFF
Xetr_Mhc1           LGLNSALLKGLCYPRVKVGN EFVTKGQTVPQVYNSVGALC KSVFEKLFLWMVTRINQQLD TKQRQFFIGVLDIAGFEIFD FNSLEQLCINFTNEKLQQFF
Pd_Mhc1             MGVNASMLKCLLKPKVKVGN EFVTKSQNMNQVNYAVAALA KSLYDRMFKWLVARVNLTLD TKNRQYFIGVLDIAGFEIFE YNTFEQLCINYTNEKLQQFF
Pd_Mhc2             -------------------- -------------------- -------------------- -------------------- --------------------

Pifu_CS             MDYIFPRERREYEIEGIEWR DIEYHCNDDVLELIFKKPDG ILPLLDEESNFPQSTDNSFV QKINKYCEENPRYVANVYNK PVFTVKHYAEQVQYDANGFL
Atri_CS             MDYIFPRERREYELEGIEWR DIVYHCNDEVLEVLFKRPDG VLSLLDEESHFPQSNDSSLV QKLNKYCHDSTRYVAQMGNR VCFGIRHYAEQVTYNADGFL
Myga_CS             MEYIFPREQREYDIEGIEWR NIMYHSNDDVLDLLFKKPDG ILSLMDEESHFPQSSDKSFV QKLNKYCSESDRYIPSLRNK TCFEIQHYAEQVVYNADGFL
Logi_CS1            MDYVFPREQKEYEIEGINWT DITYRSNEDVLELIFLKSYG ILPLLDEESTFPQSSDASLV HKLNRYCSPNDRYIPPKGNH CIFGVSHYAEKVYYDAEGFL
Logi_CS2            LRQAFKEEEREFLSEGVRFD SVQYRDNTDVISLLCRPPVG IIHILNDSVSVNPS-VRHFI QHVQQIHHQNPNFSIIKGNH TMFSINHYASEVIYDASNFL
Cate_CS1            NQHVFAEEQRELDAEGVVCG DVIYDDNSDVIDLIMCRLAG VLQMLDEESRHPQTTDQALI MQLHEHFELHPCYQRANKGA PLFTIKHFAGPVKYNAVGFL
Logi_CS4            NEHVFVLDRQEYEREGLSWK NIQYRNNNDLLNLFMMKPLG IFALMDEEIRFPRSTDSSLV EKITTHCKNTKLVSKSKSGD LAFCVHHYVGTVEYHAAGFL
Macr_CS             --------------------G IFDLMDEETRFPQSTDTTLV QKLNANCHKNSYFKQRPGSS KMFSIHHYAGEVCYNAHGFL
Logi_CS3            YDKIFKYERSDYEKEGVDVS LVEFQDNQDLIDLFFKKPLG LFYLIDEESSLPRANDKSLV QKLENICGSSPHYKTVPGDQ AAFTISHYAGKVSYCAKGMV
Logi_CS5            NKQVFIKEQEIYESEGLTLE DVNYPDNAKIIEMFTMNRYG IIALIDEDTTFAQATSMSLI TKLQDKWGHHDHFVSPVYDI PEFGIKHFAGEVTYDADDFI
Miye_Myo            NRNVMDYEMSIYKEEGIHVT GIKFKNNDALLDLFMKKTFG LLPLLDEESKLGQGSNERFV KKLNDKYDTHPCFTESPHGR VEFGVRHFAAQVWYDGSLFI
Dapu_Myo21          NQSLFSWEMQDYKDEEIRMG KFAFRDNRNVLNAMLSKPDG LLSVVDEQSRAPDASDESIL SHFEKVISPSSLQVE----N GSFVVHHSGGDVRYRMEDYP
Dapu_Myo3           NQHVFTWEQQEYMAEGINVD VVEYTDNRPVLDMFLAKPLG LLALLDEESRFPGASDNSLI EKFQKNIR-SDYYIRHKSNG LDFTVQHYAGRVSYDATSFL
Xetr_Myo3A          NQHIFAWEQNEYLNEDVDVR LIEYEDNRPILDMFLQKPMG LLSLLDEESRFPRATDQTLV EKFEDNLKSKNFWR-PKRVD LTFGIHHYAGKTYHPR----
Xetr_Myo16          TEVLFFQEQTECAQEGIPME TIYCGNHTTVLDFFFQKPMG FMSVLDEESQSIHTMEHSLS KRLQSLLETNAIYFSSKDGH PTFAVMHYAGRVTYNIFGAT
Pd_Myo3             NQHIFALEKKECEEEGVVHN DLFYNNNESTLNLF-LNSMG ILAILDEESRFPKATDHTLA TKLHNPGRDHEMYIAPPDKG LSFGIKHYAGTVIYNLNGVL
Xetr_Myo28          NHHIFLMEEQHYKEEGMERE AVPFSNNQATLELFLARPWG ILSLLDEQSSFPQATDVSFV EKLNSTYRESPLYERGRGFE PGFAIHHYAGKIKYSAAGFL
Dapu_Myo7           VRHIFKLEQEEYNQENISWQ HIEFVDNQDSLDLIAIKQMN IMALIDEESKFPKGTDQTML AKLNKTHGSNRNYVKPRSDL ASFGFNHFAGVVFYDARGFL
Xetr_Myo7A          VRHVFKLEQEEYNLENINWQ HIEFTDNQDALDMIAIKPMN IISLIDEESKFPKGTDTTML NKLNVQHKLNTFYIPPKNNY TQFGINHFAGIVYYETKGFL
Xetr_Myo10A         NKHIFSLEQLEYSREGLAWN AIDWIDNGECLDLIEKK-LG LLALINEESYFPQATDNTLL EKLHKQHANNHFYVKPRVAV HNFGVKHYAGEVLYDVRGTL
Dapu_Myo9           NQHVFKYEQEEYKREGIHWV NIQFVDNTGCLQLYESKPNG LLCILDDQCNFPGATNETLL QKFNSVNKDNDYYEVPQKRE AAFIVRHYAGKVKYQSNEFR
Pd_Myo9             -------------------- -------------------- -------------------- -------------------- --------------------
Xetr_Myo9A          NQHLFQLEQELYRSEGISWS NITYSDNGGCINLISKKPTG LLQLLDEESNFPQATHQTLL EKFKRHNERNPYIEFPAVME PAFIIHHYAGKVKYGVKDFR
Xetr_Myo15          NKIIFKEEQEEYIREQIDWR EISFSDNQACIDIISQKPHG ILRILDDQSCFPQATDHTFL QKCHYHHGTNDLYCKPKMPL PEFGIKHFAGKVTYQVHKFL
Xetr_Myo35          KQRLVIQEQAEYAREGLRWV PLNQESVGPCLGLLTEPPHG IFHMLEEQSSLAQATDHTFL QKCHYHHGNSGCYIKPKLPL PVFTLQHYAGPVTYQVHNFL
Dapu_Myo15          NKHIFKLEQQEYAKERIEWQ TITFTDNQPVISLIAKKPIG VLHLLDDESNFPKATDGSFL EKCHYNHALNELYFRPRMSS MEFGIKHYAGQVWYNADGFL
Leed_CS1            TRQTFEDTSNHIVSDGVSIP AISTMDNTACVEMLQRKPGG LLSLMNKASSAHKQRNEDLL QEMQAKFGVHASFVATSGNR MQFGVNHYAGVCMYDVSDFV
Cogr_CSc            LQNFFDRKADMLENEEVSVA ATSYFDNSDAVRGLLKPGNG LLSILDDQTRRNRS-DLQFL ESLRKFEGKNPAITPGSANA ASFTVKHFAGEVDYPVKGII
Dapu_Myo20          NTHIFKSSIESCREEGIQCE EPDYVDNVPCIDLISLRTG LLSMLDVECSA-RGTPESYV QKVRLQHKHNSRLFESKSVA SDFGIQHFAGRVVYDASDCL
Pd_Myo20            NDVIFRDPMVACMDEGIQAN DCSFFDNSPIVNVICAPQTG ILSLLEAECSLQRGNTESFL QKVRVQHSRSKHFYEAEFGN HHFGIIHYGSRIPYDAANFL
Dapu_Myo22          NHYVFALEQELYSTEEIQFS HIQFTDNSQCVELIEKPPLC ILKLLSEQCHMPKGCDLSYI TNLHAEFENHAYYIKRRKWE EVFGIRHYAGAVMYNVSGFV
Pd_Myo1             IQLTLKSEQEEYLSEGIEWE PVQYFNNKIICDLVEALPIG IISLLDEECLRPEANDRTFL EKLSKNLGGHPHFLSHSTGR EEFRLIHYAGEVTYNVQGFV
Xetr_Myo1A          IQMTLKEEQEEYTREGIKWT RIVYTDNEVICTLIEDNTNG ILAVLDEECLRPKVSDATFL AKLGTKFKNHKHFQ-SKVTD NCFRIEHYAGQVTYNTDGFL
Dapu_Myo1A          IELVLKQEQEEYRREGIAWQ NIDYFNNQIICDLVEQQHKG ILAIMDEACLNVRVTDQMLL EALDQKLSQHKHYTSRKLSH SDFRIRHYAGDVVYDVGGFL
Dapu_Myo19          TINYLKEQQVHLANEGLNSI DLDELDKHNLIS-FLDGPVS VFGVLNEESCYLNRGCDDSQV SSLNERTKGVRSMPPSPRMP VEFAISHYAGLVSYSTVAML
Xetr_Myo19          VSHYLKSQQDEYAAEGLEWS FISYQDNQSCVDLIEGSPIS IFSLLNEECRLNRSSDAGQL SRLEKALSHNKSIGRDKFSK PNFIVSHYAGKVQYQIEDMA
Pd_Myo19            -------------------- -------------------- -------------------- -------------------- --------------------
Dapu_Myo5           NQHVFKLEQEEYVREKIDWT FIDFYDNQPCIDLIE-KPLG ILDLLDEECRVPKGADNAWV EKLYTQCKKYEQFVKPRLSN TGFIIVHFADRVEYQCAGFV
Xetr_Myo5A          NLHVFKLEQEEYMKEQIPWT LIDFYDNQPCINLIEAK-MG ILDLLDEECKMPKGSDSTWA QKLYTHLKKSALFEKPRLSN VAFIIKHFADKVEYQCEGFL
Xetr_Myo6           NERILKEEQELYQKEGLGVN EVHYVDNQDCIDLIEAKLVG ILDILDEENRLPQPTDQHFT SAVHQKHKNHFRLTIPRKSD EGFIVRHFAGAVCYETTQFV
```

```
Pd_Myo6          NERILKDEQNLYEKEGLNVK RIDYVDNQDCIDLIEYKGVG IFDILDEENKLTKPNPEHFT SEVHKKTKNHFRIDLPRKSD EGFLIRHFAGAVCYRTALFI
Dapu_Myo6        NERILKEEQLLYDKEGLGVK KIAYVDNQDCIDLIEARTSG IFYLLDEESKLPKSNHTHFT STVHSANPNHFRLALPRKSD EGFLIRHFAGAVCYQTAKFI
Dapu_Myo18       HDTHITAQLNRYAQEHIDIT EEEMPSPEPLVDLLDRAPRG FLWLLDDEALVSTTSEDQLL DKLSAMYNDEKLFTRPEGTA GSLMIHHGQGTVLYDLDGWL
Xetr_Myo18A      HEKTFVQELERYKEENIELA AVDQASHQTLVRTLARTDRG LLWLLEEEALQPGGNEDTLL ERLFTYYGPQESEEPPLLQS SKFLLGHSYGTVEYDATGWL
Dapu_Mhc1        NHHMFVLEQEEYKKEGIDWV FMDFMDLQACIELME-KPMG VLSILEEESMFPKATDQTFA EKLNNHLGKSASFVKPKPAK AHFAIAHYAGTVPYNITGWL
Xetr_Mhc1        NHHMFVLEQEEYKKEGIDWE FIDFMDLAACIELIE-KPLG IFSILEEECMFPKATDTSFK NKLYQHLGKCKNFEKPKPGA AHFSLVHYAGTVDYNISGWL
Pd_Mhc1          NHHMFVLEQEEYKKEGIQWE FIDFMDLQQCIDLI-EKPMG ILSILEEECMFPKASDDSFK AKMYEHLGKSPNFGKPRPNE AHFELKHYAGTVGYNISGWL
Pd_Mhc2          -------------------- -------------------- -------------------- -------------------- --------------------

Pifu_CS          ERNRDTLSADLVGCLLNSNN EFIKDLFTSIAGNSSPTVTR HFKRSLTDLMTKLSQAQPLF VRCIKSNQHLAANKFDSELV RRQLLCNGLMEIAELRRDGY
Atri_CS          EGNRDSLSSDLVGCLLNSNN EFIKDLFTTLYARSSPTVTK HFKRSLSDLMTKLSQAQPLF VRCVKPNLHLSSGKFDSDLV RRQLLCNGLMEIAELRRDGY
Myga_CS          ERNRDNLSSDLVGCMLNSNN EFIKDLFTTINGRSSPTVTN HFKRSLSDLMTKLTQSQPLF VRCIKPNKTIAADKFETELV RRQMLCNGLMEIAELRRYGY
Logi_CS1         ERNRDKLSSDIVNCLVNCQN EFIRELLTSTSLRSKPTVTS HFKRSLADLISKLNRAQPLF VRCVKPNTTLAPNQFDTELV RRQLRCNGLMEITQLRRDGF
Logi_CS2         KKNRDTVSAQLSEIVQRSDN MFIRNYFQSRTGKLVGRVSD SFQKSLKRMMDKVQTLPPLF IRCIKPNPKLQPSKFDGELV RQQLERSGIMDAAKVRCSGY
Cate_CS1         EKNRDDMSANLQQCLEQSSM NLIKNLFKDKTKRWKQTLCR TFRNSLWELLSKLKETEPLF VRCIKPNLDQVPHKFAEGLV SKQMRYNGLLELARIRRQGL
Logi_CS4         EKNRDNLSICIVDCMKSSQS QLISDLFKDRSLKNQTTVAQ YFKNSLVDLMTRMNSSDPSF VRCIKPNDQKVPEKFQHDQV VNQLRSSGVFEVSRIRQTDY
Macr_CS          EKNRDRLGANLLEVMKKSSN ELVKELFTETLSRTTHTVAA TFRNSLSELMDKVNDSQPWF VRCIKPNHEKLSNKFQSSMV MSQLRYNGILEMTKIRKAGY
Logi_CS3         EKNRDSISQNVLDCLLKSSN DMVQMLFSGSEIQTKSFVSR YFRKSLDDLMTKISSVNPWF VRCVRPNSKNRANSFDAELV MNQLRCNGLLEITKIRRDGF
Logi_CS5         EKNADNLSSGIFSCIQDSEN DLISDLFKQKTSQDHKTQLS YFTESIKDLLRKMNSADPLF IRCIKPNNEQREKLFVDEKV NEQLSYLGIKEIAKIRKMCY
Miye_Myo         EKNRDMLSQDVTSCMRESDN PFVSDLFTVYNPKDHKTVIS YFQSSMNELLQKLQRADPYY VRCIKPNMFLKPDNFDDEKV LEQMLYNGISEVAKIRKLGL
Dapu_Myo21       NRNRDFLSAELINVMRKSTD SRLVNIFLYSQTRQMQTMAM KTRQSLIQLLKNLVHGTPHF IRCLRRSKSGAANVLDKAYL AEQIRFCNLVETIKIRQNGY
Dapu_Myo3        EKNRNFLPHEVVQLLRSSSN PTVQFLFNASQTRAQQTVST YFRHSLMELLQKMIGGEPHF IRCIKPNDSRRAGYVDRTKI AQQLRYTGVLETVRIRKQGF
Xetr_Myo3A       -------------------- --------ETTNMKTQTVSS YFRYSLMDLLSKMVAGQPHF VRCIKPNNDRLANKFDREKV LVQLRYTGILETARIRRLGY
Xetr_Myo16       EKNKDVLSQNLLFVMKTSEN LVINQMFQKLERGGPTTVSI QLRKSFTEIIGKVQSSTPHL MHCIKPNSTKLAETFDNFYV SAQLQYIGVLEMVKIIRYGY
Pd_Myo3          DKNRDTFPNSLVSTMRASEN PLISDLFQATTRKGPASVAY HFKNSLTELINKMKNSAPHF VRCIKPNGSSAPNYFNQEYI AEQLAYSGILETIRIRKHGF
Xetr_Myo28       EKNRDALPTNINRLFISSET SLISILFTEKGSSRKISVGA QFRHSLAVLMEKMYLAEPHF IRCIKPNSQNEPGLLETQTV LNQ-SCKRIAETAQ------
Dapu_Myo7        DKNRDSFSADLMQLVHVTSN KFLRTLFASETRKKAPTLSA QFKKSLDSLMRTLSACQPFF VRCIKPNELKQSMVFDRELC CRQLRYSGMMETIRIRRAGY
Xetr_Myo7A       EKNRDTLHGDIIQLVHSAKN KFIKQIFQAETRKRSPTLSS QFKRSLELLMRTLSVCQPFF VRCIKPNEYKKPMLFDRELC VRQLRYSGMMETIRIRRAGY
Xetr_Myo10A      EKNRDTFRDDLLNLLRESSL DFIYDLFESGNKHHKKPTVSL QFKDSLHSLMAKLSTSNPFF VRCIKPNIHKMPDQFDDTVV LNQLRYSGMLETVRIRKAGF
Dapu_Myo9        EKNLDLMKADIVAVLKSSQY AFVRELVGHKNVKKAQTVGA QFQHSLQSLMVALNTANPFF VRCIKSNRLKGPNQFDEEIV LRQLRYTGMLETVRIRQAGF
Pd_Myo9          -------------------- -------------------- -------------------- -------------------- --------------------
Xetr_Myo9A       EKNTDHMRPDIVALLRSSRN AFIRGMISHHKKKKPPSISA QFQTSLNKLMETLGQAEPYF VKCIRSNAEKVPMRFDDALV LRQLRYTGMLETVRIRQSGY
Xetr_Myo15       DKNYDQVRQDVLDLFVNSRI KVVANLFFVTRKYKASTVAA KFQQSLLELVEKMERCNPFF VRCIKPNSRKEPSLFESDIV TSQLKYSGILETIRIRKEGF
Xetr_Myo35       NKNRDRLPAAATEVLSQSRL RLLSELFTGGALAKELTLTG RFQASLQELTSCLERGRTLF VRCIAPNPKQLRGIFDVDFV SSQLRNSGILETVQLMKEGY
Dapu_Myo15       DKNRDTLRPDVVDLLISSRI QMLSRMFQVTMKPRTATVAA RFHDSLQALLESMSKCNPWF VRCLKPNREKLPMSFDLPVV LEQLRYTGMLETIRIRKNGY
Leed_CS1         EKDTDLLDPAFVPLLRNSSE SFVAKLFSGKSYPVTTQITS LFPSSLPTSPKPVSGHCLAF VRTT-PAHQILLTNADVEYI AD------------------
Cogr_CSc         EENGEVISGDLLNMINSTKS EFVARLFGSKLQASEQGASG QFLASLDNVTKSVASTNSYF VFCLKPNDRRIANQFDSKCV RTQVQTFGIAEISRLRSADF
Dapu_Myo20       ETNRDVLPDDLVSVFYKQSC QFATHLFGLNGDDPVSTLTQ DFHTRLDNLLRTLVHAKPHF IRCVRANNTETTGHFDRSVV SRQVRALQVLETVNLMAGGY
Pd_Myo20         GVNRDQIPDDVVSMFSKQSC NFISHLFNSVGTDQQESYTG DFRSRLDHLLQVLRNCKPHF IRCIKSNAGGAGEIFERELV VRQLRSLQLLETVHIMSGGF
Dapu_Myo22       DKNRDVHQEVFLDLLTNSQK ALVREFTSRGTNKAKPTKSD TFRIQLTQLVQSLQSTNPWY IRCIKPNTGKLANTFHEPMV LDQLRYLGMLDLIRIRREGY
Pd_Myo1          EKNNDRLYRDLKEVMCLTTN VIASSCFX------------ --------------XXEPSY VRCIKPNDSKQSGVFNDEIV KHQVKYLGLMENLRVKRAGF
Xetr_Myo1A       DKNNDLLFRDLSQAMWNAKH QLLKTLFPNSSLKRPPTVGF QFRNSVSTLMKNLYAKNPNY IRCLKPNSSKKPSLFDDVLV KTQILYLGLLENVKVRRAGF
Dapu_Myo1A       DKNKDTLFQDFKRLLYSSKD PNIRHMWPHKTTKRPLTAGT LFKNSMIALVKNLTSKEPFY VRCIKPNDQKSPVLFDEERV RHQVSYLGLIENVRVRRAGF
Dapu_Myo19       DKNRDQVPGEMMSLLSKSKS PFIKLLLGHTMNGRRHTVLA KFKNSLDNLMRILNACDVHY VRCIRPSTSSPSQRWDQDYV DAQLKACGVIDTIKVSSFGY
Xetr_Myo19       EKNKDPVPPELIQLLQESDD HLLQILFPTNNRAVGVTVVS KFKGSLESLMQILHSTTPHY IRCIKPNVDCQALVFRQEEV LMQLEACGIVETINISAAGF
Pd_Myo19         -------------------- -------------------- -------------------- -------------------- --------------------
Dapu_Myo5        EKNRDTVLEEQVQVLRSSSN GIVRQLIVTMTKQNRRTVGS QFRESLTLLMNTLNATTPHY VRCIKPNDSKESFVFEPRRA VQQLRACGVLETVRISAAGF
Xetr_Myo5A       EKNKDTVFEEQIKVLKASKF TLLTELFQQTSKEHKKTVGH QFRNSLHLLMETLNATTPHY VRCIKPNDFKYPFTFDSKRA VQQLRACGVLETIRISAAGF
Xetr_Myo6        EKNNDALHMSLESLICESKD KFVRELFEKTGKLSFISVGN KFKTQLNLLLEKLRSTGSSF IRCIKPNLKMTSHQFEGAQI LSQLQCSGMVSVLDLMQGGF
Pd_Myo6          EKNNDALHLSLETLVMESKN TLVMTIYQSMGKLGFISVGS KFRDQLTILLQKLRSTGTSF IRCIKPNLKMVDSLFEGTQI LTQLQCSGMTFVLELMQQGF
Dapu_Myo6        EKNNDALHASLECIIQESHN PFLKSLFVWKGKLSFISVGS KFRMQLNELMEKLRSTGTNF IRCVKPNLKMVDHLFEGGQI LSQLQCSGMTSVLELMQQGY
Dapu_Myo18       KSAREMTTRNAALALTESTK EEMASLFRGTAGIKRKSLAL QLKFQVDGLIETLRRTKPRF VYCFIPNHRAGDDSLQVPLL RSQLRGSELLPAIRLYRQGY
Xetr_Myo18A      SYVRQNASQNAAALLQDSQK KIISGLFAGVAAVKKKSLCI QIKLQVDALIDTIKRSKIHF VHCFLSKDARLLMQVDVPFL RAQIRGSRLLDALHMYRQGY
Dapu_Mhc1        EKNKDPLNDTVVDQFKKGSS KLVQEIFARTKGSGFQTVSA LYREQLNGLMKTLNATSPHF IRCIIPNETKSPGVIDSHLV MHQLTCNGVLEGIRICRKGF
Xetr_Mhc1        EKNKDPLNESVVQLYQKSSV KLLSLLYSKKKGSSFQTVSG LFRENLNKLMTNLRSTHPHF VRCLIPNETKTPGIMDNHLL IHQLRCNGVLEGIRICRKGF
Pd_Mhc1          DKNKDPINETVVTLLSQSKE PIVAHLFQKGKGSAYQTISA THKESLFRLMHNLHSTYPHF VRCLIPNEIKTAGVIDAMLV QNQLQCNGVLEGIRICRKGF
Pd_Mhc2          -------------------- -------------------- -------------------- -------------------- --------------------

Pifu_CS          PIRIKYEDFVERYRDISDFV NPKSDLGQTLVILKSL--NI EGFKAGRSKIFLKNWHGDIL ETKLR
Atri_CS          PIRIKFEDFAARYKDICDFG NTNSDLGKCLDILKTE--RI EGFKVGRSKIFLKDYQKDML EDTLR
Myga_CS          PVRVKFEDFAARYAMLCDSS DEMYGFGRCVDILKAA--HI EGAQFGKSKIFMKSWEKDLL EETLS
Logi_CS1         PIRIHYQDFSERYGELLNNL DLEDVGSRCQEICDKN--GI KDYQLGRTKVFLKLEHKELL DTVLK
Logi_CS2         PVRMFFKDFVKRYSILQNTE TPRSTKSDCENILNSL--KI FDYGIGKTKVFLKMKHRDKL EDQVV
Cate_CS1         PVRIKRNVFVDRYLCLVKRK HKHQENENCHQILSSV--NI TDAQIGRGKVFLKNWHKDVL DEQLK
Logi_CS4         PIRMAFGKFIDRYKIIGFRP NMYVEPINCVAILKTL--GL SDFETGQTKVFLRYSHLDRL NEYLQ
Macr_CS          PVRMLFKDFIYRYQLVGFES TAVVQVSNCEVILMMI--ST DGYCMGKSKVFMKDWFVEEL DARLQ
Logi_CS3         AIRIPVQDFVKSYREICYKA KEVLPWQAAQKIMQVA--KA KDYVFGHTKVFLRHWHGPLL NNTLE
Logi_CS5         PIRSQYTDFSQWYKDIVPTS EKSRSADQSKFILNKFIKSN KEYKFGKTQIFMMEAVKKTL DLILT
Miye_Myo         PIRKRYDDFTKRYRPLFLDC RKARSRAGAELLLKKTLPDM SGIQFGKTRVFMQEDVSIWL EKCRG
Dapu_Myo21       AHRIPFMEFLRRYQFLAFDY EETVDPDNCRLLLLRLKMPQ EGWMIGKTKIFLKYHVEEYL ARMYE
Dapu_Myo3        SHRLNFVDFLRRYCFLAFNF EERVVRETCQKLLVRLK--L DGWAVGKTKVFLKYYHVEYL SKVYD
Xetr_Myo3A       SHRILFANFIKRYRLLCFKS NEDPPPESCAAILEKAN--V DNWVLGKTKVFLKYYHAEHL DRMIK
Xetr_Myo16       PIRLAFSDFLSRYKNLVEIL AWEKQEEKCRHVLLQCK--Q QGWQMGTRKVFLKYWQADQL SDMCL
Pd_Myo3          PLRLTPEEFLERYELLSSCL IGSNAIAKCKDILGQC--QV QNYKIGKTRIFLR------- -----
Xetr_Myo28       -------------------- -------------------L IGWHCGKSRLFFKYWHQDQL QRRLV
Dapu_Myo7        PIRHTFREFVERYRFLISGC PPPHRRQATARICSATLG-K TDYQLGQTKVFLKDAQDLFL EQERD
Xetr_Myo7A       PIRYTFVEFVDRYRVLMPGV KPAYKRGTCERIAESVLGKD DDWQIGKTKIFLKDHHDMLL EIERD
Xetr_Myo10A      PVRRLFNDFCFRYKVLMKNL SLPEDKGKCEVLLHLHDNTD TEWQIGKTKVFLRESLEHKL EKQRD
Dapu_Myo9        NVRLTFEEFIHHYRILLPRG LLSS-QSDVHDFLCRMNLNA EHYQIGHQKIFLRESEKAKL DYRLH
Pd_Myo9          -------------------- -------------------- -------------------- -----
Xetr_Myo9A       SCKYPFQDFVTQFSVLLPKD ISPV-HQSIKEFFWRIGHSP DQYQVGRTMVFLKEQERQCL MEMLH
Xetr_Myo15       PVRIPFHFFLSRYRCVMDLG PNVRVGQICVTVLKKLCPSP SMYCIGVSKLFMKENVYQQL ESKRD
Xetr_Myo35       PIRIPLKDFVARYGYLVRSS ASFANQECCTAVLTKAVGDS KLYQIGLSKVFLKEEGKEML QGYLD
Dapu_Myo15       PVRMKFHHFIERYRCLLTRR ERRNLPDICRIVLDRHAQ-G DQFQLGTTKVFMREALEQQI ERKRL
Leed_CS1         ---FDLAQFCGRYVPTMRGS ESERICARANGWIEGV---- -DYVVGHRSIWLSYSAWKMV EDTVR
Cogr_CSc         SLFLPFGEFLG----LADSD TILVGRERVELVVEDKHWPS NEVRIGSTGVFLSERCWMEI AQLSD
Dapu_Myo20       PHRMRFKAFNARYKMLAPFK RLKRVLDDCSLILQSFQANS TSWSLGKRHIFLSEGARQQL EALRA
Pd_Myo20         PHKIRYKTFIARYK------ -------------------- -------------------- -----
Dapu_Myo22       PVHIDFDSFISSYKCLCKGT HFPVGKDIVHSILKYLKYPT DQWQVGKRKVFLRLQVYEPL ESSRI
Pd_Myo1          AYRGEYAQFLKRYKPLCPDT WPHFNARGVATLVAHLQYGE DDYRLGRTKIFIR------- -----
Xetr_Myo1A       AYRQLYKPFLERYKMISKKT WPQWARDGVAAMMAELGIPD EEAALGHTKIFIRTPSTLFL EDRRR
Dapu_Myo1A       AYRQSYDRFLRRYKMVSQYT WPNFRRDGVRILIQQQNF-D EDVVYGKTKIFVRSPQTLFL EQARA
Dapu_Myo19       PIRMSHEEFATRYQYLSSSR RDRNLKEICRRISENYPQMA SEVRVGNSQMYATEKGIEMA ERWKL
Xetr_Myo19       PIRISFESFVERYSVIAPRQ IHMENPSKLQTILHAVLSKN TLVHCGTTKVFLTHLMVELL EERRL
Pd_Myo19         -------------------- -------------------- -------------------- -----
```

```
Dapu_Myo5            PSRLTYEEFIVRYRVLFHSR QCQRKRESCETVLATLITED DKFKFGASKIFFRAGQVAYL EKRRT
Xetr_Myo5A           PSRWTYQEFFSRYRVLMKQK DVLSDKQTCKNVLEKLILDK DKYQFGKTKIFFRAGQVAYL EKIRA
Xetr_Myo6            PSRASFHELYNMYKKYMPAK MSRLDRLFCKALFKALGLNE NDYKFGLTKVFFRPGKFAEF DQIMK
Pd_Myo6              PSRTQFSELYNMYKQYMPPA IARLDRLFCKALFKALGLNE NDFKFGMTKVFFR------- -----
Dapu_Myo6            PSRAPFVDLYNIYKSILPPQ LARLDRLFCRSLFQALSLND QDFKFGLTKVFFRPGKFAEF DTLMR
Dapu_Myo18           PDYMPLGEFVRRFNVLVPPD ALPASKQAAEILLLHIDMER SSYRLGLSQVFFRPGVLSQL EDQRD
Xetr_Myo18A          PDHMVFCEFRRRFDILVPHL TKKHG--------------R NYIVTDEKRVFFRAGTITKL EEQRD
Dapu_Mhc1            PNRMVYPDFKHRYMILAPNE MKAEPRKAAKICLEKIALDP EWYRIGHTKVFFKAGVLGQL EEMRD
Xetr_Mhc1            PSRILYGDFKQRYKVLNASA IPEGQKKACEKLLGSIDVDH TQYKFGHTKVFFKAGLLGTL EEMRD
Pd_Mhc1              PNRMIYSEFKQRYSILAPNA IPQSGDGATEKVLGALDLG- DNYRLGHTKVLFKGWY---- -----
Pd_Mhc2              -------------------- -------------------- -------------------- -----
```

Alignment for evolutionary analysis of MLCK homologs

```
Pd_MLCK1                 -------------------- -----GTPEPTITWFFNKKL LKSSNDITQTYNGRRAKLTF VEVFPDDKGVYECVATNTGG EVKSSCTLTVEGRNTGTIRN
Pd_MLCK2                 -------------------- -----GSPEPTVTWYFGNRK LPNCGEFEQTYNGNVAKMVI AEIFLDDEGEYSCVAENVSG RVKSSCWIAVE---------
Cate_MLCK                -------------------- -------------------- -------------------- -------------------- --------------------
Macr_MLCK                -------------------- -------------------- -------------------- -------------------- --------------------
Logi_MLCK                PQFIIKPRRQLVDEGDSTKF KASFEGSSNTQVTWSKDGKE ILPDDHCKMFVKDDFHYLEI CKVLPEHGGSYTCVLSNPAG STSATVGLEV----FESYDA
Myci_MLCK                PTFVTQLEGAHVAEGQRFTF ECRIECEVTPEVSWFKDDVP LSP-PAYVMSFKDGVALLTI EETFGEDTALYTCKASTRDG EATTSAYLKV----------
Hero_MLCK                PIFTEALRDQQVKEGTKVVL SVKFSGEPAPQIKWLRNDGQ IIPSSSFKIVVDHGYSALEI KEFFPEDAGLYTVVARNLGG ETRTSCHLDIEGLSLSSGIP
Strongylocentrotus_MLCK2 PFFSLPLRRKTVYDGEPARI SCNVQGTPKPKISWSKNGVP IYDGTFYNITYRWGVAMVEI RHASKSDTGTFRCTAKNAEG EATSACELVVEQRVPRSE--
Strongylocentrotus_MLCK1 PEFLALPTQRTITEPNSAKF NCRVVGDPKPTVTWSRNGVV LSDSGRYELYEEKDEFVLEI FDTTTDDSGVYVCTATNLAG QKAAETMLTVES--------
Drosophila_MLCK          -------------------- -------------------- -------------------- -------------------- --------------------
C.brenneri_MLCK2         -------------------- -------------------- -------------------- -------------------- --------------------
Ascaris_MLCK             -------------------- -------------------- -------------------- -------------------- --------------------
C.brenneri_MLCK1         -------------------- -------------------- -------------------- -------------------- --------------------
C.elegans_MLCK           -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_SHEEP               -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_BOVIN               TPKPVSNAKPAETLKPVGNA KPAETPKPLSNVKPAETPKL VGNAKPAETSKPLDNAKPAE APKPLGNAKPAEIPKPTGKE ELKKEIKNDVNCKKGHAGAT
MYLK_RABIT               TLKPIANTKPAETLKPVGNA KPAETLKPVGNAKPAETLKP VGNAKPAETLKPVGNAKPAE TLKAVANAKPAETPKPAGKE ELKKEVQNDVNCKREKAGAA

Pd_MLCK1                 GEDKPDVPQDVAPVFSKMLE NITVQDGTMVSLECNVTANP APKVIWYKAGVQVKANDD-F IMTYDGNLARLLIQEVYPED EGEYTCVAENPAGDARSKCK
Pd_MLCK2                 -----DDTDDGTPSIDKPLK DLEAFDGDQLTLVCEVSGSP KPDLYWFHDEKRLQQSDD-F ELSYDGKVATLVIDEVFPED SGIYKLVVENSKGKTSTQCK
Cate_MLCK                ----------APQIIEEMS DISVRDGCSGALSCKLNATA DTEITWYRNDKVVKASRD-F KPSFDGFVAKLEITEVYPED AGVYVCKAVNSSGESSSTAQ
Macr_MLCK                -------------------- ----------SISCFL-GTP EPAIRWLKNGKEIRENKD-F HATYTDKMAKLVIQEVFPED SATFECIAKNSAGEIQSSAN
Logi_MLCK                FSEKAKPVKVAQPILVKPLE DQTVTAGDRVKLTCSASDAD AATVTWYKDGKELRQGRQ-H RIRFDGRVASLGLTQAASTD GGLYECVIKNNGHEVRTSAT
Myci_MLCK                ----------IPEFTRMLR SSEVPEGSHHVMECHVTGIP EPEITWLRDDSPLPTNDNVV SISREGNACVLRLRNLKRTQ SGRYTCLAQNEAGRASASAK
Hero_MLCK                GSQVP-----CKPKFVDFLQ NKDVQEGSRAHFKFHVSGKP VPEITWYKNGVEVRPDKN-H DIIKKEALVQLIIRHATQED TGRYVCVAKNEGGQVQSSAQ
Strongylocentrotus_MLCK2 ------TEDYIRPSFAKPLR RIRAKEGRKVRLEIQVMGWP EPKVTWLKDNVDVVHSDHMY VEKFPNGLQSLVIYRVVPTD AGRYQAIASSPAGKATCSAD
Strongylocentrotus_MLCK1 -----SAPKGTIPEFIEKMS GTAVRDADKAKFTCKLRGIP SPSVHWYYRNQEIVGSDEVY ELYHEGEIASLCLPEVLPED EGEYSCTIKNDMGETSCSAF
Drosophila_MLCK          -------------------- -------------------- -------------------- -------------------- --------------------
C.brenneri_MLCK2         -------------------- -------------------- -------------------- -------------------- --------------------
Ascaris_MLCK             -------------------- -------------------- -------------------- -------------------- --------------------
C.brenneri_MLCK1         -------------------- -------------------- -------------------- -------------------- --------------------
C.elegans_MLCK           -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_SHEEP               -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_BOVIN               DSEKRPESRGTAPTFEEKLQ DLHVAEGQKLLLQCRVSSDP PATITWTLNGKTLKTTKF-I VLSQEGSLCSVSIEKALPED RGLYKCVAKNSAGQAESSCQ
MYLK_RABIT               DNEKPPASPGTAPTFKEKLQ DVRVAEGEKLLLQCQVSSEP PATITWTLNGKTLKTTKF-V ILSQEGSLCSVSIEKALPED RGLYKCVAKNAAPEAECSCH

Pd_MLCK1                 LIVKVKPPTVIKKLKDMNVA DGERVELVVEVEGSEPIDVV WLHGEKEIKKTSAVYDLFDE GTTHRLVVPEVYPEDSGLYV CEIYNAFGDAETNMTLTIDD
Pd_MLCK2                 VTIHDVPPKSAQEVQDTEAT DGDSLTLSCKLMGGKGATIC WYKSEKVLKGVDFKQSF--D GQTAKLEIREAFPEDRGTYS CVAKLGSSEARTECKLTVTD
Cate_MLCK                VSVQG---------------- -------------------- -------------------- -------------------- --------------------
Macr_MLCK                LSVIA---------------- -------------------- -------------------- -------------------- --------------------
Logi_MLCK                LTVENEPPYFVQELEDMEVD DGDKVELVVEVKGSEPIDVF WIHNNKEVAKDDPDYQQAAN KNRFKLTIGKVVPGDAGAYV CEAYNEFGESDTFCNLSI--
Myci_MLCK                LSVHVQKPEMTSSLANFTAT EGQSVTLEVTFKADPTPQIT WFKDELPVVSTSG-LTITID NNKSQLTIHEALYTSS--YS VVLRNVGGEVRSSCLITV--
Hero_MLCK                LNVKAEPPRFLDRLQSTSVR EGDTVKLSVRVTGKPEPEVT WFRQGQQIMNTK-DFQISKS GDEHILLIPEVFGEDQGKIT VKATNSAGQSQCTADLNV--
Strongylocentrotus_MLCK2 VDVPAGPPSFDKKFQDMSIR DGERVVLEVVITGSRPMDVG WLRNGQDIVDCK-EFRYVSR GNIYQLEIAEIFPEDSGKYT CEALNDHGESECSAMITVNE
Strongylocentrotus_MLCK1 LKVQSTMPEFLQKLKDIQAI DGSPLSLPVRIKGSPAPDVQ WFFNKEQIKEDN-DFKFVVD GDRLSLVIAEVYPDDAGIYT CKIFNSAGSAECACKVFVQA
Drosophila_MLCK          -------------------- -------------------- -------------------- -------------------- --------------------
C.brenneri_MLCK2         -------------------- -------------------- -------------------- -------------------- --------------------
Ascaris_MLCK             -------------------- -------------------- -------------------- -------------------- --------------------
C.brenneri_MLCK1         -------------------- -------------------- -------------------- -------------------- --------------------
C.elegans_MLCK           -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_SHEEP               -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_BOVIN               VTVDDAPTSENAKAPEMKA- -------------------- -------------------- -------------------- --------------------
MYLK_RABIT               VTVHDAPASENAKAPEMKS- -------------------- -------------------- -------------------- --------------------

Pd_MLCK1                 REVQNNDRSHGHGKNGAEDL DGEDDEFYPPEFIRKPRSVN VEEGNAARFQCQIDAYPPPT VCWETDGKILENGGRYRISE DGEMFTFEIPHSLATDAGRY
Pd_MLCK2                 RSQESSSASSSDSSNSNADL DADDKSRTPPGGNKDKRDSS PPGG-----QAKIKRTWPPT -------------------- --------------------
Cate_MLCK                -------------------- -------------------- -------------------- -------------------- --------------------
Macr_MLCK                -------------------- -------------------- -------------------- -------------------- --------------------
Logi_MLCK                REIEKSL------------- --------PPDFITKPKLTK VTEGADAIFSCTVVGFPLPN VYWERNGKILQDCTKYQISQ VHDKHTLVISKATKDDGGKY
Myci_MLCK                -------------------- --------PPSFTGKLHNKA GLEGSRVRMDCVCVGHPEPE VIWYHNQSPVKESRDFQLLF EGDKCSLIIRELYLEDAGEY
Hero_MLCK                EALPNPF------------- --------PPEFSVPLSDQV IQEGNPVSFMTEASGFPYPQ YIWQKDGRALDDCSRYNVEQ --------------------
Strongylocentrotus_MLCK2 ESISQSMDVDVSLSTDMPCT ATMDEDHIPPEFFEEPQSQM VEEGTTAVFTCDVDGDPVPS LTWSKDGRPFTDGGRFSIRS DKFTHTLEIQNARSTDAGVY
Strongylocentrotus_MLCK1 S-------------------- ---DIEGYPPKFVQKPRSIH IDEGSSVTLSCKIDGEPLPT VTWIKDGRPIEAGRRFKMES SGTTRSLNIPTVLATDAGSY
Drosophila_MLCK          -------------------- -------------------- -------------------- -------------------- --------------------
C.brenneri_MLCK2         -------------------- -------------------- -------------------- -------------------- --------------------
Ascaris_MLCK             -------------------- -------------------- -------------------- -------------------- --------------------
C.brenneri_MLCK1         -------------------- -------------------- -------------------- -------------------- --------------------
C.elegans_MLCK           -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_SHEEP               -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_BOVIN               -------------------- -------------------- -------------------- -------------------- --------------------
MYLK_RABIT               -------------------- -------------------- -------------------- -------------------- --------------------

Pd_MLCK1                 VVIATNSLGTEKYSVSLMIK EESTEVTDFRGVLKDQVDFR KVLQPNTITSEADYKVQPDA EQIDFRNVLTRHVKTKERPT VKKELRDTQAREGEPIKLEC
Pd_MLCK2                 -------------------- -------------------- -------------------- -------------------- --------------------
Cate_MLCK                -------------------- -------------------- -------------------- -------------------- --------------------
Macr_MLCK                -------------------- -------------------- -------------------- -------------------- --------------------
```

```
Logi_MLCK                ICRLDNQMVSMVHSTSLIVQ EKSANTTDFRSVLKDSVDFR SVLKPKVPMNLPDCKLPPSG AQHDFRHLLSKKSEISTEPK LLQTLTNQTVKYGGQAVLFC
Myci_MLCK                ACVAKNTSGTAQTQCKLTVE PLS------------------ ---------------------- ---------------------- ----------------------
Hero_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK2 ACTARNSEGEVVCEFTLDIL PTGADDKGITASDVSSAAAR DSAESAKSSSAPSAKGPKDG ESTTGVSGSDTDSQEGGPPK FLRGLQEEYAWEEDEVVLEC
Strongylocentrotus_MLCK1 TCVLHNQSGGDQCSVNVVVK PLEEEQTDFRSLLKSRPRLQ NISNSPNNVS----SDKSDA EQVDFRHVLTRHVNTRKRPS FIKPLTDIEVVEGESVTFEC
Drosophila_MLCK          ------------------- ---------------------- ---------------------- ---------------------- ----------------------
C.brenneri_MLCK2         ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Ascaris_MLCK             ------------------- ---------------------- ---------------------- ---------------------- ----------------------
C.brenneri_MLCK1         ------------------- ---------------------- ---------------------- ---------------------- ----------------------
C.elegans_MLCK           ------------------- ---------------------- ---------------------- ---------------------- ----------------------
MYLK_SHEEP               ------------------- ---------------------- ---------------------- ---------------------- ----------------------
MYLK_BOVIN               ------------------- ---------------------- ---------------------- ---------------------- ----------------------
MYLK_RABIT               ------------------- ---------------------- ---------------------- ---------------------- ----------------------

Pd_MLCK1                 IIIGTPIPEIRWYHGKKEVK ESKYFKMSYERAVAVLYMAE VYKEDEGEYLCMAVNTAGTV HSSCDLKIVESVEDSPRVRR RKVDPSAMEPPLIEDLQXDN
Pd_MLCK2                 ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Cate_MLCK                ------------------- ---------------------- ---------------------- ---------------------- --------------DLSPEN
Macr_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Logi_MLCK                EVRGKPEPDIKWSVNQRDIQ QSRYLQLKYKDNIASLTINE AFSEDEGEYAVVASNCNGSI KTSCTLSIEDSSSMSSRSDS RSLEGPVAFPPKILNVTPTT
Myci_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Hero_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK2 HVQGVPMPTVKWFVHNKEIE SDKFVVITMDN--------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK1 HVDGIPEPIIIWTANKKEIK ESKYFQMSYKDTVAKLLIAE AFAEDEGDYACTATNGVGSV TCAAELTVQDCKAVNSCGTK ESTAEIIVEAPARIKDAPKT
Drosophila_MLCK          ------------------- ---------------------- ---------------------- ---------------------- ----------RSCILTRPED
C.brenneri_MLCK2         ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Ascaris_MLCK             ------------------- ---------------------- ---------------------- ---------------------- ----------------------
C.brenneri_MLCK1         ------------------- ---------------------- ---------------------- ---------------------- ----------------------
C.elegans_MLCK           ------------------- ---------------------- ---------------------- ---------------------- ----------------------
MYLK_SHEEP               ------------------- ---------------------- ---------------------- ---------------------- ----------------------
MYLK_BOVIN               ------------------- ---------------------- ---------------------- ---------------------- ----------------FPED
MYLK_RABIT               ------------------- ---------------------- ---------------------- ---------------------- ----------------FPED

Pd_MLCK1                 LHILRGTTAQLKAKFAGEPR PNVVWTHGKTTVSTGGRLKI DVVRDTTTLTIQDAQPDDSG IYTLSVDNELGCDTCSSSLT VEDKPNPPAGQPAASNITGT
Pd_MLCK2                 ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Cate_MLCK                IHIIRRQSFSLVASFVASPA PDVAWFRDRVQLADGGRIRI ENKKKQSVLHVDCAEGDDTG CYTLQLTNELGSDRVSSSVT VEDKPEPPVGQPAISNITGT
Macr_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Logi_MLCK                VSATKNKSTEIKAAFNGEPT PSISWYKYKHELESEGRVHI KTLTNSSVLTIDDLQQADSG KYTIIAENDLGYDKAKIFIS VEDVPYSPMGCPSVSEITSN
Myci_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Hero_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK2 ------------------- ---------------------- ---DGVSRLTISEAFAEDSG VYRCEATNSSGTASTSGELM VAGKPKAPGGKPKVEQASKT
Strongylocentrotus_MLCK1 LQVMAGHDLSVTCNFVAIPE PDVVWRKNRKTVQESDRVKI EETDTSTTLTVTKATHDDSG RYTLQVENDLGSDDTAIYVS VLDKPDPPVGTPIASNISRQ
Drosophila_MLCK          CTALIGGHVRLSVRYEPFPG TKIIWYKACHPIVESSNVTI RTTSQQSTLYITDISADDSG KYTVEVMNDYGVEAAAASVA VEGPPEPPSGQPSVS-LGPD
C.brenneri_MLCK2         ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Ascaris_MLCK             ------------------- ---------------------- ---------------------- ---------------------- ----------------------
C.brenneri_MLCK1         ------------------- ---------------------- ---------------------- ---------------------- ----------------------
C.elegans_MLCK           ------------------- ---------------------- ---------------------- ---------------------- ----------------------
MYLK_SHEEP               ------------------- ---------------------- ---------------------- ---------------------- ----------------------
MYLK_BOVIN               QKVRAGESVELFGKVAGTQP ITCTWMKFRKQIQDSEHIKV ENSEQGSKLTIRAARQEHCG CYTLLVENKLGSRQAQVNLT VVDKPDPPAGTPCASDIRSS
MYLK_RABIT               QKVRAGERVELFGKVAGTQP ITCTWMKFRKQIQDSEHIKV ENSEAGSKLTILAARQEHCG CYTLLVENKLGSRQAQVNLT VVDKPDPPAGTPCASDIRSS

Pd_MLCK1                 SLTLSWYGPSFDGGSTITDY RVELMKVGEVDWKTLTSNCK YTSFQVRNLEKNTEYLFRVS AANKHGFSDPGKISEVI-VT IDRRRS-------STKLDEL
Pd_MLCK2                 ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Cate_MLCK                SLVLSWSGPSYDGGSSITGY KVEMAE-KSAEWQTISETCR NTTFEVSDLLPNKEYEFRVS VGNQHGFSDPGLPSEVA-IT IDGRNS-------SNSADEL
Macr_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Logi_MLCK                AVSLAWYGPAYDGGSPITAY TVECCKGNSRKWTSLTNNCR NTFYQAQSLEPNTQYCFRIR AGNKHGMSEASEPSELI-TT YEDCSS-------ESIEDEL
Myci_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Hero_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK2 SFKITWAAPAEDGGSPIFNY KVDGRKAKETGWTEV-ANVT QTSFEVTGIVPGVEYVYRVS AENEAGWSPSTAISDRAVIT ESESPP--------------
Strongylocentrotus_MLCK1 SLTLSWTGSSYDGGSRITGY VVEMRTVDDEAWTKA-DVVS HTSHTIDNLKPETKYLFRVS SENEHGISKPGQESDPI-LT VVEEKKKVDMRKPVTENDAE
Drosophila_MLCK          RVAVAWCGPPYDGGCMITGF IIEMQTIGDESWQQVTRVVD SLAYTVKNLQPERQYRFRVR AENIHGRSAPGQASELVQIT NTPQRS-------TSSDASD
C.brenneri_MLCK2         ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Ascaris_MLCK             ------------------- ---------------------- ---------------------- -VFQGTLSASTSADTSGRIV VDESDY-------RLEEDKL
C.brenneri_MLCK1         ------------------- ---------------------- ---------------------- ---------GTEPSGSGRIK VDEH-Y-------PSELDEP
C.elegans_MLCK           ------------------- ---------------------- ---------------------- ---------GTEPSGSGKIK VDEH-Y-------PSELDEP
MYLK_SHEEP               ------------------- ---------------------- ---------------------- ---------------------- ----------------------
MYLK_BOVIN               SLTLSWYGSSYDGGSAVQSY SVEIWDSVDKTWKEL-ATCR STSFNVQDLLPDREYKFRVR AINVYGTSEPSQESELTALG EKPEEEPKDEV-EVSDDDEK
MYLK_RABIT               SLTLSWYGSSYDGGSAVQSY SVEIWDSVDKMWTEL-ATCR STSFNVRDLLPDREYKFRVR AINVYGTSEPSQESELTTVG EKP-EEPKDEVEEVSDDDEK

Pd_MLCK1                 -AFEPRDV-TI-SDRVFKDN YVLSDEIGRGKFGSVCLCIE KSSGRELAVKFIRI-KATER DQAQQEVNIMNQLHHPKLLM LYDAFQTP-REVILVMEHIA
Pd_MLCK2                 ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Cate_MLCK                -AFEPRQV-ELDTTRNFTDF YTTGGEIGKGRFGSVFLCKE KQSNREFAAKFIKV-KMGQR DELRNEVMIMNALHHPKLLL LWDAFETR-REMVLIMEHVA
Macr_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Logi_MLCK                -PFEPKYV-DIDKSRKFEDY FDTMEEIGRGKFGNVFKCKD KKTKKIWAAKILKC-REKEK DNIRNEINVMNKLHHPKVLM LWDAFEAA-KSMVLVMEYIQ
Myci_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Hero_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK2 ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK1 -EVEFRHV-TIKPDKDFSDK YQIKEVLGKGRFGTVHKCIE KVTGKAYAAKMIKTIKSTDK ESVKNEIEIMNKLHHAKLLQ CLDAFESP-KQMIMVLEIVN
Drosophila_MLCK          -RFGQATV-SVQSGGDFKSR FEIIEELGKGRFGIVYKVQE RQPEQLLAAKVIKCIKSQDR QKVLEEISIMRALQHPKLLQ LAASFESP-REIVMVMEYIT
C.brenneri_MLCK2         ------------DDTPFEEV YDIETELGSGQFAVVRRVRD RKTGERYAAKFIKSRRGVTR QNIEREVRVLQKIRHSNVVE LHAVYETA-SDVIIVLELVS
Ascaris_MLCK             P-FEIKEVVRIRANEKFSKY YECFDEIGEGKFGKVYRCRE KATGLELAAKRIKIRRDADR AQVEKEVAIMTQMRHPRIAQ IYDAFSTPDNDIILVMEVVR
C.brenneri_MLCK1         PVFDVKKIENIRANVKFDTL YQVTKLLGDGKFGKVYCVIE KATGKEFAAKFIKIRKEADR TEVEREVSILTGLRHPRIAQ IYDAFYTTNDVVLIMEIVR
C.elegans_MLCK           PVFDVKKIENIRANVKFDTL YQVTKLLGDGKFGKVYCVIE KETGKEFAAKFIKIRKEADR AEVEREVSILTQLRHPRIAQ IYDAFYTTNDVVLIMEIVR
MYLK_SHEEP               ------------------- ---------------FRLVE KKTGKVWAGKFFKAYSAKEK ENIRQEISIMNCLHHPKLVQ CVDAFEEK-ANIVMVLEIVS
MYLK_BOVIN               PEVDYRTV-TVNTEQKVSDF YDIEERLGSGKFGQVFRLVE KKTGKIWAGKFFKAYSAKEK ENIRQEISIMNCLHHPKLVQ CVDAFEEK-ANIVMVLEIVS
MYLK_RABIT               PEVDYRTV-TVNTEQKVSDF YDIEERLGSGKFGQVFRLVE KKTGKIWAGKFFKAYSAKEK ENIPAEIGIMNCLHHPKLVQ CVDAFEEK-ANIVMVLEIVS

Pd_MLCK1                 GGELFERIVDDDYILTEREC VHFMRQIIDGVRFMHERNIL HLDLKPENILCVTRNSNEIK IIDFGLARKFDPKKNFRVMF GTPEFVAPEVINYDQISYAT
Pd_MLCK2                 ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Cate_MLCK                GGELFERIIDEDYILTERES IHFMRQIVSGVHYMHENNIL HLDLKPENILCISKNSNEIK IIDFGLARKYDPTKSAKVMF GTAEFVAPEVVNYDPISYTT
Macr_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Logi_MLCK                GGELFERVISDDFELTEGDV IHFMRQICDGLGYIHQQNVL HLDLKPENILCISSKTNQIK IIDFGLAQNFKRGESVKVLF GTPEFIAPEVVNYDEISFAT
Myci_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Hero_MLCK                ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK2 ------------------- ---------------------- ---------------------- ---------------------- ----------------------
Strongylocentrotus_MLCK1 GGELFERVIDDDFGLTESDV IEFMRQICAGVHHMHSTNIL HLDLKPENILCIDKTGSRIK LIDFGLARDFNPAQSTKVMF GTPEFVAPEVINYDVIGFTT
```

```
Drosophila_MLCK          GGELFERVVADDFTLTEMDC ILFLRQVCDGVAYMHGQSVV HLDLKPENIMCHTRTSHQIK IIDFGLAQRLDTKAPVRVLF GTPEFIPPEIISYEPIGFQS
C.brenneri_MLCK2         GGELFDHVCAKE-CLDEVEA AAFIKQILLAVRHLHSLHVV HLDIKPENVMLKQRGESHIK IIDFGLSREIEPGATVKDMV GTPEFVAPEVVNYEALSPAT
Ascaris_MLCK             GGELFDRVVDDNYILTEMAV VMIVCQLCEAVSYIHSKNIV HLDIKPENIMCVSQTGNRIK LIDFGLAQYYDGSSNLLFMA GTPEFVAPEVIKFEPIDFHT
C.brenneri_MLCK1         GGELFDRVAEESYELSELAV VMIICQLCEAIDYIHKQNIL HLDVKPENIMCVSLTGNRIK LIDFGLARHYDGTQELKYMA GTPEFAAPEVIKFEKLDYHT
C.elegans_MLCK           GGELFDRVAEESYVLSELAV VMIICQLCEAIDYIHKQNIL HLDVKPENIMCVSLTGNRIK LIDFGLARHYDGTQELKYMA GTPEFAAPEVIKFEKLDYHT
MYLK_SHEEP               GGELFERIIDEDFELTEREC IKYMKQISEGVEYIHKQGIV HLDLKPENIMCVNKTGTRIK LIDFGLARRLENAGSLKVLF GTPEFVAPEVINYEPIGYAT
MYLK_BOVIN               GGELFERIIDEDFELTEREC IKYMKQISEGVEYIHKQGIV HLDLKPENIMCVNKTGTRIK LIDFGLARRLENAGSLKVLF GTPEFVAPEVINYEPIGYAT
MYLK_RABIT               GGELFERIIDEDFELTEREC IKYMRQISEGVEYIHKQGIV HLDLKPENIMCVNKTGTRIK LIDFGLARRLENAGSLKVLF GTPEFVAPEVINYEPISYAT

Pd_MLCK1                 DMWSIGVICYIL-------- -------------------- -------------------- -----------------
Pd_MLCK2                 -------------------- -------------------- -------------------- -----------------
Cate_MLCK                DMWSVGVICYMLLSGLSPFM GENDAETFVNVTLTKWDFDD DIFDDISDDAKEFIETLLML DPRSRQSASEALQHKWL
Macr_MLCK                -------------------- -------------------- -------------------- -----------------
Logi_MLCK                DLWSVGVICYVLLSGLSPFL GDSENETLSNVTLGEYDFDD DAFTEISDNAKDFIQKLLIK KKQKRHTIDQCKNHQWL
Myci_MLCK                -------------------- -------------------- -------------------- -----------------
Hero_MLCK                -------------------- -------------------- -------------------- -----------------
Strongylocentrotus_MLCK2 -------------------- -------------------- -------------------- -----------------
Strongylocentrotus_MLCK1 DMWSVGVICYILLSGLSPFM GDNDAETLNNVTLAEWDFED EAFDAISEDAKTFIEGLLIQ KKEERMTAAECLQHHWL
Drosophila_MLCK          DMWSVGVICYVLLSGLSPFM GDTDVETFSNITRADYDYDD EAFDCVSQEAKDFISQLLVH RKEDRLTAQQCLASKWL
C.brenneri_MLCK2         DMWAVGVVTYILLSGGSPFL GDNRDETFSNITRVRYHFSD RYFKNTSKHAKDFISRLFVR DVDQRATVEECLQHPWI
Ascaris_MLCK             DMWSIGVITYILLSGISPFL GETLGDTYCAVEKGEWEFDE EAFEGISEAAKDFISKLLVY DQKKRMLPEQCLQHEWI
C.brenneri_MLCK1         DMWSIGVITYILLSGYSPFL GDNLGETYCNVEKGVWEFTE E-FDTVSEEAKDFVTKLLIY DQSKRMLPHECLQHPWI
C.elegans_MLCK           DMWSIGVITYILLSGYSPFL GDNLGETYCNVEKGVWEFTE E-FDTVTEEAKDFVTKLLVY DQSKRMLPHECLQHPWI
MYLK_SHEEP               DMWSIGVICYILVSGLSPFM GDNDNETLANVTSATWDFDD EAFDEISDDAKDFISNLLKK DIKNRLNCTQCLQHPWL
MYLK_BOVIN               DMWSIGVICYILVSGLSPFM GDNDNETLANVTSATWDFDD EAFDEISDDAKDFISNLLKK DMKNRLNCTQCLQHPWL
MYLK_RABIT               DMWSIGVICYILVSGLSPFM GDNDNETLANVTSATWDFDD EAFDEISDDAKDFISNLLKK DMKNRLDCTQCLQHPWL
```

Alignment for evolutionary analysis of SCA2 homologs

```
Pd_SCA2L                 LAVLAACIAVSHGLRCYVC- -NETSLAAGDTCADPFDTSS DAAKGYVE--LCTAPE---D KFCKKEKTGS---SKIAQYV VRSCA-----NQCSDLNLG-
Perinereis_SCA2L         LAFLAACIAVSHGLKCYVC- -NETALAKGDTCADPFD--X XXRQGFLQ--TCVAPA---D KFCMKEKIGS---SKIVQNV VRSCA-----NQCEDYNAG-
Drosophila_SCA2L1        IVLCSGSLGFADGLKCHMCG QYNEAVGSITPCTNYTTDIA HLYLK-----ECTKKS---E KFCVKYVSEL--------ST VRDCA-----TECVEKEIW-
Drosophila_SCA2L2        IILCSGNFGFADGLKCHMCG QYNEGVGSITPCTNYTKDIA HLYLK-----ECTKKS---E KYCVKYVSEL--------ST VRDCA-----TECVEKEIW-
Haematobia_SCA2L         LTVLVCGLVSVNGLKCHMCG QYNEGVGSITPCLNYSDQYA HLYLK-----ECSKKS---E KYCVKYVSEL--------ST VRDCA-----TECAEKEIW-
Schistosoma_SCA2L        LIFNGINCIKNKKVKCYRC- ---------SDCPNPFDKTQ ITELG-----NCN------- -FCRTVYTYR---DEDNYRI AKDCV-----ASCVPQDRRG
Alvinella_SCA2L          LVVFATVLTTTYGLKCYVC- NDIDNIVENSECGDTFEKDA HQNLLK----ECNASAG--E VWCKKQKTGI---NNLVTRI ERSCA-----AQCEDQTGAY
Myci_SCA2L               -------------------- -------------------- -------------------- WFCTKVKYGY---KHVFQHV DRDCS-----ESCTERNWF-
Saccoglossus_SCA2L       VVVGSGTAATPTNLTCYAC- ---MYISEREACIYNVTTME TR--------LCRKH----E KFCEVERFIE---AGELITF TRRCS-----SSCEEGCVTA
Cate_SCA2L               FLFVLAAASDVHGLRCYTC- -LHNQLQENDACVNNATLVD VT--------TCLSN----Q KYCMVEEIVIPTERGTFGKF HRKCS-----SECSSECSEA
Patiria_SCA2L            TGILLASVGYLFATECHFC- -SYVSGTMGEECKDPFTSNG NSSE------TCQGT----- -YCLKVVSKV---SGELTSI TRSCS-----EACSEACLSL
Paracentrotus_SCA2L      LFVLAACVAPLYALDCYVC- -GYVSGVGGEVCKDIFDTAT LNSSSEVTTGSCSGS----- --CQKTVVTV---SGEVTAT ARACS-----SSCTAGCISL
Strongylocentrotus_SCA2L VTVLISMVAGTWAIECYTC- -------TGAGCTDPFDSSG SGVSN-----TCPGSTLITY NYCAKAVSGS--------TV VRTCA-----PACTEACTG-
Asterina_SCA2L           LLTLSACVAGSYAVQCYECN GAIQPFKSNDWCNDPFDADS AGAKASIK--SCNGP----- --CVKLYSKI---GSSIEGY VRQCSNDTDQSLCINDCGSV
Crassostrea_SCA2L        LVLPGLLRILVSPLRCYEC- -DLIGNASMYSCFDPLDTTH DDVRVT----QCQDP----R ETCLKVKTRF---EDGDVFY QRKCGKAKHKDSCKDLHLN-
Xenopus_SCA2L            LLVAALCIGTAVPLQCYTC- ---LSATSSADCMNPT---- ----------TCSST----D TSCTKIVGSA----LSSTTV VKSCA-----PSCSPGSSS-
Notophthalmus_SCA2L      ---AITAGGAAHSLRCFSC- ---VGQSTSEACTTRT---- ----------DCAPA----A EYCAKVKVTA----SGTSLI SKDCA-----LICNEGSVSV
Gallus_SCA2L1            LLVAVLCAEQASSLFCYTC- ---ENEHSNWNCLKTY---- ----------KCEDH----E KYCTTTYSAAGFGKDMGYRI TKKCS-----ADCPETNTDF
Gallus_SCA2L2            VLVIALCTESAFSLRCFSC- ---KDAPSNIHCFSTT---- ----------TCADH----E KYCLTTYSTKGLGDDRQQSI TKKCA-----AFCPSIDLNI
Gallus_SCA2              VLAAVLCVERAHTLICFSC- ---SDASSNWACLTPV---- ----------KCAEN----E EHCVTTYVGVGIGGKSGQSI SKGCS-----PVCPSAGINL

Pd_SCA2L                 ------VHTVDCCDTDNCN- -GASSVWSFSLLTLLAFVSF LAYFNK
Perinereis_SCA2L         ------FVRKDCCEKDGCN- -TATSLWSLSPLLIIAFASL IAYFK-
Drosophila_SCA2L1        ------ETQTYCCTEDGCN- -SGTQLAYSVILLTLPILSV AWPTFV
Drosophila_SCA2L2        ------ETQTYCCTEDGCN- -SATQFAYSTLLLILPLMSI IWPTLV
Haematobia_SCA2L         ------ETQTYCCTEDGCN- -GASAHDISSLVLMVPLMVY ICQAIW
Schistosoma_SCA2L        GKA---GLVTECCDEDYCN- -ASPKHYSISFLLITSFTIF ITYTNK
Alvinella_SCA2L          ------VDRVDCCSEDGCN- -AGDHVTPVLNVLAICYMLF AIVNRL
Myci_SCA2L               ------IYASSCCQTEFCN- -ISSVLKQSEILFITSAIII LIVFKK
Saccoglossus_SCA2L       GSAHKYESCVSCCSENLCN- -VDNKAGILFSLRNLFELTV VAYFVH
Cate_SCA2L               DGN---KRCISCCQQDLCN- -GCGSVMERLLRLTVPLILA SNVVTL
Patiria_SCA2L            FGI---EGCTYCCQSDNCX- -GAGSVTFSLSVMVAMVTAV LKLSAL
Paracentrotus_SCA2L      FGI---KTCTNCCDTEYCN- -GATNVNMSIVSVLAAFLAA LVFVR-
Strongylocentrotus_SCA2L ------DTCLYCCNSNLCN- -GGNTVTTSFVAMGVAMLAA FILARQ
Asterina_SCA2L           GDI---YSCGYCCTGDLCN- -SAASVTFNLLAAVAMLVSA WVFTS-
Crassostrea_SCA2L        GTRYMLDVEVCSCQTDLCN- -TADSISASSSRTTAFITFL WAFL--
Xenopus_SCA2L            ------VATIFCCNTDLCN- -GSTAVKYSFPAIVLSLGFL LVLLRS
Notophthalmus_SCA2L      AGA---THTTSCCKTDLCN- -GASGARISYTLLSLAAAIS ALVVRA
Gallus_SCA2L1            GVV---AISTKCCKTSLCNF SGANSVKLSYAVMFLGTVAS LICVIR
Gallus_SCA2L2            GIA---GLATSCCQTSLCNI SGASSVKTSYTVIAVGVLAS LACILR
Gallus_SCA2              GIA---AASVYCCDSFLCNI SGSSSVKASYAVLALGILVS FVYVLR
```

7 Summary

Annelida are a well-known group of segmented, worm-like organisms, whose most characteristic features are prominent, often bizarrely formed chaetae. These structures have long been discussed as being an autapomorphy of the annelid ancestor, as similar structures of other taxa, such as Myzostomida and Brachiopoda, were interpreted to be of independent evolutionary origin. The process of chaetae formation (chaetogenesis) has long been studied on the ultrastructural level, but many aspects remained poorly understood, as data from other levels of investigation are missing. Thus, the aim of this study is to characterize chaetae-specific cells on a novel level using a combined strategy of structural and molecular methods. In the process of the *Platynereis dumerilii* transcriptome and genome project, random *in situ*-hybridization screening revealed a set of marker candidates that provide the first-time opportunity to gain new insights into chaetogenesis. The cellular localization of these markers was investigated using ultrastructural, immunhistochemical and various *in situ*-hybridization methods, revealing fifteen molecular markers that specifically characterize two different cell types that are involved in chaetogenesis, namely the chaetoblast and the proximal follicle cells. Furthermore, this set of highly specific markers enabled the first comparative analyses of chaetae-producing cells of the annelid *Capitella teleta*, the myzostomid *Myzostoma cirriferum* and the brachiopod *Macandrevia cranium*. For some of the *P. dumerilii* markers, orthologous sequences were found in the other study organisms where their expression is localized in homologous cell types, thus providing the first molecular evidence for homology of chaetae of Annelida, Myzostomida and Brachiopoda. Computational analyses of the marker sequences reveal a variety of amino acid motifs and domains that might indicate potential functions. Among these, the chitin synthases linked to a myosin motor domain (MMD) appear to be of special interest, as this study shows that the sampled lophotrochozoans possess an unexpected diversity of these enzymes, of which representatives of two chitin synthase groups appear to have a chaetoblast-specific expression in the studied organisms. As evolutionary analyses of chitin synthases indicate that the last common ancestor of Annelida, Myzostomida, Brachiopoda and Mollusca featured three MMD-linked chitin synthases, it is hypothesized that the ancestral function of these enzymes is the formation of homologous chaetae and other chitinous structures, potentially even shells. As the phylogenetic relationships among Lophotrochozoa are far from being solved, the investigation of all remaining lophotrochozoan taxa, using the combined methodological procedures established in this study, is necessary to infer the exact evolutionary origin of the homologous chaetae of Annelida, Myzostomida and Brachiopoda.

8 Zusammenfassung

Annelida sind eine gut bekannte Gruppe von segmentierten, wurmförmigen Organismen, deren charakteristischste Merkmale ausgeprägte und oft bizarr gestaltete Borsten (Chaetae) sind. Diese Strukturen wurden seit langem als Autapomorphie der Anneliden-Stammart diskutiert, da ähnliche Strukturen anderer Taxa, z.B. Myzostomida und Brachiopoda, als Resultat eines unabhängigen evolutiven Ursprungs interpretiert wurden. Der Prozess der Borstenbildung (Chaetogenese) wurde seit langem auf ultrastruktureller Ebene untersucht, aber viele Aspekte blieben unklar, da Daten anderer Vergleichsebenen fehlen. Das Ziel dieser Studie ist es daher, Borsten-spezifische Zellen auf einer neuen Ebene zu untersuchen, und zwar durch eine kombinierte Strategie aus strukturellen und molekularen Methoden. Im Zuge des Transkriptom- und Genom-Projektes von *Platynereis dumerilii* wurde eine Reihe an Marker-Kandidaten durch zufällige *in-situ*-Hybridisierung identifiziert, die die erstmalige Möglichkeit bieten, neue Erkenntnisse über die Chaetogenese zu gewinnen. Die zelluläre Lokalisierung dieser Marker erfolgte durch ultrastrukturelle, immunhistochemische, sowie verschiedenste Verfahren der *in-situ*-Hybridisierung und lieferte fünfzehn molekulare Marker, die zwei verschiedene in der Chaetogenese involvierte Zelltypen spezifisch charakterisieren, nämlich den Chaetoblasten und die proximalen Follikelzellen. Außerdem ermöglicht dieses Set von hochspezifischen Markern die ersten vergleichenden Analysen von Borsten-bildenden Zellen des Anneliden *Capitella teleta*, des Myzostomiden *Myzostoma cirriferum* und des Brachiopoden *Macandrevia cranium*. Im Fall mehrerer Marker aus *P. dumerilii* wurden orthologe Sequenzen in den anderen untersuchten Organismen gefunden, wo deren Expression in homologen Zelltypen lokalisiert ist und somit die ersten molekularen Indizien für die Homologie der Borsten der Annelida, Myzostomida und Brachiopoda liefert. Computer-gestützte Analysen der Markersequenzen zeigen eine Reihe an Aminosäure-Motiven und -Domänen, die auf potentielle Funktionen hindeuten könnten. Unter diesen erscheinen vor allem die mit einer Myosin-Motordomäne (MMD) verbundenen Chitinsynthasen von besonderem Interesse, da diese Studie eine unerwartete Diversität dieser Enzyme in den verfügbaren Lophotrochozoen nachweist, wobei die Vertreter von zwei Chitinsynthase-Gruppen anscheinend eine Chaetoblasten-spezifische Expression in den untersuchten Organismen besitzen. Da die evolutiven Analysen der Chitinsynthasen nahelegen, dass der letzte gemeinsame Vorfahr von Annelida, Myzostomida, Brachiopoda und Mollusca drei MMD-gekoppelte Chitinsynthasen besaß, wird hypothetisiert, dass die ursprüngliche Funktion dieser Enzyme die Bildung homologer Borsten und anderer chitinöser Strukturen ist, möglicherweise sogar die Bildung von Schalen. Da die phylogenetischen Beziehungen innerhalb der Lophotrochozoa noch nicht einmal annähernd geklärt sind, ist die Untersuchung aller verbleibenden Subtaxa der Lophotrochozoa unter Verwendung der in dieser Studie etablierten kombinierten Methodik nötig, um den genauen evolutiven Ursprung der homologen Borsten von Annelida, Myzostomida und Brachiopoda abzuleiten.

Don’t waste your time or time will waste you!

Matthew Bellamy (MUSE)