Investigating the molecular basis of notochord loss in *Molgula occulta* via transcriptome sequencing

Kristin Andrykovich[1, 2], Billie J. Swalla[1, 3]

BEACON Research Fellowship 2011
Summer 2011

[1]Friday Harbor Laboratories, University of Washington, Friday Harbor, WA 98250
[2]Department of Biology, Carleton College, Northfield, MN 55057
[3]Department of Biology, University of Washington, Seattle, WA 98195

Contact Information:
Kristin Andrykovich
Biology Department
Carleton College
300 N. College St
Northfield, MN 55057
andrykok@carleton.edu

Introduction

Ascidian tunicates, commonly known as 'sea squirts', are sessile hermaphroditic filter-feeders found in marine environments (Huber *et al.*, 2000). Tunicates are *uro*chordates (i.e. 'tail' chordates) because they possess the defining chordate structure, the notochord, in a larval tail that is reabsorbed upon metamorphosis. This tail is believed to benefit progeny by facilitating motility, which in turn permits a wider variety of anchoring environments and a higher probability of finding favorable settling conditions (Huber *et al.*, 2000). Despite the utility of this adaptation the larval notochord has been independently lost several times among the ascidians (Huber *et al.*, 2000). These anural, or tailless, tunicates occur with the highest frequency in the Molgulidae family (Huber *et al.*, 2000).

This study uses a newly sequenced transcriptome of closely related molgulid ascidians to evaluate notochord-specific gene expression in the anural tunicate *Molgula occulta*, the tailed tunicate *Molgula oculata* and *M. occulta* X *M. oculata* hybrids. Highly conserved genes expressed in the notochords of the urodele species *Ciona intestinalis* and *Molgula tectiformis* were examined for expression in *M. occulta*, *M. oculata* and *M. occulta* x *M. oculata* hybrids via transcriptome analysis. The results of this study may further elucidate the expression patterns essential to notochord development and implicate the nature of the genes involved in the evolution of taillessness. This study may also facilitate creating a clearer tunicate phylogeny through phylogenetic analyses that assesses the conservation of these key developmental genes in a Molgulid genome.

One compelling gene identified in the *C. intestinalis* and *M. tectiformis* transcriptomes is *prickle*, a gene family that plays a pivotal role in the planar cell-polarity

(PCP) pathway of Wnt-signaling. Studies in mice and *Drosophila* have shown that the PCP pathway is involved in coordinating cell movements during development in both vertebrates and invertebrates (Tao *et al.*, 2009). Within *C. intestinalis* and *C. savignyi*, *prickle* is necessary for notochord intercalation and is most highly expressed in notochord cells (Kulger *et al.*, 2011). Knock-out studies have also shown that prickle is essential in mediolaterial (M/L) polarization during notochord morphogenesis; *prickle* is expressed in notochord cell membranes, where cells detect interactions with neighboring tissues (Jiang *et al.*, 2005). As *C. intestinalis* contains two *prickle* genes— *prickle1* (pk1) and *prickle2* (pk2)—this study will search for expression of either *prickle* in *M. occulta* and *M. oculata*.

Other notochord-localized genes identified in the *C. intestinalis* transcriptome are *noto6* and *noto17*. These genes are enigmatic in that little is known of their products aside from their presence in urochordate notochords. Studies in the larvacean *Oikopleura dioica*, a member of another class of urochordata, showed diffuse expression of *noto17* in the notochord (Kulger *et al.*, 2011). *Noto6* has been detected in *Molgula tectiformis*, an anural ascidian, (Gyoja *et al.*, 2007) and a homologue has been isolated from human neural tissue, though its function has yet to be described (Hotta et al., 2000). The implications of these results are still under investigation, but the variety of tunicates expressing these genes suggest that they may allow an interesting analysis of phylogeny.

In *C. intestinalis*, *leprecan* is a highly conserved, notochord-localized transcriptional target of Brachyury, a zinc-finger transcription factor that is essential in ascidian notochord development (Di Gregorio *et al.*, 2002). *Ci-leprecan* encodes collagen-producing prolyl-3 hydroxylases and is regulated by *Ci-Bra* via a compact *cis*-regulatory

module (CRM) (Dunn and Di Gregorio, 2009). Dunn and Di Gregorio found that *C. intestinalis* with selectively knocked-out *Ci-leprecan* develop a suboptimal notochord phenotype, which suggests *Ci-leprecan* is essential in notochord development. Another study in *C. intestinalis* found that *Ci-leprecan* is most highly expressed in the notochordal sheath (NS) (Hotta *et al*., 2000); as the notochord affects specification of surrounding tissue (Liem *et al*., 2000), this may mean *Ci-leprecan* contributes to tissue specification. While *M. occulta* has been shown to have normal early expression of *Brachyury* in the notochord precursor cells (Takada *et al*., 2002), Molgulid *leprecan* expression remains unexplored. This study will then build upon past studies by assessing the extent to which the *Ci-Bra*/ *Ci-leprecan* association is conserved in *M. occulta* and *M. oculata*.

Finally, we will examine the expression of *merlin*, the product of the NF2 tumor suppressor gene. *Merlin* is a linker between proteins embedded in the cell membrane and the actin-cytoskeleton, making it an important contributor to the orchestration of cell-cell interactions (McClatchey and Giovannini, 2005). Probing for *merlin* in *M. oculata* will give insight into the non-autonomous expression patterns at play during the tissue development of ascidian embryos. *Merlin* is also interesting because it is closely linked with nervous tissue, as is evident by how nonfunctional *merlin* correlates with human neurofibromatosis type II, the symptoms of which are schwannomas, ependymomas and meningiomas (Stamenkovic and Yu, 2010).

Materials and Methods

Animal Samples

M. occulta and *M. oculata* specimens were collected on the sand flats of Pointe de Bloscon at Station Biologique in Roscoff, France by Dr. Billie Swalla in 1999. Eggs and sperm extracted from the animal gonads were fertilized to raise *M. occulta* and *M occulta* x *M. oculata* hybrids until embryos hatched. *In situ* samples of 10 life stages (F+1, F+1.5, F+2, F+3, F+4, F+5, F+6, F+7, F+8, F+9) of the *M. occulta*, *M. oculata* and hybrids were fixed in 4% paraformaldehyde and preserved in 70% ethanol; these samples were used in this study. The procedures for insemination, embryo culture and production of hybrids are previously described (Swalla and Jeffery, 1990).

Molgulid transcriptome

Genes selected for analysis were chosen from *M. oculata*, *M. occulta* and *M. occulta* x *M. oculata* hybrid transcriptomes sequenced by Illumina sequencing at Michigan State University. Nucleotide fragments were between 200-250 bp. RNAs were extracted from several different stages of development, sequenced, then reassembled; the lengths of reassembled *prickle*, *noto6*, *noto17*, *leprecan* and *merlin* were 1077 bp, 1154 bp, 960 bp, 888 bp and 1121 bp respectively. Data from the gastrula (F+3), F+4, and F+6 stages were used for analysis.

Sequence Alignments

Protein and nucleotide BLAST was used to confirm the identity of genes selected and to assess protein conservation. All genes were found to be highly conserved in *Molgula tectiformis* and *Ciona intestinalis*. BLASTx was used to derive protein sequences from the *M. occulta*, *M. oculata* and hybrid DNA sequences and to check the orientation of the

DNA sequences. Reverse complements of the transcripts were obtained using Sequencher. All forward nucleotide sequences and protein sequences from the *M. occulta*, *M. oculata* and hybrid transcriptomes were aligned using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/ clustalw2/). For some genes, there were sections in which one transciptome did not have a reading; this is being attributed to mistakes inherent in DNA sequencing assembly.

PCR and Subcloning

Primers were designed using DNA and protein alignments. Highly conserved sections at the start and end of each protein sequence were translated using the genetic code. This DNA sequence was then compared with the actual DNA sequences derived by Dr. Titus Brown *et al*. (2011, unpublished data). Each primer is between 20-23 bp and was selected for highly conserved 5' and 3' end sequences and minimal degeneracy. The forward and reverse primers used are listed in Table 1:

Table 1. Forward and Reverse Primer Sequences

Gene	Forward	Tm (°C)	Reverse	Tm (°C)
prickle	5'-GATAGCGGATGTTCCTTGGAAG-3'	55.4	5'-CGCTTCGGATGGGTACAGCTC-3'	60.1
noto6	5'-ATGTATGAAATAACAAGATGG-3'	45.7	5'-TCAGTCAAATAAAGAATATTTAA-3'	43.8
noto17	5'-GAGAGTCAAAACATGGATCCG-3'	53.4	5'-TCACTGCGTTTCCCAGGAAGCAT-3'	61.2
leprecan	5'-GAACCTGAAACTGAAGAATTAGC-3'	51.9	5'-GCGTCAGTATCAAGGAACAT-3'	52.4
merlin	5'-AAGAGCATTAACGTTTGCATTTC-3'	52.6	5'—CTTCTTTAGCTTGTGTTTTCAT-3'	49.4

Primer melting points (Tm) ranged from 43.8°C to 61.2°C (see Table 1). PCR conditions were 35 cycles of: 1. Denature at 94°C for 4 minutes, 2. Denature at 94°C for 1 minute, 3.

Anneal at 40°C for 1 minute, 4. Extend at 72°C for 2 minutes. The final PCR cycle was at 72°C for 10 minutes for extension.

PCR inserts were manually extracted from gels and isolated using the GenElute™ Plasmid Miniprep Kit. Following isolation, inserts were incorporated into 3956 bp 4-TOPO® bacterial vectors with T3 and T7 priming sites on either end. These priming sites facilitate linearization for the transcription of sense and anti-sense strands; *Not I* was used to linearize the T3 strand and *Pst I* was used for the T7 strand. On the immediate 6-10 bp following either end of the insert there are also *EcoRI* restriction sites useful for preliminary digests to confirm insert length. Standard transformation procedures were used to ligate plasmids and transform *E. coli* via temperature shock. Subclones were transformed in INVαF' strain of *E. coli* and grown on Kanamycin plates at 37°C. Successfully transformed and proliferated *E. coli* were stored in sterile glycerol and stored at -70°C.

Results

Transcriptome Alignments

BLASTp searches (Altschul *et al.*, 1997) using protein sequences derived from the *M. occulta* and *M. oculata* transcriptomes showed that *prickle* is highly conserved among the ascidians, as well as in more distantly related species like *Homo sapiens* (see Figure 1). The highest BLASTp scores were *Mt-prickle*, *Ci-prickle2* and *Ci-prickle1*. For homologues in other organisms with more than one *prickle* type, *prickle2* was generally the closest match to *Mo-prickle*. Three zinc-binding sites were identified: LIM1 between the 105th and 160th amino acid, LIM2 between the 170th and 220th amino acid and LIM3 between the 230th and 285th amino acid; these are all members of the LIM superfamily, a

group characterized by two highly conserved zinc-finger motifs. A fourth domain, PET_prickle from the Prickle Espinas Testin (PET) superfamily, is found between the 1st and 100th amino acid; this is also characterized by a zinc-finger protein-protein interaction domain. As the molgulid transcriptome sequences are approximately 400 amino acids shorter than that of *M. tectiformis*, it is likely that the full molgulid sequence for prickle has not been isolated (Figure 1). Within the molgulid sequences, there are no significant gaps, meaning that *Mo-prickle* has been highly conserved in the chordates (Figure 1).

Noto6 is well conserved among molgulids but not in more distantly related species like human (Figure 2). The human protein sequence is much longer because a section is absent in all the aligned ascidian sequences (Figure 2); this most likely merely indicates an evolutionary divergence between ascidians and human. Insertions were apparent in the lancelet and human *noto6* homologues approximately halfway into the human protein sequence (Figure 2). No identifiable binding regions were found in BLASTp, though the full sequence was identified as Glyco-tranf-GTA-type superfamily (Marchler-Bauer et al. 2011). Domain hits include the single domain Chondroitin N-acetylgalactosaminyltransferase and multi-domain riboflavin synthase subunit alpha.

A *noto17* BLASTp search yielded one nonspecific hit in the Thioredoxin_like superfamily with a CXXC motif between the 40th and 90th amino acid (Kikuchi et al., 2002). The molgulid *noto17* sequence may be incomplete because the sequence lengths of the transcriptomes are significantly shorter than that of *M. tectiformis*; more rounds of sequencing *Mo-noto17* will help in resolving the full length sequence.

Leprecan was highly conserved among the molgulids, the non-molgulid ascidian *C. intestinalis* and human (Figure 4). A sizable portion of the full sequence appears to be present, but the fact that the transcriptome sequences are shorter than *M. tectiformis* suggests the ends of the sequence may be missing (Figure 4). The molgulid *leprecan* protein sequence had a 20G Fe(II) oxygenase superfamily match between the 110th and 296th amino acid.

Merlin is also highly conserved among the molgulids; there is a low level of conservation with human, but it is worth noting that there were no insertions or deletions present in the human protein sequence. We do not appear to have the full protein isolated, as there are large gaps of information missing between the molgulid transcriptomes themselves and between the transcriptomes and *M. tectiformis* (Figure 5). BLASTp identified an ezrin/radixin/moesin (ERM) protein domain between the first and 210th amino acid; this domain characterizes a phosphatase, which supports the previously established function of *merlin* (McClatchey and Giovannini, 2005). There is a strongly conserved PH-like core between the 210th and 320th amino acid within a FERM-C domain; this site is capable of both peptide and lipid binding (García-Alvarez et al., 2003, Hamada et al., 2000). Upstream there is also a FERM_N and FERM_M domain between the 20th-90th amino acid and 100th-210th amino acid respectively.

PCR and Subcloning

The first PCR reaction had an annealing temperature of 45°C and produced some bands, but none with a proper length. *M. occulta Mo-prickle*, *M. oculata Mo-prickle*, *M. occulta Mo-noto17*, *M. oculata Mo-noto17*, *M. occulta Mo-leprecan*, *M. oculata Mo-leprecan*, *M. occulta Mo-merlin*, and *M. oculata Mo-merlin* were amplified by PCR with

an annealing temperature of 40°C (Figure 7). PCR was repeated for *Mo-noto6* at 38°C, but this temperature also failed to amplify *Mo-noto6* (Figure 8). The failure of my *Mo-noto6* primers may be attributed to the apparent incompleteness of the *Mo-noto6* sequence; further sequencing may reveal more inclusive primer sites.

Subclones were successfully grown for *Mo-prickle*, *Mo-noto17*, *Mo-leprecan* and *Mo-merlin* for both *M. occulta* and *M. oculata*. When 100 µl of *E. coli* transformed with these genes were grown on Kanamycin plates, *Mo-prickle* for *M. oculata* produced the most colonies (31) and *Mo-leprecan* for *M. oculata* produced the least colonies (2). Minipreps digested with *EcoRI* revealed that the inserts for *Mo-noto17*, *Mo-leprecan* and *Mo-merlin* were shorter than the anticipated lengths of 960 bp, 888 bp, and 1121 bp respectively (Figure 9). *Mo-prickle* for *M. oculata* consistently appeared at approximately 1077 bp (Figure 10), as anticipated from the *Mo-prickle* sequence. *Mo-prickle* for *M. occulta*, however, appeared notably shorter than *Mo*-prickle for *M. oculata* (Figure 10).

RNA probes were successfully made for the T7 strand of *Mo-prickle* (Figure 11), but not for the T3 strand; a low concentration of DNA in the T3 transcription reaction may account for the insignificant yield of T3 probes.

Discussion

Prickle *in the Molgulids*

The *M. occulta* and *M. oculata prickle* sequences aligned in this study may imply the presence of more than one *prickle* in the molgulids (Figure 1). *Prickle* is well conserved between *M. occulta* and *M. oculata*, but there are several permutations in the aligned amino acid sequences that may justify a distinction between *prickle* in these species. The relative sizes of prickle in *M. occulta* and *M. oculata* appear to be different

in the DNA segments seen in the miniprep *EcoRI* digestion (Figure 10). This could potentially imply a deletion in the *prickle* sequence of *M. occulta*, but only sequencing the isolated genes may give immediate conclusive results. A search for restriction sites within *prickle* using Sequencher found that *prickle* in *M. occulta* and *M. oculata* have *Hind III* and *Xba I* restriction sites in common, but otherwise have several unshared restriction sites; *prickle* in *M. occulta* has *Kpn I*, *BamHI*, and *Pst I* restriction sites (Figure 12) while *prickle* in *M. oculata* has *Sac I* and *Bst XI* restriction sites (Figure 13). The difference in *prickle* sequences causing these different restriction sites may also support the defining of *Mo-prickle1* and *Mo-prickle2*. The presence of *Ci-prickle1* and *Ci-prickle2* in the ascidian *Ciona intestinalis* (Hotta *et al.*, 2000) strengthens the feasibility of more than one *prickle* being present in molgulid ascidians.

Molecular nature of Molgulid genes

The primers designed to isolate *prickle*, *noto6*, *noto17*, *leprecan* and *merlin* for *M. occulta* and *M. oculata* required exceptionally low annealing temperatures. Many A-T base-pairs translate to lower annealing temperatures because A-T hydrogen bonding is less efficient than C-G hydrogen bonding at overcoming the initial inertia of free-moving primers. It is then possible that these low annealing temperatures are due to the ubiquity of A-T bonds in the gene sequences. High A-T content in the DNA sequences of *prickle*, *noto6*, *noto17*, *leprecan* and *merlin* in *M. occulta* and *M. oculata* is readily noticeable at the ends of the genes, where primers were designed. It is possible that the ascidians are A-T rich due to environmental pressures in the same way thermophiles are C-G rich (Basak *et al.*, 2010). A-T ubiquity in this sense may then be a consequence of being native to benthic, low-temperature environments.

Future Studies

The immediate next step in this project is the sequencing of the genes isolated during this study. This will confirm their identity and allow the project to move forward into an investigation of gene expression via *in-situ* hybridization. The genes I focused on do not contain the restriction sites used to linearize the T3 (*Not I*) and T7 (*Pst I*) strands for probe synthesis, with the exception of *prickle* in *M. occulta*. This *prickle* appears to have a downstream *Pst I* restriction site (Figure 12). Future researchers may resolve this conflict by choosing a different bacterial vector, in which, the restriction sites used to linearize sense and anti-sense strands are not found in the gene.

Once probes are synthesized for *Mo-prickle*, *Mo-noto17*, *Mo-leprecan* and *Mo-merlin*, *in-situ* hybridizations may be done in the neurula and tail-bud stage embryos of *M. occulta, M. oculata* and *M. occulta* x *M. oculata* hybrids. Dr. Billie Swalla has preserved neurula and tail-bud stage embryos of these organisms available for such a study; animals were collected in Roscoff, France in 1999. The *in-situ* results for *M. oculata* are expected to resemble those found in the ascidian *Ciona intestinalis* (Hotta *et al.*, 2000); in the neurula stage, the probes will appear down the A/P axis of the embryo, while in the tail-bud stage the probes will mainly illuminate the notochord. An *M. occulta* x *M. oculata* hybrid *in-situ* is also expected to show an illuminated notochord; though there are fewer notochord cells present (Swalla and Jeffery, 1996), these hybrids have nonetheless completed convergence and elongation and thus produce the molecular transcripts required to orchestrate this event.

Results of the *M. occulta in-situ* are more elusive. No expression of these genes in the tail-bud stage of *M. occulta* would support the hypothesis that these genes are

required for the full development of a notochord. However, if these genes *are* expressed, it may mean they are not necessarily a requirement for late-stage notochord development. Knock-out studies have shown that without *prickle*, proper development in ascidians does not occur (Jiang *et al.*, 2005); suboptimal development has also been described in *leprecan* knock-out experiments (Dunn and Di Gregorio, 2009). If this holds true for *noto6*, *noto17* and *merlin*, it would mean these genes are a *requirement* for notochord development, but not necessarily the key genes involved in the converge-and-extend event absent in *M. occulta*.

This second scenario immediately raises a question of efficiency: why produce a transcriptional product that doesn't lead to a morphological change? It is possible that expression of these genes in late stages of anural embryonic development may be vestigial signals; natural selection does not necessarily select for complete elimination of gene expression if the expression does not significantly impede the organism.

Another important question to consider while studying molgulid tail-loss is what kind of factors facilitated tail-loss. These factors may be active or passive: Active selective factors mean taillessness was *selected for* by pressures in the environment, while passive factors mean tail-loss simply was not *selected against*. Passive selection would be made possible by the genetic plasticity observed in molgulids (Iannelli *et al.*, 2007).

In-situ hybridization is the ultimate end of this particular study, but it is not the only or even the most conclusive way to investigate the overarching question of this project: are *prickle*, *noto6*, *noto17*, *leprecan* and *merlin* required for the convergence and extension of notochordal cells? Aside from *in-situs*, a knock-out gene experiment for

each of these genes in anural and urodele molgulids would be compelling as a way to directly determine the importance of each gene in the converge-and-extend event.

Acknowledgements

I would like to thank BEACON, National Science Foundation and the American Society for Cell Biology for their financial support, as well as everyone at Friday Harbor Laboratories who has made my research possible. I would especially like to thank my mentor, Dr. Billie J. Swalla, Dr. Sophie George, Peter X. Wu, Paul Hausch, Joie Cannon, Dr. Kenneth Halanych, Dr. Brad Schuster, Silvia Sipulveda, Dr. Titus Brown, Elijah Lowe and Max Maliska for all of their input and guidance.

Literature Cited

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Basak S., Mukhopadhyay P., Gupta S., Ghosh T. (2010) Genomic adaptation of prokaryotic organisms at high temperature. Bioinformation 4 (8): 352-256

Dehal P., Satou Y., Campbell R. K., Chapman J., Degnan B., De Tomaso A., Davidson B., Di Gregorio A., Gelpke M., Goodstein D.M., Harafuji N., Hastings K.E., Ho I., Hotta K., Huang W., Kawashima T., Lemaire P., Martinez D., Meinertzhagen I.A., Necula S., Nonaka M., Putnam N., Rash S., Saiga H. Satake M., Terry A., Yamada L., Wang H.G., Awazu S., Azumi K., Boore J., Branno M., Chin-Bow S., DeSantis R., Doyle S., Francino P., Keys D.N., Haga S., Hayashi H., Hino K., Imai K.S., Inaba K., Kano S., Kobayashi K., Kobayashi M., Lee B.I., Makabe K.W., Manohar C., Matassi G., Medina M., Mochizuki Y., Mount S., Morishita T., Miura S., Nakayama A., Nishizaka S., Nomomoto H., Ohta F., Oishi K., Rigoutsos I., Sano M., Sasaki A., Sasakura Y., Shoguchi E., Shin-i T., Spagnuolo A., Stainier D., Suzuki M.M., Tassy O., Takatori N., Tokuoka M., Yagi K., Yoshizaki F., Wada S., Zhang C., Hyatt P.D., Larimer F., Detter C., Doggett N., Glavina T., Hawkins T., Richardson P., Lucas S., Kohara Y., Levine M., Satoh N., and Rokhsar D.S. (2002) The draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins. Science 298: 2157-2167

Di Gregorio A., Harland R.M., Levine M., Casey E.S. (2002) Tail morphogenesis in the ascidian, *Ciona intestinalis*, requires cooperation between notochord and muscle. Developmental Biology 244 (2): 385–395

Dunn M.P. and Di Gregorio A. (2009) The evolutionarily conserved leprecan gene: its regulation by Brachyury and its role in the developing *Ciona* notochord. Developmental Biology 328 (2): 561-574

García-Alvarez B., de Pereda J.M., Calderwood D.A., Ulmer T.S., Critchley D., Campbell I.D., Ginsberg M.H., Liddington R.C. (2003) Structural determinants of integrin recognition by talin. Molcular Cell 11(1): 49-58

Gyoja F., Satou Y., Shin-i T., Kohara Y., Swalla B.J., Satoh N. (2007) Analysis of large scale expression sequenced tags (ESTs) from the anural ascidian, *Molgula tectiformis*. Developmental Biology 307: 460-482

Hamada K., Shimizu T., Matsui T., Tsukita S., Hakoshima T. (2000) Structural basis of the membrane-targeting and unmasking mechanisms of the radixin FERM domain. European Molecular Biology Organization, Japan 19(17): 4449-62

Hotta K., Takahashi H., Asakura T., Saitoh B., Takatori N., Satou Y., Satoh N. (2000) Characterization of Brachyury-Downstream Notochord Genes in the *Ciona intestinalis* Embryo. Developmental Biology 224: 69-80

Huber J.L., Burke da Silva K., Bates W.R., Swalla B.J. (2000) The evolution of anural larvae in molgulid ascidians. Cell & Developmental Biology 11 (6): 419-426

Iannelli F., Griggio F., Pesole G., Gissi C. (2007) The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level. BMC Evolutionary Biology 7 (155). (http://www.biomedcentral.com/1471-2148/7/155)

Jiang D., Munro E.M., Smith W.C. (2005) Ascidian prickle regulates both mediolateral and anterior-posterior cell polarity of notochord cells. Current Biology 15 (1): 79-85

Kikuchi M., Doi E., Tsujimoto I., Horibe T., Tsujimoto Y. Functional analysis of human P5, a protein disulfide isomerase homolog. (2002) J Biochem 132(3): 451-5

Kulger J.E., Pierre K., Bouquet J., Jiang D., Di Gregorio A. (2011) Evolutionary changes in the notochord genetic toolkit: a comparative analysis of notochord genes in the ascidian *Ciona* and the larvacean *Oikopleura*. BMC Evolutionary Biology 11: 21

Liem K.F. Jr, Jessell T.M., Briscoe J. (2000) Regulation of the neural patterning activity of sonic hedgehog by secreted BMP inhibitors expressed by notochord and somites. Development 127 (22): 4855-66

Marchler-Bauer A., Lu S., Anderson J.B., Chitsaz F., Derbyshire M.K., DeWeese-Scott C., Fong J.H., Geer L.Y., Geer R.C., Gonzales N.R., Gwadz M., Hurwitz D.I., Jackson J.D., Ke Z., Lanczycki C.J., Lu F., Marchler G.H., Mullokandov M., Omelchenko M.V., Robertson C.L., Song J.S., Thanki N., Yamashita R.A., Zhang D., Zhang N., Zheng C., Bryant S.H. CDD: a Conserved Domain Database for the functional annotation of proteins. (2011) Nucleic Acids Research (Database Issue): D225-9

McClatchey A.I. and Giovannini M. (2005) Membrane organization and tumorigenesis-the NF2 tumor suppressor, Merlin. Genes & Development 19 (19): 2265-2277

Milligan M. (1946) Trichrome stain for formalin-fixed tissue. American Journal of Clinical Pathology, Technical Section 10: 184-85

Scholz C.B. and Technau U. (2003) The ancestral role of Brachyury: expression of NemBra1 in the basal cnidarian *Nematostella vectensis* (Anthozoa). Development and Genes Evolution 212 (12): 563-570

Stamenkovic I. and Yu Q. (2010) Merlin, a "Magic" Linker Between the Extracellular Cues and Intracellular Signaling Pathways that Regulate Cell Motility, Proliferation, and Survival. Current Protein & Peptide Science 11 (6): 471-484.

Swalla B.J. and Jeffery W.R. (1990) Interspecific hybridization between an anural and urodele ascidian: Differential expression of urodele features suggests multiple mechanisms control anural development. Developmental Biology 142: 319-334

Swalla B.J., Makabe K.W., Satoh N., Jeffery W.R. (1993) Novel genes expressed differentially in ascidians with alternate modes of development. Development 119: 307-318

Swalla B.J. and Jeffery W.R. (1996) Requirement of the Manx gene for expression of chordate features in a tailless ascidian larva. Science 274 (5290): 1205-8

Swalla B.J. (2004) Protochordate Gastrulation: Lancelets and Ascidians, in Gastrulation: From Cells to Embryo, Chapter 10 (Stern CD, ed) p.139-149. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Takada N., York J., Davis J.M., Schumpert B., Yasuo H., Satoh N., Swalla B.J. (2002) Brachyury expression in tailless Molgulid ascidian embryos. Evolution & Development 4 (3): 205-211

Tao H, Suzuki M, Kiyonari H., Abe T., Sasaoka T., Ueno N. (2009) Mouse prickle1, the homolog of a PCP gene, is essential for epiblast apical-basal polarity. PNAS 106 (34): 14426-14431

FIGURE LEGENDS

Figure 1. Protein alignment for *prickle* including *M. occulta* and *M. oculata* transcriptome sequences and sequences from another anural molgulid (*M. tectiformis*), a urodele non-molgulid ascidian (*Ciona intestinalis*) and human. Yellow indicates a LIM1 domain, blue indicates a LIM2 domain, green indicates a LIM3 domain and fuchsia indicates a PET_prickle domain. Amino acids conserved between all species are indicated by an asterisk (*), amino acids with the same charge are indicated by a period (.) and significant but imperfect matches are indicated by a semicolon (:).

Figure 2. Protein alignment for *noto6* including *M. occulta* and *M. oculata* transcriptome sequences and sequences from another anural molgulid (*M. tectiformis*), a urodele non-molgulid ascidian (*C. intestinalis*), lancelet and human. Amino acids conserved between all species are indicated by an asterisk (*), amino acids with the same charge are indicated by a period (.) and significant but imperfect matches are indicated by a semicolon (:).

Figure 3. Protein alignment for *noto17* including *M. occulta* and *M. oculata* transcriptome sequences and sequences from another anural molgulid (*M. tectiformis*) and two urodele non-molgulid ascidians (*C. intestinalis* and *Oikopleura dioica*). Yellow indicates a Thioredoxin_like domain. Amino acids conserved between all species are indicated by an asterisk (*), amino acids with the same charge are indicated by a period (.) and significant but imperfect matches are indicated by a semicolon (:).

Figure 4. Protein sequence alignment for *leprecan* including hybrids of *M. occulta* and *M. oculata*. In addition there are sequences for *M. tectiformis*, *C. intestinalis* and human. Yellow indicates a 20G Fe(II) oxygenase domain. Amino acids conserved between all species are indicated by an asterisk (*), amino acids with the same charge are indicated by a period (.) and significant but imperfect matches are indicated by a semicolon (:).

Figure 5. Protein sequence alignment for *merlin* including *M. occulta* and *M. oculata*, as well as *M. tectiformis*, *C. intestinalis*, an acorn worm (*Saccoglossus kowalevskii*) and human. Yellow indicates an ezrin/radixin/moesin (ERM) domain. White diamonds (◊) indicates a FERM_N domain, black dots (●) indicate a FERM_M domain and white pentagons (⌂) indicate a FERM_C domain with a PH-like core. Amino acids conserved between all species are indicated by an asterisk (*), amino acids with the same charge are indicated by a period (.) and significant but imperfect matches are indicated by a semicolon (:).

Figure 6. Results of PCR with 45°C annealing temperature. Lanes starting from the far left are: 1 kb ladder, *prickle* for *M. occulta*, *prickle* for *M. oculata*, *noto17* for *M. occulta*, *noto17* for *M. oculata*, *noto6* for *M. occulta*, *noto6* for *M. oculata*, *leprecan* for *M. occulta*, *leprecan* for *M. oculata*, *merlin* for *M. occulta* and *merlin* for *M. oculata*. Expected bands for *prickle*, *noto17*, *noto6*, *leprecan* and *merlin* are 1077 bp, 960 bp, 1154 bp, 888 bp and 1121 bp respectively; the bands that appear are too small to be correct.

Figure 7. Results of PCR with 40°C annealing temperature. Lanes starting from the far left are: 1 kb ladder, *prickle* for *M. occulta*, *prickle* for *M. oculata*, *noto6* for *M. occulta*, *noto6* for *M. oculata*, *noto17* for *M. occulta*, *noto17* for *M. oculata*, *leprecan* for *M. occulta*, *leprecan* for *M. oculata*, *merlin* for *M. occulta* and *merlin* for *M. oculata*. Expected bands for *prickle*, *noto17*, *noto6*, *leprecan* and *merlin* are 1077 bp, 960 bp, 1154 bp, 888 bp and 1121 bp respectively.

Figure 8. Results of PCR with 38°C annealing temperature. Lanes starting from the far left are: 1 kb ladder, *noto6* for *M. occulta*, *noto6* for *M. oculata*, uncut *merlin* for *M. oculata* (miniprep DNA) and *EcoRI* digest of *merlin* for *M. oculata*.

Figure 9. Results of first miniprep. Lanes starting from the far left of Gel 1 are: 1 kb ladder, uncut *noto17* for *M. occulta* culture 1, *EcoRI* digest of *noto17* for *M. occulta* culture 1, uncut *noto17* for *M. occulta* culture 2, *EcoRI* digest of *noto17* for *M. occulta* culture 2, uncut *noto17* for *M. oculata* culture 3, *EcoRI* digest of *noto17* for *M. oculata* culture 3, uncut *noto17* for *M. oculata* culture 4, *EcoRI* digest of *noto17* for *M. oculata* culture 4, uncut *noto17* for *M. oculata* culture 5, *EcoRI* digest of *noto17* for *M. oculata* culture 5, uncut *merlin* for *M. occulta* culture 6, *EcoRI* digest of *merlin* for *M. occulta* culture 6, uncut *merlin* for *M. occulta* culture 7 and *EcoRI* digest of *merlin* for *M. occulta* culture 7.

Lanes starting from the far left of Gel 2 are: 1 kb ladder, uncut *merlin* for *M. occulta* culture 8, *EcoRI* digest of *merlin* for *M. occulta* culture 8, uncut *merlin* for *M. oculata* culture 9, *EcoRI* digest of *merlin* for *M. oculata* culture 9, uncut *merlin* for *M. oculata* culture 10, *EcoRI* digest of *merlin* for *M. oculata* culture 10, uncut *merlin* for *M. oculata* culture 11, *EcoRI* digest of *merlin* for *M. oculata* culture 11, uncut *prickle* for *M. oculata* culture 12 and *EcoRI* digest of *prickle* for *M. oculata* culture 12.

Figure 10. Results of second miniprep. Lanes starting from the far left of Gel 1 are: 1 kb ladder, uncut *prickle* for *M. oculata* culture 1, *EcoRI* digest of *prickle* for *M. oculata* culture 1, uncut *prickle* for *M. oculata* culture 2, *EcoRI* digest of *prickle* for *M. oculata* culture 2, uncut *prickle* for *M. oculata* culture 3, *EcoRI* digest of *prickle* for *M. oculata* culture 3, uncut *prickle* for *M. oculata* culture 4, *EcoRI* digest of *prickle* for *M. oculata* culture 4, uncut *prickle* for *M. oculata* culture 5, *EcoRI* digest of *prickle* for *M. oculata* culture 5, uncut *prickle* for *M. oculata* culture 6, *EcoRI* digest of *prickle* for *M. oculata* culture 6, uncut *prickle* for *M. occulta* culture 7 and *EcoRI* digest of *prickle* for *M. occulta* culture 7.

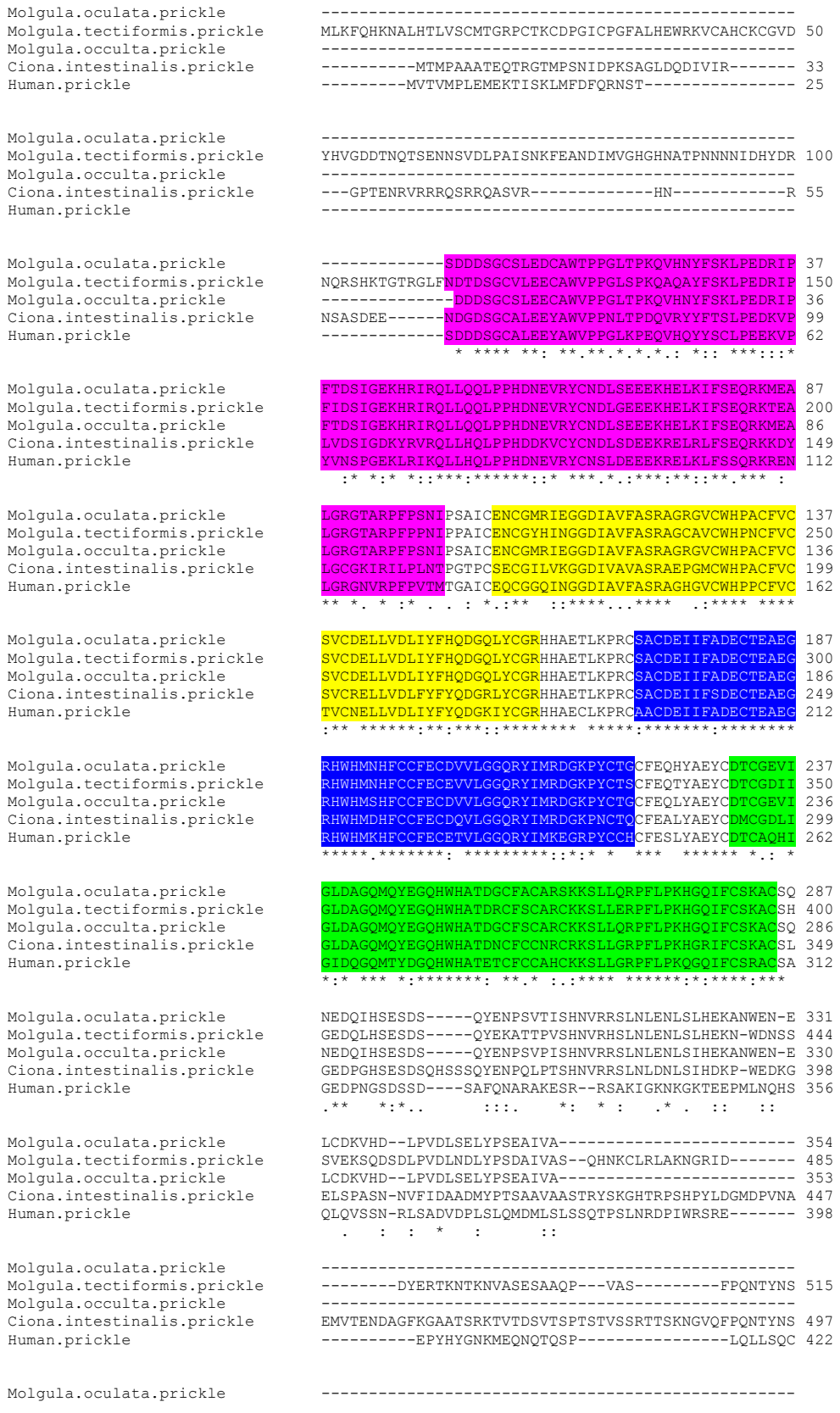
Lanes starting from the far left of Gel 2 are: 1 kb ladder, uncut *prickle* for *M. occulta* culture 8, *EcoRI* digest of *prickle* for *M. occulta* culture 8, uncut *prickle* for *M. occulta* culture 9, *EcoRI* digest of *prickle* for *M. occulta* culture 9, uncut *leprecan* for *M. occulta* culture 10, *EcoRI* digest of *leprecan* for *M. occulta* culture 10, uncut *leprecan* for *M. occulta* culture 11, *EcoRI* digest of *leprecan* for *M. occulta* culture 11, uncut *leprecan* for *M. occulta* culture 12 and *EcoRI* digest of *leprecan* for *M. occulta* culture 12.

Figure 11. Dot blot test for T3 and T7 RNA probes. Y axis shows 1/10, 1/50, 1/100 and 1/500 dilutions for the probes, while the X axis shows, from left to right, the control probe, T3 probe and T7 probe. A purple dot signifies a successful probe.

Figure 12. Sequencher representation of restriction sites in *Mo-prickle* from *Molgula occulta*.

Figure 13. Sequencher representation of restriction sites in *Mo-prickle* from *Molgula oculata*.

Figure 1


Molgula.oculata.prickle ----------
Molgula.tectiformis.prickle MLKFQHKNALHTLVSCMTGRPCTKCDPGICPGFALHEWRKVCAHCKCGVD 50
Molgula.occulta.prickle ----------
Ciona.intestinalis.prickle ----------MTMPAAATEQTRGTMPSNIDPKSAGLDQDIVIR------- 33
Human.prickle ---------MVTVMPLEMEKTISKLMFDFQRNST---------- 25
Molgula.oculata.prickle ----------
Molgula.tectiformis.prickle YHVGDDTNQTSENNSVDLPAISNKFEANDIMVGHGHNATPNNNNIDHYDR 100
Molgula.occulta.prickle ----------
Ciona.intestinalis.prickle ---GPTENRVRRRQSRRQASVR----------HN----------R 55
Human.prickle ----------
Molgula.oculata.prickle ----------SDDDSGCSLEDCAWTPPGLTPKQVHNYFSKLPEDRIP 37
Molgula.tectiformis.prickle NQRSHKTGTRGLFNDTDSGCVLEECAWVPPGLSPKQAQAYFSKLPEDRIP 150
Molgula.occulta.prickle ----------DDDSGCSLEECAWVPPGLTPKQVHNYFSKLPEDRIP 36
Ciona.intestinalis.prickle NSASDEE------NDGDSGCALEEYAWVPPNLTPDQVRYYFTSLPEDKVP 99
Human.prickle ----------SDDDSGCALEEYAWVPPGLKPEQVHQYYSCLPEEKVP 62
Molgula.oculata.prickle FTDSIGEKHRIRQLLQQLPPHDNEVRYCNDLSEEEKHELKIFSEQRKMEA 87
Molgula.tectiformis.prickle FIDSIGEKHRIRQLLQQLPPHDNEVRYCNDLGEEEKHELKIFSEQRKTEA 200
Molgula.occulta.prickle FTDSIGEKHRIRQLLQQLPPHDNEVRYCNDLSEEEKHELKIFSEQRKMEA 86
Ciona.intestinalis.prickle LVDSIGDKYRVRQLLHQLPPHDDKVCYCNDLSDEEKRELRLFSEQRKKDY 149
Human.prickle YVNSPGEKLRIKQLLHQLPPHDNEVRYCNSLDEEEKRELKLFSSQRKREN 112
Molgula.oculata.prickle LGRGTARPFPSNIPSAICENCGMRIEGGDIAVFASRAGRGVCWHPACFVC 137
Molgula.tectiformis.prickle LGRGTARPFPPNIPPAICENCGYHINGGDIAVFASRAGCAVCWHPNCFVC 250
Molgula.occulta.prickle LGRGTARPFPSNIPSAICENCGMRIEGGDIAVFASRAGRGVCWHPACFVC 136
Ciona.intestinalis.prickle LGCGKIRILPLNTPGTPCSECGILVKGGDIVAVASRAEPGMCWHPACFVC 199
Human.prickle LGRGNVRPFPVTMTGAICEQCGGQINGGDIAVFASRAGHGVCWHPPCFVC 162
Molgula.oculata.prickle SVCDELLVDLIYFHQDGQLYCGRHHAETLKPRCSACDEIIFADECTEAEG 187
Molgula.tectiformis.prickle SVCDELLVDLIYFHQDGQLYCGRHHAETLKPRCSACDEIIFADECTEAEG 300
Molgula.occulta.prickle SVCDELLVDLIYFHQDGQLYCGRHHAETLKPRCSACDEIIFADECTEAEG 186
Ciona.intestinalis.prickle SVCRELLVDLFYFYQDGRLYCGRHHAETLKPRCSACDEIIFSDECTEAEG 249
Human.prickle TVCNELLVDLIYFYQDGKIYCGRHHAECLKPRCAACDEIIFADECTEAEG 212
Molgula.oculata.prickle RHWHMNHFCCFECDVVLGGQRYIMRDGKPYCTGCFEQHYAEYCDTCGEVI 237
Molgula.tectiformis.prickle RHWHMNHFCCFECEVVLGGQRYIMRDGKPYCTSCFEQTYAEYCDTCGDII 350
Molgula.occulta.prickle RHWHMSHFCCFECDVVLGGQRYIMRDGKPYCTGCFEQLYAEYCDTCGEVI 236
Ciona.intestinalis.prickle RHWHMDHFCCFECDQVLGGQRYIMRDGKPNCTQCFEALYAEYCDMCGDLI 299
Human.prickle RHWHMKHFCCFECETVLGGQRYIMKEGRPYCCHCFESLYAEYCDTCAQHI 262
Molgula.oculata.prickle GLDAGQMQYEGQHWHATDGCFACARSKKSLLQRPFLPKHGQIFCSKACSQ 287
Molgula.tectiformis.prickle GLDAGQMQYEGQHWHATDRCFSCARCKKSLLERPFLPKHGQIFCSKACSH 400
Molgula.occulta.prickle GLDAGQMQYEGQHWHATDGCFSCARCKKSLLQRPFLPKHGQIFCSKACSQ 286
Ciona.intestinalis.prickle GLDAGQMQYEGQHWHATDNCFCCNRCRKSLLGRPFLPKHGRIFCSKACSL 349
Human.prickle GIDQGQMTYDGQHWHATETCFCCAHCKKSLLGRPFLPKQGQIFCSRACSA 312
Molgula.oculata.prickle NEDQIHSESDS-----QYENPSVTISHNVRRSLNLENLSLHEKANWEN-E 331
Molgula.tectiformis.prickle GEDQLHSESDS-----QYEKATTPVSHNVRHSLNLENLSLHEKN-WDNSS 444
Molgula.occulta.prickle NEDQIHSESDS-----QYENPSVPISHNVRRSLNLENLSIHEKANWEN-E 330
Ciona.intestinalis.prickle GEDPGHSESDSQHSSSQYENPQLPTSHNVRRSLNLDNLSIHDKP-WEDKG 398
Human.prickle GEDPNGSDSSD----SAFQNARAKESR--RSAKIGKNKGKTEEPMLNQHS 356
Molgula.oculata.prickle LCDKVHD--LPVDLSELYPSEAIVA---------- 354
Molgula.tectiformis.prickle SVEKSQDSDLPVDLNDLYPSDAIVAS--QHNKCLRLAKNGRID------- 485
Molgula.occulta.prickle LCDKVHD--LPVDLSELYPSEAIVA---------- 353
Ciona.intestinalis.prickle ELSPASN-NVFIDAADMYPTSAAVAASTRYSKGHTRPSHPYLDGMDPVNA 447
Human.prickle QLQVSSN-RLSADVDPLSLQMDMLSLSSQTPSLNRDPIWRSRE------- 398
Molgula.oculata.prickle ----------
Molgula.tectiformis.prickle --------DYERTKNTKNVASESAAQP---VAS---------FPQNTYNS 515
Molgula.occulta.prickle ----------
Ciona.intestinalis.prickle EMVTENDAGFKGAATSRKTVTDSVTSPTSTVSSRTTSKNGVQFPQNTYNS 497
Human.prickle ----------EPYHYGNKMEQNQTQSP----------LQLLSQC 422
Molgula.oculata.prickle ----------

```
Molgula.tectiformis.prickle      TDSSGYNSSSTIDAMDHKSRGRNGQTKVVPSDTLNSTCSTA-SQATTCSA 564
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       TDSSGYNSSSTLDAIEH-------QQNAALKAAMGSNYSYGKSKQTPCSK 540
Human.prickle                    NIRTSYSPGG--------------QGAGAQPEMWGKHFSNPKRSSSLAMT 458


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      MSNSDKAFCMGGHVPASEFVPYRVNCSESNKNNNIVQKRIESILSSTQD- 613
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       RPQNGE----DGHVSATEFTPFHPAAPRASPPTIIGSRKLAPEIKKTIDS 586
Human.prickle                    -----------GHAGS-----FIKECREDYYPGRLRSQESYSDMSSQSFS 492


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      TQNSSVWQTRRPSVEKIS---AFKSVSDSNNILKHPIINGPIPAPRAKYP 660
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       LTKATEIDNKSPPVNVAS---MLPKSAVPIPAPRARYAPSLTPSPPSTAA 633
Human.prickle                    ETRGSIQVPKYEEEEEEEGGLSTQQCRTRHPISSLKYTEDMTPTEQTPRG 542


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      SKQSD-WSDANQSLSKNEPARR------VSDITPPSNYFKDTP------- 696
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       SELTSPWMHKSHARTDSPPDSREFPSPPVPVRSPPTESKEHSSPLQRSVS 683
Human.prickle                    SMESLALSNATGLSADGGAKRQE----HLSRFSMPDLSKDSGMN----VS 584


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      -RGATRRLSENNIGL---------RLNT-----TSSTGSNNSQPRGILKK 731
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       ERLANKRRSREPISLPEQTISEHPRLRSDDKHVSVENDKTSPELKSILKK 733
Human.prickle                    EKLSNMGTLNSSM-----------QFRSAESVRSLLSAQQYQEMEGNLHQ 623


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      SRSQV------ESGNIS-------DLHTPTDETPVSP------IF--PEP 760
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       SRNPSKSFRNRERGSLSGSLDRLEEFHRKSDVMKYASDDEDGAGFGDAQG 783
Human.prickle                    LSNPIGYRDLQSHGRMHQS----FDFDGGMAGSKLPGQE----GVRIQPM 665


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      NTPPFNRAGRLNQSARFP----------NAPGSPESKSTTNYFS------ 794
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       DFSSFQRGQRLYSSARFPEEVTEKPRSQNQGGRPRSQHRTRFKDNSALDR 833
Human.prickle                    SERTRRRATSRDDNRRFR------PHRSRRSRRSRSDNALHLASER---- 705


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      ------------------------ECEKK----TCSKKLRRTKSTDFTSK- 816
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       THSALNLDELDCAIARRNPKPGKTCSKLSGKSTCSKKLKRTRSTDFAFER 883
Human.prickle                    -------EAISRLKDRPPLRAREDYDQFMRQRSFQESMGHGSRRDLYGQC 748


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      ----ATGASKKKRQARFANDVPDEHDSWCSTCTSSSDDSDYERWDKFDDN 862
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       SAATPTSSRKNRRTKRFVED--EEEDGWCSTCTSSNDDSDYERWDGLGTS 931
Human.prickle                    PRTVSDLALQNAFGDRWGPY--FAEYDWCSTCSSSSE-SDNE---GYFLG 792


Molgula.oculata.prickle          --------------------------------------------------
Molgula.tectiformis.prickle      VSMTSSPHHQRRST----EFNLTHLQNLQ--------QQAKLRYGVQST- 899
Molgula.occulta.prickle          --------------------------------------------------
Ciona.intestinalis.prickle       PPTSPLSAMRRGSAPVGVRVNMTRRQPPHPFLANADSALAASAAGFNSNG 981
Human.prickle                    EPIPQPARLR----------YVTSDELLH----------KYSSYGLPKS- 821


Molgula.oculata.prickle          ------------------------------
Molgula.tectiformis.prickle      ---SALPK----YHHSRHSRRHHKKNCVIM 922
Molgula.occulta.prickle          ------------------------------
Ciona.intestinalis.prickle       VYRPSMPRNFSTTSHMRYRRRQQKKHCIVM 1011
Human.prickle                    ---STLGG----RGQLHSRKRQKSKNCIIS 844
```

Figure 2


Molgula.oculata.noto6 ---------- 50
Molgula.occulta.noto6 ---------- 50
Molgula.tectiformis.noto6 ---------- 50
Ciona.intestinalis.noto6 ---------- 50
Lancelet.noto6 ---------- 50
Human.noto6 MAVRSRRPWMSVALGLVLGFTAASWLIAPRVAELSERKRRGSSLCSYYGR 50
Molgula.oculata.noto6 ---------- 100
Molgula.occulta.noto6 ---------- 100
Molgula.tectiformis.noto6 ---------- 100
Ciona.intestinalis.noto6 ---------- 100
Lancelet.noto6 ---------- 100
Human.noto6 SAAGPRAGAQQPLPQPQSRPRQEQSPPPARQDLQGPPLPEAAPGITSFRS 100
Molgula.oculata.noto6 ---------- 150
Molgula.occulta.noto6 ---------- 150
Molgula.tectiformis.noto6 ---------- 150
Ciona.intestinalis.noto6 ---------- 150
Lancelet.noto6 ---------- 150
Human.noto6 SPWQQPPPLQQRRRGREPEGATGLPGAPAAEGEPEEEDGGAAGQRRDGRP 150
Molgula.oculata.noto6 ---------- 200
Molgula.occulta.noto6 ---------- 200
Molgula.tectiformis.noto6 ---------- 200
Ciona.intestinalis.noto6 ---------- 200
Lancelet.noto6 ---------- 200
Human.noto6 GSSHNGSGDGGAAAPSARPRDFPYVGVMTAQKYLGSRALAAQRTRARFIP 200
Molgula.oculata.noto6 ---------- 250
Molgula.occulta.noto6 ---------- 250
Molgula.tectiformis.noto6 ---------- 250
Ciona.intestinalis.noto6 ---------- 250
Lancelet.noto6 ---------- 250
Human.noto6 GHVEFFSSQQPPNAGQPPPPLPVIALPGVDDSYPPQKKSFMMIKYMHDHY 250
Molgula.oculata.noto6 ---------- 300
Molgula.occulta.noto6 ---------- 300
Molgula.tectiformis.noto6 ---------- 300
Ciona.intestinalis.noto6 ---------- 300
Lancelet.noto6 ---------- 300
Human.noto6 LDKYEWFMRADDDVYIKEDKLEEFLRSLNSSKPLYLGQTGLGNIEELGKL 300
Molgula.oculata.noto6 ---------- 350
Molgula.occulta.noto6 ---------- 350
Molgula.tectiformis.noto6 ---------- 350
Ciona.intestinalis.noto6 ---------- 350
Lancelet.noto6 ---------- 350
Human.noto6 GLEPGENFCMGGPGMIFSREVLRRMVPHIGECLREMYTTHEDVEVGRCVR 350
Molgula.oculata.noto6 ---------- 400
Molgula.occulta.noto6 ---------- 400
Molgula.tectiformis.noto6 ---------- 400
Ciona.intestinalis.noto6 ----------MRFLRGLNILRIMTGIMLGFLSRIA 25
Lancelet.noto6 ----------MFGRSLRRPGFPLVPVCAGIVVGYILRGL 29
Human.noto6 RFGGTQCVWSYEMQQLFHENYEHNRKGYIQDLHNSKIHAAITLHPNKRPA 400
Molgula.oculata.noto6 ---------- 450
Molgula.occulta.noto6 ---------- 450
Molgula.tectiformis.noto6 ---------- 450
Ciona.intestinalis.noto6 FEQ---LRGSTRLTVGTHMDNIKSLDCPKCPIDDHCTLCEPTHVKNPTEN 72
Lancelet.noto6 FDS---TFLS--IIEEKRYKKMSSEKESTPAHCKPCPLCALCSDCWRHDA 74
Human.noto6 YQYRLHNYMLSRKISELRYRTIQLHRESALMSKLSNTEVSKEDQQLGVIP 450
Molgula.oculata.noto6 ---------MYEITRWYYFDDNYLYDIINEKP-KVPLIGQWKEETQAVTQ 40
Molgula.occulta.noto6 ----------EETQAVTQ 8
Molgula.tectiformis.noto6 ---------MYEITRWYYFDDNYLYDIINDEP-KVLLIGQWAEEITAVTQ 40
Ciona.intestinalis.noto6 IPSAYKPKGMYEVTRWLQFDEHYVYDIINEEP-KIPLVGHWLEEVRAVTH 121
Lancelet.noto6 IEEPYRPKTMYEVIRWFHFTEKHLYENLDDDP-RIGLIGRWKEDVQDIQE 123
Human.noto6 SFNHFQPRERNEVIEWEFLTGKLLYSAAENQPPRQSLSSILRTALDDTVL 500

```
Molgula.oculata.noto6          IALEYINI-NKMMGRTWKLLKRENGYIQSDVKHGTEYMIDLKVKSMQDDV 89
Molgula.occulta.noto6          IALEYINI-NKMMGRTWKLLKRENGYIQSDVKHGTEYVIDLKVKSMQDDV 57
Molgula.tectiformis.noto6      VALEYINT-NEILGRTWKIIKRENGYLKSDVFQGREYIFDIKAESMQDGI 89
Ciona.intestinalis.noto6       VAMDYINA-NRMLGQSWSLVKLHDGYLRRDALRGTEYILDVVVQPASHNF 170
Lancelet.noto6                 QAVRIISK-EEKT--QWRLAKLNYGYKRIDAFRGMDYVLDMEVAKADGTE 170
Human.noto6                    QVMEMINENAKSRGRLIDFKEIQYGYRRVNPMHGVEYILDLLLLYKRHKG 550
                                .:  *.   .       : : . ** : :  :* :*::*:

Molgula.oculata.noto6          --------------------------------------------------
Molgula.occulta.noto6          --------------------------------------------------
Molgula.tectiformis.noto6      --------------------------------------------------
Ciona.intestinalis.noto6       S------------------------------------------------- 171
Lancelet.noto6                 KGRKLVHLVRSTRPHVRMLSLQNVAQQTVVHFIVPLSRVTDRLRDFMKVY 220
Human.noto6                    RKLTVPVRRHAYLQQLFSKPFFRETEELDVN------------------- 581


Molgula.oculata.noto6          --------------------------EETFRTTVVHPL--QSSYEVQSFI 111
Molgula.occulta.noto6          --------------------------EETFRTTIVHPL--ESTYEVESFI 79
Molgula.tectiformis.noto6      -------------------------VRSFRIQLLHPFNHQNEVEVVSYV 113
Ciona.intestinalis.noto6       ----------------------YFINERMFRVNLVHPLKSMGEAEEVRVQ 199
Lancelet.noto6                 EKVCLETGAAVSLLGVLFLHSFIQSFTHSFTPSFLHSFT-HSPLPTCRCM 269
Human.noto6                    -----------------SLVESINSETQSFSFISNSLKILSSFQGAKEMG 614
                                                  . *          .

Molgula.oculata.noto6          QDTQHPIQIIIPISGDLTNTVETLNNFKEVE---LEDVCLT-------LI 151
Molgula.occulta.noto6          QDTQHQIQIIIPLSGDLTYTMETLNNFKEVK---LEYICLT-------LI 119
Molgula.tectiformis.noto6      QNIQKPVHIIVPLTGVLNNAKNIMLVFQQTAKL-KEGLCLT-------LV 155
Ciona.intestinalis.noto6       RHVEGRLKIIVPAKITSPGIVASIDNFVKSKP--KPQVHLV-------LV 240
Lancelet.noto6                 RRFAWRAALPSASSGFCSFTHSFSHSLTHSLLC-MKXVCLESGAAVSLLG 318
Human.noto6                    GHNEKKVHILVPLIGRYDIFLRFMENFENMCLIPKQNVKLV-------II 657
                                      :  .             : .       : *        :

Molgula.oculata.noto6          SFTGNGMWGYSFDETIRTHIDKWMLENNNLCAKQQTVIGDYNYVRALKT- 200
Molgula.occulta.noto6          LFTGNGMWGYSFDETILTHIDKWKMENNNLCIKVETATGDYNYARALRT- 168
Molgula.tectiformis.noto6      LFKKRGIYQQLFDPNIWEYLENFKQNNKEMCFHMREINSDYNFFDGIKQ- 204
Ciona.intestinalis.noto6       LFLGSNKKEISGYKNKLQELETKIKSYQKVNIQIYPVEGLYSYFKGVRYG 290
Lancelet.noto6                 VLFRDSPEDQTSLDQCKQLFDQYAAKYPAAAIRYVEADGEFSRAVGLDT- 367
Human.noto6                    LFSRDSGQD---SSKHIELIKGYQNKYPKAEMTLIPMKGEFSRGLGLEM- 703
                                :   .            :.    .            . :.   .:

Molgula.oculata.noto6          VSSSFDDNSIVLTLKDDIEFGPNVVESCRSIAIPRKRVFIPLAFAQHDPI 250
Molgula.occulta.noto6          VSSSFDDNSIVLTLKDDMEFSPDVVESCRSIAVPQKRVFIPLAFAQHDPI 218
Molgula.tectiformis.noto6      SIEEIEENNIILTLKDDIEFNSDAIENCQNLALQNIRVYFPLSFSQFNPK 254
Ciona.intestinalis.noto6       IKKATISTDVVLLATIEAAFTEQLYHHCQANAIPKKRVYFPIAFGQFNPD 340
Lancelet.noto6                 GARQLAPDALLFFCDIDVDFDQGFLQRCRNNAEVGRQVFYPAVFSQYDPD 417
Human.noto6                    ASAQFDNDTLLLFCDVDLIFREDFLQRCRDNTIQGQQVYYPIIFSQYDPK 753
                                        :::    :  *     . *:  :    :*: *  *.*.:*

Molgula.oculata.noto6          NIENGMPVGKPKNINKKDINKFTGHWLYQVHDFICAYAADLDSVVQKIFS 300
Molgula.occulta.noto6          NIERGMPVGKPKNINKKDINKFTGHWLYQVHDFVCAYAADFDNVLRKTLS 268
Molgula.tectiformis.noto6      NIEDGMPFGKPKNVDKKSFTKFTGYWMHNMHDFICAYSDDLKSILSKTES 304
Ciona.intestinalis.noto6       IIKKGMPPGKPKDVNKRDHNKFTGYWMHDTFDFVCANQADMVGILNNLSS 390
Lancelet.noto6                 LVMEGIP-GKPKPTNLRDINKYNGKDTHKPTNLRDINKYNGKDTHKPTNL 466
Human.noto6                    VTNGGNP----PTDDYFIFSKKTGFWRDYGYGITCIYKSDLLGAGGFDTS 799
                                   * *       :    .* .*       .:      :  .

Molgula.oculata.noto6          EEPEKS-NEIYSEISEYGTAVYN-TFLSKGFKVVSSVEPGLLRYYKNHNC 348
Molgula.occulta.noto6          EAPEKS-NEIYFEMSEYGTDMYN-AFLSKGIKVVSSVEPGLLRYYRNHNC 316
Molgula.tectiformis.noto6      LVYEENPNNIFIEQSENGDNVYN-KFLLANYKVISSVEPGLLRLYNNINC 353
Ciona.intestinalis.noto6       LGRRMG-----STTEALGTKIYD-AFLRSDIEVICAVEPGLYRTYGMTCS 434
Lancelet.noto6                 RDINKY----------NGKDTHKPTNINKSLTVFRAMDPALVHRYHAKTC 506
Human.noto6                    IQGWGL----------EDVDLYN-KVILSGLRPFRSQEVGVVHIFHPVHC 838
                                              .   :.   :  .   . : : .: : :    .

Molgula.oculata.noto6          SKIKEAPSKRNCIIRHLESAGSKQALNILYLTEKAS-------- 384
Molgula.occulta.noto6          SKIKEATSKRKCVIRHLESGGSKQALNILYLTENASKM------ 354
Molgula.tectiformis.noto6      TNIRNKLDRRKCKIQMMETSGSKQELNILYLNENASKLR----- 392
Ciona.intestinalis.noto6       KGELTDKSYQHCMRTKAMSLGSKPALGILYINEVLQKKN----- 473
Lancelet.noto6                 DPDLTRDQYRMCVGALSETMASKAQLGVKLLTLQGKLPPAA--- 547
Human.noto6                    DPNLDPKQYKMCLGSKANTFASTMQLAELWLEKHLGVRYNRTLS 882
                                      . : *      : .*.  *    :
```

Figure 3

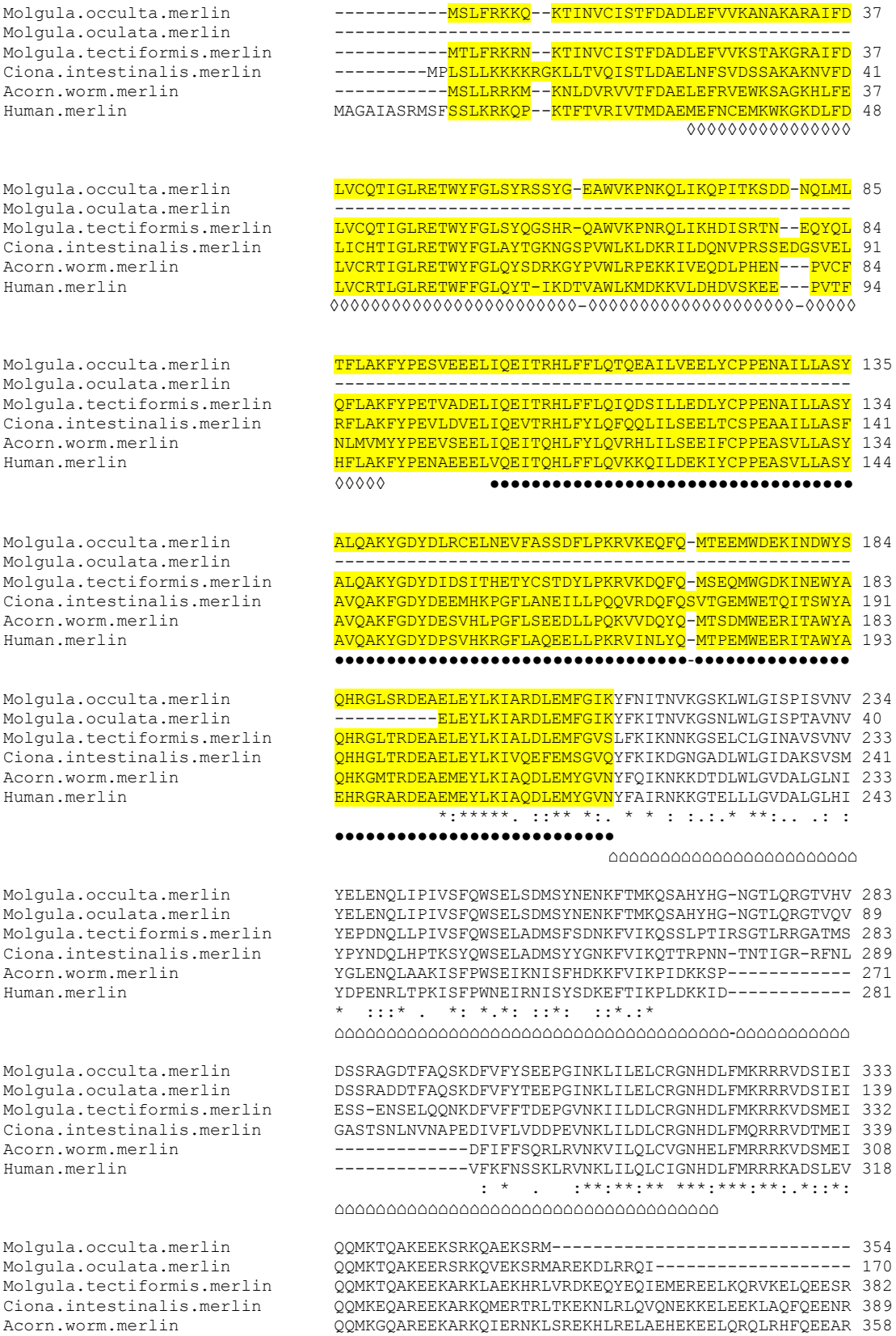

Molgula.oculata.noto17 ----------
Molgula.occulta.noto17 ----------FQSTLSNAKKSFESQNMDPVSASKWSL 27
Molgula.tectiformis.noto17 --------MLLLRCILFVFGFACLNSPIDAKKAKYESQNMDKVSAEKWDL 42
Ciona.intestinalis.noto17 MLTAYGWIMWKLNLLVTLLVLLLVQRGTAEDEKPKISLNMDPVDVKTWGL 50
Oikopleura.dioica.noto17 ----------MRILLFIFAAVAALNSSSLKDTTMDKPRLKKFFL 34
Molgula.oculata.noto17 ----------
Molgula.occulta.noto17 PDQSSYEIMFRLTDKNYEDVLSESEDPWIIVFY-KSKIDLKWIQHSKIVG 76
Molgula.tectiformis.noto17 PDKSNYETIFKITDGNYEDSIMEMEEPWIVLFY-KTKIDRKWLQHSKRVG 91
Ciona.intestinalis.noto17 PLDSKYPDVVQINEANFKPKVLESRDAWIVVVF-KDKLHGKWWHHAGHVK 99
Oikopleura.dioica.noto17 PEVDQYIDWWKITDQNFDQRVLESVDPWIIVIAPEHQIPVAWKEYATRYR 84
Molgula.oculata.noto17 ----------
Molgula.occulta.noto17 GTLWYGKVDLDSNPTLAKKLDFNDKITNFIGGVGVIFPS-KSSEKMKDLN 125
Molgula.tectiformis.noto17 GAIWYGKVDLEKNLHLANEMNFDHKTTNYVGGIAVVFPS-KSKEKKIQFK 140
Ciona.intestinalis.noto17 GAVWYGSVDIEKQENFAKLLAPESTAENIGNGVAVIFPFGVTPKEKIKFT 149
Oikopleura.dioica.noto17 GQFWFGKFFNKWGENLEERFGYEGKPK------VLVYSWTREAKKKGPFI 128
Molgula.oculata.noto17 ----------AENMAAETLPDWTTKFDFTPATMPKFNEWVINSYY 35
Molgula.occulta.noto17 SKKLKK---VIEPGLAENMAAETLPDWTTKFDFTPATMPKFNEWVINSYY 172
Molgula.tectiformis.noto17 AKKLIK---VSEPHLAEKFASETMEDYTTELDFLPSNMQVFNDWVVNTYY 187
Ciona.intestinalis.noto17 GKKKKRPLIVKSPTKAEEIISSSIPNQIERMNFKPSRMKKFQDWVVDSYY 199
Oikopleura.dioica.noto17 TDNFSV----ALEKAIESLPVTSLQRLS--FDTTIEDPSNIDAFLTNALY 172
*.: :: :: :: :: :: *
Molgula.oculata.noto17 QSEI-TRYPVICVS--DEPG--VPLLVKSLSVHFKSHFSFAVVSKASVPR 80
Molgula.occulta.noto17 QSEI-ARYPVMCVT--DEPG--VPLLVKSLAIHFKSHFTFAVVSKASVPR 217
Molgula.tectiformis.noto17 KSPI-AKFPILCVT--DTGD--VPAVVKSLSKLMKSHFVIGVVQKSSAHR 232
Ciona.intestinalis.noto17 KSSP-PRFPVLCIT--DEPEPTIPVSMLVLSNYFSRFFSFALVRKEDVSR 246
Oikopleura.dioica.noto17 GEEKNPKWPVIFLLGKEDEEEGLPPITRVISHYFSDSFKFAFVKFDQLKA 222
. .::*:: : : :* :: :. * :..* .
Molgula.oculata.noto17 MIANFKEVP---ESTQFPMFYLLMGREPSSDEMNRGLYNMNFNIFPYIKE 127
Molgula.occulta.noto17 MMANFKEVP---EPTHFPMFYLLIGREPSTDEMNRGMYNMHFNIFPYIKE 264
Molgula.tectiformis.noto17 MISIFKEVP---MPKVFPMYYTLLGREPTSEEMNRGSYKLHFDIFPYLSE 279
Ciona.intestinalis.noto17 MRLQFKEVP---EPKVYPHYLMLMGKEPSKEEMDRGSFGMHFNIFPFLAD 293
Oikopleura.dioica.noto17 LKKVFPDYYPTDMKDEFPDTFIVAGKEPS-DDAPEDEFELTFSVVRMSKQ 271
: * : :* : *:**: :: .. : : *.:. :
Molgula.oculata.noto17 KYGDSTIFHNVANYLFVANHDFRSMLPGKRSDE---------- 160
Molgula.occulta.noto17 KYGDSNIFHNVANYLFVANHDYRSMLPGKRSDETEKVKKMKGISQALSKR 314
Molgula.tectiformis.noto17 KYGDSSLFHNVLNYMFLVNQDYRADLPGKRRGESSAIKNMEEIKEKMNRK 329
Ciona.intestinalis.noto17 KYGRSELFTSVVKFLFTANEDYRNELPGRRSDEAISEARMSLVRDSLRNR 343
Oikopleura.dioica.noto17 KFGMSTKFNDVAAFFFWVNEQIRETLPGRDPNEDMDKKTMGSIRNKLWKR 321
: * * .* ::* .*.: * ***: .*
Molgula.oculata.noto17 ----------
Molgula.occulta.noto17 VEILMK---------- 320
Molgula.tectiformis.noto17 IKLLLKSQHKKLNKSEL---------- 346
Ciona.intestinalis.noto17 VELVNKHFRKQKNVEL---------- 359
Oikopleura.dioica.noto17 LEIIKPAVMKAQKESDVMQAAKRFENLNAAASKLLEEKEKEAERLKNELE 371
Molgula.oculata.noto17 ----------
Molgula.occulta.noto17 ----------
Molgula.tectiformis.noto17 ----------
Ciona.intestinalis.noto17 ----------
Oikopleura.dioica.noto17 ETQAKLEEAEGKIAESEDSNSTEDPADAKADVEQKDEL 409

Figure 4

```
Molgulid.hybrid.67122.leprecan      ------------------------------------------------
Molgulid.hybrid.67123.leprecan      ------------------------------------------------
Molgula.tectiformis.leprecan        MIQNKKKMLIESVICCILSIIIAANRIQATSVTYQPFEENFETGVTAYWN 50
Ciona.intestinalis.leprecan         ------------------------------------------------
Human.leprecan                      ------------------------------------------------

Molgulid.hybrid.67122.leprecan      ------------------------------------------------
Molgulid.hybrid.67123.leprecan      ------------------------------------------------
Molgula.tectiformis.leprecan        NDWENAIVEIEKSLVMHRLFESGIYDCYMKCKDTRNIENTDRHSQFADIK 100
Ciona.intestinalis.leprecan         ------------------------------------------------
Human.leprecan                      ------------------------------------------------

Molgulid.hybrid.67122.leprecan      ------------------------------------------------
Molgulid.hybrid.67123.leprecan      ------------------------------------------------
Molgula.tectiformis.leprecan        FIISKASCSRSCLDEIFGDARLITQVSEHIKHLFHIKRPYVYLQFAYFQV 150
Ciona.intestinalis.leprecan         ------------------------------------------------
Human.leprecan                      ------------------------------------------------

Molgulid.hybrid.67122.leprecan      ------------------------------------------------
Molgulid.hybrid.67123.leprecan      ------------------------------------------------
Molgula.tectiformis.leprecan        GNIEQAVKTAHTYWLRHVGDDVIEDSIEYYRNLTEVKPEFFIDLEASEHI 200
Ciona.intestinalis.leprecan         ------------------------------------------------
Human.leprecan                      --------------------MEMQQNIENYRATAGVEALQLVDREAKPHM 30

Molgulid.hybrid.67122.leprecan      ------------------------------------------------
Molgulid.hybrid.67123.leprecan      ------------------------------------------------
Molgula.tectiformis.leprecan        YVYDTAVEIYLDEDYEAAIPLFMRALEAYYTASDKCNALCEG-EHVFDYD 249
Ciona.intestinalis.leprecan         ------------------------------------------------
Human.leprecan                      ESYNAGVKHYEADDFEMAIRHFEQALREYFVEDTECRTLCEGPQRFEEYE 80

Molgulid.hybrid.67122.leprecan      ------------------------------------------------
Molgulid.hybrid.67123.leprecan      ------------------------------------------------
Molgula.tectiformis.leprecan        YP-DVPEFHMQTADHYIQVTECSLECVKKIAT-DSSGVHISDFLPLHYHY 297
Ciona.intestinalis.leprecan         ----------------------MMVTQKKIIY-ESSE------------- 14
Human.leprecan                      YLGYKAGLYEAIADHYMQVLVCQHECVRELATRPGRLSPIENFLPLHYDY 130

Molgulid.hybrid.67122.leprecan      ------------------------------------------------
Molgulid.hybrid.67123.leprecan      ------------------------------------------------
Molgula.tectiformis.leprecan        LQFAIYKEGNISAALSNAKTYLLFHPEDEVMVNNVNLLSKKVDSKTEAIP 347
Ciona.intestinalis.leprecan         -NVADSTENNADLAEN-----------EDSIVGEMGDISKNTEDTEDNGD 52
Human.leprecan                      LQFAYYRVGEYVKALECAKAYLLCHPDDEDVLDNVDYYESLLDDS----- 175

Molgulid.hybrid.67122.leprecan      ------------------------------------------------
Molgulid.hybrid.67123.leprecan      ------------------------------------------------
Molgula.tectiformis.leprecan        EAVLYKKRIDGQIALLRYSYQAFGYEYRTNLLTPFADLTENEDGIDEFFS 397
Ciona.intestinalis.leprecan         SAENIELKEN------------------------------SDKSTKNVAA 72
Human.leprecan                      ------------------------------------------------

Molgulid.hybrid.67122.leprecan      -----EPETEELAPEDSFYQSFQEKQADIIEKKSIENEILKEIEAKKTES 45
Molgulid.hybrid.67123.leprecan      -----EPETEELAPEDSFYQSFQEKQADIIEKKSIENEILKEIEAKKTES 45
Molgula.tectiformis.leprecan        GHTLQKPNILPSNPEEQFYQSPDEKQKDIIEKKSMEEHILNELEEQRSQQ 447
Ciona.intestinalis.leprecan         TLKQVEPPKESKTQAKVEYNSMKDKTEDMIEAKDAYNARMEVIRKKRE-- 120
Human.leprecan                      ----IDPASIEAREDLTMFVKRHKLESELIKS-AAEGLGFSYTEPNYWIR 220
                                         .*          : . ..   ::*:       :.  . :

Molgulid.hybrid.67122.leprecan      TTRETEDDV-VDDARSNTHLITSEGMVTIGRTDSVS--SNAEGKILSEKT 92
Molgulid.hybrid.67123.leprecan      TTRETEDDV-VDDARSNTHLITSEGMVTIGRTDSVS--SNAEGKILSEKT 92
Molgula.tectiformis.leprecan        KVAKPDNDVDVSNVSLGEALVDSKDMVTIRDAANET--PTAEGNLLFEDY 495
Ciona.intestinalis.leprecan         --AESDVDISHMSSDIGHEVIDSNSIISIIDSTSLPQPPTPEGPLLHNKV 168
Human.leprecan                      YGGRQDENRVPSGVNVEGAEVHGFSMGKKLSPKIDR-DLREGGPLLYENI 269
```

```
                                    . : :    .      : . .: .   .          * :* :.

Molgulid.hybrid.67122.leprecan      LMVANSTQLNGPTRFVVDHLITENQCQDLINLELSAGVEGDGYQGKKQPH 142
Molgulid.hybrid.67123.leprecan      LMVANSTQLNGPTRFVVDHLITENQCQDLINLELSAGVEGDGYQGKKQPH 142
Molgula.tectiformis.leprecan        VLVANSTQLNGQERFLVDNLITQNDCDDLINLELSRGLFGDGYQDKKYPH 545
Ciona.intestinalis.leprecan         VMLSNSTQMAGPLRFAVDGFASEQQCQDLIDLELSGGVLGDGYMGRVSPH 218
Human.leprecan                      TFVYNSEQLNGTQRVLLDNVLSEEQCRELHSVASGIMLVGDGYRGKTSPH 319
                                     :: ** *: *  *. :* . ::::* :* .:  .  : **** .:  **

Molgulid.hybrid.67122.leprecan      TKFEMFEGLSIHQTAMLAKKGDIPFRIAKLYYDVSEVVRAQVQAYFNLD- 191
Molgulid.hybrid.67123.leprecan      TKFEMFEGLSIHQTAMLAKKGDIPFRIAKLYYDVSEVVRAQVQAYFNLD- 191
Molgula.tectiformis.leprecan        TEYESFQGISIYTAVELAKKSEIPLRIARLYFDTSEKVRAQVQEYFHLP- 594
Ciona.intestinalis.leprecan         SEHELFQGMTVYLAAKLAEKGKVPPQTAALYYKLSEEARLQVKMYFKLTQ 268
Human.leprecan                      TPNEKFEGATVLKALKSGYEGRVPLKSARLFYDISEKARRIVESYFMLNS 369
                                    :  * *:* ::  :   . :. :* : * *::. ** .*  *: ** *

Molgulid.hybrid.67122.leprecan      QLYFDYTHLVCRTALSSASAQDREDLSHPVHADNCLLG-EGGECLKKKPA 240
Molgulid.hybrid.67123.leprecan      QLYFDYTHLVCRTALTSASDEERKDLSHPVHADNCLLG-EGGECLKKKPA 240
Molgula.tectiformis.leprecan        NLYFDYTHLVCRTALPKSSS-EREDLSHPVHADNCLLQ-DNGECLKKPPA 642
Ciona.intestinalis.leprecan         ELYFDYTHLVCRTTV-KGKPVKRTDLSHPVHSDNCLLK-ENGSCLKERPA 316
Human.leprecan                      TLYFSYTHMVCRTALSGQQD-RRNDLSHPIHADNCLLDPEANECWKEPPA 418
                                     ***.***:****::   .   * *****:*:*****  : ..* *: **

Molgulid.hybrid.67122.leprecan      YIWRDYSALLYLNENFEGGNLMFLDTDARHILAEIEPRCGRMVSFCAGKE 290
Molgulid.hybrid.67123.leprecan      YIWRDYSALLYLNENFEGGNLMFLDTDARHILVSFNPHSGL--------- 281
Molgula.tectiformis.leprecan        YVWRDYSAILYLNNKFEGGNLIFVDSTAKRISAQVEPKCGRMAAFCAGKE 692
Ciona.intestinalis.leprecan         YTWRDYSAILYLNDEFEGGEFIMTDATARRVKVQVRPKCGRLVSFSAGKE 366
Human.leprecan                      YTFRDYSALLYMNDDFEGGEFIFTEMDAKTVTASIKPKCGRMISFSSGGE 468
                                    * :*****:**:*:.****:::: :  *: : ....*:.*

Molgulid.hybrid.67122.leprecan      CFHGVK---------------------------------------------- 296
Molgulid.hybrid.67123.leprecan      ----------------------------------------------------
Molgula.tectiformis.leprecan        CFHGVKPVSKGQRCAMALWFTTKKEKQEVQRQIAEETLASLE-------- 734
Ciona.intestinalis.leprecan         CLHGVKPVTKGRRCAMALWFTMDPTHNEESRQKAEKILKR---------- 406
Human.leprecan                      NPHGVKAVTKGKRCAVALWFTLDPLYRELERIQADEVIAILDQEQQGKHE 518


Molgulid.hybrid.67122.leprecan      ---------
Molgulid.hybrid.67123.leprecan      ---------
Molgula.tectiformis.leprecan        ---NSKDEL 740
Ciona.intestinalis.leprecan         ---LLKDEL 412
Human.leprecan                      LNINPKDEL 527
```

Figure 5


Molgula.occulta.merlin ----------MSLFRKKQ--KTINVCISTFDADLEFVVKANAKARAIFD 37
Molgula.oculata.merlin ----------
Molgula.tectiformis.merlin ----------MTLFRKRN--KTINVCISTFDADLEFVVKSTAKGRAIFD 37
Ciona.intestinalis.merlin ---------MPLSLLKKKKRGKLLTVQISTLDAELNFSVDSSAKAKNVFD 41
Acorn.worm.merlin ----------MSLLRRKM--KNLDVRVVTFDAELEFRVEWKSAGKHLFE 37
Human.merlin MAGAIASRMSFSSLKRKQP--KTFTVRIVTMDAEMEFNCEMKWKGKDLFD 48
Molgula.occulta.merlin LVCQTIGLRETWYFGLSYRSSYG-EAWVKPNKQLIKQPITKSDD-NQLML 85
Molgula.oculata.merlin ----------
Molgula.tectiformis.merlin LVCQTIGLRETWYFGLSYQGSHR-QAWVKPNRQLIKHDISRTN--EQYQL 84
Ciona.intestinalis.merlin LICHTIGLRETWYFGLAYTGKNGSPVWLKLDKRILDQNVPRSSEDGSVEL 91
Acorn.worm.merlin LVCRTIGLRETWYFGLQYSDRKGYPVWLRPEKKIVEQDLPHEN---PVCF 84
Human.merlin LVCRTLGLRETWFFGLQYT-IKDTVAWLKMDKKVLDHDVSKEE---PVTF 94
Molgula.occulta.merlin TFLAKFYPESVEEELIQEITRHLFFLQTQEAILVEELYCPPENAILLASY 135
Molgula.oculata.merlin ----------
Molgula.tectiformis.merlin QFLAKFYPETVADELIQEITRHLFFLQIQDSILLEDLYCPPENAILLASY 134
Ciona.intestinalis.merlin RFLAKFYPEVLDVELIQEVTRHLFYLQFQQLILSEELTCSPEAAILLASF 141
Acorn.worm.merlin NLMVMYYPEEVSEELIQEITQHLFYLQVRHLILSEEIFCPPEASVLLASY 134
Human.merlin HFLAKFYPENAEEELVQEITQHLFFLQVKKQILDEKIYCPPEASVLLASY 144
Molgula.occulta.merlin ALQAKYGDYDLRCELNEVFASSDFLPKRVKEQFQ-MTEEMWDEKINDWYS 184
Molgula.oculata.merlin ----------
Molgula.tectiformis.merlin ALQAKYGDYDIDSITHETYCSTDYLPKRVKDQFQ-MSEQMWGDKINEWYA 183
Ciona.intestinalis.merlin AVQAKFGDYDEEMHKPGFLANEILLPQQVRDQFQSVTGEMWETQITSWYA 191
Acorn.worm.merlin AVQAKFGDYDESVHLPGFLSEEDLLPQKVVDQYQ-MTSDMWEERITAWYA 183
Human.merlin AVQAKYGDYDPSVHKRGFLAQEELLPKRVINLYQ-MTPEMWEERITAWYA 193
Molgula.occulta.merlin QHRGLSRDEAELEYLKIARDLEMFGIKYFNITNVKGSKLWLGISPISVNV 234
Molgula.oculata.merlin ----------ELEYLKIARDLEMFGIKYFKITNVKGSNLWLGISPTAVNV 40
Molgula.tectiformis.merlin QHRGLTRDEAELEYLKIALDLEMFGVSLFKIKNNKGSELCLGINAVSVNV 233
Ciona.intestinalis.merlin QHHGLTRDEAELEYLKIVQEFEMSGVQYFKIKDGNGADLWLGIDAKSVSM 241
Acorn.worm.merlin QHKGMTRDEAEMEYLKIAQDLEMYGVNYFQIKNKKDTDLWLGVDALGLNI 233
Human.merlin EHRGRARDEAEMEYLKIAQDLEMYGVNYFAIRNKKGTELLLGVDALGLHI 243
Molgula.occulta.merlin YELENQLIPIVSFQWSELSDMSYNENKFTMKQSAHYHG-NGTLQRGTVHV 283
Molgula.oculata.merlin YELENQLIPIVSFQWSELSDMSYNENKFTMKQSAHYHG-NGTLQRGTVQV 89
Molgula.tectiformis.merlin YEPDNQLLPIVSFQWSELADMSFSDNKFVIKQSSLPTIRSGTLRRGATMS 283
Ciona.intestinalis.merlin YPYNDQLHPTKSYQWSELADMSYYGNKFVIKQTTRPNN-TNTIGR-RFNL 289
Acorn.worm.merlin YGLENQLAAKISFPWSEIKNISFHDKKFVIKPIDKKSP---------- 271
Human.merlin YDPENRLTPKISFPWNEIRNISYSDKEFTIKPLDKKID---------- 281
Molgula.occulta.merlin DSSRAGDTFAQSKDFVFYSEEPGINKLILELCRGNHDLFMKRRRVDSIEI 333
Molgula.oculata.merlin DSSRADDTFAQSKDFVFYTEEPGINKLILELCRGNHDLFMKRRRVDSIEI 139
Molgula.tectiformis.merlin ESS-ENSELQQNKDFVFFTDEPGVNKIILDLCRGNHDLFMKRRKVDSMEI 332
Ciona.intestinalis.merlin GASTSNLNVNAPEDIVFLVDDPEVNKLILDLCRGNHDLFMQRRRVDTMEI 339
Acorn.worm.merlin ----------DFIFFSQRLRVNKVILQLCVGNHELFMRRRKVDSMEI 308
Human.merlin ----------VFKFNSSKLRVNKLILQLCIGNHDLFMRRRKADSLEV 318
Molgula.occulta.merlin QQMKTQAKEEKSRKQAEKSRM---------- 354
Molgula.oculata.merlin QQMKTQAKEERSRKQVEKSRMAREKDLRRQI---------- 170
Molgula.tectiformis.merlin QQMKTQAKEEKARKLAEKHRLVRDKEQYEQIEMEREELKQRVKELQEESR 382
Ciona.intestinalis.merlin QQMKEQAREEKARKQMERTRLTKEKNLRLQVQNEKKELEEKLAQFQEENR 389
Acorn.worm.merlin QQMKGQAREEKARKQIERNKLSREKHLRELAEHEKEELQRQLRHFQEEAR 358

```
Human.merlin                    QQMKAQAREEKARKQMERQRLAREKQMREEAERTRDELERRLLQMKEEAT 368
                                **** **:**::**  *: ::

Molgula.occulta.merlin          --------------------------------------------------
Molgula.oculata.merlin          --------------------------------------------------
Molgula.tectiformis.merlin      MAMEALNRSEETAKLLAEKAQIAEEETVLLKKKANHAEQEMQRIKAEFAK 432
Ciona.intestinalis.merlin       SAADTLRRSEETAELLGEKAKVAEEEAQLLRAKLNKSDQEIQSLKLEIGK 439
Acorn.worm.merlin               YANEALMRSEETADLLGEKASVAEEEALLLAQKATEAEQETQRIKMQAIK 408
Human.merlin                    MANEALMRSEETADLLAEKAQITEEEAKLLAQKAAEAEQEMQRIKATAIR 418


Molgula.occulta.merlin          --------------------------------------------------
Molgula.oculata.merlin          --------------------------------------------------
Molgula.tectiformis.merlin      TQEANIQLQQNLQNYDQVTKQLAEDSKTNAAEAEKLRKELKRAMQAKNEA 482
Ciona.intestinalis.merlin       LQEANHQLPQKFLKYEQYVSQLNEQAQARAGELKNVREELYATKAALNDA 489
Acorn.worm.merlin               SEEEKMLMEQKALEAEMIASKIAKDSERRNKEADQLKEELIRAKEAEKSA 458
Human.merlin                    TEEEKRLMEQKVLEAEVLALKMAEESERRAKEADQLKQDLQEAREAERRA 468


Molgula.occulta.merlin          --------------------------------------------------
Molgula.oculata.merlin          --------------------------------------------------
Molgula.tectiformis.merlin      KEKLTNASMVNSNNTSTDMSYGHYIPDTPGMASTPLSNAMGESVMVTGAS 532
Ciona.intestinalis.merlin       NIKLN---LLANQHVHHSPNTSNSVPMN--MGFLPINNQIHSQVISMPG- 533
Acorn.worm.merlin               KNRLIEITRTTWTTQP---------------------------------- 474
Human.merlin                    KQKLLEIA--TKPTYP---------------------------------- 482


Molgula.occulta.merlin          --------------------------------------------------
Molgula.oculata.merlin          --------------------------------------------------
Molgula.tectiformis.merlin      VSPLQIQPLAPGVPIVSHKHIGSNGSKASNGSIGGANNGGIGQNGRPGNV 582
Ciona.intestinalis.merlin       -INLNNTSQSNHSPAVSN-AQYYHMYAASNGSAHSVKTDGSNVTDRNSNE 581
Acorn.worm.merlin               -------MYDVMSHDLSDLHLDTDLTANSEPACDLMTHG----------- 506
Human.merlin                    -------PMNPIPAPLPPDIPSFNLIGDS-LSFDFKDT------------ 512


Molgula.occulta.merlin          --------------------------------------------------
Molgula.oculata.merlin          --------------------------------------------------
Molgula.tectiformis.merlin      GMHKVGSNGGLQTIGGDYDDSNGDFARVSEMQHLSQEIEKERMEYQAKSK 632
Ciona.intestinalis.merlin       -----------------FTSAPGN----SDMQQLSQEIEKERMEYHVKSR 610
Acorn.worm.merlin               ----------------------------DVEQLSLEIEKERVEYMEKSK 527
Human.merlin                    ----------------------------DMKRLSMEIEKEKVEYMEKSK 533


Molgula.occulta.merlin          --------------------------------------------------
Molgula.oculata.merlin          --------------------------------------------------
Molgula.tectiformis.merlin      HIEQQLAMLKTEIEGLKVDEKMTPLDHMYTENSLKG-AVKYQSLNQAKES 681
Ciona.intestinalis.merlin       NIEQQLFNLRSEIEVLKVDESMTGFDQK--QDSNQP-HTHEISTFQGHKE 657
Acorn.worm.merlin               HLQAQLNDLKTEIEVLKIEEKQTHLDQLHNSMVQKG-DTKFSTLKKIKAG 576
Human.merlin                    HLQEQLNELKTEIEALKLKERETALDILHNENSDRGGSSKHNTIKKLPLQ 583


Molgula.occulta.merlin          ------------
Molgula.oculata.merlin          ------------
Molgula.tectiformis.merlin      TPEQRLQLYNKL 693
Ciona.intestinalis.merlin       TPQYYDGL---- 665
Acorn.worm.merlin               TTRARVAFFEEL 588
Human.merlin                    SAKSRVAFFEEL 595
```

Figure 6

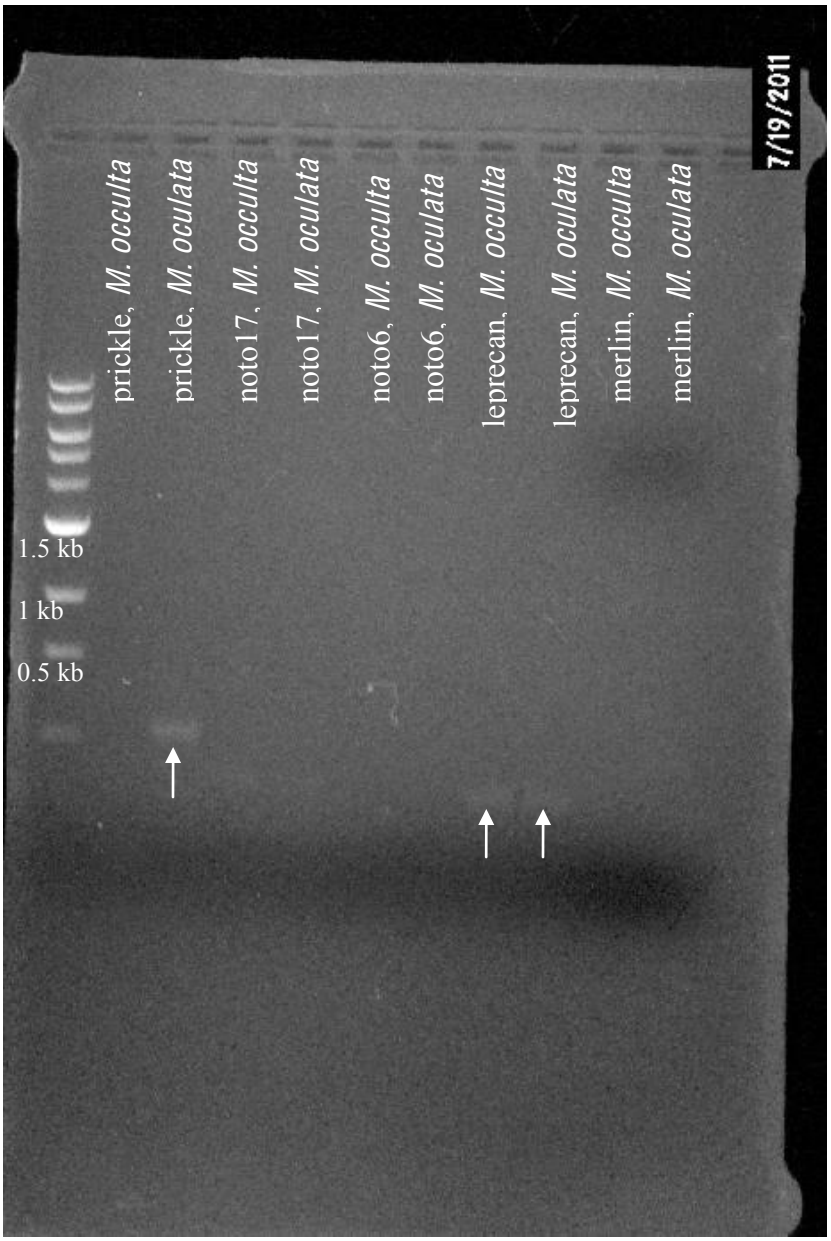

7/19/2011
prickle, M. occulta
prickle, M. oculata
noto17, M. occulta
noto17, M. oculata
noto6, M. occulta
noto6, M. oculata
leprecan, M. occulta
leprecan, M. oculata
merlin, M. occulta
merlin, M. oculata
1.5 kb
1 kb
0.5 kb

Figure 7

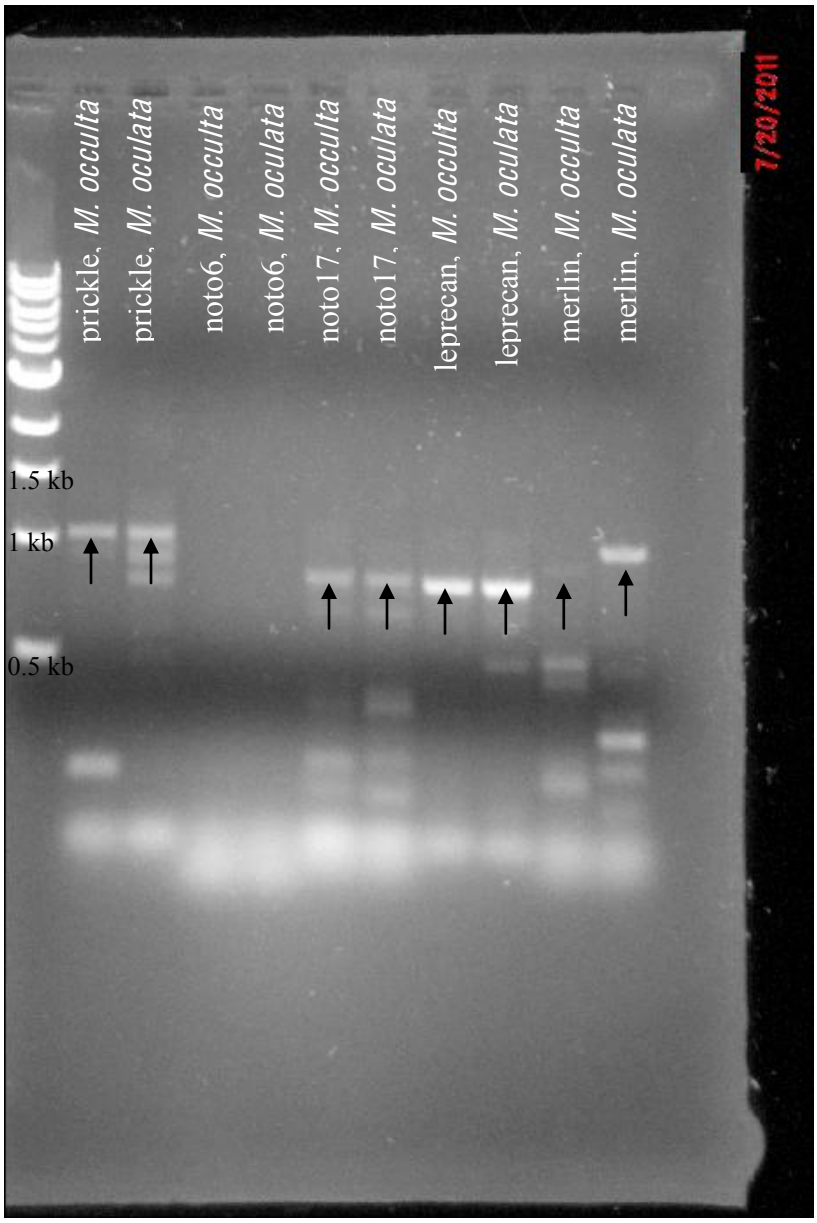

prickle, M. occulta
prickle, M. oculata
noto6, M. occulta
noto6, M. oculata
noto17, M. occulta
noto17, M. oculata
leprecan, M. occulta
leprecan, M. oculata
merlin, M. occulta
merlin, M. oculata
1.5 kb
1 kb
0.5 kb
7/20/2011

Figure 8

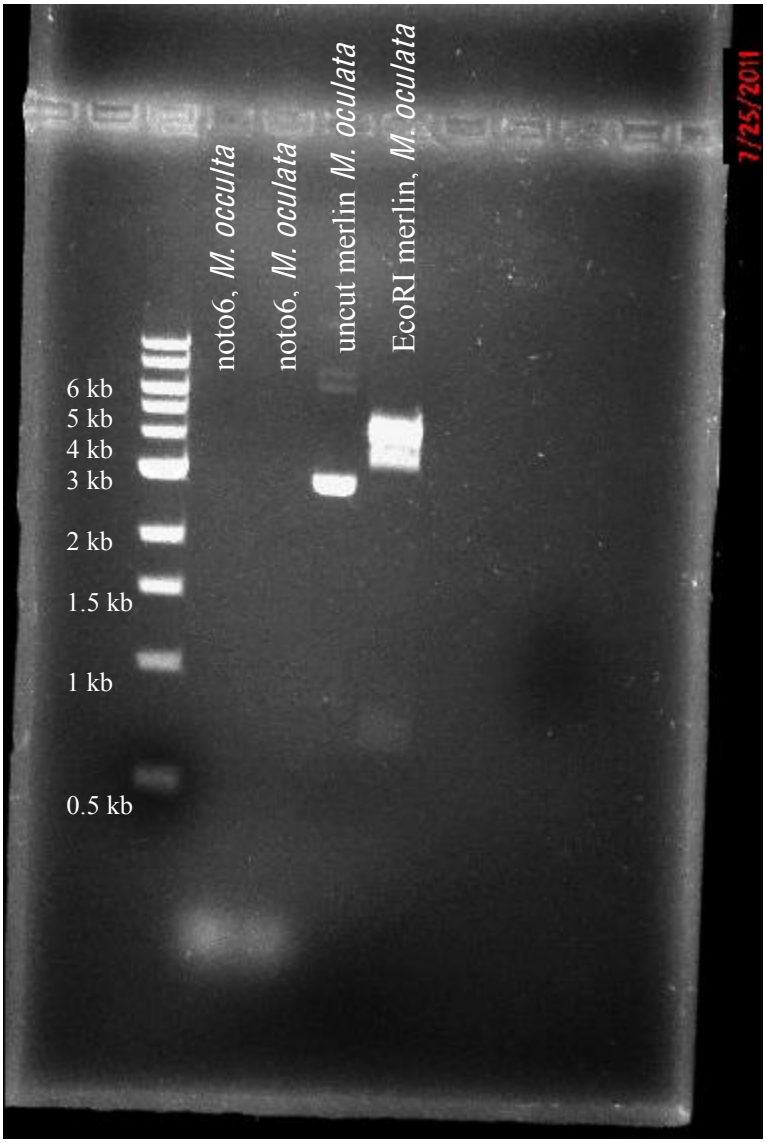
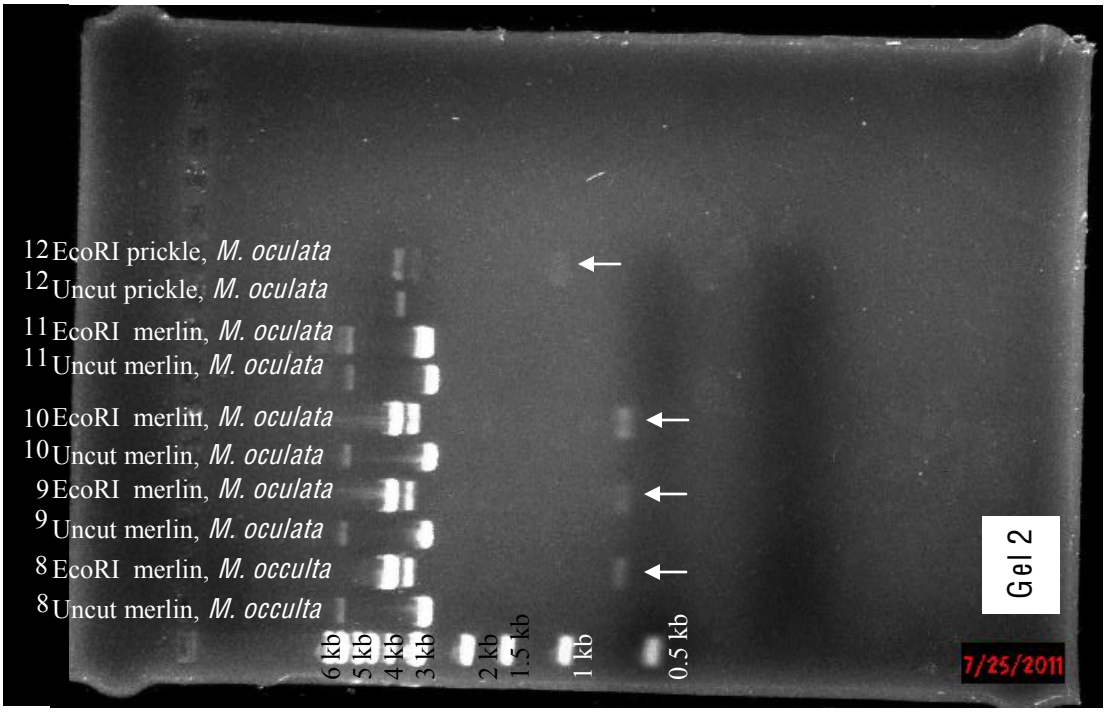

noto6, M. occulta
noto6, M. oculata
uncut merlin M. oculata
EcoRI merlin, M. oculata
6 kb
5 kb
4 kb
3 kb
2 kb
1.5 kb
1 kb
0.5 kb
7/25/2011

Figure 9


12 EcoRI prickle, M. oculata
12 Uncut prickle, M. oculata
11 EcoRI merlin, M. oculata
11 Uncut merlin, M. oculata
10 EcoRI merlin, M. oculata
10 Uncut merlin, M. oculata
9 EcoRI merlin, M. oculata
9 Uncut merlin, M. oculata
8 EcoRI merlin, M. occulta
8 Uncut merlin, M. occulta
6 kb
5 kb
4 kb
3 kb
2 kb
1.5 kb
1 kb
0.5 kb
Gel 2
7/25/2011

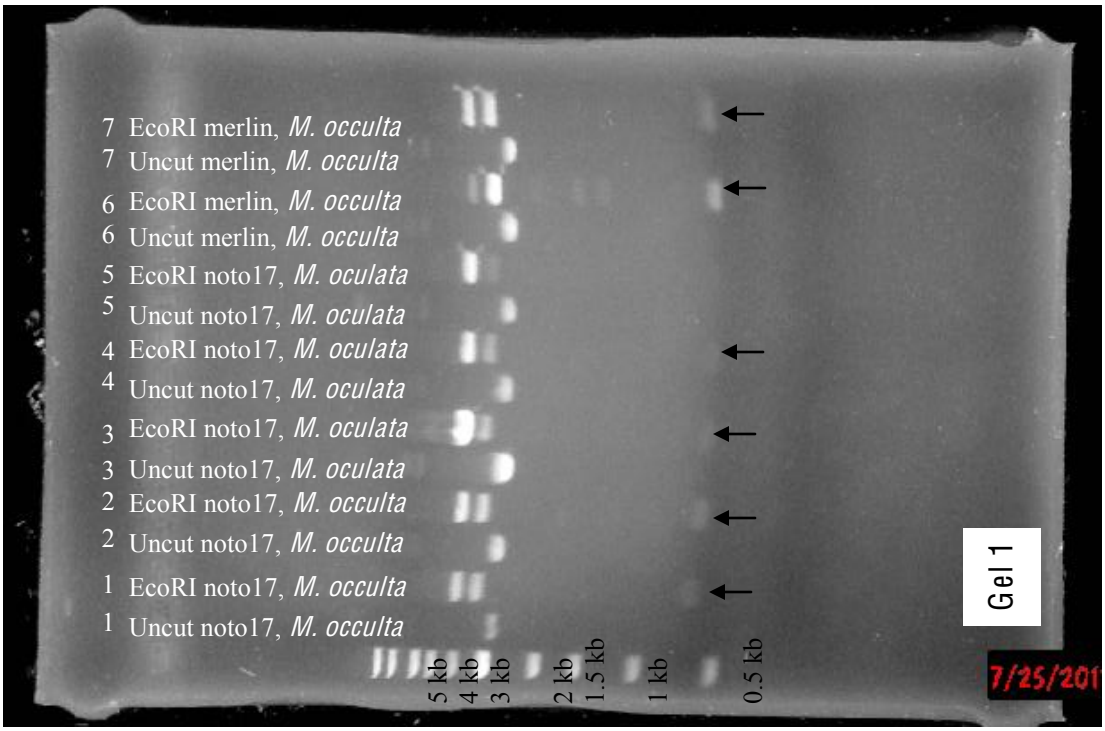
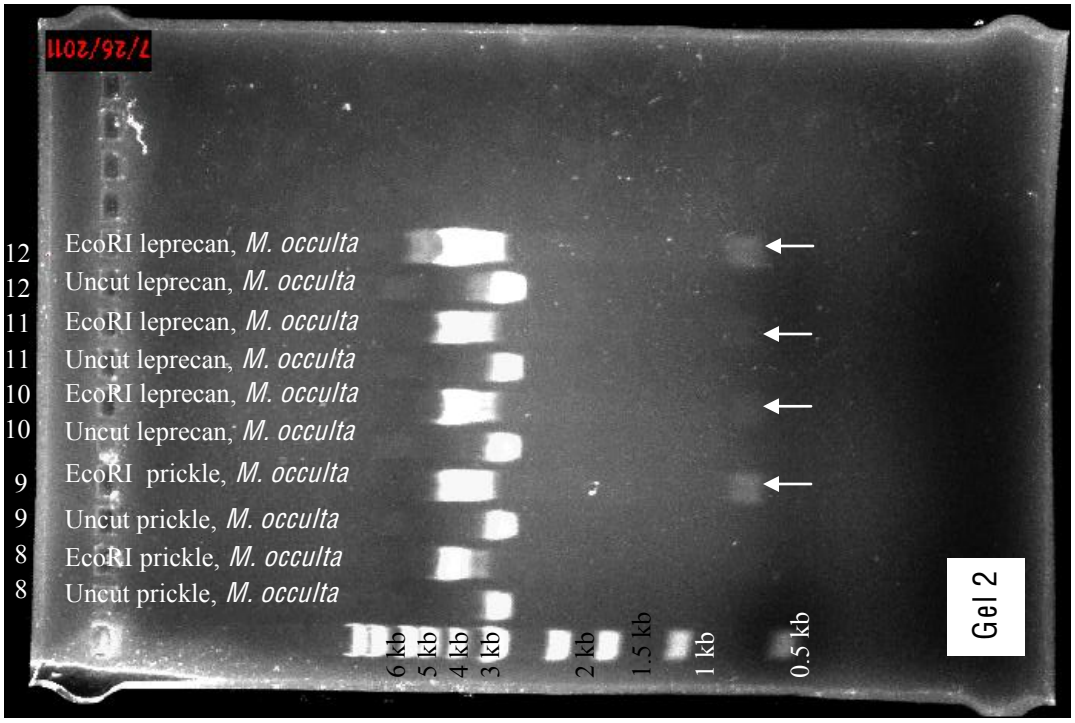

7 EcoRI merlin, M. occulta
7 Uncut merlin, M. occulta
6 EcoRI merlin, M. occulta
6 Uncut merlin, M. occulta
5 EcoRI noto17, M. oculata
5 Uncut noto17, M. oculata
4 EcoRI noto17, M. oculata
4 Uncut noto17, M. oculata
3 EcoRI noto17, M. oculata
3 Uncut noto17, M. oculata
2 EcoRI noto17, M. occulta
2 Uncut noto17, M. occulta
1 EcoRI noto17, M. occulta
1 Uncut noto17, M. occulta
5 kb
4 kb
3 kb
2 kb
1.5 kb
1 kb
0.5 kb
Gel 1
7/25/201

Figure 10


7/26/2011
12 EcoRI leprecan, M. occulta
12 Uncut leprecan, M. occulta
11 EcoRI leprecan, M. occulta
11 Uncut leprecan, M. occulta
10 EcoRI leprecan, M. occulta
10 Uncut leprecan, M. occulta
9 EcoRI prickle, M. occulta
9 Uncut prickle, M. occulta
8 EcoRI prickle, M. occulta
8 Uncut prickle, M. occulta
6 kb
5 kb
4 kb
3 kb
2 kb
1.5 kb
1 kb
0.5 kb
Gel 2

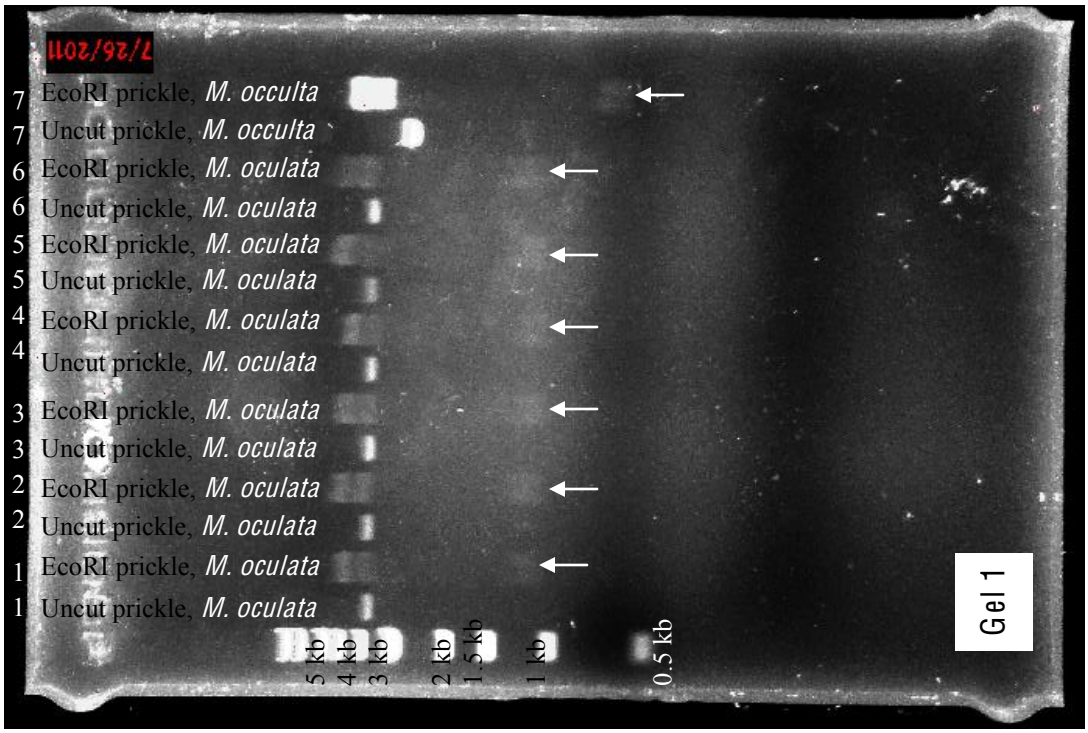

7/26/2011
7 EcoRI prickle, M. occulta
7 Uncut prickle, M. occulta
6 EcoRI prickle, M. oculata
6 Uncut prickle, M. oculata
5 EcoRI prickle, M. oculata
5 Uncut prickle, M. oculata
4 EcoRI prickle, M. oculata
4 Uncut prickle, M. oculata
3 EcoRI prickle, M. oculata
3 Uncut prickle, M. oculata
2 EcoRI prickle, M. oculata
2 Uncut prickle, M. oculata
1 EcoRI prickle, M. oculata
1 Uncut prickle, M. oculata
5 kb
4 kb
3 kb
2 kb
1.5 kb
1 kb
0.5 kb
Gel 1

Figure 11

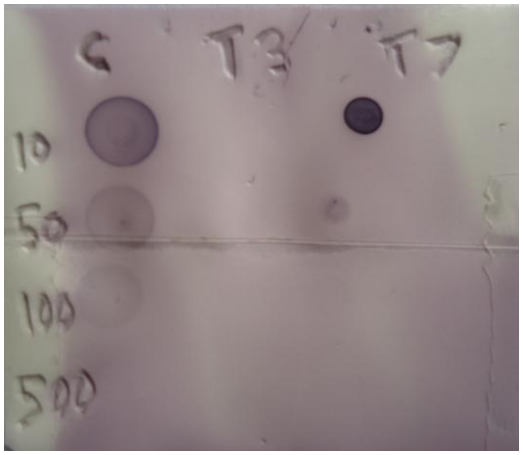


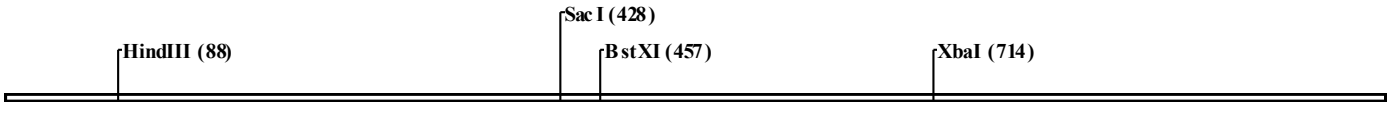
Figure 12

Mapping all cutsites.

Cutters : BamHI, HindIII, KpnI, PstI & XbaI

Non-Cutters : ApaI, Bsp106, BstXI, DraII, EcoRI, EcoRV, NotI, SacI, SacII, SalI, SmaI, SpeI, XhoI & XmaIII

Figure 13



Mapping all cutsites.

Cutters : BstXI, HindIII, SacI & XbaI

Non-Cutters : ApaI, BamHI, Bsp106, DraII, EcoRI, EcoRV, KpnI, NotI, PstI, SacII, SalI, SmaI, SpeI, XhoI & XmaIII