

Developmental Studies of Appendage Patterning and Formation in Spiders

Dissertation
for the award of the degree
“Doctor rerum naturalium”
of the Georg-August-University Göttingen

within the doctoral program Genes and Development of the Göttingen Graduate
School for Neurosciences, Biophysics, and Molecular Biosciences
Georg-August University Göttingen

submitted by
Natascha Zhang
Holzminden

Göttingen 2016

Thesis Committee

Dr. Nikola-Michael Prpic-Schäper (reviewer and advisor),
Department Developmental Biology,
Georg-August-University Göttingen

Prof. Dr. Martin Göpfert (reviewer),
Department Cellular Neurobiology,
Georg-August-University Göttingen

Prof. Dr. Andreas Wodarz
Institute I of Anatomy
University of Cologne

Members of the Examination Board

Referee:

Dr. Nikola-Michael Prpic-Schäper,
Department Developmental Biology,
Georg-August-University Göttingen

2nd Referee:

Prof. Dr. Martin Göpfert ,
Department Cellular Neurobiology,
Georg-August-University Göttingen

Further members of the Examination Board

Prof. Dr. Andreas Wodarz
Institute I of Anatomy
University of Cologne

Dr. Gerd Vorbrüggen,
Guest Group Molecular Cell Dynamics
Max Planck Institute for Biophysical Chemistry Göttingen

Prof. Dr. Daniel J. Jackson
Evolution of Metazoa
Courant Research Centre Geobiology
Georg-August-University Göttingen

Dr. Sven Bradler
Department Morphology, Systematic, Evolutionary biology
Georg-August-University Göttingen

Date of oral examination: 17.03.2016

Declaration

Herewith I ensure, that the thesis “Developmental Studies of Appendage Patterning and Formation in Spiders” has been written independently and with no other sources and aids than quoted. Apart from the publications listed below the thesis has not been published so far. I also declare that I have not previously applied for a doctoral degree at another university.

Göttingen, January 31th, 2016

Natascha Zhang

Publications that are part of this thesis:

Turetzek, N., Pechmann, M., Schomburg, C., Schneider, J., and Prpic, N.-M. (2016). Neofunctionalization of a Duplicate *dachshund* Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. *Mol Biol Evol* 33, 109–121.

Schomburg, C., Turetzek, N., Schacht, M.I., Schneider, J., Kirfel, P., Prpic, N.-M., and Posnien, N. (2015). Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *EvoDevo* 6, 15.

Janssen, R., Schönauer, A., Weber, M., Turetzek, N., Hogvall, M., Goss, G.E., Patel, N.H., McGregor, A.P., and Hilbrant, M. (2015). The evolution and expression of panarthropod frizzled genes. *Frontiers in Ecology and Evolution* 3, 96.

Pechmann, M., Khadjeh, S., Turetzek, N., McGregor, A.P., Damen, W.G.M., and Prpic, N.-M. (2011). Novel Function of *Distal-less* as a Gap Gene during Spider Segmentation. *PLoS Genetics* 7, e1002342.

Manuscript in preparation:

Turetzek, N., Prpic, N –M. Observations on germ band development in the cellar spider *Pholcus phalangioides*. In preparation.

Für meine Tochter
Reya Eliza Zhang

“Tu die Dinge die du tust, weil du sie liebst und
nicht weil ein Anderer dich dafür lieben könnte”

Torben Nehrhoff 2009

Danksagung

Als Erstes danke ich meinem Doktorvater Nikola-Michael Prpic-Schäper für die wissenschaftliche Betreuung meiner Doktorarbeit. Ich danke für eine stets harmonische und sehr produktive Zusammenarbeit. Ganz besonders dankbar bin ich für seine Unterstützung, seine Ermutigung mir auch mal was zuzutrauen und sein großes Engagement. Schon während der Anfertigung meiner Masterarbeit hat er mich immer darin unterstützt eigenen Ideen nachzugehen. Vielen Dank für ein offenes Ohr und die vielen interessanten wissenschaftlichen Diskussionen.

Als nächstes möchte ich mich herzlich bei Nico Posnien für die Betreuung des Transkriptomprojektes bedanken. Aber auch für zahlreiche Diskussion über mein Projekt und generelle Fragestellungen der Evolutionsbiologie, welche mich immer wieder dazu gebracht haben meine Arbeit neu zu fokussieren, bin ich sehr dankbar.

Ich danke den Mitgliedern meines Thesis Committes, Andreas Wodarz und Martin Göpfert, für richtungsweisende Kommentare und wissenschaftliche Diskussion meiner Daten.

Ein ganz besonderer Dank geht auch an meine Kollegen aus Labor 2 und mittlerweile sehr guten Freunde Christoph und Montse für ihre nette Aufnahme nach meinem Umzug, die zahlreichen Diskussionen und ihre Unterstützung bei bioinformatischen Fragen. Aber auch Felix danke ich für ein offenes Ohr und sein Engagement mir bei der Optimierung der mikroskopischen Aufnahmen zu helfen.

Desweiteren möchte ich mich bei Ernst Wimmer und Gregor Bucher bedanken. Ihr Rat, Diskussionen und die kritische Hinterfragung der präsentierten Daten in zahlreichen Progress reports hat mich immer wieder dazu angehalten mein Projekt aus anderen Blickwinkeln zu sehen und auch damit dazu beigetragen die Projekte voranzubringen.

Nicht zu vergessen danke ich Sara und Matthias für ihre Einführung in das Arbeiten mit den Spinnen und vielen wissenschaftlichen Diskussionen. Ein ganz besonderer Dank geht auch ein meine lieben Hiwis, Julia Schneider und Julia Golldamm, die mir bei der

Anfertigung zahlreicher Färbungen helfend zur Seite standen und ohne die, die Arbeit niemals diesen Umfang hätte annehmen können.

Ein großer Dank geht an die ganze Abteilung der Entwicklungsbiologie für ein sehr harmonisches Umfeld und Unterstützung bei jeglichen Dingen, vor allem Birgit, Angelika, Elke und Beate möchte ich herzlich danken. Während der Jahre habe ich in dieser Abteilung viele Menschen kennen gelernt die mir sehr ans Herz gewachsen sind und auch dafür bedanke ich mich. Wie schon genannt grade Niko, Nico, Montse, Christoph und Felix aber auch Dawid, Dippel, Kolja, Peter und Janna haben immer dazu beigetragen den Laboralltag interessant zu gestalten und noch mehr Spaß an der Arbeit zu haben.

Nicht zuletzt danke ich meinen guten Freunden, die mir während der ganzen Zeit den Rücken gestärkt, immer an mich geglaubt und mir Mut gemacht haben: Laura, Conny, Kathi, Marie und vor allem Torben, ich bin euch unfassbar dankbar!

Schlußendlich danke ich meiner Familie für ihre Geduld und ihre Liebe während der ganzen Jahre.

Am meisten danke ich meiner Tochter, dass sie mich immer wieder auf so wundervolle Weise in den Alltag zurück geholt hat.

TABLE OF CONTENTS

1. SUMMARY	1
2. GENERAL INTRODUCTION	3
2.1 MORPHOLOGICAL DIVERSITY	3
2.2 CHELICERATE MODEL ORGANISMS	6
2.3 APPENDAGE DEVELOPMENT.....	9
2.3.1 <i>Proximo-distal axis formation in arthropods</i>	13
2.3.2 <i>Tarsus segmentation gene network</i>	15
3. MATERIALS AND METHODS.....	19
3.1 ANIMAL CULTURE	19
3.2 IDENTIFICATION OF SPIDER ORTHOLOGS AND PHYLOGENETIC ANALYSIS	19
3.3 MOLECULAR METHODS	20
3.3.1 <i>RNA isolation</i>	20
3.3.1 <i>cDNA synthesis</i>	20
3.3.2 <i>PCR</i>	21
3.3.3 <i>Cloning, transformation and plasmid preparation</i>	24
3.3.1 <i>in vitro transcription of RNA probes and dsRNA for RNAi</i>	25
3.4 EMBRYOLOGICAL METHODS	25
3.4.1 <i>Embryo fixation and cuticle preparations</i>	25
3.4.2 <i>Whole mount in situ hybridization</i>	27
3.4.3 <i>Live imaging of embryonic development of P. phalangoides</i>	28
3.4.4 <i>Parental RNAi</i>	28
3.5 RNA SEQ, NGS AND <i>DE NOVO</i> TRANSCRIPTOME ASSEMBLY AND ANALYSIS.....	32
3.5.1 <i>Read mapping and reduction of transcriptome complexity</i>	32
3.5.2 <i>Differential gene expression analysis with DESeq</i>	33
4. PHOLCUS PHALANGIODES: A HAPLOGYNE SPIDER MODEL.....	35
4.1 OBSERVATIONS ON GERM BAND DEVELOPMENT IN THE CELLAR SPIDER <i>PHOLCUS PHALANGIODES</i>	36
4.1.1 <i>Formation of the germ band in P. phalangoides</i>	37
4.1.1 <i>Segmentation and further differentiation of the germ band</i>	39
4.1.1 <i>Development of the opisthosoma</i>	40
4.1.2 <i>Inversion and dorsal closure</i>	42
4.1.3 <i>Growth of the legs</i>	43

4.2	GENERATING TRANSCRIPTOMIC RESOURCES FOR <i>P. PHALANGIOIDES</i>	45
4.2.1	<i>de novo</i> transcriptome assembly and differential gene expression.....	45
4.2.2	<i>Leg</i> candidate genes	50
4.3	DISCUSSION AND FUTURE WORK.....	57
4.3.1	<i>Two early waves of cell dispersal</i>	57
4.3.2	<i>"Preabdomen-postabdomen" subdivision of the opisthosoma</i>	58
4.3.3	<i>Extreme leg length in pholcid spiders</i>	59
4.3.4	<i>Conserved developmental steps for comparative expressional studies</i>	59
4.3.5	<i>General transcriptome optimization</i>	60
4.3.6	<i>Up-regulated genes in the walking legs of P. phalangioides</i>	61
5.	LEG GAP GENES IN APPENDAGE DEVELOPMENT OF SPIDERS	66
5.1	HOMOTHORAX AND EXTRADENTICLE	66
5.1.1	<i>Duplicated hth and genes in P. tepidariorum and P. phalangioides</i>	67
5.1.2	<i>Phylogenetic analysis reveals duplication of hth in spiders</i>	69
5.1.3	<i>Expression of hth1 and hth2 during P. tepidariorum and P. phalangioides</i> <i>embryogenesis</i>	69
5.1.4	<i>hth</i> expression in the prosomal appendages of spiders.....	73
5.1.5	<i>Duplication of exd genes in P. tepidariorum and P. phalangioides</i>	78
5.1.6	<i>Phylogenetic analysis of exd homologs in spiders and chelicerates</i>	79
5.1.7	<i>Expression of exd during embryonic development of P. tepidariorum and P.</i> <i>phalangioides</i>	80
5.1.8	<i>Expression of exd in prosomal appendages</i>	83
5.2	NEOFUNCTIONALISATION OF A DUPLICATE DACHSHUND GENE UNDERLIES THE EVOLUTION OF A NOVEL LEG SEGMENT IN ARACHNIDS.....	87
5.2.1	<i>Introduction</i>	87
5.2.2	<i>Duplicated dachshund genes in spiders</i>	89
5.2.3	<i>Phylogenetic analysis of dachshund gene evolution and duplication</i>	91
5.2.4	<i>Expression of dac1 in Parasteatoda tepidariorum and Pholcus phalangioides</i>	92
5.2.5	<i>Expression of dac2 in Parasteatoda tepidariorum and Pholcus phalangioides</i>	94
5.2.6	<i>Expression profile of dac1 and dac2 in the prosomal appendages</i>	97
5.2.7	<i>dac2</i> is required for morphogenesis of the patella leg segment in <i>P. tepidariorum</i>	100
5.3	DISCUSSION AND FUTURE WORK.....	102
5.3.1	<i>Gene duplication of spider dachshund genes</i>	102

5.3.2	<i>Neofunctionalisation of dac2 in the arachnids</i>	102
5.3.3	<i>Evolution of the arachnid patella: neofunctionalisation of dac2 underlies the evolution of a morphological novelty</i>	103
5.3.4	<i>Conservation of exd after duplication</i>	105
5.3.5	<i>Spider specific hth duplication</i>	106
5.3.6	<i>Rapid diversification of hth2 after neofunctionalization</i>	107
5.3.7	<i>Leg gap gene conservation and duplication</i>	109
6.	TARSUS SEGMENTATION CANDIDATE GENES IN SPIDERS	110
6.1	<i>APTEROUS IN P. TEPIDARIORUM AND P. PHALANGIOIDES</i>	110
6.2	<i>LIM1 IN P. TEPIDARIORUM AND P. PHALANGIOIDES</i>	114
6.3	<i>SPINELESS IN P. TEPIDARIORUM AND P. PHALANGIOIDES</i>	117
6.4	<i>TRACHEALESS IN P. TEPIDARIORUM AND P. PHALANGIOIDES</i>	121
6.5	<i>ARISTALESS IN P. TEPIDARIORUM AND P. PHALANGIOIDES</i>	125
6.6	<i>CLAWLESS</i>	129
6.6.1	<i>Phylogenetic analysis of clawless duplicates in spider</i>	129
6.6.2	<i>Expression of clawless</i>	131
6.6.3	<i>Expression of clawless in prosomal appendages</i>	135
6.6.4	<i>Function of cll1 in P. tepidariorum</i>	140
6.7	<i>BARH IN P. TEPIDARIORUM AND P. PHALANGIOIDES</i>	147
6.7.1	<i>Phylogenetic analysis of Bar in P. tepidariorum and P. phalangioides</i>	147
6.7.2	<i>Expression of BarH homologs in P. tepidariorum</i>	149
6.7.3	<i>Function of BarH1 in P. tepidariorum</i>	150
6.7.4	<i>Expression of head and leg developmental genes in Pt-BarH1 RNAi embryos</i>	155
6.8	<i>DISCUSSION AND FUTURE WORK</i>	158
6.8.1	<i>Distal limb development is not conserved for insects and spiders</i>	158
6.8.2	<i>Cross-regulation of al, cll and Lim1 is not conserved in arthropods</i>	159
6.8.3	<i>Conservation of distal function of clawless</i>	161
6.8.4	<i>Other conserved functions of tarsus segmentation genes</i>	162
6.8.5	<i>Possible role for ss and trh in growth regulation during late embryogenesis</i>	164
6.8.6	<i>Novel role for BarH1 in head patterning of P. tepidariorum</i>	165
7.	GENERAL DISCUSSION	167
7.1	<i>HOURLASS MODEL OF WALKING LEG DEVELOPMENT</i>	167
7.2	<i>DUPLICATION OF GENES FACILITATES DIVERSIFICATION</i>	171

8. REFERENCES	174
9. REFERENCES OF PUBLICATIONS AND MANUSCRIPTS.....	191
9.1 TURETZEK ET AL. (2016). NEOFUNCTIONALIZATION OF A DUPLICATE <i>DACHSHUND</i> GENE UNDERLIES THE EVOLUTION OF A NOVEL LEG SEGMENT IN ARACHNIDS.	191
9.2 TURETZEK AND PRPIC IN PREPARATION. OBSERVATIONS ON GERM BAND DEVELOPMENT IN THE CELLAR SPIDER <i>PHOLCUS PHALANGIOIDES</i>	194
10. ACKNOWLEDGMENTS.....	196
11. SUPPLEMENTAL MATERIAL	1
11.1 ENGRAILED	1
11.2 TIMING OF EMBRYONIC DEVELOPMENT OF <i>P. PHALANGIOIDES</i>	3
11.3 RNA SEQ GENES	4
11.3.1 <i>Evi1</i>	4
11.3.2 <i>c105514</i>	5
11.3.3 <i>c95092</i>	5
11.3.4 <i>c109329</i>	6
11.3.5 <i>c109662</i>	6
11.3.6 <i>oddskiped</i>	7
11.4 LEG GAP GENES.....	9
11.4.1 <i>extradenticle</i>	9
11.4.2 <i>homothorax</i>	20
11.4.3 <i>dachshund</i>	28
11.5 TARSUS SEGMENTATION GENES.....	34
11.5.1 <i>aristaless</i>	34
11.5.2 <i>Lim</i>	38
11.5.3 <i>apterous</i>	43
11.5.4 <i>spineless</i>	49
11.5.5 <i>trachealess</i>	56
11.5.6 <i>clawless</i>	60
11.5.7 <i>BarH</i>	65

List of Figures

Figure 1: Phylogeny of Arthropoda	3
Figure 2: Diversity of arthropod appendages	4
Figure 3: Spider models used in this study	7
Figure 4: Simplified leg patterning cascade of <i>D. melanogaster</i>	12
Figure 5: Conservation of leg gap genes in panarthropods and spiders.....	14
Figure 6: Overview of PD leg patterning in <i>D. melanogaster</i> and <i>T. castaneum</i>	17
Figure 7: Consecutive stages of germ band formation and differentiation in <i>P. phalangioides</i>	38
Figure 8: Expression of <i>engrailed (en)</i> in the germ band of <i>P. phalangioides</i>	41
Figure 9: Formation of the tail.....	42
Figure 10: Formation of the opisthosoma by folding down tail segments on the yolk.....	43
Figure 11: Allometric growth of the legs in <i>P. phalangioides</i> , and ventral closure	44
Figure 12: Overview <i>de novo</i> transcriptome assembly.....	46
Figure 13: Comparison of gene and transcript numbers after corset.....	47
Figure 14: Dependence of transcript number and transcript length	48
Figure 15: Principal component analysis of the triplicates for the different sample tissues	49
Figure 17: Expression of <i>c102331 Pp-odd skipped</i> in <i>P. phalangioides</i>	53
Figure 18: Expression of <i>c109624</i> in <i>P. phalangioides</i>	54
Figure 19: Expression of <i>c95092</i> in <i>P. phalangioides</i>	55
Figure 20: Expression of <i>c105514</i> and <i>c109329</i> in <i>P. phalangioides</i>	55
Figure 21: Expression of <i>c109662</i> in <i>P. phalangioides</i>	56
Figure 22: Unrooted 50% phylogenetic majority rule consensus tree of <i>hth</i> orthologs ...	70
Figure 23: Expression of <i>hth</i> paralogs in <i>P. tepidariorum</i>	71
Figure 24: Expression of <i>hth</i> paralogs in <i>P. phalangioides</i>	72
Figure 25: Expression of <i>hth</i> paralogs in prosomal appendages of <i>P. tepidariorum</i>	75
Figure 26: Expression of <i>hth</i> paralogs in prosomal appendages of <i>P. phalangioides</i>	76
Figure 27: Expression of <i>hth2</i> in prosomal appendages of <i>C. salei</i>	77
Figure 28: Expression of <i>exd1</i> in <i>P. tepidariorum</i>	81

Figure 29: Expression <i>exd1</i> in <i>P. phalangioides</i>	82
Figure 30: Expression of <i>exd2</i> in <i>P. tepidariorum</i>	82
Figure 31: Expression <i>exd2</i> in <i>P. phalangioides</i>	83
Figure 32: <i>exd</i> expression in prosomal appendages of <i>P. phalangioides</i>	85
Figure 33: <i>exd</i> expression in prosomal appendages of <i>P. tepidariorum</i>	86
Figure 34: Alignment of the duplicated <i>dac</i> genes from <i>P. tepidariorum</i> (abbreviated PTEP in the figure) and <i>P. phalangioides</i> (abbreviated PPHA in the figure).....	90
Figure 35: Unrooted 50% majority rule consensus tree after Bayesian Markov chain Monte Carlo analysis.....	92
Figure 36: Expression of <i>dac1</i> in <i>P. tepidariorum</i>	94
Figure 37: Expression of <i>dac1</i> in <i>P. phalangioides</i>	95
Figure 38: Expression of <i>dac2</i> in <i>P. tepidariorum</i>	96
Figure 39: Expression of <i>dac2</i> in <i>P. phalangioides</i>	96
Figure 40: Detection of <i>dac1</i> (A-F) and <i>dac2</i> (G-L) expression in dissected appendages of <i>P. tepidariorum</i>	98
Figure 41: Detection of <i>dac1</i> (A-F) and <i>dac2</i> (G-L) expression in dissected appendages of <i>P. phalangioides</i>	99
Figure 42: Malformation of the patella after <i>dac2</i> RNAi in <i>P. tepidariorum</i> . (A-D) Wild type nymph.....	101
Figure 43: Evolution of a novel leg segment in arachnids linked to <i>dac</i> duplication and neofunctionalisation.	104
Figure 44: Overview spider <i>hth2</i> walking leg expression.....	108
Figure 45: Expression of <i>ap1</i> in <i>P. tepidariorum</i>	111
Figure 46: Expression of <i>ap1</i> in <i>P. phalangioides</i>	112
Figure 47: Expression of <i>ap2.2</i> in <i>P. phalangioides</i>	114
Figure 48: Expression of <i>Lim1</i> paralogs during embryonic development of <i>P. tepidariorum</i>	115
Figure 49: Expression of <i>Lim1.2</i> in <i>P. phalangioides</i>	116
Figure 50: Expression of <i>spineless</i> paralogs in <i>P. tepidariorum</i>	118
Figure 51: Expression of <i>spineless</i> paralogs in <i>P. phalangioides</i>	120
Figure 52: Expression of <i>trh1</i> in <i>P. tepidariorum</i>	122

Figure 53: Expression of <i>trh1</i> in <i>P. phalangioides</i>	122
Figure 54: Expression of <i>trh2</i> in <i>P. tepidariorum</i>	124
Figure 55: Expression of <i>trh2</i> in <i>P. phalangioides</i>	125
Figure 56: Expression of <i>aristaless</i> in <i>P. tepidariorum</i>	127
Figure 57: Expression of <i>aristaless</i> in <i>P. phalangioides</i>	128
Figure 58: Unrooted 50% phylogenetic majority rule consensus tree of <i>c11</i> orthologs ...	130
Figure 59: Expression of <i>c11</i> in <i>P. tepidariorum</i>	132
Figure 60: Expression of <i>c11</i> in <i>P. phalangioides</i>	133
Figure 61: Expression of <i>c12</i> in <i>P. tepidariorum</i>	134
Figure 62: Expression of <i>c12</i> in <i>P. phalangioides</i>	135
Figure 63: Expression of <i>c11</i> in prosomal appendages of <i>P. tepidariorum</i>	136
Figure 64: Expression of <i>c11</i> in prosomal appendages of <i>P. phalangioides</i>	137
Figure 65: Expression of <i>c12</i> in prosomal appendages of <i>P. tepidariorum</i>	139
Figure 66: Expression of <i>c12</i> in prosomal appendages of <i>P. phalangioides</i>	140
Figure 67: Overview of phenotypes in the different cocoon after <i>c11</i> RNAi and control injection	142
Figure 68: Overview of phenotype in nymphs of <i>P. tepidariorum</i> after <i>c11</i> RNAi	143
Figure 69: Weak <i>Pt-c11</i> RNAi phenotype in the walking legs.....	144
Figure 70: Strong <i>Pt-c11</i> RNAi phenotype in the walking legs	145
Figure 71: <i>Pt-c11</i> RNAi effect on nymphal pedipalps.....	146
Figure 72: Expression of <i>BarH1</i> throughout embryonic development of <i>P. tepidariorum</i>	151
Figure 73: Expression of <i>BarH1</i> in the head of <i>P. tepidariorum</i>	151
Figure 74: Expression of <i>BarH2</i> in <i>P. tepidariorum</i>	152
Figure 75: Phenotype of <i>Pt-BarH1</i> knockdown in the different cocoons.....	153
Figure 76: Embryonic and nymphal phenotypes from <i>Pt-BarH1</i> knockdown	154
Figure 77: Overview of embryonic phenotypes in cocoon 2-4 after <i>Pt-BarH1</i> knockdown	155
Figure 78: Expression of <i>otd</i> , <i>Dll</i> and <i>Scr</i> in embryos after <i>Pt-BarH1</i> RNAi.....	157
Figure 79: Hourglass model of walking leg diversification development.....	171

List of Tables

Table 1: Spider orthologs.....	21
Table 2: Standard oligonucleotides	22
Table 3: Oligonucleotides for RACE-PCR.....	22
Table 4: Gene specific PCR oligonucleotides.....	23
Table 5: List of gene clones used in this study	26
Table 6: dsRNA injection procedure for RNAi.....	29
Table 7: Differently expressed genes in the different tissues sequenced for <i>P.</i> <i>phalangioides</i>	50
Table 8: Filtering criteria to reduce the number of up-regulated leg gene candidates.....	50
Table 9: Final candidates and their annotation	51

1. Summary

In the animal kingdom arthropods are unparalleled in species number and diversity. Especially their appendages show a great morphological diversity and are adapted for different purposes like feeding, walking, flying, swimming, breathing and courtship. The segmentation of the appendages into podomers is one characteristic trait shared by all extant arthropod groups and, at least in part, might be the key to their evolutionary success. Comparative studies of the genetic mechanisms involved in the formation and patterning of the appendages in different representatives of the four major arthropod clades is important to understand the evolution of this diversity. After the specification of the appendage primordia, appendage patterning in general requires the establishment of an additional axis, the proximo-distal axis. The formation of this axis is highly conserved in the arthropods, including *homothorax/extradenticle* defining the proximal, *dachshund* the medial and *Distal-less* the distal leg area. Further mechanisms of proximo-distal patterning involved in tarsus patterning are largely conserved in the insects, but not much is known for arthropods outside the insect clade. To study these processes in non-insect arthropods, I used the main spider model *Parasteatoda tepidariorum* which has an undivided tarsus and the long-legged cellar spider *Pholcus phalangioides*, which shows tarsus segmentation that is morphologically similar to the insect tarsus. The conserved leg gap genes were already studied in two other spider species and *homothorax* and *extradenticle* were found to be present as duplicates. I studied homologs of all leg gap genes in *Parasteatoda tepidariorum* and *Pholcus phalangioides* to clarify the origin of the duplication event and their impact on appendage patterning differences leading to morphological diversification. My data suggest that all proximal and medial leg gap genes were duplicated in the lineage leading to the spiders. Whereas the initial leg gap genes all show the conserved expression as found for other panarthropods and thus might be a synapomorphic trait of the entire phylum, the expression of the paralogs differs indicating that the duplicates have undergone neofunctionalization. However, the origin of duplication as well as the conservation of the new function differs for these

genes. The duplicate of *dachshund* is arachnid specific. It is required for patella formation, which is a morphologically novel walking leg segment only present in spiders and allies. The analysis of candidate patterning genes involved in tarsus segmentation of insects shows that these genes have some conserved functions within the phylum of the arthropods, but that their tarsus segmentation function is not conserved between spiders and insects. These data show that the initial steps of proximo-distal axis formation are highly conserved, but that more downstream patterning networks can be altered to drive the development new morphologies. In addition the presence of gene duplications is beneficial for functional gene diversification leading to novel traits whereas the fundamental developmental role of one paralog remains conserved to ensure proper development.

2. General Introduction

2.1 Morphological diversity

The animal kingdom impresses by the vast diversity of species and morphologies. This diversification is probably the result of novel traits, which facilitate adaption to ecological niches causing a survival benefit and thus are positively selected over time. Especially the arthropods are unparalleled in species number and the great morphological disparity of the species. The phylogenetic interrelationships of the major arthropods clades (i.e. chelicerates, myriapods, crustaceans and insects) are an intensively debated issue and the great diversity hampers the phylogenetic resolution despite the growing number of available genomic resources (Caravas and Friedrich, 2010; Edgecombe, 2010). The most widely accepted phylogeny (Figure 1) is that the chelicerates, including mites, ticks, scorpions, harvestmen, spiders and the "living fossils" horseshoe crabs, are the most basally branching clade. Insects and crustaceans are grouped together forming a monophyletic group, the pancrustaceans, distally in the arthropod tree. The myriapods form the sister group to the pancrustaceans and together the three clades are referred to as Mandibulata (Regier et al., 2010; Rota-Stabelli et al., 2011).

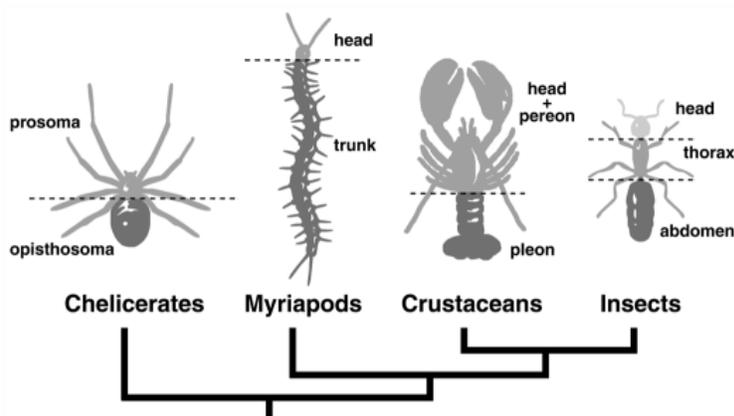


Figure 1: Phylogeny of Arthropoda

Current view of the phylogeny of all four extant arthropod groups. Dark grey denotes the hind body in the different clades. Light grey in contrast labels the more proximal head and thoracic body segments. Picture taken from (Hughes and Kaufman, 2002a).

The serially homologous appendages of the arthropods are present in a varying number and morphology and have been adapted for special purposes like feeding, sensory function, walking, flying, mating and courtship (Prpic and Damen, 2008)(Figure 2). The appendages are not only diverse between different species, but even in an individual the different appendages on the distinct body segments are morphologically different from each other. One example are isopod crustaceans like the woodlouse, which can have up to 9 different appendage types along their body axis (Kensley and Schotte, 1989). But also in members of the other taxa we find a number of different appendage types (Prpic and Damen, 2008). So the question arises how are the arthropods so successful in terms of evolution, what is the key to their great species number and disparity, especially of the appendages. The disparity probably is the reason for the great species number and this disparity is in all likelihood the result of the fragmentation of their body and appendages into segments (e.g. Angelini and Kaufman, 2005). The subdivision of body and appendages into smaller units facilitates morphological changes of these independent modules, which might be positively selected without disturbing the complete organization of the body. Every multicellular organism arises from one single cell, which divides and gives rise to cells that differentiate during embryonic development and finally form certain organs or tissues.

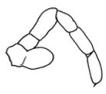
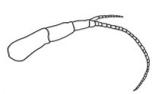
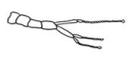
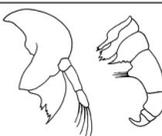
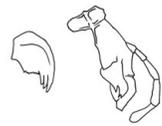
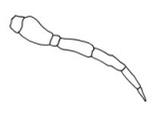
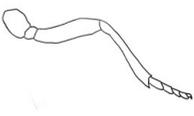
	Chelicerata	Crustacea	Insecta	Myriapoda
Perception				
Feeding				
Walking				

Figure 2: Diversity of arthropod appendages

Examples of different appendage types in the four major arthropod clades serving different purposes. The picture is taken from (Prpic and Damen, 2008)

Even more interesting is that all cells inherit the same genetic information and different use of this information decides the fate of the cell. To understand the genetic level of this different use and where alterations are possible to develop novel morphological traits, it is necessary to understand general developmental processes involved in body and appendage segmentation. In case of the commonly used arthropod model organism the fruit fly, *Drosophila melanogaster* (*D. melanogaster*), many of these processes are well studied. For example, the early gene cascade of body axis patterning involves morphogen gradients setting up the anterior posterior axis, segmentation by the pair-rule genes and positional information provided by the Hox genes (Harding et al., 1986; Johnston and Nüsslein-Volhard, 1992; McGinnis and Krumlauf, 1992). Also the detailed genetic network driving the development of the different fruit fly appendages is extensively studied (reviewed in Estella et al., 2012; Kojima, 2004), but will be described in more detail below. In order to understand how the great morphological disparity was achieved during evolution it is not sufficient to study only one representative of a phylum because this could represent a derived mode (Bolker, 1995; Wilkins, A. S., 2002). Studies of fossil ancestors from the stem-line of the phylum may provide hints how certain morphologies have changed over time (Prpic and Damen, 2008; Whittington and Almond, 1987), but it is not possible to study the gene functions involved in the formation and patterning of these morphologies during embryonic or larval development. The only way to infer the ancestral state and reconstruct the level of conservation and/or diversification through evolution is to compare genetic mechanisms involved in the formation and patterning of morphologies during embryonic development of representatives from different branches of one phylum (Damen, 2007; Diogo, 2015; Peel et al., 2005). Especially the technological advances in sequencing leading to the Next Generation Sequencing (NGS) era facilitate working with a diverse set of non-model organisms of different taxa (Liu et al., 2015).

Diversification of morphological traits during evolution has to be based on diversification at the molecular level. Two possibilities of gene function diversification are changes on the regulatory level or alterations in the coding region of the gene, which are then positively selected over time (Wittkopp et al., 2004). However, most genes fulfil more than one specific function (pleiotropic genes), thus any change in their sequence usually causes multiple effects in different organs or tissues (Rose, 1982). In order to be

positively selected, the genetic change must have a beneficial or neutral effect on all the different functions of a pleiotropic gene. Thus, genes cannot easily change their function and lead to morphological novelties, because genetic changes bear the risk of being lethal or disturbing finely tuned gene regulatory networks important for proper development or physiology of an individual. Current research has shown that gene duplications can facilitate evolution of novel traits (Innan and Kondrashov, 2010; Kondrashov et al., 2002), by providing genetic material for new functions, but without losing or changing the original gene. In vertebrates duplication of genes or complete genomes is intensively studied (Baskaran et al., 2015; Holland, 2013; Holland et al., 1994; Kasahara, 2013; Marlétaz et al., 2015), but in the arthropods duplication events appear to be less common and intensive research is still pending. Interestingly, recent studies indicate gene or genome duplication events in chelicerates (Di et al., 2015; Nossa et al., 2014; Sharma et al., 2014). Indeed, the pervasive gene duplications have been linked to the evolution of the repertoire of silk and venom genes in spiders, and to specific sensory organs in scorpions (Haney et al., 2016). Thus, extensive gene duplication in arachnids apparently provides the starting material for the evolution of morphological novelties. But not only the extensive duplication events in the lineage leading to spiders makes them good model organisms to study the evolution of the great arthropod disparity, other features of this clade and their importance in evolutionary developmental research are summarized in the following section.

2.2 Chelicerate model organisms

Chelicerates have an interesting position in the phylogenetic tree of the arthropods representing the most basally branching group (Caravas and Friedrich, 2010; Friedrich and Tautz, 1995; Giribet et al., 2001). Thus they are the most distantly related relatives to insects including *D. melanogaster* in the group of the arthropods (McGregor et al., 2008; Regier et al., 2010). In addition they have diverse appendages with different functions on the distinct body segments. Spiders for example have a pair of chelicerae for feeding, the pedipalps, which can be called the Swiss-knife of this taxon full-filling tasks ranging from sensing to sperm transfer, and the walking legs (Foelix, 1996). The

appendages on the hind body (opisthosoma), give rise to the invaginated book lungs and tubular tracheae as well as the spinnerets for silk production (Foelix, 1996). Spiders also fulfil many requirements important for model organisms (summarized in McGregor et al., 2008). They are easy to keep in the lab, have a relatively short generation time (depending on the species) and the eggs are laid into cocoons, which are easy to access. In contrast to holometabolous insects their embryonic development ends with the generation of nymphs, which largely resemble the adult body morphologies, except for some specialized features used for mating. Thus the genetic background leading to the adult morphology can be studied directly during embryonic development, and not like in *D. melanogaster* in imaginal discs of the larva.

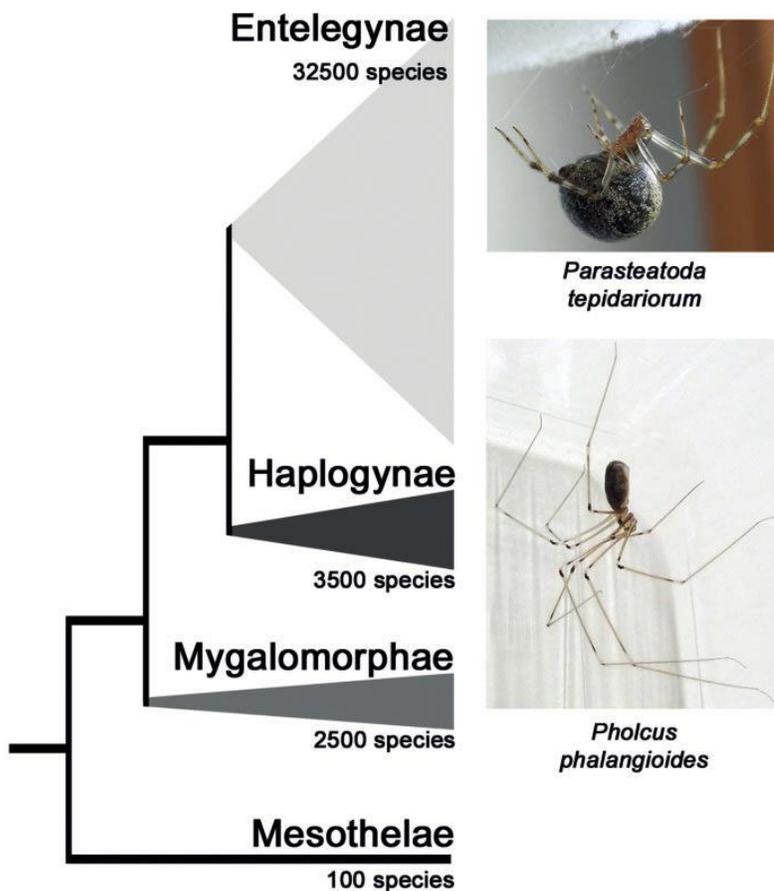


Figure 3: Spider models used in this study

The tree to the left shows a simplified spider phylogeny to show the relationships between the species *Parasteatoda tepidariorum* (an entelegyne spider) and *Pholcus phalangioides* (a haplogyne spider). Approximate numbers of described species are given for each clade. Picture taken from Turetzek et al. 2016

Early genetic studies were performed in the Mexican wandering spider *Cupiennius salei* (*C. salei*). *C. salei* is a member of the most derived and largest spider group (32,500 species, Figure 3), the Entelegynae (Turetzek et al., 2016). These studies already fostered the impact of chelicerate arthropod models for evolutionary developmental biology. Studies of Hox genes in *C. salei* helped to resolve the homology of the arthropod mouthparts and revealed the presence of duplicated Hox genes (Prpic and Damen, 2004; Schwager et al., 2007). Drawbacks of this species are its large size, its relatively long generation time and a limited toolkit for the study of gene function. Therefore another entelegyne spider species went into the focus of developmental studies, namely *Parasteatoda tepidariorum* (*P. tepidariorum*) (Hilbrant et al., 2012; McGregor et al., 2008). This spider species has a shorter life cycle and it is suited for gene function studies by parental RNAi (pRNAi). In pRNAi the double stranded RNA is injected in adult females and possible knockdown phenotypes can be observed in the numerous offspring of the cocoons laid by this female. So far many techniques, like lineage tracing, whole mount *in situ* hybridization and exogenous protein expression were established in this species (reviewed in Hilbrant 2012). In addition a full transcriptome sequence (Posnien et al., 2014) is available and a genome sequence resource is in preparation (unpublished data, courtesy of Alistair McGregor, coordinator of the *P. tepidariorum* sequencing consortium). As stated above, one species is not sufficient to understand the changes in evolution leading to novel morphologies. Findings from only one representative always bear the risk being a derived specialty of the used species and must be compared to other members of the same clade. The best option would be to validate the findings from entelegyne spiders with data of a spider species branching off earlier in the lineage leading to the entelegyne spiders (depicted in Figure 3). For that there are three possibilities: (1) Segmented spiders (*Mesothelae*, approximately 100 Species), which are rare and difficult to keep under laboratory condition. (2) Bird spiders (Mygalomorphae, 2500 Species), which also require a lot of lab space and have long generation times. However, some comparative studies were performed with a member of the bird spiders, *Acanthoscurria geniculata* (*A. geniculata*). (3) Haplogyne spiders (Haplogynae, 3,500 species), which are the sister group of the entelegyne spiders and were already in focus of early spider research (Claparède, 1862; Holm, 1940). One of the most widespread and

best-known member of the haplogyne spider group is the cellar spider *Pholcus phalangioides* (*P. phalangioides*, Order: Araneae; Family: Pholcidae). In addition to its phylogenetic position this spider is characterized by special appendage morphologies. This makes it interesting for comparative appendage patterning studies. Other names for this spider species are “daddy long legged” or “vibrating” spider and refer to the development of extremely long walking legs compared to its overall body size. The other characteristic morphological novelty is the segmentation of the tarsus into approximately twenty smaller segments, called tarsomeres (Roberts, 1995). This usually is not found in spiders. These characteristic walking leg features presumably influenced the evolutionary success of this widespread spider species, which is found in nearly every basement around the world. The extremely long legs can be used to catch prey of a much bigger size by throwing the spider silk from a safe distance. The long legs also facilitate vibrating behaviour, which makes the spider blur and thus distracts predators. For these reasons I performed comparative studies of appendage patterning (with a focus on walking leg development) in the two spider species *P. tepidariorum* and *P. phalangioides* and compared them to findings from other arthropod groups studied so far. Thus the following section is a summary about the main mechanisms of appendage development in the fruit fly *D. melanogaster* and its conservation/ or differentiation in comparison to other species.

2.3 Appendage development

D. melanogaster has a very derived indirect appendage development from imaginal discs (Cohen, 1993; Fristrom, and Fristrom, 1993) and thus might not be the best representative to study formation and patterning of these outgrowths. However, most knowledge about appendage formation is based on intensive studies of this species. The fly leg is subdivided into five distinct segments: coxa, trochanter, femur, tibia and tarsus (Kojima, 2004), which develop from the leg imaginal disc. The tarsus is composed of five tarsal segments (so-called tarsomeres), and the pretarsus, which carries the claw. The common chelicerate leg shows some differences compared to the insect leg. The distal part of the spider leg consists of two segments, the metatarsus and tarsus (Foelix, 1996), but usually

lacks the subdivision of the tarsus (except for *P. phalangioides* see above). The homology of these distal leg segments between different arthropods is still under debate and requires further genetic studies (Bitsch, 2001). In contrast to all other arthropods the spider leg contains a short segment between the femur and tibia, the patella.

No matter how versatile the different types of appendages are, they all have to undergo the same fundamental processes of development: (1) Specification of appendage primordia. (2) Specification of appendage type and position. (3) Specification of 3D Structure (anterior-posterior axis, dorso-ventral axis and proximo-distal axis). (4) Production of anatomically distinct regions (Segmentation). (5) Allometric growth (Angelini and Kaufman, 2005). These processes are studied in great detail in the fruit fly *D. melanogaster*. The specification of appendage primordia requires the onset of *Dll* expression by *wingless (wg)* during embryonic development (Cohen et al., 1993). Dorsal and ventral appendage primordia are then distinguished dorsally by *decapentaplegic (dpp)* and ventrally by EGF signalling (Goto and Hayashi, 1997; Kubota et al., 2000). Knockout and ectopic expression experiments with the genes *buttonhead* and *Sp1* (homolog of the vertebrate *Sp8*, also involved in leg development) showed complete loss and induction of leg structures, respectively and are thus also required to induce leg identity in the dorsal primordia (Estella and Mann, 2010; Estella et al., 2003). *Dll* and *Sp8* both are known to be conserved factors in early development and onset of appendage development in arthropods and vertebrates (Bell et al., 2003; Kawakami et al., 2004; Panganiban et al., 1997; Treichel et al., 2003). The interactions and hierarchy of these genes are still a matter of debate. In *D. melanogaster* it was shown that early *Dll* expression from the *Dll304* enhancer does not require *Sp1/btd* function and that *Sp1* and *Dll* are initially activated in parallel (Estella and Mann, 2010). However, maintenance of later *Dll* expression driven by another *DLL* enhancer *DllLT* is dependent on the *Sp* gene function. The appendage types along the anterior posterior body axis is largely influenced by the Hox genes (Hughes and Kaufman, 2002a, 2002b). The thoracic Hox gene *Antennapedia (Antp)* drives the proper development of the thoracic legs in the fruit fly (Schneuwly et al., 1987; Struhl, 1981). The posterior Hox genes *Ultrabithorax (Ubx)* and *abdominal-A (abd-A)*, on the contrary repress the development of appendages in the abdomen of *D. melanogaster* and are important for the formation of pleuropodia in the

beetle *Tribolium castaneum* (*T. castaneum*) and the milkweed bug *Oncopeltus fasciatus* (Angelini et al., 2005; Bennett et al., 1999; Hughes and Kaufman, 2002b). The conservation and exact function of these early mechanisms in appendage specification are, however, still not fully understood for the other arthropod groups. Results from the spider *P. tepidariorum* show that legs develop in the absence of the *Antp*, but opisthosomal limb buds develop in the presence of Hox gene co-expression of *Antp*, *Ubx* and *abd-A* (Khadjeh et al., 2012). In addition, *Dll* is required for the development of the first and second walking leg segment and distal leg development in *P. tepidariorum* (Pechmann et al., 2011). The mechanisms of appendage allocation and initiation probably do not have a significant impact on diversification of appendage morphology, because these are fundamental processes likely to be virtually identical in all appendages. However, the formation of the proximal distal appendage axis has a significant influence on morphological differences of different appendage types and divergent species. Thus my work focused on the proximo-distal axis (PD axis) patterning and refinement mentioned in steps 3 and 4 above. The genetic networks involved in these processes show, as already observed for body segmentation, a certain level of modularity (simplified in Figure 4). Combined *wg* and *dpp* interactions initiates the leg gap genes (*hth/exd*, *dac* and *Dll*), which generally subdivide the leg into proximal, medial and distal regions. Afterwards the EGFR pathway and several downstream targets are activated by *Dll*, *wg* and *dpp* in the distal portion to refine this patterning in the developing fly tarsus, called the secondary PD patterning (summarized in Damen and Prpic 2008; Estella et al., 2012; Kojima, 2004). The joints are finally developed under the control of the Notch pathway (Bishop et al., 1999; Celis et al., 1998; Rauskolb and Irvine, 1999). Data from the spider *C. salei*, the cricket *Gryllus bimaculatus* and the beetle *T. castaneum* show that the role of joint formation is a conserved function fulfilled by the Notch pathway (Angelini et al., 2012; Mito et al., 2011; Prpic and Damen, 2009). Preliminary expressional data for Notch pathway genes in the spiders *P. tepidariorum* and *P. phalangioides* further support the hypothesis that the joint formation itself is not altered between the different species and thus is also not involved in diversification of appendage morphologies ((Pechmann, 2011), and unpublished data).

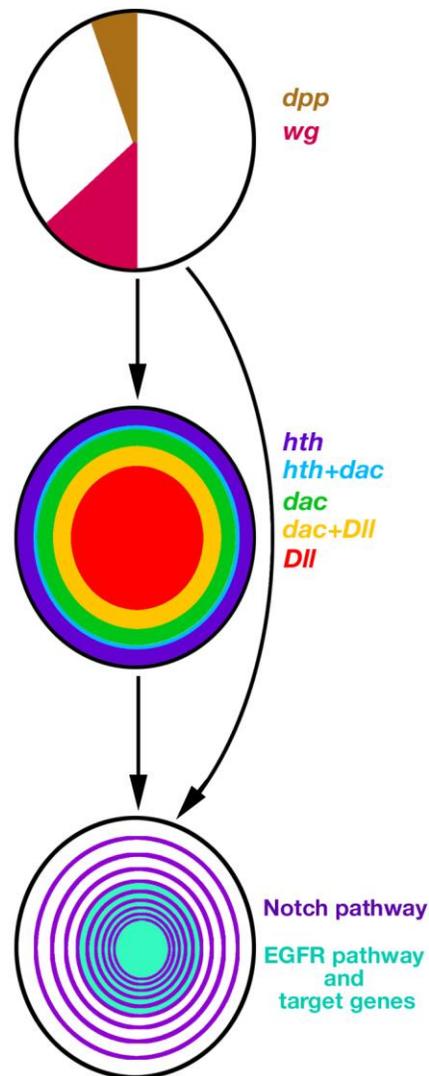


Figure 4: Simplified leg patterning cascade of *D. melanogaster*

Schematic overview of the walking leg patterning cascade in leg imaginal discs of *D. melanogaster*. The colour code in the disc represents the gene functions given in the same colour on the right. The rings of the Notch pathway represent the joints formed in the leg and tarsus segments. The picture was modified from (Prpic and Damen, 2008) with findings reviewed in (Estella et al., 2012; Kojima, 2004).

To investigate whether the leg gap genes or the genes involved in secondary PD axis patterning are altered and thus may drive differences between spiders and insects as well as the characteristic leg morphology of *P. phalangioides* I studied candidate genes known to be involved in these patterning networks from *D. melanogaster*. Findings on primary and secondary PD axis patterning in *D. melanogaster* are summarized and compared to other species in the following.

2.3.1 Proximo-distal axis formation in arthropods

Key players in proximo-distal axis formation are the genes *hth*, *dac* and *Dll*, which are required to pattern the leg into proximal, medial and distal parts, respectively (Abu-Shaar and Mann, 1998; Lecuit and Cohen, 1997). They were called leg gap genes because the corresponding part of the leg was lost after loss of function experiments (summarized in Kojima, 2004 and Estella, 2012; Cohen and Jürgens, 1989; Cohen et al., 1989; Gonzalez-Crespo and Morata, 1995; Mardon et al., 1994; Rauskolb et al., 1995). Already nearly 20 years ago it was suggested that different levels of the morphogens *wg* and *dpp* are responsible for activation of *dachshund* and *Dll* and repression of *hth*. This hypothesis was challenged by the finding that mutant discs lacking the *dpp* gradient, can still form a PD axis (Estella and Mann, 2008). This has led to a number of subsequent studies that have further refined our understanding of the role of *wg* and *dpp* and their targets like *Dll* during the formation of the proximo-distal axis of the leg imaginal disc.

Current research on the enhancer elements of *Dll* showed that co-expression of *wg* and *dpp* indeed induces the expression of *Dll* in the most distal part in early leg imaginal discs, but that it is not required for the maintenance of *Dll* (reviewed in Estella et al., 2012). In addition it was shown that *dac* is not activated by intermediate *wg* and *dpp* levels, but by direct activation of *Dll* (Giorgianni and Mann, 2011; McKay et al., 2009). *wg* *dpp* coexpression thus only play a role in repressing *dac* in the most distal part. During leg growth the *Dll* and *dac* domains broaden and getting distinct from each other, which is probably maintained by auto regulatory mechanisms, transcriptional memory mechanism (Kim et al., 2008) and other downstream factors involved in secondary PD patterning (see below). The proximal part is defined by the homeobox gene *hth* distally repressed by *wg*, *dpp* and *dac* (Abu-Shaar and Mann, 1998; Wu and Cohen, 2000). Conservation of PD axis patterning by the leg gap genes was intensively studied in many panarthropod species including onychophorans, chelicerates, myriapods, crustaceans and various insects (Angelini and Kaufmann 2005; Janssen et al., 2010; Prpic and Damen, 2004; Prpic and Telford, 2008; Prpic et al., 2001, 2003; Sharma et al., 2012).

These studies showed that *Dll* and *dac* have conserved expression patterns in the developing walking legs of all panarthropods (Figure 5). The expression of *hth* and *exd*

on the other hand differs in myriapods and chelicerates when compared to the pancrustaceans and onychophorans. Studies in the spiders *C. salei* and *A. geniculata*, and the myriapod *Glomeris marginata* showed a reversal of *hth* and *exd* expression (Pechmann and Prpic, 2009; Prpic and Damen, 2004; Prpic and Tautz, 2003). In these arthropod species *hth* is expressed in the entire walking leg excluding the most distal tip and *exd* is restricted to the proximal part. In the onychophoran *Euperipatoides kanangrensis* (*E. kanangrensis*), the crustacean *Parhyale hawaiiensis* and different insects *hth* is restricted to the proximal part and *exd* expression expands distally (Janssen et al., 2010; Prpic and Telford, 2008). That this reversal of expression could happen and did not lead to morphologically different PD axis is caused by the dependency of *hth* on *exd*. Hth needs *exd* to be transported to the nucleus (Berthelsen et al., 1999; Jaw et al., 2000; Rieckhof et al., 1997). Thus only in the area of co-expression these genes are functional and can fulfil their proximal patterning function.

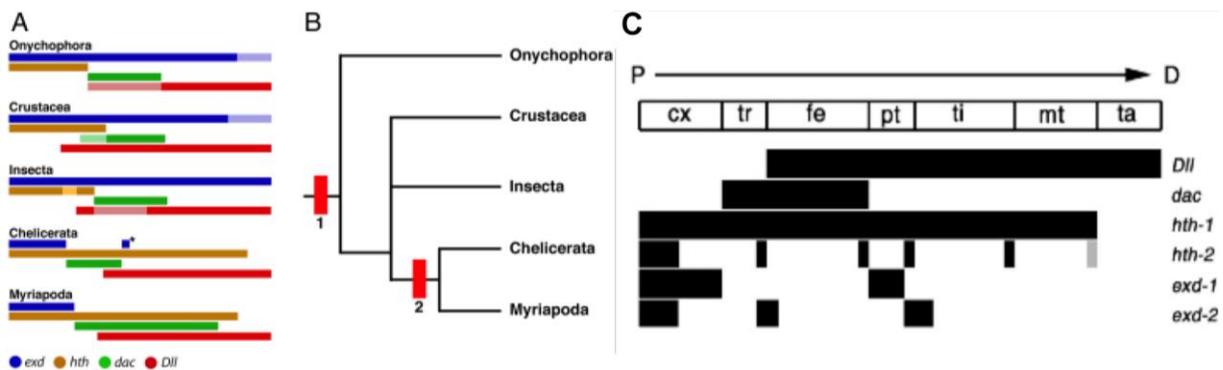


Figure 5: Conservation of leg gap genes in panarthropods and spiders

A: Expression patterns of *hth* (orange), *exd* (blue), *dac* (green) and *Dll* (red) in the four major arthropod clades in comparison with onychophorans. **B:** Phylogenetic relationships of the panarthropods supported by findings from leg gap gene expression. The red box 1 marks the plesiomorphic state of the *hth/exd* expression pattern typical for onychophorans, crustaceans and insects, and the appearance of the *dachshund* domain within the *Dll* domain. The red box 2 denotes the synapomorphic reversal of *hth* and *exd* expression present in chelicerates and the partially non-overlapping appearance of *dac* and *Dll*. **A** and **B** are taken from (Janssen et al., 2010). **C:** Summary of leg gap gene expression including duplicates of *hth* and *exd* in spiders. The grey box marks differing expression of *hth2* in *C. salei* compared to *A. geniculata*. modified from Pechmann et al., 2010. cx: coxa, tr: trochanter, fe: femur, pt: patella, ti: tibia, mt: metatarsus and ta: tarsus.

Since the co-expression of *hth* or *exd* is at all times restricted to the proximal leg, the divergent expression patterns found in the panarthropods are nevertheless functionally equivalent and the patterning function of *hth* and *exd* in the proximal leg is evolutionarily

highly conserved. Interestingly studies from the spiders revealed that *hth* and *exd* are duplicated in the entelegyne spider *C. salei* and the bird spider *A. geniculata* (Figure 5, C) (Pechmann and Prpic, 2009; Prpic et al., 2003). The expression of these duplicates suggests that they differentiated from their original gene function, and probably underwent neofunctionalisation. The paralog of *hth*, *hth2* even shows different expression patterns in the walking legs of *C. salei* and *A. geniculata*. Thus studies in more chelicerate species could provide insight in the origin of gene duplication events and subsequent evolution of gene function of the duplicates. Furthermore one ring of expression in the area of the patella was found for the paralogs *exd1* and *exd2* in *C. salei* and *A. geniculata* in addition to the conserved proximal expression. Further studies showed that this ring represents a later function of *exd*, acting downstream of the Notch pathway after joint formation (Prpic and Damen, 2009). The coincidence of this ring with the chelicerate specific patella suggests that late function of the leg gap genes and their duplicates might have a role in the development of morphological novelties. Studies in the mite *Archezogozetes longisetosus* (*A. longisetosus*) challenge this hypothesis. Although it is a member of the chelicerates and also has a patella only one *exd* paralog has been found so far and this gene has no medial expression ring (Barnett and Thomas, 2013). Findings for another chelicerate the harvestman *Phalangium opilio* (*P. opilio*), however, revealed the expression of *exd1* as a ring in the medial part of the leg (Sharma et al., 2012). To further clarify the evolution of these leg patterning genes in the spiders I have performed additional studies of these genes in *P. tepidariorum* and *P. phalangioides*.

2.3.2 Tarsus segmentation gene network

The secondary segmentation process that occurs in the tarsus of *D. melanogaster* requires *Dll* and the EGFR pathway, which is initiated by *wg/dpp* (Galindo, 2002; Galindo et al., 2005). The EGFR signaling pathway creates a gradient in the tarsal segment 3- 5 (t3-5), and the pretarsus. The pretarsus segment is defined by the transcription factors *aristaless* (*al*), *clawless* (*cll*) and *Lim1*, which are activated upon high levels of EGFR signaling (Campbell and Tomlinson, 1998; Kojima et al., 2005; Pueyo, 2004; Pueyo et al., 2000; Tsuji et al., 2000). The development of tarsal segment 4 and 5 in contrast requires

low levels of EGFR signaling activating the homeodomain genes *BarH1* and *BarH2* and *apterous (ap)* (Kojima et al., 2000; Pueyo, 2004; Pueyo et al., 2000). The more proximal tarsal segments 1–5 are defined by *bric-a-brac (bab)*, *tarsal-less (tal)*, *spineless (ss)* and *rotund (rn)* (summarized in Kojima 2004 and Estella 2012; Baanannou et al., 2013; Emmons et al., 1999; Godt et al., 1993; Kozu et al., 2006).

The exact gene expression patterns that subdivide the tarsus segment of *D. melanogaster* into several smaller units are established by complex cross-regulatory interactions of these tarsal genes and have been intensively studied in over a decade, but are still not fully understood (Overview in Figure 6). In the pretarsus the proteins Al and Cll form a complex, which activates the expression of *Lim1* (Campbell, 2005; Miyazono et al., 2010). The *Bar* homologs are then repressed distally by *Lim1* and its cofactor *Chip* and proximally by *ss* and *rn* (Kojima et al., 2005; Pueyo, 2004). Also the genes *trachealless (trh)* and *tango*, both predominantly known for the development of the tracheal system during embryonic development (Isaac and Andrew, 1996; Sonnenfeld et al., 1997), were shown to be involved in tarsus patterning by fine tuning the function of *BarH* in the fifth tarsal and pretarsal segment (Tajiri et al., 2007). For proper development of the more proximal tarsus segments the expression of *dac* needs to be restricted to the medial part. It was found that this is performed by the *Bar* homologs in the early third instar leg imaginal disc (Giorgianni and Mann, 2011). After the onset of *ss* and *rn* expression these genes mediate the distal repression of *dac* (Pueyo and Couso, 2008). *bab* expression in concentric rings is fine tuned by *Dll* and *rn* binding to the leg and antennal enhancer and is then also involved in repression of *dac* (Baanannou et al., 2013).

The study of the gene battery involved in secondary PD axis formation in the red flour beetle *T. castaneum*, showed that most genes involved in *D. melanogaster* tarsus segmentation are also present and have a similar function in the tarsus of the beetle (Overview Figure 6) (Angelini et al., 2009; Angelini et al., 2012; Smith et al., 2014). However, not all genes known from *D. melanogaster* were included in this study and especially the genes involved in the pretarsus segment development led to contrasting results. Whereas the RNAi of *al* led to alterations of the claw (Beermann and Schröder, 2004), no obvious leg phenotypes were found after *cil* RNAi, although expression of this gene is present in the distal tip during embryonic development of *T. castaneum* (Cande et

al., 2009; Grossmann and Prpic, 2012). *Lim1* RNAi showed the fusion of femur and tibia but in contrast no distal walking leg phenotype. Even simultaneous knockdown of *Lim1*, *al* and *cil* did only lead to an additive combination of the single RNAi of *al* and *Lim1* (Angelini et al., 2009). Data for *Gryllus bimaculatus* and *Manduca sexta* showed that also EGFR signalling and distal expression of *bab* and *al* are similar to *D. melanogaster* (Miyawaki et al., 2002; Nakamura et al., 2008; Tanaka and Truman, 2007). This further supports the conservation of distal patterning in insects.

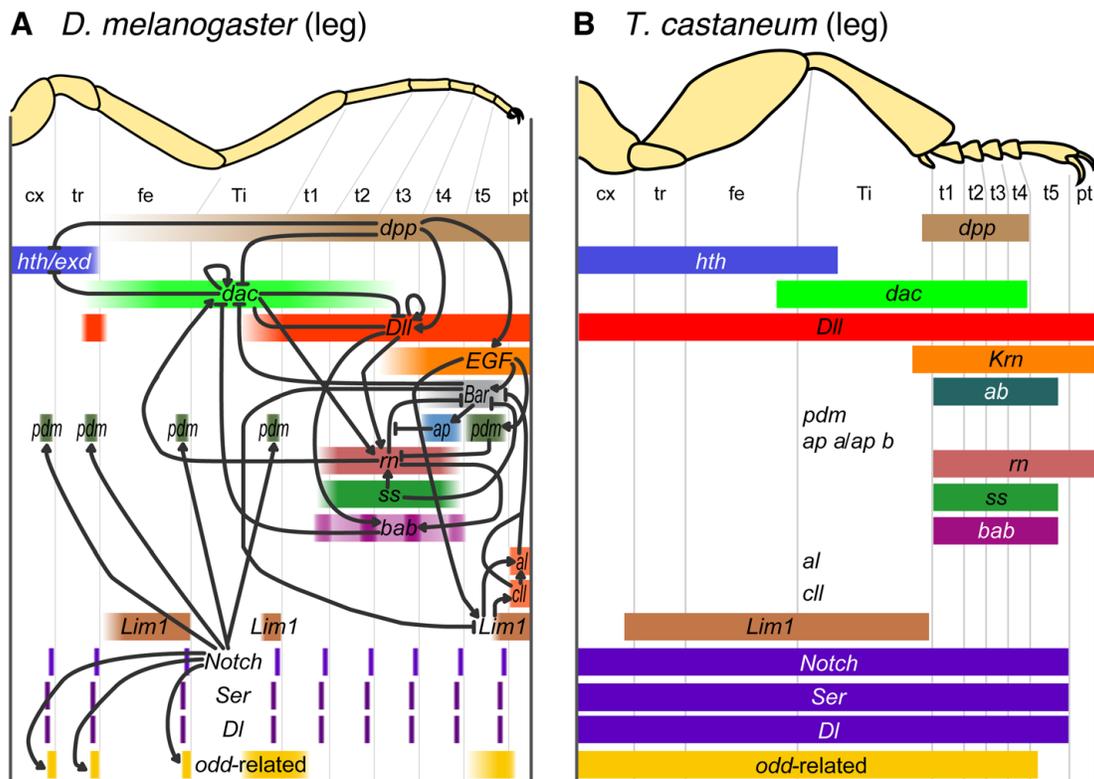


Figure 6: Overview of PD leg patterning in *D. melanogaster* and *T. castaneum*

A: Overview of genetic cross-regulation involved in patterning the walking leg and tarsus segmentation of *D. melanogaster* (Summarized from Kojima 2004 and Estella 2012). **B:** Summary of the patterning cascade in *T. castaneum*. The picture was modified from (Smith et al., 2014).

Contrasting findings from the crustacean *Artemia*, however, showed that *trh* is absent from the distal tip of all appendages (Mitchell and Crews, 2002), thus suggesting that conservation of distal patterning in the insects cannot be generalized for all arthropods. Recently six distal limb genes were studied in the onychophoran *Euperipatoides*

kanangrensis (Oliveira et al., 2014), which represents the more basal sister group of arthropods (Borner et al., 2014). Comparative studies from insects and the members of the sister group of arthropods could reveal a possible conservation of distal patterning in the last common ancestor even prior to segmentation of the limb. The study shows that the genes *cll*, *al*, *rn*, *Lim1* and *ss* are present in all appendages, but with a different expression in the anterior appendages compared to *D. melanogaster*. However, it was also shown that at least *cll*, *al*, *Lim1* and *ss* are expressed in the distal part of the legs during late embryonic development. This suggests that the distal patterning of the locomotory appendages already involved these genes in the last common ancestor. Another scenario could be that distal limb development independently evolved in onychophorans and higher insects using the same set of genes. The findings for *Lim1* of *T. castaneum* and *trh* of *Artemia* would support this scenario. The only possibility to identify the level of conservation of distal limb development and tarsus segmentation and determine whether it was already present in the last common ancestor of all arthropods is to study the distal patterning in more arthropod species. Therefore we studied these genes in the two spider species *P. tepidariorum* and *P. phalangioides*. Preliminary data on the expression of *EGFR* in the spider *P. tepidariorum* (see Supplemental Figure 2) already indicates that the distal organization of the leg is probably independent of an EGFR gradient signalling in the distal part. However, this does not exclude the possibility that the downstream targets were ancestrally used to pattern the distal limb area and are just switched on by another gradient.

3. Materials and Methods

3.1 Animal culture

Parasteatoda tepidariorum (*P. tepidariorum*) and *Pholcus phalangioides* (*P. phalangioides*) embryos were obtained from our laboratory stocks in Göttingen, and treated as previously described (Turetzek et al., 2016).

3.2 Identification of spider orthologs and phylogenetic analysis

Unless stated otherwise the identification and cloning of candidate genes was based on the transcriptomic resources available for both spiders. To identify genuine spider gene orthologs a blast search was performed using a translated nucleotide query of the gene of interest from *Drosophila melanogaster* (usually the canonical isoform from uniprot (Consortium, 2015)) in the two spider transcriptomes of either *P. phalangioides* (Janssen et al., 2015) or *P. tepidariorum* (Posnien et al., 2014) using Geneious® (version 8.1.5). The best five transcriptome hits were selected as long as their e-value was below e^{-9} . The largest open reading frame was translated into protein sequences. Afterwards a protein blast (<http://blast.ncbi.nlm.nih.gov>) against the non-redundant protein sequence database including the following organisms and taxa (*Drosophila melanogaster* (taxid:7227), *Mus musculus* (taxid:10090), *Danio rerio* (taxid:7955), *Onychophora* (taxid:27563), *Chelicerata* (taxid:6843), *Tribolium castaneum* (taxid:7070), *Apis mellifera* (taxid:7460), *Anopheles gambiae* (taxid:7165)) was performed to identify further homologues. The best fifty hits from the non-redundant protein sequence database with an e-value below e^{-15} (whereas isoforms of one gene and redundant predicted genes were discarded) were then used to generate a protein alignment with default settings using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/> (Sievers et al., 2011)). Based on this Clustal protein alignment a phylogenetic tree was calculated using the parallel version of MRBAYES ((version 3.2.5) (Ronquist and Huelsenbeck, 2003)). After testing mixed amino acid substitution models the correct model was chosen to generate topological convergence using Metropolis coupling for

maximum 3.000.000 generations or after reaching an average standard deviation of split frequencies below 0.01. The resulting 50% majority rule consensus tree was visualized with Geneious® (version 8.1.5). Genuine spider homologs were identified forming a monophyletic group with the candidate gene homologs from other species. Unless stated otherwise the identified *P. tepidariorum* paralogs were named according to their similarity to the *D. melanogaster* protein, so that the paralog with lower e-value was referred to as first (e.g. *clawless1*). The *P. phalangioides* paralogs were named according to their homologs in *P. tepidariorum* identified with the phylogenetic tree (see Table 1). After this identification of the spider homologs an additional phylogenetic tree was generated for some genes to investigate the phylogenetic relationships between the chelicerate paralogs in more detail. These phylogenetic trees were generated with a subset of homologs for chosen species and calculated in the same way as described above.

3.3 Molecular methods

3.3.1 RNA isolation

Total RNA for cDNA synthesis was isolated from a combination of different embryonic stages of either *P. phalangioides* or *P. tepidariorum* with TRIzol® (Life Technologies, Thermo Fisher Scientific, Carlsbad, CA, USA) according to the manufacturer's instructions. Total RNA extraction for library preparation of Illumina sequencing of *P. phalangioides* was previously described in Janssen et al., 2015.

3.3.1 cDNA synthesis

After RNA isolation cDNA was generated with either the SMARTer™ PCR cDNA Synthesis Kit or the SMARTer™ RACE cDNA Synthesis Kit (Clontech, Mountain View, CA, USA). cDNA synthesis of specific stages, organs or tissues, which were needed in lower amounts was performed with the Maxima® First Strand cDNA Synthesis Kit (Life Technologies, Thermo Fisher Scientific, Carlsbad, CA, USA).

Table 1: Spider orthologs

Organism	Gene name	Transcript ID	e-value	
<i>Pt</i>	<i>apterous1</i>	Locus_4700_Transcript_14/16	2.26e ⁻³⁸	
<i>Pp</i>	<i>apterous2.1</i>	comp120245	7.35e ⁻⁴⁵	not cloned
<i>Pt</i>	<i>apterous2.1</i>	Locus_11195_	8.31e ⁻⁴³	not cloned
<i>Pt</i>	<i>apterous2.2</i>	Locus_1_Transcript_14384/166847	3.28e ⁻²³	not cloned
<i>Pp</i>	<i>apterous1</i>	comp117589	4.98e ⁻⁴⁰	
<i>Pp</i>	<i>apterous2.2</i>	comp118341	1.17e ⁻⁴¹	
<i>Pp</i>	<i>aristaless</i>	comp123836		
<i>Pt</i>	<i>aristaless</i>	Locus_18982_Transcript_4/4		
<i>Pp</i>	<i>BarH1</i>	comp116578_c0_seq1	5.00e ⁻⁴⁴	bad probe
<i>Pp</i>	<i>BarH2</i>	comp120407_c0_seq1	8.14e ⁻³³	bad probe
<i>Pt</i>	<i>BarH2</i>	Locus_1_Transcript_157093/166847	2.31e ⁻³⁰	
<i>Pt</i>	<i>BarH1</i>	Locus_3013_Transcript_10/10	6.74e ⁻⁴³	
<i>Pp</i>	<i>clawless1</i>	comp116449	9.68e ⁻⁴⁹	
<i>Pt</i>	<i>clawless1</i>	Locus_15589_Transcript_1/1	4.69e ⁻⁵³	
<i>Pp</i>	<i>clawless2</i>	comp121099	3.21e ⁻⁴⁸	
<i>Pt</i>	<i>clawless2</i>	Locus_21582_Transcript_1/1	9.04e ⁻⁵²	
<i>Pp</i>	<i>Lim1.2</i>	comp121897	6.45e ⁻⁵⁸	
<i>Pt</i>	<i>Lim1</i>	Locus_1_Transcript_73444/166847	5.70e ⁻⁶⁴	
<i>Pp</i>	<i>Lim1</i>	comp125826	1.43e ⁻⁹⁴	
<i>Pt</i>	<i>Lim1.2</i>	Locus_2322_Transcript_1/1	1.23e ⁻⁴⁷	
<i>Pp</i>	<i>spineless1</i>	comp121950	0	
<i>Pt</i>	<i>spineless1</i>	Locus_18705_Transcript_3/6	0	
<i>Pp</i>	<i>spineless2</i>	comp125157	0	
<i>Pt</i>	<i>spineless2</i>	Locus20228_Transcript_1/2	6.80e ⁻¹⁵²	
<i>Pp</i>	<i>trachealess2</i>	comp119399	5.07e ⁻¹⁶⁵	
<i>Pp</i>	<i>trachealess1</i>	comp128285	1.08e ⁻¹⁶³	
<i>Pt</i>	<i>trachealess2</i>	Locus_17705_Transcript_7/7	4.43e ⁻¹⁵⁸	
<i>Pt</i>	<i>trachealess1</i>	Locus_1_Transcript_160439/166847	1.02e ⁻¹⁶⁸	

3.3.2 PCR

The strategies to amplify gene fragments from cDNA differed depending on the availability of transcriptomic resources. When no transcriptome was available first small fragments of the gene of interest were amplified with degenerated primers, usually in combination with nested PCR. Subsequently, additional sequence information was obtained by RACE (Rapid Amplification of cDNA Ends) PCR with primers designed depending on these small fragments. All these PCR types were performed with the Advantage® 2 Polymerase Mix (Clontech, Mountain View, CA, USA), according to the

manufacturer's recommendations. For genes identified from the transcriptomes of either *P. phalangioides* or *P. tepidariorum* primers were designed with help of Primer3 (Untergasser et al., 2012) and gene fragments were amplified by standard PCR using one of the following thermo stable DNA-polymerases in the given order: Taq DNA Polymerase (New England Biolabs, Ipswich, MA, USA or Life Technologies, Thermo Fisher Scientific, Carlsbad, CA, USA), Phusion®High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) or Advantage®2 Polymerase Mix (Clontech, Mountain View, CA, USA). Standard PCRs were performed with standard primers and the Phusion®High-Fidelity DNA Polymerase to provide sufficient amounts of template DNA from the cloned gene fragments prior to *in vitro* transcription of probes or double stranded RNA (dsRNA).

3.3.2.1 Standard oligonucleotides

Table 2: Standard oligonucleotides

Name	Sequence 5' - 3'
Topo-1	TCG GATC CAC TAG TAA CG
Topo-2	GTG TGA TGG ATA TCT GC
T7-Sp6	GTA ATA CGA CTC ACT ATA GGG ATT TAG GTG ACA CTA TAG AAT ACT CAA GC
T7	GTA ATA CGA CTC ACT ATA GGG CGA AT
M13 forward	GTA AAA CGA CGG CCA G
M13 reverse	CAG GAA ACA GCT ATG AC

3.3.2.2 RACE-PCR oligonucleotides

Table 3: Oligonucleotides for RACE-PCR

Name	Sequence 5' - 3'
Pp_exd_1_3'RACE_new_Nt251	GCG GCC GGT GGC CCC AAT CAG C
Pp_exd_1_5'RACE Nt249	TGT CTG ATT TGG GCC AGT TTC GCC CGG
Pp_exd_2_3'RACE_Nt310	CGA AGG GGT TGC AGG ACC CGA GAA AG
Pp_hth1_3'RACE_new_Nt374	CAG GGA CTG CGA CGG GGG CCT C
Pp_hth1_5'RACE_new_Nt 267	GGT GGT GGC GCC GGT GGT GGA AG
Pp-En-3RACE:	CGG CGA CAG GAT TTG GCC CG
Pp-En-5RACE	CGG GCC AAA TCC TGT CGC CG
UPM	Combination of long[2µM] and short [10µM]
UPM long T3	ATT AAC CCT CAC TAA AGG GAA AGC AGT GGT ATC AAC GCA GAG T
UPM short T3	ATT AAC CCT CAC TAA AGG GA

3.3.2.3 Gene specific PCR oligonucleotides

Table 4: Gene specific PCR oligonucleotides

Name	Sequence 5' - 3'
Pp_al_for	CCG GTT CTC TCC TTC GAT ATT GTT GTC TC
Pp_al_rev	TCA ACT GAC TTC GCC GTT GGA CTT CCT C
Pp_apt_2.2_comp118341_f	TCG GAT GAC CAG TGC CAA GAC CTG
Pp_apt_2.2_comp118341_r	TCG CTC TTT GTC AGA ACA TCT CGT GGA
Pp_apt1_for	ACC GTT ACG TGA CAA CAG TGA GGC GG
Pp_apt1_rev	AGC ACC CAT TGT CCC CGA ACC GCT AA
Pp_c102331_for	CTT CCT TCG AGC GGC TCT TC
Pp_c102331_rev	CTG AAC TCT TTG CCG CAG GA
Pp_c105514_g1_i1_for	CGA CAG CCG TCC AAG CTA TC
Pp_c105514_g1_i1_rev	CAG GGG TCA TTG TCG TCT CG
Pp_c109329_for	GCC ACG ATC TCA ACC ACC AC
Pp_c109329_rev	TGC GCT ACC TGT TCG ACC TC
Pp_c109624_for	GAC GCT GAG TCC GAT CCA GA
Pp_c109624_rev	CTC CGG GTC TTT CGT CGA GT
Pp_c109662_for	CGG TGA ACC GTC TTC CAG TG
Pp_c109662_rev	GTC TCG GGA CCG TCC TCT GT
Pp_c95092_g2_i1_for	CTG GTC TAG TCG CCG CAG TC
Pp_c95092_g2_i1_rev	TCC ACG TCC GCA GTC TTT GT
Pp_cll_1_comp116449_for	CAT GGA CCA CGG TGA TAC GGA CGA
Pp_cll_1_comp116449_rev	CCT CCT CCA TTT CGT CCG CCT GTT
Pp_cll_2_comp121099_f	GGA ATC AAT TCC CCG CAA CCG GAC
Pp_cll_2_comp121099_r	GTC AGA GAT GTC GGT GGT TGC AGG
Pp_hth2_for	CGG TTA TCG GGT GGA CTT CGG
Pp_hth2_rev	GGT CCA TGA TGT TCG GAG GCG AA
Pp_Lim1_comp121897_f	GGA GGC GTA AGA GGA AGC GGA GG
Pp_Lim1_comp121897_r	CAC CGA GGG CAC CGG GGA AC
Pp_ss1_for	GGA ACA GCA AAA CCG TCC GTG TCC CG
Pp_ss1_rev	TGT AGC GGC GCG TTC TCT GAC ACC TT
Pp_ss2_comp125157_f	CGT CGG TGG CTG CCT TTC TAT GTA C
Pp_ss2_comp125157_r	CCG ATA TCG ACA GCA GGG GAT TAC C
Pp_trh1_comp128285_for	GCG GCA GCA GAA TCG GGC CTA TCC
Pp_trh1_comp128285_rev	CCT CAT TAT CAG GCG TTG ATC CAC GC
Pp_trh2_comp_119399_for	ATG CAG CCT GAT ACT TAT GGA TCG TTA CC
Pp_trh2_comp_119399_rev	GCT CTG TAT GTA ATG GCC GTT ATC CAA GG

Pt_al_for	GCC TCC AAT CCC TAC ACG TAA CAG G
Pt_al_rev	GCT TCC CAA AAC ACC ACG CAC
Pt_apterous_1_for	GGA GAT ATC GCA GTT GGG TGT CCC
Pt_apterous_1_rev	GTG ATG CGA GCC AGA GAA CAC G
Pt_Bar_Locus_1_157093_for	GAG ATA GTA GCC CTT CAC TGA AGA ACA
Pt_Bar_Locus_1_157093_rev	GAC ACG CAC GAA TGT CCA GTT GC
Pt_Bar_Locus_3013_for	CGA TCA TTC TGG TCT GGA AGA AAC AAA C
Pt_Bar_Locus_3013_rev	CCT GCG CCA TAG ACT AGT AGT CAG CC
Pt_cll_f_L15589	GCTC TGT GAT CGT ACT GCT CGC GT
Pt_cll_r_L15589	GGT GAA GAT AAA GAG TGT GGA GGT GC
Pt_cll2_for	AAC AAT TCG TCG CCG CTA GTT G
Pt_cll2_rev	GGT TGT AGA TTT TGA AGT GCG TGC
Pt_Lim_1.2_Locus2322_f	CGA CCC AGT CCT CTT GCC TCT TGA
Pt_Lim_1.2_Locus2322_r	TCG TGG TGG TCC GTA AGC GTC TGA
Pt_Lim1_for	GGT GGC TGT TCC CAA GGC ATA TGT
Pt_Lim1_rev	CGA ATG TTC GAG AGT ATT ATT TCC TCC
Pt_odds_f_L17047	GGA AGT TGA GCG GCG ACG TAC C
Pt_odds_r_L17047	CCA GGG TCT TGA TCT CCA CTC TCC
Pt_spineless_1_for	GAG AAC TCC GCG CAG TAG CAT GT
Pt_spineless_1_rev	ATC TCT TTG TCC TCT GGC AGC GG
Pt_ss2_Locus_20228_f	CGT GTC CCT CCA AAA GAT GGC CAA AC
Pt_ss2_Locus_20228_r	CCA AGC AAA TCC CTT CCT TCT TCT TCC
Pt_trh1_for	CCT GTC AAC TCA TCG GCT GTG TAG
Pt_trh1_rev	CCT GTT GAT CTG TCC GGT CAT TAT GC
Pt_trh2_for	GC AAC CTG AAG GAT ATG GGT CCA TAC
Pt_trh2_rev	GGT TTC TTC TTC GGC ATT CTT GGT GTT

3.3.3 Cloning, transformation and plasmid preparation

After gel-electrophoresis, fragments of interest were extracted from the gel with the help of the NucleoSpin®Gel and PCR Clean-up (MACHEREY-NAGEL, Düren, Germany) or MinElute Gel Extraction Kit (Qiagen, Hilden, Germany). Gene fragments amplified with either Taq DNA Polymerase or Advantage® 2 Polymerase Mix were cloned into the pCR®II vector using the TA Cloning® Kit Dual Promoter (Life Technologies, Thermo Fisher Scientific, Carlsbad, CA, USA). Gene fragments amplified with Phusion®High-Fidelity DNA Polymerase were cloned blunt end into a modified pJet2.1 vector (generated by Stefan Dippel, Department Developmental Biology, Georg-August-

University Göttingen) using a standard T4 Ligase (5 Weiss units/ μ l, New England Biolabs, Ipswich, MA, USA or Life Technologies, Thermo Fisher Scientific, Carlsbad, CA, USA). Chemically competent *Escherichia coli* dh5 α or dh10b bacteria were transformed with the vectors and positive clones were selected for sequencing and further experiments using blue-white selection after incubation on LB-Agar plates containing 40 μ l X-Gal (4% in DMF). Plasmid DNA was isolated as with a modified alkaline cell lysis protocol (Pechmann, 2011), or with one of the following kits: Zyppy™ Plasmid Miniprep Kit (Zymo Research Europe, Freiburg, Germany) or NucleoSpin®Plasmid Miniprep Kit (Macharey-Nagel, Düren, Germany). Clones were further tested by restriction digest with either EcoRV or EcoRI (New England Biolabs, Ipswich, MA, USA), in case of pJet2.1 or pCR®II vector, respectively. Clones with fragments of the expected size after digestions were send for sequencing (Macrogen Europe, Amsterdam, Netherlands or LGC Genomics, Berlin, Germany).

3.3.1 *in vitro* transcription of RNA probes and dsRNA for RNAi

After amplification by standard PCR approximately 500ng template DNA were used for *in vitro* transcription to synthesize dsRNA or Digoxigenin-labeled RNA probes from the cloned gene fragments. Synthesis of the RNA probes was performed according to standard protocols (Prpic et al., 2008a). Kits and precipitation strategy for dsRNA synthesis were previously described in Turetzek et al., 2016.

3.4 Embryological methods

3.4.1 Embryo fixation and cuticle preparations

Fixation of *P. tepidariorum* embryos after reaching germ band elongation stages was performed according to standard protocols described in pervious publications (Schomburg et al., 2015). Fixation of early embryonic stages of *P. tepidariorum* and *P. phalangioides* embryos were performed according to the fixation protocol previously published for *C. salei* (Prpic et al., 2008b). Cuticle preparations of nymphal stages of *P. tepidariorum* after RNAi experiments were performed with a mixture of lactic acid and Hoyer's medium (1:1) as previously described in Turetzek et. al, 2016.

Table 5: List of gene clones used in this study

Organism	Gene name	Transcript ID	Accession number	Clone	Reference
Pp	<i>apterous1</i>	comp117589		own cloning	this work, Appendix
Pp	<i>apterous2.2</i>	comp118341		own cloning, with help of Julia Schneider	this work, Appendix
Pp	<i>aristaless</i>	comp123836		own cloning	this work, Appendix
Pp	<i>c102331</i>	c102331		own cloning	this work, Appendix
Pp	<i>c105514</i>	c105514_g1_i1		own cloning	this work, Appendix
Pp	<i>c109329</i>	c109329		own cloning	this work, Appendix
Pp	<i>c109624</i>	c109624		own cloning	this work, Appendix
Pp	<i>c109662</i>	c109662		own cloning	this work, Appendix
Pp	<i>c95092</i>	c95092_g2_i1		own cloning	this work, Appendix
Pp	<i>clawless1</i>	comp116449		own cloning, with help of Jana Holtzheimer	this work, Appendix
Pp	<i>clawless2</i>	comp121099		own cloning	this work, Appendix
Pp	<i>dachshund1</i>		CUC08949.1	Dr. Matthias Pechmann	(Turetzek et al.,2016)
Pp	<i>dachshund2</i>		CUC08949.1	Dr. Matthias Pechmann	(Turetzek et al., 2016)
Pp	<i>engrailed</i>			own cloning	(Turetzek, 2011),Appendix
Pp	<i>extradenticle1</i>	comp102492		own cloning	this work, Appendix
Pp	<i>extradenticle2</i>	comp122137		own cloning	this work, Appendix
Pp	<i>homothorax1</i>	comp102262		own cloning	this work, Appendix
Pp	<i>homothorax2</i>	comp120146		own cloning	this work, Appendix
Pp	<i>Lim1.2</i>	comp121897		own cloning	this work, Appendix
Pp	<i>spineless1</i>	comp121950		own cloning	this work, Appendix
Pp	<i>spineless2</i>	comp125157		own cloning	this work, Appendix
Pp	<i>trachealess1</i>	comp128285		own cloning	this work, Appendix
Pp	<i>trachealess2</i>	comp119399		own cloning	this work, Appendix
Pt	<i>apterous1</i>	Locus_4700_Transcript_14/16		own cloning	this work, Appendix
Pt	<i>aristaless</i>	Locus_18982_Transcript_4/4		own cloning	this work, Appendix
Pt	<i>BarH1</i>	Locus_3013_Transcript_10/10		own cloning	this work, Appendix
Pt	<i>BarH2</i>	Locus_1_Transcript_157093/166847		own cloning	this work, Appendix
Pt	<i>clawless1</i>	Locus_15589_Transcript_1/1		own cloning	this work, Appendix
Pt	<i>clawless2</i>	Locus_21582_Transcript_1/1		Christoph Schomburg and Jana Holtzheimer	this work, Appendix
Pt	<i>dachshund1</i>		FM945397.1	Dr. Matthias Pechmann	(Pechmann et al. 2009)
Pt	<i>dachshund2</i>	Locus_18128_Transcript_3/3	AKM21240.2	own cloning	this work, Appendix
Pt	<i>Distal-less</i>		CAS89759.2	Dr. Matthias Pechmann	(Pechmann et al. 2009)

Pt	<i>extradenticle1</i>	Locus_1_Transcript_58666/166847	HE608681	own cloning	(Khadjeh et al., 2012)
Pt	<i>extradenticle2</i>	Locus_1_Transcript_58675/166847		Dr. Sara Khadjeh	this work, Appendix
Pt	<i>homothorax1</i>	Locus_81_Transcript_57/63	HE608682	own cloning	(Khadjeh et al., 2012)
Pt	<i>homothorax2</i>	Locus_10999_Transcript_20/21		Dr. Sara Khadjeh	this work, Appendix
Pt	<i>Lim1</i>	Locus_1_Transcript_73444/166847		own cloning	this work, Appendix
Pt	<i>Lim1.2</i>	Locus_2322_Transcript_1/1		own cloning, with help of David Salamanca	this work, Appendix
Pt	<i>odd-skipped</i>	Locus_17047_Transcript_1/1		own cloning	this work, Appendix
Pt	<i>odd-skipped</i>	Locus_13444_Transcript_9/13		own cloning	this work, Appendix
Pt	<i>orthodenticle</i>		AB096074	Dr. Matthias Pechmann	(Akiyama-Oda, 2003)
Pt	<i>Sex combs reduced</i>		FM956097	Dr. Sara Khadjeh	(Schwager et al., 2007), sequence extended by Sara Khadjeh
Pt	<i>spineless1</i>	Locus_18705_Transcript_3/6		own cloning	this work, Appendix
Pt	<i>spineless2</i>	Locus20228_Transcript_1/2		own cloning	this work, Appendix
Pt	<i>trachealess1</i>	Locus_1_160439/166847		own cloning	this work, Appendix
Pt	<i>trachealess2</i>	Locus_17705_Transcript_7/7		own cloning	this work, Appendix

3.4.2 Whole mount in situ hybridization

In situ hybridization of embryos was performed according to standard methods with minor modification (Schomburg et al., 2015). After *in situ* hybridization and SYTOX-Green (Life Technologies, Thermo Fisher Scientific, Carlsbad, CA, USA) staining whole embryos were imaged with a Leica M205 FA binocular equipped with a QImaging Micropublisher 5.0 RTV camera using combined UV and white light. When necessary appendages were dissected and mounted in 80% Glycerol in PBST. Images of dissected tissues were taken with a Zeiss Axioplan-2 microscope using an Intas digital camera. For colour and brightness correction of the captured images Adobe Photoshop CS5 Extended or CS6 for Apple Macintosh was used. Staging of embryos for all spiders was chosen on the basis of the staging of *P. tepidariorum* (Mittmann and Wolff, 2012).

3.4.3 Live imaging of embryonic development of *P. phalangoides*

The time-lapse movies of the embryonic development of *P. phalangoides* were captured with a Leica M205 FA binocular equipped with a QImaging Micropublisher 5.0 RTV camera. The embryos were covered with Voltalef H10S oil (Arkema) and images were taken in intervals of ten minutes. The single images were combined into a movie using iMovie version 10.0.5 (Apple Macintosh), in a way that every image is shown 0.05 seconds (time-lapse movies are attached as CD-ROM).

3.4.4 Parental RNAi

For parental RNAi dsRNA of the cloned gene fragments was injected into sexually mature adult female spiders. The synthesis and precipitation of dsRNA, as well as the injection and statistical analysis of *Pt-dachshund1* were performed as described in Turetzek et. al, 2016. For successful RNAi the injection procedure and time point of mating differs for every gene and is described in the following table (Table 6). For all genes at least 3.5-4µg/µl ds RNA were injected into sexually mature female spiders. The first five cocoons produced by injected females were opened and usually half of the embryos were fixed for further experiments. The remaining half was allowed to develop until hatching of the first nymphal stage and screened for any visible phenotypes. Unless stated otherwise only offspring from the females that survived and produced offspring showing a phenotype in at least one cocoon were included in the statistical analysis. That the offspring of some females did not show any phenotype although the females were injected in the same way, with the same batch of dsRNA and treated exactly in the same manner can have multiple reasons. Although the females were fed and mated equally from beginning of the experiment, each individual female might be in a different condition, caused by mite infection, feeding status during the juvenile period and other environmental stress factors. Furthermore it can happen that the injected liquid is pushed out of the injection hole or through the mouth, which cannot be completely avoided even when injected very carefully. Thus the amount of dsRNA might simply not be sufficient to cause an effect. In addition the exact RNAi mechanism in the spider is not yet known and here individual differences and the position of injection might have an

effect as well. Freshly hatched nymphs showing phenotypes were embedded in Voltalef H10S oil (Arkema, Colombes, France) and boiled for at least one minute. Z-stack images of these nymphs were captured using the confocal laser-scanning microscope (LSM 510, Carl Zeiss, Jena, Germany).

Table 6: dsRNA injection procedure for RNAi

Gene	# Females	# Injections	Day for recovery	Volume of dsRNA	Mating	# Phenotypes	# Females died	Included in statistic
<i>BarH1</i>	10	5	1	2,0	after 3rd injection	4	4	4
<i>BarH2</i>	5	5	1	2,0	after 3rd injection	0	0	
<i>cII1</i>	5	5	1	2,5	after 3rd injection	3	1	3
<i>cII1</i>	5	3	0	2,5	one day before injection	4	1	4
<i>cII1</i>	3	5	1	2,5	after 3rd injection	1	0	excluded from statistic
<i>dac2</i>	3	5	1	2,5	after 3rd injection	0	0	
<i>exd1</i>	3	3	1	2,0	after 3rd injection	0	1	
<i>exd1</i>	3	3	1	3,5	before injection	0	0	
<i>exd2</i>	3	3	1	2,0	after 3rd injection	0	2	
<i>exd2</i>	3	3	1	3,5	before injection	0	2	
<i>hth1</i>	3	3	1	2,0	after 3rd injection	0	0	
<i>hth1</i>	3	3	1	3,5	before injection	0	1	
<i>hth2</i>	3	3	1	2,0	after 3rd injection	0	1	
<i>hth2</i>	3	3	1	3,5	before injection	0	1	
<i>lim1</i>	3	5	1	2,5	after 3rd injection	0	0	
<i>ss1</i>	5	5	1	2,0	after 3rd injection	0	0	

3.4.4.1 Phenotypic analysis of *Pt-cII1* RNAi

Freshly hatched nymphs injected with *Pt-cII1* dsRNA were observed and categorized in the following manner: “wild type” no visible effect of the nymphal cuticle can be observed; “not hatched”: embryos which either died during embryogenesis or were not able to hatch from the embryonic envelope; “weak” malformation or missing of the joint

between the metatarsus and tarsus in at least one walking leg or pedipalp; “strong” missing of either the tarsus segment or the entire part of the appendage distal to the metatarsus in at least one walking leg or pedipalp.

In freshly hatched nymphs as well as after the second moulting into second instar nymphs the joints in the prosomal appendages are not fully developed, thus especially the strong phenotype was difficult to distinguish from wild types during these stages. Therefore most of these nymphs were allowed to moult once more into third instar nymphs. These show, in the wild type, fully developed joints and have bristles as well as sensory organs, which were not present in earlier stages. To ensure correct categorization of the phenotypes cuticle preparations of these third instar nymphs were performed in combination or instead of analysing freshly hatched nymphs. The overview shown in Figure 67 (left panel) shows the percentage of these categorized phenotypes from the different cocoons laid by seven injected females.

The *c111* phenotype was observed in two independent experiments and with two different injection procedures. The *c111* phenotype was initially observed in one female (marked in blue Table 6) after injection of three females. The phenotypes from these offspring were not included in the statistical analysis, because the effect was first observed in the fourth cocoon and it was not clear if the strong phenotype was already present in previous cocoons and simply was not recognized in first nymphal stages. The statistic thus includes phenotypic observations in cocoon1-5 of seven females injected in two different ways (Figure 67).

As control five sexually mature female spiders were injected with injection buffer in the same way for both different injection procedures as described for *Pt-c111*. Two of the control spiders, which were injected three days in a row, only produced cocoons with embryos that died shortly after egg laying. This might be an effect of the injection but is also observed for unfertilized embryos after unsuccessful mating. Because all cocoons showed the effect, which usually tapers off when caused by the injection itself, it is more likely that unsuccessful mating caused the effect. Thus these females and the produced cocoons were excluded from the following statistical analysis (shown in Figure 67, right panel).

3.4.4.2 Phenotypic analysis of *Pt-BarH1* RNAi

The *Pt-BarH1* phenotypes were observed in two independent RNAi experiments (5 females in each experiment, from which each time two showed effects in the offspring). Unfortunately all control females injected with injection buffer died, thus the experiment needs to be repeated including a control.

After RNAi of *Pt-BarH1* freshly hatched nymphs up to the seventh cocoon from the four females were analyzed (depicted in Figure 76). The category “wild type” is similar as explained for the *Pt-cll1* RNAi. The category “not hatched”, was further subdivided into “not hatched” and “dead”. “not hatched” included all embryos, which finished ventral closure stages indicated by tightening of the embryonic membrane, which makes visible some embryonic structures like the prosomal appendages, but failed to open the egg membrane, which usually bursts above the tooth pic of the pedipalps. “dead” included the embryos, which died during the course of embryonic development, indicated by aggregation of the cells to one spot. In some cases it was needed to distinguish between these categories, then the unhatched embryos were covered with Voltalef H10S oil, which makes the vitelline membrane transparent. The only RNAi phenotype category present after moulting was the missing of the first walking leg segment and was therefore named “L1 missing”.

Due to the high number of unhatched but fully developed embryos the fixated embryos for the second, third and fourth cocoon of all four females were analysed and included in the statistical analysis (Figure 77). For a better visualization the embryos were stained with Sytox-Green after fixation. In addition to the categories described above defects involved in segmentation of the embryo were found and described as follows: “head” missing of at least the head segments up to all anterior segments until the first walking leg; “chelicera/Pedipalps/L1”: missing or fusion of either chelicera, pedipalps, one of the walking legs or combinations of these effects; “other defects” massive malformation of the embryo, including duplicated axis, duplicated opisthosoma or malformed opisthosoma.

3.5 RNA Seq, NGS and *de novo* Transcriptome assembly and analysis

RNA isolation, library preparation and sequencing with Illumina HiSeq2000 of *P. phalangioides* samples were previously described in Janssen et al. 2015. In summary, triplicates of 4 different samples (a mixture of all embryonic stages as well as legs (excluding the coxa), prosoma and opisthosoma of freshly hatched first instar nymphs) were sequenced on one lane of an Illumina HiSeq2000 sequencer (Tanskriptom Analyse Labor, TAL, GZMB, Göttingen, Germany). This resulted in 472,770,758 paired-end reads of 100 bp lengths each. The assembled transcriptome that was used in Janssen et al., 2015 represents only a preliminary version. To improve the *P. phalangioides* transcriptome all paired-end reads were *de novo* assembled with a newer version of the Trinity assembler, using the following settings: `--seqType fq --JM 240G --run_as_paired --CPU 6` (version r20140717, (Haas et al., 2013)). Quality trimming and filtering of the raw reads was performed prior to the assembly with trimmomatic (Bolger et al., 2014) as part of the Trinity pipeline. Translated sequences of the entire improved transcriptome version were blasted against the common protein databases SwissProt ((Bairoch and Apweiler, 2000)) and Uniref90 (Suzek et al., 2007). In addition a protein blast search was performed against the gene set of the *P. tepidariorum* genome (unpublished data, courtesy of Alistair McGregor, co-ordinator of the *P. tepidariorum* sequencing consortium). The e-value cut-off was set to $5e^{-2}$ in all cases. This annotation was performed following the instructions of the freely available annotation pipeline Trinotate (Trinotate Release 2.0.2, <https://trinotate.github.io>, with kind help of Dr. Nico Posnien).

3.5.1 Read mapping and reduction of transcriptome complexity

For differential gene expression analysis it is necessary to map the raw reads of the different tissues against the combined transcriptome. This was performed using bowtie2 (Version: 2.1.0, (Langmead and Salzberg, 2012)). First of all, the *de novo* assembled transcriptome was indexed as reference transcriptome. Then the raw reads of the triplicates from the four different samples (embryo, legs, prosoma and opisthosoma) were multi-mapped (100 alignments recorded) end-to-end against this indexed

reference transcriptome, no mismatches allowed, using the following parameters: `--no discordant --no mixed`. The generated .sam files were converted into .bam files using Samtools (Li et al., 2009). Corset (version 1.0.3, (Davidson and Oshlack, 2014)) was used to reduce the number of transcripts prior to differential gene expression calling. Based on the bowtie2 alignments Corset clusters Trinity transcripts into potential gene models (Cluster). Subsequently the reads for each library and cluster were counted and can be used for downstream differential expression analysis. Corset was run in general with default settings and the information, which samples represent replicates, was defined by the command `-group`.

3.5.2 Differential gene expression analysis with DESeq

A principle component analysis was performed for the sample tissues and replicates based on counts from the Corset gene clusters and depicted with the plotMDS command in R (Version 0.98.1087, including the following packages: limma Version 3.22.1, RColorBrewer, (Neuwirth, 2011)). Afterwards, significantly up- and down regulated ($p_{adj} < 0.05$) corset gene clusters were identified on the basis of the library read counts generating four different comparisons: (1) Leg, prosoma and opisthosoma (all from freshly hatched nymphs) read counts were compared to embryonic read counts (referred to as 1st instar larvae). (2) The leg read counts were compared to the opisthosoma and prosoma read counts (3) The prosoma read counts were compared to the leg and opisthosoma read counts. (4) The opisthosoma read counts were compared to the leg and prosoma read counts. This analysis was performed following the instructions of the R package DESeq2 (Version 1.6.2, (Anders and Huber, 2010)) from bioconductor (Gentleman et al., 2004). The GO-term enrichment analysis of the candidate genes up-regulated in the nymphal legs was performed by Dr. Nico Posnien with Blast2Go (Conesa et al., 2005). To reduce the number of candidate genes we filtered the candidates as follows. First, all candidates with a base mean lower than 10 were discarded, because these usually include gene clusters with very low read counts in every tissue. Then the candidates were sorted for the mean read counts in the three embryonic samples and all below 50 were rejected. This had three reasons: (1) Staining

Materials and Methods

of gene expression in nymphal stages of *P. phalangioides* is not yet established. (2) The segmentation of the leg presumably already takes place during embryogenesis and (3) previous data revealed extreme growth of the walking legs during late embryonic development in *P. phalangioides*. After sorting for the log₂Fold change the best 15 candidates were chosen and further studied. Finally all candidates without an annotation in any other organism were excluded.

4. *Pholcus phalangioides*: a haplogyne spider model

A detailed study of the embryogenesis of a new model organism provides the groundwork to perform comparative studies prior to comparative expressional and functional studies. Comparisons of the general embryonic development of *Parasteatoda tepidariorum* (*P. tepidariorum*) and *Pholcus phalangioides* (*P. phalangioides*) will clarify if these spiders develop entirely different or if they share comparable tissue organization. Earlier studies already revealed major differences during germ band stages and opisthosomal organization in the embryonic development of the haplogyne spider *P. phalangioides* (Claparède, 1862; Pechmann et al., 2011), which I analyzed in more detail in this study.

The first part of this chapter “Observations on germ band development in the cellar spider *Pholcus phalangioides*” thus is the detailed description of the embryonic development of the haplogyne spider. It will be submitted as a manuscript to Development, genes and evolution. The data collection was already started during my master thesis and was finalized during this work. My supervisor and I performed the analysis of the data with equal contribution. The text of the manuscript was largely written or rewritten by my direct supervisor Dr. Prpic-Schäper.

In the era of Next generation sequencing (NGS) it is easier to sequence the transcriptome of a new species than facing the problems of fishing for genes in cDNA pools with degenerated primers to clone genes. Thus the second chapter will focus on the *de novo* transcriptome assembly for the haplogyne cellar spider *P. phalangioides*. In addition the transcriptomic resources were used for differential expressional analysis to identify genes fulfilling important roles during spider developmental processes. This project was performed in close collaboration and with supervision by Dr. Nico Posnien.

4.1 Observations on germ band development in the cellar spider *Pholcus phalangioides*

Spider development involves, after cleavage and blastoderm formation, the formation of the so-called germ disc (Andersson 1973). The germ disc represents the anlage of the embryo proper, whereas the remaining cells that are not part of the disc will either degrade or form extra embryonic tissue or yolk. The germ disc is radially symmetric and therefore needs to undergo a transition from radial to bilateral symmetry to form the germ band and ultimately the bilaterally symmetric spider. This process of symmetry break at the transition from germ disc to germ band involves the migration of a group of cells, called the cumulus, from the center to the perimeter of the disc. A few genes that are required for this migration of the cumulus have been identified (Akiyama-Oda and Oda 2003; Akiyama-Oda and Oda 2006, Akiyama-Oda and Oda 2010), but the mechanisms of symmetry break are still only poorly understood.

The majority of recent studies of spider development focused on entelegyne spiders (e.g. Abzhanov and Kaufman 2000, Akiyama-Oda and Oda 2003, McGregor et al. 2008). Although this group of spiders includes the vast majority of all spider species (ca. 33000 species (i.e. over 80%)) it is unclear whether the processes described on the basis of entelegyne spiders are also representative of the other, smaller spider groups. We have therefore initiated comparative studies in the haplogyne spiders which are phylogenetically the sister group of the entelegyne spiders, but comprise only about 3500 species worldwide. We have chosen the species *Pholcus phalangioides* as a very common and therefore easily accessible species. The majority of the species of the genus *Pholcus* are distributed in Asia, but a few species also have a western Palearctic distribution, and *P. phalangioides* is even a cosmopolitan species (Huber 2011). The species of *Pholcus* are characterized by their extremely long walking legs that make the animals superficially similar to long-legged species of harvestmen, and that have earned them their English nickname "daddy long leg spiders". So far, we have studied gene expression and function in appendage development in *P. phalangioides* (Pechmann et al. 2011; Turetzek et al. 2016). Here we focus on the description of germ band formation and further differentiation of the germ band.

4.1.1 Formation of the germ band in *P. phalangioides*

The first sign of the transformation of the radially symmetric germ disc into the bilaterally symmetric germ band is the aggregation of cells at the center of the germ disc, thus forming the primary thickening (Figure 7, A). At 25 °C the primary thickening is visible approximately 48 hours after egg deposition (see Supplemental table 1). About 10 hours later a smaller cell mass, the so-called cumulus, separates from the primary thickening and migrates towards the rim of the germ disc (Figure 7, B). During migration the cumulus initially remains connected with the primary thickening via a thinner stalk of cells. As soon as the cumulus cell mass separates from the primary thickening, the cells of the primary thickening start dispersing across the entire germ disc. This leads to the dissolution of the primary thickening until it is barely visible anymore. However, at approximately 70 hours after egg deposition, the connection between the dissolving primary thickening and the migrating cumulus ruptures and at the same time the remaining cells of the primary thickening congregate again, thus forming the secondary agglomeration (Figure 7, C). At approximately 72 hours after egg deposition the cumulus cells have reached the rim of the germ disc and the cells of the secondary agglomeration have attained their maximum density (Figure 7, D). Now the secondary agglomeration starts dissolving, the cells segregate and form a posterior lobe, a new and enlarged structure, which will give rise to the posterior growth zone (also referred to as segment addition zone). This event marks the symmetry break that defines the posterior and anterior ends of the germ band and the formation of a bilaterally symmetric germ band is completed at approximately 82 hours after egg deposition, when the germ band shows the first morphological signs of segmentation (Figure 7, E).

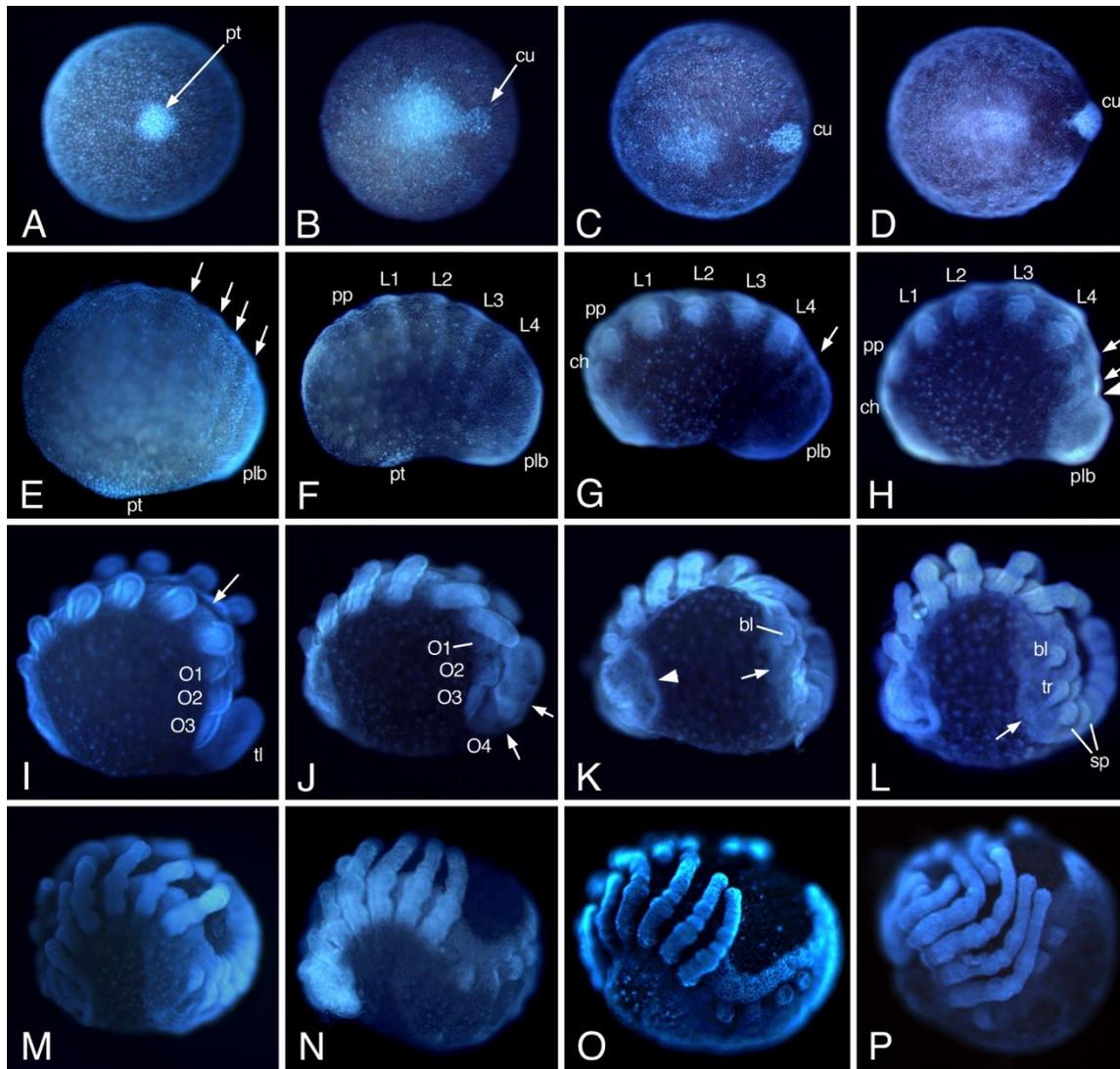


Figure 7: Consecutive stages of germ band formation and differentiation in *P. phalangioides*

A: Germ disc with the primary thickening of cells in its center. **B:** Cumulus cells separate from the primary thickening. **C:** The cumulus is fully separated and secondary aggregation has formed. **D:** Cumulus has reached the rim of the germ disc. **E:** Pear-shaped germ band has formed; segments are visible in the posterior part (arrows). **F:** Anterior segments are visible; embryo attains bean-shape. **G:** First opisthosomal segment is visible (arrow). **H:** Tail bud and prosomal limb buds form. More opisthosomal segments are present (arrows) or are in statu nascendi (arrowhead). **I:** Ventral sulcus forms (arrow); tailbud elongates into tail. **J:** Opisthosomal limb buds appear; tail segments are visible. **K:** Brain differentiation begins (arrow); dorsal tissue develops on the opisthosomal segments (arrowhead). **L:** Dorsal tissue grows further. **M:** Dorsal tissue overgrows the yolk (arrowhead); tail segments are sequentially folded down on the yolk (arrow). **N:** Dorsal tissue is almost closed (arrowhead); most opisthosomal segments are now folded down on the yolk (arrow). **O:** Dorsal closure complete (arrowhead). **P:** Prosomal shield forms (arrowhead). For further details for each stage please see text. All embryos shown in E-P are shown in lateral view and oriented with anterior on the left. Abbreviations: bl, book lung bud; ch, cheliceral segment; cu, cumulus; L1-L4, leg segments 1 to 4; O1-O10, opisthosomal segments 1 to 10; plb: posterior lobe; pp, pedipalpal segment; pt, primary thickening; sp, spinnerets buds; tl, tail; tr, trachea bud.

4.1.1 Segmentation and further differentiation of the germ band

In the early germ band the anterior region is poorly defined, but in the posterior half four segment precursors are already visible (Figure 7, E). The embryo now appears pear-shaped because the posterior growth zone slightly protrudes and takes up the entire posterior end of the embryo. The cumulus is still visible on the opposite side of the germ band (the future dorsal side, but no dorsal tissue has developed yet) (Figure 7, E), but will gradually disappear (Figure 7, F) until it has completely vanished at around 96 hours after egg deposition (Figure 7, G). At approximately 92 hours after egg deposition, the germ band is more differentiated also in the anterior portion and the whole embryo is now bent into a bean-shape (Figure 7, F). The segmental anlagen of the cheliceral, pedipalpal and the four walking legs are morphologically visible, and at the molecular level a further segmental anlage is already present in the posterior growth zone as visualized by the expression of the segmental gene *engrailed* (*en*) (Figure 8, A). At approximately 96 hours after egg deposition the prosomal segments are fully delineated and the first opisthosomal segment is now also morphologically separated from the posterior growth zone (Figure 7, G). At this stage also the specification of the prosomal limb buds starts and slightly later, at around 106 hours after egg deposition, the buds are formed (later referred to as limb bud development Figure 7, H). This stage also marks the formation of the tailbud (Figure 7, H), which is the posterior growth zone now strongly bulging outward. By generating more opisthosomal segments over the following stages it produces a "tail" that grows away from the embryo and hangs freely in the perivitelline space (Figure 7, I-L) (see also next chapter). All segments, including those in the tube shaped tail, express the segmentation gene *engrailed* in their posterior portion (Figure 8, B-D). The prosomal appendages continue growing (limb bud elongation Figure 7, I) and at approximately 130 hours after egg deposition also the first opisthosomal limb buds, the buds of the book lungs, are specified (Figure 7, J, K). Later also the buds of the tracheal system and the spinnerets are visible (Figure 7, L). At approximately 143 hours after egg deposition the head lobes enlarge indicating the beginning differentiation of the brain (Figure 7, K). The appendages on the prosoma further elongate during these stages. Although the development of the opisthosoma differs compared to the entelegyne

spiders (described below) general spider tissue rearrangements like inversion (Figure 7, L and M) and dorsal closure (Figure 7, N and O) are also observed for *P. phalangioides*. During these rearrangements the prosomal appendages further differentiate and elongate. One of the last major embryonic movements is the closure of the ventral prosomal tissue (ventral closure, Figure 7, P). During this step the podomers of the prosomal appendages differentiate and become visible and the walking legs undergo intensive elongation (discussed below).

4.1.1 Development of the opisthosoma

An unusual feature of germ band development of *P. phalangioides* is the mode of opisthosoma formation. The posterior growth zone generates the first three opisthosomal segments while it is still closely attached to the yolk (Figure 9, A and B). These three segments are therefore lying directly on the yolk in the same way as all of the more anterior segments (Figure 9, B). Then the growth zone bulges outward and thus all further segments are formed without connection to the yolk in a tube-shaped tail (Figure 9, C-E). The posterior segments in this tail are closed dorsally, but in fact do not have any true dorsal tissue yet. This is revealed by a detailed sequence of segment formation in the tail (Figure 10). The O4 segment is the first opisthosomal segment that is not formed in direct contact with the yolk, but is included in the tail. However, during further development it opens on its dorsal side and the two halves are then folded down on the yolk (Figure 10, A). Now the O4 segment is very similar to the segments O1 to O3, and there is no indication anymore that would reveal that it has not been formed in permanent contact with the yolk like O1, O2, and O3. In particular, like the segments O1 to O3, the O4 segment also has no proper dorsal tissue yet; this tissue forms afterwards during the later phases of inversion (see below). The posterior growth zone continuously adds segments to the tail (Figure 10, B) whereby the tail initially becomes longer. But at the same time more segments are opened "dorsally" and folded down on the yolk (Figure 10, C and D). In this way, the tail becomes shorter after the posterior growth zone has generated the O10 segment and thus terminated its activity (Figure 10, D and E). Finally, all segments are folded down on the yolk and the tail has thus disappeared (Figure 10, F).

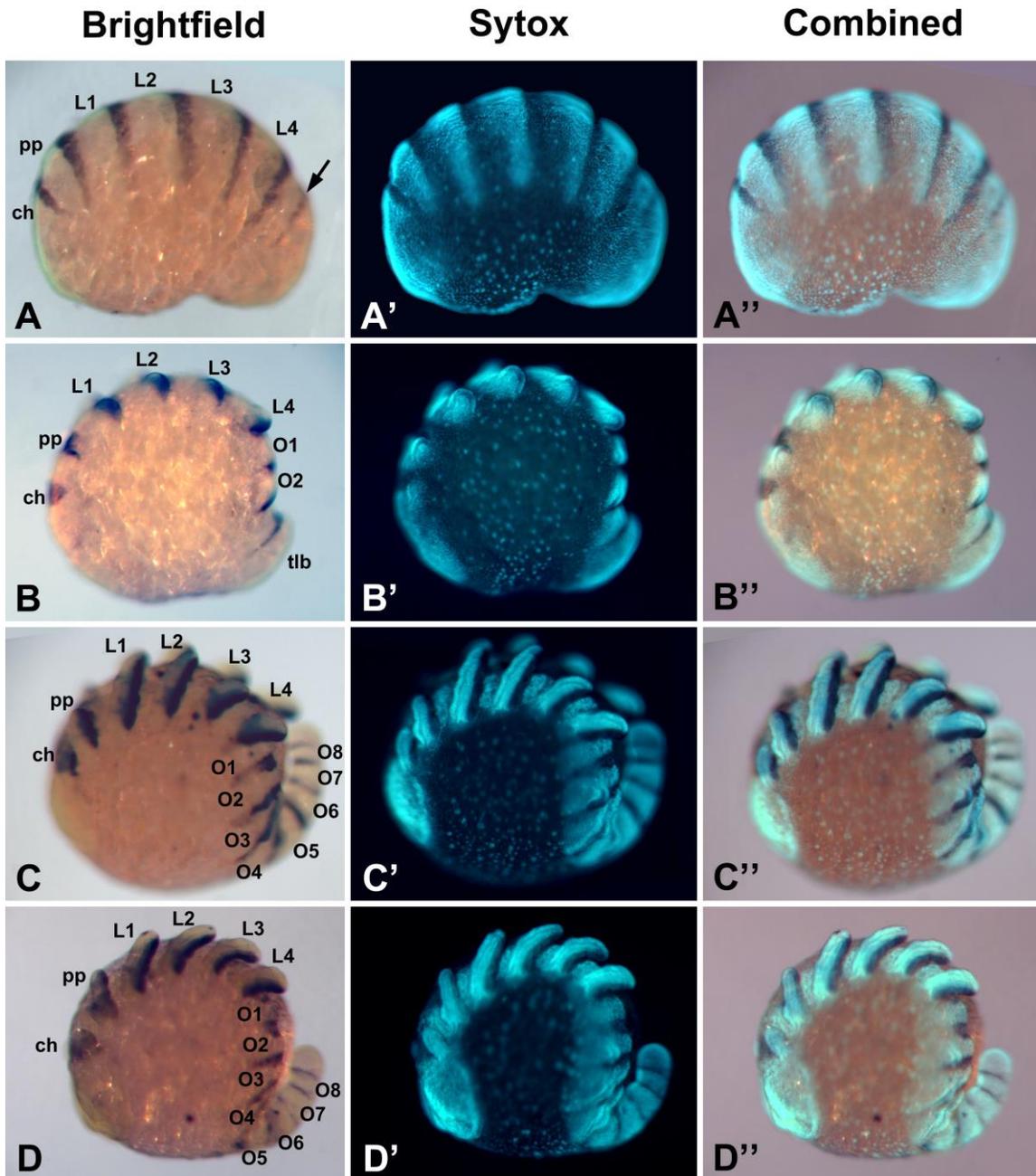


Figure 8: Expression of *engrailed* (*en*) in the germ band of *P. phalangioides*

A: Embryo at the bean-shaped stage. All prosomal segments are visible and express *en* in their posterior portion. Expression of *en* also prefigures the first opisthosomal segment formed by the posterior growth zone (arrow). **B:** After tailbud formation, further opisthosomal segments are formed, all denoted by *en* expression. **C-D:** Segments in the tube shaped tail also express *en*. **A'-D':** The same embryos as in A-D, respectively, cell nuclei have been stained with Sytox Green and visualized with UV light. **A''-D'':** The same embryos as in A-D, respectively. Combined detection of cell nuclei (UV light) and gene expression (white light). All embryos are shown in lateral view with anterior pointing to the left. Abbreviations: see Figure 7.

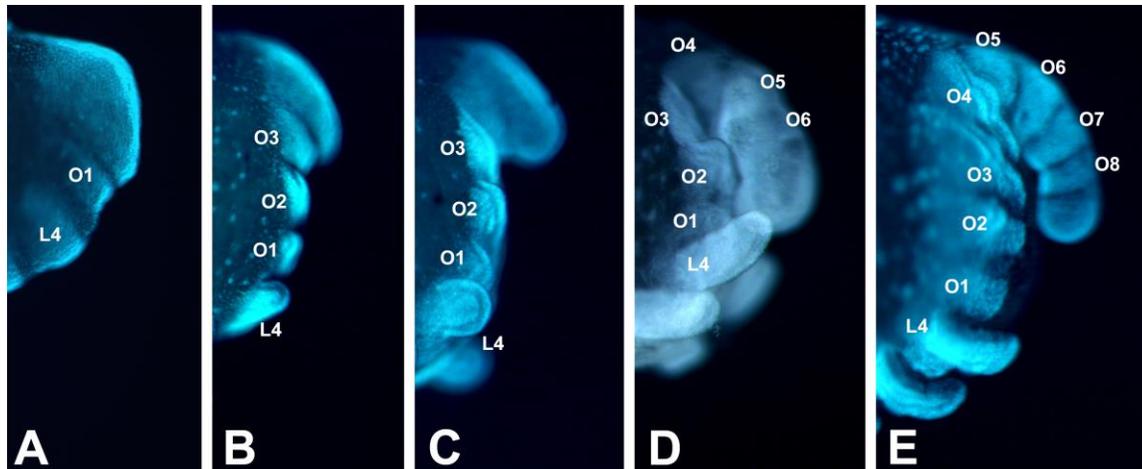


Figure 9: Formation of the tail

A: The posterior growth zone starts bulging outward after it has initiated the formation of the O1 segment. **B-C:** After the formation of the segments O1 to O3, the growth zone orients away from the body (B) and elongates (C). **D-E:** All following segments are then formed in the tail (D) that therefore elongates (E) before it shortens again, because the more anterior segments (i.e. more proximal in the tail) are successively folded down on the yolk (see Figure 9). All embryos are shown in lateral view and oriented with the ventral side to the bottom. Abbreviations: see Figure 7.

4.1.2 Inversion and dorsal closure

As a preparation for inversion, the germ band starts splitting along the ventral midline at approximately 119 hours after egg deposition (Figure 7, I). This formation of the ventral sulcus initiates the ventral opening necessary for inversion, but on the dorsal side no inversion specific events occur until approximately 143 hours after egg deposition when genuine dorsal tissue develops on the opisthosomal segments (Figure 7, K). This dorsal tissue overgrows the yolk during the following stages (Figure 7, L-N) until the two leading edges meet and fuse on the dorsal side during dorsal closure (Figure 7, O). After the dorsal side of the embryo has closed over the yolk there is further differentiation of this tissue when it further contracts and forms the uniform prosomal shield (Figure 7, P) at approximately 215 h after egg deposition. At this stage the tail has vanished and all opisthosomal segments have been folded down on the yolk. Although this process also leads to a partial closure of the ventral side of the embryo (Figure 11, A and B), the ventral side is still open at the time when the dorsal side is fully closed and the prosomal shield forms (Figure 11, C). However, ventral closure proceeds quickly and the prosoma is closed first (Figure 11, D) and shortly after, the ventral side is closed entirely (Figure 11, E).

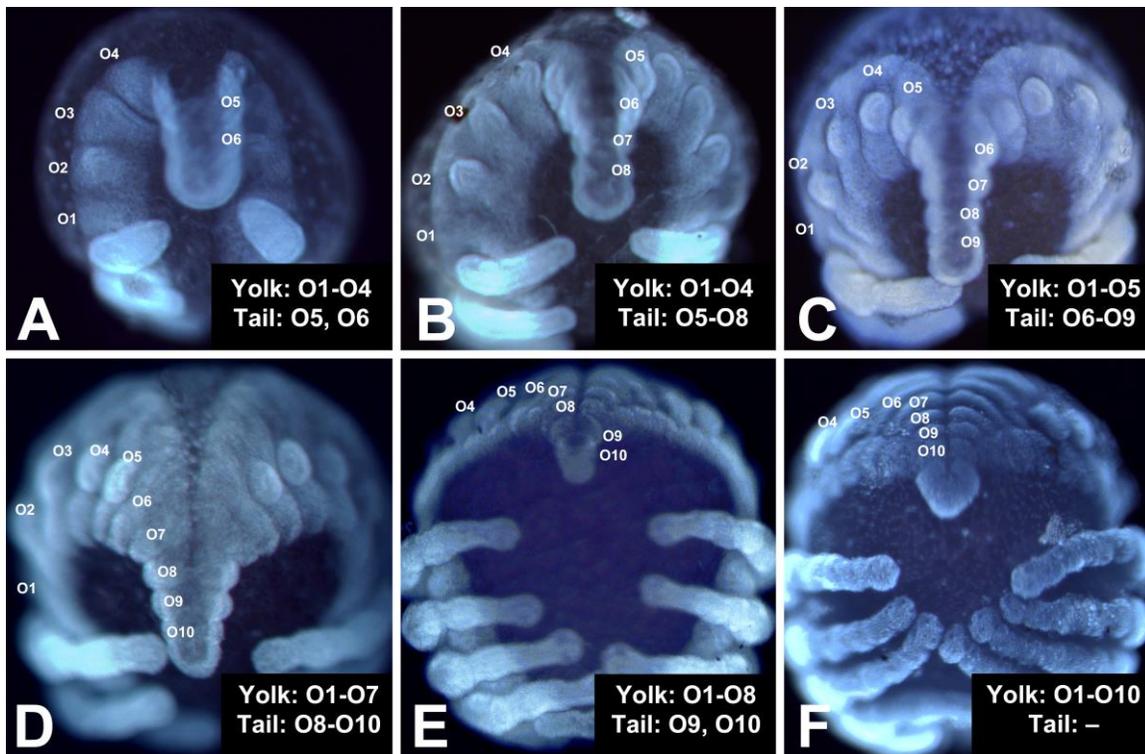


Figure 10: Formation of the opisthosoma by folding down tail segments on the yolk

A-B: The number of segments in the tail grows continuously, but the most proximal tail segment (O4) is already at the "hinge" between the segments on the yolk and segments in the tail (A) or fully folded down on the yolk (B). C-E: The growth zone produces more segments, but the speed of folding down proximal tail segments also increases, thus leading to a shortening of the tail. (F) Because the growth zone stops producing segments at one point, but the folding down of segments at the proximal end of the tail continues, the tail is eventually gone. The number of segments already lying on the yolk and those still in the tail is given in each panel at the lower right corner. All embryos are shown in posterior view with the dorsal side pointing upwards. Abbreviations: see Figure 7.

4.1.3 Growth of the legs

The legs first appear as small limb buds at approximately 106 hours after egg-deposition, and then continuously grow throughout embryogenesis. The legs are not extraordinarily long before dorsal closure (Figure 11, A). However, around the time of dorsal closure the legs have grown considerably and now especially the distal portion of the first leg pair is overlapping, thus giving the impression that the legs are "too long" to fit next to each other and the embryo looks like "folding its arms" (Figure 11, B). This impression is lost again when the dorsal side compacts during the formation of the prosomal shield (Figure 11, C) and this contraction of the dorsal side also pulls apart the entangled legs (Figure 11, D).

Coinciding with ventral closure, there is a sudden increase in leg growth that leads to the legs overlapping again and the tips are often twisted, obviously because there is not enough space now to accommodate the long legs (Figure 11, E). This length growth continues throughout the rest of embryonic development and in embryos shortly before hatching the legs are so long that they are fully wrapped around the body (time lapse movie provided on CD-Rom, Figure 11, F). In the time lapse movie of embryos just before hatching no rearrangement or development of the remaining embryonic tissue can be observed anymore, but instead only the walking legs are growing further.

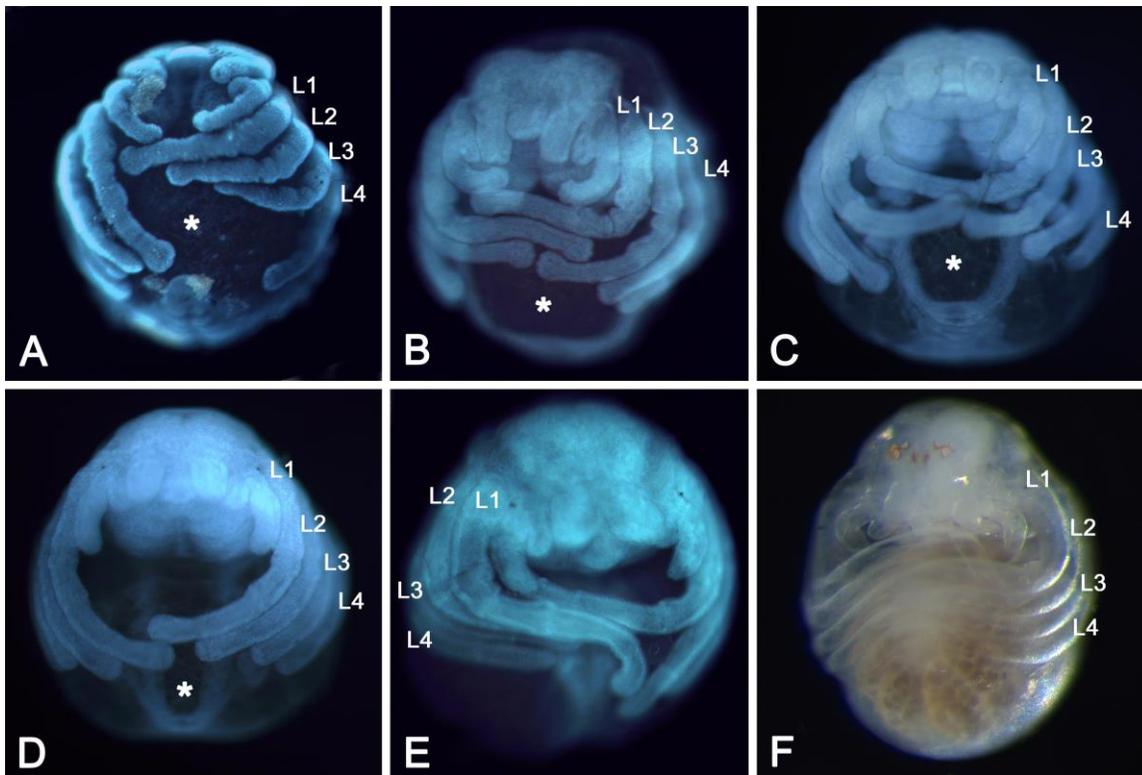


Figure 11: Allometric growth of the legs in *P. phalangioides*, and ventral closure

A: Shortly before dorsal closure the legs are moderately long and the ventral side is still wide open (asterisk). **B:** When the dorsal side is closed the legs (especially the first leg pair) interlace. Ventral closure has started, and the gap between left and right half is closing (asterisk). **C-D:** The formation of the prosomal shield on the dorsal side pulls apart the interlaced legs. Ventral closure proceeds (asterisk) and is first complete in the prosoma (arrow in D). **E-F:** When ventral closure is complete (arrow in E), a sudden increase in leg growth occurs, and in stages shortly before hatching the legs are fully wrapped around the body (F). All embryos are shown in ventral view with anterior pointing upwards. Abbreviations: see Figure 7.

4.2 Generating transcriptomic resources for *P. phalangioides*

Technical advances giving rise to the next generation sequencing (NGS) era facilitate the fast and relatively cheap generation of transcriptomic and genomic resources for various tissues and species and thus revolutionized biological research (Gilad et al., 2009; Oppenheim et al., 2015; Stapley et al., 2010). This helps to overcome labour-intensive cloning strategies to identify homologous genes in non-model organisms. Whereas genomic information represents the general gene set of an organism, transcriptomes have the ability to provide gene expression data for different tissues or developmental stages of an organism. Thus unbiased transcriptomic approaches are in contrast to the candidate gene approach an auspicious stepping-stone to identify species or lineage specific genes or novel functions of pleiotropic genes driving the evolution of morphological differences between species.

This study was among others performed to uncover the genetic basis causing the special walking leg characteristics of *P. phalangioides*. Therefore we performed RNAseq, *de novo* transcriptome assembly and differential gene expression analysis for embryonic and different nymphal tissue. After isolation of a convenient number of leg gene candidates their expression pattern was observed in the haplogyne spider model *P. phalangioides*.

4.2.1 de novo transcriptome assembly and differential gene expression

A preliminary version of the *P. phalangioides* transcriptome was assembled *de novo* and previously published (Janssen et al., 2015). For the unbiased identification of spider or even *P. phalangioides* specific leg gene candidates a second assembly was performed using a newer version of the assembler Trinity. With this improvement a full assembly of 284.577 transcripts and 233.353 predicted genes with a GC content of 37.14% was generated. The N50 of the transcriptome is 1032 bp with a median transcript length of 348 bp and an average transcript length of 659 bp. The *P. phalangioides* transcriptome completeness was assessed using the freely available software BUSCO (Simão et al., 2015). BUSCO uses the comprehensive OrthoDB database (divided into six major phylogenetic clades) of orthologs to assess the completeness of genomes and transcriptomes by comparing the presence or absence of common orthologous genes.

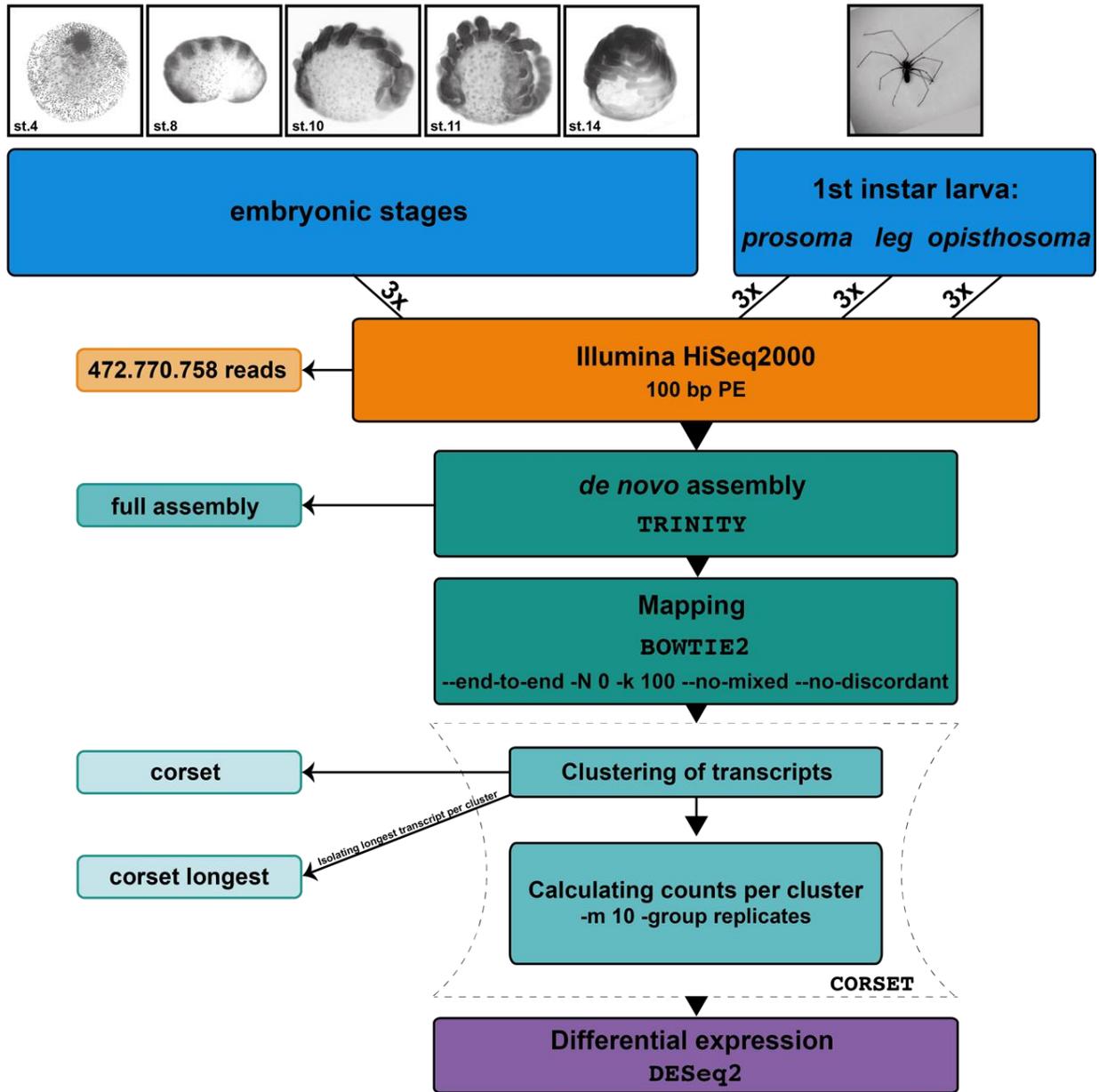


Figure 12: Overview *de novo* transcriptome assembly

Work flow of RNAseq, *de novo* transcriptome assembly and quality optimization prior to differential gene expression analysis. Triplicates of four different samples were sequenced with the Illumina HiSeq2000 generating 100 base pair paired end reads. Afterwards assembly and mapping were performed with the software Trinity and bowtie2 respectively. To reduce the number of overall transcripts the program corset was used to cluster transcripts into gene models and calculating the read counts per cluster. Finally differential gene expression was performed with the R package DESeq2.

82% of the arthropod OrthoDB database orthologs (n= 2.675) were found in the *de novo* assembled transcriptome, and even 89% in case of the metazoan database (n= 843). 33% of the arthropod orthologs are duplicated in the *P. phalangioides* transcriptome. After this quality assessment the program corset was used to cluster transcripts, which might represent isoforms of one gene into gene clusters and counts the reads for these clusters. This clustering reduced the complexity of the reference transcriptome (Figure 13, A) to 142,919 transcripts (Figure 13, panel B) and only 109,640 transcripts (Figure 13, C) if only the longest transcript of each cluster was considered. The reduction of the transcriptome complexity mainly minimized the number of very short transcripts compared to the full transcriptome (Figure 14).

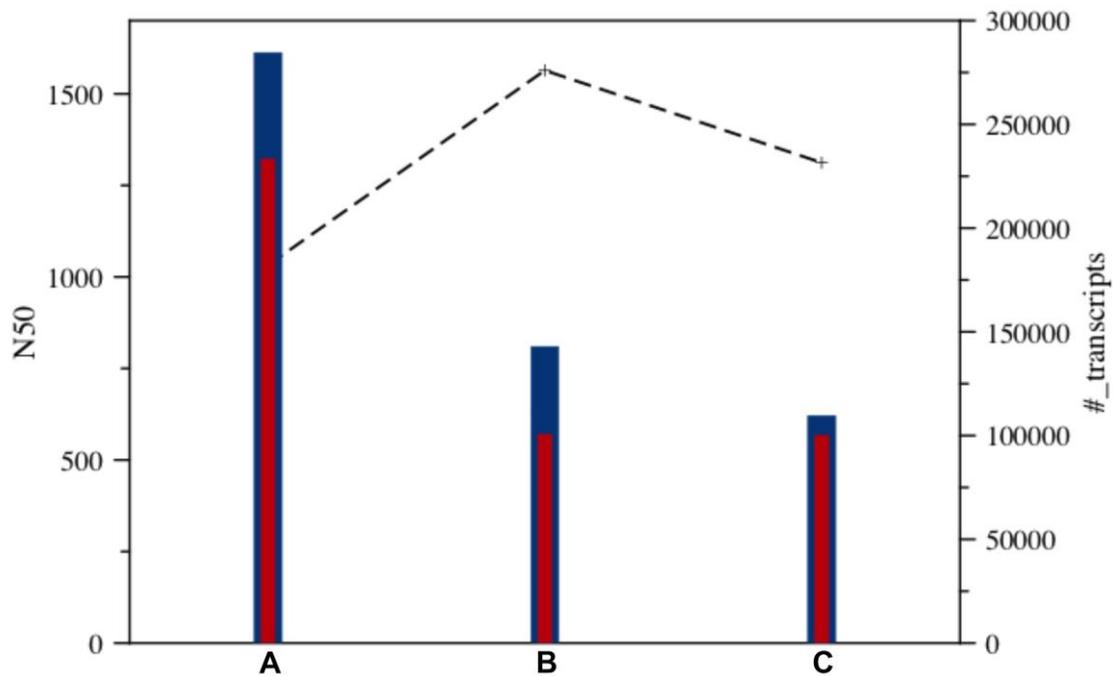


Figure 13: Comparison of gene and transcript numbers after corset

Depicted are the N50 (dashed line) and the number of transcripts (labeled in blue) and genes (red) in the *de novo* assembled transcriptome before (A) and after optimization with corset (B). C: Total numbers of transcripts and genes when only the longest transcript of the gene cluster was considered.

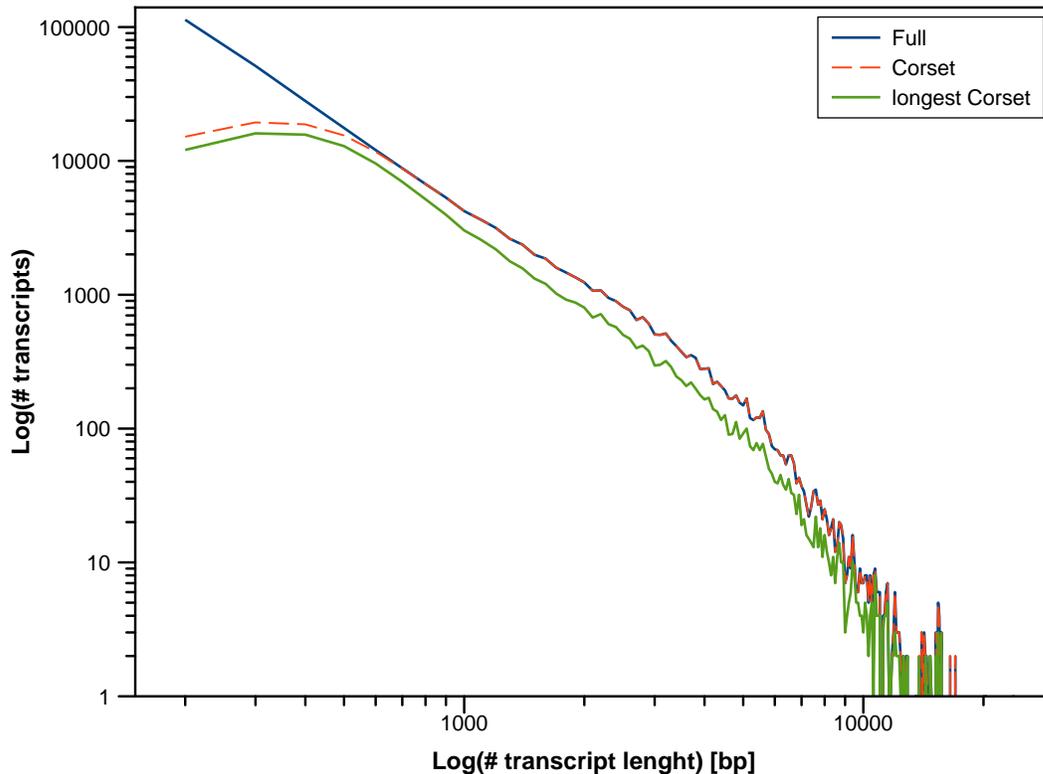


Figure 14: Dependence of transcript number and transcript length

This graph shows the relation of transcript length and number of the *de novo* assembled transcriptome before and after optimization. The blue line represents the distribution of the length of all transcripts without optimization. The dashed red line shows that after optimization with corset especially short transcripts were discarded what resulted in the overall reduction of transcript number. In contrast the green line shows the length distribution of only the longest transcript of the corset gene cluster. This shows a general reduction of transcript number independent for the length of the transcripts.

Based on the read counts the distribution and relatedness of the triplicates and the different sample tissues was analysed with a principle component analysis. The three biological replicates of the different tissues are always closely grouped together. Also biologically more similar tissues like the samples of the legs (green Figure 15) and the prosoma (magenta, Figure 15 (coxa, pedipalps and chelicera attached)) are grouped more closely, than less related tissues like the opisthosoma (blue, Figure 15).

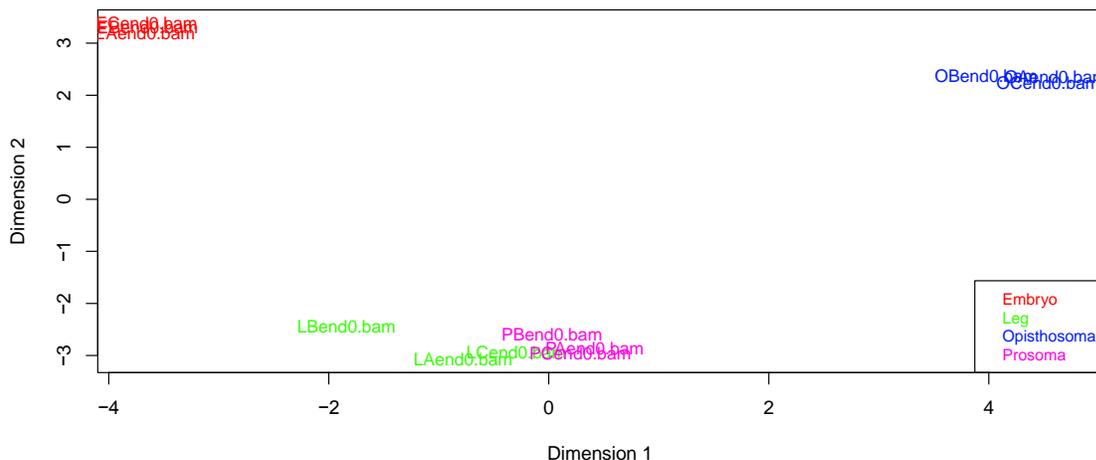


Figure 15: Principal component analysis of the triplicates for the different sample tissues

The principle component analysis based on read counts shows that the triplicates of the different sample tissues are always grouped closely together. The triplicates are named as the following: Embryo tissue (red): EAend0.bam, EBend0.bam, ECend0.bam; Prosoma (magenta): PAend0.bam, PBend0.bam, PCend0.bam; Opisthosoma (blue): OAend0.bam, OBend0.bam, OCend0.bam and Walking legs (green): LAend0.bam, LBend0.bam, LCend0.bam. The embryonic samples and the opisthosomal samples are found in upper proximity of principle component 2 but are far away from each other along principle component 1. Walking leg and prosomal samples are grouped more closely together in the most negative part of principle component 2.

In order to identify genes with different expression levels in the different tissue samples, the read counts of the different stages and tissues were compared using the program DESeq2. The comparison of mixed embryonic tissue samples to all nymphal tissue samples revealed the highest number of up regulated genes (14.212) (Table 7). Comparisons of prosomal tissue samples to walking leg and opisthosomal tissue samples revealed around 7.000 up-regulated genes. A similar amount of up-regulated genes was observed for the opisthosomal tissue samples compared to the other two nymphal tissues. Comparison of the walking legs samples to the prosomal and opisthosomal tissue samples indicated only 1.490 up regulated genes. The number of down-regulated genes ranges between 3.000 and nearly 5.000 for the different comparisons of the nymph tissue samples, and is lower than the identified 9.000 down-regulated genes in the comparison of the embryonic tissue samples to the samples of nymphal tissues. In this study I focused on the 1.490 genes up-regulated in the legs and their possible role in the specialized leg formation of *P. phalangioides*.

Table 7: Differently expressed genes in the different tissues sequenced for *P. phalangioides*

The list shows the number of genes, which were significantly differentially expressed ($\text{padj} < 0.05$) in the different comparisons. i) Comparison of the embryonic samples to all nymphal samples. ii) Comparison of freshly hatched nymph prosoma samples to freshly hatched nymph walking leg and opisthosoma samples. iii) Comparison of the walking leg samples to the samples of the prosoma and opisthosoma. 1.490 genes were identified to be significantly up regulated in the walking legs of freshly hatched nymphs indicated in green. iv) Comparison of the opisthosomal samples to the samples of the walking legs and prosoma.

	i) 1st instar larva	ii) prosoma	iii) leg	iv) opisthosoma
Up regulated	14.212	6.934	1.490	6.976
Down regulated	9.158	3.071	3.197	4.711

4.2.2 Leg candidate genes

In total 1.490 transcripts were identified to be up-regulated in the walking legs of freshly hatched nymphs (Table 7, highlighted in green). Functional annotation (using gene ontologies) of these overexpressed genes was performed with the Blast2Go annotation pipeline in addition to the general transcriptome annotation with SwissProt, Uniref90 and the *P. tepidariorum* genome gene set. The gene ontology (GO) term analysis (Figure 16) revealed many general processes like transcriptional regulation, cuticle formation metabolic and cellular processes. But also growth related processes like regulation of cell proliferation and negative regulation of apoptosis were found. Based on the GO term analysis developmental genes involved in the Notch signalling pathway and in imaginal-disc derived leg segmentation are especially up-regulated in the walking legs of the *P. phalangioides*.

Table 8: Filtering criteria to reduce the number of up-regulated leg gene candidates

Criteria	# candidates
leg up $\text{padj} < 0.05$	1.490
basemean >10	1.215
Embryomean >50	582
sort log2fold	best 15
Annotation in any other organsim	9
orf >300	8

To reduce the number of candidates from 1.490 to a number of genes that can be functionally tested, we applied different criteria (Table 8). The rejection of all candidates with generally low expression (base mean lower than ten) and low expression in the embryo (embryo mean lower than 50) resulted in a reduction to 582 candidates.

From these the 15 most differentially expressed candidates (indicated by the log2Fold) were analysed in more depth. Eight of these candidates had homologs in at least one other species and an open reading frame covering at least 300 base pairs (Table 9). Among these candidates were predicted homologs of the homeobox protein *unc4* from *D. melanogaster*, one *odd-skipped* (*odds*) homolog of the human (*odds-related 2 OSR2*), the Kinesin-like calmodulin-binding protein homolog (KCBP) of the Asian rice plant *Oryza sativa subsp. japonica* and the MDS1 and EVI1 complex locus protein, Evi1 from the mouse. The other four candidates only had homologs in the gene set of *P. tepidariorum*. Except for *c105236* and the *unc4* homolog these candidates were successfully cloned (Table 9 columns filled with green) and subsequently analysed for their expression patterns during embryonic development of *P. phalangioides* (see below).

Table 9: Final candidates and their annotation

Cluster	baseMean	log2Fold Change	padj	transcript	length	aug_pred.	sprot_Top_BLASTX
Cluster-82327.0	17607,64	-5,75	3,13E-04	c95092_g2_i1	2143	aug3.g14955.t1	.
Cluster-60373.0	2035,79	-4,72	3,47E-04	c109662_g1_i1	4554	aug3.g26302.t1	.
Cluster-88654.1	23875,43	-4,72	8,71E-03	c109329_g1_i2	2467	aug3.g27504.t1	.
Cluster-91098.5	85,93	-4,69	1,73E-04	c102331_g1_i4	2211	aug3.g27262.t1	OSR2_HUMAN
Cluster-84832.0	273,76	-4,6	6,76E-04	c110656_g1_i4	3115	aug3.g2943.t2	UNC4_DROME
Cluster-73819.0	545,56	-4,31	2,93E-22	c105514_g1_i1	3137	aug3.g4311.t1	KCBP_ORYSJ
Cluster-67258.0	915,66	-4	3,68E-03	c105236_g9_i1	1256	aug3.g22859.t1	.
Cluster-72087.1	736,78	-3,91	2,55E-03	c109624_g3_i2	5505	aug3.g11431.t2	EVI1_MOUSE

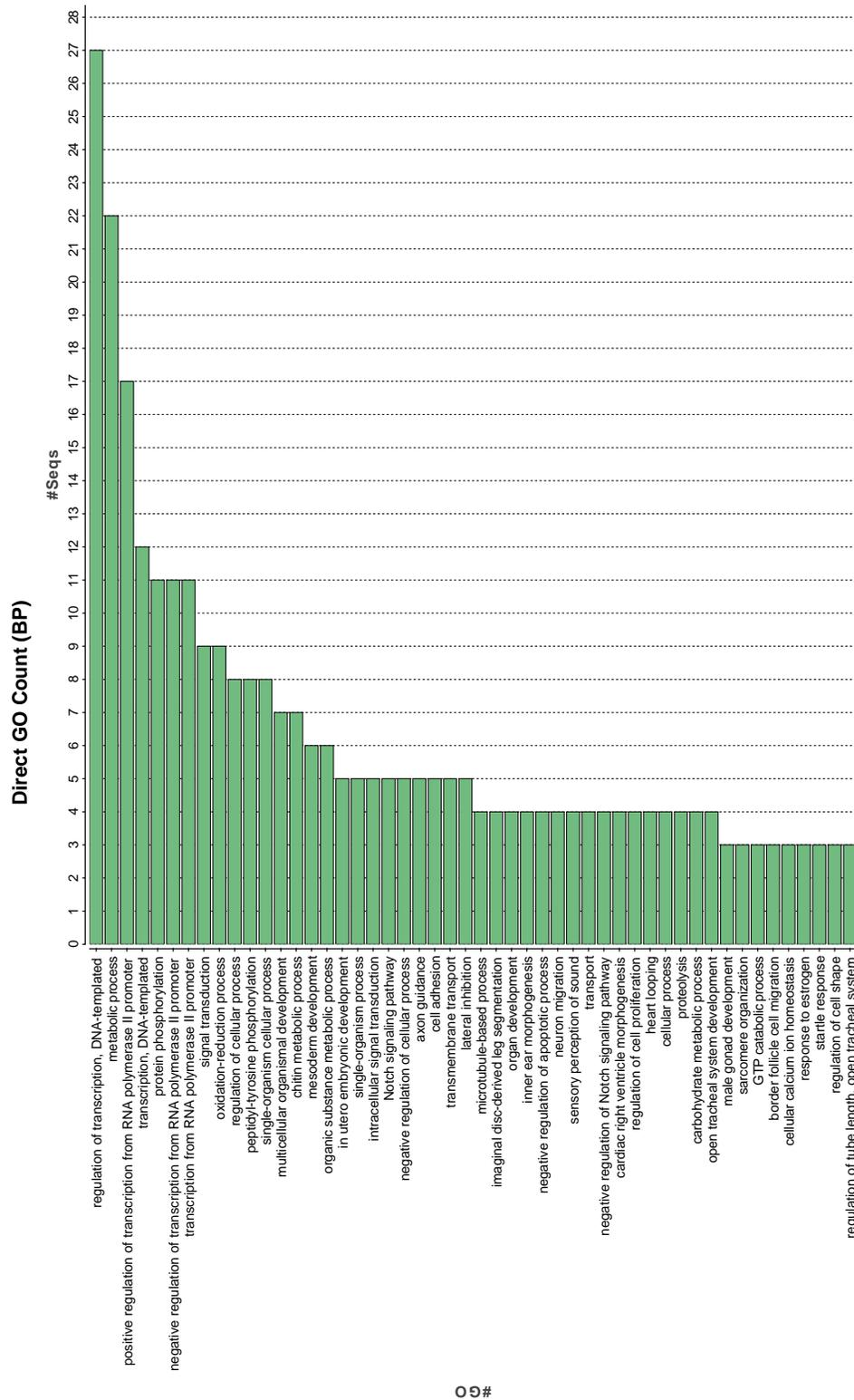


Figure 16: GO-term results for biological processes of the genes up-regulated in the legs
 Overview of most prominent biological processes found in the 1,490 up-regulated leg genes after differential gene expression analysis. On the x-axis the number of sequences found in the set of up-regulated gene genes with the certain GO term is depicted. In the y-axis the most prominent biological processes are listed.

4.2.2.1 Expression of *c102331* in *P. phalangioides*

Expression of the *odd-skipped* ortholog with the transcript ID *c102331* starts as a distal ring in the pedipalps and the first walking leg pair with beginning of prosomal limb bud elongation (Figure 17, A arrowhead). During inversion stages this expression also starts in the remaining walking legs (Figure 17, B arrowheads) and broadens during dorsal closure stages (Figure 17, C arrow). This expansion of expression continues during ventral closure stages covering the area proximal to the distal thickening (Figure 17, D arrow), but is not present in the pedipalps, where only the ring of expression remains (arrowhead).

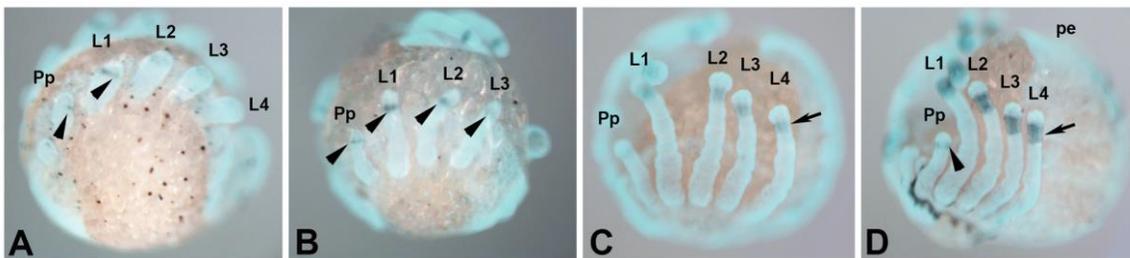


Figure 17: Expression of *c102331 Pp-odd skipped* in *P. phalangioides*

A: Expression of *c102331* in the pedipalps and L1 (arrowhead) of an embryo during prosomal limb bud elongation, lateral view. **B:** Ventro-lateral view of an embryo during inversion. Some of the walking legs were lost during the washing steps of the *in-situ* hybridization. **C:** Ventro-lateral view of an embryo expressing *c102331* at dorsal closure. **D:** Embryo during ventral closure, ventro-lateral view. For detailed explanation of expression see text. Embryos are oriented with the anterior side to the left. Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, pe: posterior end.

4.2.2.2 Expression of *c109624* in *P. phalangioides*

The expression of the *P. phalangioides* Evi homolog *c109624* starts in the ventral neuroectoderm during limb bud elongation stages (Figure 18, A arrow). This expression in the ventral neuroectoderm remains throughout the entire embryonic development, but is always stronger in the head and prosomal part (Figure 18, B and C arrowhead) than in the opisthosoma (Figure 18, G and H arrowhead). During dorsal closure the gene is expressed in a single spot in the distal part of the prosomal appendages and the opisthosomal limb buds on O2, O4 and O5 starts (Figure 18, C and H arrow). In the chelicera *c109624* expression resolves into a complex pattern during ventral closure (Figure 18, D arrowhead), but expression in the pedipalps remains as a single spot. In the walking legs however, additional spots of expression appear in the distal thickening (Figure 18, D arrow).

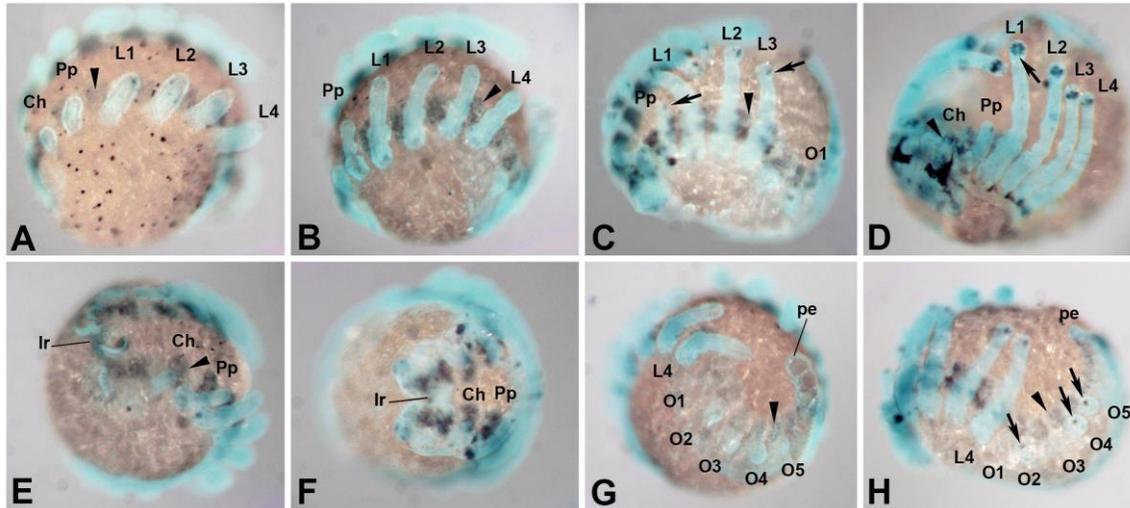


Figure 18: Expression of *c109624* in *P. phalangioides*

A: Expression of *c109624* in the ventral neuroectoderm (arrowhead) of an embryo during prosomal limb bud elongation, lateral view. **B:** Ventro-lateral view of an embryo during inversion. **C:** Lateral view of an embryo at dorsal closure expressing *c109624* in the ventral neuroectoderm (arrowhead) and in a spot in pedipalps and walking legs (arrow). L4 was lost during the staining of the embryo. **D:** Embryo during ventral closure, anterior-lateral view. For detailed explanation of expression in the prosomal appendages see paragraph text. **E:** Expression of *c109624* in the neurogenic ectoderm of the head at inversion of the embryo, frontal-lateral view. **F:** Neurogenic expression of *c109624* in the head of an embryo during dorsal closure, frontal view. **G:** Posterior-lateral view of the opisthosoma during inversion. Weak *c109624* expression in the opisthosomal-ventral neuroectoderm (arrowhead). **H:** Lateral view of *c109624* expression in the ventral neuroectoderm (arrowhead) and limb buds (arrow) of the opisthosoma. A-D and G-H: Embryos are oriented with the anterior side to the left. E and F: Embryo is shown in frontal aspect, walking legs pointing right. Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

4.2.2.3 Expression of *c109329*, *c105514* and *c95092* in *P. phalangioides*

The genes *c109329*, *c95092* and the *ncd* homolog *c105514* are not expressed in the walking legs of *P. phalangioides* during embryonic development. For *Pp-c105514* no embryonic expression was observed at all. *c95092* and *c109329* are expressed in the chelicera in late embryonic stages. *c95092* is expressed in the distal part of the chelicera when dorsal closure begins (Figure 19, A and B arrow), and later during ventral closure this expression restricts to a small spot (Figure 19, C arrow). *c109329* is expressed as a spot in the distal part of the chelicera and starts during ventral closure (Figure 20, C arrow).

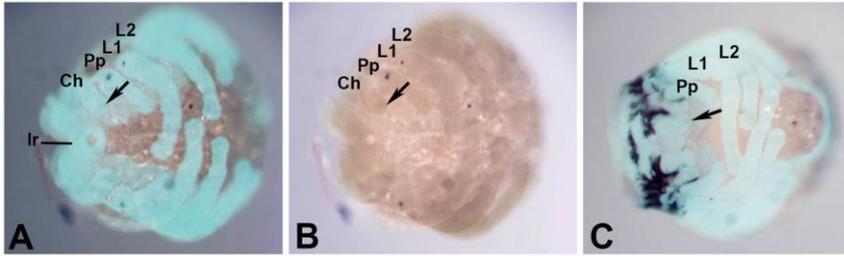


Figure 19: Expression of *c95092* in *P. phalangioides*

A: Dorsal closure of an embryo expressing *c95092* in the chelicera and nuclear staining with Sytox Green, frontal view. B: The same embryo without nuclear staining. C: Frontal view of an embryo during ventral closure. Dark areas are unspecific staining of the forming cuticle in the head and lateral at the base of the walking legs. All embryos are oriented with the anterior side to the left. Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4.

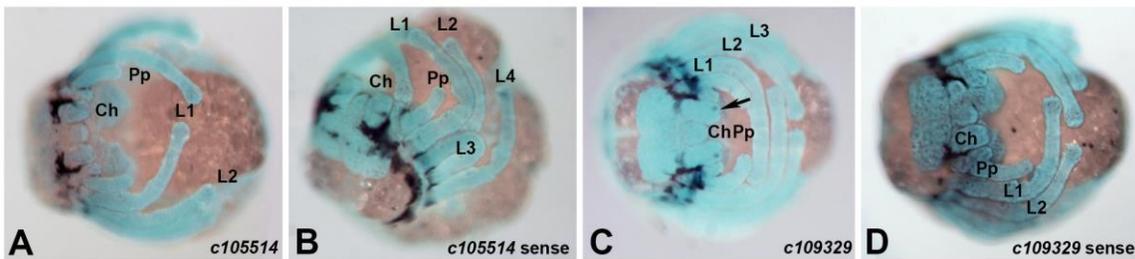


Figure 20: Expression of *c105514* and *c109329* in *P. phalangioides*

A-D: Embryos undergoing ventral closure shown in frontal view. A: Embryo after *in-situ* hybridization with an antisense probe of *Pp-c105514* and nuclear staining with Sytox-Green. B: Embryo after *in-situ* hybridization with a sense probe of *Pp-c105514* and nuclear staining with Sytox-Green. C: Embryo with expression in the chelicera after *in-situ* hybridization with an antisense probe of *Pp-c109329* and nuclear staining with Sytox-Green. D: Embryo after *in-situ* hybridization with a sense probe of *Pp-c109329* and nuclear staining with Sytox-Green. Dark areas in the head and lateral at the base of the walking legs depict unspecific staining of the forming cuticle also present after negative control *in-situ* hybridizations with sense probes. All embryos are oriented with the anterior side to the left. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4.

4.2.2.4 Expression of *c109662* in *P. phalangioides*

The expression of *c109662* starts during dorsal closure very strong covering the labrum and the anterior part of the head lobes. In addition a weak ring of expression in the medial part of the walking legs starts (Figure 21, A arrow). During ventral closure the strong labral expression remains unchanged and additional expression domains appear in the gnathendite and two rings in the pedipalps itself. With proceeding ventral closure more rings at the joints of the developing walking legs appear.

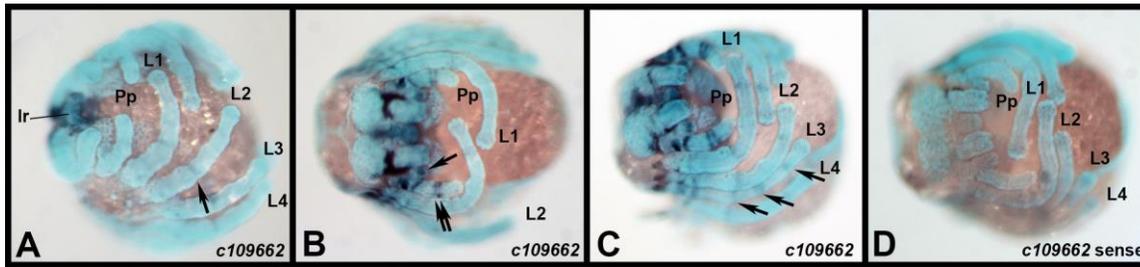


Figure 21: Expression of *c109662* in *P. phalangioides*

Shown are embryos during dorsal (A) or ventral closure (B-D). A: Frontal-lateral view of an embryo expressing *c109662* in the walking legs (arrow). B: Frontal view of an embryo with *Pp.c109662* expression in the pedipalps and gnathendite (arrow). C: Additional expression of *Pp-c109662* starts in the walking legs during ventral closure, frontal-lateral view. D: No expression can be detected after *in-situ* hybridization with a *Pp-c109662* sense probe. All embryos are oriented with the anterior side to the left. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

4.3 Discussion and future work

4.3.1 Two early waves of cell dispersal

The formation of a primary thickening and the migration of the cumulus are features common to all spiders and have previously been described for *P. phalangioides* as well (Claparède 1862; Emerton 1872). However, to our knowledge we provide here the first evidence that this process involves two independent waves of cell aggregation and dispersal in *P. phalangioides*. These waves are difficult to be appreciated from still images, because they directly follow each other and thus the transition between them is easily missed. Our time-lapse recording of this process shows that the first wave leads to the formation of the primary thickening and its dissolution by cell dispersal. The second wave includes some former primary thickening cells, that reverse their movement, come together a second time at the center of the germ disc (secondary aggregation), but then disperse into the posterior growth zone.

Although this second wave of accumulation and segregation of former primary thickening cells has not been described previously in entelegyne spiders, we think that mechanistically there may be no huge difference between germ band formation in haplogyne and entelegyne spiders. This is because the result of this process is virtually identical in haplogyne and entelegyne spiders: (1) In both spider groups, the majority of primary thickening cells, by dispersion, end up as a "second layer" below the germ disc epithelium (summarized in Edgar et al. 2015). (2) In both groups, the posterior growth zone is formed by a small portion of former primary thickening cells at the center of the germ disc, either because these cells do not disperse like the other primary thickening cells (Akiyama-Oda and Oda 2003), or because already dispersed cells return to the disc center and then disperse again into the larger structure of the posterior growth zone (this study). (3) In both groups, the migrating cumulus ends up at a position in the extra embryonic area of the egg and then disappears (overview in Andersson 1973). Note, however, that Edgar et al. (2015) have recently devised a new model of spider development that proposes that the cells of the migrating cumulus also end up in the posterior growth zone. This model is based on studies of germ band formation in the entelegyne species *Latrodectus mactans*, *Latrodectus geometricus*, and *Cheiracanthium*

mildei, but is not compatible with the results from *Parasteatoda tepidariorum* (Akiyama-Oda and Oda 2003) and *P. phalangioides* (this study).

4.3.2 "Preabdomen-postabdomen" subdivision of the opisthosoma

Most studies of spider germ band development have focused on entelegyne spiders and this also includes all recent models for developmental genetic studies (e.g. Abzhanov and Kaufman 2000, Akiyama-Oda and Oda 2003, McGregor et al. 2008). In these spiders the opisthosoma develops from a growth zone that remains attached to the yolk surface throughout development. Therefore all opisthosomal segments that are formed by the growth zone lie on the yolk surface as well, and the opisthosoma gradually grows around the yolk as segmentation proceeds. By contrast, in *P. phalangioides* the formation of the opisthosoma is a two-phase process: the first three segments are formed like in entelegyne spiders, but then the growth zone bulges away from the body and therefore produces a "segmented tail". Intriguingly, this developmental mode has not only been reported from *P. phalangioides* (Claparède 1862; Emerton 1872), but also from several other haplogyne spiders (Holm 1940), Mygalomorphae (bird spiders) (Yoshikura 1958; Crome 1963; Crome 1964) and even from the Mesothelae (segmented spiders), the most primitive spider group (Yoshikura 1954). Thus, although this mode appears to be more complicated and is unusual when compared to other chelicerates and other arthropods, it appears to be the primitive mode of posterior segment addition in spiders.

Holm (1940) has coined the terms "preabdomen" and "postabdomen" for those segments on the yolk and those segments in the tail, respectively. He suggested that these two parts of the opisthosoma in non-entelegyne spiders are a remnant of a phylogenetically older subdivision of the arachnid opisthosoma into two subtagmata, and he compares the spider "pre-" and "postabdomen" to the mesosoma and metasoma of scorpions. Our data do not support this notion, because the initial border between "preabdomen" and "postabdomen" is between the third and the fourth opisthosomal segment, and this border does not correspond to the subdivision of the scorpion opisthosoma. In addition, the border between "preabdomen" and "postabdomen" shifts during development (see Figure 10) thus making a clear distinction of the two parts

anyway difficult. We note, however, that an opisthosomal border between the third and fourth opisthosomal segment is indeed present in Ricinulei (hooded tickspiders). This border could thus be homologous to the initial border between "pre-" and "postabdomen" in the non-entelegyne spiders, but to our knowledge no embryological data is so far available for Ricinulei.

4.3.3 Extreme leg length in pholcid spiders

The length of spider legs in relation to body length varies significantly between different species. Pholcid spiders are remarkable for their extremely elongate walking legs. Interestingly, the allometric growth that leads to the extremely long legs sets in relatively late in embryogenesis. Before dorsal closure no remarkable difference in leg growth to species with shorter legs is detected, e.g. *Parasteatoda tepidariorum* and *Cupiennius salei*. A dramatic increase in leg growth then, however, occurs after dorsal closure and the formation of the prosomal shield. Therefore, future studies of the genetic control of leg length in *P. phalangioides* will have to focus on these stages.

4.3.4 Conserved developmental steps for comparative expressional studies

The detailed study of *P. phalangioides* embryogenesis revealed several differences between the embryonic development of entelegyne spiders, like *P. tepidariorum* and *C. salei* and haplogyne spiders (Mittmann and Wolff, 2012; Wolff and Hilbrant, 2011). These differences include the early organization of the germ disc, differently shaped germ band stages, as the result of the ancestral mode of opisthosoma development and *P. phalangioides* specific extreme leg growth starting during ventral stages. Also the overall developmental time is different for the different spider species, thus hours after egg deposition is not suited to compare temporal expression dynamics of genes. It is, however, important to identify general processes, which are similar during embryogenesis to perform spatial and temporal comparison of candidate genes involved in appendage patterning processes in the two spiders. In both *P. tepidariorum* and *P. phalangioides* prosomal limb bud development starts after germ band formation, which

is followed by limb bud elongation (Mittmann and Wolff, 2012). During inversion and dorsal closure of the embryo the prosomal appendage are further elongated and start differentiating in both species. This differentiation is continued during ventral closure and ends with growth of the appendages, especially the walking legs, which is more drastically observed in *P. phalangioides* compared to the entelegyne spider species observed so far.

In summary this embryonic development study shows, that *P. phalangioides*, *C. salei* and *P. tepidariorum*, share enough main developmental tissue rearrangements, which are sufficient to compare the onset and dynamic of expression pattern between species, especially at early leg development.

4.3.5 General transcriptome optimization

The genome of the fruit fly *D. melanogaster* consists of about 13,000 genes (Adams et al., 2000). Previous findings in the group of chelicerates indicate that their genomes contain a large number of duplicated genes (Di et al., 2015; Nossa et al., 2014; Schwager et al., 2007). A total number of about 30.000 genes were predicted for the genomes of the velvet spider *Stegodyphus mimosarum* (Sanggaard et al., 2014) and the scorpion *Mesobuthus martensii* (Di et al., 2015) and in the current transcriptome of the entelegyne spider *P. tepidariorum* roughly 40.000 genes were predicted (Posnien et al., 2014). The preliminary transcriptome of *P. phalangioides* (Janssen et. al, 2015) contained 444.103 transcripts and 320.428 predicted genes. This transcriptome was sufficient to search for gene homologs, thus a majority of the genes presented in my thesis are based on this preliminary version of the transcriptome. The number of predicted genes is too high to represent the actual gene set of this species and would hamper the identification of differentially expressed genes. The number of predicted genes was reduced to 233.353 in the second assembly, but still higher than observed for the other spider species. Busco analysis of the improved *P. phalangioides* transcriptome revealed that 33% of the genes were duplicated. A similar percentage of duplicated genes was found in the currently annotated genome of the main spider model organism *P. tepidariorum* (personal communication with Dr. Nico Posnien and unpublished data, courtesy of Alistair

McGregor, co-ordinator of the *P. tepidariorum* sequencing consortium). Thus the increased number of genes in the *P. phalangioides* transcriptome compared to the other spiders is unlikely caused by these duplicated genes. Two possible scenarios could have led to this large number of predicted gene: (1) wrongly or incomplete assembled transcripts. (2) Isoforms of one gene were assigned as different genes. To improve the number of genes and transcripts lowly supported transcripts were removed and transcripts probably representing isoforms were grouped into possible gene models based on the raw reads mapped to the assembly. This resulted in the reduction of the overall transcript number by especially losing weakly supported short transcripts, and in the reduction of predicted genes to 100.000. This is still higher than observed in the other chelicerate species and is probably caused by a larger amount of input sequences and the use of different sequencing methods.

4.3.6 Up-regulated genes in the walking legs of *P. phalangioides*

Transcriptomic resources can, in contrast to genomes, also be used to reveal the actually transcribed set of genes in different stages and tissues. To use the transcriptome of *P. phalangioides* not only to facilitate gene cloning, but also for differential gene expression studies, libraries of different nymphal tissues, including walking legs, the prosoma and the opisthosoma were generated and sequenced in addition to the mixture of embryonic stages. The differential expression analysis, comparing all these nymphal tissues to the embryonic data set, showed the largest number of differentially expressed genes, indicating the drastic expressional shift from embryonic to post-embryonic development. Thus this dataset provides a useful resource to study changes in gene regulation during moulting (for example cuticle formation) and has the potential to be used as a platform to identify genes involved in the transition from embryonic to postembryonic development. Comparisons of different the nymphal tissues identified a large number of differentially expressed genes for the prosoma and the opisthosoma. This can be used for further studies focusing on special characters of these particular tissues, like the pedipalps on the prosoma, or the spinnerets and breathing organs on the

opisthosoma. The most restricted sample tissue were the walking legs of the nymphs. Accordingly, the number of up and down regulated genes is lowest for this tissue.

Since my work has focused on the analysis of the genetic cues involved in the development of the special leg characteristics found in *P. phalangoides*, the 1.490 up-regulated genes were studied in more depth. Previous studies from *D. melanogaster* and *C. salei* showed that the Notch pathway and its downstream factors are required for joint formation and leg growth (Celis et al., 1998; Prpic and Damen, 2009). This agrees perfectly with the fact that differential transcriptome analysis of the leg tissue in *P. phalangoides* also identified members of the Notch signalling pathway.

However, not all of these genes could be studied in greater detail and the number of candidate genes was filtered using the above-mentioned cut-offs and restrictions. It could be discussed if the filter criteria were appropriate to find possible gene targets, or if other criteria would have revealed better candidates. Alternative filtering with less restricted cutoffs (e.g. embryo mean =1.000) largely resulted in candidate genes already known for general leg development function during embryonic development including *Dll*, *Notch* and *dpp* (data not shown) (Celis et al., 1998; Goto and Hayashi, 1997; Wu and Cohen, 1999). This again reveals the huge potential of this transcriptomic resource to identify genes involved in leg development. The decision to discard all candidates lacking any annotation in any other organisms was based on two facts: First of all, many of the presumptive *P. phalangoides* specific genes did not show a proper ORF and thus might represent wrongly assembled transcripts; secondly, functional analysis in *P. phalangoides* is challenging and comparison to the expression and function in *P. tepidariorum* would be beneficial.

Four of the identified candidates genes did not show any similar sequences in any other species except for the spider *P. tepidariorum* genome. Also manual blast search did not lead to the identification of homologs in other species. Thus these candidates might represent newly evolved spider specific genes. Three of these candidates were successfully cloned. The expression pattern of the spider specific genes *c95092* and *c109239* did not show any expression in the walking legs, but had a similar expression in the cheliceres. The cutoff for the mean embryo counts was set to a relative low value (50). It is possible that the up regulation of these genes in the walking legs starts at

nymphal stages, which would explain the lack of walking leg expression during embryonic stages. Comparisons of reverse transcriptase PCR for late embryonic stages and legs of early nymphal stages could support this hypothesis. However, it would be interesting to study a possible expression of these genes in walking legs of nymphs. This demands the establishment of an *in situ* hybridization protocol for these tissues, which is not available yet. The spider specific *c109662* is expressed in a ring like manner in the pedipalps and walking legs. Expression in a ring like manner suggests a role in leg patterning or segmentation (Prpic and Damen, 2009). A detailed phylogenetic study including more chelicerate and arthropod species as well as comparison to the expression in *P. tepidariorum*, and functional studies would thus be interesting to clarify the function and origin of this possibly newly evolved leg developmental gene.

The remaining four final candidates were predicted to be homologs of *unc-4*, *odd-skipped* (*odds*), a Kinesin-like calmodulin-binding protein homolog (KCBP) and the MDS1 and EVI1 complex locus protein, Evi1. The homeodomain protein *unc-4* is not intensively studied in *D. melanogaster*; the only findings are based on embryonic expression data, where it is found in postmitotic neurons (Tabuchi et al., 1998). Consistent with these findings, data from *C. elegans* shows that *unc-4* regulates the synaptic input of motor neurons by regulation of neurotransmitter vesicles (Lickteig et al., 2001; Miller et al., 1992). More recent findings show that the initial *unc-4* gene was duplicated in the *Drosophila* genome and gained a new function in the male testis and is responsible for hybrid male sterility (Ting et al., 2004). Thus a functional comparison to the fly homologs would be interesting. Unfortunately cloning of this gene was not successful.

KBPC probably represents a gene with a conserved role during cell division indicated by microtubule motor activity and supported by findings from the plants *Arabidopsis* and tobacco as well as the unicellular flagellate green alga *Chlamydomonas* (Bowser and Reddy, 1997; Dymek et al., 2006). The *P. phalangioides* homolog *c105514* is the only gene with no detectable expression during embryonic development at all. As mentioned above further studies will be performed to analyse a possible late onset of this gene during nymphal stages. In addition a phylogenetic analysis with similar genes from other species would be interesting to identify animal homologs and their function.

The homolog of *Evi1 c109624* was expressed in a spot-like manner in the legs. This kind of expression pattern usually hints to a neuronal function of the gene. Also the expression in the ventral neuroectoderm in the head, prosomal and opisthosoma suggest a role in neurogenesis. Previous findings for the zinc finger transcription factor *Evi1* in mouse showed that this gene controls cell proliferation processes (Hirai, 1999) and it thus plays an important role in general embryonic development. Strong expression was for example found in the urinary system, bronchial epithelium, focal areas within the nasal cavities and cardiac tissues, but also in the developing limbs (Perkins et al., 1991). Furthermore it was found to be essential for proliferation/maintenance of hematopoietic stem cells (Goyama et al., 2008; Yuasa et al., 2005) and thus an interesting oncogene to study leukemogenesis (Hirai, 1999; Mucenski et al., 1988). Also in zebrafish it is involved in hematopoietic stem cells regulation (Konantz et al., 2014, 2015) and nephrogenesis together with Notch signalling (Li et al., 2014). Results for the chicken suggests that its function is required for the size of cartilage elements in the limbs (Celá et al., 2013). These functions in various vertebrates hint to a conserved role of this gene in the control of growth via regulation of cell proliferation and thus it is a promising candidate to study the extreme leg growth of *P. phalangioides*. Many studies were performed to understand its oncogene function in vertebrates (Goyama et al., 2008; Hinai and Valk, 2016). Studies in arthropod species could reveal general regulatory functions of this gene with other developmental genes. However, functional studies are required to analyse the exact function of the spot-like expression of *c109624* in *P. phalangioides*.

The other candidate is a homolog of *odd-skipped (odds)*. *odds* has a conserved function during leg patterning and development (Angelini et al., 2012), especially in the distal part of the fly leg (Hao et al., 2003). The expression pattern in the tarsal region of the walking legs of *P. phalangioides* thus agrees perfectly with previous findings. Because *P. phalangioides* is one of only a few spider species with tarsus segmentation we performed preliminary expressional studies of two *odds* related genes in *P. tepidariorum*. Both genes are expressed in the walking legs of *P. tepidariorum*. The paralogs are expressed in differing number of rings and throughout the entire walking legs (Supplemental Figure 2), but always different to the expression found in *P. phalangioides*. This expressional difference of the genes in *P. phalangioides* and *P. tepidariorum* suggests that they could

indeed drive tarsus developmental differences. However, three *odds* related homologs are found for the other entelegyne spider species *C. salei* (Damen et al., 2005). Thus it cannot be excluded that there is a possible third *P. tepidariorum odds*, which is more similar to the one studied in *P. phalangioides*. Unfortunately expressional data for the *odds* homologs the walking legs of *C. salei* are lacking. However, a preliminary phylogenetic analysis (Supplemental Figure 1) grouped the *odds* homolog *Locus_17047* of *P. tepidariorum* and the *P. phalangioides odds* homolog studied here into one monophyletic group. Thus functional studies of the *odds* homologs in both spider species would be interesting to study a potential function involved in the different tarsus morphologies found for the two spiders.

In summary the expression pattern analysis showed that three of the six studied candidates showed expression in the legs. More generally the outcome of the GO enrichment and the identification of gene candidates involved in cell division, cell proliferation and limb development and the suggested unique *odds* expression in the *P. p* tarsus indicate that the unbiased transcriptome approach and the methodical pipeline led to previously neglected interesting candidates.

5. Leg gap genes in appendage development of spiders

Previous research showed that the formation of the proximo-distal axis in walking legs of arthropods is conserved and requires the leg gap gene function of the genes, *homothorax* (*hth*) and *extradenticle* (*exd*), *dachshund* (*dac*) and *Distal-less* (*Dll*) (e.g. Janssen et al., 2010). Previous findings from the spiders *Cupiennius salei* (*C. salei*) and *Acanthoscurria geniculata* (*A. geniculata*) showed that *exd* and *hth* are duplicated in these spiders (Pechmann and Prpic, 2009; Prpic et al., 2003). In the following I present the phylogenetic relationships, expression patterns and gene function analysis of *hth*, *exd* and *dac* in two additional spider species *Parasteatoda tepidariorum* (*P. tepidariorum*) and *Pholcus phalangioides* (*P. phalangioides*). *Dll* is in contrast to *exd* and *hth* not duplicated and was already studied in *P. tepidariorum* and *P. phalangioides* (Pechmann et al., 2011.)

5.1 homothorax and extradenticle

The gene *hth* in *Drosophila melanogaster* (*D. melanogaster*) encodes a protein that contains two highly conserved protein domains. The MEIS domain is a protein-protein binding domain that mediates the binding between *Hth* and its co-factor *Exd* (Pai et al., 1998; Rieckhof et al., 1997). In addition to the MEIS domain the *Hth* protein also contains a modified homeodomain (TALE-HD) that facilitates the binding of *Hth* to DNA (e.g. Jacobs et al., 1999; Mann and Affolter, 1998). In many cellular contexts *Hth* first binds to *Exd* in the cytoplasm, and the pair is then translocated to the nucleus where they bind DNA together with several Hox proteins. The different Hox proteins recognize very similar DNA sites and thus must rely on protein partners like *Hth* and *Exd* to obtain additional binding specificity. *Hth* and *Exd* cooperate with most Hox genes and are also known to team up with other DNA binding proteins as well. Therefore, *Hth* and *Exd* are involved in the formation and function of a large number of organs and tissues, for example leg and antenna specification, eye development, renal tubule growth, muscle fiber identity, and regulation of neuron differentiation (Bryantsev et al., 2012; Casares

and Mann, 2001; Dong et al., 2002; Hasegawa et al., 2011, 2011; Pai et al., 1998; Wang et al., 2004).

Hth and Exd are widely conserved in the Metazoa and homologs have been described from all major arthropod clades, onychophorans, and from vertebrates, where *hth* is called Meis and *exd* is called PBX (e.g. Janssen et al., 2010). Interestingly, *hth* and *exd* are duplicated in two spider species *C. salei* and *A. geniculata* and the duplicates are expressed in a pattern differing from the first paralog in the walking legs, which is most drastic for *hth2* (summarized in the Introduction and Pechmann et al., 2010). The novel expression pattern of the *exd2* duplicate is conserved in these two spider species, but the duplicated *hth* gene is expressed in contrast to its paralog in numerous segmental rings of differing number in the walking legs of *C. salei* and *A. geniculata*. This expression pattern suggests that one *hth* copy has retained the original functions of this highly pleiotropic gene, whereas the second *hth* copy has undergone diversification after neofunctionalization. To study the grade of diversification and to confirm the conservation of the *exd* after duplication we studied these genes in two additional spider species *P. tepidariorum* and *P. phalangioides*. In the next section first the phylogenetic analysis and expression of *hth* will be described and followed by the study of *exd*.

5.1.1 Duplicated *hth* and genes in *P. tepidariorum* and *P. phalangioides*

Previous studies in the spiders *Cupiennius salei* (*C. salei*) and *Acanthoscurria geniculata* (*A. geniculata*) have shown that *hth* is duplicated in these spider species (Pechmann and Prpic, 2009; Prpic et al., 2003). In order to investigate whether a set of two *hth* genes is a general character of spiders and probably of arachnids as a whole, we searched for *hth* sequences in the transcriptomes of two additional spiders and scorpions, the cellar spider *P. phalangioides* (Janssen et al. 2015) and the common house spider *P. tepidariorum* (Posnien et al. 2014). And in the genomes of the Chinese scorpion *Mesobuthus martensii* (*M. martensii*) (Di et al., 2015) and the Arizona bark scorpion *Centruroides sculpturatus* (*C. sculpturatus*) (I5K Project, Bark Scorpion Genome Project). These searches and the subsequent large phylogenetic analysis to identify only true chelicerate *hth* orthologs were performed as explained in the methods section. As

expected from previous studies in spiders, we found two *hth* homologs in the transcriptomes of *P. phalangioides* and *P. tepidariorum* (group of *hth* homologs is depicted in red, Supplemental Figure 5 and Supplemental Figure 6). In the large phylogenetic analysis we also included previously cloned sequence fragments of *Pt-hth1* (Khadjeh et al., 2012), *Pt-hth2* (cloned by Sara Khadjeh with degenerated primers described in Prpic et al., 2003) and *Pp-hth1* (also cloned from degenerated primers and prolonged with RACE PCR). The phylogenetic analysis shows that they are largely identical with the sequences found in the transcriptomes of *P. tepidariorum* and *P. phalangioides*. In the genome of *C. sculpturatus* we were able to identify three sequences with similarity to *hth*, from one of which, CSCU004313, is largely comparable to the previously described *C. sculpturatus hth* homolog (Sharma et al., 2015). In the genome of the scorpion *M. martensii* only two sequences were identified with similarity to *hth*. An additional sequence was found for *M. martensii* when searched with the *hth* paralogs of *P. phalangioides* MMa00258, which is only short and thus might be an incompletely assembled fragment.

In the previous studies of *hth* genes in *C. salei* and *A. geniculata* (Pechmann and Prpic, 2009; Prpic et al., 2003), the *hth* genes were assigned to two groups, named *hth1* and *hth2*, on the basis phylogenetic analyses and similarities of the expression patterns. The duplicated *hth* genes from *P. phalangioides* and *P. tepidariorum* can also be placed in these two groups (see next chapters). Interestingly, however, the *hth1* genes from all four spider species are much more similar amongst themselves than the *hth2* genes: pairwise alignments of the full-length spider *hth* protein sequences show that the *hth1* homologs have a sequence similarity ranging from 84.1% to 95.5%, whereas the *hth2* homologs range between 72.5 to 90.3% (Supplemental Table 8 and Supplemental Table 9). The same trend is true for pairwise comparisons with *hth* of *Drosophila* that shows less sequence similarity to the *hth2* paralog protein sequences than to *hth1* (Supplemental Table 6 and Supplemental Table 7). In summary, the *hth1* genes form a more homogeneous group than the *hth2* group, which is more diverse, and the genes in the *hth1* group are also more similar to the single *Drosophila* ortholog than the genes in the *hth2* group.

5.1.2 Phylogenetic analysis reveals duplication of *hth* in spiders

In order to investigate the relationship of the *hth* sequences from *P. phalangioides* and *P. tepidariorum* with the previously known arachnid sequences and the *hth*-related genes from other metazoans, we performed a more detailed phylogenetic analysis with the *hth* homologs found in the transcriptomes of *P. phalangioides*, *P. tepidariorum* (Supplemental Figure 6) and *hth* protein sequences from other chelicerates (*M. martensii*, *Limulus polyphemus*, *Phalangium opilio*, *Centruroides sculpturatus*, *Archezogetes longisetosus*), mandibulates, onychophorans and vertebrates (Figure 22). The vertebrate Meis sequences form a monophyletic group in the tree and are clearly separated from the *hth* sequences from arthropods and the onychophoran. The insect and crustacean sequences also form a monophyletic group that is well separated from the group that contains the chelicerate sequences. Interestingly, the chelicerate sequences are divided into two separate monophyletic groups. The first group contains all the sequences from the horseshoe crab, scorpions, harvestman, and mite, plus the previously described *hth1* copy from *C. salei* and *A. geniculata*, and one *hth* copy each from *P. tepidariorum* and *P. phalangioides*, that we therefore also denote as *hth1* (blue shading, Figure 22).

The second group (green shading, Figure 22) exclusively contains the *hth2* copy from *C. salei* and *A. geniculata*, and the second copy from *P. tepidariorum* and *P. phalangioides* (therefore denoted as *hth2*). These data indicate that, although other existing *hth* duplicates are found in other chelicerates (e.g. in scorpions), the *hth2* gene is an *hth* duplication specific to spiders.

5.1.3 Expression of *hth1* and *hth2* during *P. tepidariorum* and *P. phalangioides* embryogenesis

Previous studies in *C. salei* and *A. geniculata* have already suggested that the two *hth* duplicates differ significantly in their expression patterns (Pechmann and Prpic, 2009; Prpic et al., 2003). We therefore investigated whether this distinction can be confirmed by the gene expression patterns of *hth1* and *hth2* in *P. tepidariorum* and *P. phalangioides*. In the developing germ band of *P. tepidariorum* and *P. phalangioides* *hth1* is strongly

expressed throughout the ventral neuroectoderm, in the anterior part of the head and in the proximal part of the prosomal limb buds.

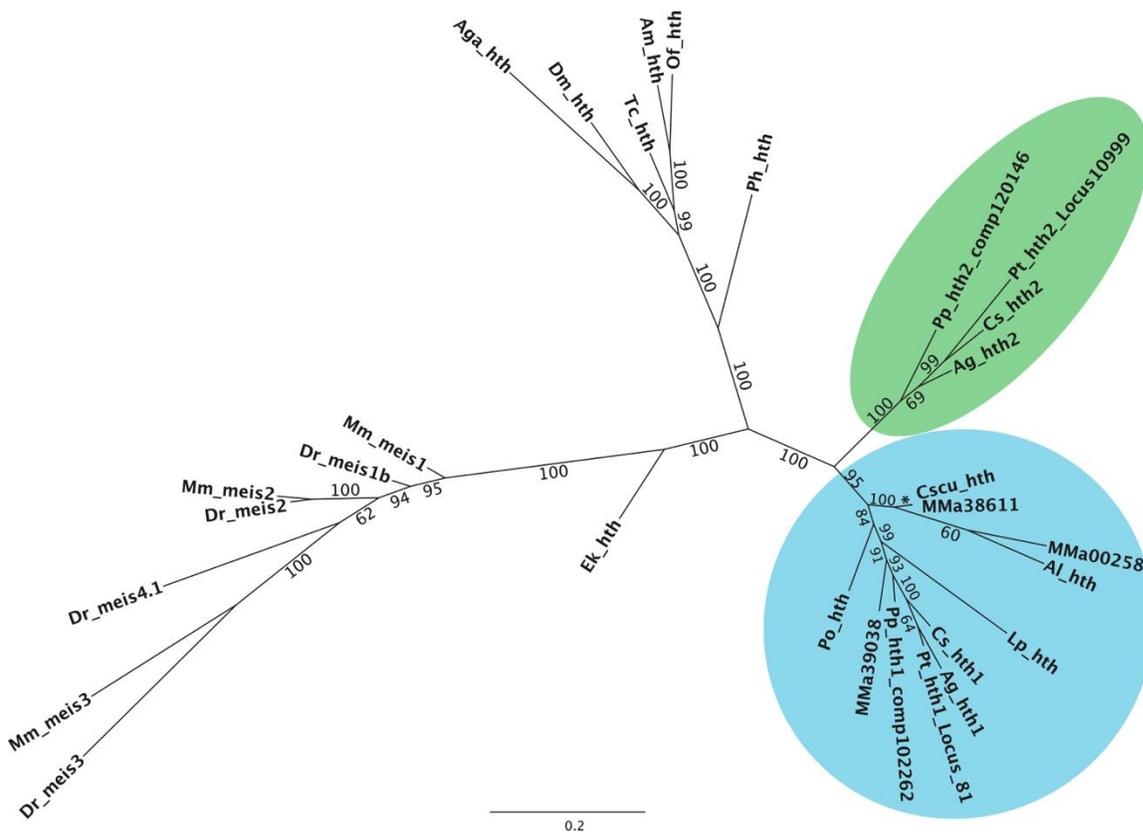


Figure 22: Unrooted 50% phylogenetic majority rule consensus tree of *hth* orthologs

The phylogenetic analysis resulted in three major groups of *hth*. The vertebrate group is shown on the left, the insect *hth* group is at the top and the chelicerate homologs of *hth* are forming two branches in the right. The two paralogs of *hth* after duplication in spiders are present in two distinct phylogenetic groups. The conserved *hth1* forms a monophyletic group with all other chelicerate *hth* homologs (blue circle). The spider *hth2* paralogs cluster together forming a spider specific *hth2* group (green circle) separate from the remaining chelicerate *hth1* paralogs. For *C. sculpturatus* only the previously described *hth* homolog (Sharma et. al. 2015) was used due to missing knowledge about the quality of the assembled *C. sculpturatus* transcriptome. Branch lengths of the tree represent the substitutions per site. Numbers at the edges of the phylogram give the probability of each clade indicated by clade credibility values. The Jones amino acid substitution model was chosen. The average standard deviation of split frequencies dropped below 0.01 after reaching 179.000 generations. A total number of 718 trees were written in two files (each file contained 359 trees of which 270 were sampled). Abbreviations are given in Supplemental Table 5

During brain differentiation the expression in the head becomes restricted to the non-neurogenic ectoderm, the labrum, the stomodeum and two stripes next to the stomodeum (Figure 23 D and Figure 24 D). When the prosomal limb buds start developing, *hth1* is expressed in the entire prosomal appendages excluding the most distal part (Figure 23 A and Figure 24 A). This strong expression in the prosomal appendages remains until late dorsal closure stages, but then decreases slightly, especially in the distal portion, mainly in the metatarsus (Figure 25 F and Figure 26 F). The expression of *hth1* in the head and prosoma is thus virtually identical in both species throughout embryonic development. However, we found differences in the opisthosomal expression patterns at the beginning of inversion. *Pt-hth1* is expressed in the entire opisthosoma at same strength except for the posterior end and one spot next to the limb bud of the second opisthosomal segment (O2 Figure 23 E). In *P. phalangoides* the expression in the opisthosoma is more restricted to presumptive ventral tissue and no expression can be detected in the developing heart tissue (Figure 24 E).

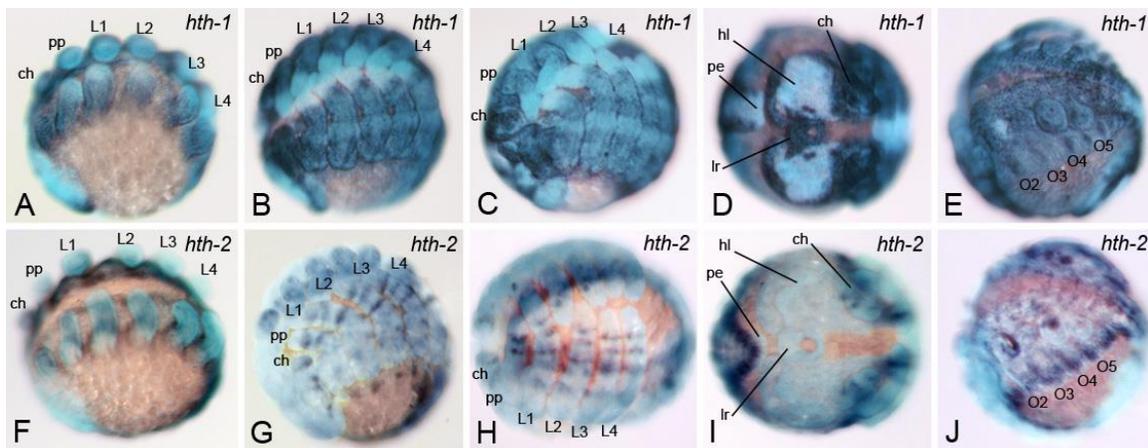


Figure 23: Expression of *hth* paralogs in *P. tepidariorum*

Expression of *Pt-hth1* in the prosomal appendages of embryos during limb bud elongation (A), inversion (B), and dorsal closure (C), all in lateral view. D: Frontal view of *Pt-hth1* expression in the head of an embryo during inversion. E: Posterior view of opisthosomal expression of *hth1* in *P. tepidariorum* during inversion. F: Lateral view of *Pt-hth2* expression in the ventral neuroectoderm during limb bud elongation. G: Rings of *hth2* are expressed in the prosomal appendages during inversion, lateral view. H: Ventral view of *hth2* ring-like expression in the pedipalps and walking legs during dorsal closure. I: No expression of *Pt-hth2* in the head during inversion, frontal view. J: Opisthosomal expression of *hth2* in *P. tepidariorum* during inversion, posterior view. Embryos shown in A-D and F-I are oriented with the anterior side to the left. hl: head lobes, Ir: labrum, ch: Chelicera, pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

Expression of *hth2* differs from *hth1* in terms of timing and pattern. At germ band stages in *P. tepidariorum* *hth2* is expressed more ventrally and weaker as *Pt-hth1* in the neuroectoderm and an expression domain in the head tissue or the developing appendages is absent (Figure 23 F). First expression of *Pt-hth2* in the prosomal appendages is present at early inversion stages, which then differentiates until ventral closure stages (see next chapter), but is always different from *hth1* expression (Figure 23 G). In the opisthosoma *Pt-hth2* has a more diffuse expression pattern and can be found adjacent to the limb buds, with strongest expression near the second and fourth opisthosomal limb bud (O2 and O4) and dorsally in the tissue of the developing heart (Figure 23 J). In *P. phalangioides* *hth2* expression starts in the ventral neuroectoderm at germ band stages comparable to *Pt-hth2* (Figure 24 F). Also the onset of expression in the prosomal appendages of *Pp-hth2* is comparable to *Pt-hth2*, starting during the inversion stage, whereas the pattern of expression is different (Figure 24 G, see also next chapter).

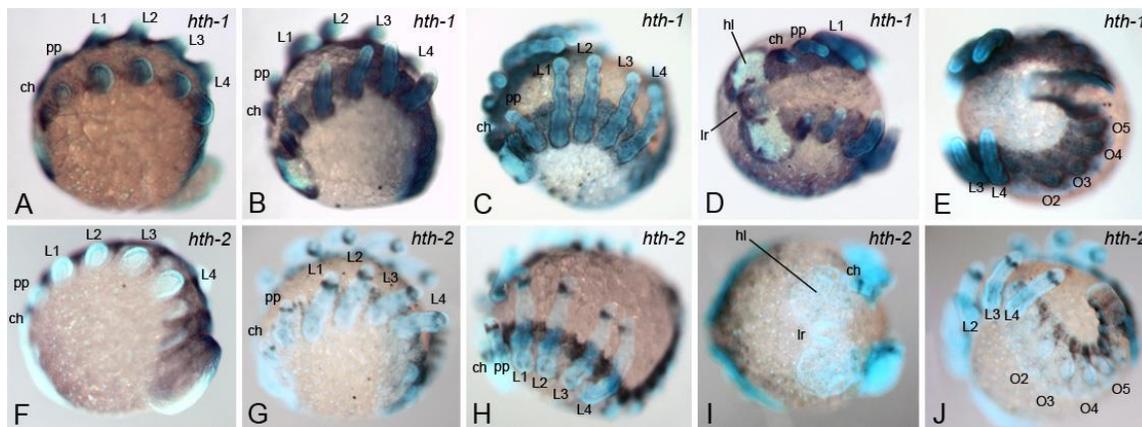


Figure 24: Expression of *hth* paralogs in *P. phalangioides*

Embryonic expression of *Pp-hth1* in the prosomal appendages during germ band stages (A), inversion (B), and dorsal closure (C), all in lateral view. D: Frontal view of *Pp-hth1* expression in the head of an embryo during inversion. E: Posterior view of opisthosomal expression of *hth1* in *P. phalangioides* during inversion. F: Lateral view of *Pp-hth2* expression in the ventral neuroectoderm during limb bud elongation. G: One distal ring of *hth2* is expressed in the prosomal appendages during inversion, lateral view. H: Ventral view of the distal *Pp-hth2* ring in the pedipalps and walking legs during dorsal closure. I: No expression of *Pp-hth2* in the head during inversion, frontal view. J: Opisthosomal expression of *hth2* in *P. phalangioides* during inversion, posterior view. All embryos (except panel E) are oriented with the anterior side to the left. hl: head lobes, lr: labrum, ch: Chelicera, pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5.

The expression of *hth2* in the opisthosoma differs significantly between *P. phalangioides* and *P. tepidariorum*. In *P. phalangioides* *hth2* is not expressed in heart tissue on the presumptive dorsal side, but expression is restricted to the ventral side of the neuroectoderm and a spot ventrally to each opisthosomal limb bud (Figure 24 J).

Because *hth* expression not necessarily also reflects the region of the functional protein and no antibodies are available for *hth* homologs of spiders, we tried to examine the gene function using gene knockdown via parental RNAi. Although two different dsRNA injection procedures were tried no obvious phenotype was observed so far.

5.1.4 *hth* expression in the prosomal appendages of spiders

The largest expression differences of the *hth* paralogs in spiders were found in the prosomal appendages (chelicerae, pedipalps and walking legs). To analyse these differences in more detail we studied dissected prosomal appendages of *P. tepidariorum* and *P. phalangioides* to compare them with the known expression profiles in *C. salei* and *A. geniculata* (Pechmann and Prpic, 2009; Prpic et al., 2003). Expression of *hth1* is very similar in *P. tepidariorum* and *P. phalangioides*. Until early dorsal closure stages *hth1* is strongly expressed in the proximal and medial domain of the pedipalps and the walking legs, but is excluded from the tip of both appendage types. This expression profile is retained throughout most of development; only in embryos undergoing dorsal closure the medial expression becomes slightly weaker (Figure 25 D and F and Figure 26 D and F). Strong expression of *hth1* can also be observed within almost the entire chelicera throughout embryonic development (Figure 25 A-B and Figure 26 A-B).

Expression of *hth2* in the chelicera of *P. tepidariorum* starts as a relatively diffuse domain (Figure 25 H), which later becomes more restricted forming two stripes (Figure 25 I) and finally forms an irregular ventral domain of expression at late ventral closure stages (Figure 25 J). In the pedipalps of *P. tepidariorum* *hth2* expression starts as a weak ring at the distal end of the medial portion of the pedipalp (asterisk, Figure 25 L), that increases in expression levels during further development. In addition, a second ring appears (arrow, Figure 25 M), and also becomes stronger during development (Figure 25 N). In

contrast to the expression in the pedipalps, *Pt-hth2* in the legs is activated as two weakly expressed rings (asterisks, Figure 25 P). The remaining dynamics of the expression pattern, however, are similar in pedipalps and legs: the initial rings become stronger during further development and an additional ring appears proximal to the initial two rings.

The expression pattern of *hth2* in *P. phalangioides* also comprises ring domains, but differs in its dynamics from *hth2* in *P. tepidariorum*. In the chelicera *Pp-hth2* is first expressed as a ring in the medial part, which broadens and partially dissolves during development (Figure 26 G and H). In the pedipalp *Pp-hth2* expression starts as one ring in the distal part, which then dissolves into a diffuse patchwork of distal expression, and in addition a number of weak segmental rings appear more proximally (Figure 26 I and J). In the walking legs *Pp-hth2* expression also starts as a ring in the distal part and is in addition expressed in a spotted manner along the ventral side (Figure 26 K). The strongly expressed distal *hth2* ring in the walking legs broadens during further development, and additional very weak rings appear near the future leg joints (Figure 26 L).

We have also restudied here the expression profile of *hth2* in *C. salei*, because in the original publication (Prpic et al., 2003), only the expression pattern in the fully formed embryonic leg was shown, but the expression profile of this gene is actually more complex. *Cs-hth2* has a diffuse expression in the chelicera with a faint medial ring (Figure 27 A). In early pedipalps it is weakly expressed in two rings in the medial pedipalp part (arrowhead and arrow, Figure 27 B). The proximal one of these two rings broadens (arrow and bar, Figure 27 D) during further development and the more distal ring divides giving rise to two rings (arrow and bar, Figure 27 C-D). Additionally one distal expression domain appears de novo (asterisks, Figure 27 C). In later stages these four rings of expression in the pedipalp become stronger and better distinguishable (Figure 27 D). *Cs-hth2* in early walking legs is expressed in a proximal, medial and distal ring (arrowhead, arrow and asterisks, respectively, Figure 27 E). All three rings broaden during development (indicated by bars in Figure 27 F and G), and then split into two rings (Figure 27 H). This results in the presence of one ring near each of the developing leg joints in late stages of embryonic development

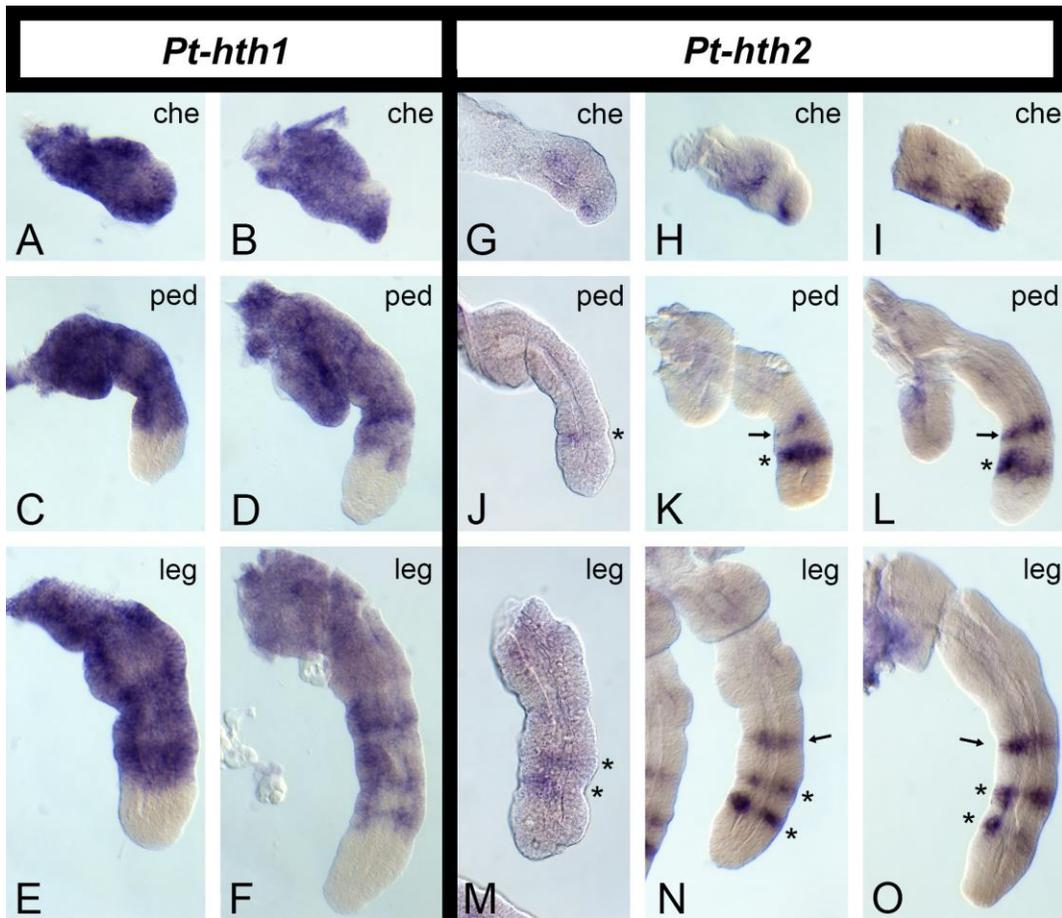


Figure 25: Expression of *hth* paralogs in prosomal appendages of *P. tepidariorum*

A-F: Expression of *hth1* in prosomal appendages of *P. tepidariorum*. A and B: Preparations of chelicera during inversion (A) and ventral closure (B). C and D: Expression of *hth1* in the pedipalps during inversion (C) and ventral closure (D). E and F: Walking legs during inversion (E) and ventral closure (F). G-O: Expression of *hth2* in prosomal appendages of *P. tepidariorum*. G-I: Preparations of chelicera during early inversion (G), late inversion (H) and ventral closure (I). J-L: Preparations of pedipalps during early inversion (J), late inversion (K) and ventral closure (L). The asterisk indicates the *hth2* ring in the distal part of the pedipalp starting during early inversion. The arrow points to the second *hth2* ring appearing during late inversion. M-O: Preparations of walking legs during early inversion (M), late inversion (N) and ventral closure (O). The asterisk marks the start of *hth2* walking leg expression starting directly with two rings during early inversion. The arrow marks the start of expression in a third *hth2* in the walking legs proximal to the two initial rings. che: chelicera. ped: pedipalp. leg: walking leg.

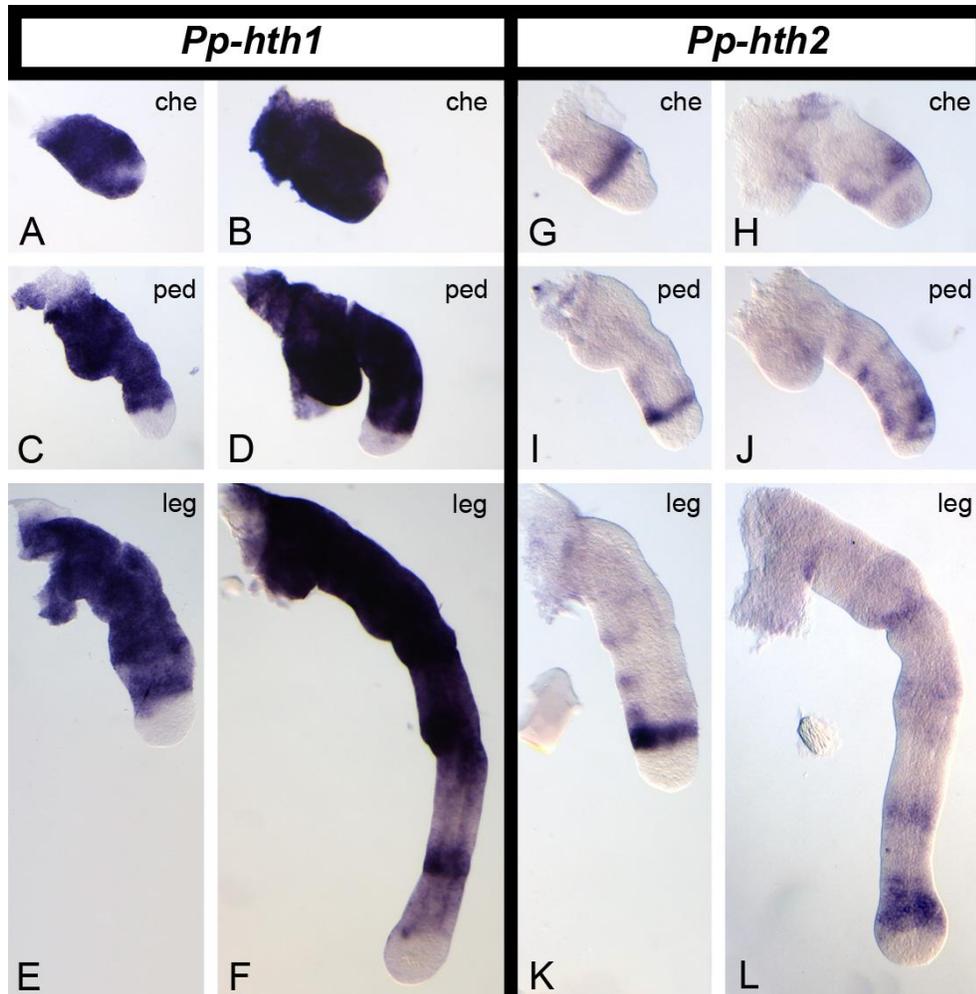


Figure 26: Expression of *hth* paralogs in prosomal appendages of *P. phalangioides*

A-F: Expression of *hth1* in prosomal appendages of *P. phalangioides*. A and B: Preparations of chelicera during limb bud elongation (A) and ventral closure (B). C and D: Expression of *hth1* in the pedipalps during limb bud elongation (C) and ventral closure (D). E and F: Walking legs during limb bud elongation (E) and ventral closure (F). **G-L:** Expression of *hth2* in prosomal appendages of *P. phalangioides*. G and H: Preparations of chelicera during inversion (G) and ventral closure (H). I and J: Expression of *exd1* in the pedipalps during inversion (I) and ventral closure (J). K and L: Walking legs during inversion (K) and ventral closure (L). For detailed description of the expression pattern please see text. che: chelicera. ped: pedipalp. leg: walking leg.

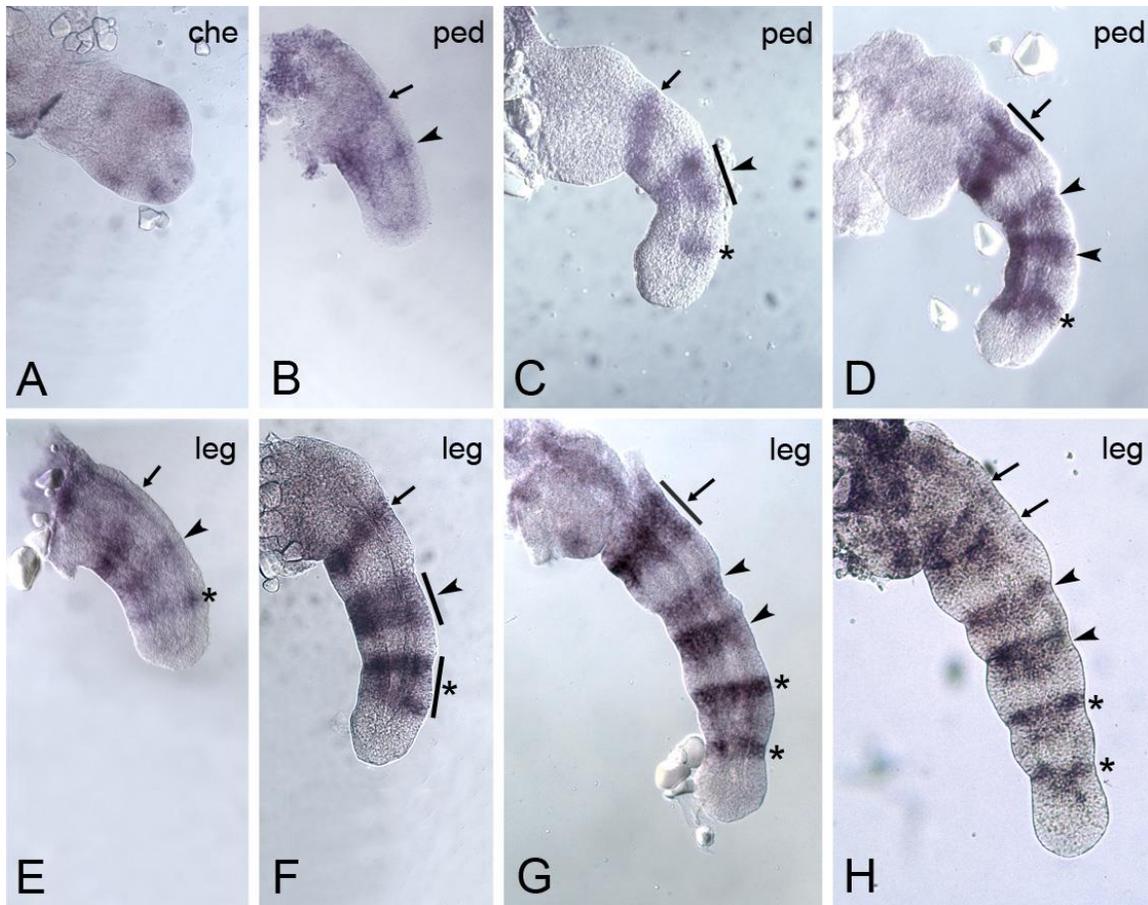


Figure 27: Expression of *hth2* in prosomal appendages of *C. salei*

A: Weak expression of *hth2* in the medial and distal domain of the chelicera during late inversion.
B-D: Expression of *hth2* in the pedipalps of *C. salei* during limb bud elongation (B), during germ band elongation (C) and during mid inversion (D). The arrow points to the proximal *hth2* ring of the pedipalp, which broadens prior to presumptive division during mid-inversion (indicated by the bar). The arrowhead points to the medial *hth2* ring, which broadens shortly before inversion (bar in C) and is then present as two rings during mid-inversion (two arrowheads in D). The asterisk marks the distal *hth2* ring starting to be expressed shortly before inversion. **E-F:** Walking legs expressing *hth2* during germ band elongation (E), shortly before inversion (F), during mid-inversion (G), and during dorsal closure (H). The arrow marks the proximal *hth2* ring and its descendent. The arrowhead marks the medial *hth2* ring and the asterisk marks the distal *hth2* ring and its descendants after divisions. Bars always indicate broadening of the ring expression. A division of the initial ring follows this broadening and results into two independent *hth2* rings. Staining and preparation of the appendages was performed by Dr. Nikola-Michael Prpic-Schäper. che: chelicera. ped: pedipalp. leg: walking leg.

5.1.5 Duplication of *exd* genes in *P. tepidariorum* and *P. phalangioides*

Previous studies of the spiders *A. geniculata* and *C. salei* also found a duplication of *exd* (Pechmann and Prpic, 2009; Prpic et al., 2003). Cloning and expressional analysis of the *exd* homologs in *P. tepidariorum* and *P. phalangioides* was already started before transcriptomic resources were available. With the use of degenerated primers (as described in Prpic et al., 2003) short fragments of two *exd* homologs were obtained in each spider used here. For *P. phalangioides* it was also possible to prolong these sequences using RACE PCR. To identify longer fragments of these *exd* homologs, they were searched in the transcriptomes of *P. phalangioides* (Janssen et. al 2015) and *P. tepidariorum* (Posnien et al. 2014), after they became available. With the standard search procedure merely two possible candidates were found in both spider transcriptomes. Thus the e-value threshold was lowered to e^{10} . To reveal the time-point of the *exd* duplication in the lineage of the chelicerates we also searched for *exd* homologs in the genomes of two scorpions *M. martensii* (Di et al. 2014) and *C. sculpturatus* (15K Project, Bark Scorpion Genome Project). The large phylogenetic tree to identify true homologs of *hth* (indicated in red) and *exd* (indicated in black) clearly shows that the cloned *exd* paralogs are largely identical with the sequences found in the transcriptomes of *P. tepidariorum* and *P. phalangioides* and that there are the only two *exd* homologs present in these spiders (Supplemental Figure 6). In the genomes of the scorpions *C. sculpturatus* and *M. martensii* I in contrast identified five sequences with similarity to *exd*. Whether this is the result of additional scorpion specific duplication events or caused by incompletely assembled transcripts or if these are simply isoforms of one gene was not further investigated here. The paralogs found for *P. tepidariorum* and *P. phalangioides* were named according to sequence identity (Supplemental Table 2) to the sequences previously described for *C. salei*, which is also reflected by similar expression patterns described in the following parts. Thus comp_102492 and Locus_1_Transcript_5866/166847 and the according clones are referred to as *exd1* and comp122137 as well as Locus_1_Transcript_58675/166847 as *exd2*. Similar as found for the conserved *hth1* homologs we see large sequence similarities when comparing sequence identities after multiple *exd* alignment. But in case of *exd* this is true for both

paralogs. Pairwise alignments of the spider *exd* protein sequences show that the *exd1* homologs have a sequence similarity ranging from 87.33% to 96.09% (Supplemental Table 3), and also the *exd2* homologs range within 87.56 to 90.39% (Supplemental Table 4). Even in the multiple alignments of both spider *exd* paralogs the similarities are at least 83.40 % (Supplemental Table 2).

5.1.6 Phylogenetic analysis of *exd* homologs in spiders and chelicerates

This high sequence similarity of and also between the *exd* paralogs is also reflected in the large phylogenetic tree that was performed to identify *exd* and *hth* orthologs in the spiders *P. tepidariorum* and *P. phalangioides*. Although 9002 trees were sampled after 3.000.000 generations the phylogenetic relationships are not properly resolved for the different chelicerate sequences. All chelicerate *exd* (indicated in black, Supplemental Figure 5 and Supplemental Figure 6) are resulting in a polytomy without forming proper paralog clusters indicated by low support values of the posterior probabilities. This is an indication that the paralogs among themselves but also across species have not diversified as rapidly as *hth2* and are more similar instead. To clarify the phylogenetic relationships in more detail we performed a second tree from a subset of *exd* protein sequences (Supplemental Figure 3 and Supplemental Figure 4). This tree is slightly better resolved and three major groups are formed: the vertebrate group containing all PBX paralogs and one outlier from the horseshoe crab *L. polyphemus*; the pancrustacean group including the insect and *Parhyale hawaiiensis* *exd* homologs; and one group including all chelicerate sequences. However, all branches are only supported with low posterior probabilities. Within the chelicerate group, the *exd1* paralogs cluster together, but the support value for this clade is with 50 also very low. In summary also this phylogenetic tree shows that *exd* is in general very conserved on the protein level, as already indicated by similarities after pairwise comparisons, which is true for both paralogs. Thus no prediction of the timing of the duplication event in the lineage leading to the spiders can be clearly stated.

5.1.7 Expression of *exd* during embryonic development of *P. tepidariorum* and *P. phalangioides*

We observed the expression of the two *exd* paralogs, *exd1* and *exd2* in *P. tepidariorum* and *P. phalangioides*. Especially the expression in the prosomal appendages differs between the two paralogs, but this is conserved for both spider species. Expression of *exd1* starts with early limb bud development in the ventral neuroectoderm, the entire chelicera and pedipalps as well as the proximal base of the walking legs (detailed description below) in both spider species (Figure 28, A and arrow, Figure 29, A). In contrast to similar expression in the prosomal appendages, the expression in the head and opisthosoma shows differences. *exd1* is strongly expressed in the labrum of *P. tepidariorum* (Figure 28, E and F), but only very weak in *P. phalangioides* (Figure 29, E). In the opisthosomal neuroectoderm *exd1* is expressed in a more defined way for *P. tepidariorum* compared to *P. phalangioides*. *Pt-exd1* is expressed in a ventral (H, arrow) and a dorsal (H, arrowhead) stripe as well as in a specific pattern for each opisthosomal limb buds, *Pp-exd1*, however, is expressed in a broad dorsal domain. The expression of the second *exd* paralog, *exd2*, differs compared to its paralog, *exd1*, but is very much conserved between the two observed spider species. The expression of *exd2* starts at early limb bud development in *P. phalangioides*, covering the proximal part of the chelicera and as a medial stripe in the other prosomal appendages (arrow, Figure 31, A). In *P. tepidariorum* expression starts slightly later during limb bud elongation, but in a visual identical pattern (detailed prosomal appendage expression description see below). *Exd2* expression in the head of both spiders starts with early inversion in the labrum and as a rim surrounding the head lobes (arrow, Figure 30 F and arrowhead, Figure 31 F). The rim remains in this early anterior head tissue during rearrangements of the head tissue throughout embryonic development (arrow, Figure 30 D and arrowhead, Figure 31 E). The opisthosomal expression of *exd2* in *P. tepidariorum* is first covering the entire dorsal neuroectoderm (arrowhead, Figure 30 G) and gets restricted to a medial stripe in the neuroectoderm covering the opisthosomal limb buds during inversion (arrowhead, Figure 30, H and asterisk, Figure 31, H). In the opisthosoma of *P. phalangioides* *exd2* is expressed similarly to *Pt-exd2* prior to inversion (Figure 31 G), but

differentiates differently during dorsal closure. *Pp-exd2* is then expressed in the opisthosomal neurogenic tissue dorsally to the limb buds (asterisk, Figure 31 H) and ventrally on the limb bud on the fourth opisthosomal segment (arrowhead, Figure 31 H). In summary the expression of each of the *exd* paralogs is conserved in most tissues between the two spider species, but differs when the two paralogs are compared to each other.

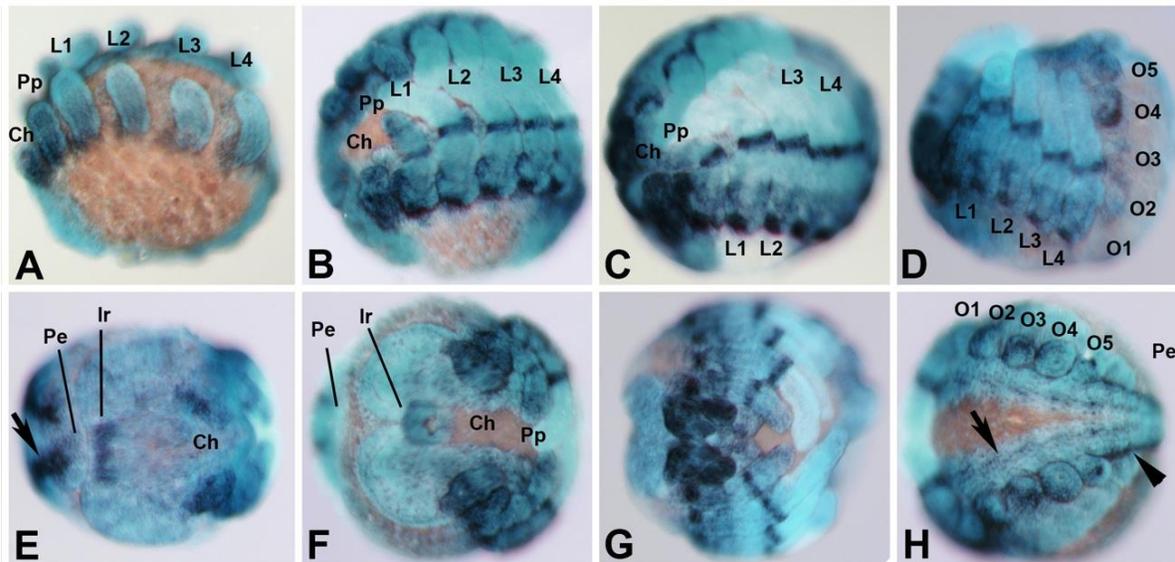


Figure 28: Expression of *exd1* in *P. tepidariorum*

A: Lateral view of a *P. tepidariorum* embryo during early limb bud development expressing *exd1* in the ventral neuroectoderm and the prosomal appendages. Ring-like *Pt-exd1* expression in the medial part of the pedipalps and walking legs during inversion (**B**) and dorsal closure (**C**), lateral view. **D:** Lateral view of *Pt-exd1* expression in the prosomal appendages and the opisthosoma during ventral closure. **E:** Frontal view of *Pt-exd1* expression in the labrum during early limb bud development. **F:** This *Pt-exd1* expression in the labrum stays but becomes weaker during inversion, frontal view. **G:** Strong *exd1* expression in the chelicera and labrum after head rearrangements in *P. tepidariorum* during ventral closure, frontal view. **H:** Posterior view of opisthosomal *Pt-exd1* expression during inversion. The arrow points to expression in the ventral neurogenic ectoderm, the arrowhead to the expression in the dorsal neurogenic ectoderm. Expression is present in all four opisthosomal limb buds, but in a distinct pattern. All embryos are oriented with the anterior side to the left, except panel H. lr: labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

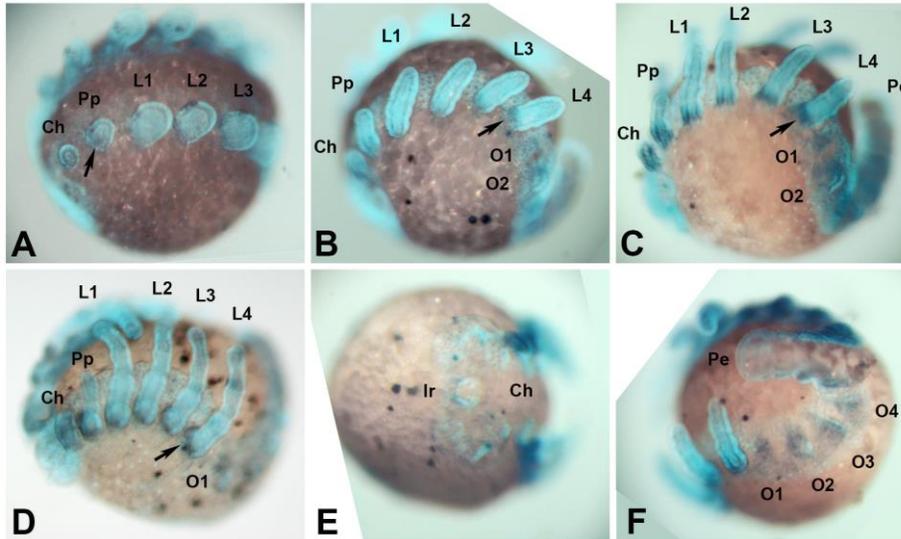


Figure 29: Expression *exd1* in *P. phalangioides*

A-D: Lateral view of embryos expressing *Pp-exd1* on the proximal base of the prosomal appendages (arrow) during early limb bud development (A), limb bud elongation (B), inversion (C) and dorsal closure (D). The embryos in A-D are oriented with the anterior side to the left. E: Frontal view of a *P. phalangioides* embryo during inversion with weak *exd1* expression in the labrum. F: Posterior view of opisthosomal *Pt-exd1* expression in the dorsal part. Ir: Labrum Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, Pe: posterior end.

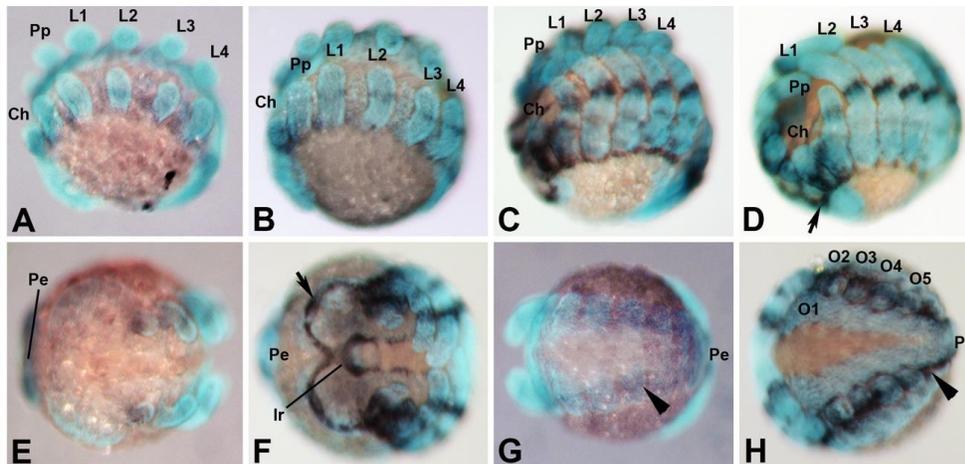


Figure 30: Expression of *exd2* in *P. tepidariorum*

A: Lateral view of a *P. tepidariorum* embryo during early limb bud development expressing *exd2* in the ventral neuroectoderm. Ring-like *Pt-exd2* expression in the medial part of the pedipalps and walking legs during limb bud elongation (B), inversion (C) and dorsal closure (D), lateral view. The arrow in D points to *Pt-exd2* expression on the former anterior rim of the non-neurogenic ectoderm overgrowing the neurogenic ectoderm. E: Frontal view of the head during early limb bud development, no visible *Pt-exd2* expression. F: Onset of *Pt-exd2* expression during inversion surrounding the head lobes anteriorly (arrow) as well as in the labrum, frontal view. G: Posterior view of broad opisthosomal *Pt-exd2* expression (arrowhead) during early limb bud development. H: Posterior view of medial *Pt-exd2* stripe (arrowhead) in the opisthosoma during inversion. All embryos are oriented with the anterior side to the left, except panels G and H. Abbreviations same as Figure 28.

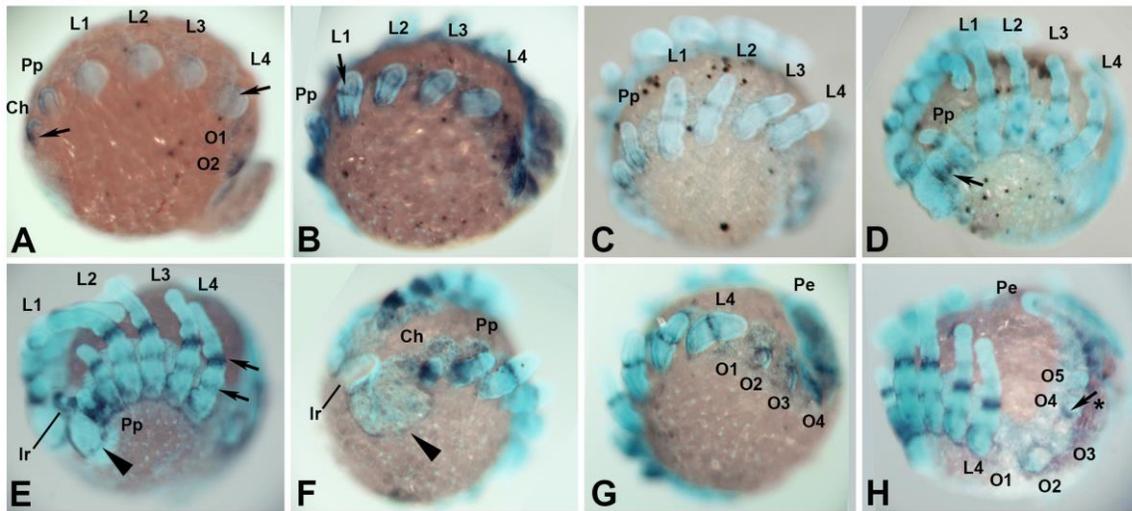


Figure 31: Expression *exd2* in *P. phalangioides*

Lateral view of *P. phalangioides* embryos with ring-like *exd2* expression in the medial part of the prosomal appendages (arrow) during early limb bud development (A), limb bud elongation (B), early inversion (C) and early and late dorsal closure (D and E), lateral view. The arrowhead in E points to *Pp-exd2* expression on the former anterior rim of the non-neurogenic ectoderm overgrowing the neurogenic ectoderm. F: Frontal view of the *Pp-exd2* head expression during inversion surrounding the head lobes anteriorly (arrow) as well as in the labrum. G: Lateral posterior view of broad opisthosomal *Pp-exd2* expression during limb bud elongation. H: Posterior view opisthosomal *Pp-exd2* expression in the dorsal part (arrowhead) and ventrally on the limb bud on O4 during inversion. All embryos are oriented with the anterior side to the left, except panels G and H. Abbreviations same as Figure 28.

5.1.8 Expression of *exd* in prosomal appendages

In the walking legs of *P. tepidariorum* and *P. phalangioides* *exd1* shows the conserved expression exclusively covering the proximal leg part with beginning of early limb bud development (bar, Figure 33, G and arrow, Figure 29, A). This proximal expression in the walking legs remains present throughout embryonic development (arrowhead, Figure 32, E and F and bar, Figure 33, H and I). With the beginning of inversion and prosomal limb bud differentiation a strong ring of *exd1* expression appears in the former patella region of the walking legs and pedipalps (arrow, Figure 32, D, E and F and arrow, Figure 33, E, F, H and I). The pedipalps of *P. phalangioides* strongly express *exd1* also in the gnathendite (arrowhead, Figure 32, C and D). This is also true for *Pt-exd1*, which is in addition moderately expressed in the distal and medial part of the pedipalps (asterisk, Figure 33, D, E, and F). From beginning of prosomal limb bud development *exd1* is

strongly expressed within the entire chelicera and only slightly vanishes during ventral closure (arrow, Figure 32, A and B, Figure 33, A, B and C). The expression of *exd2* differs from its paralog *exd1* for all three prosomal appendage types. In the pedipalps and walking legs, *exd2* expression directly starts as a ring in the former patella region at early limb bud developmental stages, but no proximal expression domain is visible (arrow, Figure 31, A and arrow, Figure 33, P). During late inversion a *Pt-exd2* and *Pp-exd2* expression domain appears in the coxa and as a ring in the developing femur in the walking legs and the pedipalps, as well as in the gnathendite (arrowhead and arrow, Figure 32, I and K and arrow, Figure 33, Q and N). In *P. tepidariorum* this expression in the pedipalps and walking legs remains unchanged until ventral closure is completed (Figure 33, R). In *P. phalangioides* by contrast all three *exd2* leg expression domains broaden during embryonic leg growth at ventral closure stages (bar, Figure 33, L). At the same time *Pp-exd2* in the pedipalps vanishes from the coxa and an additional expression appears distal to the patella ring. In the chelicera *exd2* is expressed in the proximal half similarly for both spiders until ventral closure. Whereas this expression remains unaltered throughout embryonic development of *P. phalangioides*, a distal expression domain of *exd2* appears during ventral closure in *P. tepidariorum*.

Taken together as already observed in the comparison of the expression in whole mount embryos, there is a general trend that the patterns are highly conserved for the orthologous duplicates between the two spider species studied here, but differs substantially between the paralogs within each species.

As stated above the expression pattern of a gene not necessarily also reflects the region of the functional protein. Also for *exd* no working antibodies are available for spiders, and thus we directly tried to knock down the gene function using parental RNAi in *P. tepidariorum*. Similar to *hth* different dsRNA injection procedures were tried each time in three females, but also for the *exd* paralogs no obvious phenotype was observed, so that the *exd* RNAi requires improvement as well.

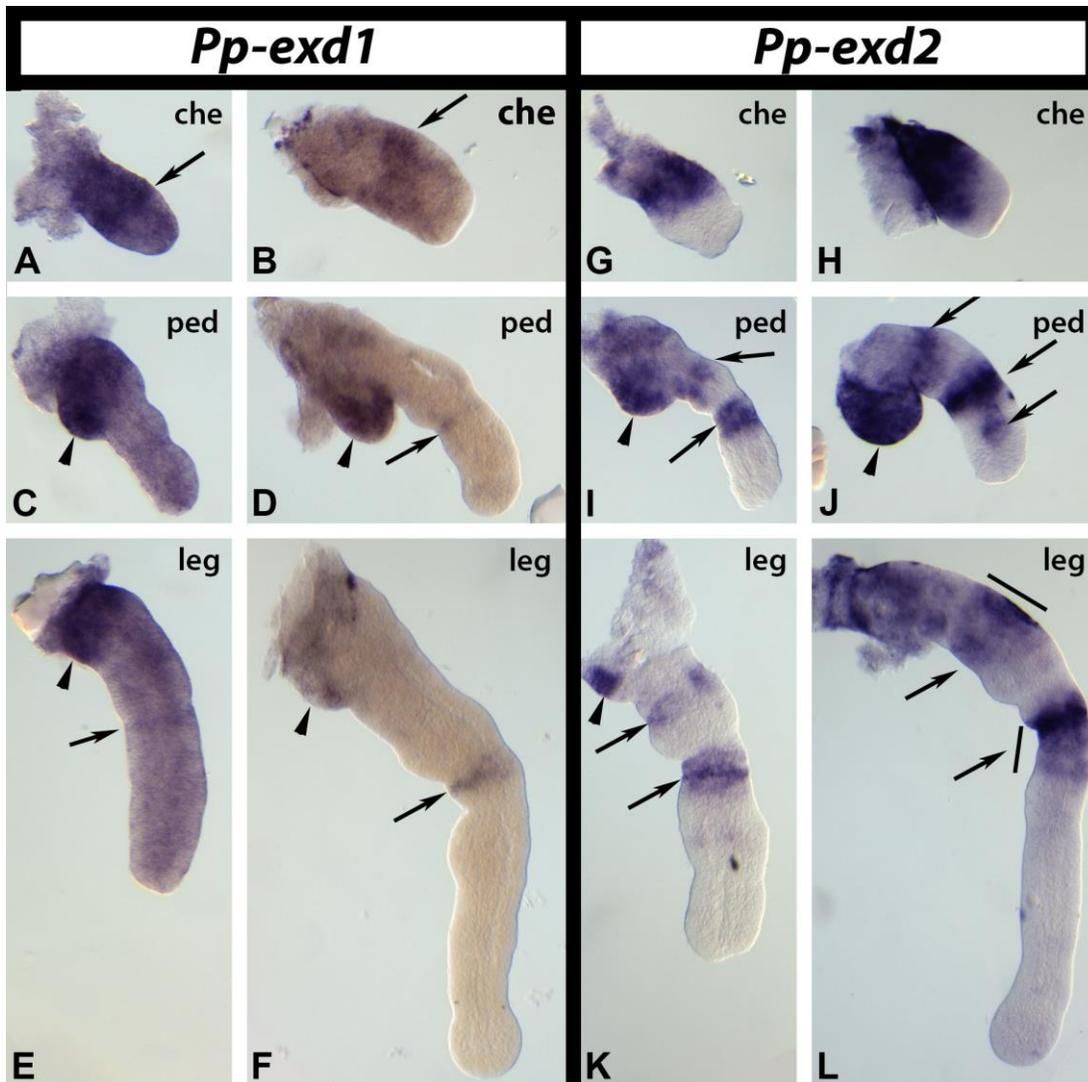


Figure 32: *exd* expression in prosomal appendages of *P. phalangioides*

A-F: Expression of *exd1* in prosomal appendages of *P. phalangioides*. A and B: Preparations of chelicera during limb bud elongation (A) and ventral closure (B). The arrow points to the proximal *exd1* expression, which later is found more medially. C and D: Expression of *exd1* in the pedipalps during limb bud elongation (C) and ventral closure (D). The arrowhead points to the expression in the gnathendite. The arrow marks *exd1* ring in the patella. E and F: Walking legs during limb bud elongation (E) and ventral closure (F). The arrowhead marks the proximal *exd1* domain and the arrow the ring of *exd1* in the patella. G-L: Expression of *exd2* in prosomal appendages of *P. phalangioides*. G and H: Preparations of chelicera during late inversion (G) and ventral closure (H). I and J: Expression of *exd1* in the pedipalps during late inversion (I) and ventral closure (J). The arrowhead points to the expression in the gnathendite. The arrows mark ring-like expression of *exd1*. K and L: Walking legs during late inversion (K) and ventral closure (L). The arrowhead marks the proximal *exd1* spot and the arrow the rings of *exd1* in the walking legs, which broaden during ventral closure (bar). che: chelicera. ped: pedipalp. leg: walking leg.

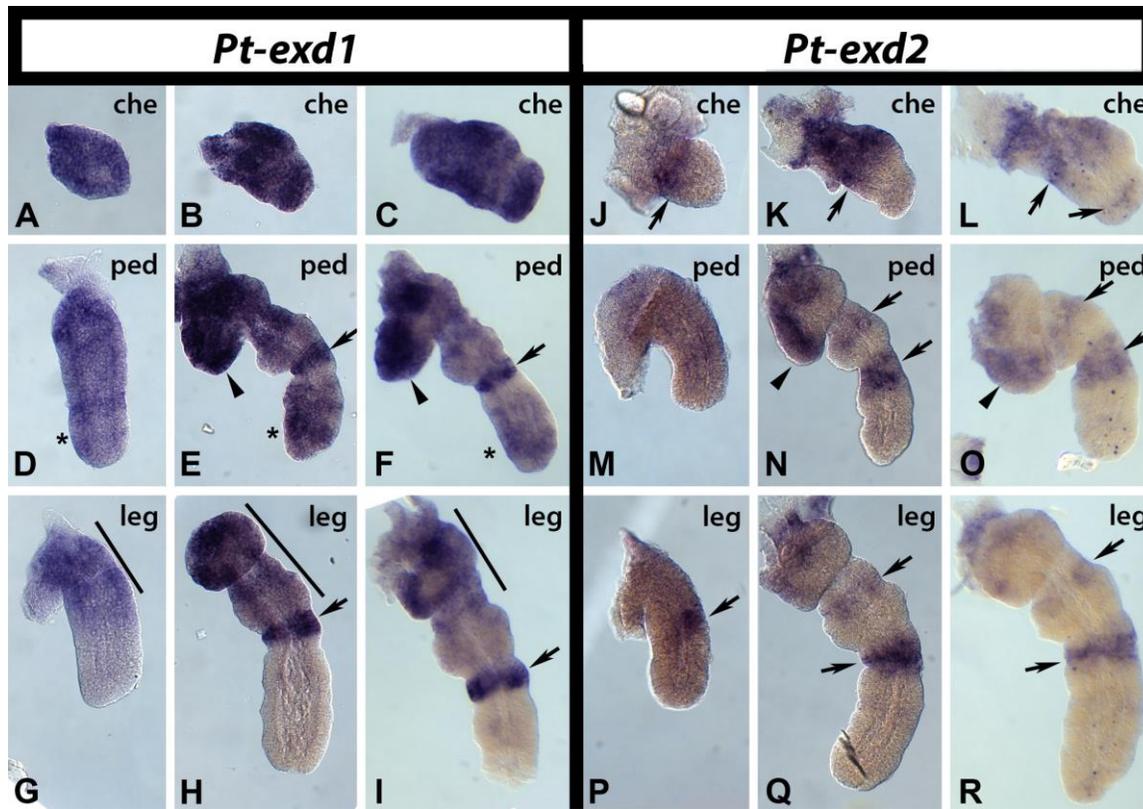


Figure 33: *exd* expression in prosomal appendages of *P. tepidariorum*

A-I: Expression of *exd1* in prosomal appendages of *P. tepidariorum*. J-R: Expression of *exd2* in prosomal appendages of *P. tepidariorum*. A-C and J-L: Expression of *exd* paralogs in the chelicera. The arrows indicate medial and distal expression of *exd2*. D-F and M-O: Expression of *exd* paralogs in the pedipalps. The arrowhead points to expression in the gnathendite. The arrow points to rings of expression. The asterisk marks broad distal *exd1* expression. G-I and P-R: Expression of *exd* paralogs in the walking legs. The bar marks the broad, conserved expression of *exd1* during limb bud development. The arrows indicate expressed ring of the *exd* paralogs. A, D, G, J, M and P show prosomal appendages during early limb bud development. B, E, H, K, N and Q show prosomal appendages during inversion. C, F, I, L, O and R: depict prosomal appendages during ventral closure. che: chelicera. ped: pedipalp. leg: walking leg.

For the following paragraph and the corresponding discussion section the text, figures and figure legends are taken from (Turetzek et al., 2016). (With kind permission Oxford Journals. Media: Molecular Biology and Evolution, Neofunctionalisation of a duplicate dachshund gene underlies the evolution of a novel leg segment in arachnids (2016), Volume 33, Issue 1. Pages: 109-121, authors N.Turetzek, M. Pechmann, C. Schomburg, J. Schneider and N. M. Prpic, copyright © by Oxford Journals and the authors.

5.2 Neofunctionalisation of a duplicate dachshund gene underlies the evolution of a novel leg segment in arachnids

5.2.1 Introduction

Gene duplication is regarded as the main process how new genes emerge during evolution. At first, the duplicate of the parental gene is superfluous and therefore most duplicates are lost again (Ohno 1970). There are, however, three principal possibilities that may lead to the retention of the duplicated gene (reviewed in Conant and Wolfe 2008; Innan and Kondrashov 2010; Magadum et al. 2013): first, the doubled amount of gene product may have beneficial effects, e.g. because the higher dosage leads to functional stability. In this case, the parental gene and the duplicate retain the same original function. Second, the duplicate may take over a subfunction of the parental gene (subfunctionalisation), thus leading to the evolution of tissue-specific genes and a reduction of pleiotropy. Third, the duplicate may gain a new function (neofunctionalisation) that makes the new gene essential for the organism.

Previous studies of the genetic basis of embryonic development in spiders have shown that many developmental genes are duplicated in spiders, for example the t-box genes *optomotor-blind* (*omb*) and *H15* (Janssen et al. 2008), or the TALE homeobox genes *homothorax* (*hth*) and *extradenticle* (*exd*) (Prpic et al. 2003). Indeed, the recent study of the complete developmental transcriptome of the spider *Parasteatoda tepidariorum* indicates that the genome contains over 40000 genes (Posnien et al. 2014), and the recently published genome sequences of the velvet spider *Stegodyphus mimosarum* (Sanggaard et al. 2014) and the scorpion *Mesobuthus martensii* (Cao et al. 2013) contain

around 30000 predicted genes. Compared to e.g. only 13000 genes in the genome of the fly *Drosophila melanogaster* (Adams et al. 2000), the high gene numbers in the genomes of arachnids suggest the presence of a large number of duplicated genes.

It has been suggested earlier that the high number of duplicated genes might be correlated with the evolution of many specific features and evolutionary novelties of arachnids (Cao et al. 2013; Sanggaard et al. 2014). Here, we provide evidence that supports this notion using the duplication of the *dachshund* (*dac*) gene. The *dac* gene has been identified first in *Drosophila* as a factor important for leg development (Mardon et al. 1994). The fly leg is composed of five principal segments, named coxa, trochanter, femur, tibia and tarsus. The *dac* gene is expressed in the developing fly leg in a broad domain in the medial portion of the leg comprising the femur and tibia. In *dac* mutant flies these leg segments are missing, thus resulting in shortened legs (Mardon et al. 1994). This patterning function of *dac* is widely conserved in the arthropods. A broad medial expression domain in the legs has been confirmed in representatives of all major arthropod groups, e.g. the myriapod *Glomeris marginata* (Prpic and Tautz 2003), the crustaceans *Porcellio scaber* (Abzhanov and Kaufman 2000) and *Triops longicaudatus* (Sewell et al. 2008), diverse insect species (e.g. Angelini and Kaufman 2005), and a number of arachnid species (e.g. Abzhanov and Kaufman 2000; Prpic et al. 2003; Sharma et al. 2013). These data indicate that *dac* is an important conserved factor in the evolution and diversification of the medial segment types in the arthropod leg.

We show here, however, that spiders also have a second *dac* gene, that has originated in the arachnid lineage, and that this duplication is linked to the evolution of a novel medial leg segment in arachnids that is not present in other arthropod groups. In this study, we use the spider species *Parasteatoda tepidariorum* and *Pholcus phalangioides*. These two spider species belong to two separate clades in the spider phylogeny: *P. tepidariorum* is a member of the Entelegynae, which comprise most of the extant spider species, whereas *P. phalangioides* belongs to the Haplogynae, which contain approximately 3500 species worldwide (see simplified cladogram in Figure 3). Thus, conserved aspects in these two species must derive from the spider lineage before the split between entelegyne and haplogyne spiders and therefore are valid for most of the extant species of spiders. We show that in both species the *dac* gene is present in two paralogous genes that trace from

a duplication event before the diversification of the extant arachnid groups. The *dac1* paralog shows the conserved patterning function that is known from other arthropods as described above. The second *dac* paralog, *dac2*, however, has undergone neofunctionalisation and is expressed in the patella, which is a leg segment that is only present in arachnids. We demonstrate that when *dac2* function is removed via RNA interference (RNAi), then the patella is joined with the tibia and does not form as a separate leg segment anymore. Thus the loss of *dac2* function removes the genetic cues to form a separate patella segment and reverts the spider leg to a state of segment composition before the evolution of the patella.

5.2.2 Duplicated *dachshund* genes in spiders

Previous studies in spiders have isolated the homolog of the *Drosophila dachshund* gene by using degenerate primers to amplify fragments from cDNA preparations (Abzhanov and Kaufman 2000; Prpic et al. 2003; Pechmann and Prpic 2009). These studies have consistently isolated a single *dachshund* homolog. However, our recent study of the full developmental transcriptome of *P. tepidariorum* (Posnien et al. 2014) has revealed the presence of a second paralogous *dac* gene, termed *Pt-dac2*, that has escaped previous gene cloning attempts. Based on the sequences from *P. tepidariorum*, we have also isolated two *dac* orthologs from another spider species, the cellar spider *Pholcus phalangioides* (termed *Pp-dac1* and *Pp-dac2*). An alignment of the conceptually translated amino acid sequences of the two Dachshund proteins from both species shows that *dac1* and *dac2* from both spider species encode very similar proteins. All proteins contain the conserved domains that were identified previously (Figure 34). The DD1 and DD2 domains were defined on the basis of sequence similarity between murine and *Drosophila* Dachshund proteins (Davis et al. 1999) and contain more restrictively defined domains based on similarities between human and murine homologs (Dachbox-N, Dachbox-C; Hammond et al. 1998), or between Dachshund proteins from diverse metazoan animals and the closely related Ski/Sno proteins (DS domain; Kozmik et al. 1999). Interestingly, the DD1 domain of *Dac2* in both spider species contains a novel insertion of about 30 amino acid residues (Figure 34), which is not known from other

5.2.3 Phylogenetic analysis of *dachshund* gene evolution and duplication

All previously isolated *dac* genes from spiders appear to be orthologs of *dac1*, whereas *dac2* orthologs were missed in gene cloning studies. Duplicated *dac* genes are also known from vertebrates (named *dach1* and *dach2*; e.g. Davis et al. 2001; Davis et al. 2006; Davis et al. 2008), and the available genome sequence of the scorpion *Mesobuthus martensii* (Cao et al. 2013) contains three (albeit short and therefore likely incomplete) predicted *dac* homologs. In order to establish the orthology between the spider *dac* genes and their relationships with other *dac* genes, we have performed a phylogenetic analysis of the duplicated *dac* genes in spiders, and *dac* genes from other protostomes and deuterostomes (Figure 35). The *dac1* genes from *P. tepidariorum* and *P. phalangioides* cluster together with the previously known spider *dac* genes, demonstrating that these are indeed genuine *dac1* genes (dark grey shading in Figure 35), whereas the *dac2* genes form a separate clade (light grey shading in Figure 35). The sequences MMa07100 and MMa52425 from the *M. martensii* genome cluster together with the spider *dac1* orthologs, but the MMa07092 sequence is an ortholog of spider *dac2*. The spider *dac* genes do not form a cluster with the vertebrate *dach* genes. Instead, the spider *dac* genes form a monophyletic group with the *dac* genes from other arthropods and are the sister group of the clade comprising *dac* genes from myriapod, crustaceans and insects. By contrast, the vertebrate *dac* genes form a separate clade in the tree and cluster together with the *dac* genes from other deuterostomes. This shows that the duplication events that produced the paralogous genes in spiders and vertebrates are entirely unrelated.

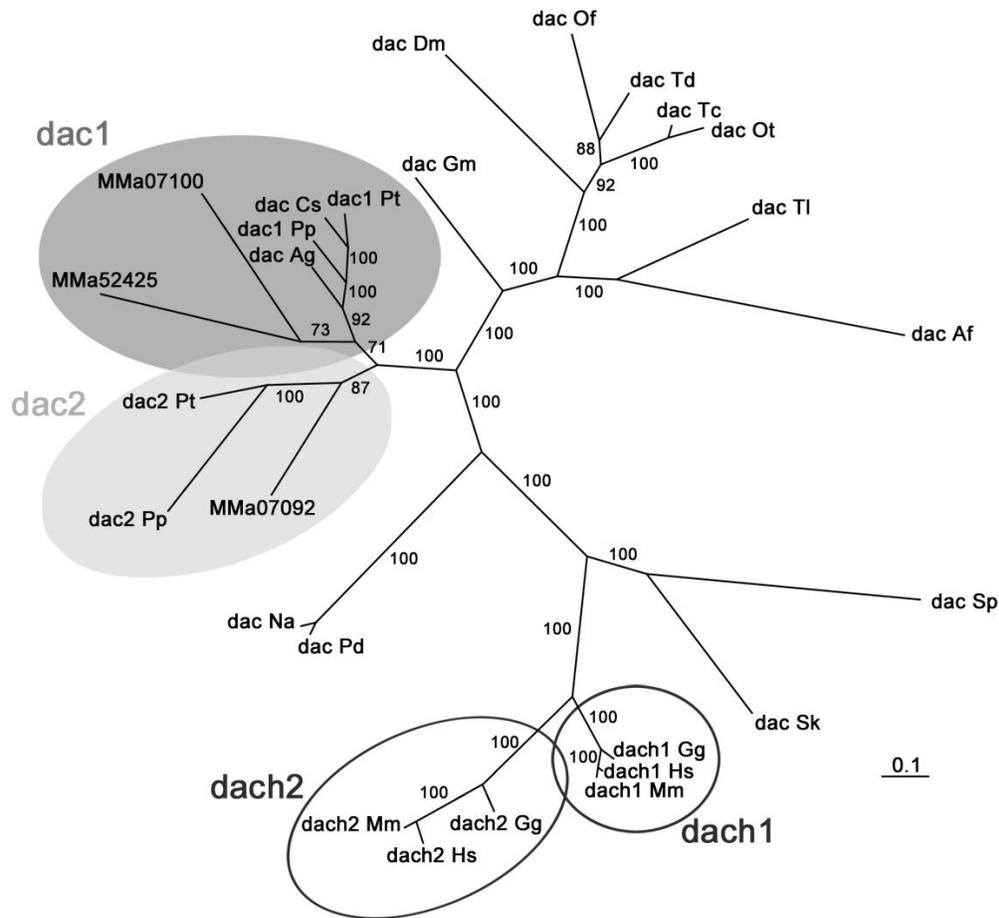


Figure 35: Unrooted 50% majority rule consensus tree after Bayesian Markov chain Monte Carlo analysis.

Branch lengths in the phylogram give the expected substitutions per site. Numbers at the tree edges are clade credibility values giving the probability of each clade in the tree. The duplicated *dac* genes in arachnids are indicated in dark grey (*dac1*) and light grey (*dac2*). The duplicated *dach* genes in vertebrates (*dach1* and *dach2*) are denoted by open circles. For species abbreviations and sequence accession numbers please see Supplemental Material 11.4.3 dachshund.

5.2.4 Expression of *dac1* in *Parasteatoda tepidariorum* and *Pholcus phalangioides*

The phylogenetic analysis shows that the *dac1* genes of *P. tepidariorum* and *P. phalangioides* are orthologs of the previously isolated *dac* genes from other spider species that show an evolutionarily conserved expression pattern which is also known from other arthropods (summarized in Angelini and Kaufman 2005). We therefore studied the *dac1* expression pattern in *P. tepidariorum* and *P. phalangioides* to confirm that they show the evolutionarily conserved expression pattern, too. The most prominent expression domain of *dac1* in *P. tepidariorum* is in the medial part of the

pedipalps and walking legs. This expression is initiated as soon as the limb buds have formed in the prosoma and is retained throughout embryonic development (Figure 36, A-D). At the beginning of germ band extension *Pt-dac1* expression is also initiated in the central nervous system (Figure 36, E) (see also Schomburg et al. 2015). The pattern in the head lobes comprises two domains in the center of the brain primordium and a domain at the rim of the head lobe (Figure 36, E). Expression in the ventral neurogenic ectoderm is initially restricted to the pedipalpal segment, the four walking leg segments and the first three opisthosomal segments (Figure 36, E, F). During further development, however, the neurogenic ectoderm of the remaining opisthosomal segments expresses *Pt-dac1* as well (Figure 36, G). Additional expression domains of *Pt-dac1* appear later in development in dorsal cells that will contribute to the heart, and in cells of the hindgut primordium (Figure 36, H).

The expression of *dac1* in *P. phalangioides* is very similar to the *dac1* ortholog of *P. tepidariorum*. There is prominent expression in the developing pedipalps, legs and in the central nervous system throughout development (Figure 37, A-G). The expression pattern in the head lobes is virtually identical to the expression in *P. tepidariorum*, except that the expression domain at the lateral rim of the head lobes is very strong (Figure 37, C). During germ band inversion there is a dorsal expression in the presumptive heart tissue (Figure 37, H), similar to the expression in *P. tepidariorum*.

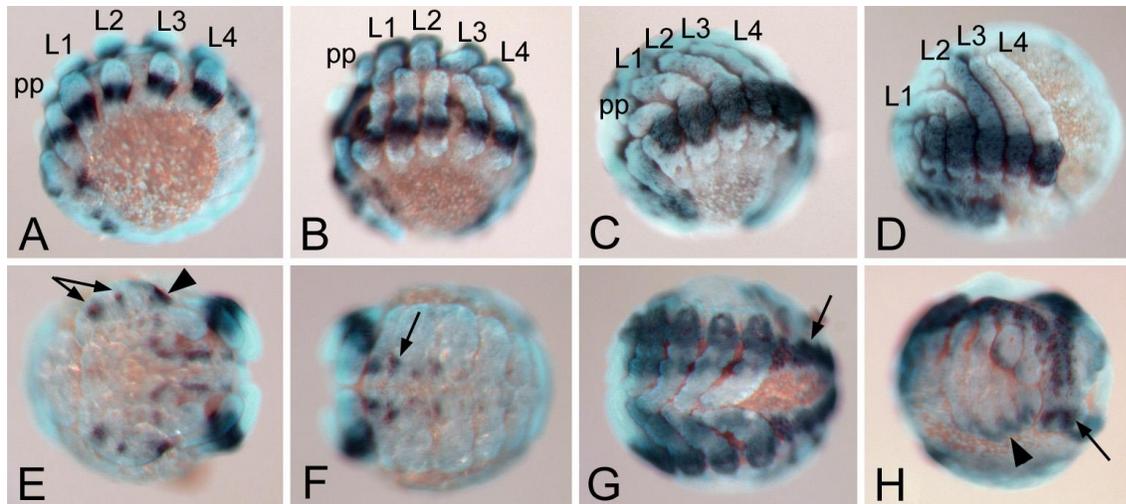


Figure 36: Expression of *dac1* in *P. tepidariorum*

(A) Embryo during germ band elongation. Lateral view. (B) Fully elongated embryo. Lateral view. (C) Embryo at the beginning of germ band inversion. Lateral view. (D) Embryo at the end of germ band inversion and after dorsal closure. Lateral view. (E) Embryo during germ band elongation. Frontal view of the head. The arrows point to domains in the center of the brain primordium. The arrowhead denotes expression at the rim of the head lobe. (F) Embryo during germ band elongation. Ventral view of the opisthosoma. The arrow points to expression in the ventral neurogenic ectoderm. (G) Embryo at the beginning of germ band inversion. Ventral view of the prosoma and anterior portion of the opisthosoma. The arrow points to expression in the ventral neurogenic ectoderm. (H) Embryo at mid-germ band inversion, before dorsal closure. Posterior view of the opisthosoma and the hindgut primordium. The arrow points to expression in the hindgut primordium. The arrowhead denotes expression in presumptive dorsal tissue that will contribute to the heart. All embryos are shown with anterior to the left. Abbreviations: pp, pedipalps; L1-L4, walking legs 1 to 4.

5.2.5 Expression of *dac2* in *Parasteatoda tepidariorum* and *Pholcus phalangioides*

In order to establish whether the *dac2* paralog in spiders also shows the evolutionarily conserved expression pattern, or has changed its expression during evolution, we have studied the expression of *dac2* in *P. tepidariorum* and *P. phalangioides*. *Pt-dac2* is expressed in the pedipalps and walking legs during germ band extension and during further development (Figure 38, A-D). This appendicular expression pattern, however, differs significantly from the expression of *Pt-dac1* (see details below). Expression of *Pt-dac2* in the ventral neuroectoderm is entirely restricted to the pedipalpal segment and the anterior half of the first walking leg segment, and there is no early expression in the head lobes (Figure 38, E-F). After dorsal closure, however, expression of *Pt-dac2* is activated in the primordia of the lateral eyes (Figure 38, D; see also Schomburg et al. 2015). During early inversion of the embryo there is expression in presumptive dorsal

ectoderm near the bases of the opisthosomal limb buds (Figure 38, G), and after dorsal closure there is very faint expression in dorsal segmental stripes that possibly correspond to the embryonic septa that divide the yolk mass within the developing opisthosoma (Figure 38, H).

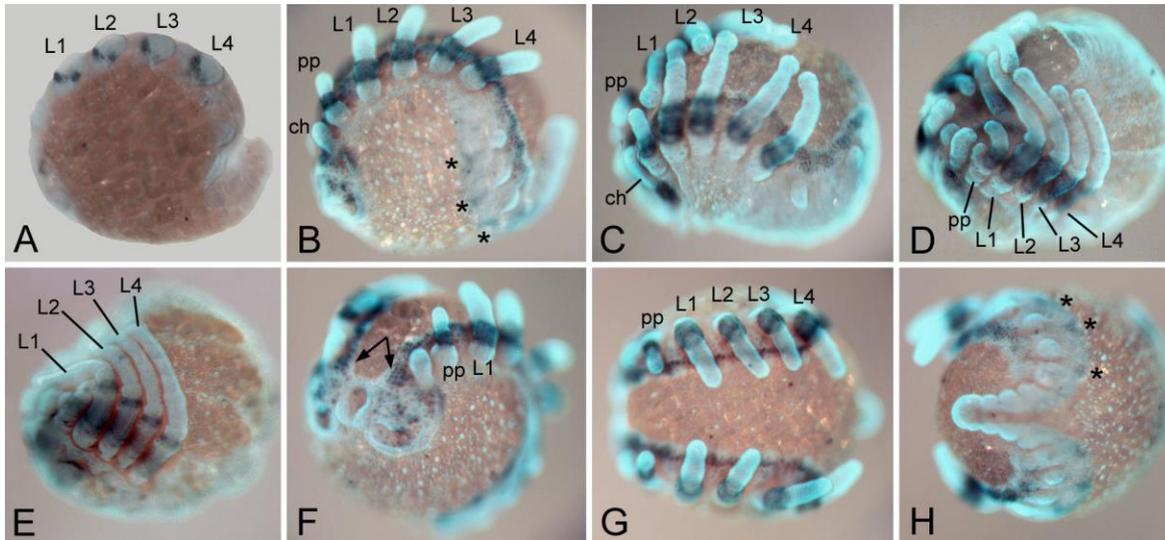


Figure 37: Expression of *dac1* in *P. phalangioides*

(A) Embryo during germ band elongation. Lateral view. (B) Fully elongated embryo at mid-germ band inversion. Lateral view. The asterisks denote expression in presumptive dorsal tissue that will contribute to the heart. (C) Embryo approaching the end of germ band inversion. Lateral view. (D) Embryo at the end of germ band inversion and after dorsal closure. Ventro-lateral view. (E) Embryo after ventral closure. Lateral view. (F) Embryo at mid-germ band inversion. Latero-frontal view. The arrows point to expression in the ventral neuroectoderm. (G) Embryo at mid-germ band inversion. Ventral view of the prosoma. (H) Embryo at mid-germ band inversion, before dorsal closure. Posterior view of the opisthosoma and the hindgut primordium. All embryos are shown with anterior to the left. Abbreviations: ch, chelicera; pp, pedipalp; L1-L4, walking legs 1 to 4.

The expression of *Pp-dac2* is very similar to the expression of *dac2* in *P. tepidariorum* (Figure 39, A-E). The expression in the appendages is virtually identical (see details below), and the expression in the ventral neurogenic ectoderm is also restricted to the pedipalpal segment and the anterior part of the first walking leg segment (Figure 39, F, G). At stages after dorsal closure, *Pp-dac2* is expressed in the primordia of the lateral eyes (Figure 39, E). Similar to the homologous gene in *P. tepidariorum*, *Pp-dac2* is also expressed in faint stripes in the opisthosoma (Figure 39, H).

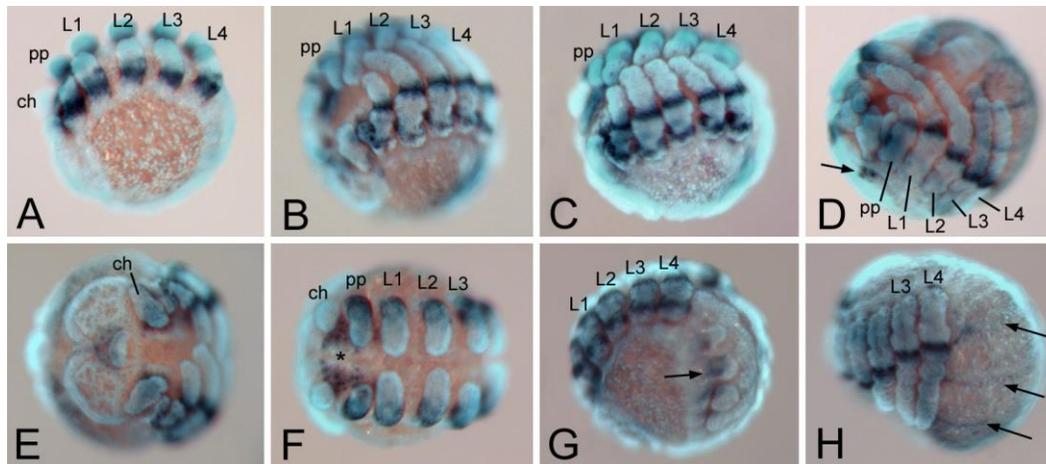


Figure 38: Expression of *dac2* in *P. tepidariorum*

(A) Embryo during germ band elongation. Lateral view. (B) Fully elongated embryo at the beginning of germ band inversion. Lateral view. (C) Embryo at mid-germ band inversion. Lateral view. (D) Embryo after dorsal closure and shortly before ventral closure. Ventro-lateral view. The arrow points to expression in the primordia of the lateral eyes. (E) Fully elongated embryo at the beginning of germ band inversion. Frontal view of the head. (F) Embryo during germ band elongation. Ventral view. The asterisk denotes expression in the ventral neuroectoderm of the pedipalpal segment. (G) Embryo at mid-germ band inversion. Ventro-lateral view of the opisthosoma. The arrow points to expression near the bases of the opisthosomal limb buds. (H) Embryo after dorsal closure and shortly before ventral closure. Lateral view. The arrows point to expression on the dorsal side, presumably in the intralecithal septa. All embryos are shown with anterior to the left. Abbreviations: ch, chelicera; pp, pedipalp; L1-L4, walking legs 1 to 4.

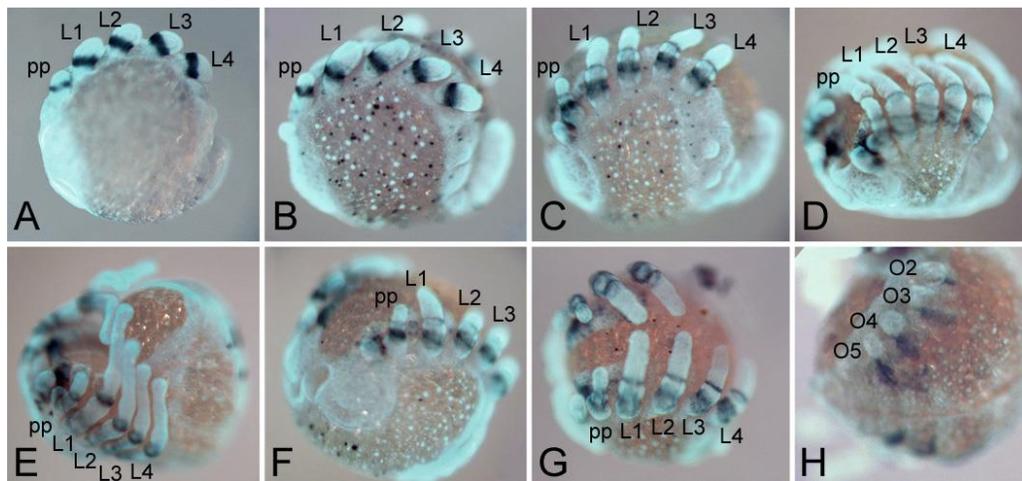


Figure 39: Expression of *dac2* in *P. phalangioides*

(A) Embryo during germ band elongation. Lateral view. (B) Fully elongated embryo. Lateral view. (C) Embryo at mid-germ band inversion. Lateral view. (D) Embryo after dorsal closure. Lateral view. (E) Embryo after dorsal closure and shortly before ventral closure. Ventro-lateral view. (F) Embryo at mid-germ band inversion. Frontal-lateral view. (G) Embryo at mid-germ band inversion. Ventro-lateral view of the opisthosoma. (H) Embryo after dorsal closure. Dorsal view of the opisthosoma. All embryos are shown with anterior to the left. Abbreviations: pp, pedipalp; L1-L4, walking legs 1 to 4; O2-O5, opisthosomal segments 2 to 5.

5.2.6 Expression profile of *dac1* and *dac2* in the prosomal appendages

The most significant differences in the expression profile of the two *dac* paralogs are found in the developing prosomal appendages and we have studied these in more detail. The expression of the homolog of the *dac1* gene in the prosomal appendages has been analysed previously in three other spider species, *Acanthoscurria geniculata* (Pechmann and Prpic 2009), *Cupiennius salei* (Prpic et al. 2003), and *Steatoda triangulosa* (Abzhanov and Kaufman 2000). Our results in *P. tepidariorum* and *P. phalangioides* show that the expression of this gene is highly conserved in spiders (Figure 40, A-F and Figure 41, A-F). In the appendages, the *dac1* gene of both species is expressed in the pedipalps and all four walking leg pairs in a broad ring-shaped domain that comprises the trochanter and the femur (Figure 40, C-F and Figure 41, A-F). The *dac1* gene is not expressed in the gnathal endite of the pedipalps (Figure 40, D and Figure 41, D). In the chelicera, there is only very faint expression within the base of the chelicera in *P. tepidariorum* (Figure 40, A). A similar expression domain has been reported from *C. salei* (Prpic and Damen 2004), and is possibly related to the development of the peripheral nervous system within the chelicera. This expression domain disappears during germ band inversion in *P. tepidariorum* (Figure 40, B) and is completely lacking in the chelicera of *P. phalangioides* (Figure 41, A and B). In both species *dac2* is expressed in all prosomal appendages, including the chelicera (Figure 40, G-L and Figure 41, G-L). In early stages, there is a single broad basal domain in the pedipalps and the legs comprising the tissue of the distal coxa and the trochanter (Figure 38, A and Figure 39, A and B). A second domain appears in both species in the tissue of the patella in early germ band inversion stages (Figure 40, I and K and Figure 41, I and K). The patella domain strongly increases its expression intensity during further development, whereas the expression in the distal coxa and trochanter becomes weaker (Figure 40, J and L and Figure 41, J and L). In the pedipalp, the vanishing proximal domain extends into the gnathendite and there retains a higher expression level (Figure 40, J and Figure 41, J). In the chelicera there is a proximal ectodermal domain along the ventral side in both species (Figure 40, G and H and Figure 41, G and H). This domain is weak in earlier stages (Figure 40, G and Figure 41, G), but becomes stronger in older stages (Figure 40, H and Figure 41, H).

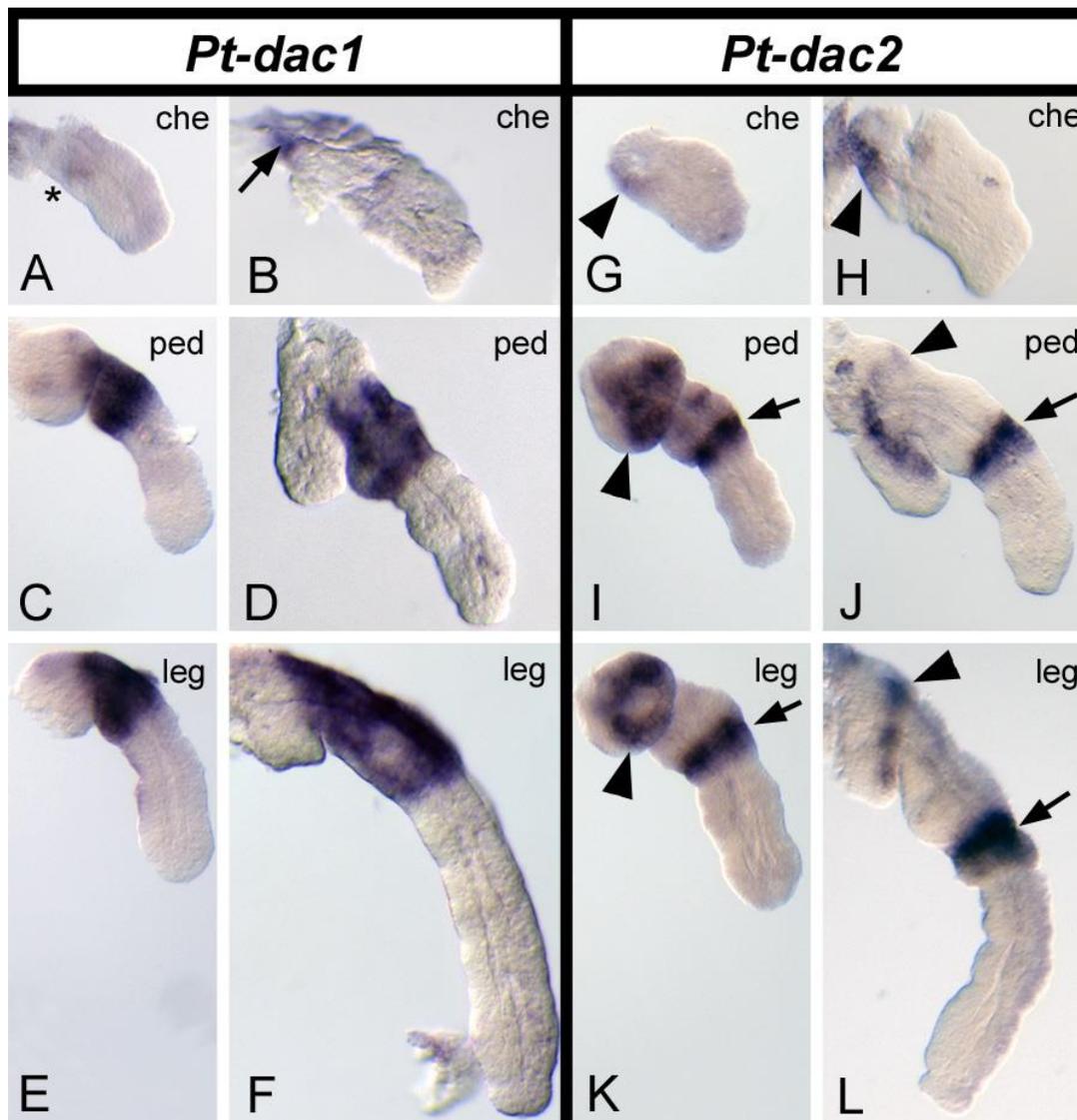


Figure 40: Detection of *dac1* (A-F) and *dac2* (G-L) expression in dissected appendages of *P. tepidariorum*.

(A, B) Chelicera of an embryo during germ band elongation (A) and after dorsal closure (B). The asterisk in A denotes expression of *dac1* in tissue at the base of the chelicera. The arrow in B points to expression in brain tissue that has not been completely removed during dissection. (C, D) Pedipalp of an embryo during germ band elongation (C) and after dorsal closure (D). (E, F) Leg of an embryo during germ band elongation (E) and after dorsal closure (F). (G, H) Chelicera of an embryo during germ band elongation (G) and after dorsal closure (H). The arrowheads denote expression of *dac2* in ectodermal tissue at the base of the chelicera. (I, J) Pedipalp of an embryo during germ band elongation (I) and after dorsal closure (J). The arrowheads denote a proximal domain in the coxa. The arrows point to a distal domain in the patella. (K, L) Leg of an embryo during germ band elongation (K) and after dorsal closure (L). The arrowheads denote a proximal domain in the coxa. The arrows point to a distal domain in the patella.

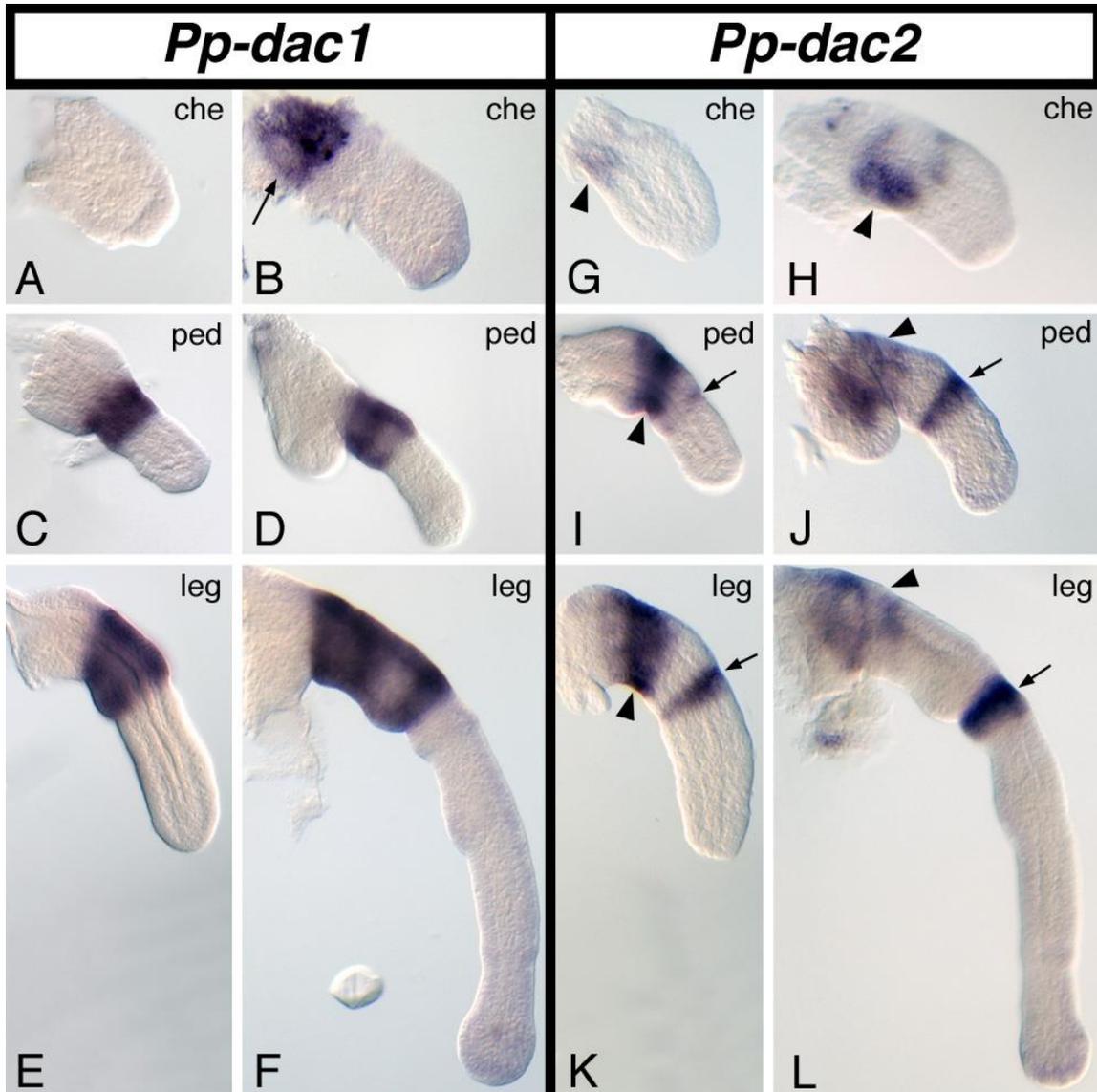


Figure 41: Detection of *dac1* (A-F) and *dac2* (G-L) expression in dissected appendages of *P. phalangioides*.

(A, B) Chelicera of a fully elongated embryo (A) and after dorsal closure (B). The arrow in B points to expression in brain tissue that has not been completely removed during dissection. (C, D) Pedipalp of a fully elongated embryo (C) and after dorsal closure (D). (E, F) Leg of a fully elongated embryo (E) and after dorsal closure (F). (G, H) Chelicera of fully elongated embryo (G) and after dorsal closure (H). The arrowheads denote expression of *dac2* in ectodermal tissue at the base of the chelicera. (I, J) Pedipalp of a fully elongated embryo (I) and after dorsal closure (J). The arrowheads denote a proximal domain in the coxa. The arrows point to a distal domain in the patella. (K, L) Leg of a fully elongated embryo (K) and after dorsal closure (L). The arrowheads denote a proximal domain in the coxa. The arrows point to a distal domain in the patella.

5.2.7 *dac2* is required for morphogenesis of the patella leg segment in *P. tepidariorum*

The strong and persistent expression of *dac2* in the patella of both spider species studied here is unique to this *dac* paralog and suggests that *dac2* might have a role in the development of this leg segment. To test this idea, we have used parental RNA interference (pRNAi) that is available in *P. tepidariorum* to downregulate *dac2* during embryonic development, and have studied the effects of *dac2* loss for the morphology of the legs in the first nymphal stage. At this stage, the cuticle is fully differentiated and the joints between the leg segments are complete and functional. In the wild type nymphs the patella is a short leg segment between the long leg segments femur and tibia (Figure 42, A-D). The patella has a characteristic shape that is reminiscent of a bell, with the narrow end towards the femur and the wide end towards the tibia (Figure 42, B-D). After *dac2* parental RNAi 20% of the nymphs show a strong malformation of the patella leg segment (Figure 42, I). In the *dac2* RNAi animals, the shape of the patella is no longer bell-shaped (Figure 42, E-H). In most cases, the proximal end of the patella is almost as wide as the distal end, much like e.g. in the tibia (Figure 42, G and H). In addition to the changes in shape, the patella in *dac2* RNAi animals is also shortened and the joint between patella and tibia is fused. At the fused junction between patella and tibia most legs show one or two short bulges (Figure 42, F-H), pointing to an excess of cells at the fused joint.

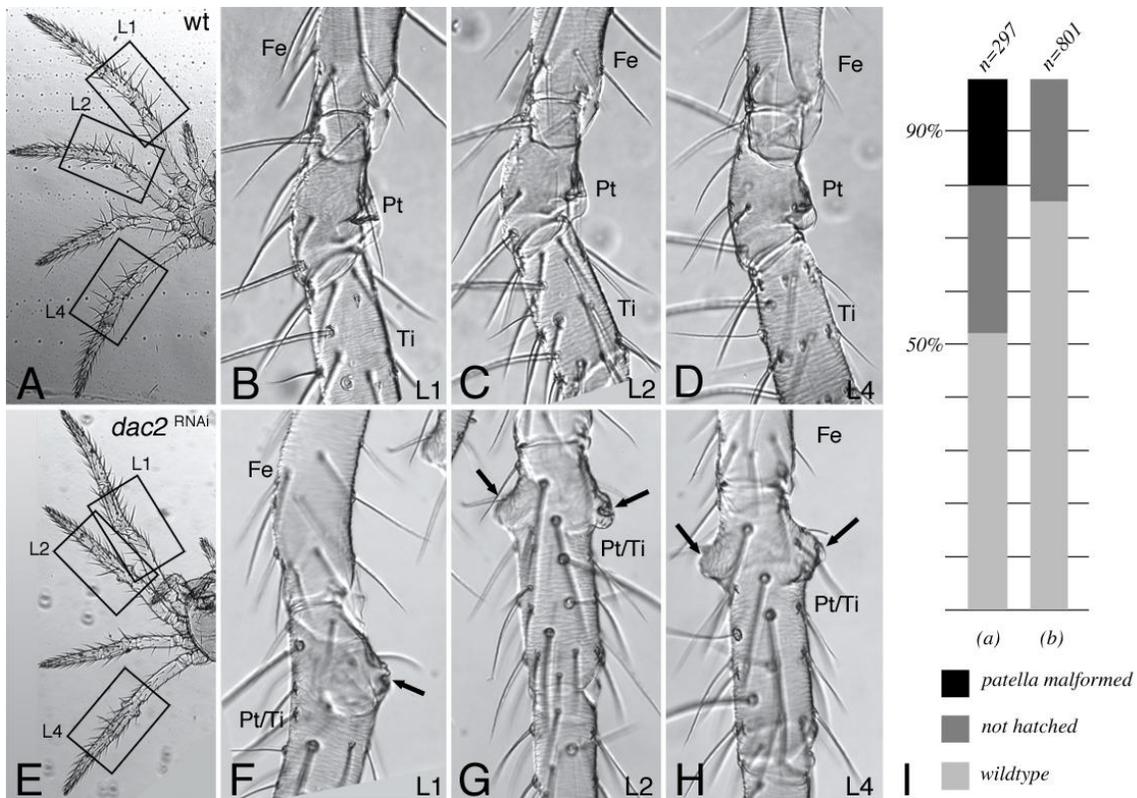


Figure 42: Malformation of the patella after *dac2* RNAi in *P. tepidariorum*. (A-D) Wild type nymph. (A) Overview of the right half in ventral view. Boxed are those leg regions that are magnified in B-D. (B-D) Magnifications of the area between femur, patella, and tibia of the L1 leg (B), L2 leg (C) and L4 leg (D). (E-H) Nymph with malformations of the patella after *dac2* RNAi. (E) Overview of the right half in ventral view. Boxed are those leg regions that are magnified in F-H. (F-H) Magnifications of the area between femur, and the fused patella/tibia of the L1 leg (F), L2 leg (G) and L4 leg (H). (I) Summary of the *dac2* RNAi experiment. The diagram in (a) shows the phenotype distribution after *dac2* dsRNA injection. The diagram in (b) shows the phenotype distribution after GFP dsRNA injection as a control. For both analyses the offspring that developed in the second cocoon of the injected females were pooled and scored for phenotypes. Please see the Materials and Methods section for details.

5.3 Discussion and future work

5.3.1 Gene duplication of spider dachshund genes

Our search for genes with sequence similarity to *dac* in the spiders *P. tepidariorum* and *P. phalangioides* revealed the presence of two paralogs in both species. The phylogenetic analysis clearly shows that the duplicated *dach* genes in vertebrates and the duplicated *dac* genes in spiders did not originate in the same gene duplication event (see Figure 35). The duplicated *dac* genes are present in spiders from two main spider lineages, the Entelegynae and the Haplogynae (see Figure 3 for a simplified overview of spider phylogeny), and this indicates that the duplication event is older than the split between haplogyne and entelegyne spiders. Indeed, the presence of three *dac*-related fragments in the genome sequence of a basal arachnid, the scorpion *M. martensii*, which are distributed in the phylogenetic analysis to the two *dac* clades, strongly suggests that the duplication of the parental *dac* gene into *dac1* and *dac2* occurred before the radiation of the arachnid orders: the original *dac* gene has been duplicated in the arachnid lineage even before the split between scorpions and spiders giving rise to *dac1* and *dac2*, and *dac1* has been duplicated again more recently in the scorpion lineage.

5.3.2 Neofunctionalisation of *dac2* in the arachnids

The *dac1* and *dac2* genes are thus the product of a phylogenetically old duplication event and are common to all extant arachnid orders. The expression of *dac1* in spiders is similar to *dac* genes in other arthropods, for example in the myriapod *Glomeris marginata* (Prpic and Tautz 2003) and the insects *Tribolium castaneum* (Prpic et al. 2001) and *Oncopeltus fasciatus* (Angelini and Kaufman 2004). Thus, *dac1* preserves the original patterning function of the parental *dac* gene. By contrast, *dac2* expression differs from the conserved expression pattern, and especially in the appendages shows novel expression domains that are not present in the *dac1* pattern. This strongly suggests that *dac2* has acquired a novel function and therefore escaped gene loss by neofunctionalisation. The novel expression domains indicate that neofunctionalisation involved evolutionary changes in the regulatory region of *dac2*. However, the insertion

within the DD1 domain of the *dac2* protein that is conserved in both species (see Figure 34) suggests that the properties of this functional domain are altered in *dac2*, and thus this indicates that evolutionary changes in the coding region of *dac2* have contributed to neofunctionalisation as well.

5.3.3 Evolution of the arachnid patella: neofunctionalisation of *dac2* underlies the evolution of a morphological novelty

The appendages of the arthropods are divided into segments that are separated by flexible joints. Usually, the appendages are composed of a combination of long and short segments to improve the mechanical properties of the appendages. Especially the short segments determine the degree of appendage flexibility and maneuverability, by providing pivotal points for the movement of the longer segments. In addition, short segments increase the degrees of freedom within a short portion of the appendage. All arthropod groups have a proximal short segment, the coxa, and in most groups this segment is followed by a second short segment, the trochanter (Figure 43). Arachnids, however, have a third short segment, the patella, that is intercalated between two long segments, the femur and the tibia (indicated in black in Figure 43). The patella is therefore a morphological novelty that has evolved in the arachnid lineage. Our results show that *dac2* is expressed in the developing patella of two spider species and is required for the proper formation of the patella. Moreover, if *dac2* function is removed via RNAi, then the patella is malformed and is no longer morphologically separated from the following segment, the tibia. Thus, removing *dac2* function experimentally restores the segment composition of the arthropod leg before the origin of *dac2* in the arachnid lineage. In this way, we are able to provide a link between the origin and neofunctionalisation of *dac2* and the evolution of a morphological novelty in the arachnid appendage. The alternative interpretation that the duplication of *dac2* already occurred in the ancestral arthropod, is less parsimonious than the interpretation that *dac2* has evolved in the arachnid lineage, because it requires two evolutionary events: the origin of *dac2* at the base of the arthropods, and the subsequent loss of *dac2* in the Mandibulata lineage. In addition, the early origin of *dac2* and thus the emergence of the

patella at the base of the arthropods is not supported by the leg morphology of the known Palaeozoic arthropods (overview in Ortega-Hernandez et al. 2013): neither marelomorphs (Whittington 1971; Garcia-Bellido and Collins 2006) or trilobites (Whittington 1975; Ramsköld and Edgecombe 1996), nor other basal arthropods like *Sidneyia inexpectans* (Stein 2013) or *Shankouia zhenghei* (Waloszek et al. 2005) show evidence for a patella segment in their legs.

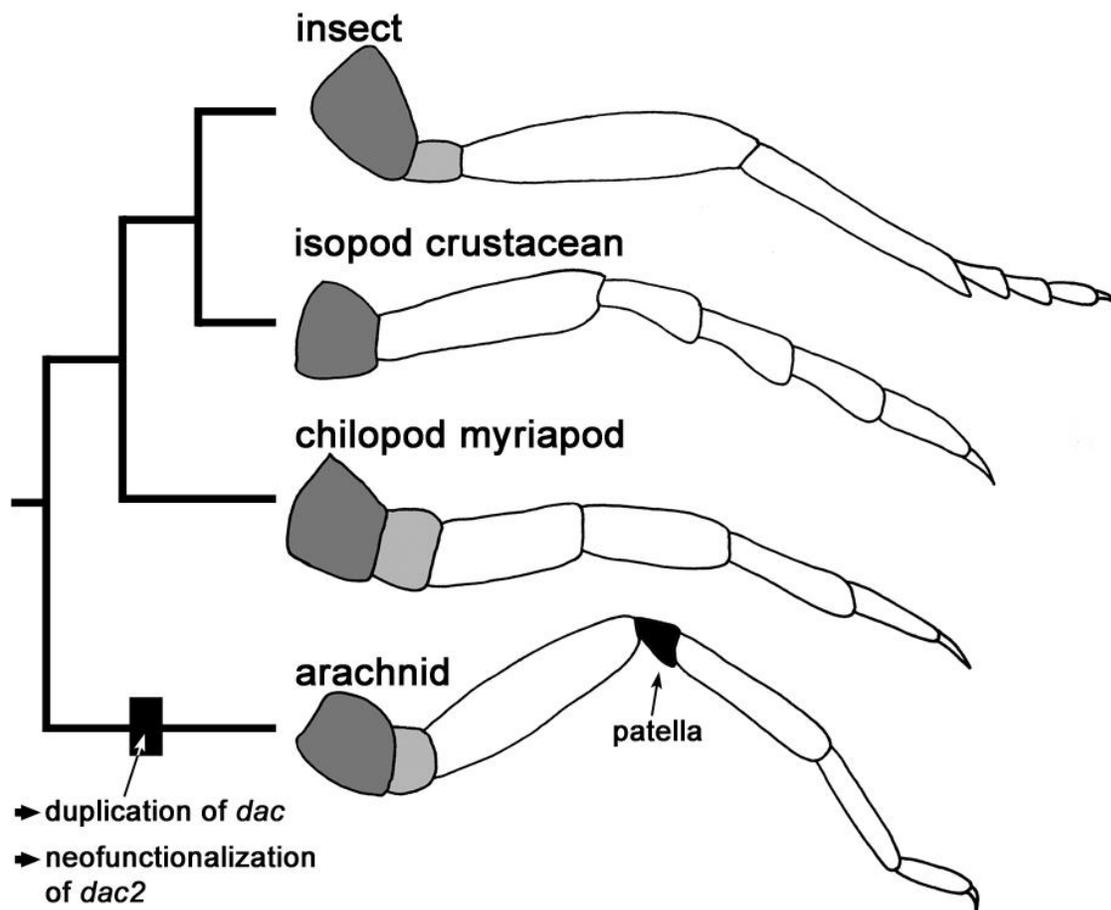


Figure 43: Evolution of a novel leg segment in arachnids linked to *dac* duplication and neofunctionalisation.

The drawings on the right show the leg segment composition in insects, isopod crustaceans, chilopod myriapods and arachnids (including spiders). The tree on the left depicts a simplified phylogenetic tree showing the interrelationships of these arthropod taxa (after Rota-Stabelli et al. 2011). All taxa possess a proximal short segment (coxa; shown in dark grey), and most taxa also have an additional short segment (trochanter; shown in light grey). But only arachnids have a third short leg segment (patella; shown in black) that is intercalated between two longer leg segments. Our data show that the proper formation of this unique leg segment requires the function of *dac2* and thus link the origin of the patella to the duplication and neofunctionalisation of the *dachshund* gene in arachnids (black rectangle in the tree; autapomorphy of the arachnid clade).

5.3.4 Conservation of *exd* after duplication

As already found for the bird spider *A. geniculata* and the entelegyne spider *C. salei*, and for the previously described leg gap gene *dac*, the study of *exd* in the two additional spider species here also revealed the duplication of this leg gap gene. For the two scorpions included in the phylogenetic study even five *exd* were found. In contrast to the duplication event of *dac*, the phylogenetic relationship of these duplicated *exd* paralogs resulted in a polytomy and thus the origin of the gene duplication events could not be properly resolved. Alignments of the protein sequences of the spider *exd* homologs show a high similarity of the protein also indicating a strong sequence conservation, and this not only true for the first paralog, but also for *exd2*. The expression in the walking legs was similar to previous findings in spiders (Pechmann and Prpic, 2009; Prpic et al., 2003). Expression of *exd1* is restricted to the proximal part during early limb bud developmental stages, reflecting its conserved role in proximal patterning function. The later expression as a ring in the patella, which was found in chelicerate species including *C. salei*, *A. geniculata* and the harvest man *Phalangium opilio* but not in the mite *Archezogozetes longisetosus*, was also present in both spiders studied here (Barnett and Thomas, 2013; Sharma et al., 2012). Thus this ring, which was found to be regulated by the Notch pathway, is a conserved trait of chelicerates (Prpic and Damen, 2009), lost or not detected in the mite. The more detailed expressional analysis of *exd1* in the pedipalps of *P. tepidariorum* showed an additional domain of expression in the distal pedipalp. Such domains were also found in the previously studied spiders, especially in *C. salei* (personal communication Dr. Prpic-Schäper) but were thought to be the cause of background staining. The presence of this overall pedipalpal staining in an additional spider hints to a differing role of *exd1* in the pedipalps and legs.

The second paralog of *exd*, *exd2*, differs in expression pattern and dynamics in the walking legs compared to its paralog. The expression starts as a ring in the patella and only later two additional rings arise more proximally. These differences in expression pattern and especially timing suggest a derived function of the duplicate different from the conserved leg gap gene function. To understand if it has also undergone neofunctionalisation similar to e.g. *dac*, functional studies are required. Although

different RNAi procedures were tried no visible knockdown phenotypes for *exd2* were observed so far, thus future experiments will include the optimization of *exd2* RNAi in *P. tepidariorum*. Similar to *dac* the diverged expression of *exd2* after the duplication event is conserved in all four spider species studied so far. This also hints to the evolution of a new function, important for walking leg development in chelicerate species, which was therefore conserved as usual after neofunctionalisation.

5.3.5 Spider specific *hth* duplication

hth is well studied in many arthropod species and is involved in several developmental processes e.g. in eye and limb development (e.g. (Pai et al., 1998; Prpic et al., 2003; Wu and Cohen, 2000)). Consistent with previous studies in the bird spider *A. geniculata* and *C. salei*, a member of the entelegyne group from the higher spiders (Pechmann and Prpic, 2009; Prpic et al., 2003), we also found two *hth* homologs in *P. tepidariorum* and in *P. phalangioides*. Although gene duplications in the chelicerates are also found in more basal groups like scorpions and horseshoe crabs (Di et al., 2015; Nossa et al., 2014) it is still not certain how they are related to the duplication found in spiders. Findings for *dac* paralogs revealed that this duplication is ancestral within the arachnids and that the new function has given rise to an innovative morphologic trait in these species. The study on the phylogeny of the *hth* duplication in contrast showed that although a duplication event also occurred in scorpions it occurred independent from *hth* duplication in spiders and therefore *hth* duplication is not ancestral to the arachnids. The phylogenetic tree shows that the spider *hth2* form a well supported separate clade, not connected with all other chelicerate *hths*. All *hth1* orthologs show the extremely conserved expression in the proximal and medial part of prosomal appendages excluding the chelicera. This shows that the original gene kept its function after gene duplication. Already in the protein sequence level the spider specific *hth2* paralog in contrast shows less conservation than its paralog *hth1*. Fittingly, like observed for *exd* and *dac*, also *hth2* is differentially expressed in the walking legs when compared to the conserved expression of *hth1*. This again indicates a possible neofunctionalisation of this leg gap gene after duplication.

5.3.6 Rapid diversification of *hth2* after neofunctionalization

In accordance with diversification at the protein sequence level after the young duplication event in the spider lineage, we found that also the expression pattern of *hth2* differs for the three spider species used in this study and the previously published data for the bird spider *A. geniculata* (Pechmann and Prpic, 2009; Prpic et al., 2003). The four spider species represent different groups with different levels of phylogenetic relatedness (see Figure 3). The largest expression differences for *hth2* are found in the prosomal appendages. Our detailed analysis shows that although *P. tepidariorum* and *C. salei* belong to the same group, they do not show a more similar expression pattern compared to the bird spider and the haplogyne spider. This indicates that the possible neofunctionalization and following rapid diversification could have led to the development of diversified morphologies of spider legs. All four spider species have *hth2* expressed in a different final number of rings in the fully differentiated walking legs comparing embryos during ventral closure. *C. salei* has six rings in every leg joint, *P. tepidariorum* in contrast has only the three distal rings. In *P. phalangioides* we find strong *hth2* expression only in the most distal ring, whereas in the bird spider we can find ring in every joint, except for the most distal leg joint (Figure 44). Analysis of the expression dynamics further shows that not only the final embryonic walking leg expression is different for the *hth2* homologs in the different species, but that also the development of the rings differs. Usually genes conserve their new function after neofunctionalization and thus do not show diversification after neofunctionalisation (Ohno, 1970). The functional and expressional data for the duplication of *exd* and *dac* nicely fit into this picture. Findings for *hth* in contrast show that genes can rapidly diversify after neofunctionalization and somehow escape the pressure of conservation, which is even more surprising taking into account that the duplication event of *hth* is spider specific and thus very young in terms of evolutionary age. Interestingly a recent study in the mouse and the bat *Miniopterus schreibersii* also showed a diversified function after duplication of the *hth* vertebrate homologs *Meis* in the development of the limbs in mammals (Dai et al., 2014). However, functional studies of the *hth2* homologs are needed to clearly state that these genes actually have undergone neofunctionalisation. Another

possible scenario causing the differing expression patterns of *hth2* is the loss of gene function after duplication. A gene without any function could actually show an extreme fast change during evolution because no selective pressure lasts on it. Thus the young *hth* duplication might serve as playground for evolution, and being on its way of being lost or gaining a novel function. Genes without function are thought to be rapidly lost during evolution, which usually begins with the loss of the open reading frame due to accumulations of mutations caused by lacking selective pressure. The presence of proper open reading frames for all the spider *hth2* homologs contradicts this scenario. However, functional studies are required to reveal the different *hth2* functions in spider. Different RNAi experimental settings were tried, but did not lead to any result yet. A possibility to improve the RNAi for *hth* might be an even earlier mating prior to the injection and the injection into more females.

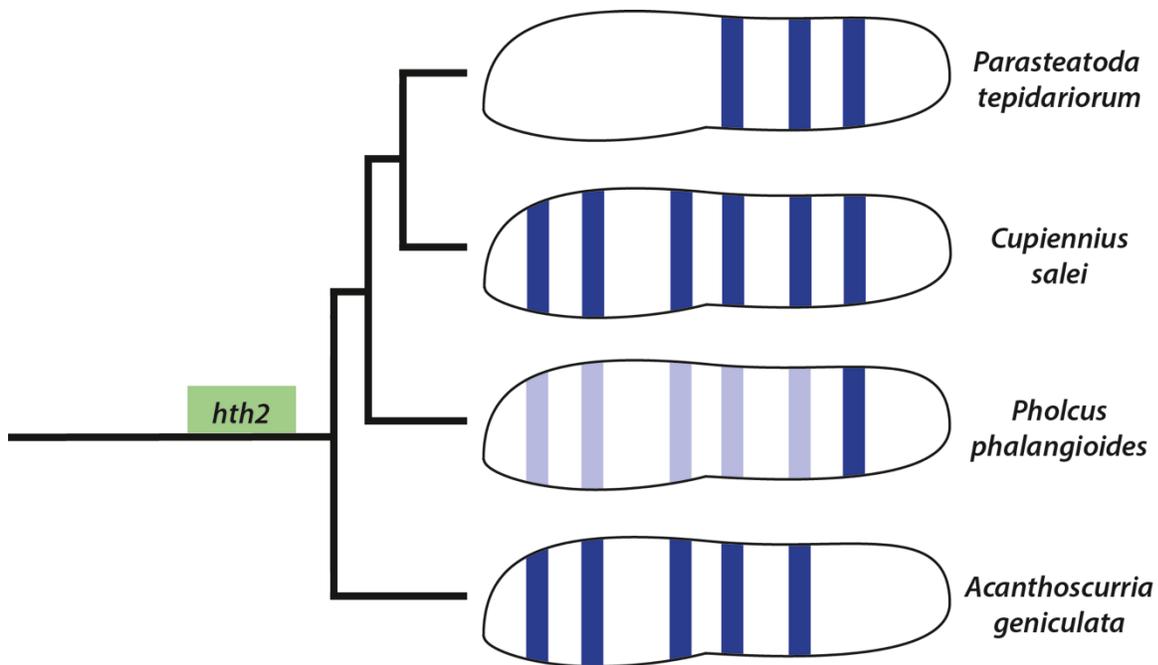


Figure 44: Overview spider *hth2* walking leg expression

The drawings represent a schematic overview of *hth2* expression in the walking legs during late embryogenesis in four different spider species. Blue bars indicate strong expression in the waling legs, light blue bars represent weaker expression doamins. Indicated by the green box is the spider specific duplication event of *hth* as plesiomorphic trait of all spiders.

5.3.7 Leg gap gene conservation and duplication

Findings in several arthropod species already revealed a strict mode of leg gap gene conservation in patterning the main proximo-distal axis in walking legs (Angelini and Kaufman, 2005; Janssen et al., 2010; Pechmann and Prpic, 2009; Prpic and Tautz, 2003; Prpic and Telford, 2008; Prpic et al., 2003; Sharma et al., 2012; Wu and Cohen, 1999). I studied the medial and proximal leg gap genes *hth*, *exd* and *dac* in the appendages of two additional spider species *P. tepidariorum* and *P. phalangioides*. I found that all three original leg gap genes observed here were duplicated. The findings also indicate that after the duplication the more conserved first duplicate always kept the conserved leg gene function required for proper PD axis development, but that the other paralog probably had the ability to evolve new functions undergoing neofunctionalisation after the duplication event. This again shows the importance of gene duplications for development novel or alteration of conserved morphological traits. The time point of the duplication event in the chelicerate lineage as well as the conservation after the possible neofunctionalisation, however, was found to be different in all three cases.

6. Tarsus segmentation candidate genes in spiders

In the first part of the following chapter I will present the phylogenetic and expressional analysis in the two spider models *P. tepidariorum* and *P. phalangioides* of the gene candidates *apterous*, *Lim1*, *spineless*, *trachealess* and *aristaless* that are known to be involved in tarsus segmentation of *D. melanogaster* (Kojima, 2004; Tajiri et al., 2007). The genes *rotund*, *bric-a-brac*, *chip* and *tango* are known to be involved in the distal patterning of the fly leg as well, but did not show any staining during embryonic development of the two spiders and therefore were excluded from the thesis. The second part of this chapter is concerned with the phylogenetic and detailed expressional analysis of the spider homolog of the *D. melanogaster* pretarsus gene *clawless*, also known as C15. The function in the legs of the developing nymphs of one *P. tepidariorum* *c11* paralog was studied after gene knockdown via RNAi and is included in that second part as well. The third part describes the phylogenetic relationships of the spider paralogs of another distal segmentation gene candidate, *BarH* and their expression as well as its function in *P. tepidariorum*. Finally all findings will be discussed in a separate part at the end of this chapter.

6.1 *apterous* in *P. tepidariorum* and *P. phalangioides*

apterous (*ap*) is a homeodomain transcription factor from the Lim-type family (Cohen et al., 1992). A search for homologs in the transcriptomes of *P. tepidariorum* and *P. phalangioides* and subsequent phylogenetic analysis (Supplemental Figure 11 and Supplemental Figure 12) identified three homologs for each spider, which were named according to their phylogenetic relationship with the *ap* homologs from *C. salei* (Damen et al., 2002). The phylogenetic analysis showed that the *ap* duplicate *ap2* from *C. salei*, is present in two copies for both *P. tepidariorum* and *P. phalangioides*. Only *ap1* of *P. tepidariorum* and *P. phalangioides*, as well as *Pp-ap2.2* could be successfully cloned and used for whole mount *in situ* hybridization.

Expressional analysis of the cloned paralogs was performed to investigate their expression in the prosomal appendages and especially in the tarsus during embryonic development. Interestingly contrary to the phylogenetic analysis the expression pattern of *Pt-ap1* is more similar to *Pp-ap2.2*, than to *Pp-ap1*. *Pp-ap2.2* and *Pt-ap1* expressions both start with the development of the prosomal limb buds and stay at the base of these appendages during entire limb bud development (arrow, Figure 45 A-C and Figure 47 A-C).

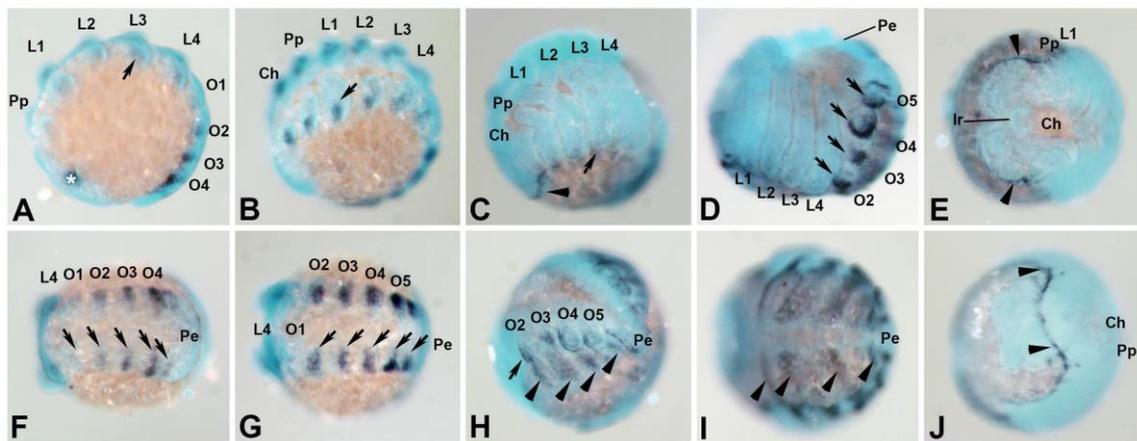


Figure 45: Expression of *ap1* in *P. tepidariorum*

A: Embryo at full germ band elongation, lateral view. The arrow indicates beginning *ap1* expression in the prosomal appendages. **B:** Lateral view of an embryo during limb bud elongation with expression at the base of the appendages (arrow). **C:** Embryo during inversion expressing *ap1* laterally on the head (arrowhead) and proximally in the prosomal appendages (arrow), lateral view. **D:** Embryo during ventral closure, lateral view. The arrows point to *ap1* expression in the opisthosomal limb buds. **E:** Embryo during inversion, shown in frontal aspect with the walking pointing to the right. Arrowheads indicate *ap1* expression laterally surrounding the head lobes. **F:** Posterior view of an embryo with fully elongated germ band. **G:** Posterior view of an embryo during limb bud elongation. The arrows mark segmental expression of *ap1* in the opisthosoma. **H:** Embryo during inversion, posterior view. The arrow points to *ap1* expression in the opisthosomal limb buds, the arrowhead marks stripes of *ap1* expression in the dorsal part of the opisthosoma. **I:** Dorsal view of an embryo after dorsal closure with stripes of *ap1* expression in the dorsal part excluding the heart tissue. **J:** Frontal view of an embryo during ventral closure. The arrowheads point to the expression of *ap1* on the edge of the non-neurogenic ectoderm overgrowing the neural head tissue. The embryos in A-E are oriented with anterior on the left. The embryos in F-I are oriented with posterior to the right. Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

With beginning of ventral closure the *P. phalangioides* ortholog, *ap2.2* is additionally expressed as two spots in the patella segment and one spot near the joint of the metatarsus and tarsus (arrow, Figure 47 D and E). *Pp-ap1* is in contrast not expressed in any of the prosomal appendages at all. Like expression in the prosomal appendages also

the opisthosomal expression is more similar between *Pt-ap1* and *Pp-ap2.2* and strongly resembles that of the *ap1* ortholog studied in *C. salei*. Opisthosomal expression of *Pt-ap1* starts at late germ-band stages as stripes in every opisthosomal segment (arrow, Figure 45 F). During limb bud elongation expression in the first opisthosomal segment vanishes and additional expression in O6 appears (arrow, Figure 45 G). During dorsal closure the expression pattern becomes more differentiated and is now present in all opisthosomal limb buds and as stripes in the dorsal areas of the opisthosomal segments. This expression of *Pt-ap1* stays during ventral closure of the embryo, but is absent from the most dorsal opisthosomal tissue, the developing heart. In *P. phalangioides* strong *ap2.2* expression in the opisthosomal limb buds begins during limb bud elongation and is similar to *Pt-ap1* (arrowhead, Figure 47 G). The expression in the dorsal opisthosomal tissue is weaker and only visible during dorsal closure stages (arrowhead, Figure 47 H). In addition we observe a specific opisthosomal expression for *Pp-ap2.2*, which was not previously reported for *C. salei* or observed for the *P. tepidariorum* paralog (Damen et al., 2002).

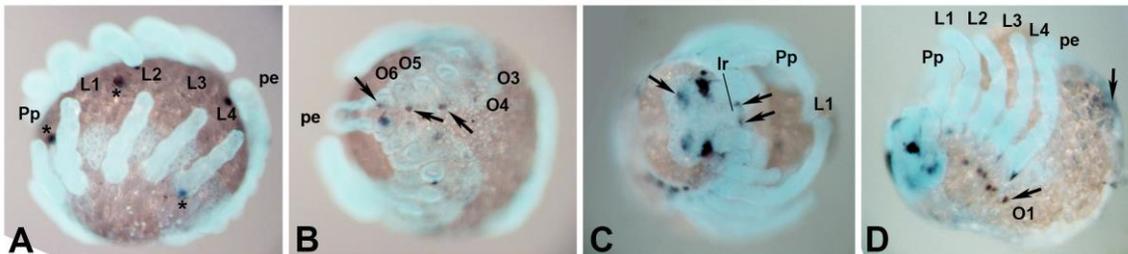


Figure 46: Expression of *ap1* in *P. phalangioides*

A: Lateral view of an embryo during inversion without *ap1* expression. The asterisk indicates unspecific staining caused by precipitation. **B:** Posterior view of an embryo during inversion. The arrows indicate spots of *ap1* expression. **C:** Embryo during ventral closure shown in frontal aspect. The arrow points to expression in the neurogenic ectoderm of the head and two *ap1* spots in the labrum. **D:** Embryo during ventral dorsal closure, lateral view. The arrows mark *ap1* spots laterally to the walking legs and a diffuse domain of *ap1* expression in the posterior part of the opisthosoma. Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

Early during formation and prolongation of the opisthosomal tail, *Pt-ap2.2* is expressed as a semicircle near the posterior end (arrow, Figure 47, F and G). During inversion when the opisthosomal tail begins to be folded down on the yolk this semicircle disappears and dorsal spots of expression are found in the opisthosomal segments posterior to O5

(arrow, Figure 47 H). A similar expression pattern is observed in the opisthosomal segments O5-O7 for the *Pp-ap1* paralog (arrow, Figure 46, B). During ventral closure, however, distinct *Pp-ap1* expression is present as spots on the lateral side of the prosoma, dorsal to the last three walking legs and in a broad domain in the posterior opisthosoma (arrow, Figure 46, D). The absence of this specific opisthosomal expression pattern in both entelegyne spiders might be correlated with the haplogyne specific opisthosomal organization, but this awaits functional tests in *P. phalangioides*. The only expression, which is more similar for the two *P. phalangioides* paralogs compared to *Pt-ap1* is found in the head. *Pt-ap1* is expressed as an outer rim surrounding the head region during inversion (arrowhead, Figure 45, E). This *Pt-ap1* expression stays on the rim of the non-neurogenic ectoderm when this overgrows the neurogenic ectoderm during head tissue rearrangement at ventral closure (arrowhead, Figure 45, J). In contrast *Pp-ap1* expression in the head starts not prior to ventral closure and is present in two defined areas in the neurogenic ectoderm of the head and in the labrum. The expression of *Pp-ap2.2* in the head starts earliest. First expression is found in a medial head region during limb bud elongation stages (arrowhead, Figure 47, I). During dorsal closure this expression is found near the anterior furrow (arrowhead, Figure 47, J) and later during ventral closure *Pp-ap2.2* is expressed in a broad domain in the neurogenic ectoderm and strongly in two defined lateral head regions anterior to the chelicera (arrowhead, Figure 47, E).

In summary expression in the walking legs is only found for *Pt-ap1* and *Pp-ap2.2* and is very similar for the two genes. Expression in the head is different for all three genes studied here, but more similar for the two paralogs of *P. phalangioides*. The opisthosomal expression of *Pp-ap2.2* is the most complex and shares pattern elements with both other genes observed here.

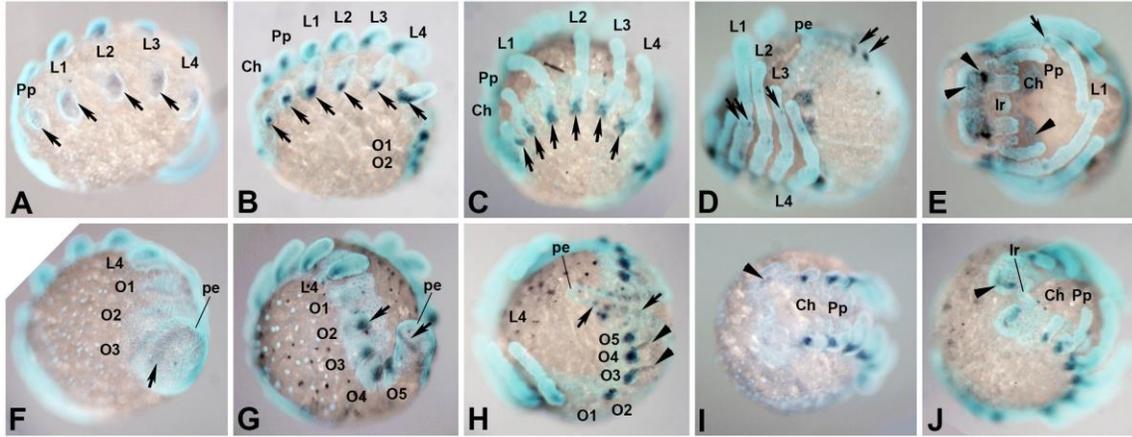


Figure 47: Expression of *ap2.2* in *P. phalangioides*

A: Embryo at beginning of prosomal limb bud development, lateral view. The arrow marks starting expression of *ap2.2*. **B:** Lateral view of an embryo during limb bud elongation. **C:** Embryo during late inversion, lateral view. The arrows in B and C indicate *ap2.2* at the base of the prosomal appendages. **D:** Embryo during ventral closure, lateral view. The arrows indicate spotted *ap2.2* expression in the legs and two *ap2.2* stripes in the posterior opisthosoma. **E:** Frontal view of an embryo during ventral closure. The arrow points to the spotted *ap2.2* in L1, the arrowheads point on the expression of *ap2.2* in the head and the ventral neurogenic tissue. **F:** Posterior view of opisthosomal expression of *ap2.2* in the posterior end (arrow) during beginning of prosomal limb bud development. **G:** Posterior view of an embryo during limb bud elongation, walking legs pointing to the top. The arrows mark *ap2.2* expression in the opisthosomal limb buds and as a semicircle on the posterior end. **H:** Posterior view of an embryo during late inversion, walking legs pointing to the left. The arrows point to the spotted *ap2.2* expression in the dorsal part of the opisthosomal segments. The arrowheads mark the opisthosomal *ap2.2* stripes expanding dorsally from the limb buds. **I:** Embryo during limb bud elongation, shown in frontal aspect. **J:** Frontal view of an embryo during late inversion. The arrowhead in I and J points to the *ap2.2* expression near the anterior furrow. Embryos are oriented with anterior to the left, exceptions are H-J. Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

6.2 *Lim1* in *P. tepidariorum* and *P. phalangioides*

Lim1 from *D. melanogaster* is the ortholog of Lhx1/Lhx5 from vertebrates and like *ap* another member from the Lim-type gene family (Lilly et al., 1999).

In the transcriptomes of *P. tepidariorum* and *P. phalangioides* two paralogs were identified. The phylogenetic tree (Supplemental Figure 9 and Supplemental Figure 10) shows that this duplication is probably common for spiders, because also two predicted paralogs were found for another spider *S. mimosarum*. However, to answer the question if the duplication is ancestral in arachnids or chelicerates a more detailed phylogenetic analysis is needed. Nevertheless the phylogenetic tree performed reveals that the

duplication events in spiders and vertebrates are unrelated, because vertebrate Lhx1 and Lhx5 form a separate monophyletic group.

Both paralogs for *P. tepidariorum* and one for *P. phalangioides* (*Pp-Lim1.2*) were successfully cloned and their expression was studied. The two paralogs of *Lim1* found in *P. tepidariorum* are similarly expressed during early embryonic development. With beginning of limb bud development expression starts at the proximal base of the legs and the pedipalps (early limb bud stages not shown) (asterisk, Figure 48, A and F) including the gnathendite (arrowhead, Figure 48, E, I and J). This expression at the proximal base remains unchanged until the end of embryonic development for both paralogs.

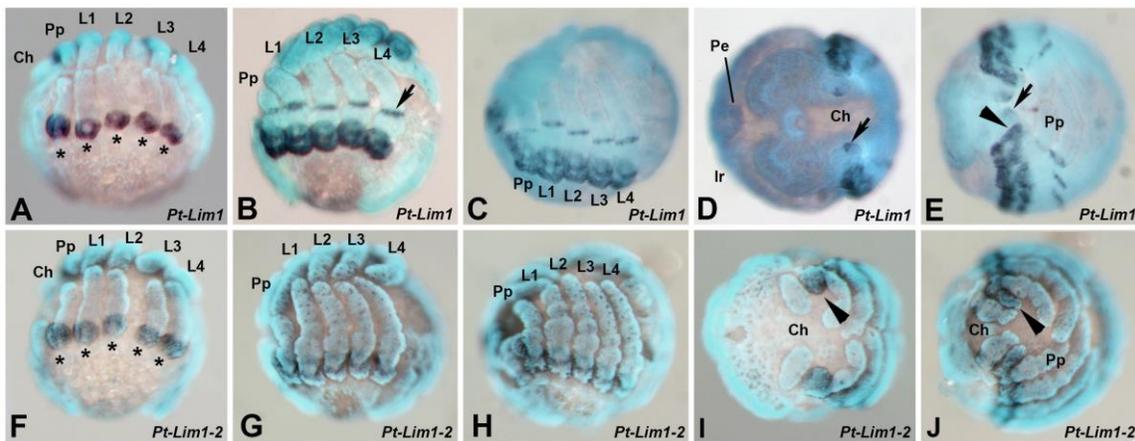


Figure 48: Expression of *Lim1* paralogs during embryonic development of *P. tepidariorum*

Expression of *Lim1* in the prosomal appendages of *P. tepidariorum* embryo during limb bud elongation (A), inversion (B) and ventral closure (C), all in lateral view. The arrow points to the *Lim1* ring in the pedipalps and walking legs. D: Frontal view of the head during inversion. The arrow points to *Lim1* expression in the ventral part of the chelicera. E: Embryo during ventral closure, frontal view. The arrow points to *Lim1* expression in the ventral part of the chelicera. The arrowhead marks the expression of *Lim1* in the gnathendite. Expression of *Lim1.2* in the prosomal appendages of *P. tepidariorum* embryo during limb bud elongation (F), inversion (G) and ventral closure (H), all in lateral view. I: Frontal view of the head during inversion. J: Embryo during ventral closure, frontal view. The arrowhead in I and J mark the expression of *Lim1* in the gnathendite. The asterisk in A and F indicate the start of expression at the base of the pedipalps and walking legs. All embryos are oriented with anterior to the left. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, pe: posterior end.

However, during inversion stages an additional ring of *Pt-Lim1* expression begins in the medial part of the legs and the pedipalps (arrow, Figure 48, B). At the same time *Pt-Lim1* expression starts on the ventral part of the chelicera (arrow, Figure 48, D and E). The *Lim1* paralog of *P. phalangioides*, *Pp-Lim1.2*, is expressed largely similar to its *P.*

tepidariorum paralog *Pt-Lim1.2* in the pedipalps and walking legs. The only expressional difference in these appendages is a ventral stripe of expression (red arrow, Figure 49 B, C and H). The expression of *Pp-Lim1.2* in the other embryonic tissues is more distinct between the two spider species. Before strong expression starts at the proximal base of the pedipalps and walking legs it is already expressed in the ventral neurogenic ectoderm anterior to the chelicera (arrowhead, Figure 49, A). This neurogenic expression of *Pp-Lim1.2* stays and gets further defined during head differentiation (arrowhead, Figure 49, E and F). Another expression uniquely found for the *P. phalangioides* gene starts just before inversion in the proximal base of the chelicera. Also in the developing limb buds of the opisthosoma *Pt-Lim1.2* is expressed throughout development. RNAi of *Pt-Lim1* did not lead to any obvious phenotype in the freshly hatched nymphs. A repetition of this RNAi experiment would be required to clarify the gene function.

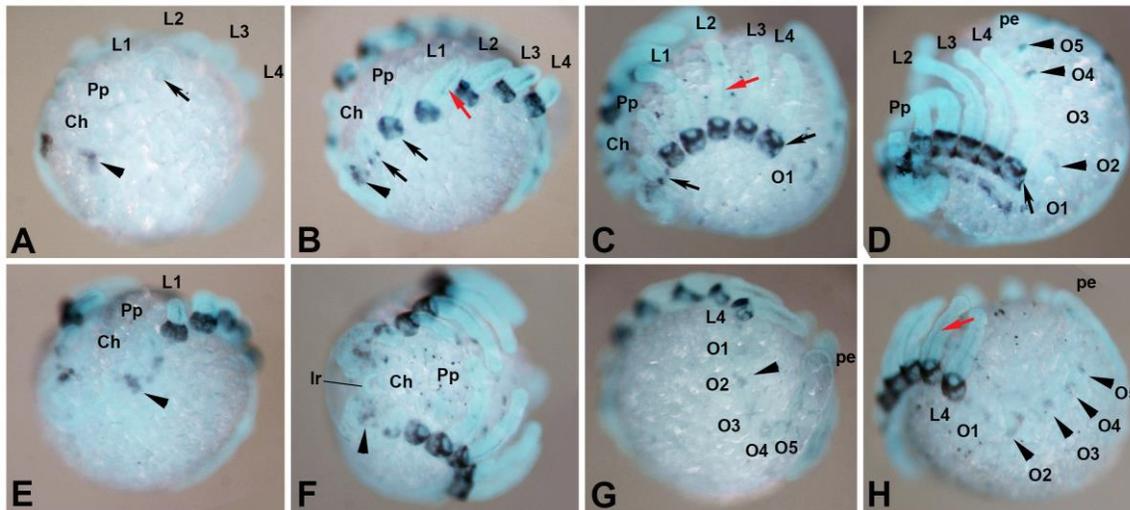


Figure 49: Expression of *Lim1.2* in *P. phalangioides*

Expression of *Lim1.2* in during embryonic development of *P. phalangioides* at early limb bud development (A), inversion (B), dorsal (C) and ventral closure (D), all in lateral view. E: Embryo during inversion shown in frontal aspect. F: Embryo during dorsal closure, frontal view. G: Posterior view of an embryo during inversion. H: Embryo during dorsal closure in posterior view. The arrowhead in A, B, E and F points to a domain of *Lim1.2* expression anterior to the chelicera, later giving rise to part of the neurogenic ectoderm. The black arrows in A, B, C and D point to the proximal expression domain of *Lim1.2* in the prosomal appendages. The arrowheads in D, G and H mark *Lim1.2* expression in the opisthosomal limb buds. All embryos are oriented with anterior to the left. The staining was performed by Jana Holtzheimer. Ir: Labrum Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

6.3 *spineless* in *P. tepidariorum* and *P. phalangioides*

spineless (*ss*) is a transcription factor and belongs to the family of basic-helix-loop-helix-PAS (bHLH-PAS) domain proteins. The name PAS derives from the founding members of this family, namely: Period, Aryl hydrocarbon receptor (Ahr) and single-minded (for review see (Crews, 1998; Crews and Fan, 1999)). Early studies in *D. melanogaster* found that it is, as its name says, involved in correct bristle formation and also development of distal antenna and leg regions (Emmons et al., 1999; Struhl, 1981).

Due to high sequence similarity caused by the PAS domain the phylogenetic analysis for *ss* was performed in combination with another bHLH-PAS gene, *tracheless* that is also known to be involved in tarsus segmentation of *D. melanogaster* (Tajiri et al., 2007). The phylogenetic analysis (Supplemental Figure 13 and Supplemental Figure 14) indicates that two *spineless* paralogs are present in the spider species *P. tepidariorum* and *P. phalangioides* (depicted in red). This duplication is probably ancient for spiders because two paralogs are also present in the spider *S. mimosarum*. Consistent with their gene family relationships the spider genes form a monophyletic group with *ss* homologs from onychophorans and insects, as well as with their vertebrate homologs, the Ahr proteins. The other candidates found in the transcriptomes of the two spiders studied here cluster together with other bHLH-PAS domain proteins like *clock* (depicted in violet), *similar* (depicted in green), *tracheless* and its vertebrate homolog neuronal-PAS-containing protein (dark red) and *single minded* (depicted in blue).

I cloned and performed *in situ* hybridization for all four *ss* paralogs identified in the phylogenetic tree. In *P. tepidariorum* expression of *ss1* in the prosomal appendages starts during limb bud elongation with one spot at the distal part of the chelicera and two spots in the walking legs (arrow, Figure 50, A). During limb bud differentiation additional expression in the gnathendite and corresponding tissue in the ventral neuroectoderm under the legs starts and medial spots of expression arise in the walking legs (arrowhead, Figure 50, B). The expression in the distal part becomes stronger and more spots of expression arise, covering the distal part of the walking legs and pedipalps in a ring like manner (arrowhead, Figure 50, B). During inversion and dorsal closure stages a broader expression domain arises underneath the spotted expression in the distal part of

the walking legs and pedipalps (red bar, Figure 50, C and D). In addition more spots of expression arise in the medial part of the pedipalps, the walking legs and the chelicera (Figure 50, D). Expression of *ss1* in the prosomal appendages during early embryonic development of *P. phalangioides* is comparable but not identical to *P. tepidariorum*. Just after beginning of inversion *Pp-ss1* is expressed in the ventral neuroectoderm at the basis of the walking legs (arrowhead, Figure 51, B) and in a spotted manner in the distal part of the walking legs and pedipalps (arrow, Figure 51, B). During dorsal closure additional medial spots arise in the walking legs, whereas the distal expression does not broaden as in *P. tepidariorum* (arrow, Figure 51, C). In addition only one single spot of *Pp-ss1* is expressed in the chelicera during dorsal closure (arrow, Figure 51, G). However, during ventral closure the expression in the chelicera is virtually identical to *P. tepidariorum* showing a complex spotted expression pattern in the chelicera (Figure 51, H).

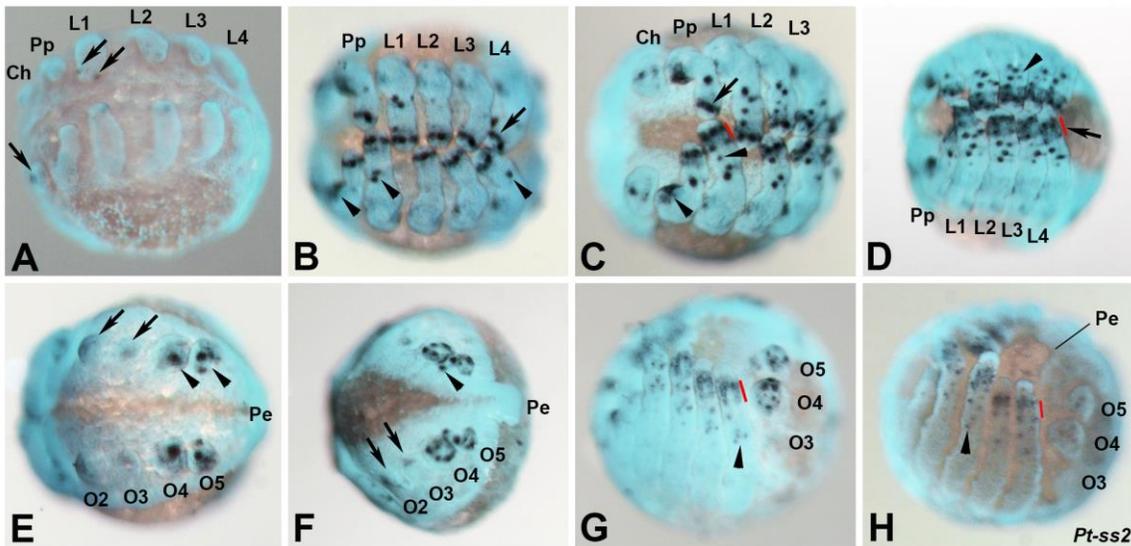


Figure 50: Expression of *spineless* paralogs in *P. tepidariorum*

A-G: Expression of *ss1* in *P. tepidariorum*. **A:** Embryo during limb bud elongation, lateral view. **B:** Ventral view of the prosomal appendages during limb bud differentiation. **C:** Inversion. **D:** Dorsal closure. **E:** Embryo during limb bud elongation, ventral view. **F:** Embryo during dorsal closure, posterior view. **G:** Lateral view of an embryo during ventral closure. **H:** Embryo expressing *Pt-ss2* during ventral closure, lateral view. Detailed expression of *ss* paralogs is described in the text. Embryos in A-D and G and H are oriented with anterior to the left, E and F are oriented with posterior to the right. Jana Holtzheimer helped by the stainings of *Pt-ss2*. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

In contrast to the similar expression of *ss1* in the chelicera of late embryonic stages, the expression in the remaining prosomal appendages differs drastically between the two spiders during ventral closure. Whereas the *ss1* expression pattern in *P. tepidariorum* remains similar to earlier stages, it changes significantly for *P. phalangioides*. With the beginning of ventral closure *Pp-ss1* is expressed in a complex ring like manner in the pedipalps and very strongly in the entire walking legs excluding the areas of joint formation (Figure 51, H and D). In the opisthosoma the expression pattern of *ss1* is more comparable for the two spider species, and differences are only observed concerning the temporal expression profile. In *P. phalangioides* opisthosomal *ss1* expression already starts after early prosomal limb bud formation and covers the dorsal part of the developing opisthosomal tail bud (arrow, Figure 51, A). During inversion this opisthosomal expression is restricted to the four opisthosomal limb buds (arrows, Figure 51, E) in *P. phalangioides*. Later expression in the second and third opisthosomal limb bud, however, first fades (dorsal closure, Figure 51 F, arrow) and finally completely disappears for the limb bud on O3 (ventral closure, Figure 51, D). The expression in the limb buds on O4 and O5 in contrast differentiates into a spotted expression pattern (arrowhead, Figure 51, D). This differentiation (arrowhead) and fading (arrows) can also be observed in *P. tepidariorum* but starts already during limb bud elongation and continues during dorsal closure (Figure 50, E and F).

The expression of the second paralog found in both spider species starts remarkably later and weaker. *Pt-ss2* starts during ventral closure stages and is virtually identical to its paralog *Pt-ss1*. *Pp-ss2* expression already starts during dorsal closure similar to *Pp-ss1* in the distal part of the walking legs (arrow, Figure 51, I) and in the limb buds on O4 and O5 (arrow, Figure 51, J). During ventral closure the distal expression in the walking legs becomes stronger and additional spots in the medial part arise (arrow, Figure 51, K). Expression also arises in the chelicera (arrow) and the gnathendite of the pedipalps (arrowhead, Figure 51, L). Hence, *Pp-ss2* does not change so dramatically as its paralog *Pp-ss1* and is thus more comparable to the expression found for the two *P. tepidariorum* paralogs. RNAi of *Pt-ss1* did not lead to any obvious phenotype in the freshly hatched nymphs. A repetition of this RNAi experiment including analysis of older nymphal stages is currently in progress.

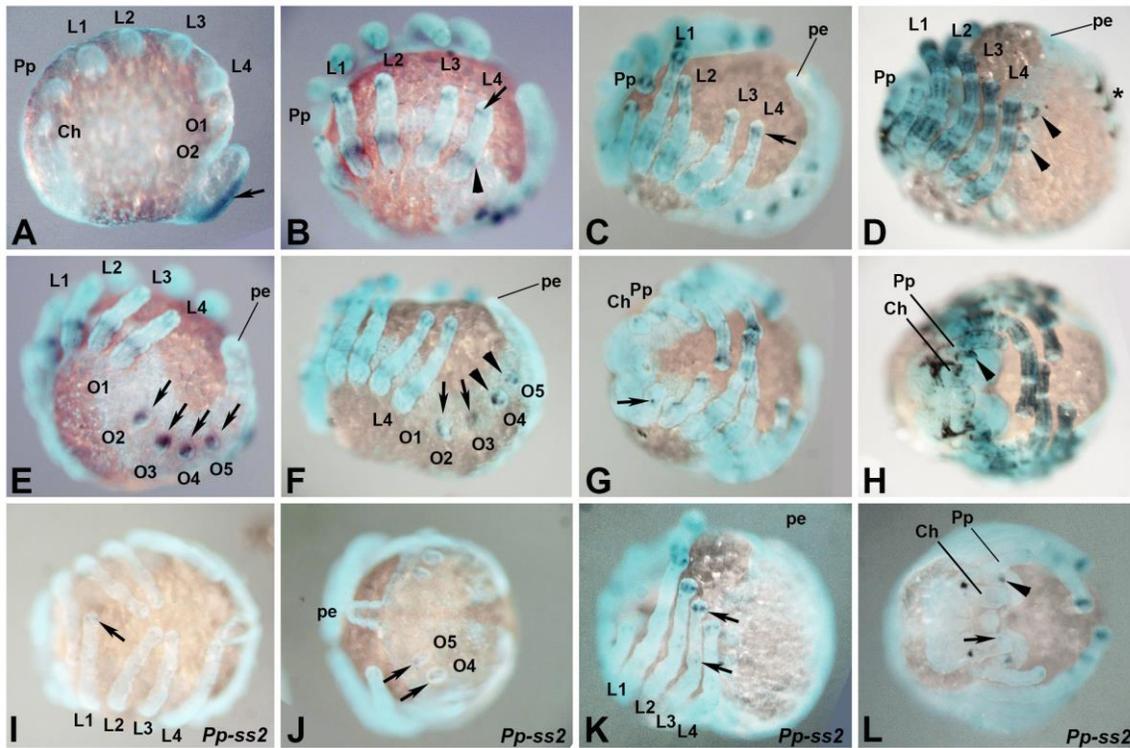


Figure 51: Expression of *spineless* paralogs in *P. phalangioides*

A-H: Expression of *ss1* during embryonic development of *P. phalangioides*. I-L: Expression of *ss2* in embryos of *P. phalangioides*. A: Embryo during early limb bud formation. B: Embryo at inversion. C: Embryo during dorsal closure. D: Embryo during ventral closure. The asterisk marks unspecific staining in the forming cuticle tissue. E: Opisthosoma of an embryo during inversion. F: Opisthosoma of an embryo during dorsal closure. The embryos in A-F are all shown in the lateral view. G: Head of an embryo in frontal view during dorsal closure. H: Frontal view of an embryo during ventral closure. I: Ventral view of an embryo during dorsal closure. J: Posterior view of an embryo during dorsal closure. K: Embryo during dorsal closure in lateral view. L: Embryo during ventral closure shown in the frontal aspect. Detailed expression of *ss* paralogs is described in the text. All embryos are shown with anterior to the left, except for panel J. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

6.4 *trachealess* in *P. tepidariorum* and *P. phalangioides*

trachealess (*trh*), encodes like *ss*, a PAS domain protein and is thus a member of the bHLH–PAS family (Córdoba and Estella, 2014).

Two paralogs of *trh* were identified after detailed phylogenetic analysis in the transcriptomes of *P. tepidariorum* and *P. phalangioides* (Supplemental Figure 13 and Supplemental Figure 14). Although not many representatives of other chelicerate species were found in the blast search the phylogeny nevertheless supposes that the paralogs of *P. tepidariorum* and *P. phalangioides* cluster together into two distinct chelicerate *trh* groups. The time point of a possible *trh* duplication in the lineage leading to the spiders needs to be further studied including more representatives from the chelicerate group. The phylogenetic tree suggests that the duplication of *trh* in chelicerates and the duplication of the vertebrate homologs of the neuronal PAS proteins were two independent events.

According to the clustering found for the spider *trh* homologs, the expression is more similar for orthologs of the different species, than for paralogs of one spider species. *trh1* is expressed in a less complex pattern than its paralog and shows a very comparable pattern in the walking legs of both spider species. In *P. tepidariorum* expression starts in two medial rings in the walking legs during early inversion (arrow, Figure 52, A). Later during inversion one ring of *Pt-trh1* expression also starts in the pedipalps (arrowhead) and the distal ring in the walking legs broadens and starts to split (bar, Figure 52, B). With the onset of dorsal closure and during ventral closure, thus three rings of *Pt-trh1* are present in the medial part of the legs (arrows, Figure 52 C and D). The proximal ring is expressed in the joint between the leg segments femur and patella, whereas the two distal rings are both expressed distally in the tibia. For *P. phalangioides* this late expression pattern of *trh1* in the walking legs is similar, but shows differences in terms of timing in earlier stages as well as in the other embryonic tissues. *Pp-trh1* already starts prior to inversion as one medial ring in the first walking leg pair (arrow, Figure 53, A), which is later during inversion visible in the other legs pairs, but never in the pedipalps (arrow, Figure 53, B). Also splitting of the *Pp-trh1* distal ring occurs slightly later, during ventral closure. However, the three rings of *Pp-trh1* are expressed in

comparable leg segments as observed for *P. tepidariorum* (arrows, Figure 52 and 53, C). In addition to these comparable *trh1* expression patterns, *Pp-trh1* has also unique expression different to *Pt-trh1*. During ventral closure *Pp-trh1* is also expressed at the proximal part of the pedipalps and walking legs as well as in the limb bud on the third opisthosomal segment during inversion (arrow, Figure 53, D). In summary *trh1* is expressed in a very conserved pattern in the walking legs of these two spiders and only the exact timing and expression in the opisthosoma is different.

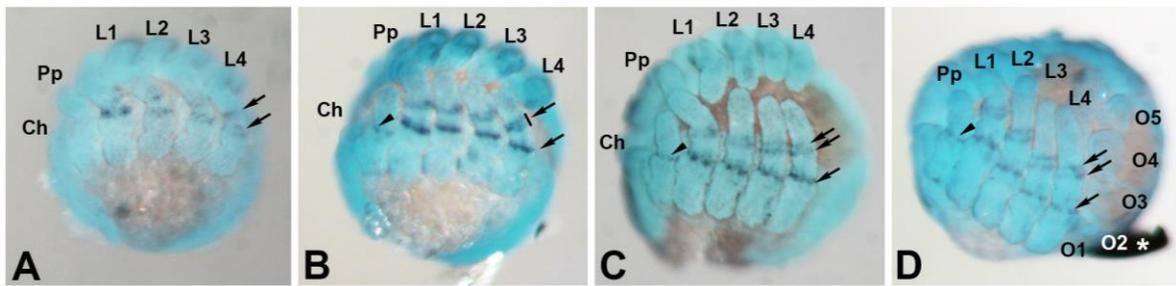


Figure 52: Expression of *trh1* in *P. tepidariorum*

Expression of *trh1* in prosomal appendages during embryonic development of *P. tepidariorum* at limb bud elongation (A), inversion (B), dorsal closure (C) and ventral closure (D). D: The asterisk marks remaining vitelline membrane which was accidentally not removed before the *in-situ* hybridization and shows unspecific staining. Detailed explanation of the expression pattern is found in the text. All embryos are shown in lateral view and oriented with anterior to the left. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5.

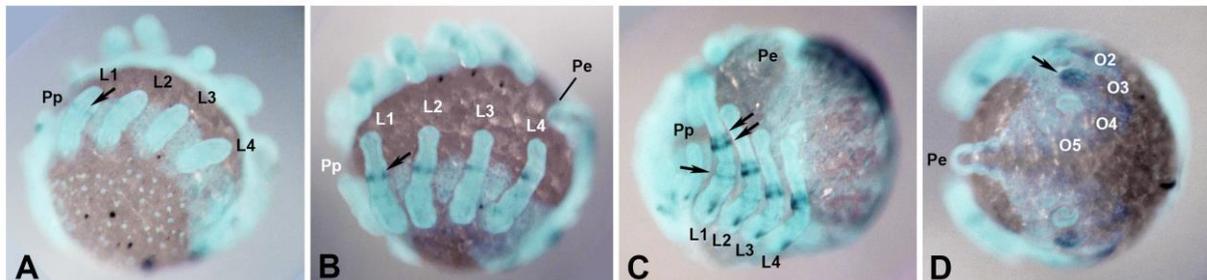


Figure 53: Expression of *trh1* in *P. phalangioides*

Expression of *trh1* in prosomal appendages of *P. phalangioides* during limb bud elongation (A), inversion (B) and ventral closure (C). A-C: All embryos are shown in lateral view and oriented with anterior to the left. D: Dorsal view of the opisthosoma during inversion. Detailed explanation of the expression pattern is found in the text. Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

The expression pattern of *trh2* is in general more complex in the prosomal appendages and the opisthosoma when compared to its paralog *trh1*. For *trh2* ring-like expression in the prosomal appendages starts only very weak prior to dorsal closure (arrow, Figure 54, B and Figure 55, B). At later stages an extensive expression throughout the entire pedipalps and walking legs starts for both spider species. Strong *Pt-trh2* expression starts during dorsal closure stages with a complex pattern in the entire walking legs, including a ring-like pattern in the medial part (arrows, Figure 54, C). During ventral closure this complex *Pt-trh2* expression in the walking legs remains ring-like but is now strongest in the distal part (arrow, Figure 54, D). In *P. phalangioides* a medial ring of expression becomes slightly stronger during dorsal closure (arrow, Figure 55, C). During ventral closure a sudden increase of expression starts and is then even stronger and more defined than for *P. tepidariorum*. *Pp-trh2* during these late stages is expressed in a ring-like manner in the pedipalps and in the entire walking legs, excluding the areas of joint formation (arrow, Figure 55, D). Not only in the walking legs but also in the chelicera *trh2* is differently expressed in the two spiders. Whereas *Pp-trh2* starts very late in the medial part of the chelicera during ventral closure (arrow, Figure 55, D), *Pp-trh2* expression already begins with a spot proximal on the ventral side during inversion (arrow, Figure 54, H). Afterwards a second spot of *Pt-trh2* appears more medially (arrows, Figure 54, I and J). In contrast to the prosomal appendages the expression pattern of *trh2* in the opisthosoma is more comparable in *P. tepidariorum* and *P. phalangioides* and differs only in timing and strength. First *trh2* expression starts at the ventral rim of the opisthosoma during limb bud elongation stages in *P. tepidariorum* (arrowhead, Figure 54, F). In *P. phalangioides* *trh2* expression starts slightly later, during inversion, also at the ventral part of the opisthosoma (asterisk and arrows, Figure 55, B and E), but is already additionally present in the third opisthosomal segment (arrowhead, Figure 55, E). During dorsal closure the ventral expression in *P. phalangioides* vanishes in the more anterior stages, but remains in the more posterior ones and the limb bud on O3 (arrowhead and arrows, Figure 55, F). During ventral closure the expression in opisthosomal limb bud is covered by the legs and cannot be observed. The expression of *Pt-trh2* is stronger compared to *Pp-trh2*, but also found in the third opisthosomal limb bud during later stages (arrow on O3, Figure 54, G). In

contrast to *Pp-trh2* it stays also in the ventral part of the opisthosoma, as well as in the posterior end and is additionally expressed in the more posterior opisthosomal limb buds after inversion. During dorsal closure *Pt-trh2* is expressed on the ventral rim of the limb bud on O4 and remains there during ventral closure (arrow on O4, Figure 54, G and E). At this stage two additional spots of *Pt-trh2* are present in the limb bud on O4 and O5 (arrows on O5, Figure 54, E). Unique is the expression of *Pt-trh2* at the anterior rim of the head during these late embryonic stages (arrowhead, Figure 54, I and J).

In summary *trh2* in both spiders is expressed stronger in the prosomal appendages and the opisthosoma compared to its paralog, especially during late stages. The pattern in these tissues is similar but not entirely identical and differs in timing.

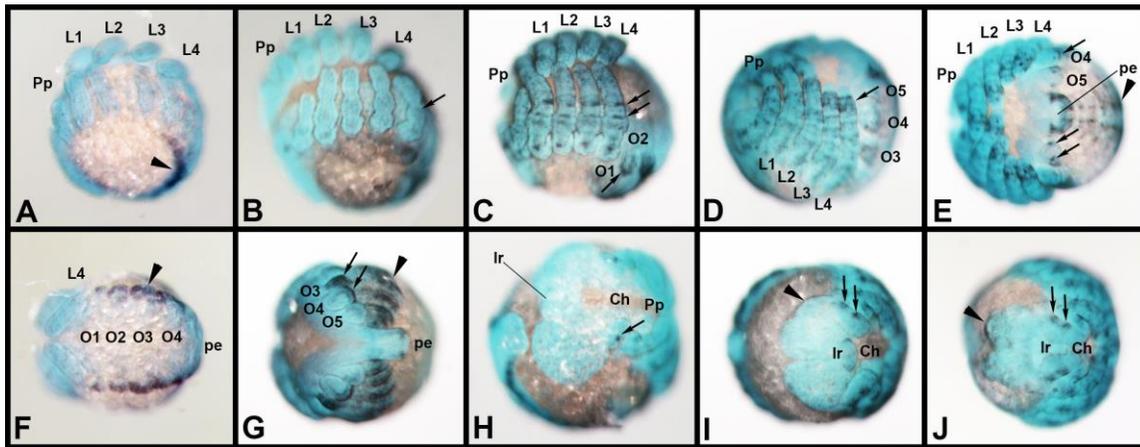


Figure 54: Expression of *trh2* in *P. tepidariorum*

Expression of *trh2* in prosomal appendages during embryonic development of *P. tepidariorum* at limb bud elongation (A), inversion (B), dorsal closure (C) and ventral closure (D). The embryos are shown in lateral view. E: Ventral view of an embryo during ventral closure. F: Ventral view of the opisthosoma during limb bud elongation. G: Opisthosoma of an embryo during dorsal closure in posterior view. H: Embryo during inversion shown in frontal aspect with posterior in the left. I: Frontal view with legs pointing to the right of an embryo during dorsal closure. J: Embryo during ventral closure, frontal view walking legs pointing to the right. Detailed explanation of the expression pattern is found in the text. The embryos in A-E are oriented with anterior to the left, in F and G the posterior end is pointing to the right. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

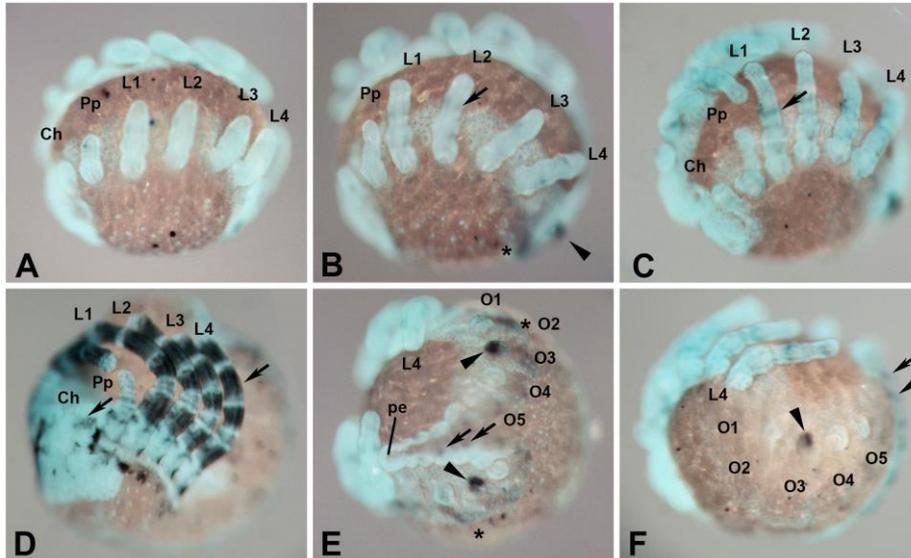


Figure 55: Expression of *trh2* in *P. phalangioides*

Expression of *trh2* in prosomal appendages of *P. phalangioides* during limb bud elongation (A), inversion (B) and beginning dorsal closure (C) and ventral closure (D). The embryos are shown in lateral view. F: Posterior view of an embryo during inversion, the walking legs pointing to the left. F: Dorso-lateral view of the opisthosoma during dorsal closure. Detailed description of expression in the text. All embryos are oriented with anterior to the left, except for panel E. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5.

6.5 *aristaless* in *P. tepidariorum* and *P. phalangioides*

The paired-type homeobox gene *aristaless* (*al*) is required for the development of the distal elements of the antenna and walking legs (Campbell and Tomlinson, 1998).

The phylogenetic analysis of *al* homolog candidates found in the transcriptomes *P. phalangioides* (Janssen et al., 2015) and *P. tepidariorum* (Posnien et al., 2014) showed that both spiders have only one genuine *al* homolog (depicted in red, Supplemental Figure 7 and Supplemental Figure 8). Only one homolog was found for another spider *S. mimosarum*. A number of additional genes that are similar to *aristaless* were found in the transcriptome sequences, but have been excluded from further analysis, because their relationship to *aristaless* is not clear from the phylogenetic analysis.

The *al* homologs were cloned for *P. tepidariorum* and *P. phalangioides*. Expression of *Pt-al* starts as spots in the distal part of the first, third and fourth walking legs at beginning of limb bud development (arrowhead, Figure 56, A). During limb bud elongation stages the spot-like expression of *Pt-al* is visible in the distal part of all legs and the pedipalps

(arrowhead, Figure 56, B) and a broader distal expression domain in the chelicera. Additional expression starts as a ring in the medial part of the walking legs (arrows, Figure 56, B) and pedipalps and a spot on the proximal base of these appendages (asterisk, Figure 56, B). During early inversion stages a second ring of expression appears proximal to the earlier ring (arrow, Figure 56, C). Beginning in late inversion stages until ventral closure these two medial rings and the proximal spots diffuse and give rise to a very complex expression pattern covering the proximal and medial part of the pedipalps and legs (Figure 56, D-F). The spot-like *Pt-al* expression in the distal part of the appendages is then also present as three spots at the very tip of the legs and pedipalps. By contrast, *Pt-al* expression in the chelicera and the opisthosomal limb buds is more uniform during embryonic development (see below). Early expression of *al* in *P. phalangioides* is visually identical with *P. tepidariorum*. With the start of limb formation on the germ band *Pp-al* is expressed as spots at the distal tip of pedipalps and legs (arrowheads, Figure 57, A). As in *P. tepidariorum* the medial ring and proximal spot of *Pp-al* expression (arrows, Figure 57, B) starts during limb bud elongation. With the beginning of inversion this medial ring broadens and proximal to it a domain of *Pp-al* starts, but not so clearly ring shaped as in *P. tepidariorum*. Starting with dorsal closure until ventral closure the expression of *Pp-al* becomes more diffuse covering the proximal and medial region of pedipalps and legs, similar to *P. tepidariorum*. In contrast to the similar expression of *al* in the pedipalps and the walking legs, *Pp-al* never is expressed in a broad distal domain in the chelicera but as a spot which later divides identical to the remaining prosomal appendages (arrowheads, Figure 57, E and F). The expression in the head and opisthosoma, however, is mainly similar for the two observed spider species.

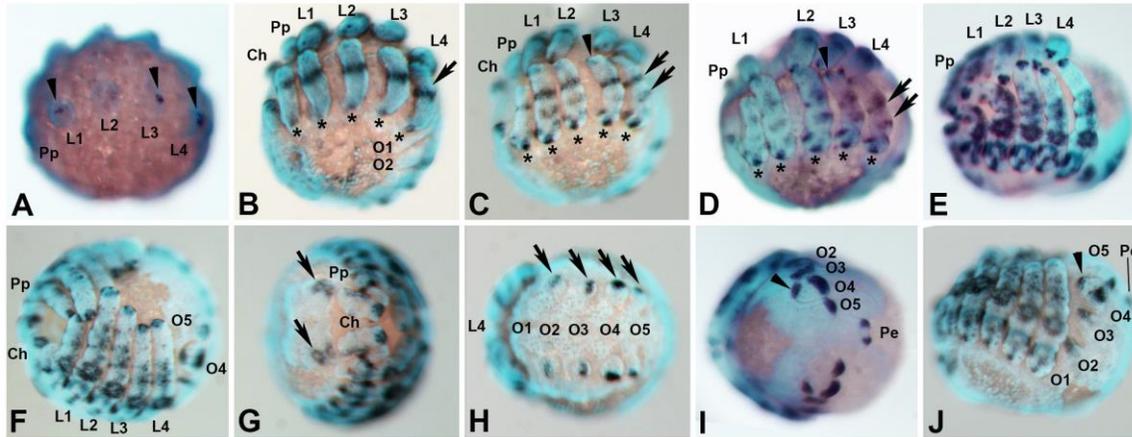


Figure 56: Expression of *aristaless* in *P. tepidariorum*

A: Fully elongated germ band embryo, lateral view. Arrowheads point starting *al* expression on the forming limb buds. Expression of *al* during limb bud elongation (**B**), early inversion (**C**), late inversion (**D**) and dorsal closure (**E**), all embryos in lateral view. Asterisk mark *al* expression at the proximal base of the pedipalps and walking legs. The arrows point to ring-like *al* expression in these appendages. The arrowhead marks the distal *al* spot in the walking legs and pedipalps. **F:** Embryo during ventral closure expressing three distal *al* spots in the walking legs, ventral view. **G:** Frontal view of embryo during ventral closure with *al* expression around the eye field. **H:** Ventral view of the opisthosoma, posterior to the right. The arrows point to *al* expression in the opisthosomal limb buds. **I:** Posterior view of an embryo during dorsal closure. **J:** Embryo during ventral closure, lateral view. The arrowhead in **I** and **J** points to the additional ventral expression domain on the limb bud on O4. All embryos are oriented with anterior to the left, except for the embryos shown in **H** and **I**. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

Until ventral closure *al* is not expressed in the head, but then starts surrounding the fields of lateral eye development in the head lobes (Figure 56, G and Figure 57 F, arrowheads). With the beginning of opisthosomal limb bud development *al* is expressed as a stripe in the opisthosomal limb buds on the second and third opisthosomal segment in both spiders (the precursors of the book lung and tubular trachea) (arrow, Figure 56, H). In the opisthosomal limb bud of the fourth and fifth opisthosomal segment, which will give rise to the spinnerets, *al* is expressed exclusively on the dorsal part. With beginning of dorsal closure in *P. tepidariorum* an additional ventral expression of *al* is visible at the limb bud of the fourth opisthosomal segment (arrowhead, Figure 56, I and J). During the same stage expression of *al* starts as two spots at the posterior end of the *P. tepidariorum* embryo. In *P. phalangioides* expression of *al* on the limb bud on O3 disappears with invagination of the tubular trachea (asterisks, Figure 57, H).

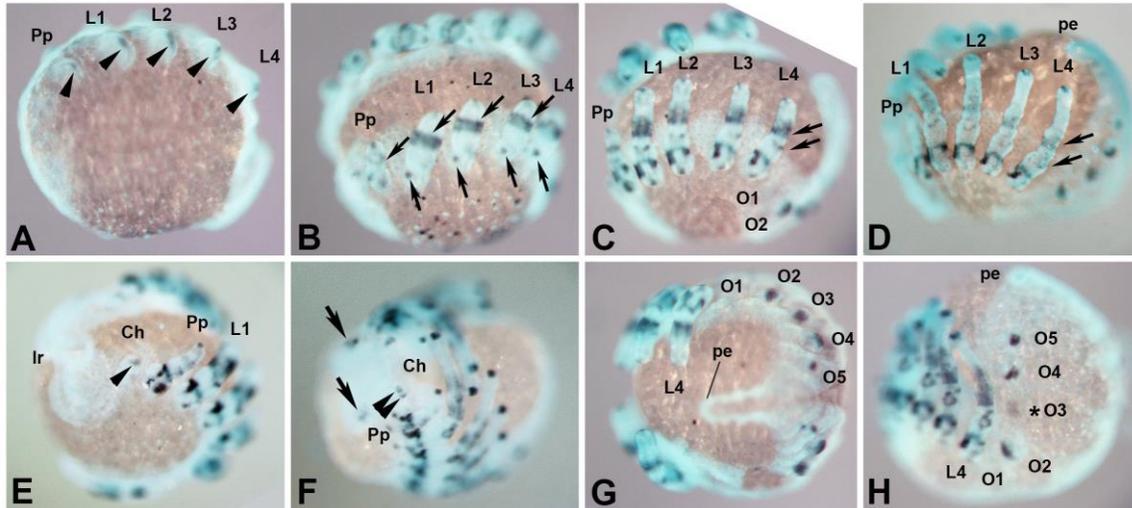


Figure 57: Expression of *aristaless* in *P. phalangioides*

A: Embryo during early limb bud development, lateral view. The arrowheads point to distal *al* expression in the pedipalps and walking legs. **B:** Embryo of *P. phalangioides* during limb bud elongation, lateral view. The arrows mark proximal *al* spot in the walking legs and appearing *al* ring in the pedipalps and walking legs. **C and D:** Lateral view of embryos during inversion and dorsal closure. The arrows point to two *al* ring-like expressions. **E:** Embryo during inversion shown in frontal aspect. The arrowhead points to distal *al* expression in the chelicera. **F:** Frontal view of an embryo during ventral closure. The arrows point to *al* expression around the developing eyes and the arrowheads indicate the division of the distal *al* spot. **G:** Posterior view of the embryo during inversion. **H:** Embryo during ventral closure in lateral view. The asterisk marks the disappearing *al* expression during invagination of the opisthosomal limb bud on O3. All embryos are oriented with anterior to the left, except for the embryo shown in G: here ventral is located left and dorsal right. Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

6.6 Clawless

clawless (*cll*) encodes for a homeodomain protein, and was first identified as a member of the 93DE homeobox cluster located on the third chromosome of *D. melanogaster* and initially named C15 (Dear and Rabbitts, 1994; Dear et al., 1993). The name *cll* was given in accordance to its *D. melanogaster* mutant phenotype lacking the tarsal claws of the walking legs (Campbell, 2005; Kojima et al., 2005).

6.6.1 Phylogenetic analysis of *clawless* duplicates in spider

I searched for *cll* sequences in the transcriptomes of *P. phalangioides* (Janssen et al. 2015) and *P. tepidariorum* (Posnien et al. 2014) and performed a large phylogenetic analysis for all candidates with an e-value below e^{-15} . This phylogenetic tree is shown in Supplemental Figure 15 and Supplemental Figure 16 and revealed a duplication event of *cll* in the two spiders used in this study (depicted in red). In this large phylogenetic analysis a monophyletic group with the highest support value is formed for the *cll* homologs, which are further subdivided into two distinct groups: *cll* homolog sequences from insects and vertebrates (for vertebrates three homologs are present and called T-cell-leukaemia homeobox (TLX) protein) form one group which is further subdivided giving rise to two branches within this group; the second group includes the *cll* homolog of an onychophoran branching off from another subgroup including all chelicerate sequences. The paralogs found for three spiders (*P. tepidariorum*, *P. phalangioides* as well as *S. mimosarum*) form two separate clusters. The paralogs in the distinct clusters were named according to sequence similarity to the *D. melanogaster* homolog. Thus comp116449 and Locus_15589 are referred to as *cll1* and comp 121099 and Locus_21582 as *cll2*. For the other spider *S. mimosarum* a third *cll* is found which clusters together with the other spider *cll2* sequences, but was not further studied. To analyse the origin of the *cll* duplication event in chelicerates in more detail *cll* homologs were additionally searched in the genome of the scorpion *M. martensii* (Di et al., 2014). Then a less inclusive Bayesian phylogenetic tree was computed with a subset of the *cll* homologs identified in the large phylogenetic tree, including also sequences from one more basal bilaterian, the purple sea urchin *Strongylocentrotus purpuratus*.

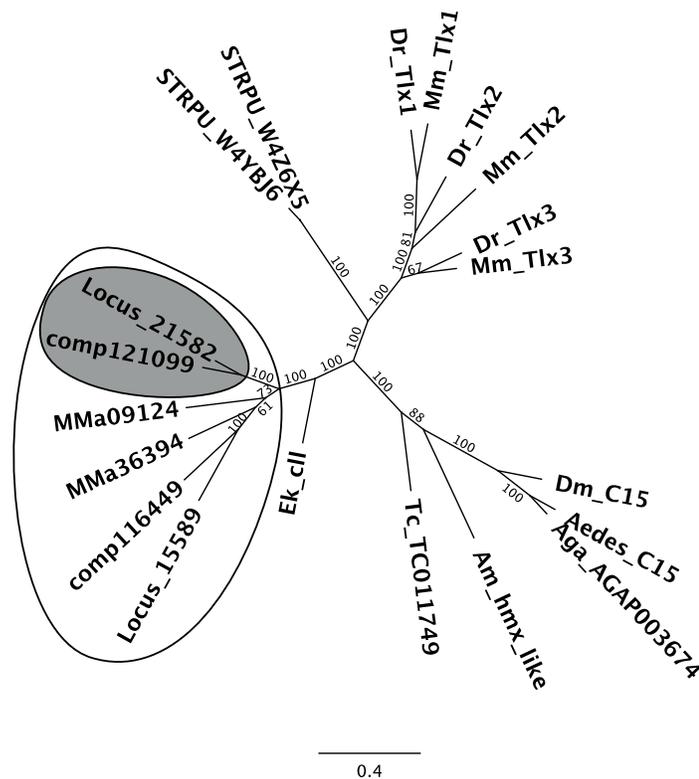


Figure 58: Unrooted 50% phylogenetic majority rule consensus tree of *cll* orthologs

The phylogenetic analysis resulted in three major groups of *cll* homologs. One branch includes the sea urchin homologs branching off from the TLX homologs of vertebrates, shown on the top. The insect *cll/C15* group is located at the right. The third branch includes the *cll* homolog of the onychophoran *Euperipatoides kanangrensis* forming a sister branch to all chelicerate *cll* (marked by black circle) located in the left. The two *cll* paralogs from spiders are present in two distinct phylogenetic groups after duplication. *cll1* forms a monophyletic group with the scorpion homologs, whereas *cll2* of *P. tepidariorum* and *P. phalangioides* cluster together more basally branching in the chelicerate group (Grey circle). Branch lengths of the tree represent the substitutions per site. Numbers at the edges of the phylogram give the probability of each clade indicated by clade credibility values. Abbreviations and accession numbers are given in Supplemental Table 10.

In the unrooted tree (Figure 58) the *cll* sequences form three major clades supported with the highest posterior probability value: the two *cll* homologs of the sea urchin cluster together and form one clade with the vertebrates TLX paralogs, the other clade includes all insects *cll* sequences and the onychophoran *cll* homolog is predicted as the sister group of the chelicerates forming the third group. This phylogenetic analysis clearly shows that the duplication events of the vertebrate TLX and spider *cll* paralogs are entirely unrelated. As already observed in the large phylogenetic analysis, the spider paralogs cluster together in two separate groups. The two scorpion *cll* homologs,

MMa09124 and MMa36394, cluster together with the spider *c11* sequences, but do not form a separate group, which leads to the assumption that the duplication occurred twice within the lineages leading to scorpion and spider and is not ancestral for arachnids. To clearly state whether *c12* has evolved in a spider specific duplication event additional chelicerate sequences are required.

6.6.2 Expression of *clawless*

The expressional analysis of the two *c11* paralogs revealed that both paralogs are strongly expressed in the prosomal appendages with beginning of early limb bud development until late ventral closure stages in both spiders. Their specific pattern, however, differs between the two paralogs. The first paralog, *c11*, is only similarly expressed in the prosomal appendages until ventral closure. The second paralog, *c12*, however, is similarly expressed throughout entire walking leg development, but not in the pedipalps and chelicera (detailed description of *c11* expression in the prosomal appendages in the next section). *Pt-c11* and *Pp-c11* expression starts in the distal tip of all prosomal appendages at very early limb bud development (arrowhead, Figure 59, A and Figure 60, A). This distal expression stays during limb bud elongation (arrowhead, Figure 59, B and Figure 60, B), but then divides and gives rise to two rings of distal expression in the walking legs during inversion stages in both species (arrows, Figure 59, C and Figure 60, C). Whereas this expression of *c11* remains unchanged in the walking legs of *P. tepidariorum* until the end of embryonic development (arrows, Figure 59 D and E), *Pp-c11* expression changes to a single distal spot during ventral closure (arrow, Figure 60, D). The only expression of *c11* found in the head of both spider species is present in the developing labrum with beginning of inversion until ventral closure (Figure 59, F arrow, G arrowhead and H, and Figure 60, E and D). In the opisthosoma *c11* is also similarly expressed in both spiders. Opisthosomal expression starts during inversion stages covering the most dorsal opisthosomal tissue in the presumptive area of the heart (asterisk, Figure 59, I and arrowhead, Figure 60, F). In *P. phalangioides* the dorsal *c11* expression is strongest in the first opisthosomal segment already at these stages. Slightly later with beginning of dorsal closure opisthosomal *c11* expression is restricted to this

segment in both spiders (asterisk, Figure 59, J and arrowhead, Figure 60, G). After rearrangement of the embryonic tissue during dorsal and ventral closure, *c11* expression covers the former first opisthosomal tissue in the field of the developing heart (asterisk, Figure 59, E and arrowhead, Figure 60, H). In summary *c11* is similarly expressed in the head, opisthosoma and early walking legs in both spider species studied here, but differentiates during late walking leg development.

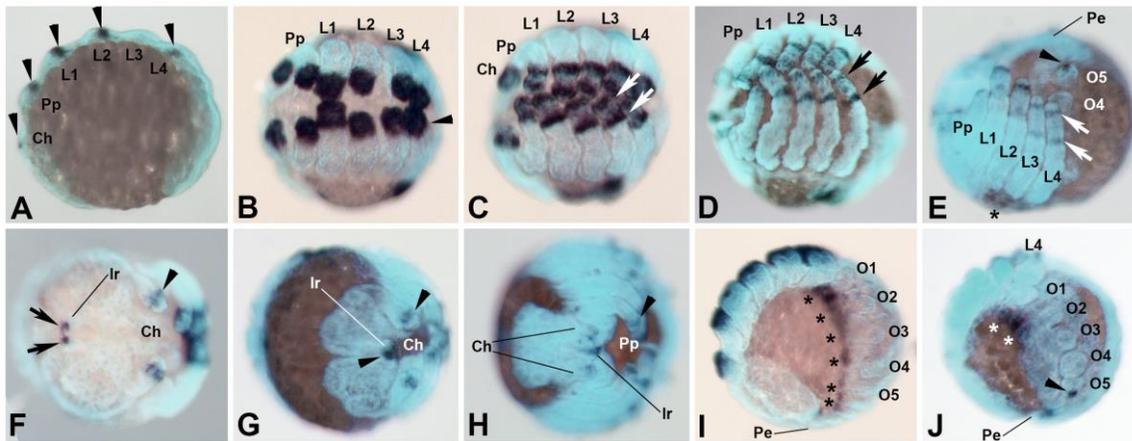


Figure 59: Expression of *c11* in *P. tepidariorum*

A: Lateral view of a germ band stage embryo with *Pt-c11* expression in all developing prosomal appendages excluding L3 (arrowheads). **B:** Ventral view of an embryo during limb bud elongation with strong distal *c11* expression in all prosomal appendages. **C:** Ventral view during of an embryo inversion. **D:** Ventral view of an embryo during dorsal closure. **E:** Lateral view of an embryo during ventral closure. **F:** Anterior view of the head during inversion. **G:** Anterior view of the head during beginning of dorsal closure of the embryo. **H:** Anterior view of the head of an embryo undergoing ventral closure. **I:** Lateral view of the embryo depicted in B. **J:** Lateral view of an embryo during dorsal closure. More detailed description of expression in the text. Anterior is always to the left. Jana Holtzheimer was involved in the staining of the embryos. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

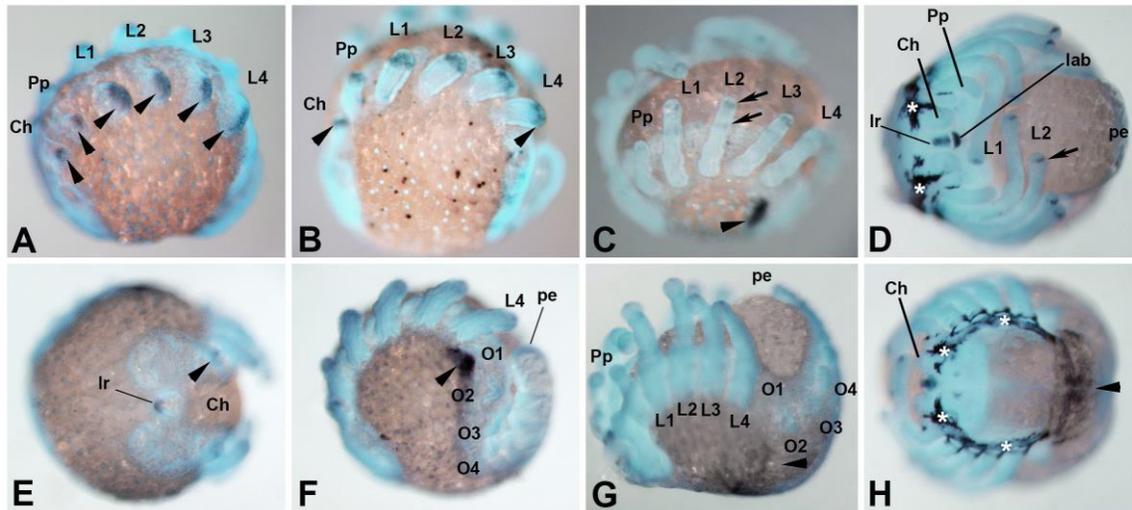


Figure 60: Expression of *cll1* in *P. phalangioides*

Expression of *cll1* in prosomal appendages of *P. phalangioides* during limb bud development (A) and limb bud elongation (B), lateral view. C: Ventro-lateral view of an embryo during inversion D: Ventral view of an embryo during ventral closure. E: Embryo during inversion shown in frontal aspect. F: Lateral view of an embryo during inversion. G: Lateral view of an embryo during dorsal closure. Dorsal view of an embryo during ventral closure. The asterisk in D and H marks unspecific staining of the probe in the forming cuticle. The other marks are described in the according text section. Anterior is always to the left, except for the embryo shown in E, this embryo is oriented with the walking legs to the right. lr: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O4: Opisthosomal segments 1-4, pe: posterior end.

The next part will focus on the comparison of the expression patterns between the paralogs of both spiders (Cloning, probe synthesis and whole mount *in situ* hybridization of *Pp-cll2* and *Pt-cll2* was performed with the help of Jana Holtzheimer). In contrast to *cll1*, *cll2* is not expressed in the opisthosoma of *P. phalangioides*. In *P. tepidariorum* *cll2* is expressed in the opisthosoma, but different to its paralog *cll1*. *Pt-cll2* is expressed at the posterior end of the opisthosoma starting at limb bud elongation until the end of ventral closure (Figure 61, F, G and E). In the head *cll2* is also differently expressed compared to *cll1*, but similar for both spiders. With beginning of limb bud elongation *cll2* is expressed at the anterior rim of the head (arrowhead, Figure 61, H and Figure 62, F). This expression remains unchanged during early dorsal closure stages (arrowhead, Figure 61, I and Figure 62, G). *cll2* expression in the prosomal appendages, especially the walking legs, is like the expression in the head on the one hand different for the two paralogs but on the other hand largely comparable for the two spider species. Although the start of *cll2* expression at the distal tip of the walking legs in *P. phalangioides* (arrowhead, Figure

62, A) still somehow resembles that of *c11*, it is quickly modified and becomes different during limb bud elongation. Directly at this stage not the complete distal tip is expressing *c12*, as observed for *c11*, but in contrast one distal ring of expression is present in the pedipalps and walking legs of both spiders (expression in the chelicera will be explained in the following section) (arrows and arrowhead in Figure 61, A and Figure 62, B). This distal ring of expression stays throughout the rest of embryonic development (arrows, Figure 61, B-E and Figure 62, B-D). Taken together *c12* is expressed in a comparable pattern in the head and the walking legs for both spiders, but not in the opisthosoma and the other prosomal appendages and is not comparable to its paralog *c11*.

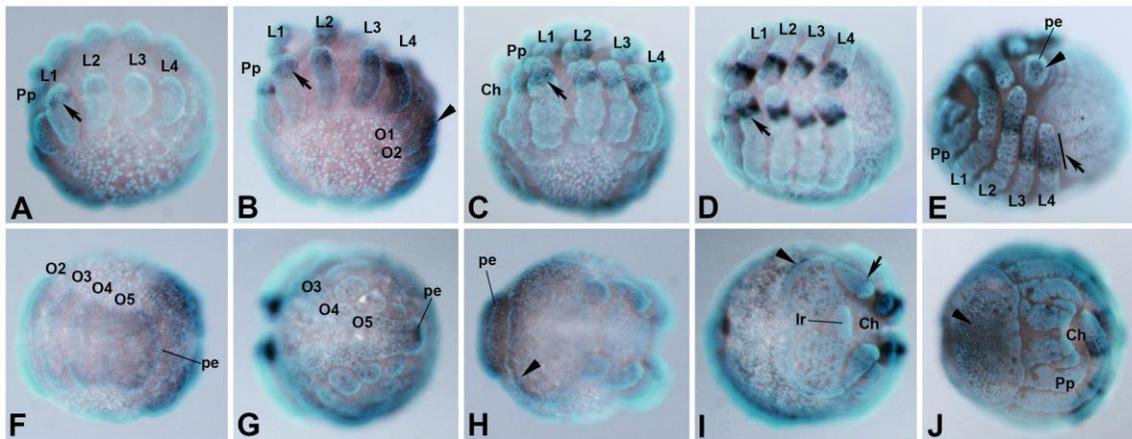


Figure 61: Expression of *c12* in *P. tepidariorum*

A, F and H show the same embryo during early limb bud elongation in lateral view, ventral view of the opisthosoma and ventral view of the head. B: Embryo during late limb bud elongation, lateral view. C: embryo during inversion, lateral view. D, G and I show the same embryo during early dorsal closure in ventral view of the walking legs, opisthosoma and head. E and J: Embryo during ventral closure in Ventro-lateral and frontal view. Detailed explanation of the expression pattern is found in the text. Anterior is pointing to the left for A-E, H and J. F and G posterior is on the right and in I posterior points to the left. Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

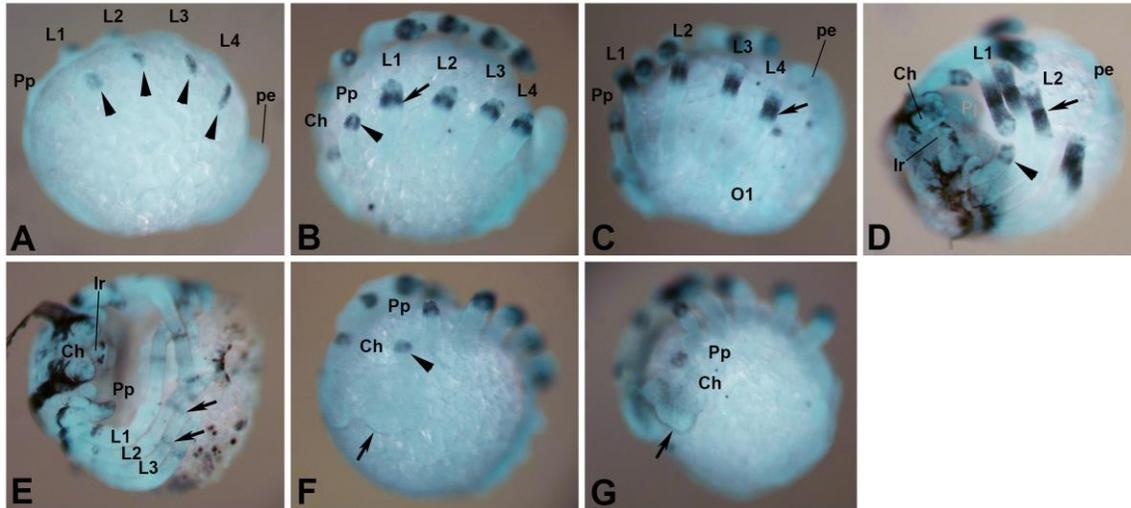


Figure 62: Expression of *cll2* in *P. phalangioides*

Expression of *cll2* in prosomal appendages of *P. phalangioides* during limb bud development (A), limb bud elongation (B) and dorsal closure (C), all in lateral view. D: Embryo during ventral closure shown in anterior-ventral view. E: Embryo after ventral closure also in ventro-frontal view. F: Embryo during limb bud elongation shown in frontal aspect. G: Embryo during dorsal closure shown in dorsal-anterior view. Detailed description of *cll2* expression in the text. All embryos are oriented with anterior to the left. Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1: first opisthosomal segment, pe: posterior end.

6.6.3 Expression of clawless in prosomal appendages

In the previous section it was already stated that the expression in the early walking legs is different for the two intraspecific paralogs but similar for interspecific orthologs. In the following the expression in the prosomal appendages will be described and compared in greater detail. As observed for the walking legs the expression of *Pt-cll1* is present also as a strong distal domain in the more anterior prosomal appendages during limb bud elongation (bars and arrow, Figure 63, A, E and I). This domain divides during early inversion in the walking legs and pedipalps (bar and arrow, Figure 63, F and J), but not in the chelicera, where it only moves more proximal (bar, Figure 63, B). In the pedipalps the more proximal portion of the dividing domain quickly vanishes during inversion and only one distal *Pt-cll1* ring is present in this appendage during dorsal and ventral closure (arrow, Figure 63, G and H). In the legs, however, the division of the distal domain gives rise to two distal rings in the area of the joints forming between tibia and metatarsus and metatarsus and tarsus (arrows, Figure 63, K and L). In the chelicera *Pt-cll1* is expressed in a medial ring-like domain (arrowhead, Figure 63, C). In *P.*

phalangioides the chelicera specific expression in the medial part starts already during limb bud elongation, when *Pp-cll1* expression is still in the distal domain of the pedipalps and walking legs. Later during ventral closure *Pp-cll1* expression is different for all three prosomal appendage types compared to *Pt-cll1*. In the pedipalps it remains absent and in the chelicera and walking legs the expression is restricted to a small spot ventrally on the distal tip of these appendages.

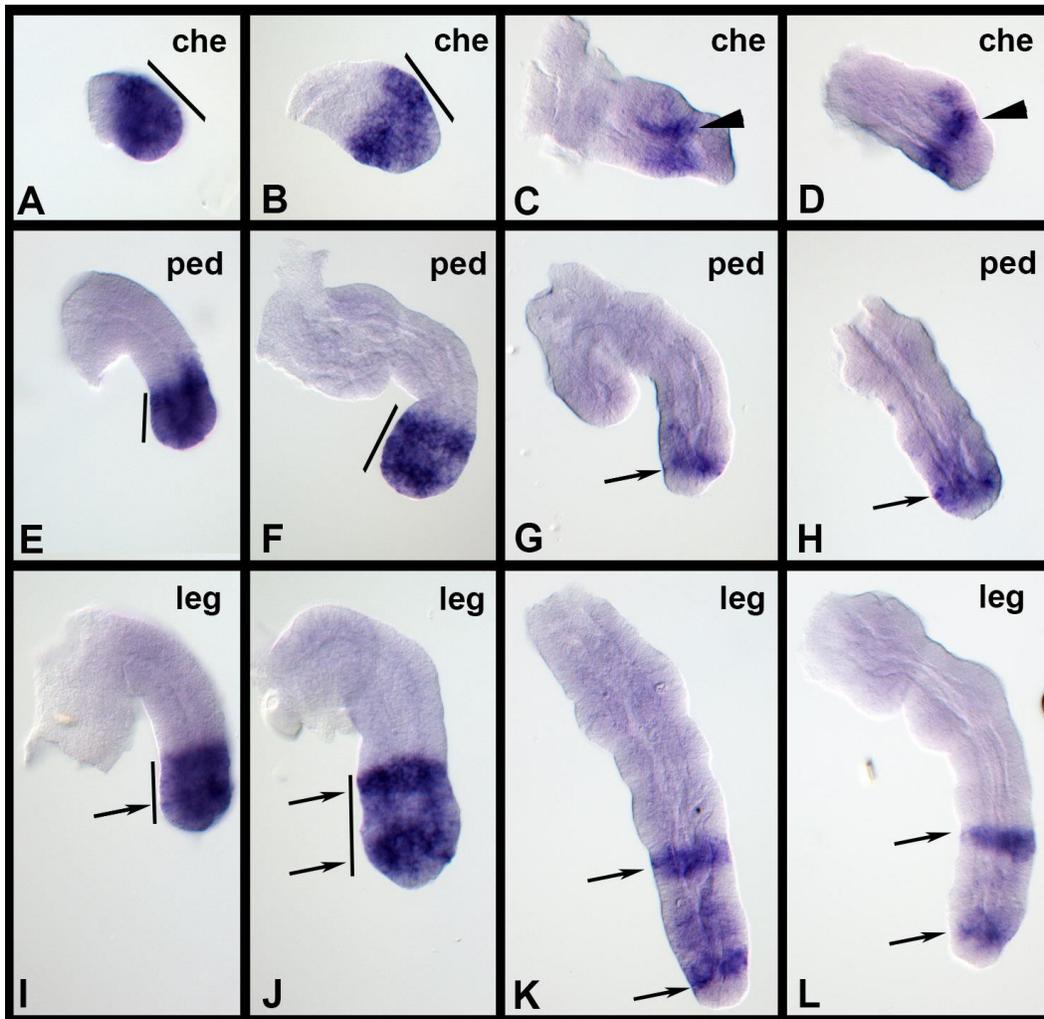


Figure 63: Expression of *cll1* in prosomal appendages of *P. tepidariorum*

A-D: Expression *cll1* in the chelicera of *P. tepidariorum*. E-H: *Pp-cll1* expression pattern in the pedipalps. I-J: Expression of *Pt-cll1* in the walking legs. A, E and I shows prosomal appendages during limb bud elongation. B, F and J show prosomal appendages during early inversion of the embryo. C, G and K depict expression in prosomal appendages during dorsal closure. D, H and L show prosomal appendages during ventral closure. For an explanation of the bars and arrows, please see text. che: chelicera. ped: pedipalp. leg: walking leg.

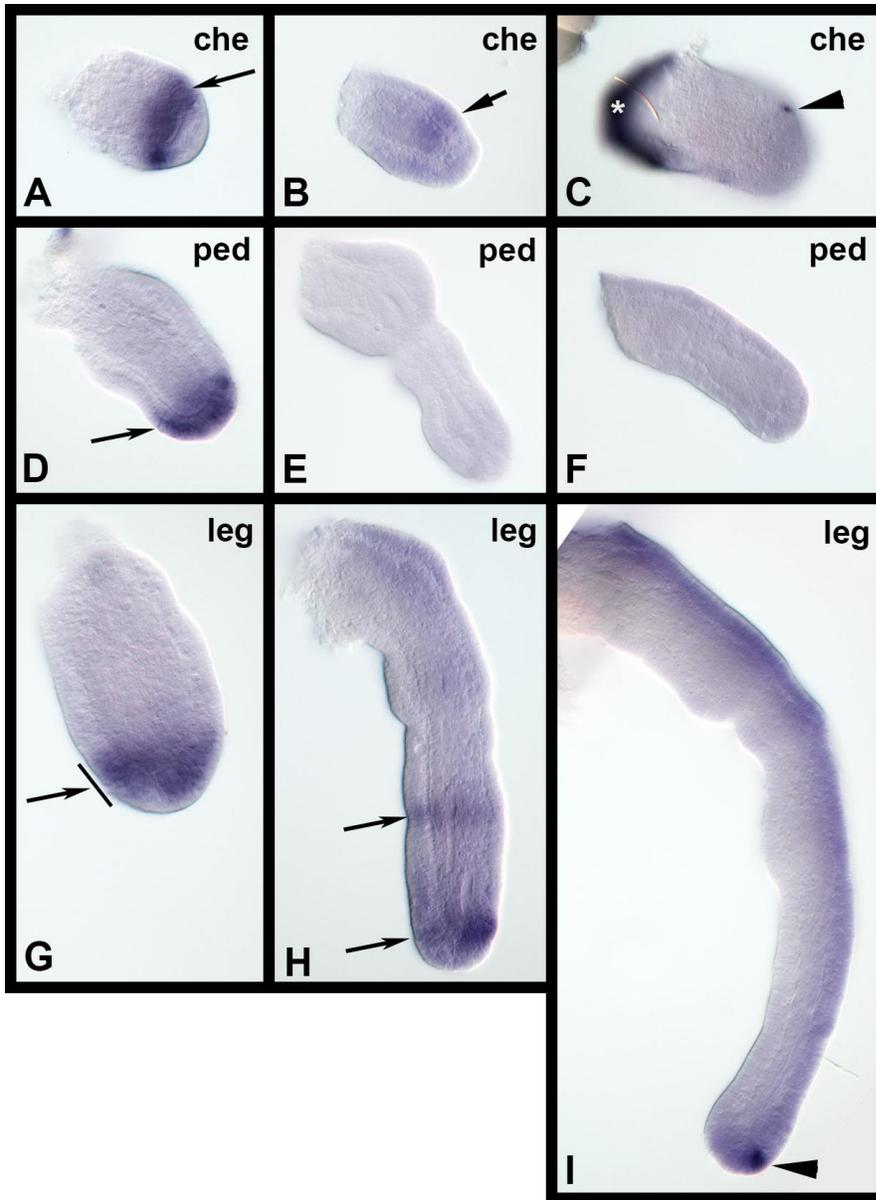


Figure 64: Expression of *cll1* in prosomal appendages of *P. phalangioides*

A-C: Expression *cll1* in the chelicera of *P. phalangioides* after *in situ* hybridization. D-F: Preparations of pedipalps after *in situ* hybridization with *Pp-cll1*. G-I: *Pp-cll1* expression in the walking legs. A, D and G show prosomal appendages during limb bud elongation. B, E and H depict these appendages during inversion of the embryo. C, F and I present the stained prosomal appendages during ventral closure stages. Asterisk in picture C marks unspecific staining of the forming cuticle at the chelicera. Detailed explanation of the expression pattern is found in the text. che: chelicera. ped: pedipalp. leg: walking leg.

As stated before the expression of *cll2* is more similar also in late stages of the developing walking legs in the two spiders, which it not true for the more anterior appendages. Whereas it is expressed in the entire chelicera of *P. tepidariorum* during

limb bud elongation (Figure 65, A), it is at comparable stages expressed in the distal part of this appendage in *P. phalangioides* (arrow, Figure 66, A) and thus more similar to *Pt-cll1*. During dorsal closure *cll2* is expressed in a medial domain similarly to its paralog *cll1* in *P. tepidariorum* and *P. phalangioides* (arrow, Figure 65 and Figure 66, B). Later this expression broadens in the chelicera of *P. phalangioides* embryos during ventral closure (arrow, Figure 66, C), but shows a ubiquitous expression in the chelicera of *P. tepidariorum* embryos of comparable stages (Figure 65, C).

cll2 is expressed in the distal domain of the pedipalps of *P. phalangioides* during limb bud elongation (bar, Figure 66 D) and stays there until dorsal and ventral closure of the embryo (bar, Figure 66, E and F). During these stages the distal domain of *Pp-cll2* in the pedipalps differentiates slightly (arrow). The expression of *cll2* in the pedipalps of *P. tepidariorum* is a bit more differentiated already during limb bud elongation stages but similar beginning with inversion (arrow, Figure 65, D and E). During ventral closure the same ubiquitous expression as in the chelicera is present. In the walking of both spiders *cll2* is expressed in a distal ring throughout entire prosomal limb bud development, which broadens during later stages (arrow, Figure 65 and Figure 66, G-I). This ring is present in the joint between the metatarsus and tarsus segment. Only subtle differences are present for the walking leg expression of *cll2*: in *P. tepidariorum* weak ubiquitous expression in the entire leg is present during prosomal limb bud elongation and ventral closure stages; in *P. phalangioides* strongest expression is as stated found in the most distal joint but also in neighbouring tissues diffusing proximally and distally from this joint (bar, Figure 66, G-I).

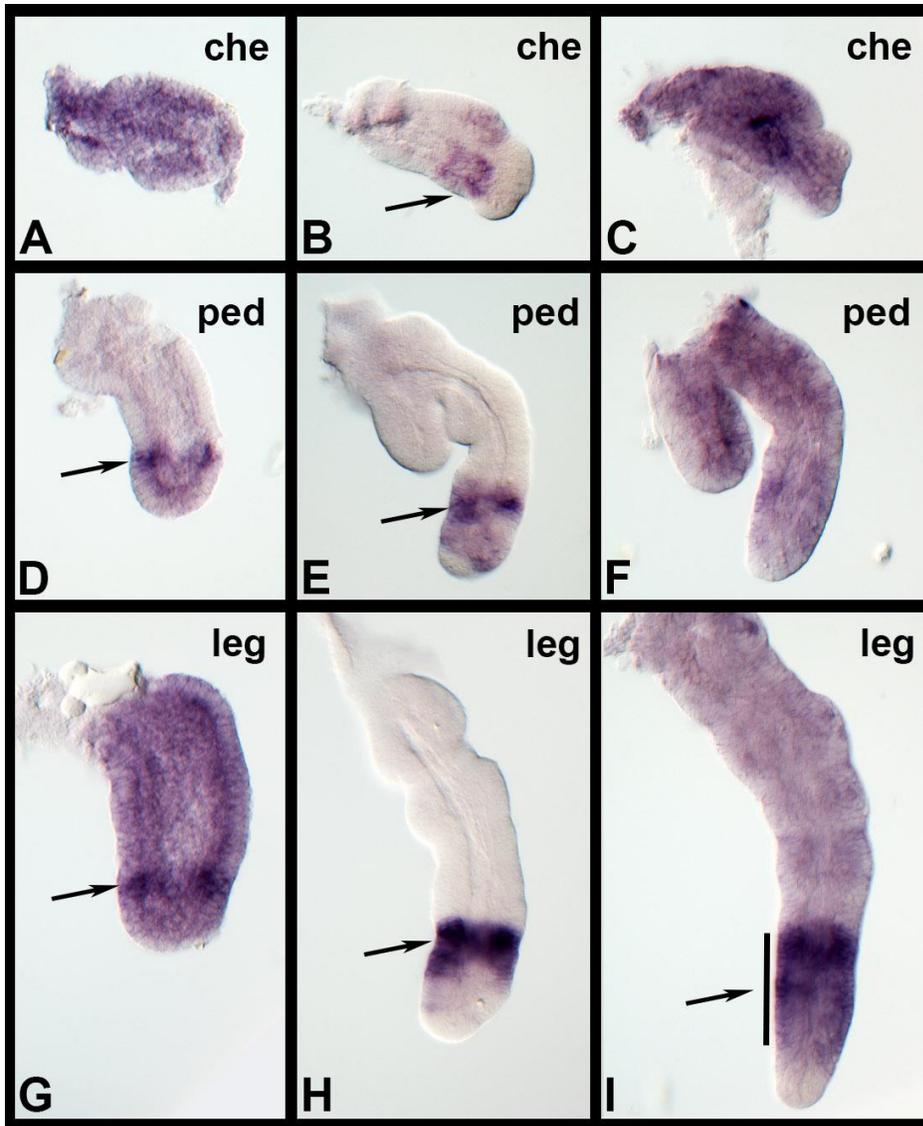


Figure 65: Expression of *cll2* in prosomal appendages of *P. tepidariorum*

A-C: Expression *cll2* in the chelicera of *P. tepidariorum*. D-F: Preparations of pedipalps after *in situ* hybridization with *Pt-cll2*. G-I: *Pt-cll2* expression in the walking legs. A, D and G show prosomal appendages during limb bud elongation. B, E and H depict these appendages during inversion of the embryo. C, F and I present the stained prosomal appendages during ventral closure stages. Detailed explanation of the expression pattern is found in the text. che: chelicera. ped: pedipalp. leg: walking leg.

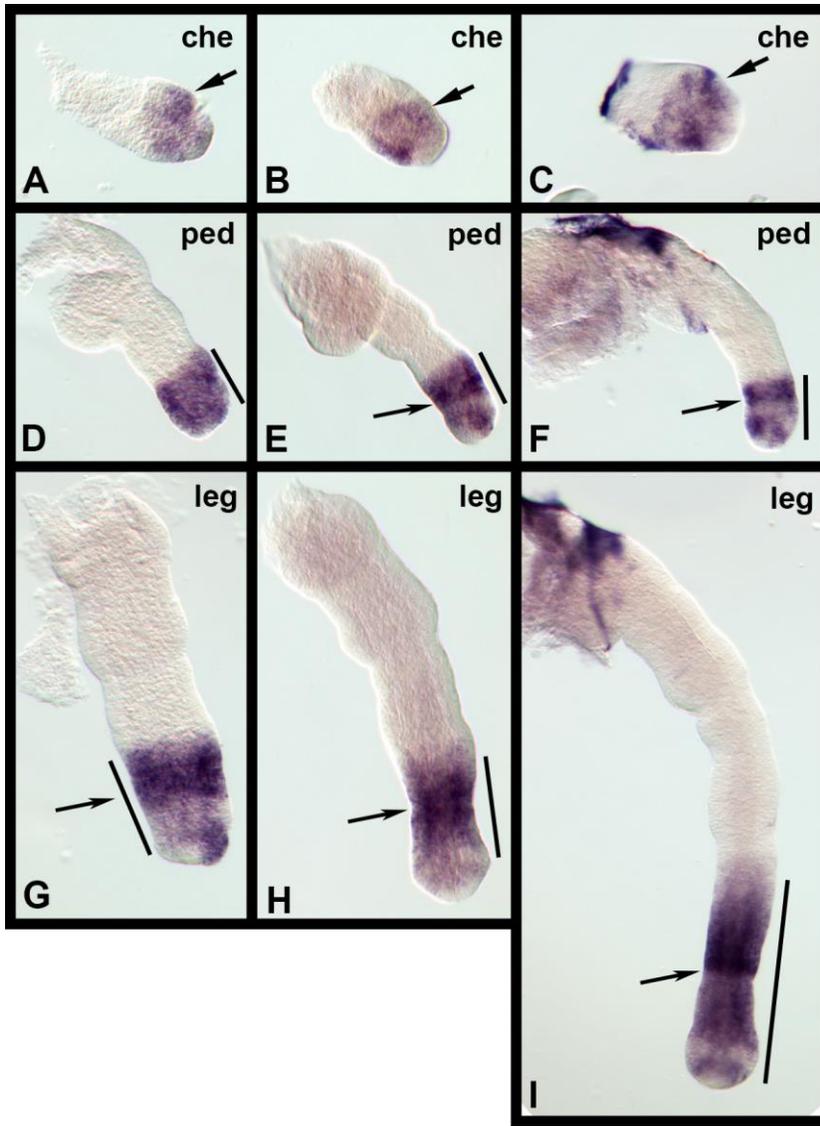


Figure 66: Expression of *cll2* in prosomal appendages of *P. phalangioides*

A-C: Expression *cll2* in the chelicera of *P. phalangioides*. D-F: Preparations of pedipalps after *in situ* hybridization with *Pp-cll2*. G-I: *Pp-cll2* expression in the walking legs. A, D and G show prosomal appendages during limb bud elongation. B, E and H depict these appendages during inversion of the embryo. C, F and I present the stained prosomal appendages during ventral closure stages. Detailed explanation of the expression pattern is found in the text. che: chelicera. ped: pedipalp. leg: walking leg.

6.6.4 Function of *cll1* in *P. tepidariorum*

To study the function of *cll1* in the walking leg of *P. tepidariorum* parental RNAi was performed using the cloned gene fragment (1161 Bp). The exact procedure of injection is described in the chapter Materials and Methods. Observation of freshly hatched nymphs

from the cocoons produced by injected females revealed two different phenotypes after *Pt-cll1* knockdown (Overview in comparison to control, Figure 68).

In weak phenotypes (Figure 68, C and D) the joint between the metatarsus and tarsus segment is missing, which leads to a fusion of these distal leg segments. For this weak phenotype two different variations were found: (1) lateral outgrowths on the position of the presumptive joint area (arrow, Figure 68, C and red arrow, Figure 69, B and E), comparable to the phenotypes after *Pt-dac2* knockdown, which are missing the joint separating the patella from the tibia. These outgrowths are probably caused by missing cues for the dividing cells in this area. (2) bending of reminiscent tarsal tissue resulting in a foot-like shaped distal leg part (arrowhead, Figure 68, C and D and red arrow, Figure 69, C and F), probably caused by uncontrolled growth of cells on one side of the leg usually forming the joint. In both variations a disturbed bristle pattern in the area where normally the joint should be was observed. This suggests that the joints are needed to guide normal development of the tissue in the segments and missing the joint removes these cues.

In strong phenotypes (Figure 68, E and F and Figure 70, B and D) the complete tarsus segment and its neighbouring joints are entirely lost. In more severe cases of this strong phenotype the tissue distal to the metatarsus including all bristles, sensory organs and the claw, was absent completely (arrowhead, Figure 70, E). In all other phenotypic variations the claw located on the most distal tip of the walking legs was unaltered (indicated by Cl plus arrow, Figure 69 E and F and Figure 70 D) and develops the typical hook and claw shape needed for holding on the silk thread (Foelix, 1996). In accordance to *cll1* expression in the pedipalps, defects in the distal segments of this appendage were observed as well. Resembling the weak leg phenotypes, affected pedipalps show fusion of the tibia and tarsus segment and disrupted joint formation causing lateral outgrowths (black arrow Figure 71, C and D). Like in the legs the trident shaped claw of the pedipalpus is not altered in these phenotypes (red arrow Figure 71, D). Penetrance of the *Pt-cll1* knockdown was analysed for the different cocoons produced by females after dsRNA injections (left panel, Figure 67) and compared to control females, which were equally treated with injection buffer (right panel, Figure 67). The amount and onset of weak and strong phenotypes was very diverse for the different females (ranging from

only one affected cocoon to strong effect in all of the observed cocoons), but did not vary in accordance to the different injection numbers and mating time points (data not shown). The characteristics of the observed phenotypes were, however, comparable. Thus the results of all females were pooled and included in the statistic. This combination shows that weak (green) and strong (light blue) phenotypes are present in all first five cocoons produced by females treated with *Pt-cll1* dsRNA. The amount of these phenotypes is lowest in the first cocoon (around 20% Figure 67, C1) and highest in the third cocoon (more than 40%, Figure 67, C3). The number of normally developed nymphs (Wt, light violet, Figure 67) is lower compared to the control in each of the observed cocoons. In contrast the amount of embryos which were not able to hatch from the embryonic egg-shell (not hatched, dark violet, Figure 67) lies within the usual range of 20-30% and is consistent with previous findings for both knockdown and control offspring.

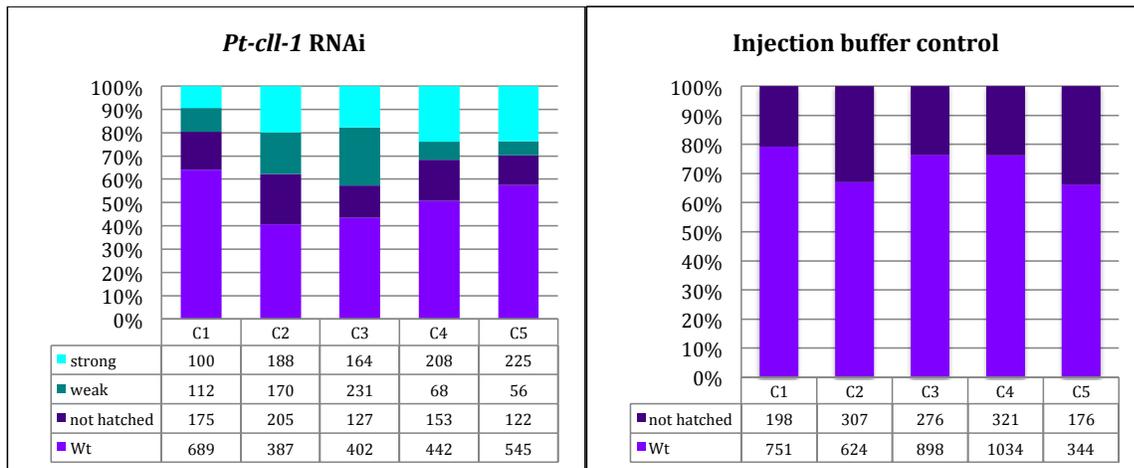


Figure 67: Overview of phenotypes in the different cocoon after *cll1* RNAi and control injection

The graph on the left gives an overview of the phenotypes in freshly hatched and third instar nymphs after knockdown of *Pt-cll1* in the first five cocoons produced by females after injection. The graph on the right shows the first five cocoons produced by females after control injections with injection buffer. In the control the amount of unhatched phenotypes ranges between 20-30%, thus usually at least 70% of the offspring in each cocoon developed without any visible abnormalities (indicated by Wt: wild-type appearance marked in light violet). After *cll1* the highest amount of normally developed nymphs is found in the first cocoon with approximately 60%. The other cocoons, especially the second and third cocoon, are more severely affected by the RNAi.

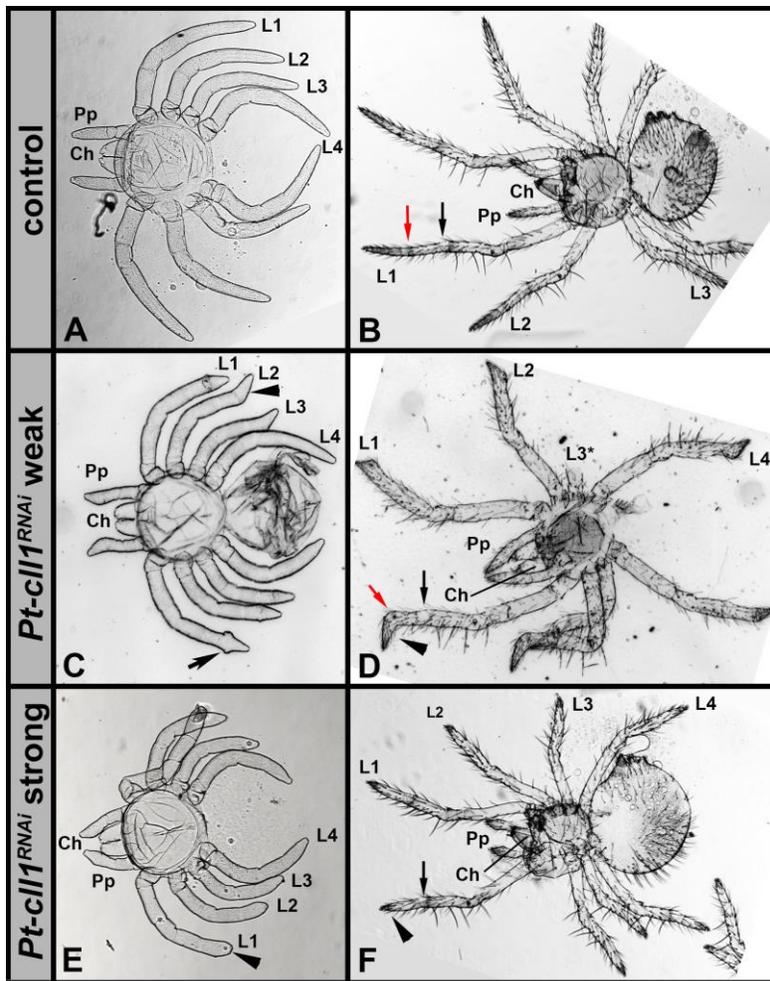


Figure 68: Overview of phenotype in nymphs of *P. tepidariorum* after *cll1* RNAi

Depicted are cuticle preparations with Hoyer's medium of different nymphal stages of *P. tepidariorum*. A: Prosoma of a freshly hatched nymph from a control female treated with injection buffer according to the parental RNAi procedure. B: Third instar nymph from a female injected in the same way as in the *cll1* RNAi with injection buffer. The black arrow points to the fully developed joint between the tibia and metatarsus. The red arrow marks the most distal joint between the metatarsus and tarsus segment. C: Prosoma of a freshly hatched nymph after parental with *Pt-cll1*, showing the weak *cll1* phenotype in the pedipalps and walking legs. The arrow indicates the malformed joint between the metatarsus and tarsus leading to lateral outgrow at the presumptive side of the joint. The arrowhead points to another variation of a weak phenotype where the disrupted distal joint causes bending of the reminiscent tarsus tissue. D: Cuticle preparation of a third instar nymph after *cll1* RNAi with weak phenotype in all walking legs. The red arrow indicates the disrupted joint, which is not properly formed between the metatarsus and tarsus. The black arrow point to the intact joint of the tibia and metatarsus. The arrowhead points to the bending of presumptive tarsal tissue after fusion. E: Prosoma of a freshly hatched nymph after parental with *Pt-cll1*, showing the strong *cll1* phenotype in the walking legs. The arrowhead marks the tip of the leg missing the tarsal segment. F: Cuticle preparation of a third instar nymph after *cll1* RNAi with strong phenotype in all walking legs. The black arrow point to the intact joint of the tibia and metatarsus. The arrowhead point to the intact claw although missing of the tarsus segment. Freshly hatched nymphs are shown in ventral view and third instar nymphs in ventro-lateral view, but all nymphs are oriented with anterior to the left. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4.

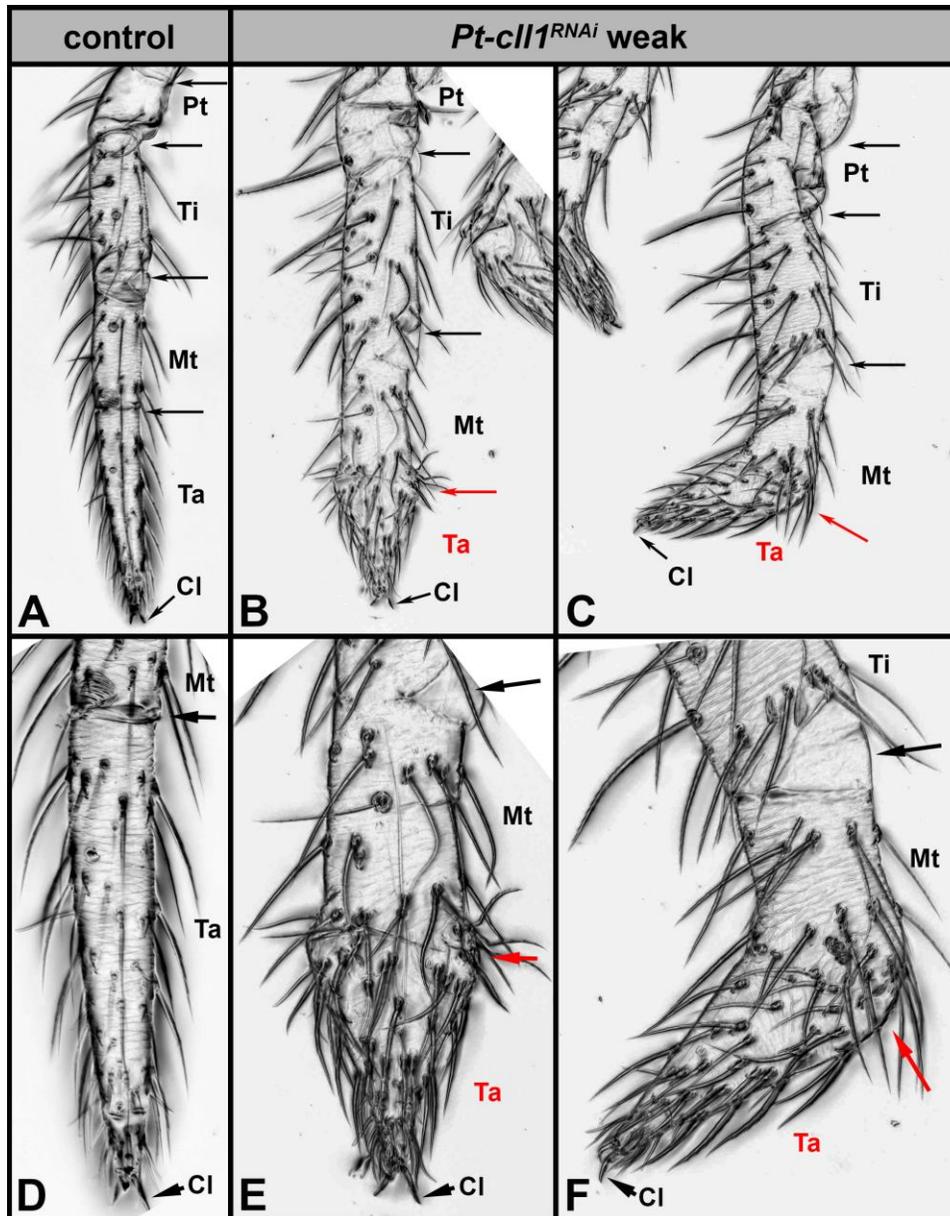


Figure 69: Weak *Pt-cll1* RNAi phenotype in the walking legs

Depicted are close ups from cuticle preparation of walking legs from the third nymphal stage of *P. tepidariorum*. A: Walking leg of third instar nymph offspring from control females injected with injection buffer instead of dsRNA. B and C: Walking legs of third instar nymph offspring after *cll1* RNAi showing two different forms of weak phenotypes, indicated by fusion of the metatarsus and tarsus segment. B: This legs shows lateral outgrowth in the presumptive area of the missing joint. C: Bending of reminiscent tarsal tissue after distal segment fusion. D: Close up of the leg shown in A. E: Close up of the leg shown in B. F: Close up of the leg shown in C. The red arrow indicates the missing joint between the metatarsus and tarsus segment and malformed tissue causing outgrowths in this area. This causes fusion of the tarsus and metatarus segment, the remaining tissue, which would usually belong to the tarsus, is indicated by Ta written in red. The black arrows in all pictures point to properly shaped and formed joints between the femur and patella, patella and tibia, tibia and metatarsus and metatarsus and tarsus. All leg preparations are oriented with proximal at the top and distal to the bottom. Cl: Claw, Pt: patella, Ti: tibia, Mt: metatarsus, Ta: tarsus.

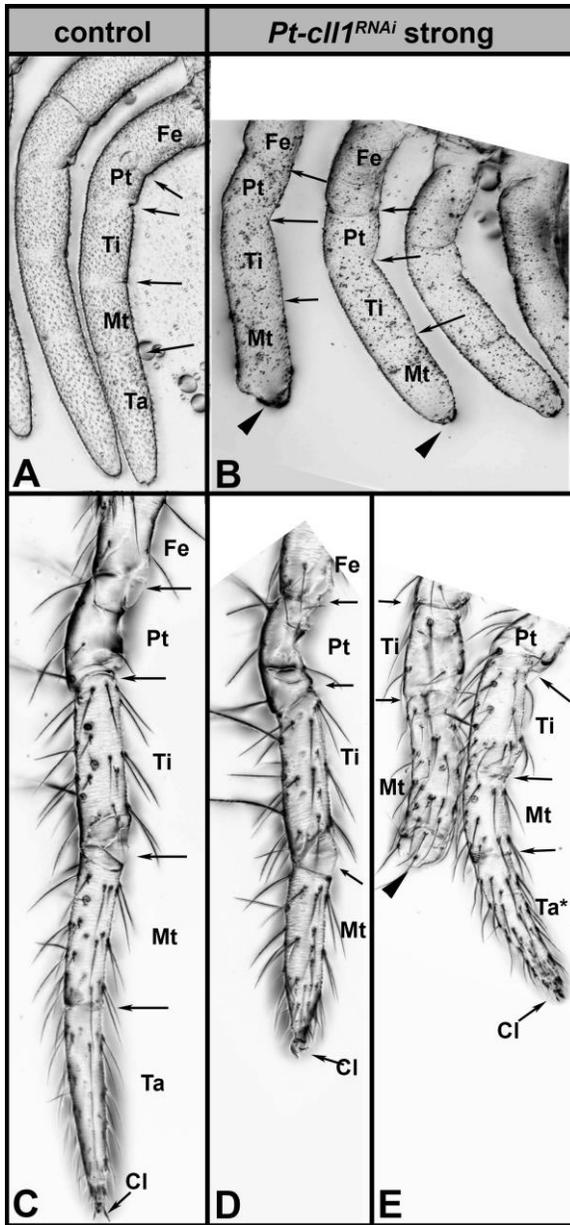


Figure 70: Strong *Pt-c11* RNAi phenotype in the walking legs

Depicted are close ups from cuticle preparations of walking legs of different nymphal stages of *P. tepidariorum*. A: Walking leg of second instar nymph offspring from control females injected with injection buffer instead of dsRNA. B: Walking legs of second instar nymph offspring after *c11* RNAi with strong phenotypes, indicated by absence of the tarsus and disrupted distal tip. C: Walking leg of third instar nymph offspring from a control female. D and E: Walking legs of third instar nymph offspring after *c11* RNAi with different variations of the strong phenotype. D: The walking leg is missing the entire tarsal segment, whereas the claw is still present and properly shaped. E: Depicts two variants of walking legs observed in a specimen with strong *c11* knockdown phenotype; on the right normally developed leg with all segments; on the left affected leg missing the entire distal tissue distally of the metatarsus. The black arrows in all pictures point to properly shaped and formed joints between the different leg segments. The arrowhead marks disrupted distal-most leg structures in most severe phenotypes. All leg preparations are oriented with proximal at the top and distal to the bottom. Cl: Claw, Pt: patella, Ti: tibia, Mt: metatarsus, Ta: tarsus.

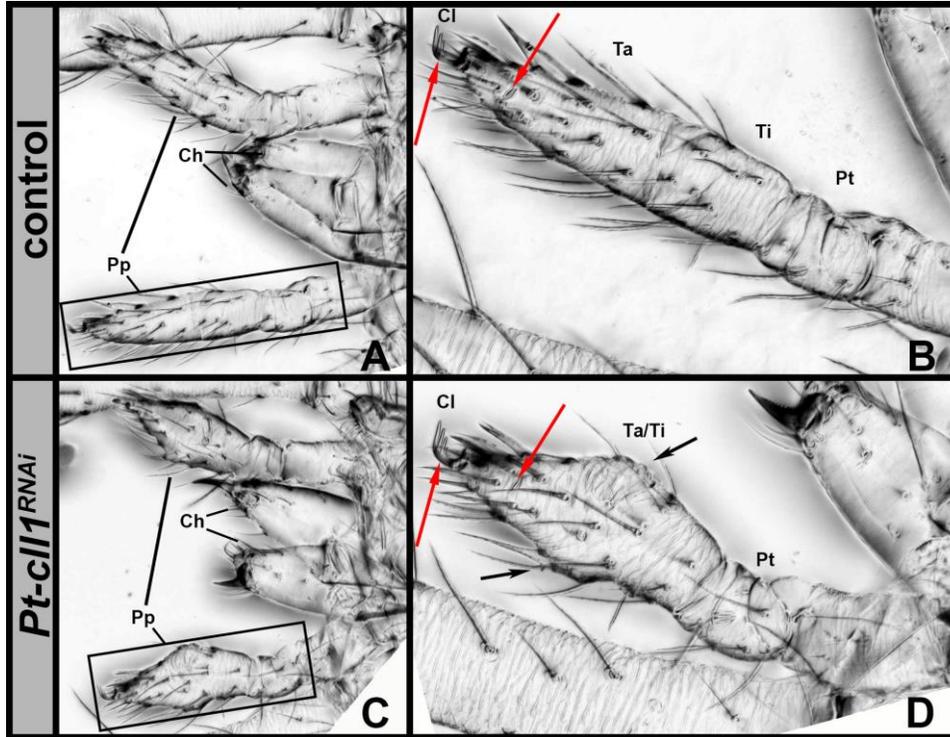


Figure 71: *Pt-cll1* RNAi effect on nymphal pedipalps

Depicted are close ups from cuticle preparation of pedipalps from the third nymphal stage of *P. tepidariorum*. A: Chelicera and pedipalps of third instar nymph offspring from control females injected with injection buffer instead of dsRNA. The rectangle marks the area including one pedipalp magnified in B. B: Magnification of the pedipalp shown in A. The segments are marked by the according abbreviations. The red arrows point to the lyriform organ and the claw on the distal tip of the pedipalps. C: Chelicera and pedipalps of third instar nymph offspring from females injected *cll1* dsRNA. The rectangle marks the area including one pedipalp magnified in D. D: Magnification of the pedipalp shown in C. The segments are marked by the according abbreviations. The black arrow points to the malformed joint between the tibia and tarsus segment. The red arrows point to the unaltered lyriform organ and the properly shaped claw with tree hooks on the distal tip of the pedipalps. Ch: chelicera, Pp: pedipalps, Cl: Claw, Pt: patella, Ti: tibia, Ta: tarsus, Ta/Ti: fusion of the leg segments tibia and tarsus.

6.7 BarH in *P. tepidariorum* and *P. phalangioides*

BarH transcription factors are members of the Bar class of homeobox genes and part of the highly conserved group of Q50-homeoproteins. Homologs of the *BarH* genes were first identified in *D. melanogaster* (Kojima et al., 1991) and also found in a number of vertebrates (reviewed in (Reig et al., 2007)). Homeodomain transcription factors can bind to a variety of downstream targets, whereas specificity is mediated by post-translational modifications and co-factor interactions depending on the cellular environment. This fine-tuning of transcription factor specificity serves as a versatile toolkit for functional evolution. Accordingly, aside from their conserved function in cell fate determination during retinal neurogenesis in *D. melanogaster* and vertebrates (Kojima et al., 1991, reviewed in Reig et al., 2007), *BarH* genes show various unique functions in distinct species. These include tarsus segmentation in distal leg of *D. melanogaster* (Kojima et al., 2000; Kozu et al., 2006), cell survival in either ear or mesoderm development in the mouse or *Xenopus laevis* respectively, as well as cell migration in the mouse cerebellum.

6.7.1 Phylogenetic analysis of *Bar* in *P. tepidariorum* and *P. phalangioides*

Genuine *BarH* homologs were searched in the transcriptomes available for the two spider species observed in this study and, in contrast to the other genes used in this study, three transcripts with high similarity to *D. melanogaster BarH* were found. The large phylogenetic study (Supplemental Figure 17 and Supplemental Figure 18) includes all candidates found in the spider transcriptomes as well as their blast hits from the non-redundant protein sequence database of chosen taxa (for details see methods section). It shows, that one of the *P. phalangioides* candidates rather clusters into the sister group of *BarH-like* proteins branching off the true *Bar* clade more basally (depicted in dark red, see Supplemental Figure 17). The true *Bar* group (depicted in red) includes all sequences from vertebrates, insects and chelicerates. These homologs are further subdivided forming a monophyletic group of the vertebrate paralogs, which cluster together forming two distinct vertebrate *Bar* groups (*BarH-like-1* and *BarH-like-2*) and a polytomy for the other proteins. In this polytomy the insect homologs form a cluster, although missing the

T. c homolog. In the insect cluster the *Dm-Bar* paralogs are closely related. Next to the insect cluster two distinct clusters for the spider *Bar* homologs were found, including *P. tepidariorum*, *P. phalangioides* and *S. mimosarum*. The grouping of the *D. melanogaster*, vertebrate and spider paralogs into taxa specific groups indicates that a *Bar* gene duplication evolved several times independently, but before the split of the Entelegynae and Haplogynae spider groups.

In addition many predicted sequences were found e.g. from the horseshoe crab *L. polyphemus*, the scorpion *M. occidentalis*, and a sequence from *T. castaneum* that all proved to be difficult to place in the phylogenetic tree and usually formed separate but weakly supported clades. Together with the insect polytomy and the difficulties to properly resolve the insect homologs this strongly suggests that a more specific phylogenetic analysis of the *BarH* homologs with additional arthropod species is necessary to clarify their phylogenetic relationships on a greater level.

However, *Pp-comp120407* clusters together with the transcript Locus_1_Transcript_157093/166847 of *P. tepidariorum* (in the following referred to as *BarH2*) and *Pp-comp116578* with Locus_3013, Locus_3015 and Locus_3016 (referred to as *BarH1*), which were named according to sequence similarity to *D. melanogaster* paralogs. The existence of three *BarH1* transcripts in *P. tepidariorum* and thus a possible species-specific duplication was examined in more detail. RACE PCR was performed to rule out that these transcripts might represent two isoforms of one *BarH1* paralog, wrongly identified as three separate genes during *de novo* assembly of the *P. tepidariorum* transcriptome (data not shown). Unfortunately this did not help in understanding the origins of these sequences. As mentioned in the Introduction current work includes genome assembly of *P. tepidariorum* (unpublished data, courtesy of Alistair McGregor, co-ordinator of the *P. tepidariorum* sequencing consortium), thus the corresponding gene was searched in a preliminary version of this genome. The full-length protein found in the genome revealed that *Locus_3013*, *Locus_3015* and *Locus_3016* represent parts of two isoforms of one *BarH* homolog (Augustus prediction *Pt-aug3.g8250.t1* and *Pt-aug3.g8250.t2*). One of these augustus predictions was included into the large phylogenetic analysis subsequently. There it resolves with the three previously found transcripts, validating the assumption that only one *Pt-BarH1* paralog

is present in *P. tepidariorum*. The gene fragment cloned from the transcriptome sequence *Locus_3013* was used for the expressional and functional analysis.

6.7.2 Expression of *BarH* homologs in *P. tepidariorum*

Expressional analysis of the two *BarH* paralogs in *P. tepidariorum* showed that they are similarly expressed in the prosomal appendages and parts of the head, but that *BarH1* has in addition unique expression details during early embryonic development and in parts of the head and opisthosoma. The expression in the prosomal appendages starts with their formation on the germ band (Figure 72, E and F, data not shown for *Pt-BarH2*). During limb bud elongation both paralogs are expressed in a medial and distal domain in the chelicera and remains unchanged throughout appendage development (arrowhead, Figure 74, D-F and Figure 73, A-C). During limb bud elongation in the walking legs and pedipalps both paralogs are expressed as a distal spot (arrow, Figure 72 G and Figure 74, A). This expression in the distal part of the legs further changes during development. At the beginning of inversion an additional spot of expression appears (arrow, Figure 72, H and Figure 74, B). Later during dorsal closure the expression pattern in the distal part changes to an intensive expression domain in the most distal part (arrow, Figure 72, I and Figure 74, C) and a diffuse spot-like pattern proximal to this distal domain. Also similar for both *Bar* homologs is the strong expression in the gnathendites of the pedipalps and the corresponding endite region at the base of the walking legs (arrowhead, Figure 72, G-J and Figure 74, A-C). Although the expression pattern for both *P. tepidariorum* paralogs is strongly conserved in the prosomal appendages, the expression in the head shows some differences between *Pt-BarH1* and *Pt-BarH2*. Both paralogs are expressed in the neurogenic ectoderm posterior to the anterior furrow in the head and later during inversion also in the labrum (Figure 73 and Figure 74, arrow E and F). After the non-neurogenic ectoderm has overgrown the neurogenic ectoderm the most anterior domain in the neurogenic ectoderm is hardly visible for *Pt-BarH1* (Figure 73, anterior arrow C) and has fully disappeared for *Pt-BarH2*. *Pt-BarH1* is expressed adjacent to the lateral furrow and as a stripe in the neurogenic ectoderm anterior to the cheliceral segment at prosomal limb bud developmental stages. These expression details

are not present in the *Pt-BarH2* pattern and start already during early germ band elongation stages (arrowhead, Figure 72, F). In addition, *Pt-BarH1* shows several other unique expression details that differentiate it from the expression pattern of *Pt-BarH2*. Starting with opisthosomal limb bud development it is strongly expressed in the limb bud developing on the second opisthosomal segment (red arrow, Figure 72, G and I), which will give rise to the book lung after invagination during further development.

Pt-BarH1 is also expressed very early as a ring surrounding the outer rim of the germ disc (Figure 72, A). This expression remains in this tissue and then opens on one side (asterisk, Figure 72 B) during the break of radial symmetry. During early germ band stages the ring of expression is therefore present as a stripe (arrow, Figure 72, C), which is moving posteriorly (arrow, Figure 72, D). However, no segmental stripe of *Pt-BarH1* expression can be detected after germ band elongation. The two paralogs found in the transcriptomes of *P. phalangioides* were cloned (Jerome Janssen) but *in situ* hybridization of whole mount embryos resulted in staining of low quality, which is probably attributable to problems during probes synthesis and thus needs to be repeated.

6.7.3 Function of *BarH1* in *P. tepidariorum*

Parental RNAi was performed with the cloned *BarH* paralog fragments of *Pt-BarH1* (813 Bp) and *Pt-BarH2* (726 Bp). 4µg ds RNA was injected five times in adult females spiders, which were mated after the third injection. The knockdown of *Pt-BarH2* did not lead to any obvious phenotype (data not shown). Knockdown of *Pt-BarH1* resulted in a large number of embryos, which were unable to hatch from the embryonic envelope (depicted in green, Figure 75) or died during embryogenesis (depicted in dark violet, Figure 75). This was observed already in the first cocoon laid by RNAi females after injection with dsRNA and did not change for at least seven cocoons. The most prominent phenotype was fully developed embryos that were unable to hatch, especially in cocoon 4 (over 90 %), cocoon 5 (over 70%) and cocoon 6 (more than 80% in Figure 75).

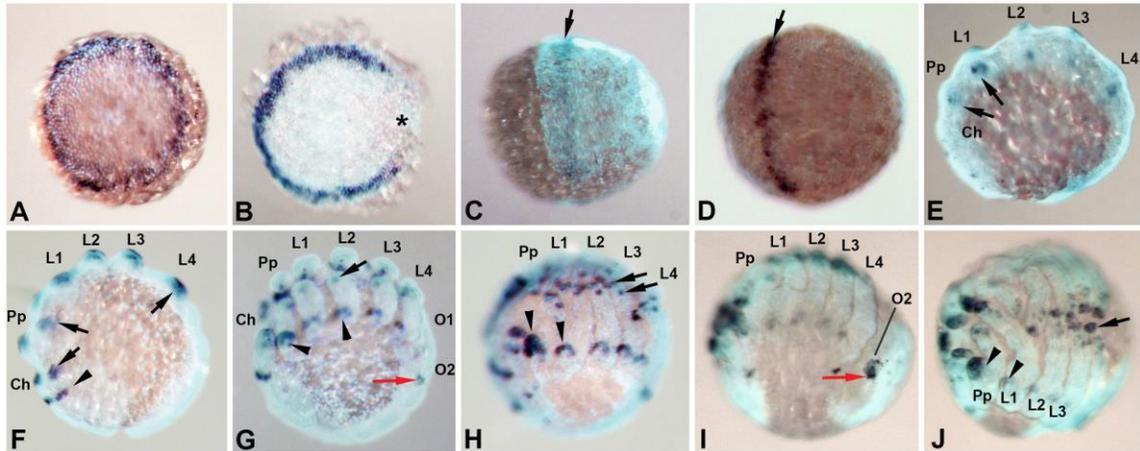


Figure 72: Expression of *BarH1* throughout embryonic development of *P. tepidariorum*

A: Germ disc surrounded with expression of *BarH1*. **B:** Opening of germ disc. The asterisk marks the area of opening. **C:** Embryo during early germ band formation. **D:** Embryo during late germ band formation. **E:** Fully elongated germ band embryo, lateral view. The arrows point to distal *BarH1* expression in the areas where prosomal limb buds will form. **F:** Embryo during inset of limb bud development, lateral view. The arrows point to *BarH1* expression distally in the limb buds. The arrowhead indicates *BarH1* bilateral stripe anterior to the chelicera. **G-I:** Lateral view of embryos during limb bud elongation (**G**), inversion (**H**) and dorsal closure (**I**). **J:** Ventral view of the shown in **I**. The arrowheads in **G-J** point to *BarH1* expression in the gnathendite and the corresponding tissue at the base of the legs. The back arrows point to distal *BarH1* spots in the prosomal appendages. The red arrow marks *BarH1* expression in in the second opisthosomal limb bud. All embryos are oriented with anterior to the left, except for the germ disc stages which are oriented in top view of the germ disc. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O2: Opisthosomal segments 1-2.

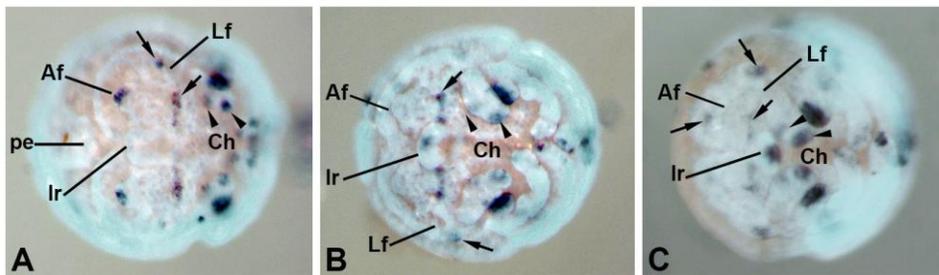


Figure 73: Expression of *BarH1* in the head of *P. tepidariorum*

A: Frontal view of the head during limb bud elongation with *Pt-BarH1* expression in the chelicera (arrowheads), the lateral furrow and two stripes anterior of the chelicera segment (arrow) as well as in the anterior furrow. **B:** Head of the embryo during inversion of *P. tepidariorum*, shown in frontal aspect. The arrowheads point to expression in the chelicera, the arrowheads indicate segmental *BarH1* stripes anterior to the chelicera and a domain lateral to the lateral furrow. **C:** Frontal view of an embryo during dorsal closure. Arrowheads as described before, arrows indicate same expression as in **B** and additional expression anterior to the anterior furrow. All embryos are oriented with anterior to the left. Ir: Labrum, Lf: Lateral furrow, Af: Anterior furrow, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

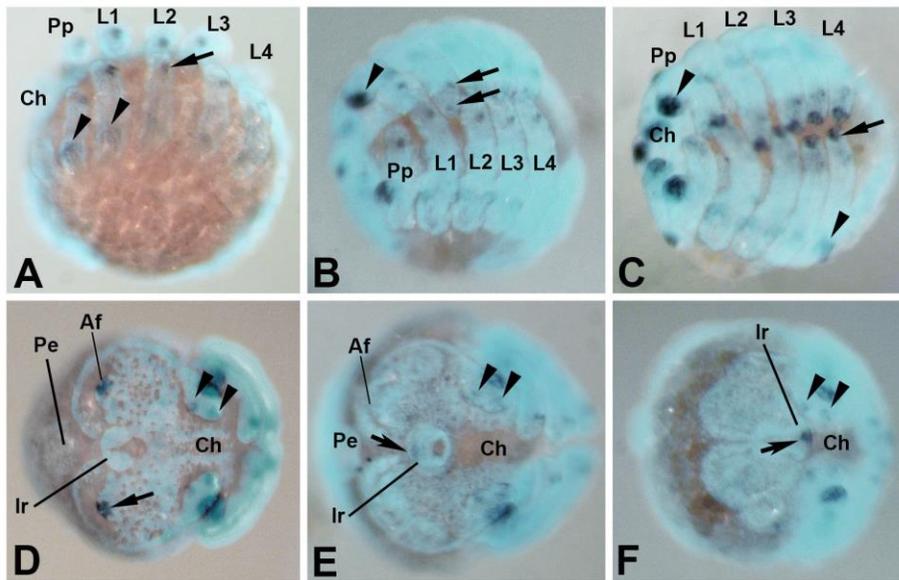


Figure 74: Expression of *BarH2* in *P. tepidariorum*

A: Lateral view of an embryo during limb bud elongation. **B:** Embryo during inversion shown in ventro-lateral view. **C:** Embryo during dorsal closure also in Ventro-lateral view. **A-C:** The arrowheads point to expression of *Pt-BarH2* in the gnathendite of the pedipalps and corresponding tissue at the base of the legs. The Arrows point to distal spotted *Pt-BarH2* expression in the pedipalps and walking legs. **D:** Embryo during limb bud elongation with *Pt-BarH2* expression in the anterior furrow (arrow), shown in frontal aspect. **E and F:** Embryo during inversion and dorsal closure shown in frontal aspect. The arrowheads point to two spots of *Pt-BarH2* expression in the chelicera. The Arrow indicates *Pt-BarH2* expression in the labrum. All embryos are oriented with anterior to the left. Af: Anterior furrow, Ir: Labrum, Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4, O1-O5: Opisthosomal segments 1-5, pe: posterior end.

The weak phenotype found in freshly hatched nymphs showed the absence of the first walking leg pair or the entire the first walking leg segment (depicted in Figure 76, C and D, strongest in cocoon six, but less than 10%).

In general the RNAi of *Pt-BarH1* showed a high penetrance with over 60% phenotypes in the cocoons 2-7 (Figure 75). More detailed analysis of the unhatched embryos from the affected cocoons showed that these embryos had several defects concerning anterior segmentation or showed completely disrupted embryonic development (other defects indicated in black, Figure 77). Strongest phenotypes were missing the head segments, including the cheliceral, pedipalpal and at least first leg segment (Figure 76, E and F). More common (approximately 20-40% in Cocoon 2-4, Figure 77) were phenotypes with ventral fusion or absence of at least one or different combinations of several prosomal appendages (usually chelicera, pedipalps or the first walking leg, example in Figure 76,

D), which resembles the weak phenotype observed in the hatched nymphs. The reason why animals missing the first walking leg were able to hatch, but animals lacking the pedipalps were not able to hatch is the absence/presence of the egg tooth. This structure is located on the pedipalps and is required for the initial bursting of the embryonic envelope (Mittmann and Wolff, 2012).

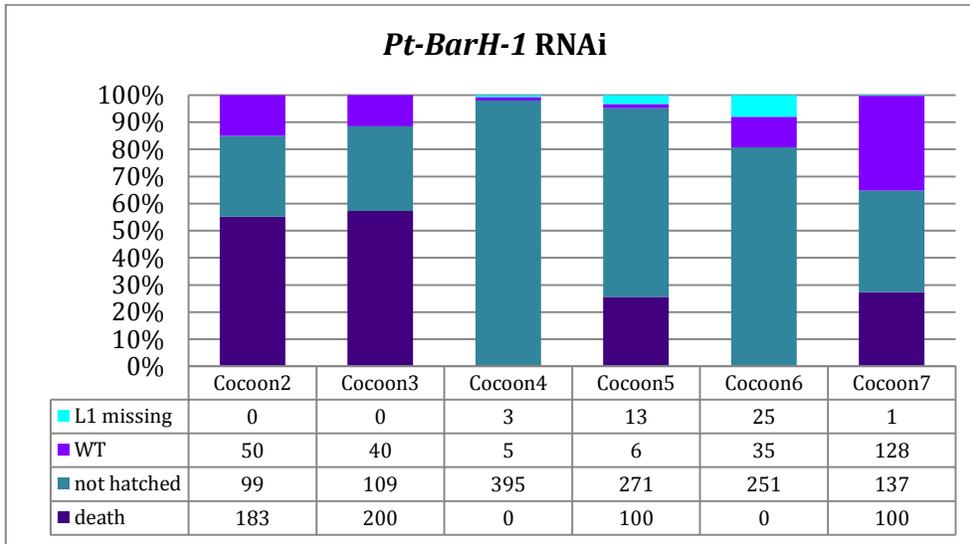


Figure 75: Phenotype of *Pt-BarH1* knockdown in the different cocoons

The graph gives an overview of the phenotypes in freshly hatched nymphs after knockdown of *Pt-BarH1* in the cocoons 2-7 produced by females after injection. Up to the sixth cocoon a majority of offspring are affected (over 80%) and were not able to develop normally (Wt: indicated in light violet). The offspring of cocoon 4 were nearly completely unable to hatch from the embryonic envelope. But also the cocoons 5 and 6 are severely affected with total number of unhatched embryos ranging between 60-80%. The only phenotype which was observed after moulting into first instar nymphs shows absence the first walking legs segment as depicted in Figure 76 C. This phenotype is found in offspring of the cocoons 4 to 7, but in a relatively low amount (highest in cocoon 6 with less than 10%).

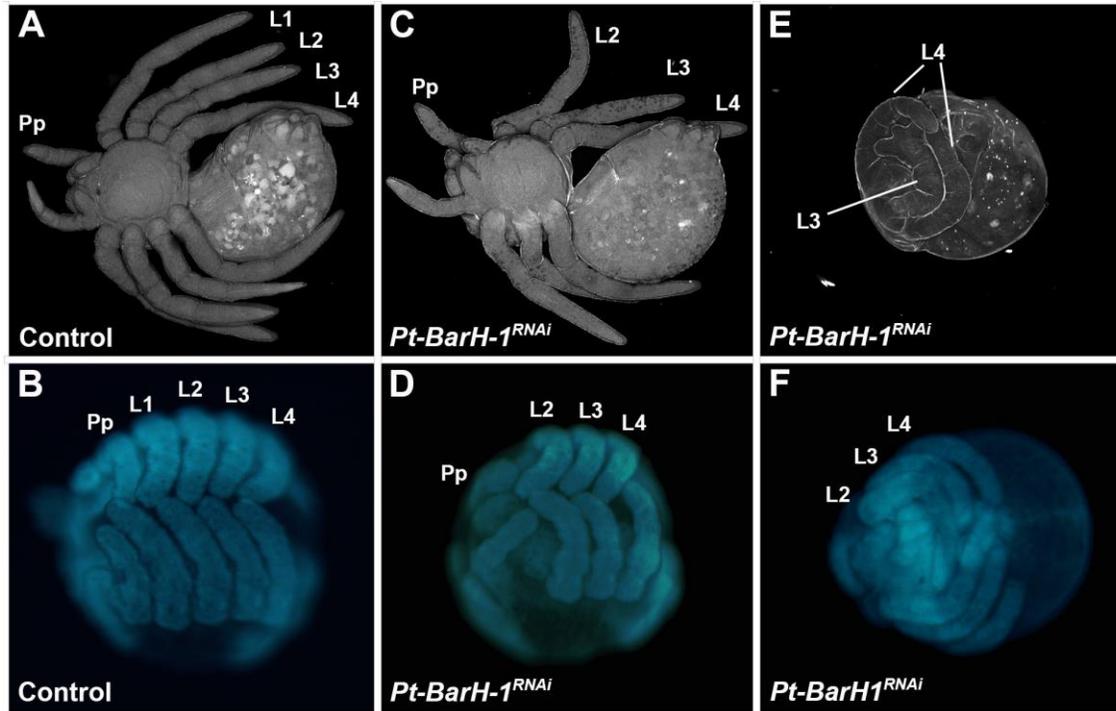


Figure 76: Embryonic and nymphal phenotypes from *Pt-BarH1* knockdown

A: Freshly hatched first instar nymph from a cocoon produced by an untreated female, mounted in Voltalef oil. **B:** Embryo during inversion stained with Sytox-Green from a female without any treatment. **C:** Freshly hatched first instar nymph (mounted in Voltalef oil) showing the weak *Pt-BarH1* phenotype indicated by absence of the first walking leg segment. **D:** Embryo after knockdown of *Pt-BarH1* with weak phenotype missing the first walking leg pair on the corresponding segment stained with Sytox-Green. **E:** Embryo after completion of embryonic development unable to hatch from the embryonic envelop due to missing egg tooth usually located on the pedipalps. The individual was mounted in Voltalef oil after the embryonic eggshell was removed by hand. It represents a very strong phenotype missing all segments anterior to the third walking leg segment. The third walking leg segment carries only one leg, which is transformed into a trident at the distal portion. **F:** Embryo probably during or shortly after ventral closure after knockdown of *Pt-BarH1*. This embryo represents a weaker variation of the strong phenotype, missing all segments anterior to the second walking leg segment, but with the three posterior walking leg pairs fully developed. All specimens are oriented with anterior to the left and in ventro-lateral view. Ch: Chelicera, Pp: Pedipalps, L1-L4: Walking legs pairs 1-4.

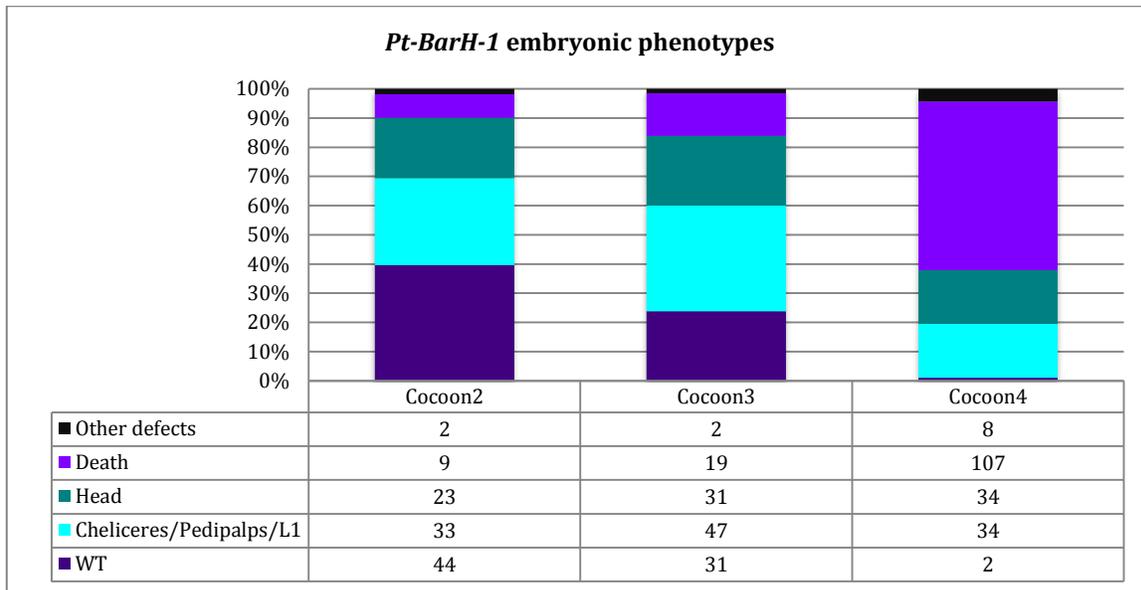


Figure 77: Overview of embryonic phenotypes in cocoon 2-4 after *Pt-BarH1* knockdown

The graph shows the analysis of embryos from the cocoons 2-4 to reveal the reason of the hatching problem observed in a majority of individuals after knockdown via parental RNAi of *Pt-BarH1*. The number of properly developed embryos is indicated by Wt and marked in dark violet. It decreases from nearly 40% in the second cocoon to nearly no individual in the fourth cocoon. In the fourth cocoon the majority of embryos died during the course of development. In the two previous cocoons the majority of individuals show either strong or weak defects, including missing the complete head or deletion or fusions of the anterior prosomal appendage segments. Other defects, like axis duplication and multiple opisthosoma represent the minority of phenotypes in these cocoons with less than 10% in each.

6.7.4 Expression of head and leg developmental genes in *Pt-BarH1* RNAi embryos

To verify which walking legs were missing in the different weak phenotypes we examined the expression of the Hox gene *Sex-combs-reduced* in embryos after *Pt-BarH1* knockdown. Expression of this gene is absent in the first walking legs pair, and has a leg-specific pattern for each of the walking leg pairs 1-3 of spiders (shown in Figure 78, D and in previous findings (Khadjeh et al., 2012; Schwager et al., 2007)). The expression of *Pt-Scr* supports the conclusion that the first walking leg is missing, whereas the other three walking legs are able to develop normally in weakly affected *Pt-BarH1* knockdown phenotypes (Figure 78, H). Furthermore the expression in these more posterior walking legs shows that they can develop normally although the complete head including the first walking leg segment is missing (Figure 78, L). The expression of *Pt-Dll* in very weak phenotypes (Figure 78 G) shows that in some cases the segment of the first walking leg is

still intact and only the pair of legs is reduced. In contrast *Pt-Dll* expression in strong *Pt-BarH1* knockdown phenotypes indicates that no appendage primordia are present after depletion of the head tissue.

orthodenticle (Pechmann et al., 2009) and the early function of *Pt-Dll* (Pechmann et al., 2011), were previously found to be involved in the head development of *P. tepidariorum*, their interaction and the genetic mechanism, however, is not known. To ascertain the position of *Pt-BarH1* in the head patterning network of *P. tepidariorum* the expression of these genes was observed in early embryos after *Pt-BarH1* knockdown. Due to lacking anterior tissue in the *Pt-BarH1* phenotypes it was challenging to predict in which stage of development the observed embryo were fixed. The embryos (Figure 78 (E-F and I-J)), could either represent late stage six or late stage seven embryos, which is hard to elaborate due to missing anterior tissue. However, in general the staining of *Pt-otd* in these phenotypes shows a disrupted and patchy *otd* expression after knockdown of *Pt-BarH1* (Figure 78, E and F). The expression of *Pt-Dll* in *Pt-BarH1* knockdown embryos seems at first glance less affected compared to expression of *Pt-otd*, indicated by a strong anterior stripe of expression. Comparison to early stage seven control embryos, however, indicates the complete absence of the head tissue anterior to the *Pt-Dll* stripe in the knockdown phenotypes. Another possible alteration of *Pt-Dll* caused by *Pt-BarH1* is the suppression of the splitting of the *Dll* stripe (not shown but previously described in (Pechmann et al., 2011)) later during stage seven.

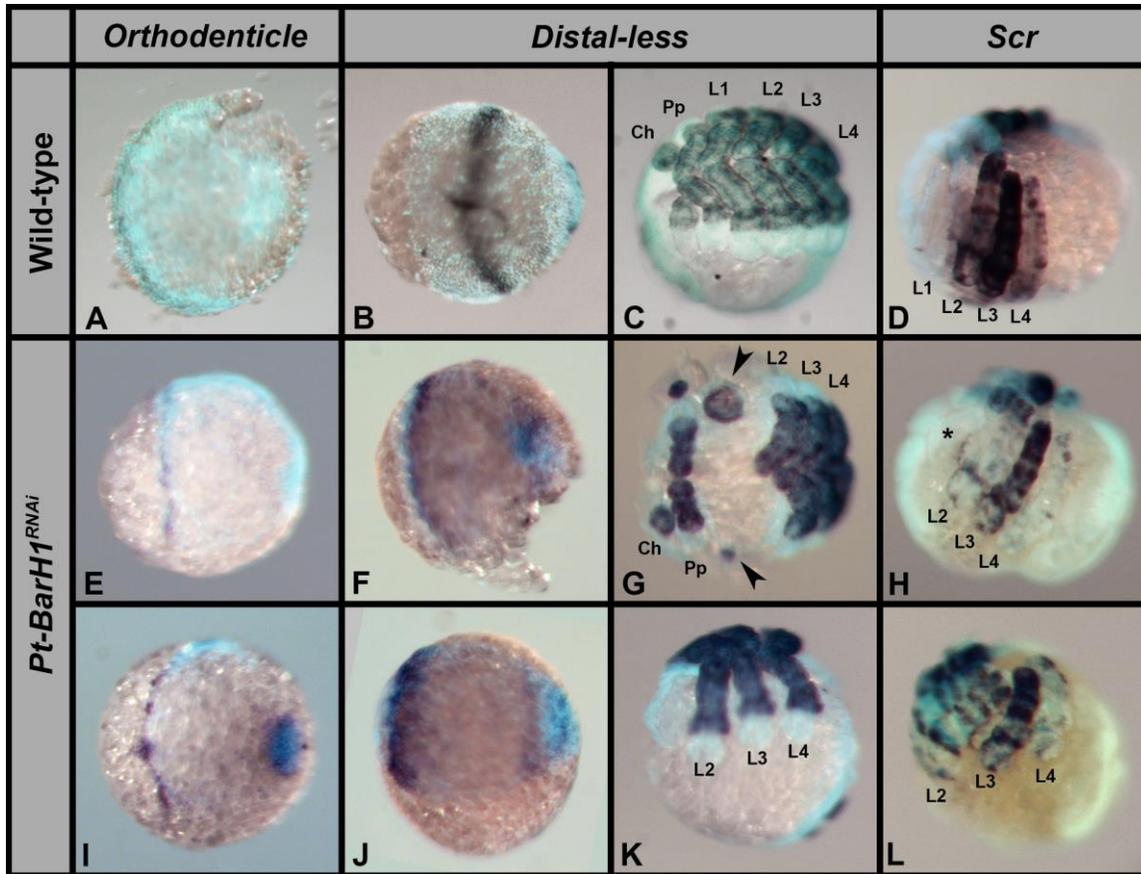


Figure 78: Expression of *otd*, *Dll* and *Scr* in embryos after *Pt-BarH1* RNAi

A-D: Expression of *otd* (A), *Dll* (B and C) and *Scr* (D) in untreated *P. tepidariorum* embryos at different stages. E-L: Expression of *otd* (E and I), *Dll* (F-G and J-K) and *Scr* (H and L) after knockdown of *BarH1* at different stages of embryonic development in *P. tepidariorum*. The embryos shown in A and B, E and F as well as I and J represent embryos after break of radial symmetry on the way to germ band stages. C, G and K: Expression of *Pt-Dll* in prosomal appendages during early inversion. The arrowheads point to remaining *Pt-Dll* expression in reminiscent leg tissue on an intact first walking leg pair segment. The embryos are shown in lateral (C and K) and ventral view (G). D, H and L: lateral view of embryo with expression of *Pt-Scr* during dorsal closure. The asterisk indicates the fusion of the first walking leg pair into a single appendage. All embryos are shown with anterior to the left. Ch: chelicera, Pp: pedipalps, L1-L4: walking leg 1-4.

6.8 Discussion and future work

6.8.1 Distal limb development is not conserved for insects and spiders

Findings from the insects *D. melanogaster* and *T. castaneum* and the onychophoran *E. kanangrensis* (Angelini et al., 2012; Estella et al., 2012; Kojima, 2004; Oliveira et al., 2014), the sister group of arthropods suggested that patterning of the distal limb development is conserved in panarthropods and thus similar genes could be used for tarsus segmentation in *P. phalangioides*. However, the lack of any tarsus segmentation in most arthropods and especially in chelicerates instead suggests that an unsegmented tarsus is the ancestral state in arthropods. Therefore I studied several tarsus segmentation candidate genes in this spider and compared their expression to *P. tepidariorum*, a spider without segmentation of the tarsus. The first evidence that the genetic cascade involved in distal limb patterning in insects was not already present in the last common ancestor of the panarthropods (including onychophorans and arthropods) is provided by the lack of expression of the genes *rotund*, *bric-a-brac*, *chip* and *tango* in *P. tepidariorum*.

This is further supported by the absence of distal expression of all *ap* homologs, the *Lim1* homologs and the *trh1* orthologs in the walking legs of both spiders. Only very few of the candidate genes revealed a possible function in the distal part of the walking leg. The homologs of *al*, both paralogs of *ss* in *P. tepidariorum* and *P. phalangioides*, as well as both *BarH* homologs in *P. tepidariorum* are expressed in the distal tips of the walking legs. To define sharp borders between the segments, genes involved in leg or tarsus segmentation should be expressed in broad domains in the segment of their function or in rings at the joints (Angelini and Kaufman, 2005; Prpic and Damen, 2009). All these genes, however, show a spotted expression in the distal leg, that suggests some function in this portion of the legs, but does not immediately suggests a role in tarsus segmentation. Spot-like expression usually is found for genes involved in neurogenesis, defining neural precursors (Doeffinger et al., 2010), which could develop into sensory organs, like bristles or lyriform organs. Thus also these genes are probably not conserved for distal leg development. Only in two cases a ring-like, or broad domain of

expression was observed. Expression of *trh2* in *P. tepidariorum* is present in a ring-like manner. However, this expression was also observed in more proximal part of the developing walking legs. Thus *trh* homologs might have a general role in leg development but not exclusively in the distal segment. The other case are the two *cll* paralogs in both spiders. These were expressed exclusively in the distal part of the developing prosomal appendages. In addition, knockdown of *cll1* in *P. tepidariorum* revealed that it is required for proper tarsus formation. In summary my data suggest that the similarities between insects and onychophorans are rather the results of a parallel co-option of a plesiomorphic gene regulatory network or caused by misinterpreted leg expression in the onychophoran. The expression patterns of the tarsus segmentation candidates in the distal part of the spider walking legs, except for *cll*, are clearly different from *D. melanogaster* and do not suggest any role in distal segmentation. Thus neither the distal patterning functions known from insects nor tarsus segmentation are phylotypic traits of arthropods or panarthropods.

6.8.2 Cross-regulation of *al*, *cll* and *Lim1* is not conserved in arthropods

Only the gene *al* was not duplicated in the spiders studied in this thesis and in the predicted genome sequences of the spider *S. mimosarium*. Although many *al* related genes were found these were not grouped into the clade containing genuine *al* homologs from other model organisms. The high support value of the monophyletic *al* group including homologs from insects, chelicerates and vertebrates rather suggests that no duplication event occurred for this gene in the lineage leading to spiders or that a possible duplicate was secondarily lost. *al* has a complex expression pattern in the entire walking legs during embryonic development of *P. tepidariorum* and *P. phalangioides* and might therefore be used for different purposes. The early ring-like expression hints to a possible role in segmentation, because rings of expression are often found for genes involved in patterning and segmentation processes (Prpic and Damen, 2009). The spot-like *al* expression at the base and in the tip of the walking legs probably define neural precursor cells. Although the expression pattern of *al* drastically changes during embryogenesis and shows a very complex pattern in the late walking legs, the pattern

and temporal profile of the expression is virtually identical in both spider species. Comparisons to *al* expression in *Gryllus bimaculatus*, *D. melanogaster*, *Manduca sexta* and *T. castaneum* show two general patterns found within insects (Beermann and Schröder, 2004; Campbell and Tomlinson, 1998; Miyawaki et al., 2002; Tanaka and Truman, 2007). Whereas the expression in *G. bimaculatus* and *T. castaneum* also shows a complex pattern in the developing legs, *al* expression in *M. sexta* and *D. melanogaster* is restricted to the coxa and the most distal part of the leg. Because *G. bimaculatus* and *T. castaneum* are less derived insect species it is most probably that the loss of medial *al* leg expression was lost in the lineage leading to *D. melanogaster* and *M. sexta*. Furthermore *G. bimaculatus* is the only representative from the hemimetabolous group that does not involve the derived pupal stage, and thus its general mode of development might be more similar to spiders than to holometabolous insects. Despite the divergent expression patterns, however, functional studies in *T. castaneum* showed that knockdown of *al* results in the absence of the pretarsus segment carrying the claw as it was also shown for *D. melanogaster* (Angelini et al., 2012; Beermann and Schröder, 2004). If knockdown of the distal expression in spider legs would lead to similar phenotypes, one could suggest that the spot-like expression is required for claw formation and that this function is conserved in all arthropods. Otherwise the function in claw development has evolved after the split of spiders and insects. The expression in the distal tip of the onychophoran hints to a phylogenetically old conservation of this gene in the leg (Oliveira et al., 2014). I note, however, that the expression was only observed in very young embryos and is extremely weak. It might therefore be premature to draw any conclusions about gene function in onychophorans. Rather, functional studies of *al* in spiders and onychophorans are required to understand the role of *al* in distal leg development. It is also interesting that activation of *Lim1* by *al* is dependent on protein complex formation with *c11* demonstrated in *D. melanogaster* (Campbell, 2005; Miyazono et al., 2010). In *P. tepidariorum* and *P. phalangioides* *al* and *c11* are never co-expressed and it is not likely that transcription factors diffuse over larger distances in cellularised environments. This could be explained in two ways: (1) If the formation of the Al and C11 protein complex is ancestral and required for correct binding of specific DNA targets neither *al* nor *c11* would be functional in the leg of spiders. (2) In spiders Al and C11 have other co-factors, which

mediate binding to DNA targets. The second scenario is more likely, because genes without function would not show such a conserved expression pattern like *al* and *c11*. In addition, the tarsal phenotypes obtained after knockdown of *c11* demonstrate that at least *c11* is not an unfunctional gene in the tarsal segment.

Lim1 was not expressed distally in the walking legs, in contrast to *E. kanangrensis* and *D. melanogaster* (Kojima, 2004; Oliveira et al., 2014; Pueyo, 2004). As stated above data from the onychophoran are based on expression patterns only and are in addition not observed for the entire walking leg development as performed in this thesis. Knockdown of *Lim1* in *T. castaneum* resulted in fusion of femur and tibia but no defects in the distal part of the leg (Angelini et al., 2012). Even double or triple knockdown together with *al* and *c11* only showed additive defects, but no role in distal patterning (Angelini et al., 2009). Unfortunately expressional data for *Lim1* during embryonic development of *T. castaneum* is lacking. In summary this suggests that *Lim1* in spiders and *T. castaneum* likely represent the ancestral function, which is independent from interaction with *al* and *c11*.

6.8.3 Conservation of distal function of *clawless*

Both *c11* paralogs in *P. tepidariorum* and *P. phalangioides* show expression in the tarsus segment. However, the exact pattern differs in late embryonic stages for both paralogs. Tarsus segmentation presumably starts at late embryonic stages. Thus differentiation of the paralogs in the two spiders might reflect differing roles in tarsus segmentation of *P. tepidariorum* and *P. phalangioides*. The distal *c11* expression in the walking leg of both spiders divides into two rings during inversion, but not in the pedipalps, which is consistent with the fact that the pedipalps lack the metatarsal segment. Interestingly, the further pedipalpal and walking leg expression profiles of *c11* differ between *P. tepidariorum* and *P. phalangioides* indicating that *c11* might be involved in the formation of species specific differences in pedipalp and leg morphology between the two species. *c12* is more similarly expressed in the pedipalps and walking legs and thus is unlikely to be involved in the development of morphological differences between these two appendage types. Unfortunately, RNAi is not yet available in *P. phalangioides* and

therefore a possible role of either *cll* gene of this species in the special tarsus segmentation uniquely found in *P. phalangioides* is unclear at the moment and awaits future gene function studies.

Knockdown of *cll1* in *P. tepidariorum* demonstrate that it is necessary for proper development of the tarsal tissue. Depending on the timing of RNAi effect, we see either a disruption of the formation of the joint or the ablation of the entire tarsal segment. This shows that *cll* is conserved in defining most distal tissues in insects and at least one spider (Campbell, 2005; Cande et al., 2009; Miyazono et al., 2010). Together with findings from the onychophoran (Oliveira et al., 2014), where *cll* expression in the distal tip is more defined and stronger than the other distal genes observed in the same study, this hypothesis is supported. Lack of distal phenotypes after RNAi in *T. castaneum* might be caused by problems of the RNAi procedure itself. This is additionally supported by strong expression found in the distal tip of developing embryos of this species (Cande et al., 2009; Grossmann and Prpic, 2012).

6.8.4 Other conserved functions of tarsus segmentation genes

In contrast to different or absent expression of tarsus candidate genes in the tarsus of *P. phalangioides* and *P. tepidariorum*, conservation of these genes involved in other processes was observed. Both *ss* paralogs in both spiders for example showed spot-like expression in all prosomal appendages as well as in the limb buds on the opisthosoma. As previously mentioned spot-like expression hints to a possible role for defining neural precursors, including sensory organs and bristles. Apart from its function in distal leg patterning of *D. melanogaster*, *ss* is required for the development sex combs on the first tarsal segment (Kuzin et al., 2014), correct bristle formation in the antenna and legs (Emmons et al., 1999) and for dendrite diversification in dendritic arborization of sensory neurons (Kim et al., 2006). Thus the role of *ss* in distal leg development apparently represents the derived mode of this gene, whereas its role in defining neurogenic tissues or its descendants is probably conserved. This is supported by findings in *C. elegans* where the *ss* homolog, named *aryl hydrocarbon receptor 1*, mediates

the differentiation of neuronal fates (Huang et al., 2004). Functional studies to clarify the exact role of *ss* in spiders are in progress, but did not show any results yet.

Also the expression of the *trh* homologs in the spiders showed interesting parallels to previously described functions. Both *P. phalangioides* paralogs and *trh2* from *P. tepidariorum* are expressed in the opisthosomal limb bud on the third opisthosomal segment. This limb buds invaginate during embryonic development and then gives rise to the tubular trachea of the spider. Interestingly, *trh* in *D. melanogaster* is required for tubulogenesis and was found to act up-stream regulating most genes involved in trachea development (Chung et al., 2011; Isaac and Andrew, 1996). Although functional data in the spiders are lacking, the expression in the precursor of the spider trachea suggests, that this gene has a conserved function in trachea development. This is even more surprising because morphological studies so far suggested that the tracheae of spiders and flies are analogous structures. But already comparative studies of the respiratory organs of *D. melanogaster* and crustacean gills showed that *trh* is expressed similar in both tissues and thus might originate from the same precursor, although they are morphologically diverged (Franch-Marro et al., 2006; Sánchez-Higueras et al., 2014).

For *ap* we did not find tarsal expression in either spider species. The restriction of the expression ventrally to the most proximal part suggests that *ap* in spiders does not have a significant role in walking leg patterning. However, the expression in the opisthosoma and the opisthosomal limb buds is similar to the previously studied *ap* homologs of *C. salei* (Damen et al., 2002). That study concluded that the opisthosomal appendages of spider and the wings of insects might be homologous appendages evolved from ancestral gills, but without more detailed and especially functional analysis of this gene this hypothesis remains open for debate. The co-expression *ap* and the other Lim-type family gene *Lim1* in the legs or in the opisthosomal limb buds are more compatible with a role in neural differentiation or muscle precursor diversification as already known from the annelid *Neanthes arenaceodentata* (Winchell and Jacobs, 2013) and also from *D. melanogaster* (Bourgouin et al., 1992).

Surprisingly, the expression of *Pp-ap2.2* is more similar to *ap.1* of *P. tepidariorum* and *C. salei*, whereas *Pp-ap1* shows a different expression pattern in the opisthosoma of *P. phalangioides*. This calls into question my classification of the orthology of these

duplicates in the spiders, and suggests that the phylogenetic analysis was not able to properly resolve the phylogenetic relationships of the spider *ap* homologs. The phylogenetic tree also suggests that *ap2* was duplicated once more in the *P. phalangioides* and *P. tepidariorum* but not in *C. salei*. Whether this duplication of *ap2* did not take place in *C. salei* or just escaped previous gene cloning attempts or is also the result of a problematic phylogenetic analysis is currently unclear. The second scenario is supported because also not all cloning attempts for *ap2* lead to successful cloning in *P. tepidariorum* and *P. phalangioides*. Findings based on the expressional data of the different *P. phalangioides* tissues during the transcriptomic approach further support this hypothesis: no read counts were found for *Pp-ap2.1* in the embryonic tissue, but only in the opisthosoma of the nymphs. Thus it is probably missing in the cDNA libraries synthesized from embryonic stages and thus escaped cloning attempts. In addition, a second copy of *ap2* is present in an additional spider, *Stegodyphus mimosarum* (*S. mimosarum*), as revealed during the phylogenetic analysis. *S. mimosarum*, *P. tepidariorum* and *C. salei* all belong to the entelegyne spiders. *P. phalangioides* is placed into the group of haplogyne spiders. The presence of two copies of *ap2* in members of both entelegyne and haplogyne spider, but not in *C. salei* thus can only be explained (if escaping cloning attempts could be ruled out as most likely cause) by loss of this gene in *C. salei*. More expressional and also functional analysis would be helpful, but also the study of the transcriptome of *C. salei* and additional chelicerate species would help to have more profound understanding of the phylogenetic relationships of the *ap* paralogs within this lineage.

6.8.5 Possible role for *ss* and *trh* in growth regulation during late embryogenesis

As previously mentioned unbiased approaches using transcriptomics representing a valuable opportunity to identify novel genes involved in morphological diversification. This was supported by various findings in different species (Armisen et al., 2015; Dong et al., 2015; Gilad et al., 2009; Pantalacci and Sémon, 2014; Tom et al., 2014; Yu et al., 2014) and in this work. However, the analysis of the tarsus candidate genes *ss* and *trh* revealed an extremely interesting expression during late leg development of *P. phalangioides*. The

analysis of the embryonic development of *P. phalangioides* showed that the extreme growth of the walking legs starts and continues during ventral closure stages. *ss1* and *trh2* show an extreme increase in expression at exactly that developmental step. The genes are very strongly expressed in the entire walking leg excluding the short segments and the areas of the joints, thus areas where growth should be enhanced or repressed, respectively. This expression pattern was only found in *P. phalangioides*, thus the spider species with long legged morphology and not in *P. tepidariorum*, which develops shorter leg compared to the overall body size. At least *Pt-trh2* also shows an increase of expression during late embryonic stages, however, the pattern is ring-like and not covering the entire segment. A possible role of *ss* and *trh* paralogs of *P. phalangioides* during leg growth are further supported by comparison to their expression in the other prosomal appendages. The pedipalps and chelicera of *P. phalangioides* are not prolonged like the walking legs. Also the expression of *ss* and *trh* in these appendages shows a ring-like pattern similar as observed for *Pt-trh2* or *D. melanogaster*. Thus the candidate approach might have revealed two genes, which could have a role in the extreme leg growth of the long-legged spider species, although this was not the purpose of the study. The influence of these genes in growth control, however, must now be verified by functional studies.

6.8.6 Novel role for *BarH1* in head patterning of *P. tepidariorum*

For the tarsus candidate gene *BarH1* I also found an unexpected expression apart from appendage patterning. *BarH1*, but not *BarH2*, is expressed in the periphery of the developing germ disc during early embryonic development of *P. tepidariorum*. For this gene it was even possible to verify the function caused by this expression. Knockdown of the gene resulted in embryos lacking at least the first walking leg segment or even the complete head. A similar role has been previously observed for another gene of *P. tepidariorum* usually involved in leg patterning, *Dll* (Pechmann et al., 2011). *Dll* knockdown embryos lack the first and sometimes also the second walking leg segment. The phenotype after *BarH1* RNAi thus affects also more anterior segments as *Dll*. In addition to *Dll*, *orthodenticle (otd)* is known to be involved in head development of *P.*

tepidariorum (Pechmann et al., 2009). Knockdown of *otd* resulted in the loss of all head segments anterior to the first leg segment. It was shown that the formation of a proper head requires a dynamic fine tuned expression of the anterior genes (Pechmann et al., 2009). Because the *BarH1* RNAi phenotype represents a combination of both *otd* and *Dll* phenotypes it might be involved in orchestrating these early dynamics required for proper head formation. Preliminary data of *otd* and *Dll* expression in early *BarH1* knockdown embryos indeed showed partial loss of the strong anterior *otd* expression and a possible alteration of *Dll* expression dynamics. Whether *BarH1* acts upstream of *otd* and *Dll* or if these alterations are an artefact caused by missing the anterior tissue remains to be tested in further experiments. Testing this would require a detailed examination of the early expression dynamics of these genes in the anterior part embryo of *P. tepidariorum*, as well as combinations of loss and gain of function studies.

7. General Discussion

7.1 Hourglass model of walking leg development

The aim of my work was to study which genetic mechanisms are important to drive the immense arthropod appendage diversity. Despite this extreme diversity the different appendages all need to undergo some crucial developmental steps, like specification of primordia, development of the proximo-distal axis, further patterning and segmentation as well as growth (Angelini and Kaufman, 2005). The walking leg development of *D. melanogaster* is studied in great detail. It was shown that the legs are developed following a cascade of gene networks required for these different steps (Detailed description in the Introduction) (reviewed in Estella et al., 2012; Kojima, 2004). Thus the question is which genetic modules of this cascade are free for alteration and which are crucial and need to be conserved to ensure proper walking leg development in general. To answer this question my work and previous studies in other representatives of the arthropods studied the gene networks of this cascade and compared the level of conservation.

A summary of the previous findings by Prpic and Damen 2008 supposed, that an hourglass could represent the genetic cascade involved in walking leg development. The early steps like appendage primordia specification and initialization of the proximo-distal axis at the beginning of the cascade require some similar genes, but the exact function of these genes is different between species. These different modes of early appendage development are reflected by differences of the gene regulatory networks involved in appendage primordia specification (e.g. Khadjeh et al., 2012; Pechmann et al., 2011, Angelini and Kaufman, 2005; Hughes and Kaufman, 2002b). In contrast to the early steps of the walking leg development cascade, the following step, patterning the proximo-distal axis and development of the joints are extremely conserved within different species. During this step the leg gap genes organize the regionalization of the developing limbs into proximal, medial and distal parts (Angelini and Kaufman, 2005,

Janssen et al., 2010). Previous studies showed that the leg gap gene function is conserved in all arthropods and also found to be required during vertebrate limb development (Janssen et al., 2010; Prpic and Damen, 2004; Prpic and Telford, 2008; Prpic et al., 2001, 2003; Sharma et al., 2012). The study of these genes in *P. tepidariorum* and *P. phalangioides* further supports the constrained role of these genes for the development of walking legs. It was suggested by Prpic and Damen, that this developmental stage might represent the constraint podotypic stage of locomotory legs. The correct regionalization thus represents a crucial step during walking leg development, which should not be altered to diversify appendage morphologies. The data on the leg gap gene *exd1*, however, showed that this only concerns early functions of these genes, whereas later function, like the patella ring observed for this gene are again suited to be altered (Prpic and Damen, 2009).

Now the question arises where changes could occur in genetic cascade to drive the evolution of novel morphologies. The next step in the cascade of walking leg development is the development of joints between the leg segments. The segmented leg is a common trait of all arthropods and is most likely one of the keys to their evolutionary success. Because of this important role of leg segmentation, the genetic mechanism behind it should be conserved as well and is probably not the basis for diversification. Expressional and functional data for the Notch signalling pathway largely reflect this genetic conservation and thus indicates that also this module of the cascade belongs to the podotypic leg state of arthropods (Bishop et al., 1999; Celis et al., 1998; Rauskolb and Irvine, 1999; Angelini et al., 2012; Mito et al., 2011; Prpic and Damen, 2009). Thus the morphological diversification must occur after these crucial developmental steps. The next process in the hierarchy involves additional patterning processes refining the leg gap gene borders during leg elongation, especially in the distal leg part. The distal-most segment, the tarsus, is further subdivided into numerous distal segments in many insects and the haplogyne spider model studied here. The secondary PD axis patterning gene battery is initiated by *wg* and *dpp* interactions and *Dll* and involves EGFR signalling and its downstream targets (summarized in Estella et al., 2012). In contrast to leg gap genes and Notch mediated joint formation secondary proximo-distal patterning is only intensively studied in *D. melanogaster* and *T. castaneum*

(Angelini et al., 2012; Beermann and Schröder, 2004; Grossmann and Prpic, 2012; Kojima, 2004). Data for representatives of the other three arthropod clades a scarce and data from an onychophoran species are only based on expressional data of not fully elongated legs (Hao et al., 2003), Comparison of tarsus segmentation between *T. castaneum* and *D. melanogaster* showed only some functional differences of this patterning process. The adult leg morphologies of these two species are, however, also relatively similar. Thus studies in species with morphologically more different walking legs are required to verify the hypothesis that the distal segmentation module is less constrained and therefore drives morphological diversification. Furthermore the findings of *T. castaneum* were only based on RNAi studies, mostly lacked whole mount embryonic expressional observations and did not include the complete set of tarsus segmentation genes identified in *D. melanogaster*. In my work I studied the gene battery involved in tarsus segmentation in two spider species. One, like *D. melanogaster*, with tarsus segmentation, and the other one, without any subdivision of the tarsus. When the further subdivision of the most distal part of the walking legs would be ancestral for all arthropods these genes should be conserved at least between *P. phalangioides* and *D. melanogaster* and only different in *P. tepidariorum*.

My data show that this hypothesis can clearly be rejected. Some genes were not expressed during embryonic development at all and thus are probably not conserved in distal patterning of arthropods. However, other genes were expressed in the prosomal appendages, but not in the distal part. Even if expression in the distal part was present, the pattern of expression suggested a different function than found for *D. melanogaster*. Furthermore these genes were similarly expressed in the walking legs during embryonic development of the two spiders. Thus the late steps of the genetic cascade of walking leg development is in contrast to the podotypic stage not strictly functionally conserved and represent the "edge of the hourglass", and are therefore good candidates driving the evolution of walking leg diversification. However, functional studies and more species are required to study the exact diversification of these genes and other possible factors involved in the development of novel traits. Such novel factors might be missed by candidate gene based studies. The use of unbiased transcriptomic approaches is suited to reveal genes causing developmental differences. Indeed, my study of differentially

expressed genes in the walking legs of *P. phalangioides* identified candidates with conserved roles in cell proliferation. Control of allometric growth driven by cell proliferation is one of the last steps after patterning of the walking legs. Another interesting candidate from the transcriptomic approach is the gene *odd skipped*. It is known to be involved in tarsal development of the fruit fly and shows distal expression in the walking legs of the two spider species studied here. Thus it might represent a conserved gene required for distal leg patterning, in contrast to the mostly insect specific tarsus segmentation genes. Interestingly the distal *odd-skipped* expression is different in *P. phalangioides* and *P. tepidariorum* and thus might be involved in the special tarsus segmentation of the haplogyne spider. All these hypotheses are based on expressional data and of course await future functional studies.

The gene *cll* is the only tarsus segmentation gene candidate, which is expressed in the distal part of both spider species. For this gene it was possible to perform knockdown experiments, which showed that it is required for the formation of the tarsus in *P. tepidariorum*. This indicates a conserved role for *cll* in distal patterning. The exact function of this gene must, however, be different to the one known in *D. melanogaster*, because the co-factors and downstream targets are not present in the spider legs.

During the study of the duplicates of the leg gap gene I found that the *dachshund* duplicate is necessary for proper formation of the arachnid specific leg segment the patella. Thus this duplication led to the development of a novel morphological trait. The *hth* duplicate is differentially expressed in the walking legs of four different spider. Intriguingly, gene duplications represent a further possibility how novel traits might evolve on the genetic level.

All these data indicate that the diversification of walking leg morphologies probably happens after the establishment of the crucial podotypic state that includes the highly conserved gene patterning networks and is driven by a combination of different genetic causes, namely (1) altered roles of conserved gene (*cll* and *odds*), (2) evolution of novel gene function of existing genes (most tarsus segmentation gene of *D. melanogaster*), (3) differences in cell proliferation and (4) new genes (duplications) (summarized in Figure 79). In summary my study supported the suggested hourglass model and provided hints of the previously unknown genetic causes acting at the edge of the hourglass to drive

morphological differences and showed that these are probably positioned in the most downstream area of the walking legs development cascade.

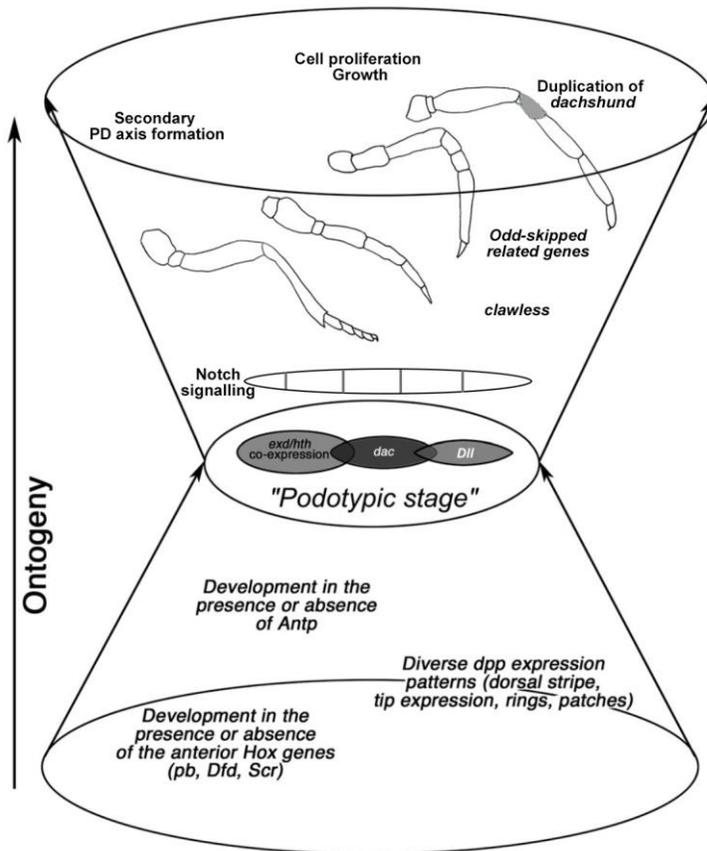


Figure 79: Hourglass model of walking leg diversification development

Illustration of the walking leg development cascade. Early steps include specification of walking leg primordia on the anterior-posterior axis, controlled by *dpp* and the Hox genes. These early steps show a certain level of genetic diversification. The patterning of the proximo-distal axis and the development of the joints requires the leg gap genes and Notch signaling. This is the most conserved step in the walking leg development cascade of arthropods and might represent a podotypic stage. During my work I was able to show that the genetic mechanisms more downstream in the walking leg development cascade involved in secondary PD axis formation and tarsus patterning are not conserved. This can explain the development of diversified walking legs despite the constrained genetic network within the podotypic stage. Modified from Prpic and Damen, 2008.

7.2 Duplication of genes facilitates diversification

One other main finding of my work was that nearly all of the studied candidate genes were duplicated in the two spider species. Diversification during evolution requires diversification also at the molecular level. To diversify gene functions, genes can be

altered on the regulatory level or in their coding region. Because most genes have several functions (so-called pleiotropic genes), changes in their sequence or in their regulation usually cause alterations in different organs or tissues (Rose, 1982). This bears the risk of disturbing finely tuned gene regulatory networks crucial for proper development of the individual. Providing new genetic material by gene duplications can facilitate evolution of novel traits, because the gene duplicates are initially redundant and therefore are free to evolve new functions (Innan and Kondrashov, 2010; Kondrashov et al., 2002; Ohno, 1970). In vertebrates extensive gene and genome duplication are in the focus of research for many years (Baskaran et al., 2015; Holland, 2013; Holland et al., 1994; Kasahara, 2013; Marlétaz et al., 2015). Although arthropods are the phylum with the largest species number displaying an enormous morphological diversity, only few examples of gene duplications are known from this phylum. Gene duplications similar in magnitude to those observed in vertebrates are so far only known in the clade of the chelicerates (Di et al., 2015; Nossa et al., 2014; Sharma et al., 2014). These gene duplications have been linked to the evolution of the repertoire of silk and venom genes in spiders (Haney et al., 2016), and to specific sensory organs in scorpions.

A general course after duplication of genes is loss, sub- or neofunctionalisation (Conant and Wolfe, 2008). The study of gene paralogs during my study showed that paralogs within one species were usually less similar than the orthologs between different spiders. Most intensively studied were the duplications of the leg gap genes and their expression and function in the walking legs. Also this study clearly shows that one gene homolog kept the original gene function. The paralog, however, was free to differentiate in expression and function after the duplication event. Most of the studied genes were again conserved after neo- or subfunctionalization due to the importance of their new function. *hth2* in contrast showed differing expression patterns in the walking legs of all four spider studied so far. Thus also differences in the conservation of duplicated genes might represent a playground for the evolution to develop novel traits. For all duplicated genes of the two spiders we found that the duplication event evolved independently from the duplication events in the vertebrates. In addition, the phylogenetic analyses revealed that the timing of gene duplications in the spiders differs between the studied genes, indicating that they were duplicated in several waves, instead of a single genome

duplication. Due to the technical advances in sequencing technology, the number of available transcriptomes and genomes is increasing steadily and provides the basis for more detailed studies of duplication events in the group of chelicerates. However, gene predictions by automatic annotation based on blast searches often lead to wrongly predicted genes and detailed phylogenetic analysis are needed to clearly identify genuine homologs. The duplication of the *dachshund* gene was found to be older than the split between scorpions and spiders and is linked with the evolution of a novel leg segment, the patella. The duplication of *hth* in contrast was found to be spider specific, which is astonishing because of the diversified expression of *hth2*. The protein sequences after duplication of *exd*, however, were shown to be so similar that they caused a polytomy even after intensive phylogenetic studies. These findings show that also the time-point of the duplication might be as diverse as the conservation and the use of the duplicates.

In summary our data show how important gene duplications are in terms of evolving diversity of morphologic traits and that it is of significant importance to study more duplication events in more depth to gain insight into how they might have been used in the evolution of diversification. In addition, this shows that the astonishing diversity in the animal kingdom presumably is caused by a similar diverse use of molecular diversity rather than by simple changes in the protein coding regions or on the regulatory level.

8. References

Abu-Shaar, M., and Mann, R.S. (1998). Generation of multiple antagonistic domains along the proximodistal axis during *Drosophila* leg development. *Development* 125, 3821–3830.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000). The Genome Sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195.

Akiyama-Oda, Y. (2003). Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. *Development* 130, 1735–1747.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106.

Angelini, D.R., and Kaufman, T.C. (2005). Insect appendages and comparative ontogenetics. *Dev. Biol.* 286, 57–77.

Angelini, D.R., Liu, P.Z., Hughes, C.L., and Kaufman, T.C. (2005). Hox gene function and interaction in the milkweed bug *Oncopeltus fasciatus* (Hemiptera). *Dev. Biol.* 287, 440–455.

Angelini, D.R., Kikuchi, M., and Jockusch, E.L. (2009). Genetic patterning in the adult capitata antenna of the beetle *Tribolium castaneum*. *Dev. Biol.* 327, 240–251.

Angelini, D.R., Smith, F.W., and Jockusch, E.L. (2012). Extent With Modification: Leg Patterning in the Beetle *Tribolium castaneum* and the Evolution of Serial Homologs. *Genes|Genomes|Genetics* 2, 235–248.

Armisen, D., Nagui Refki, P., Crumière, A.J.J., Viala, S., Toubiana, W., and Khila, A. (2015). Predator strike shapes antipredator phenotype through new genetic interactions in water striders. *Nat. Commun.* 6, 8153.

Baanannou, A., Mojica-Vazquez, L.H., Darras, G., Couderc, J.-L., Cribbs, D.L., Boube, M., and Bourbon, H.-M. (2013). *Drosophila* Distal-less and Rotund Bind a Single Enhancer Ensuring Reliable and Robust bric-a-brac2 Expression in Distinct Limb Morphogenetic Fields. *PLoS Genet.* 9, e1003581.

Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28, 45–48.

Barnett, A.A., and Thomas, R.H. (2013). The expression of limb gap genes in the mite *Archezogetes longisetosus* reveals differential patterning mechanisms in chelicerates:

Archezogozetes longisetosus reveals differential patterning mechanisms in chelicerates. *Evol. Dev.* *15*, 280–292.

Baskaran, P., Rödelberger, C., Prabh, N., Seroby, V., Markov, G.V., Hirsekorn, A., and Dieterich, C. (2015). Ancient gene duplications have shaped developmental stage-specific expression in *Pristionchus pacificus*. *BMC Evol. Biol.* *15*, 185.

Beermann, A., and Schröder, R. (2004). Functional stability of the aristaless gene in appendage tip formation during evolution. *Dev. Genes Evol.* *214*.

Bell, S.M., Schreiner, C.M., Waclaw, R.R., Campbell, K., Potter, S.S., and Scott, W.J. (2003). Sp8 is crucial for limb outgrowth and neuropore closure. *Proc. Natl. Acad. Sci.* *100*, 12195–12200.

Bennett, R.L., Brown, S.J., and Denell, R.E. (1999). Molecular and genetic analysis of the *Tribolium Ultrabithorax* ortholog, *Ultrathorax*. *Dev. Genes Evol.* *209*, 608–619.

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., and Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. *Genes Dev.* *13*, 946–953.

Bishop, S.A., Klein, T., Arias, A.M., and Couso, J.P. (1999). Composite signalling from Serrate and Delta establishes leg segments in *Drosophila* through Notch. *Development* *126*, 2993–3003.

Bitsch, J. (2001). The hexapod appendage : Basic structure, development and origin. In *Annales de La Société Entomologique de France, (Société entomologique de France)*, *37*, 175–193.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* *30*, 2114–2120.

Bolker, J.A. (1995). Model systems in developmental biology. *BioEssays* *17*, 451–455.

Borner, J., Rehm, P., Schill, R.O., Ebersberger, I., and Burmester, T. (2014). A transcriptome approach to ecdysozoan phylogeny. *Mol. Phylogenet. Evol.* *80*, 79–87.

Bourgouin, C., Lundgren, S.E., and Thomas, J.B. (1992). Apterous is a *Drosophila* LIM domain gene required for the development of a subset of embryonic muscles. *Neuron* *9*, 549–561.

Bowser, J., and Reddy, A. s. n. (1997). Localization of a kinesin-like calmodulin-binding protein in dividing cells of *Arabidopsis* and tobacco. *Plant J.* *12*, 1429–1437.

Bryantsev, A.L., Duong, S., Brunetti, T.M., Chechenova, M.B., Lovato, T.L., Nelson, C., Shaw, E., Uhl, J.D., Gebelein, B., and Cripps, R.M. (2012). Extradenticle and Homothorax Control Adult Muscle Fiber Identity in *Drosophila*. *Dev. Cell* *23*, 664–673.

- Campbell, G. (2005). Regulation of gene expression in the distal region of the *Drosophila* leg by the Hox11 homolog, C15. *Dev. Biol.* 278, 607–618.
- Campbell, G., and Tomlinson, A. (1998). The roles of the homeobox genes *aristaless* and *Distal-less* in patterning the legs and wings of *Drosophila*. *Development* 125, 4483–4493.
- Cande, J.D., Chopra, V.S., and Levine, M. (2009). Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, *Tribolium castaneum*. *Development* 136, 3153–3160.
- Caravas, J., and Friedrich, M. (2010). Of mites and millipedes: Recent progress in resolving the base of the arthropod tree. *BioEssays* 32, 488–495.
- Casares, F., and Mann, R.S. (2001). The Ground State of the Ventral Appendage in *Drosophila*. *Science* 293, 1477–1480.
- Celá, P., Balková, S.M., Bryjová, A., Horáková, D., Míšek, I., Richman, J.M., and Buchtová, M. (2013). Expression, function and regulation of *Evi-1* during embryonic avian development. *Gene Expr. Patterns* 13, 343–353.
- Celis, J.F. de, Tyler, D.M., Celis, J. de, and Bray, S.J. (1998). Notch signalling mediates segmentation of the *Drosophila* leg. *Development* 125, 4617–4626.
- Chung, S., Chavez, C., and Andrew, D.J. (2011). *Trachealess* (*Trh*) regulates all tracheal genes during *Drosophila* embryogenesis. *Dev. Biol.* 360, 160–172.
- Claparède, R. (1862). *Recherches sur l'évolution des araignées*. (Natuurkundige Verhandelingen uitgegeven door het Provinciaal Utrechtsch Genootschap von Kunsten en Wetenschappen. Deel I. Stuk 1. C. van der Post Jr., Utrecht.).
- Cohen, S.M. (1993). Imaginal disc development. In *The Development of Drosophila Melanogaster* 2, pp. 747–841.
- Cohen, S.M., and Jürgens, G. (1989). Proximal—distal pattern formation in *Drosophila*: cell autonomous requirement for *Distal-less* gene activity in limb development. *EMBO J.* 8, 2045–2055.
- Cohen, B., McGuffin, M.E., Pfeifle, C., Segal, D., and Cohen, S.M. (1992). *apterous*, a gene required for imaginal disc development in *Drosophila* encodes a member of the LIM family of developmental regulatory proteins. *Genes Dev.* 6, 715–729.
- Cohen, B., Simcox, A.A., and Cohen, S.M. (1993). Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. *Development* 117, 597–608.
- Cohen, S.M., Brönner, G., Küttner, F., Jürgens, G., and Jäckle, H. (1989). *Distal-less* encodes a homoeodomain protein required for limb development in *Drosophila*. *Nature* 338, 432–434.

- Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* *9*, 938–950.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* *21*, 3674–3676.
- Consortium, T.U. (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* *43*, D204–D212.
- Córdoba, S., and Estella, C. (2014). The bHLH-PAS Transcription Factor Dysfusion Regulates Tarsal Joint Formation in Response to Notch Activity during *Drosophila* Leg Development. *PLoS Genet.* *10*, e1004621.
- Crews, S.T. (1998). Control of cell lineage-specific development and transcription by bHLH–PAS proteins. *Genes Dev.* *12*, 607–620.
- Crews, S.T., and Fan, C.-M. (1999). Remembrance of things PAS: regulation of development by bHLH–PAS proteins. *Curr. Opin. Genet. Dev.* *9*, 580–587.
- Dai, M., Wang, Y., Fang, L., Irwin, D.M., Zhu, T., Zhang, J., Zhang, S., and Wang, Z. (2014). Differential Expression of *Meis2*, *Mab21l2* and *Tbx3* during Limb Development Associated with Diversification of Limb Morphology in Mammals. *PLoS ONE* *9*, e106100.
- Damen, W.G.M. (2007). Evolutionary conservation and divergence of the segmentation process in arthropods. *Dev. Dyn.* *236*, 1379–1391.
- Damen, W.G., Saridaki, T., and Averof, M. (2002). Diverse adaptations of an ancestral gill: a common evolutionary origin for wings, breathing organs, and spinnerets. *Curr. Biol.* *12*, 1711–1716.
- Damen, W.G., Janssen, R., and Prpic, N.-M. (2005). Pair rule gene orthologs in spider segmentation. *Evol. Dev.* *7*, 618–628.
- Davidson, N.M., and Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* *15*, 410.
- Dear, T.N., and Rabbitts, T.H. (1994). A *Drosophila melanogaster* homologue of the T-cell oncogene *HOX11* localises to a cluster of homeobox genes. *Gene* *141*, 225–229.
- Dear, T.N., Sanchez-Garcia, I., and Rabbitts, T.H. (1993). The *HOX11* gene encodes a DNA-binding nuclear transcription factor belonging to a distinct family of homeobox genes. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 4431–4435.
- Di, Z., Yu, Y., Wu, Y., Hao, P., He, Y., Zhao, H., Li, Y., Zhao, G., Li, X., Li, W., et al. (2015). Genome-wide analysis of homeobox genes from *Mesobuthus martensii* reveals Hox gene duplication in scorpions. *Insect Biochem. Mol. Biol.* *61*, 25–33.

- Diogo, R. (2015). Where is the evo in Evo-Devo (evolutionary developmental biology)? *J. Exp. Zool. B Mol. Dev. Evol.* n/a – n/a.
- Doeffinger, C., Hartenstein, V., and Stollewerk, A. (2010). Compartmentalisation of the precheliceral neuroectoderm in the spider *Cupiennius salei*: Development of the arcuate body, the optic ganglia and the mushroom body. *J. Comp. Neurol.* 518, 2612–2632.
- Dong, P.D.S., Dicks, J.S., and Panganiban, G. (2002). Distal-less and homothorax regulate multiple targets to pattern the *Drosophila* antenna. *Development* 129, 1967–1974.
- Dong, Y., Dai, F., Ren, Y., Liu, H., Chen, L., Yang, P., Liu, Y., Li, X., Wang, W., and Xiang, H. (2015). Comparative transcriptome analyses on silk glands of six silkmoths imply the genetic basis of silk structure and coloration. *BMC Genomics* 16.
- Dymek, E.E., Goduti, D., Kramer, T., and Smith, E.F. (2006). A kinesin-like calmodulin-binding protein in *Chlamydomonas*: evidence for a role in cell division and flagellar functions. *J. Cell Sci.* 119, 3107–3116.
- Edgecombe, G.D. (2010). Arthropod phylogeny: An overview from the perspectives of morphology, molecular data and the fossil record. *Arthropod Struct. Dev.* 39, 74–87.
- Emmons, R.B., Duncan, D., Estes, P.A., Kiefel, P., Mosher, J.T., Sonnenfeld, M., Ward, M.P., Duncan, I., and Crews, S.T. (1999). The spineless-aristapedia and tango bHLH-PAS proteins interact to control antennal and tarsal development in *Drosophila*. *Dev. Camb. Engl.* 126, 3937–3945.
- Estella, C., and Mann, R.S. (2008). Logic of Wg and Dpp induction of distal and medial fates in the *Drosophila* leg. *Development* 135, 627–636.
- Estella, C., and Mann, R.S. (2010). Non-Redundant Selector and Growth-Promoting Functions of Two Sister Genes, buttonhead and Sp1, in *Drosophila* Leg Development. *PLoS Genet.* 6, e1001001.
- Estella, C., Rieckhof, G., Calleja, M., and Morata, G. (2003). The role of buttonhead and Sp1 in the development of the ventral imaginal discs of *Drosophila*. *Development* 130, 5929–5941.
- Estella, C., Voutev, R., and Mann, R.S. (2012). A Dynamic Network of Morphogens and Transcription Factors Patterns the Fly Leg. *Curr. Top. Dev. Biol.* 98, 173–198.
- Foelix, R.F. (1996). *Biology of spiders* (Oxford: Oxford University Press.).
- Franch-Marro, X., Martín, N., Averof, M., and Casanova, J. (2006). Association of tracheal placodes with leg primordia in *Drosophila* and implications for the origin of insect tracheal systems. *Dev. Camb. Engl.* 133, 785–790.

- Friedrich, M., and Tautz, D. (1995). Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* 376, 165–167.
- Fristrom, D.F.A.J.W., and Fristrom, J.W. (1993). The metamorphic development of the adult epidermis. In *The Development of Drosophila Melanogaster* 2, pp. 843–897.
- Galindo, M.I. (2002). Leg Patterning Driven by Proximal-Distal Interactions and EGFR Signaling. *Science* 297, 256–259.
- Galindo, M. i., Bishop, S. a., and Couso, J. p. (2005). Dynamic EGFR-Ras signalling in *Drosophila* leg development. *Dev. Dyn.* 233, 1496–1508.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80.
- Gilad, Y., Pritchard, J.K., and Thornton, K. (2009). Characterizing natural variation using next-generation sequencing technologies. *Trends Genet.* 25, 463–471.
- Giorgianni, M.W., and Mann, R.S. (2011). Establishment of Medial Fates along the Proximodistal Axis of the *Drosophila* Leg through Direct Activation of *dachshund* by *Distalless*. *Dev. Cell* 20, 455–468.
- Giribet, G., Edgecombe, G.D., and Wheeler, W.C. (2001). Arthropod phylogeny based on eight molecular loci and morphology. *Nature* 413, 157–161.
- Godt, D., Couderc, J.L., Cramton, S.E., and Laski, F.A. (1993). Pattern formation in the limbs of *Drosophila*: *bric a brac* is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development* 119, 799–812.
- Gonzalez-Crespo, S., and Morata, G. (1995). Control of *Drosophila* adult pattern by *extradenticle*. *Development* 121, 2117–2125.
- Goto, S., and Hayashi, S. (1997). Specification of the embryonic limb primordium by graded activity of *Decapentaplegic*. *Development* 124, 125–132.
- Goyama, S., Yamamoto, G., Shimabe, M., Sato, T., Ichikawa, M., Ogawa, S., Chiba, S., and Kurokawa, M. (2008). *Evi-1* Is a Critical Regulator for Hematopoietic Stem Cells and Transformed Leukemic Cells. *Cell Stem Cell* 3, 207–220.
- Grossmann, D., and Prpic, N.-M. (2012). *Egfr* signaling regulates distal as well as medial fate in the embryonic leg of *Tribolium castaneum*. *Dev. Biol.* 370, 264–272.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence

reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* *8*, 1494–1512.

Haney, R.A., Clarke, T.H., Gadgil, R., Fitzpatrick, R., Hayashi, C.Y., Ayoub, N.A., and Garb, J.E. (2016). Effects of gene duplication, positive selection and shifts in gene expression on the evolution of the venom gland transcriptome in widow spiders. *Genome Biol. Evol.* *evv253*.

Hao, I., Green, R.B., Dunaevsky, O., Lengyel, J.A., and Rauskolb, C. (2003). The odd-skipped family of zinc finger genes promotes *Drosophila* leg segmentation. *Dev. Biol.* *263*, 282–295.

Harding, K., Rushlow, C., Doyle, H.J., Hoey, T., and Levine, M. (1986). Cross-regulatory interactions among pair-rule genes in *Drosophila*. *Science* *233*, 953–959.

Hasegawa, E., Kitada, Y., Kaido, M., Takayama, R., Awasaki, T., Tabata, T., and Sato, M. (2011). Concentric zones, cell migration and neuronal circuits in the *Drosophila* visual center. *Development* *138*, 983–993.

Hilbrant, M., Damen, W.G.M., and McGregor, A.P. (2012). Evolutionary crossroads in developmental biology: the spider *Parasteatoda tepidariorum*. *Development* *139*, 2655–2662.

Hinai, A.A., and Valk, P.J.M. (2016). Review: Aberrant EVI1 expression in acute myeloid leukaemia. *Br. J. Haematol.* n/a – n/a.

Hirai, H. (1999). The transcription factor Evi-1. *Int. J. Biochem. Cell Biol.* *31*, 1367–1371.

Holland, L.Z. (2013). Evolution of new characters after whole genome duplications: insights from amphioxus. *Semin. Cell Dev. Biol.* *24*, 101–109.

Holland, P.W., Garcia-Fernàndez, J., Williams, N.A., and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. *Development* *1994*, 125–133.

Holm, Å. (1940). Studien über die Entwicklung und Entwicklungsbiologie der Spinnen. *Zoologiska Bidrag från Uppsala* *19*, 1–214. Plates 1–11.

Huang, X., Powell-Coffman, J.A., and Jin, Y. (2004). The AHR-1 aryl hydrocarbon receptor and its co-factor the AHA-1 aryl hydrocarbon receptor nuclear translocator specify GABAergic neuron cell fate in *C. elegans*. *Development* *131*, 819–828.

Hughes, C.L., and Kaufman, T.C. (2002a). Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede. *Development* *129*, 1225–1238.

Hughes, C.L., and Kaufman, T.C. (2002b). Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* *4*, 459–499.

- Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* *11*, 4.
- Isaac, D.D., and Andrew, D.J. (1996). Tubulogenesis in *Drosophila*: a requirement for the tracheless gene product. *Genes Dev.* *10*, 103–117.
- Jacobs, Y., Schnabel, C.A., and Cleary, M.L. (1999). Trimeric Association of Hox and TALE Homeodomain Proteins Mediates Hoxb2 Hindbrain Enhancer Activity. *Mol. Cell. Biol.* *19*, 5134–5142.
- Janssen, R., Eriksson, B.J., Budd, G.E., Akam, M., and Prpic, N.-M. (2010). Gene expression patterns in an onychophoran reveal that regionalization predates limb segmentation in pan-arthropods: Appendage patterning in onychophorans. *Evol. Dev.* *12*, 363–372.
- Janssen, R., Schönauer, A., Weber, M., Turetzek, N., Hogvall, M., Goss, G.E., Patel, N.H., McGregor, A.P., and Hilbrant, M. (2015). The evolution and expression of panarthropod frizzled genes. *Evol. Dev. Biol.* *3*, 96.
- Jaw, T.J., You, L.-R., Knoepfler, P.S., Yao, L.-C., Pai, C.-Y., Tang, C.-Y., Chang, L.-P., Berthelsen, J., Blasi, F., Kamps, M.P., et al. (2000). Direct interaction of two homeoproteins, Homothorax and Extradenticle, is essential for EXD nuclear localization and function. *Mech. Dev.* *91*, 279–291.
- Johnston, D.S., and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* *68*, 201–219.
- Kasahara, M. (2013). Impact of whole-genome duplication on vertebrate development and evolution. *Semin. Cell Dev. Biol.* *24*, 81–82.
- Kawakami, Y., Esteban, C.R., Matsui, T., Rodríguez-León, J., Kato, S., and Belmonte, J.C.I. (2004). Sp8 and Sp9, two closely related buttonhead-like transcription factors, regulate Fgf8 expression and limb outgrowth in vertebrate embryos. *Development* *131*, 4763–4774.
- Kensley, and Schotte (1989). *GT MARINE ISOPOD CRUSTACEA* (Smithsonian Inst Press).
- Khadjeh, S., Turetzek, N., Pechmann, M., Schwager, E.E., Wimmer, E.A., Damen, W.G., and Prpic, N.-M. (2012). Divergent role of the Hox gene Antennapedia in spiders is responsible for the convergent evolution of abdominal limb repression. *Proc. Natl. Acad. Sci.* *109*, 4921–4926.
- Kim, M.D., Jan, L.Y., and Jan, Y.N. (2006). The bHLH-PAS protein Spineless is necessary for the diversification of dendrite morphology of *Drosophila* dendritic arborization neurons. *Genes Dev.* *20*, 2806–2819.

References

- Kim, S.-N., Jung, K.I., Chung, H.-M., Kim, S.H., and Jeon, S.-H. (2008). The pleiohomeotic gene is required for maintaining expression of genes functioning in ventral appendage formation in *Drosophila melanogaster*. *Dev. Biol.* *319*, 121–129.
- Kojima, T. (2004). The mechanism of *Drosophila* leg development along the proximodistal axis. *Dev. Growth Differ.* *46*, 115–129.
- Kojima, T., Ishimaru, S., Higashijima, S., Takayama, E., Akimaru, H., Sone, M., Emori, Y., and Saigo, K. (1991). Identification of a different-type homeobox gene, BarH1, possibly causing Bar (B) and Om(1D) mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 4343–4347.
- Kojima, T., Sato, M., and Saigo, K. (2000). Formation and specification of distal leg segments in *Drosophila* by dual Bar homeobox genes, BarH1 and BarH2. *Development* *127*, 769–778.
- Kojima, T., Tsuji, T., and Saigo, K. (2005). A concerted action of a paired-type homeobox gene, aristaless, and a homolog of Hox11/tlx homeobox gene, clawless, is essential for the distal tip development of the *Drosophila* leg. *Dev. Biol.* *279*, 434–445.
- Konantz, M., Grzywna, S., Pereboom, T.C., Carroll, K.J., Esain, V., Kanz, L., North, T.E., and Lengerke, C. (2014). Multiple Roles for the Zebrafish Homologue of the Murine Evi1 Gene during Primitive Myelopoiesis and HSC Development. *Blood* *124*, 2901–2901.
- Konantz, M., Müller, J., Lenard, A., Esain, V., North, T.E., and Lengerke, C. (2015). The zebrafish homologue of the murine EVI1 gene critically regulates hsc development. *Exp. Hematol.* *43*, S74.
- Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I., and Koonin, E.V. (2002). Selection in the evolution of gene duplications. *Genome Biol* *3*, 8–1.
- Kozu, S., Tajiri, R., Tsuji, T., Michiue, T., Saigo, K., and Kojima, T. (2006). Temporal regulation of late expression of Bar homeobox genes during *Drosophila* leg development by Spineless, a homolog of the mammalian dioxin receptor. *Dev. Biol.* *294*, 497–508.
- Kubota, K., Goto, S., Eto, K., and Hayashi, S. (2000). EGF receptor attenuates Dpp signaling and helps to distinguish the wing and leg cell fates in *Drosophila*. *Development* *127*, 3769–3776.
- Kuzin, B.A., Nikitina, E.A., Cherezov, R.O., Vorontsova, J.E., Slezinger, M.S., Zatsepina, O.G., Simonova, O.B., Enikolopov, G.N., and Savvateeva-Popova, E.V. (2014). Combination of Hypomorphic Mutations of the *Drosophila* Homologues of Aryl Hydrocarbon Receptor and Nucleosome Assembly Protein Family Genes Disrupts Morphogenesis, Memory and Detoxification. *PLoS ONE* *9*, e94975.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.

- Lecuit, T., and Cohen, S.M. (1997). Proximal–distal axis formation in the *Drosophila* leg. *Nature* *388*, 139–145.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
- Li, Y., Cheng, C.N., Verdun, V.A., and Wingert, R.A. (2014). Zebrafish nephrogenesis is regulated by interactions between retinoic acid, mecom, and Notch signaling. *Dev. Biol.* *386*, 111–122.
- Lickteig, K.M., Duerr, J.S., Frisby, D.L., Hall, D.H., Rand, J.B., and Miller, D.M. (2001). Regulation of Neurotransmitter Vesicles by the Homeodomain Protein UNC-4 and Its Transcriptional Corepressor UNC-37/Groucho in *Caenorhabditis elegans* Cholinergic Motor Neurons. *J. Neurosci.* *21*, 2001–2014.
- Lilly, B., O’Keefe, D.D., Thomas, J.B., and Botas, J. (1999). The LIM homeodomain protein dLim1 defines a subclass of neurons within the embryonic ventral nerve cord of *Drosophila*. *Mech. Dev.* *88*, 195–205.
- Liu, T., Yu, L., Liu, L., Li, H., Li, Y., Liu, T., Yu, L., Liu, L., Li, H., and Li, Y. (2015). Comparative Transcriptomes and EVO-DEVO Studies Depending on Next Generation Sequencing. *Comput. Math. Methods Med. Comput. Math. Methods Med.* *2015*, *2015*, e896176.
- Mann, R.S., and Affolter, M. (1998). Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* *8*, 423–429.
- Mardon, G., Solomon, N.M., and Rubin, G.M. (1994). dachshund encodes a nuclear protein required for normal eye and leg development in *Drosophila*. *Development* *120*, 3473–3486.
- Marlétaz, F., Maeso, I., Faas, L., Isaacs, H.V., and Holland, P.W. (2015). Cdx ParaHox genes acquired distinct developmental roles after gene duplication in vertebrate evolution. *BMC Biol.* *13*, 56.
- McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* *68*, 283–302.
- McGregor, A.P., Hilbrant, M., Pechmann, M., Schwager, E.E., Prpic, N.-M., and Damen, W.G.M. (2008). *Cupiennius salei* and *Achaearanea tepidariorum*: Spider models for investigating evolution and development. *BioEssays* *30*, 487–498.
- McKay, D.J., Estella, C., and Mann, R.S. (2009). The origins of the *Drosophila* leg revealed by the cis-regulatory architecture of the *Distalless* gene. *Development* *136*, 61–71.

- Miller, D.M., Shen, M.M., Shamu, C.E., Bürglin, T.R., Ruvkun, G., Dubois, M.L., Ghee, M., and Wilson, L. (1992). *C. elegans unc-4* gene encodes a homeodomain protein that determines the pattern of synaptic input to specific motor neurons. *Nature* 355, 841–845.
- Mitchell, B., and Crews, S.T. (2002). Expression of the *Artemia tracheless* gene in the salt gland and epipod. *Evol. Dev.* 4, 344–353.
- Mito, T., Shinmyo, Y., Kurita, K., Nakamura, T., Ohuchi, H., and Noji, S. (2011). Ancestral functions of Delta/Notch signaling in the formation of body and leg segments in the cricket *Gryllus bimaculatus*. *Development* 138, 3823–3833.
- Mittmann, B., and Wolff, C. (2012). Embryonic development and staging of the cobweb spider *Parasteatoda tepidariorum* C. L. Koch, 1841 (syn.: *Achaearanea tepidariorum*; *Araneomorphae*; *Theridiidae*). *Dev. Genes Evol.* 222, 189–216.
- Miyawaki, K., Inoue, Y., Mito, T., Fujimoto, T., Matsushima, K., Shinmyo, Y., Ohuchi, H., and Noji, S. (2002). Expression patterns of *aristaless* in developing appendages of *Gryllus bimaculatus* (cricket). *Mech. Dev.* 113, 181–184.
- Miyazono, K., Zhi, Y., Takamura, Y., Nagata, K., Saigo, K., Kojima, T., and Tanokura, M. (2010). Cooperative DNA-binding and sequence-recognition mechanism of *aristaless* and *clawless*. *EMBO J.* 29, 1613–1623.
- Mucenski, M.L., Taylor, B.A., Ihle, J.N., Hartley, J.W., Morse, H.C., Jenkins, N.A., and Copeland, N.G. (1988). Identification of a common ecotropic viral integration site, *Evi-1*, in the DNA of AKXD murine myeloid tumors. *Mol. Cell. Biol.* 8, 301–308.
- Nakamura, T., Mito, T., Miyawaki, K., Ohuchi, H., and Noji, S. (2008). EGFR signaling is required for re-establishing the proximodistal axis during distal leg regeneration in the cricket *Gryllus bimaculatus* nymph. *Dev. Biol.* 319, 46–55.
- Neuwirth, E. (2011). *RColorBrewer: ColorBrewer palettes*. R package version 1.5.
- Nossa, C.W., Havlak, P., Yue, J.-X., Lv, J., Vincent, K.Y., Brockmann, H.J., and Putnam, N.H. (2014). Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *GigaScience* 3, 9.
- Ohno, S. (1970). *Evolution by Gene Duplication*. (New York, Heidelberg, Berlin: Springer-Verlag.).
- Oliveira, M.B., Liedholm, S.E., Lopez, J.E., Lochte, A.A., Pazio, M., Martin, J.P., Mörch, P.R., Salakka, S., York, J., Yoshimoto, A., et al. (2014). Expression of arthropod distal limb-patterning genes in the onychophoran *Euperipatoides kanangrensis*. *Dev. Genes Evol.* 224, 87–96.

- Oppenheim, S.J., Baker, R.H., Simon, S., and DeSalle, R. (2015). We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. *Insect Mol. Biol.* *24*, 139–154.
- Pai, C.-Y., Kuo, T.-S., Jaw, T.J., Kurant, E., Chen, C.-T., Bessarab, D.A., Salzberg, A., and Sun, Y.H. (1998). The Homothorax homeoprotein activates the nuclear localization of another homeoprotein, Extradenticle, and suppresses eye development in *Drosophila*. *Genes Dev.* *12*, 435–446.
- Panganiban, G., Irvine, S.M., Lowe, C., Roehl, H., Corley, L.S., Sherbon, B., Grenier, J.K., Fallon, J.F., Kimble, J., Walker, M., et al. (1997). The origin and evolution of animal appendages. *Proc. Natl. Acad. Sci.* *94*, 5162–5166.
- Pantalacci, S., and Sémon, M. (2014). Transcriptomics of developing embryos and organs: A raising tool for evo–devo. *J. Exp. Zoolog. B Mol. Dev. Evol.* *00B*, 1–9.
- Pechmann, M. (2011). Comparative studies of appendage and segment formation in arthropods. PhD Thesis. Georg-August-University Göttingen.
- Pechmann, M., and Prpic, N.-M. (2009). Appendage patterning in the South American bird spider *Acanthoscurria geniculata* (Araneae: Mygalomorphae). *Dev. Genes Evol.* *219*, 189–198.
- Pechmann, M., McGregor, A.P., Schwager, E.E., Feitosa, N.M., and Damen, W.G.M. (2009). Dynamic gene expression is required for anterior regionalization in a spider. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 1468–1472.
- Pechmann, M., Khadjeh, S., Sprenger, F., and Prpic, N.-M. (2010). Patterning mechanisms and morphological diversity of spider appendages and their importance for spider evolution. *Arthropod Struct. Dev.* *39*, 453–467.
- Pechmann, M., Khadjeh, S., Turetzek, N., McGregor, A.P., Damen, W.G.M., and Prpic, N.-M. (2011). Novel Function of Distal-less as a Gap Gene during Spider Segmentation. *PLoS Genet.* *7*, e1002342.
- Peel, A.D., Chipman, A.D., and Akam, M. (2005). Arthropod Segmentation: beyond the *Drosophila* paradigm. *Nat. Rev. Genet.* *6*, 905–916.
- Perkins, A.S., Mercer, J.A., Jenkins, N.A., and Copeland, N.G. (1991). Patterns of Evi-1 expression in embryonic and adult tissues suggest that Evi-1 plays an important regulatory role in mouse development. *Development* *111*, 479–487.
- Posnien, N., Zeng, V., Schwager, E.E., Pechmann, M., Hilbrant, M., Keefe, J.D., Damen, W.G.M., Prpic, N.-M., McGregor, A.P., and Extavour, C.G. (2014). A Comprehensive Reference Transcriptome Resource for the Common House Spider *Parasteatoda tepidariorum*. *PLoS ONE* *9*, e104885.

References

- Prpic, N.-M., and Damen, W.G.M. (2009). Notch-mediated segmentation of the appendages is a molecular phylotypic trait of the arthropods. *Dev. Biol.* *326*, 262–271.
- Prpic, N.-M., and Damen, W.M. (2004). Expression patterns of leg genes in the mouthparts of the spider *Cupiennius salei* (Chelicerata: Arachnida). *Dev. Genes Evol.* *214*, 296–302.
- Prpic, N.M.P., and Damen, W.G.M. (2008). Arthropod appendages: a prime example for the evolution of morphological diversity and innovation. In *Evolving Pathways: Key Themes in Evolutionary Developmental Biology.*, (Cambridge University Press, Cambridge), pp. 381–398.
- Prpic, N.-M., and Tautz, D. (2003). The expression of the proximodistal axis patterning genes *Distal-less* and *dachshund* in the appendages of *Glomeris marginata* (Myriapoda: Diplopoda) suggests a special role of these genes in patterning the head appendages. *Dev. Biol.* *260*, 97–112.
- Prpic, N.-M., and Telford, M.J. (2008). Expression of homothorax and extradenticle mRNA in the legs of the crustacean *Parhyale hawaiiensis*: evidence for a reversal of gene expression regulation in the pancrustacean lineage. *Dev. Genes Evol.* *218*, 333–339.
- Prpic, N.-M., Wigand, B., Damen, W., and Klingler, M. (2001). Expression of *dachshund* in wild-type and *Distal-less* mutant *Tribolium* corroborates serial homologies in insect appendages. *Dev. Genes Evol.* *211*, 467–477.
- Prpic, N.-M., Janssen, R., Wigand, B., Klingler, M., and Damen, W.G.. (2003). Gene expression in spider appendages reveals reversal of *exd/hth* spatial specificity, altered leg gap gene dynamics, and suggests divergent distal morphogen signaling. *Dev. Biol.* *264*, 119–140.
- Prpic, N.-M., Schoppmeier, M., and Damen, W.G.M. (2008a). Whole-mount in situ hybridization of spider embryos. *CSH Protoc.* *2008*, pdb.prot5068.
- Prpic, N.-M., Schoppmeier, M., and Damen, W.G.M. (2008b). Collection and Fixation of Spider Embryos. *Cold Spring Harb. Protoc.* *2008*, pdb.prot5067 – pdb.prot5067.
- Pueyo, J.I. (2004). Chip-mediated partnerships of the homeodomain proteins *Bar* and *Aristaless* with the LIM-HOM proteins *Apterous* and *Lim1* regulate distal leg development. *Development* *131*, 3107–3120.
- Pueyo, J.I., and Couso, J.P. (2008). The 11-aminoacid long *Tarsal-less* peptides trigger a cell signal in *Drosophila* leg development. *Dev. Biol.* *324*, 192–201.
- Pueyo, J.I., Galindo, M.I., Bishop, S.A., and Couso, J.P. (2000). Proximal-distal leg development in *Drosophila* requires the *apterous* gene and the *Lim1* homologue *dlim1*. *Development* *127*, 5391–5402.

- Rauskolb, C., and Irvine, K.D. (1999). Notch-Mediated Segmentation and Growth Control of the *Drosophila* Leg. *Dev. Biol.* *210*, 339–350.
- Rauskolb, C., Smith, K.M., Peifer, M., and Wieschaus, E. (1995). extradenticle determines segmental identities throughout *Drosophila* development. *Development* *121*, 3663–3673.
- Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., and Cunningham, C.W. (2010). Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* *463*, 1079–1083.
- Reig, G., Cabrejos, M.E., and Concha, M.L. (2007). Functions of BarH transcription factors during embryonic development. *Dev. Biol.* *302*, 367–375.
- Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-Shaar, M., and Mann, R.S. (1997). Nuclear Translocation of Extradenticle Requires homothorax, which Encodes an Extradenticle-Related Homeodomain Protein. *Cell* *91*, 171–183.
- Roberts, M.J. (1995). Collins Field guide SPIDERS of britian and Nothern europe (London: Harper Collins Publishers).
- Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* *19*, 1572–1574.
- Rose, M.R. (1982). Antagonistic pleiotropy, dominance, and genetic variation. *Heredity* *48*, 63–78.
- Rota-Stabelli, O., Campbell, L., Brinkmann, H., Edgecombe, G.D., Longhorn, S.J., Peterson, K.J., Pisani, D., Philippe, H., and Telford, M.J. (2011). A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc. R. Soc. Lond. B Biol. Sci.* *278*, 298–306.
- Sánchez-Higueras, C., Sotillos, S., and Castelli-Gair Hombría, J. (2014). Common Origin of Insect Trachea and Endocrine Organs from a Segmentally Repeated Precursor. *Curr. Biol.* *24*, 76–81.
- Sanggaard, K.W., Bechsgaard, J.S., Fang, X., Duan, J., Dyrlund, T.F., Gupta, V., Jiang, X., Cheng, L., Fan, D., Feng, Y., et al. (2014). Spider genomes provide insight into composition and evolution of venom and silk. *Nat. Commun.* *5*, 3765.
- Schneuwly, S., Kuroiwa, A., and Gehring, W.J. (1987). Molecular analysis of the dominant homeotic Antennapedia phenotype. *EMBO J.* *6*, 201–206.
- Schomburg, C., Turetzek, N., Schacht, M.I., Schneider, J., Kirfel, P., Prpic, N.-M., and Posnien, N. (2015). Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *EvoDevo* *6*, 15.

- Schwager, E.E., Schoppmeier, M., Pechmann, M., and Damen, W.G. (2007). Duplicated Hox genes in the spider *Cupiennius salei*. *Front. Zool.* *4*, 10.
- Sharma, P.P., Schwager, E.E., Extavour, C.G., and Giribet, G. (2012). Evolution of the chelicera: a *dachshund* domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata): Harvestman leg gap genes. *Evol. Dev.* *14*, 522–533.
- Sharma, P.P., Schwager, E.E., Extavour, C.G., and Wheeler, W.C. (2014). Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B Biol. Sci.* *281*, 20140661–20140661.
- Sharma, P.P., Tarazona, O.A., Lopez, D.H., Schwager, E.E., Cohn, M.J., Wheeler, W.C., and Extavour, C.G. (2015). A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proc. R. Soc. Lond. B Biol. Sci.* *282*, 20150698.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* *7*, 539.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* *31*, 3210–3212.
- Smith, F.W., Angelini, D.R., Gaudio, M.S., and Jockusch, E.L. (2014). Metamorphic labral axis patterning in the beetle *Tribolium castaneum* requires multiple upstream, but few downstream, genes in the appendage patterning network: Labral metamorphosis in *T. castaneum*. *Evol. Dev.* *16*, 78–91.
- Sonnenfeld, M., Ward, M., Nystrom, G., Mosher, J., Stahl, S., and Crews, S. (1997). The *Drosophila tango* gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development. *Dev. Camb. Engl.* *124*, 4571–4582.
- Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., and Slate, J. (2010). Adaptation genomics: the next generation. *Trends Ecol. Evol.* *25*, 705–712.
- Struhl, G. (1981). A homoeotic mutation transforming leg to antenna in *Drosophila*. *Nature* *292*, 635–638.
- Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C.H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinforma. Oxf. Engl.* *23*, 1282–1288.
- Tabuchi, K., Yoshikawa, S., Yuasa, Y., Sawamoto, K., and Okano, H. (1998). A novel *Drosophila* paired-like homeobox gene related to *Caenorhabditis elegans unc-4* is

expressed in subsets of postmitotic neurons and epidermal cells. *Neurosci. Lett.* *257*, 49–52.

Tajiri, R., Tsuji, T., Ueda, R., Saigo, K., and Kojima, T. (2007). Fate determination of *Drosophila* leg distal regions by tracheless and tango through repression and stimulation, respectively, of Bar homeobox gene expression in the future pretarsus and tarsus. *Dev. Biol.* *303*, 461–473.

Tanaka, K., and Truman, J.W. (2007). Molecular patterning mechanism underlying metamorphosis of the thoracic leg in *Manduca sexta*. *Dev. Biol.* *305*, 539–550.

Ting, C.-T., Tsaur, S.-C., Sun, S., Browne, W.E., Chen, Y.-C., Patel, N.H., and Wu, C.-I. (2004). Gene duplication and speciation in *Drosophila*: evidence from the Odysseus locus. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 12232–12235.

Tom, M., Manfrin, C., Chung, S.J., Sagi, A., Gerdol, M., Moro, G.D., Pallavicini, A., and Giulianini, P.G. (2014). Expression of cytoskeletal and molt-related genes is temporally scheduled in the hypodermis of the crayfish *Procambarus clarkii* during premolt. *J. Exp. Biol.* *217*, 4193–4202.

Treichel, D., Schöck, F., Jäckle, H., Gruss, P., and Mansouri, A. (2003). mBtd is required to maintain signaling during murine limb development. *Genes Dev.* *17*, 2630–2635.

Tsuji, T., Sato, A., Hiratani, I., Taira, M., Saigo, K., and Kojima, T. (2000). Requirements of Lim1, a *Drosophila* LIM-homeobox gene, for normal leg and antennal development. *Development* *127*, 4315–4323.

Turetzek, N. (2011). On the role of hox genes and co-factors in body plan specification in spiders. Msc Thesis. Georg-August-University Göttingen.

Turetzek, N., Pechmann, M., Schomburg, C., Schneider, J., and Prpic, N.-M. (2016). Neofunctionalization of a Duplicate dachshund Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. *Mol. Biol. Evol.* *33*, 109–121.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* *40*, e115.

Wang, J., Kean, L., Yang, J., Allan, A.K., Davies, S.A., Herzyk, P., and Dow, J.A. (2004). Function-informed transcriptome analysis of *Drosophila* renal tubule. *Genome Biol.* *5*, R69.

Whittington, H.B., and Almond, J.E. (1987). Appendages and Habits of the Upper Ordovician Trilobite *Triarthrus eatoni*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *317*, 1–46.

Wilkins, A. S. (2002). *The evolution of developmental pathways.* (Sunderland, Massachusetts, USA: Sinauer Associates Inc.).

References

- Winchell, C.J., and Jacobs, D.K. (2013). Expression of the Lhx genes *apterous* and *lim1* in an errant polychaete: implications for bilaterian appendage evolution, neural development, and muscle diversification. *EvoDevo* 4, 4.
- Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature* 430, 85–88.
- Wolff, C., and Hilbrant, M. (2011). The embryonic development of the central American wandering spider *Cupiennius salei*. *Front. Zool.* 8, 15.
- Wu, J., and Cohen, S.M. (1999). Proximodistal axis formation in the *Drosophila* leg: subdivision into proximal and distal domains by *Homothorax* and *Distal-less*. *Development* 126, 109–117.
- Wu, J., and Cohen, S.M. (2000). Proximal distal axis formation in the *Drosophila* leg: distinct functions of *teashirt* and *homothorax* in the proximal leg. *Mech. Dev.* 94, 47–56.
- Yu, Y., Fuscoe, J.C., Zhao, C., Guo, C., Jia, M., Qing, T., Bannon, D.I., Lancashire, L., Bao, W., Du, T., et al. (2014). A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat. Commun.* 5, 3230.
- Yuasa, H., Oike, Y., Iwama, A., Nishikata, I., Sugiyama, D., Perkins, A., Mucenski, M.L., Suda, T., and Morishita, K. (2005). Oncogenic transcription factor *Evi1* regulates hematopoietic stem cell proliferation through *GATA-2* expression. *EMBO J.* 24, 1976–1987.

9. References of Publications and Manuscripts

9.1 Turetzek et al. (2016). Neofunctionalization of a Duplicate *dachshund* Gene Underlies the Evolution of a Novel Leg Segment in Arachnids.

Abzhanov A, Kaufman TC. 2000. Homologs of *Drosophila* appendage genes in the patterning of arthropod limbs. *Dev Biol.* 227:673-689.

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185-2195.

Angelini DR, Kaufman TC. 2004. Functional analyses in the hemipteran *Oncopeltus fasciatus* reveal conserved and derived aspects of appendage patterning in insects. *Dev Biol.* 271:306- 321.

Angelini DR, Kaufman TC. 2005. Insect appendages and comparative ontogenetics. *Dev Biol.* 286:57-77.

Cao Z, Yu Y, Wu Y, Hao P, Di Z, He Y, Chen Z, Yang W, Shen Z, He X, et al. 2013. The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nat Commun.* 4:2602. doi: 10.1038/ncomms3602.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nature Rev Genet.* 9:938-950.

Davis RJ, Shen W, Heanue TA, Mardon G. 1999. Mouse *Dach*, a homologue of *Drosophila dachshund*, is expressed in the developing retina, brain and limbs. *Dev Genes Evol.* 209:526- 536.

Davis RJ, Shen W, Sandler YI, Amoui M, Purcell P, Maas R, Ou CN, Vogel H, Beaudet AL, Mardon G. 2001. *Dach1* mutant mice bear no gross abnormalities in eye, limb, and brain development and exhibit postnatal lethality. *Mol Cell Biol.* 21:1484-1490.

Davis RJ, Pesah YI, Harding M, Paylor R, Mardon G. 2006. Mouse *Dach2* mutants do not exhibit gross defects in eye development or brain function. *Genesis* 44:84-92.

Davis RJ, Harding M, Moayedi Y, Mardon G. 2008. Mouse *Dach1* and *Dach2* are redundantly required for Müllerian duct development. *Genesis* 46:205-213.

Garcia-Bellido DC, Collins DH. 2006. A new study of *Marrella splendens* (Arthropoda, Marrellomorpha) from the Middle Cambrian Burgess Shale, British Columbia, Canada. *Canadian J Earth Sci.* 43:721-742.

Hammond KL, Hanson IM, Brown AG, Lettice LA, Hill RE. 1998. Mammalian and *Drosophila dachshund* genes are related to the Ski proto-oncogene and are expressed in eye and limb. *Mech Dev.* 74:121-131.

References

- Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nature Rev Genet.* 11:97-108.
- Janssen R, Feitosa NM, Damen WGM, Prpic NM. 2008. The T-box genes *H15* and *optomotor-blind* in the spiders *Cupiennius salei*, *Tegenaria atrica* and *Achaearanea tepidariorum* and the dorso-ventral axis of arthropod appendages. *Evol Dev.* 10:143-154.
- Kozmik Z, Pfeffer P, Kralova J, Paces J, Paces V, Kalousova A, Cvekl A. 1999. Molecular cloning and expression of the human and mouse homologues of the *Drosophila dachshund* gene. *Dev Genes Evol.* 209:537-545.
- Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. 2013. Gene duplication as a major force in evolution. *J Genet.* 92:155-161.
- Mardon G, Solomon NM, Rubin GM. 1994. *dachshund* encodes a nuclear protein required for normal eye and leg development in *Drosophila*. *Development* 120:3473-3486.
- Ohno S. 1970. *Evolution by Gene Duplication*. New York: Springer-Verlag.
- Ortega-Hernandez J, Legg DA, Braddy SJ. 2013. The phylogeny of aglaspidid arthropods and the internal relationships within Artiopoda. *Cladistics* 29:15-45.
- Page RD. 2002. Visualizing phylogenetic trees using TreeView. *Curr Protoc Bioinformatics* Chapter 6:Unit 6.2. doi: 10.1002/0471250953.bi0602s01.
- Pechmann M, Prpic NM. 2009. Appendage patterning in the South American bird spider *Acanthoscurria geniculata* (Araneae: Mygalomorphae). *Dev Genes Evol.* 219:189-198.
- Pechmann M, Khadjeh S, Turetzek N, McGregor AP, Damen WGM, Prpic NM. 2011. Novel function of *Distal-less* as a gap gene during spider segmentation. *PLoS Genet.* 7:e1002342. doi: 10.1371/journal.pgen.1002342.
- Posnien N, Zeng V, Schwager EE, Pechmann M, Hilbrant M, Keefe JD, Damen WGM, Prpic NM, McGregor A, Extavour CG. 2014. A comprehensive reference transcriptome resource for the common house spider *Parasteatoda tepidariorum*. *PLoS One* 9:e104885. doi: 10.1371/journal.pone.0104885.
- Prpic NM, Wigand B, Damen WGM, Klingler M. 2001. Expression of *dachshund* in wild- type and mutant *Distal-less Tribolium* corroborates serial homologies in insect appendages. *Dev Genes Evol.* 211:467-477.
- Prpic NM, Janssen R, Wigand B, Klingler M, Damen WGM. 2003. Gene expression in spider appendages reveals reversal of *exd/hth* spatial specificity, altered leg gap gene dynamics, and suggests divergent distal morphogen signaling. *Dev Biol.* 264:119-140.
- Prpic NM, Tautz D. 2003. The expression of the proximodistal axis patterning genes *Distal-less* and *dachshund* in the appendages of *Glomeris marginata* (Myriapoda: Diplopoda) suggests a special role of these genes in patterning the head appendages. *Dev. Biol.* 260:97- 112.

- Prpic NM, Damen WGM. 2004. Expression patterns of leg genes in the mouthparts of the spider *Cupiennius salei* (Chelicerata: Arachnida). *Dev Genes Evol* 214:296-302.
- Prpic NM, Schoppmeier M, Damen WGM. 2008a. Dissecting spider embryos for light microscopy. *CSH Protocols* 3, 950-954. doi: 10.1101/pdb.prot5072
- Prpic NM, Schoppmeier M, Damen WGM. 2008b. Whole-mount in situ hybridization of spider embryos. *CSH Protocols* 3:933-936 (2008b). doi: 10.1101/pdb.prot5068
- Ramsköld L, Edgecombe GD. 1996. Trilobite appendage structure - *Eoredlichia* reconsidered. *Alcheringia* 20:269-176.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
- Rota-Stabelli O, Campbell L, Brinkmann H, Edgecombe GD, Longhorn SJ, Peterson KJ, Pisani D, Philippe H, Telford MJ. 2011. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc R Soc B*. 2011 278:298-306. doi: 10.1098/rspb.2010.0590.
- Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrland TF, Gupta V, Jiang X, Cheng L, Fan D, Feng Y, et al. 2014. Spider genomes provide insight into composition and evolution of venom and silk. *Nat Commun*. 5:3765. doi: 10.1038/ncomms4765.
- Schomburg C, Turetzek N, Schacht MI, Schneider J, Kirfel P, Prpic NM, Posnien N. 2015. Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *EvoDevo* 6:15. doi: 10.1186/s13227-015-0011-9.
- Sewell W, Williams T, Cooley J, Terry M, Ho R, Nagy L. 2008. Evidence for a novel role for *dachshund* in patterning the proximal arthropod leg. *Dev Genes Evol*. 218:293-305. doi: 10.1007/s00427-008-0220-5.
- Sharma PP, Schwager EE, Giribet G, Jockusch EL, Extavour CG. 2013. *Distal-less* and *dachshund* pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman *Phalangium opilio* (Opiliones). *Evol Dev*. 15:228-242. doi: 10.1111/ede.12029.
- Sievers F, Higgins DG. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol*. 1079:105-116. doi: 10.1007/978-1-62703-646-7_6.
- Stein M. 2013. Cephalic and appendage morphology of the Cambrian arthropod *Sidneyia inexpectans* Walcott, 1911. *Zool Anz*. 253:164-178.
- Waloszek D, Chen J, Maas A, Wang Y. 2005. Early Cambrian arthropods—new insights into arthropod head and structural evolution. *Arthropod Struct Dev*. 34:189-205.
- Whittington HB. 1971. Redescription of *Marrella splendens* (Trilobitoidea) from the Burgess Shale, Middle Cambrian, British Columbia. *Geol Survey Canada Bull*. 209:1-24.

Whittington HB. 1975. Trilobites with appendages from the Middle Cambrian, Burgess Shale, British Columbia. *Fossils Strata*. 4:97-136.

9.2 Turetzek and Prpic in preparation. Observations on germ band development in the cellar spider *Pholcus phalangoides*

Abzhanov A, Kaufman TC 2000. Homologs of *Drosophila* appendage genes in the patterning of arthropod limbs. *Dev Biol* 227, 673-689.

Akiyama-Oda Y, Oda H 2003. Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. *Development* 130, 1735-1747.

Akiyama-Oda Y, Oda H 2006. Axis specification in the spider embryo: dpp is required for radial-to-axial symmetry transformation and sog for ventral patterning. *Development* 133, 2347-2357.

Akiyama-Oda Y, Oda H 2010. Cell migration that orients the dorsoventral axis is coordinated with anteroposterior patterning mediated by Hedgehog signaling in the early spider embryo. *Development* 137, 1263-1273.

Andersson DT 1973. *Embryology and phylogeny in annelids and arthropods*. Pergamon Press, Oxford.

Claparède É 1862. Recherches sur l'évolution des araignées. *Natuurkundige Verhandelingen uitgegeven door het Provinciaal Utrechtsch Genootschap von Kunsten en Wetenschappen*. Deel I. Stuk 1. C. van der Post Jr., Utrecht.

Crome W 1963. Embryonalentwicklung ohne "Umrollung" (=Reversion) bei Vogelspinnen (Araneae: Orthognatha). *Deutsche Entomologische Zeitschrift Neue Folge* 10, 83-95.

Crome W 1964. Eikokon, Embryonalstadien und frühe Jugendformen von *Conothele arboricola* Pocock (Araneae: Ctenizidae). *Zoologische Jahrbücher Abteilung für Systematik, Ökologie und Geographie der Tiere* 91, 411-450.

Edgar A, Bates C, Larkin K, Black S 2015. Gastrulation occurs in multiple phases at two distinct sites in *Latrodectus* and *Cheiracanthium* spiders. *EvoDevo* 6:33.

Emerton JH 1872. Observations on the development of *Pholcus*. *Proceedings of the Boston Society of Natural History* 14, 393-395. Plate II.

Holm Å 1940. Studien über die Entwicklung und Entwicklungsbiologie der Spinnen. *Zoologiska Bidrag från Uppsala* 19, 1-214. Plates 1-11.

Huber BA 2011. Revision and cladistic analysis of *Pholcus* and closely related taxa (Araneae, Pholcidae). *Bonner Zoologische Monographien* 58, 1-509.

McGregor AP, Hilbrant M, Pechmann M, Schwager EE, Prpic NM, Damen WG 2008. *Cupiennius salei* and *Achaearanea tepidariorum*: Spider models for investigating evolution and development. *Bioessays* 30, 487-498.

Pechmann M, Khadjeh S, Turetzek N, McGregor AP, Damen WG, Prpic NM 2011. Novel function of *Distal-less* as a gap gene during spider segmentation. *PLoS Genet.* 7:e1002342.

Turetzek N, Pechmann M, Schomburg C, Schneider J, Prpic NM 2016. Neofunctionalization of a Duplicate *dachshund* Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. *Mol Biol Evol* 33, 109-121.

Wallstabe P 1908. Beiträge zur Kenntnis der Entwicklungsgeschichte der Araneinen. Die Entwicklung der äussern Form und Segmentierung. *Zoologische Jahrbücher Abteilung für Anatomie und Ontogenie der Tiere* 26, 683-712. Plates 38-39.

Yoshikura M 1954. Embryological studies on the liphistiid spider *Heptathela kimurai*, part I. *Kumamoto Journal of Science Series B (Biology and Geology)* 3, 41-48. Plate I.

Yoshikura M 1958. On the development of a purse-web spider, *Atypus karschi* Dönitz. *Kumamoto Journal of Science Series B Section 2 Biology* 3(2), 73-86.

10. Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (grant numbers PR 1109/4-1, PR 1109/6-1 to N.M.P.). Additional financial backing has been received from the Göttingen Graduate School for Neurosciences, Biophysics and Molecular Biosciences (GGNB), the Göttingen Center for Molecular Biosciences (GZMB), and the University of Göttingen (GAU). Further support was given by Christiane-Nüsslein-Volhard-Foundation fellowship and a "Women in Science" Award by L'Oréal Deutschland and the Deutsche UNESCO-Kommission. I thank Beate Preitz and Felix Quade for help with microscopy. Cloning and *in situ* hybridization experiments were performed with support of my student assistants, Julia Golldamm and especially Julia Schneider.

11. Supplemental Material

11.1 engrailed

Pp-engrailed was cloned using degenerated Primers for the initial PCR, followed by RACE PCR (Turetzek, 2011). The probes were synthesized from the 5'RACE clone with a length of 386 bp (coloured in light blue).

```

acgCGGGGgagtgactcaagagaaaccacctcccatagaccgctcacttcaaattctgatt
ttttttgagacaggagtacagcaacattcgacctcaaggcattcattggagaaacaaga
ctatttcgTTTTcgTccacgtgctgctTTTTcggttgctattacttaattagTctgtgat
tgtatgTTTTTTTTgTTTTgTTTTTTTTcaccggtgaattattatggTggaaaagaaaaaat
agcattctcccgtcctctgtgTtagaacgTtcttcagaatcttatcttcaggcctctttg
aattttgaaaacgtagatcattttgcaagactttctatcgaatcaacgtaagtttgaaga
acttctaggcgtgTtagttgatcaacacttatttaatgatcgcttattccacatcagtga
tgaaggaatgacgagTgtgctgctggagaagacttttcccgaagagtcagcaggggctgatt
catattctcactcaagtggcaaaaacggaaaacattttgtcgccacctgcagcctcgagt
tatcagTTTTggTatcaacgtagcaacctgtTgtctaatgcaaagaatccgTttcctttc
tgggactttgctactTgctcttcgTccgctggTgagccactgaccgTggactcttctttc
cagaggtcgaaaattcagcccagcgatgaaactgatctcgataatttccagctaaagTtt
gcttcaaattattggactgTcgtctagaataaccaggatttaggactatggctctggacatg
M A L D M
gagcgaccaagctctgtgcacagTccagagcagcattcattatccggcggaagtagcctc
E R P S S V H S P E Q H S L S G G S S L
ggcggTctaccggccagTcaaatgcacgaacgTcggTcatcggaaggTcgccggTcgacg
G G L P A S Q M H E R R S S E G R R S T
agTcctgaggatgagcgatcgagcagcgaaggaacagaaccgcccacagagcacaccac
S P E D E R S S S E G T E P P H R A H H
ccagcggTccagggacagccgTaccgactactgcgattcccttaagTTTTcaatcgaa
P A V Q G Q P L P T T A H S L K F S I E
aaaactctctccggactttggcgtTggTatccagggacaggagagattcgTttcga
K I L S P D F G R L V S R D R R D S F R
gaagagacttccgtagcgtcaataataaagTtaactccaaagaatctacgaactctagc
E E T S V A L N N K V N S K E S T N S S
aagcaaagTtccgaaaaacaaaaatacggctcctttgggactcgcagaccctaccgattcc
K Q S S E N K N T A P L G L A D P T D S
ggaaaaaatgtattgtggccagcatggTgctattgtacaaggTtcagcgatcgccccca
G K N V L W P A W C Y C T R F S D R P S
tcaggTcctcgatcacgacgagTgaaaaagaaggacaagaagcccgatgaaaagaggcct
S G P R S R R V K K K D K K P D E K R P
aggactgctttcacggccgaacaattggctcggTtgaaacaagaatttcaagaaaatcgg
R T A F T A E Q L A R L K Q E F Q E N R
tatctgacagagaaaCGGCGACAGGATTTGGCCCGTGACTTGAACTTCACGAATCTCAA
Y L T E K R R Q D L A R D L K L H E S Q
attaagatattgTttcagaaccgTcgtgCGAAACTGAAAAAGCTAGTGGTCAAAGAAAT
I K I W F Q N R R A K L K K A S G Q R N
cctctcgctctccaacttatggccaaggattgtataaccatagcaccattcctttgagg
P L A L Q L M A Q G L Y N H S T I P L R
gatgaagacgatgatgagaggccaaaatcgtcttcgTcttcttaatatagatggaaactttc
D E D D D E R P K S S S S S -
caagTttctaacttgacaatgcttcaccggTggTgtaataactaaggcctagTggtagacc

```

agaaatcttggaaactccctgtgtaagaactgcacaaccgaattcacaataacatatgact
tgatgacgagtagagatgtagagtagaaaaggactagaatgtttgactaaactactcgct
aacatcactataatagcagaaaacgaaatcatcactcttagatgaacttggatttatc
ctgtacaacataatataggaaagagcccagtttaatcattcttacgagcagggacacga
tattctggttgatcatgattaatttaagtatgggggaacaatcaaccatctttgtctctt
aagcacatcgtagtgcccagttgagcagcactgttcgattcaccacgtgcagtcagcct
acgtactgaaaaaaaaatcaaactgtgataatttctgaagaaaactttttcttccttgc
tatgggtggaagtatttatcactgatttccgcagattccattttttttctgagcagcct
cttggtttccattgaaaaatagcctataaccattcttcttactataactcgtaccttca
tctaagtntcacatagtgaatataactcaagatcacatgcgtg

11.2 Timing of embryonic development of *P. phalangioides*

Supplemental table 1: Timing of developmental events during germ band formation and differentiation in embryos of *P. phalangioides* at 25 °C

Please note that all time points given in the table are approximate values. Abbreviation: h AED, hours after egg deposition

48 h AED	Primary thickening forms in germ disc center
58 h AED	Cumulus migration starts
	Primary thickening cells disperse
70 h AED	Connection between cumulus and remnant of primary thickening ruptures
	Remaining primary thickening cells congregate again: formation of secondary agglomeration
72 h AED	Cumulus reaches rim of germ disc
	Cells of the secondary agglomeration disperse: formation of the posterior growth zone
	Radial symmetry is broken: bilaterally symmetric germ band forms
82 h AED	First segments are visible in the posterior portion
92 h AED	Embryo attains bean-shape
	All prosomal segments are morphologically visible
	Cumulus slowly disappears
96 h AED	Prosomal segments well delineated
	First opisthosomal segment visible
	Cumulus not visible anymore
106 h AED	Prosomal limb buds formed
	Tail bud forms
	Three opisthosomal segment anlagen present
119 h AED	Prosomal limbs elongate
	Tail elongates
	Three opisthosomal segments well developed and lying on the yolk
	Ventral sulcus appears
130 h AED	Buds of book lungs appear
	Tail segments clearly delineated
	First tail segment (O4) opens and folds down on yolk
143 h AED	Brain differentiation begins
	Dorsal tissue develops on opisthosoma: start of inversion
	Tail adds more segments
154 h AED	Opisthosomal limb buds are well formed
	Next tail segment (O5) opens and folds down on yolk
167 h AED	Mid inversion: dorsal tissue grows
	Tail shortens by folding down more segments to the yolk
178 h AED	Late inversion: dorsal tissue overgrows yolk
	Tail shortened; only two segments
191 h AED	Dorsal closure
	Tail gone; all segments folded down on yolk
215 h AED	Prosomal shield begins differentiating
	Ventral closure begins
225 h AED	Prosomal shield fully formed; pulls legs apart
	Ventral prosoma closed
239 h AED	Ventral closure complete
	Dramatic leg length growth starts
263 h AED	Legs wrap body
	Eyes are pigmented

11.3 RNA seq genes

11.3.1 Evi1

>c109624_g3_i2

CAAAAAAGAAACCCACAGGTGAGAGAAAAAGAGTAGTAGACCTTGTTTTCTTTTTTCAATCAGAGCCTTACCTGTTCCATAGAGTAAAGT
AGAGAGAGAGAGAGACGCACTCTTAAACGCGGCTGCTCTACTTTACTTACTTTACTCTCTCTGGAAATAAGGGAGGAGTCCCCGCAAAGTGGGA
CATGAAATGCCCGATCTGATTCCAACAACCGCGGCACACACACAACAAGTAATACTCAACAACGAGACCCAAACAAAGCCGGTGTATAATT
GGGAGAGACTGCTGTTACCTGGTGGTGGTGGTGGCAATAAAAAAAAAAAGAAACAACAAGAAACGAGAAAGAGAGAGAGAGAGAGGAG
AAGAAAAAGTCCAACCTACTCTCTCTACATTTGTGTGTGTAATCTCTATCTGTGGGAGAGAAAAAAAAAAGGGGCGGTGGCAGCGGCAAG
CGATGAAACTCTGTGAAGCATGGAGTCCAGGTACGGCAGTCTGTCAATGAAATCTCGTCCGAGTAAACACAACACACACACACACACACA
ATTGACCGTTTTCTGAGGAAATTTACTCATGGTCCGTCCTACTAGCTTGACGGGTGTCGTTGATTGCTGCTGATTACTGGTGTGATAT
CCGACCGCGATCCGAAACAATATTTCTGTGCGGTGAGTGTGTTCTCGACACTTGAGACATTTGTGTGTTGTCAACTCGGCAATGTGCACATGAA
ATTGATCTCTCATGAAAGTGTGTGTCAGTGCAGCTTGAGTGTGAAATGACTCTGATGAAGAAAGGGGGTCCGATTTCCACCGAATTTGCCAC
GTGATAATGCCGAGACCTACAGTACTGACGGTCAGCAACGGATATTTCCGGTGACATTTGCTCGATCTTCGGCATCTTCGCCAGCCAAAGAAT
AGTTTTAAGCCACCGTCGATTTCAAGAAGAAGGATATTTTTGATAAGAAAAGACAAAACCTCCATCTTTCTAGGAGACTTACCAATCGATGA
ATTTTTGTCCAGGACTCGAACTGATTGAGAGGATTTCTGTTTTGCAAAAAGAACCTTATTGTATTCTATCGATATCGACCAATCGAATATTT
AAGTGTGACTTTTTGTGGATAAGAATCAAAGGGAATTGAATCTACTTAGTTGGCTCAAGGCTAAACGAGCATATTGGAGGGAATTTCCACCTCT
CTGAAGACAGCGGTCAAGGCGAGAGGGAATACAAGTGCCTGAGTGCCTGAAAGTCTTCAACTGGAATCAAACCTCATACGTCACCAGGTGCGC
CAGCAGGAGAGCGGAAGTACACCTCGGAGAACTGCAAGAAAGTCTTTACAGACCCGAGTAACTACAGCGGCACATCAGTCCAGCATATAGG
AGCTAGCTCAGCCATTTCCGAATACAGGCAAGACTTTGCCACTTCCAGCGGACTCAAGCAGCACACCCACATTACAGCAGCGTAAAGCCGT
TTAGTGTGCGAGTATGCTTCAAGGCGTATACGCACTTCTCGAACTTGTGCCGCACAAGCGCATGCACGCGACCTGCCGATGCAGATTAAGTGT
CACAAATGCGGTCAAGCGTTCAGTGCCTTACTTCTGTGCAAGCATAAGAGTTCTGTGAAGTACTTCAAGCAACAGTAACTCGGCATCGAC
GACGTGACCGGTCAACAAAAGCAGCAACAGTCCGATCAAGGACCGGTTTCAAGCTAGCAAGGCTGCTCTATGAGCTAGCAGCAC
CGCTAAACCCCTGCTGTTCTATCCTCGACTGGTTTCCCGCTCTATCCTCCATCTTTATTGCTTACCCTTTGTTTCTGGCCCGGGTTACCCTG
ACCTCCACCTTATGATCCCGGACAACAGTTCTGTGGGACAGACGACCGGAGCAACAGTACGCTGCGGCGAGCCAAAGTGTGCAAGAATTTCCAAA
GGCAACAGTTATCTACCAATAGTCTCCTCGATGTCAGGCAAAAAGCACAGGACTCTTGTCTTCCGCAAAAGCCATCGGAGAACGTGAATTCGTCG
CCAACCAAACAGGATAAGAGAAAACGACAACAACCTCGGAAGTTCTGAATCTGACGCGTCTGAGGACTTGAGTTCCGAAACGGGTGGTGAAGACA
TCAGTTCTATCAGTACGCTGAGTCCGATCCAGATAATATCAGGCTCCGAGTCTCAGGCTGTGGGAAAAGCAGCAGCACCCGACGAAAAGT
GAAATGTGAGCCATATTTCCGAATACAGGCAAGACTCACAAAGCAAGGGTCAAGTCTGCCAGACCCATCTTTGATTATGTCAACAATGAACCT
GAACGGTCACATCAGATCGATGATGACCAACGGCATGACCAAGTCAGACAACGACATGCCATTCGACTTGTCCCGAACAGCAAAGGAGGCGGTT
CCGAAGATCGGACACTTATAATTTCAAAGTGCAGGAGCACCAAGTTGGAGCACCAGCCACTGGACTTGAGAGTACCCACAAAAAATCTTTT
CAGCCCGTCCGAAAGAGGATTTGTGCAAGAGCCCAATCAGCACAAAGACAGCAACAAACCGGCAACAACGATTTCCGCTGTGCACCGAAAAG
CACCACAGCATACTATCTACGGTACTCCGCCACTCCACCGCTGGAATCTGCCAAAACCGGACAGTCTATCCCCAGTCCAGGATAAGCTACA
GATGGCTACCCAGGCCGATACATCCGATGTTCTCGAGTCCATGTACAGGCTGCAACAAGAGAAAACAGGCTACAACCTGTTCCCGGCCACG
ACCGACTGATGCGCCTTTCCCGGCACGATATCCGTTTTTGGTCCACTACTGGGTACGAACCTTCTCTCGATTTATGAGGGCCACATGGAGA
AAGTCCGCAAGCCATGCACGACATGATGTCCTTACCTGGGAAAGACGAAAGAACGTTACTCGTGAAGTTCTGCGGCAAGATTTCCCGGAG
TCCGCAAACTGACGCGGACCTCGGACGACACCGGCGAGCAGCCCTACAATGCAAGTACTGCGAGAGATCGTTACGATATCATCGAACCT
CCAGCGTCAAGTCCGCAACATACACAACAAGGAGAAAACCTTTCAAATGTCCGCTCTGCGACAGATGCTTCCGGCCAGCAGACCAATTTGGACAGG
ACCTCAAGAAGCAGAAAGCAGACGGTCCGACCACTCGACGACTCGCCAAGACCATAGAAGTCGACGACAAGGACGAGACTTACTTCCGACGAG
ATCCGGAGCTTATCGGAAAAGTACGAAACGACAACCATGGGAGCAGCGCGCGGGTCCGCTGCCGCTGATGGCGTGGCGATGGACGACAG
GAAGCGGGAACCGGACATGATGCTGTGGCTCCTTGTCCGCCAGGACAGTCCCGGAGCGGACACAGCGACAATCTGGACGAGGAAACCGGACG
GTATTAGCCCTCGGACGAGCAGGAACTTCCATCAAAAAGCACCGGCTCTCCGAGCGCTGCTCTCCAGCGGCGAAAAGCAGCAGCAAC
GGAGCCCGACTCGACGAAAAGACCCGAGCAGCAGAAACGATTCAGTCCACGCTCATACCACCCATCACAACAGCAAAACCAATCACCAT
CATCTCAGGAACGGGATTAAGTCCGCAAGTACCAGCAACAGCAGCAGAAAGCCCTCGGTCAGGATGATGACCCTGAACCATAAAGAT
CGACAACATCAACGGCATCGACAGAAAACGACGACAGAGCGTGGCCGATGCAACGACAGGAAATACGGAGAGGAGTCCAGCGACGAGGACGAG
ATCGTGGACCCGTCGACGACGACGAGGAAAGTCCGATCGACGCTCAAGGAAAGAGGTAAGTTCGACTGATGTCATTTGCGCAGTA
AATTCGGGTAGTAGAATCAAAAAGACACAACCTAAAGAAAAGGCATAGCGGGGGCCGTTCCAGGGTCAACAACCGCTCCCTCAGCCGAAAGCGG
ACGGGACGTTCACTGCAACCGAGAAAACCTCGGCGAAAACCGACTCGTGACGGACATTGAGTCTCATTTCTTGTGACTCATAGGACTCAAAAAT
CATCTTCCCTATAACGAACTCAGTCCGTTGAAATGAACTGTATCCGACTTCGGTTAAGACGAAATACGTCAGTACATTTGGGATACGTACAAT
TTTTTTCCCATATCAAAAATTTCAAAAAAATGTTGTTGTTTTTTAAATAGATGCGCGTGAAGAAATGCTAGTCACAAAAAAGTTATATATA
TACATAAAATATACATATATAATATCTATATCCCTGCTTTAGCCGTTGCAACCCCGTTTTATGTGAAGCCTAATGATGCCCTCGAAGACATTT
TGTAAGAACTGTGACGTTTCCCCCACTCGAAGAAAATCCGCTGATGAAACTGAAATGATGGACGATTGAGTCAAAATCAAAGTGAATGAAATG
AACATGATATCGTGACATGCAAGTACTCTCGTCTGTCACCGTTTTATTTCTGTTGTTTTTCTCAATCGATTATGATGGACTGGATCGAT
GTCAAAAAAATTTATCCCGTCTTTTTCTTTCTTTGTTGATGTGAACATGCGCATTCCTTGATTTCACTGTAATAACATGATAAGAAATTTAT
CAATATCCGGTTTTCCGCCACACATGAATTTGTATCCGTTGAAATATTTTGTAAATTAATCGAGAGAAAATGTATTAGATCATGTCATAGT
AGGGAGTATGTTGGCATCCATGAGGAAAAGCCATGCTACTTCCGCTGTTATTTCTTACCTAGTGTAGGCAAAATACGACCCCGGGT
TCAGGTGCGGCCACCAAGGCTCCCATCCGCCCTCCAAGGTGTTCAAGTAAAAAGTATCAATGCACATACCACAATAAATACCTAGGCTTA
CCATATTTCTTATGAAAATACGGGAAATAGACAATCAGGATCCATATGGGATATCACAAAATATCTCAAAAAAATTTGTCAAATGCAAG
TTAATTTGATAAATACATCAAAAAATTTTCAAATGCAAGCAAAGTTGTCAATTTATAACCAAAAATTTGTCAAAGACAAGAAAAGTTGCCAACT
ACCACTGAAATGCTGAAAATCGCAATAGCGGAAAGTGGGATCTTCCACAAGACCGCATGTGTAATAAATTTACTTACCAAAACAAAACG
TATGATGTTGCAAAATATGTGAGATGAATGTGGCTAAAATTAATTTATGATGAAATTTATCTTGATAAATATGACTAACCGGTGAAAAA
TCTTAATTTATGATTTCAAAGTTCAAGTAACGCTCATTGAAAATAAGAGAGAAAACAGGAAATTTTTTAAAAACATTAAGGTCAAAAATTA
GAAACAT

11.3.2 c105514

>c105514_g1_i1

GGAGCGTAAAAGGTTTACGCTCTGAGAAACCGGTGACAGGGGTCCTTAAAGTGAGGAAGAGAAAACAACATTTTCTCGTATTTGTGAGTGTGTATC
AATCATTTTCGGTAAATAGTTTTGGGCTATCGGTCCGAGGAGGATTTCTTGTATTTGATCCTGTTCTCTCCATTCATGATCGTGAATGGGCAAT
GTACGCGAGCATCAGGGATCATCATTCCGGCTAATAATCATTAAACAGTCATCGTGTTTTTCTTTAGTGTTTTTTATTCATGTAGTGTGTGGAA
ATAATACCTTTATAGGAAACGTAATCTCGTGCCTTTTCATGTGTCTAGATATGGGAAGATGGATTATCAAAGATGTTTTGGCTCTGTGAGC
TAGCACAGAGTTGTTCTGTCACCCGTCGGATGTCCTTTCTGTCGGAATTTGCCGTCCGAGAAAGGATCGTTTTTCGATAGCCAGTCTAGCATCTGCTG
TGGGCTGCAAAGTTCTCGAAAAAGTCTTTGTTCCGACAGCCGTCACAGCTATCTTGAAGCAAGAGCTAATCCAAGCGATGGTAAACGTAATTT
AAAGCCATAACATCCAGCAAGTTCTTGTCAAATGTCCGGCATCAAGAAGCAATCTCCGAGCTACGCGCTTCGAGCGCAGGCGCTACGGCCGAGGGA
GGTGCACCGCAACATCTGGCGTTTGTAGCGGAAAACTTTCCGCAAACTGGCAAACTGTCAAAGATCCCGAAACGGAGGAGAAAAATGTCAGA
CCTGAAGTACCGAAGAGATCGACATATAAGAAACGCAAGCGAGTGTCAAAGGAAAAAGCCATTGAAGTCTGTGTTAAGGAGCTCGAAGTAAAGG
TTCCATGTGCAGCGCAAATCCCGATTTCCCTGCCATCGAGATCGGAAGTCCTTCCCCCTGATGGCGATTTCATTACGTTTCAGATGCCGAAAAAGAA
CATCAATCTCCGAAGGACAGCAGATGCCATCCTGTACTTCTCGCAACTGAAGCTTTGAAAGCGTACGCTCAGAAAAGGAGCTGTTGCTGT
GACCTTGGAGAAGACTCAGAGGGATTTGAATAATGCTCACACGAGGATTAAGGATCTCGAGTTGCAATTTGGAAGACTGCAAGAAGTCCGTTGAA
GAACTCTCAGCGACCTAAAGACTTGGAAACAGAGATGTGCCAATTCAGAAAGACAAAAGCTTGTCTCCAGCAGGAATTTGATGTTCACTAGGA
AACAACTGACGAGGATTTCTGACCAGTCCGATTTCTCCGATTCATCTCATGTGAGCTCGGTGTTGTTTCCGGGTCTCAGAACCTCACTGACGAC
GGACTCGGTGACGAGACGACAATGACCCCTGACGAGGACGGTACGGACACCCGGAAGATGATGATGACGACGAGGACATCGAAGTGCCTCAC
AGCAAAGCCGCAAGAAATGCCAAGACGAGGTTGGAGCAGAATAATTTGATATATCGCGGCTAGAAGACTTGCAAATGAAAAATGAAGAGCTCGA
ATTGACGATTAGGCATTGCGAGGAACGGATAATCGAACTCGAGGAACTACTTAGACAAGCCACTACCCGTATCGATATGAGAGACGTGACGACC
CAAACATTCAGGAGACGGTACTTTACGAGACCCAGAGAAAAACAAAATTTCTAAAAGGATGCTCGGTATCACAGCTCGTTACCTGTGATAA
GTGACCCGAACAACCTTGGTAGATGCAAAACCCCACTGGTAGCTTACGAGATTGAAAAACCTACAGCGAAGCTTACGGGTGAAAGAAAAGAGA
GATCAACAACATTTGAAGAAGGTGATAGCAACTATGGAAATGAATGAAGTCCGTAAGGAATTTGAAAAAGCAGAAAGAAAAAGAGGTGATGATCCA
GAATTCGAGGATCGGTGAACCTTGCATCTGAGAACCAGGCTTTTGAATGAGTTGAGATCGTCAAGGAGAACTACCAGAGAGAGACGGTATTG
AGAAAAGATTTTCAACATTTCCGAAGATCTCAAAGGCAAAATTCGAGTTTTTTCGCGATAAAGCCAATGGCAATGATGCACAGGAATGCAC
CTGTCTTTGAAGTACCTGATGAATATACGGTTATGTTACGAACCCACAGAGCACAATACGAATTTACATTTGATGGTGTCTTCGATGGGGATCGA
TCTCAAGAGGATGTGTTCACTGATGTTCAAACCTTGGTGCAGTCTGCAGTAGACGGCTACAATGTTGTATATTTGCGTACGGCCAAACGGGATC
CGAAAAGACCTACACTCTAGTAGGAAATGAGTCCCAACCTGGAATAGGACCCGAGAGCTTTCCAGAAGATTTTCGATTTAGTCCGTGAGAACCAA
TGTGAACCTGACTTCCGAATAACTGTATCTGTTTATGAACGTACAACGAACGCTGATTTGATTTATTGAAAGATTCAACTGAAGACAACCTAA
AGATAAGGGGAGATAAATTATGGTCTAGTGTGTTGTTCAAGGTTTCAAGATCTAGAGAGGTCAAAAATGAACGGCAACTGTCAGAAATCTTTGAGGA
AGCCTGTCCGAATCTGCAGTCCGGAGCAACAAGAAATGAACACGGAGAGCTCAAGATCTCACCTCGTAATAACTATACCGCTCGAAAGTACAAACA
AAATAACTGGAACGTTTTCCCGGGGAAACTTAGCATAATTTGACTTGGCTGGAAGTGAAGAGTTGGAAGACAGCGGCTGACACAGCAGCT
GAAGGAAGCCAAATTCGATCAACAAATCATTTATCCGCTCTTGGTGACGTGATAGCCGCTGAGCTCGGAGAACAGCAGCTGCCATATAGGAATA
ACAAGCTGACACTTCTCATGAGGATTTCTCCGAAGCAATGCCAAGACTCTCATGTTTGTCAATATCTCATCATCTGAACAAAACGTCGAAGAA
TCACTCAACTCTTTAACCTATGCTTCAAGGGTCAAACAGATACCAATTCGGCATCAGCAATGGCAGAAAGCAAAAGAGTCCGAAGTTACGATC
GATTATAATGAAGCTTTCCGAAGGAAAACAAATCGAAGAATTGACGCTTTCATGAAATTTACTCAGTATTATGCACGTGATCAATCAATTTG
CTTTTTTAATTCATGAAAAA

11.3.3 c95092

>c95092_g2_i1

GCGGGGGTTTGTATGATCCAGTCTCAGGTAAGTTGTTATAAAGGTCATCCTAACAACCTCCTCACTTCTTCGTGTTCCGATTGGACCACATAAA
CGATAATGGAGGAAAAAATAAAGAGAAATTTTTTTCACATGTAATTTTTTTCGTTTCGAGAGAGTTTTTTCGGTGCCTGTGATAATAAAG
GGGTGTGTGACGTTTTTACAGTCAAGTTGACCAACTTTCCCTCAGCTGACGGGTACTAGTACGACCATACCAGCGCTATTTTCGCTAT
CGGCATCAGACCATCAGCCGAGAGTTTACAAAACACAGTCCGGTCTAATTTTTTCTTACTCTTACGACTTTTTTACGATAAAAATTCGAACAAA
AACAAAAATGACTTTCAGCTTCAAGACCATCGTTTTTCATTTCTGCTTAGTCCCGCAGTCTATGGCCAGGAGTATAGGTACCAGCAGCAGCAAC
AGGCTCAGTTCAGGCAACACAGCAGCAGAAAGAGAATTGCAACATCGAAGAATGTCATGAAGCTCAGGGACCCGACCCCTGTCAAGGAAGG
AGGAGACATCAAAATTTCTGCGCCGAACCTGAAGGATTTGGCTGCTGGACAGATGCCGACTCACCGCTGTACCAGGCCACCAGCCCAA
TCGTTATGGGAGGAGTTAAGCAGCTGCTGTCGAAATATGCAAGTCTGGATCTAGTCTTGGCAACCGTTACATGCAGGAGTCCAAGTGCCTGAAC
CACCAGAACACGACAGTCTGACTGCGCAACCTCCATGATCGAGAAGTATCCAGTGTCTCACCCGTCGCAACCCGAGGCCATCGTCCGAGTC
TTCTGTTGTTCAATCGACAGAACCGAAGATGCATCATCGAACGTTCCACAAAGACTGCGGACGTTGGAGCTGCTAAACTCGTCGACGAAATGAT
GTCCAAGGCTTTTTACCCTCAACCAAGTCAATTTGCTATTTCAATGATCCATCGCAGTGTCTCAATTCAGAGGAAAAAAGACTCAATGTTT
TAAAACCTTTCATCATTTTTAAATATGTGTAATATATTTTTTAAATATATCATCATCCTCCCTCCATCATTTTTCTTACATACAAAAA
TATTCATCTCTGCATGTGTTTTTTTTTCTTTATGTAAAAAGGAATTTCTCATCGATTATTTTGAAGGTTGAATGTAATATAAATAAGGCCGGG
ATTATATATATACACTTAAAAGTTTGTATCAATGCTCTTCCAAAACTTCAAAAATGCAACAACAAACAGCGTATCGGTTACATTTGAGC
GAATTTAACCAATAGGCTTATTGGTGTTTTTTAAAGTTTTGAAAAAGATCTTTTATCAAACCTTTCAGAGTGAATCTCGGTCTAAGGAAGATATA
ATCTTCAGTTTTTAAAATTTTCAATAAAGCTAGGATTTACGCTATAAAGTTGTTTTATGAAATTCGAAAAATTCACCAATAAGCGTTTTAATAAA
GTTAAATAAAATTTTAGCAACCAATAAGTTTTCTAGCAAAATGGAACAGATAAAAATTTGGTTGCAATTTTGTGTTTAAATTTGTCGAAAAA
ATTTTCATAAAAACCTTATAGTGAATCCCGCTTAAAGAAATTTCTTGACATATTTAAGTTCCACCCGAAAAAGAAATTTGGAGATATACATATATG
TGACCACATAAAATATTTAAATTTAGTATATGATGACATTGATACAATGACGAGGTCACATCTTTAGCTGAAAGCTAGACGGCGGAATACA
ATGTCAGAGATTTGAAAACTTGAAGACATCAAGTAGCTTGGTGACATGGCTACAAAATCTCCTCGTCAATTCGAATAAAACCAATCCATGATGA
AATGTAATTTGATCAGCTATCTGCTATACTGATTAATCAATAAAAAATTAATACTACTGCTCACTTTTCAAAGCTCTCTGTGATAAAAAATCGTGA
AAAAAAGTACAACATAATTTATGACATAATTTAGTGAACCTTTGCTGTTCTCCTCACACCAATGCGAAATGCAATATTTTTTTATGTGCA
ATCATTTTTGTAATAGCTGAGATTTGTCAATAAATATTCATCCATTCACCAAAAAA

11.3.4 c109329

>c109329_g1_i2

TTTTTTTACAGAAAACAACTCTTTTTTTTATTATTTTACAAAAATCTGTAGGAAAATAACTGCGTGGTACTCGACAACAAAAGAAATTTCCA
AGAAAGAAAAAAAATGTACAGCTCATACTGATATTTTTTTTTTACATTGACGTTGTACAAAAGCAATTCATCTTCTCGACTAAAATATTCAGTCT
GATCAACGCATACAGCAACAAAAAACAAGCGTACAGCCATATTTTGTAAATACTGCATATTTTTTGTTCACATATCTTTGTTGGTCCCATTA
AGAACAATTCATGTATAGATGAACATCATTGCAGAATTTCAAAGAAAAAAGAAAAATTTTTACACATATTTTTCGATGAAATCATTTCAAAT
TATCAAAGCGTCTCTTATCCATTTGTATATTTTCCACATTTTTCCTCTTATTTGACAAGAAAAAATTTGATTTGCCCGCAGTCGAATCTT
CGCGTCTTGGCTCAAATTCGCAAGATGGAAGACAACACTGGGTCTTATGCGTCACCGGTTTTATTTTTGAATTAACGTTCAATTTGTTTTCCGA
ATCGTGGCACCCTTGAATATGGAAGAAAGAAACTATCAATTGTTAGTAAAGTAGTATACCTAATTTTGGCCCAATCAATCATT
ACTTTGTGAAGAATTTTCGTTATCATGCAATTTGTTTTCAATTATGGCAGAATTCGGAGTGTGGGTGAATAGAGGCACCTCAGATGCGATAC
CTCCATTTTAGAGCAGAGATGCCATTTTCAATTTCAAAGAGTTTATAAAATGTACATAGTACATCGAAATATAGTTAGGACATGGTGATTTTTCT
AAATTTACAACACTCCACACCACAGAAATTTCTTGGCAGCATCTGATATCTTTTGAATGGTTTTTCAATAGGCAATTTTGTATCCGATTTTT
TTCTTTTGTGCATGAACTTATGGCATCCCTGCTGCCACTTATGGATTTTTAGAAGCACAATTTACAAAAAAGTGAGATATGTGTGTGTAGA
GAGGATAAAGGAATGAGATGCCACCCTTCTTGTATACAAATGCCTTTCGATGTCTGCGGTGGTGTCTTGTAGTGACAAAAGGGAGGACCGG
ACGGATGGGGAGATCTTTATTTACATCAAAGCTCGTGAATGTCCTCTCTTCCAAATATTTACTGAGGTGATGGAATGGTACGTAGACTGG
ATGAAGCTCACTGGCGTCGGCCCTGGACGGATGAGATACTTTGGATCACTTCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
CTATTTCAACGATCCTCGGAGATCTCATTGGGATGGCATCGGTTGCATTTGAGCAACTTGACGGTTGGCTGTCTTGGCATCAGCAGTTCTCTA
AGGAGCCTGATTGCGTGCAGAGGGCTCATCACTGGTGGCCTTACGACTGGTGGGTGGGATGAACATAGGTGGCCTCTGCGCTACCTGTTGCGAC
CTCGTGGTAGGAACTAGGAGAAGGACTTCTTGTATCGTCTGCTGCTGCTCTTCTCGGGTCTCCCTCACGAAGATGACTCGTCTCTCTCTT
GCGGACTGGGATCTGAATCATTGGCGGTGGAACATCGGCATCTGTTGCTGGACGATCGGAGCTGCTGTTGGACGATGGAACACTGGTATTTGG
TGTGGCAGTGGTGCATAAGAGGAACCTGGGATGGCATCGGTTGCATCATAAGCATGACTGGATGAAGAGGCAACTGGGGAGGCTGATCATT
TAGGAAGGACACCTGGAGCTGGTTCTGGAACCTGCTACCGCAGCATGGGCTGATGAACTGGTGTGGCACTGTTGAATGAAATGGGTCTGG
CTGAGGATCGGAACTCGTACCATGGGTTGGTACTGGCATGCGTCTGGCCATAGTTGGTGGAGTTGCTGCACGGGCATCCTCCTGGCATTG
GCCTCATGACTGGGGAGGAACTTTGGAAGAGGATCAGTATGGGGCGTGCATCGACGGTTGGGGCATAGGCATGATTGGTACTGTACGGG
TACCATGTGGTGCATGTGCATTTATCTGCTGTTGTTGCACTGGATACATATTTCTGCTGTTGTTGCACTGGATACATATTTGCTGTTGCACTG
GATACATATTTCTGCTGTTGTTGCACTGGATACATATTTCTGCTGGACTGGCATGATAGGGATGAGGTTTTGCATTGGTGGTTGGATGGGAATCAG
ATGTTGCATGGGGGAAAGGGCAACATTTGGTGGTGGTGGATCGTGGCTGTGGTGGGCTCATGATGTTGTGCATGATCTGCTGGCCAAAGGT
AGGAATAAAGCTACTGGTCCAGTGGTGGTACTGGTAGCCTGGGTGGTTTTGCAGCAGGCATTTGCTGTGGGACTGGGAAAGGTAGGGCAATG
GGAATCGTCCG

11.3.5 c109662

>c109662_g1_i1

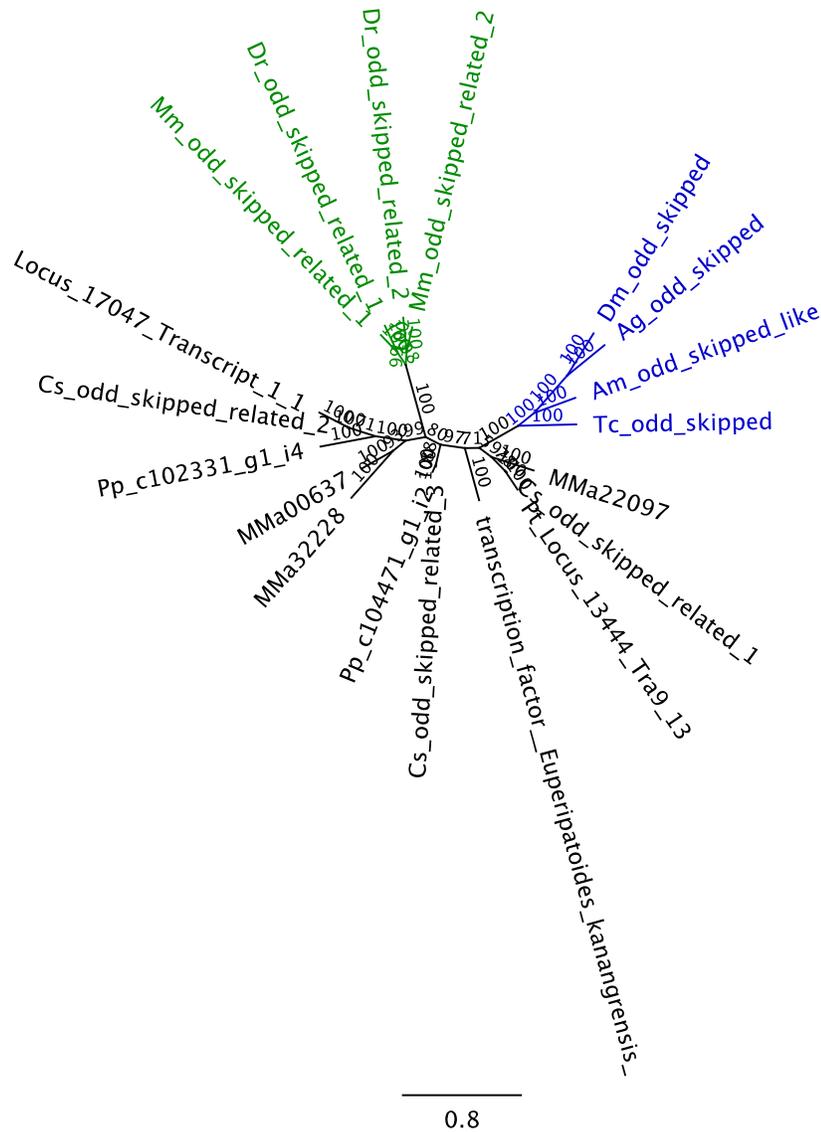
TTACACATACAGTTTTTTCTACATATAATGGGATATTGAAATCTTGAATATTTCAAGAATCATTCGGGGGAAAAAATCATTGTTGAAGG
AAAAGAACTGTTTTTTTTTCACTTGAGAAGAATCCGGTTTTATTGTGAGGAGCGTATGTGTGTGGAAAAGAGACACAAAAGGCACCAACC
ACCACCACCAACCCTAAATTCGTTAATAATAAAGAAAAGTGGAGTAGAGAAAACAAAATAAAAAACTATATACCAGTGAAGTGGAT
GTACTACTAGGCTCTTGGGAACGAAACGGCGTGGCGGCGGAGGAAGAGAAGCCGAATAAATATATAATAATAAACCATAATCGTAAGGA
GTTTACCTCTCAGGCCATGAATCTCATGTTAGCGGACGTTATAACCGTATAAGACTTTCTCGGGGTACATACAGGTTGAGAGAGAAGCTCAAG
ACGCGGGAGTAACAACGTAAGATAGACGAATTTGTCCAAATTTTGTAGAGATAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA
TTACAGCAGGGGTGAGGACAGACTGGTCACTCGTCAACCGTCTGTCAGGCAGCCATCATCGTCAATCCGAAGGAGACAGCCGTCGGTGTCCACTG
GTGGGCAACGCCAGCATTTCTGCAACGGCCGACGACGATACGGCTGTGGCCCAAAACCTCGGCTAGAGGATCGTCTACCAGTCGAGCGTGAAC
AAGGACAGCTATGACTTGAGGGGACCTATCGACAACAGGCCGTCATATCAGCAATCGGTCGTAGGTGAAGCACTGGAACAGACGCCCCACCAGC
CGACAACAGCCAGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
CTGAAGACCGCTGCCACACCCCGGCGCATCGATAGAAATGATTTACATACAGGAAACCTCAGACACCCATCAATCTGAGACTGATCTGAGAG
CAATACTCAGCACCGAGAGGGCGCAACAGCACGAATCAAACCTCAGACACCGCGCCAGAGGCAAGCAACGAGACCAGCGGTATCTCGAGGGT
TCCGGAATCGAGGTCTTCCAGGAGACAGCCTACCGCGCCGTGAACCGTCTTCCAGTGGAGCCACTAGTCTGTTCTTGGAGAACAGGCCCTCGG
TACCGGAGGCTCCGGCTAGTGTGCACCAACCTATCAGCCAGACGGCGACGACGAGCATTCTGAGGTTACGAAACGAATACGCGGTTTTCT
GAGACTCGGATCCGCTCGGCTACCAGTAAACGACGAAACTGGTCAAGCAGACTTTCTGAATCATAGGCGCCGTTACCGGTCATTCTGACGG
AATAGACTCCGCCGAAAAAGTCTGCTGGTCAACAGCACGGATCAGAGTTTTTACCTACCCCAAAATCGTCTTAACGATGCCTCAGTACCATCATCAC
CTTTCATGATCTCCACGTCATCTGCCCTCCGCCGTAATCAGAGACATCGACTTCAAAGCTGATGTGGTAACTAGGACCCAGAGGGACTGTCGG
ACCGCCTTCTAACGATCAGTACCTGAGGGCCTTCCGATCAGAGAAGCTCTACCGCACGACAGACTTCCGGAACCCGAAGGTCGAAACCCCGAA
CCGAAAGGATCGCGCTCGAGTTCCCTTGGAAATCGACGAGAGGTGGTCCCGCTACCAGGCGCGGAGCAACCCGAACCCAGCTCGCATAACAGCC
TCCGACCGACCAACGTCGTAGGCAGGTGCCCTCGTGCAGATCCAGGTCGGGTATAGAAAATGAAGAAGTCGCCAAGAATACTGTACGGTCTTCCA
ACAACGACAGGCCAGTCGGCGCTCGATTGCAAGAGGACCGGTTCCCGTTACCAGCCTGAAGATCCCGAAGTCACACGACCAGTCCCTCCAGGGGA
CCCAAAGAGTTTCTGAACAGCAGTCAACGAAAGAGCCCTCAAAGGAACCCAGGCACTCGGTCACGACAAGAAGAGCCGAAGCCCAACCCAC
ACCGACAGATCTCGAGGACGAGGTCAGATACCAACCCGAACCTGCACCCGTAAGAAGAGACTTACAGCGCCCGATAGAAGAGCCGGCTCGAC
AGAGACGGTCCCGAGACCAGACAGGTCAGGAGTCGGAGCCAGTCCGCTGCCAGGCGGAGCCAGTCCGATGACAGCGAGGAATCCAT
CGCACGGACCTCAGTATGAATCTGAGCTCGAGGTCGCCAGCACAACAGAGGCTTACGTGCTCAATACCAAGATGTTCCGAGGACAAACCA
ACGACGACCGGAACCCGCACAAGTTGGTCATTTCTGTTCTGTCACGTTGTTGACGCTTCTCCGAGGCAGAGCCTCATCTCAAGGCTCCTCA
GGAAGAGCCTCAAGCAGGTCGACGCCACGACAGAATCAGCAACAACCTCCGCAAGCAGTAGCAGCGGGTCATCTTCAATGCCGAATCCCT
ACGGCTTCTACCTGATCAAACCACTGCGACAAGTACTACGAATGCCCAACGACGACGCTGAGGAGGAGGACTTTGTCGGGACGGTCTTGTCTTC
AACCCCATCAGCGCCGACTTCTTGGTGGATTTCCGTCAGGGACATCGACTGCACTGCGAGTCCAGGCGCAATTTGCAAGCTCTCAATCGACAAA
GAACTGTCTCGACGATTTGGTCTGTTCCCGGACAAAAAAGAAATGCACCAAGTTCTGGCACTGCGTTGACGGCACGTCCTCAAGTGACCTGTC
CTCCAGGACTCGTGTACGACGATTCAAGGGAACTGTAATGGGCAGATCTCGTGAAGGATCTGTTCCACTGAAGATCTCTGGGATTCGCT

TGCCCGACGAAATACCAGCCAATATCCAGGAAGGGTTGTACCTTCGCTATCGCCATCCGGACAACCTGCCAAGTGCACCTCACCTGCATCAGGGCC
 GCTGACGGAAACCCGCAAGCCTCGCATGCTGTCTGCAACGAGGGATTTCGTGTTTGGATCCCGAAACCAGTCACTGCACTCGTCCAGAGAACGTACC
 AGGCTGCGAGGAATTCACGGACCACCCACCACCACCGTGGTCAATCGAATACGAGGCAAGAGCAACAGAAACCACCCGCGTGGCTCGCAAC
 CACCTCCACCCGGCAGAATACCGTGAAGGAGAGGACGAGGAGCAACTGTTGACAGAGACCCAGAGAAGGAGGAGCCGAGCAGTACGAGGACGAG
 ATCATGAACGATGTGAAATGTTAATAAGAAAAGCCGAATAAGCAATTAACGTTTCAAGACTATCTTAAAGAAGCCACCTGTAATGAAAACAAA
 AATTGCCTCGTCACTCTTATATAGAACAGTATTAACACACATTTTCTATTTAAAGAGGATTTCATCAGCACATTTGATCTCACAAACAGA
 ACCCTATAGGAACCTTGAACAAAATCTATGCATCAGAAATCTCACATAAACGTTATCGATGAAAATCTTCAATAGTTTGGAGGCTTTTATTTGA
 CTCTGGACTCTTTGGCTCACCCGTTAAGGTACCCTCTGCACTACAGCGTCCCTAGGCATTGCTCCAAAATCCGTCACCCGTAAGGCTGAAAAT
 CCCTCCCTGGTTCAGGATGTTTTGTGCACCCTCAGGTCGGAGCTTTTGGTTGGTTCCCTCTCTGATTACCTAATTGATTTTTTTCAGCGATTTAAA
 AATAATAAGATTCCCGGTCTCTTAGGCAACAGCCTTGTAAACCTCGTACATGTTGGGCGTCAATGCCGAAATCCCTCTAGGTGTAGATCATAT
 TAACATTAACCTTTGACTACGAATTAAGAAAGAAATTAACCTCAGCCTTTTCTAATGAACAAAATAGGTCACAAGTAAGGCGAATTTAGCATT
 GGAATTTTTCTTTTTTGGGATATAAAGTCCCGTTACTAGTCAAAGGTTGAAGGAGATATAAGAAGCAAGGAGAATGAATGCTCCTATAT
 CTCTCTAACTAACGAGGCGGTGATTTTTGTTTTGGTATTAACAGTGGCACTCAAACCTTGGACAGCTGACTCCTTTGAGATAGGGTCAAAAATA
 TCGACTGGAACAAAAAACTTACTCTGTCCGCGGGCCAGTAGCATTCTGTCCGCCACCCTTCACAATTTTGAACAGAAAAAGAGTAAACTCA
 ACAGCTTCTGTTCCCTAGTTGCGGCCCTGCATAGGATAGCAGCGGCTTGACATAGCGACGAAAAGTGCCTGCTTTGGACCTCGCTACACAATA
 TACAGGTGTGTGCACCGCCAATAATTACACATAGATTCCCTCTGAAAACTAACGTGCAAGAGTGTTCGTTTTTTTTTTTGTATTTGATCA
 AT

11.3.6 oddskiped

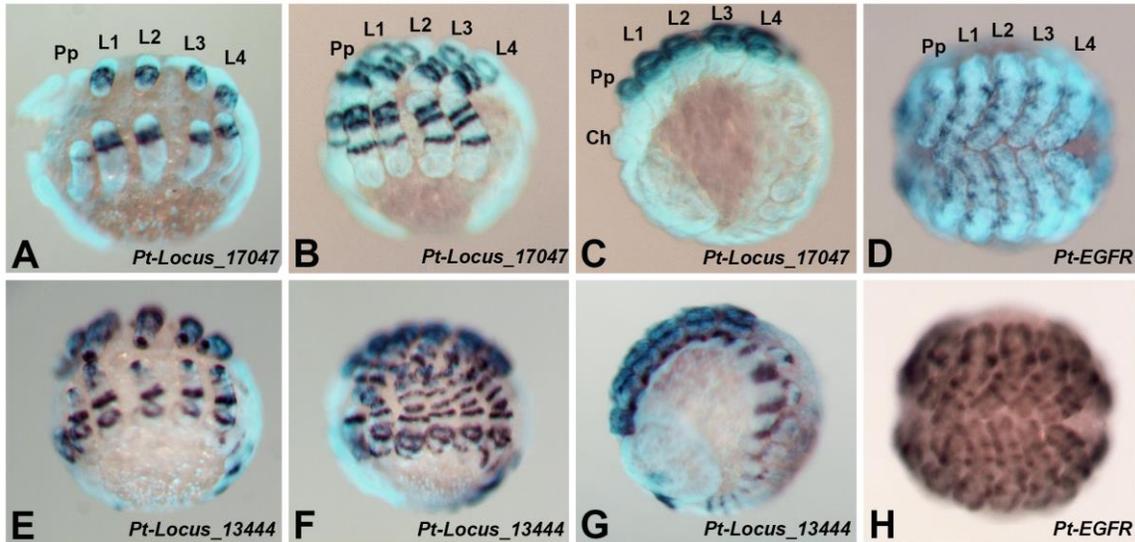
>c102331_g1_i4

AGAGGAACATGGACGAGGTGCATGTGTGGTTCAGTAGTAGTTCCTCGAGTCAAATCGGGAGCCCTACGTTATTCGAGAATTTTGTGTTGT
 TCCGAGATCATATAAATTTTTCATAAAATTTTCCCACTGAATATGATTTTTTTCATTAGTTCGAGAAATTTTTCATTAAATATGAATTTTC
 ATAAATCTTGATTACTCCCACTAAAATATCAAATTTGCCCTTTCAAAAATCATATTTGCCCTCCCCCTGTGGGGCGTGGGCCCTCAGTTGGGAA
 ACAGGGCTCTAGAAAATTTGATGTGTGAGATCGGCCCGGGTGGGCCCATGTATTAACACGGCAGTGTCTGCATTATAATGTTTCTCAGTCTCTCA
 GTTGTATAAAAAATACTTCTATATACTGTTACATTATAATGTAGCGACAGGTTAAGAATTTTCAGTTTTAATTGATACTTGAGAAGCGTTGCAAA
 TTAGACGATCACAAAATTTACATTGGCTCTATTGCACATAATTTTGTATATAACCTCGTGAATATACATTCCATTGTGTACGCGTTACATATT
 TCTTAGTATATTTGTCTCTTTAAAAATGGGTAAATCAATGAATAAGAATTTTTTTTATTAGTATTTTTTTAAGCGAATTTCTCTTCCCGCGGA
 GTAAGAAAAGAACAGCGCATGTTTCACGACAGACTTTTCGAAAAAGGACGGACGGCACAGCGGGGGTGGGAAGAGCATAATTTGATTGTTT
 CAGTTTTCAACGTCGACGATATCAGAGTCGATGTCCGAGCAGAGAGATCGCTATCGTGATCTTGATCGCCGCTCTAATCCAGAGTCTCGCGG
 AGACGGCGGCAACATTTTTTCAGCGCGTTATTGACGTCGTGGATGAGTAGATGTCTCCTCAAGTCACAATTTGCGTCTGAACCTTTTGGCCGAGG
 ACGAGCAGTTGTAAGGCTTTATGTCGGTATGCGTGAGTAGATGAGTTTTCAAATTTGCTGCGTGGTTGAAAGAAGCTCCGAGGTAGCGCACTTG
 TCGGAGAATCTCCATGTGAAGAATCCGGTGCACCGCAACGTTCTCGACTGGCAGAAACCTTTGCCGATTCGATGCACCTTGAAGGCTTTTC
 TTTGCAATGAATGTACCTGTGATCTCGAAGATGGTCTTGCCTTCTGAAAGATTTGTGACATATGTGCGAAGTATACGGTCTCTCGTCTGTGTGCG
 TCCTTTTCGTAATCAGCAAAATTTGACGATTTTCGTGAACCTTTTGGCAGAAATTTGCACACGAATTTTCCGGGGACGAGAAGTTGCAAAAGGA
 TTGCGGTGGTGTGATGCTCATCTCGCCGCTGTTGGATGCAGGATGAAGGCCCTCTTAGTTGAGGTACCGCCGAGTCCGGCGGATGCCGTGAACCAAGC
 GTTAGTTTTGGCAAAATGTCTCGAAAAGTTGTGCTGCGGATAGATGTGCTACACGGATATTCGCCGTTTATAGACGATGAATCGGCTCGACTGA
 CGTCAGTCATCGTTGCGGACTCGGCAATCGGCTGAAATCGAATTTAATGGAACGCTCATTTCCGGATTAAGAGAGCGGATACCGCGTGGGGG
 AAATTTTGTATCAACGAGAGGATAACAGTCAACCGCTGGGGGTGCACCGACTTGGCACTCTCAAGGAAAGACGACTTTTACGTCCTTCGAGACT
 AAGAGCACAATGCGATCTCGAAACTGTCGTTGATCCAGAATCTGAGGCAAACAACGCATCGTAGATTCGCCATCGTCAATGTCCGACCCAGCG
 TCCAACAGTGCATGTACAAAACATTAATACGAACGGATGCTGCCCGTACGAGCTGCTACTGGCCAGAGAATACCGTGGTTGCAAGAGCCGCTC
 GAAGGAAGATTGTGTTGATGATACATCATGTGTGACATCAAGTAAATCGACTTCACTTCATATCGTCTGCTTTTTTCAGCGTAATCTGATTGC
 AAAGGAGATAACGGTGCACCTCCCGCAGCGAAAATAACAGGAAGCAACTACATTATGGATACCACAGGCAGCTACATATGATATACCATATCCT
 ACTGCTTAAATTTGCCACAGTTACTGTAAGCACAGCGTGGTGGAGTAGTATCCAGGGTTTTGTATTTAACTCTTGGCTCATTAACCTCAGTCAAA
 TTCTGACTAGCGGTGACATCGATTCCTGTTGG



Supplemental Figure 1: Preliminary phylogenetic analysis of *odds*-related homologs

Rooted 50% majority rule consensus tree from total number of 670 trees written in two files. Each file contained 335 trees of which 252 were sampled. Mixed model testing indicated that the amino acid substitution model Blosum was the best suited model and the majority consensus tree was generated from 167000 generations after reaching the average standard deviation of split frequencies below 0.01. Shown in black are all the chelicerate and one onychophoran sequences. In blue all the insect sequences are grouped and in green the *odds* homologs of vertebrates are depicted.



Supplemental Figure 2: Expression of *odds* and *EGFR* homologs in *P. tepidariorum*

11.4 Leg gap genes

11.4.1 extradenticle

In the following you will find the nucleotide and protein sequences, which were cloned with degenerated primer and RACE PCR of the spider *exd* homologs observed in this study. To verify that these are the only paralogs of *exd* and *hth* in the two spider species observed here, a large phylogenetic analysis was performed, with these sequences and additional data obtained from the now available transcriptomes. In most of the times the sequences obtained from transcriptomic data include more sequence information and result in longer protein sequences. For the detailed phylogenetic trees to clarify phylogenetic relationships of the duplicated paralogs of spiders and other chelicerates usually these longer sequences obtained from the transcriptomic resources were used. The probe for *Pp-exd1* was synthesized from the clone Nt_251

11.4.1.1 *exd1* paralog sequences

```
>comp102492_c1_seq1
AGCGATTGGACGAGGCTTTTAGTGACGTCAGGAGGTAGCGCCACGGCACCGCTCATCTTTTATGGCGAAGTGGCTTGACTGGCATTATCGCT
CAAACCTCCTGCGCTGTCCGTCAAAAACGATTTCCCCATCAAAAACACTCAATTGAGAGCACAAAACCAAGCAGTGCTAGAGGCTCGAGCCCCAC
GGCGTGTGCGCGGGCCTCGTCTCCACATTTTTCACAATGGATGCTGCTGCTCGACTACGAACGGGCGGCTAAATGAAACATAAACACGGCTAATAA
AAAACGAACGAAGGCTTTTGTTCACAACGTTGACGTCGACGTAGAATGGACGATCAGCAAAGAATGATGCACCCGGTATCGCAGCACCCCTCA
TCGGCAGCAGCTAGCATGGGCGGCCACGTCGTGCCAGCACGGGTACGGACTACCGGGCCACACGTTGTTGACCCGGGCTCGCAACAGAGCACT
CCAGACCATCAAGATTCTCGAAAACACGACATCTCCGAGATCCTTCAACAAATCATGAACATCACAGATCAAAGCCTGGACGAAGCTCAGGCGAG
```

Supplemental Material

GAAGCACACCTGAACTGTCATCGCATGAAACCGGGCTTGTAAACGTCTCTGTGAAATCAAGGAAAAACAGTTTTAAGTCTGCGGAATACAC
AAGAAGATGAGCCACCCGATCCGCAGCTGATGCGGCTCGACAATATGCTCATAGCTGAAGGTGTCGCGGGTCCCGAGAAAGGGGGAGGCGCCGGT
CGCCCGGCAACCGCGACGGCGGACAGCGCGGGTGGCCCAATCAGCCGAAAAATGCCATTTGAACTTCCGACTACCGGGCGAACTGGCCCA
AATCAGACAGATATACCACCAAGAAATTAGAGAAATACGAACAGGGCTGCAACAGAGTTACAGCAGCAGCTCATGAACCTTACTACGAGAGCAAAGC
CGCACTCGCCATACAGCAGAAAGAGATTGAAAGAATGGTGCAGATCATCCACAAGAAATCAACTCCATCCAAGTCCAGCTGAAGCAGAGCAC
GTGTGAAGCCGTCATGATTCTCAGGTCCAGGTTTCTGGATGCGAGACGGAAGCGAAGGAATTCAGCAAGCAGGCCACAGAAATACCAATGAG
TATTTCTATTCTATCTGAGCAATCCCTATCCGAGTGAAGAAGCGAAAGAGGAATAGCCCGGAAGTGTGGTATAACAGTCTCACAGGTTTCAA
ATTGGTTTGGGAAACAAGCGTATTCGATATAAAAAAATATCGGCAAGGCCAAGAAAGAGGCGAACTGTATGCTGCGAAAAAAGCAGATTACAT
CAATGACTGTTATTCCCACCGTCTGCGGCAGGTTCTCGCCCTACAGCTTAGGTCTTCTTCTCAGGGCAAATGATGAGCCCTCCACCTGGTCC
AGCAACCGCGAAGATTCTATGTATAACATGTCTATGAACGGTAGCGACTTACTCTCTATGGGGCAAATGTGCAGTCTCAGGCAAACGCAT
TGAGGCAGCTATTTCTCAAACCTGCTGGTTACGGGTGATGGGATGCCCGCCATCTCAGCTTCTATATATACAGCCCGGCATGCATCAGACT
TCCGAACCTCACCCATAGTAAAGCTGAGCACTAGCAGCCAGGCAATCTGTGTACAGAATGGTCTGATTAAGGCCTACCAAAAAATAAGAAAA
AAAAATATAAAAAAAGAAAAATAAAATGAAACAAGGGCTTCAATCAATATACGTTTTTGTTCATACCTCAAAAAAAGAAAGATTACAA
AATAAAAAACCATGTATCATTTATAAAATTTGAAAGAAAGAAAGAAAAATGATGAAAAACGGACAATTTTACGACATTATGGGTATAT
ATGTGTATATAAAAACTCTGATGGAACGTTTTACATGATTATATATATAAATAATATACCCGGTCTGTTGACTGTGAAAGAACAATAAATATC
TGAAAAAGAAAAAATGAAAAATTTAAAAAATAAACTGGAAAGTGTGTGTGTACAGGAATTCAGCGTTTAAAGAAAAACATAAAAAAAT
TAATATGTAGACAGAATAATTGTGCTTTAAAAAATCCAGATTTCTGTGATTTTGTATGAACATCCCGTGGTGTGAATTTTTTTTTTTTTTTT
T

>comp1022492_c1_seq1

MDDQQRMMHPVSPSSAAASMGHVVPHQHYGLPGPHVDPGSQSTPDHQDSRKHDISEILQIMNITDQSLDEAQARKHTLNCHRMKPAL
FNVLCIEIKETVLSLRNTQEDEPPDPQLMRLDNMLIAEGVAGPEKGGGAGAAANATAATAAGGPNQPENAIHSDYRAKLAQIRQIYHQELEKYE
QACNEFTTHVMNLLREQSRTRPITTKIEIERMVQIIHKKFNSIQVQLKQSTCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEAK
EELARKCGITVSVSNWFGNKRIRYKKNIGKAQEEANLYAAKKADYINDCYSPPSAAGSSPYSLGPSSQGMMSPPPGPATPQDSMYNMSMNGSD
SYSSMGANVQSQANALRHVISQTAGYDGMPPSSASSIYDPGMHQTSELHP

>Pp_exd1_RACE_Nt249_Nt251

ACATGGGGATCAAACACTCAATTGAGAGCACAAACCAAGCAGGGCTAGAGGCTCGAGCCCCAGGCGTGTGCGCGGCCCTCGTCTCCACATTTT
TCACAATGGATGCTGCTGCTCGACTACGAACGGGCCGCTAAATGAAACATAAACACGCTAATAAAAAACGAACGAAGGCTTTTGTTCACAAC
GTTGACGTCGACGTAGAATGGACGATCAGCAAGAATGATGCACCCGGTATCGCAGCACCTTTCATCGGCAGCAGCTAGCATGGGCGGCCACGTC
GTGCCACAGCAGGGTACGGACTACCGGGCCACAGGTTGTTGACCCGGGCTCGCAACAGAGCACTCCAGACCATCAAGATTCTCGAAAAACACGA
CATCTCCGAGATCCTTCAACAAATCATGAACATCACAGATCAAAGCCTGGACGAAGCTCAGGCGAGGAAGCACACCCTGAACTGTCATCGCATGA
AACCGGCTTGTTTAACGTCCTCTGTGAAATCAAGGAAAAACAGTTTTAAGTCTGCGGAATACACAAGAAGATGAGCCACCCGATCCGCAGCTG
ATGCGGCTCGACAATATGCTCATAGCTGAAGGTGTCGCGGGTCCCGAGAAAGGGGGAGGCGCGGTTGCCGCGCCAACCGCAGCGCGGCCAGC
GGCCGTTGCCCAATCAGCCGAAAAATGCCATTTGAACATTCGACTACCGGGCGAACTGGCCCAATCAGACAGATATGCCACCAAGAATTAG
AGAAATACGAACAGGCGTGAACGAGTTACAGACGCACTGCAACTTACTACGAGAGCAAAGCCGCACTGCCCTATCACGACGAAAGAGATT
GAAAGAATGGTGCAGATCATCCACAAGAAATCAACTCCATCCAAGTCCAGCTGAAGCAGAGCAGCTGTGAAGCCGTCATGATTCTCAGGTCCAG
GTTTCTGGATGCGAGACGGAAGCGAAGGAATTCAGCAAGCAGGCCACAGAAATACTCAATGAGTATTTCTATTTCTATCTGAGCAATCCCTATT
CGAGTGAAGAAGCGAAAGAGGAATTAGCCCGGAAGTGTGGTATAACAGTCTCACAGGTTTCAAATTTGGTTTGGGAAACAAGCGTATTCGATATA
AAAAATATCGGCAAGGCCAAGAGGCGAACTGTATGCTGCGAAAAAAGCAGATTACATCAATGCTGTTATTTCCCAACCGTCTGCGGCA
GGTTCTCTCGCCCTACAGCTTAGTCTTCTTCTCAGGGGCAAATGATGAGCCCTCCACCTGGTCCAGCAACCGCGCAAGATTCTATGTATAACAT
GTCTATGAACGGTAGCGACTTACTCTCTATGGGGCAAATGTGCAGTCTCAGGCAAACGCATTTGAGGCAGCTATTTCTCAAACCTGCTGGTT
ACGGTGTGGGATGCCCGCCATCTCAGCTTCTATATATACAGCCCGGCATGCATCAGACTTCCGAACCTCACCCATAGTAAAGCTGAGCA
CTAGCAGCCAGGCAATCTGTGTACAGAATGGTCTGATTAAGGCCTACCGAAAAAATAAAAAAAAAAAAAAAAAAAGT

>Pp_exd1_RACE_Nt249_Nt251

MDDQQRMMHPVSPSSAAASMGHVVPHQHYGLPGPHVDPGSQSTPDHQDSRKHDISEILQIMNITDQSLDEAQARKHTLNCHRMKPAL
FNVLCIEIKETVLSLRNTQEDEPPDPQLMRLDNMLIAEGVAGPEKGGGAGAAANATAAPAAGGPNQPENAIHSDYRAKLAQIRQIYHQELEKYE
QACNEFTTHVMNLLREQSRTRPITTKIEIERMVQIIHKKFNSIQVQLKQSTCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYSSEAK
EELARKCGITVSVSNWFGNKRIRYKKNIGKAQEEANLYAAKKADYINDCYSPPSAAGSSPYSLGPSSQGMMSPPPGPATPQDSMYNMSMNGSD
SYSSMGANVQSQANALRHVISQTAGYDGMPPSSASSIYDPGMHQTSELHP

>Locus_1_Transcript_58666/166847_Confidence_1.000_Length_605

AGAAGTTCAACTCCATCCAAGTCCAGCTCAAACAGAGCACGTTGTGAAGCCGTCATGATCCTTAGGTCGAGGTTTCTAGACGCCAGGAGAAAAAGG
CGGAACCTCAGCAAGCAGCGACGGAGATCCTGAATGAGTATTTCTATTCTCATCTGAGCAATCCTTACCTTAGTGAAGAAGCAAAGGAGAGCT
TGCCAGGAAGTGTGGAATCACAGTCTCTCAGATATCAAATTTGGTTTGGTAATAAAGGATACGATATAAAGAACATAGGCAAAGCTCAAGAA
GAGGCCAACCTTTATGCTGCGAAAAAAGCAGATTACATAAAGACTGCTACTCACCTCTTACAGCAGCAGGTTCTTACACCTTACAGTCTGGCTCC
TCTTTCGAGGCGCAGATGATCAGTCCACCCCGGGTTCGGGAACCTCAAGATTTCCATGTACAATATGAGTATGAATGGTGGTACTCCTATT
CTCTATGGGAGCCAACTACAGTACAGTACAGGCAATGCGTTGAGGCAGCTGATCTTCAAACCGCGGGATATCTGATGGCCTGCCACCAACATCT
TCAGCCTCTATGTACGATCTGCAGGCATGCATCAG

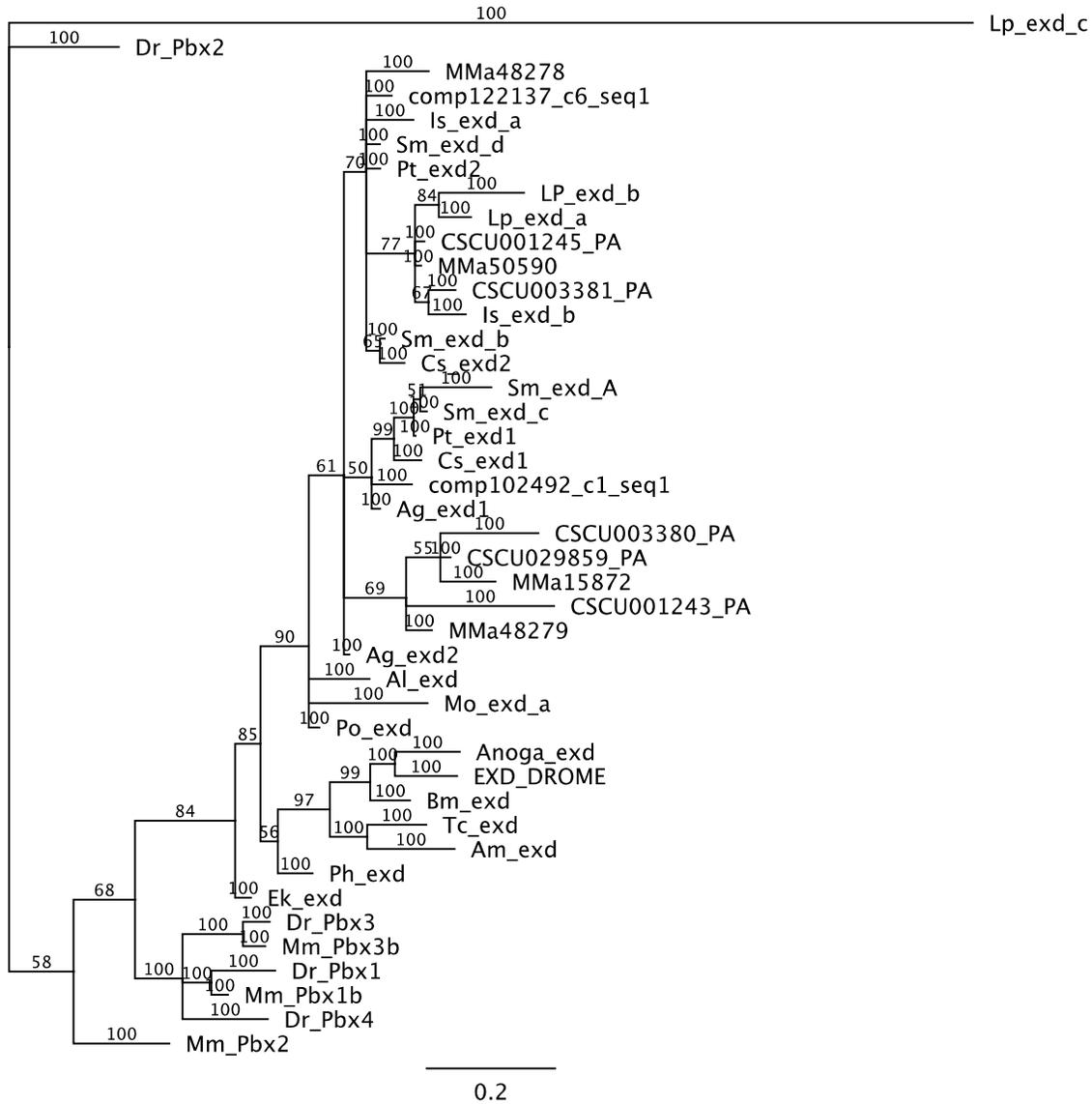
>Locus_1_Transcript_58666/166847

MILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITVSVSNWFGNKRIRYKKNIGKAQEEANLYAAKKADYINDCYSPP
SAAGSSPYSLAPSSQGMMSPPPGSGTPQDSMYNMSMNGSDSYSSMGANVQSQANALRHVISQTAGYDGLPPQSSASMYDPAGMHQ

Supplemental Material

>pt_exd_2_Sk12_-_ORF_1_(frame_2)_translation

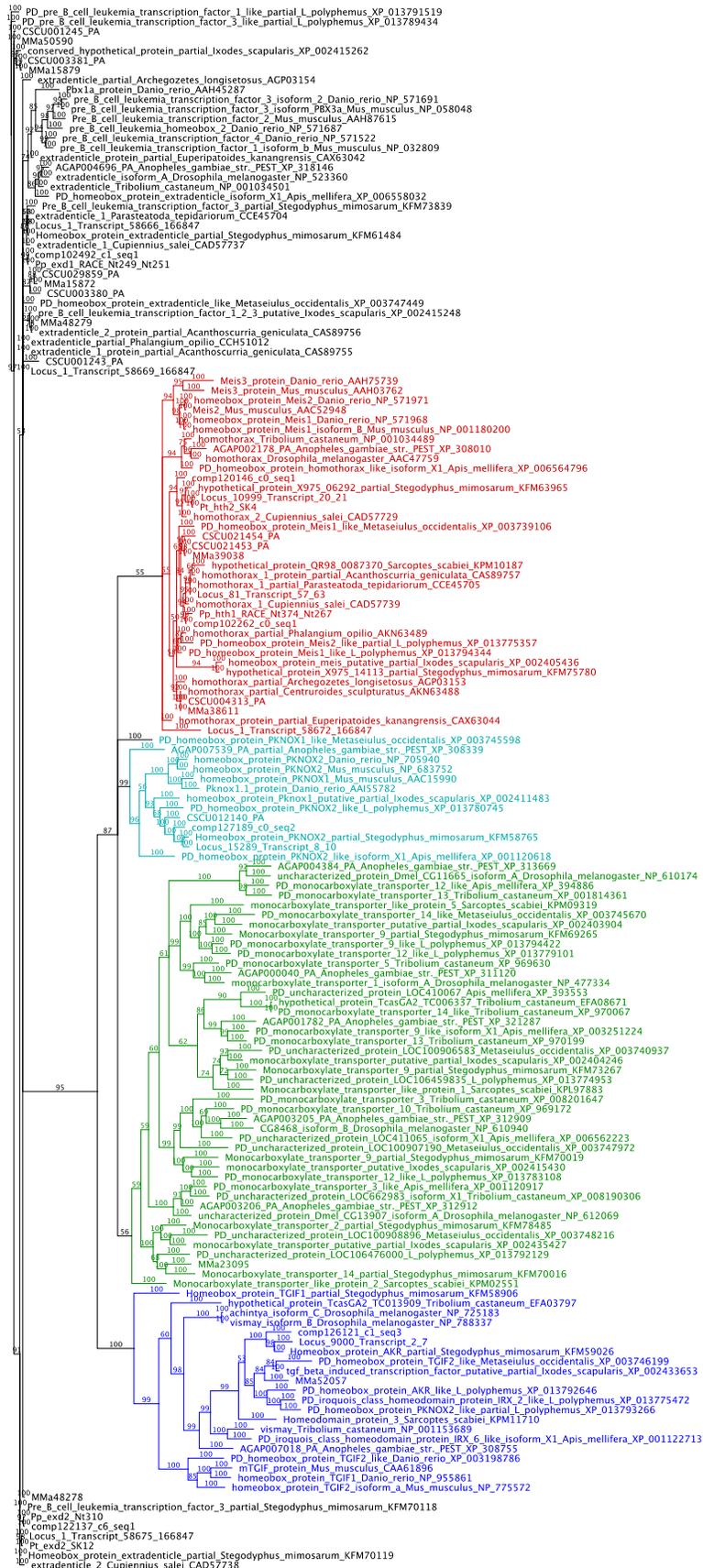
MRLDNMLLAEGVAGPEKGGGAGAAANASAAASSGPQSENAIEHSDYRAKLAQIRQIYHQELEKYEQACNEFTTHVMNLLREQRSRTRPITPKEIER
MVQIIHKKFNSIQVLKQSTCEAVMILRSRFLDTRRRRRNFSKQATEILNEYFYSHLSNPYPSEEAKEEL



Supplemental Figure 4: Unrooted 50% majority rule consensus tree to resolve *exd* relationships of the paralogs

The Jones amino acid substitution model was chosen after mixed model testing. After reaching 300000 generations a total number of 12002 trees were written in two files (each file contained 6001 trees of which 4501 were sampled).

Supplemental Material



Supplemental Figure 6: Unrooted 50% majority rule consensus tree to identify true *exd* and *hth* homologs.

The same phylogram as in Supplemental Figure 5 in a magnified view to display gene and species names. The third sequence with similarity to *hth*, *Locus_1_transcript_58672* was transcribed into a protein with only 33 amino acids and thus excluded from further studies.

General *exd* alignment:

CLUSTAL O(1.2.1) multiple sequence alignment

```

Locus_1_Transcript_58675/166847      MDDPQQRMMHPVSVQHTSV---SMAGHVVPQHGYGMPQQPHGVDPNSSQ-PPPDQDQVVRKHD
comp122137_c6_seq1                  -MDDQQTMMHPVSVQHPVSVS-----MAVPQHQGYGMPQQPP---HGVDQGPPTDQEPKHKD
EXD_DROME                             -MEDPNRML-----AHTGGMMAPQGYGLSGDDGQAGSE-----NEVRKQKD
Cs_exd2                               -MDDPQQRMMHPVSVQHTSV---VSMAGHVVPQHGYGMPQQPHDPT---GQAPPDHPDVRKHD
comp102492_c1_seq1                  -MDDQQRMMHPVSVQHPSSAAAASMGGHVVPQHGYGLPGP-HVVDPGSQSTPDHQDSRKH
Cs_exd1                               -MNDQQTMMHPVSVQMPSPS---ASMAGHVGAQHGYGLAQTSHV--PGSQGGTTPD-QDAPKHD
Ag_exd1                               -----
Pt_exd1                               -MNDQQTMMHPVSVQLPS---ASMAGHVGAQHGYGLAQPShVVNQGSGGGTTPD-QDAPKHD
Ag_exd2                               -----

Locus_1_Transcript_58675/166847      ISEILQQIMNITDQSLDEAQARKHTLNCHRMKPFALFVGLCEIKEKTVLSLRNTQEEPPD
comp122137_c6_seq1                  ISEILQQIMNITDQSLDEAQARKHTLNCHRMKTFALFVGLCEIKEKTVLSLRNTQEEPPD
EXD_DROME                             IGEILQQIMSISEQSLDEAQARKHTLNCHRMKPFALFVGLCEIKEKTVLSLRNTQEEPPD
Cs_exd2                               ISEILQQIMNITDQSLDEAQARKHTLNCHRMKPFALFVGLCEIKEKTVLSLRNTQEEPPD
comp102492_c1_seq1                  ISEILQQIMNITDQSLDEAQARKHTLNCHRMKPFALFVGLCEIKEKTVLSLRNTQEEPPD
Cs_exd1                               ISEILQQIMNITDQSLDEAQARKHTLNCHRMKPFALFVGLCEIKEKTVLSLRNTQEEPPD
Ag_exd1                               -----VLSLRNTQEEPPD
Pt_exd1                               ISEILQQIMNITDQSLDEAQARKHTLNCHRMKPFALFVGLCEIKEKTVLSLRNTQEEPPD
Ag_exd2                               -----VLSLRNTQEEPPD
                                          ***:*****:***

Locus_1_Transcript_58675/166847      PQLMRLDNMLIAEGVAGPEKGGGAGAAAANASAAAAS---GPQSENAIEHSDYRAKLAQI
comp122137_c6_seq1                  PQLMRLDNMLIAEGVAGPEKGGGAGAAAANATAAAAAS---GPMSENAIEHSDYRAKLAQI
EXD_DROME                             PQLMRLDNMLIAEGVAGPEKGGGAGAAAASAAAASQGGSLSIDGADNAIEHSDYRAKLAQI
Cs_exd2                               PQLMRLDNMLIAEGVAGPEKGGGAGAAAANASAAA-----AGSPQSENAIEHSDYRAKLAQI
comp102492_c1_seq1                  PQLMRLDNMLIAEGVAGPEKGGGAGAAAANATAATAAGG--PNQPENAIHSDYRAKLAQI
Cs_exd1                               PQLMRLDNMLIAEGVAGPEKGGGAGAAAANASAAAAS---QGSQSENAIEHSDYRAKLAQI
Ag_exd1                               PQLMRLDNMLIAEGVAGPEKGGGAGAAAANASAAA-----SSGPGQSENAIEHSDYRAKLAQI
Pt_exd1                               PQLMRLDNMLIAEGVAGPEKGGGAGAAAANASAAAASQSPGPPQSENAIEHSDYRAKLAQI
Ag_exd2                               PQLMRLDNMLIAEGVAGPEKGGGAGAAAANASAAA-----SSGPGQSENAIEHSDYRAKLAQI
                                          ***** :*****.*.* : * : ***** *****

Locus_1_Transcript_58675/166847      RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKLNSI-----
comp122137_c6_seq1                  RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITGKEIERMVQIIHKKFNSIQVQLKQ
EXD_DROME                             RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIQVQLKQ
Cs_exd2                               RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIQVQLKQ
comp102492_c1_seq1                  RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIQVQLKQ
Cs_exd1                               RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIQVQLKQ
Ag_exd1                               RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIQVQLKQ
Pt_exd1                               RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIQVQLKQ
Ag_exd2                               RQIYHQELEKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIRVQLKQ
                                          *****:*****:*****:*****:*****:*****:*****

Locus_1_Transcript_58675/166847      -----
comp122137_c6_seq1                  STCEAVMILRSRFLDAR-----
EXD_DROME                             STCEAVMILRSRFLDARRKRRNFSKQASEILNEYFYSHLSNPYPSEEAKEELARKCGITV
Cs_exd2                               STCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITV
comp102492_c1_seq1                  STCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITV
Cs_exd1                               STCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITV
Ag_exd1                               STCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYP-----
Pt_exd1                               STCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITV
Ag_exd2                               STCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYP-----

Locus_1_Transcript_58675/166847      -----
comp122137_c6_seq1                  SQVSNWFGNKRIYKKNIGKAQEEANLYAAKAA-----GASPYSMAGPPSGT
EXD_DROME                             SQVSNWFGNKRIYKKNIGKAQEEANLYAAKAGKPDLCDDFF-----
Cs_exd2                               SQVSNWFGNKRIYKKNIGKAQEEANLYAAKADYINDCYSPPSAAGSSPYSLGPSSQGG
comp102492_c1_seq1                  SQVSNWFGNKRIYKKNIGKAQEEANLYAAKADYINDCYSPPSAAGSSPYSLALSSQGG
Cs_exd1                               -----
Ag_exd1                               -----
Pt_exd1                               SQISNWFGNKRIYKKNIGKAQEEANLYAAKADYINDCYSPPSAAGSSPYSLAPSSQGG
Ag_exd2                               -----

```

Supplemental Material

```

Locus_1_Transcript_58675/166847 -----
comp122137_c6_seq1 -----
EXD_DROME T-TPMMS PAPPQDSMGYPMGS GGYDQQQPYDN-----SMGGYD-----
Cs_exd2 -----
comp102492_c1_seq1 MMSPPPGPATPQDSM-YNMSMNGSDSYSSMGANVQSQANALRHV ISQTAGYGDGMPPPSS
Cs_exd1 MISPPPGAGTPQESM-YNMSMNGGDSYSSMGANVQSQANALRHV ISQTAGYADGLPPQST
Ag_exd1 -----
Pt_exd1 MISPPPGSGTPQDSM-YNMSMNGGDSYSSMGANVQSQANALRHV ISQTAGYPDGLPPQSS
Ag_exd2 -----

```

```

Locus_1_Transcript_58675/166847 -----
comp122137_c6_seq1 -----
EXD_DROME -----PNLHQDLSP-----
Cs_exd2 -----
comp102492_c1_seq1 ASSIYDPGMHQTSELHP-----
Cs_exd1 ASMYDPAGMHQVGTDSIVSNAENGSMALKAVSHYRHAKMRNYQ
Ag_exd1 -----
Pt_exd1 ASMYDPAGMHQVRI-----
Ag_exd2 -----

```

Supplemental Table 2: Identities of *exd* spider paralogs compared to *D. melanogaster*

Locus_1_T...	comp122...	EXD_DROME	Cs_exd2	comp102...	Cs_exd1	Ag_exd1	Pt_exd1	Ag_exd2
Locus_1_Transcript_5...	85.02%	69.43%	85.15%	85.28%	84.65%	90.55%	84.85%	95.24%
comp122137_c6_seq1	85.02%	71.26%	83.40%	83.86%	83.60%	90.00%	82.75%	93.29%
EXD_DROME	69.43%	71.26%	75.81%	65.21%	64.73%	88.83%	65.81%	88.83%
Cs_exd2	85.15%	83.40%	75.81%	83.92%	83.68%	90.34%	83.78%	92.00%
comp102492_c1_seq1	85.28%	83.86%	65.21%	83.92%	87.33%	93.30%	88.89%	92.18%
Cs_exd1	84.65%	83.60%	64.73%	83.68%	87.33%	93.82%	95.09%	94.35%
Ag_exd1	90.55%	90.00%	88.83%	90.34%	93.30%	93.82%	96.09%	95.45%
Pt_exd1	84.85%	82.75%	65.81%	83.78%	88.89%	95.09%	96.09%	94.97%
Ag_exd2	95.24%	93.29%	88.83%	92.00%	92.18%	94.35%	94.97%	

Alignment of *exd1* in spider

CLUSTAL O(1.2.1) multiple sequence alignment

```

comp102492_c1_seq1 MDDQQRMMHPVPSQHPSAAAASMGGHVVPQHGYGLPGP-HVVDPGSQQSTPDHQDSRKHDI
Cs_exd1 MNDQQTMMHPVPSQMPS---ASMAGHVGAQHGYGLAQTSHV--PGSQGGTPD-QDAPKHDI
Ag_exd1 -----
Pt_exd1 MNDQQTMMHPVPSQLPS---ASMAGHVGAQHGYGLAQP SHVVNQGSQGGTPD-QDAPKHDI

```

```

comp102492_c1_seq1 SEILQQIMNITDQSLDEAQARKHTLNCHRMKPALFNVLCEIKEKTVLSLRNTQEDEPPDP
Cs_exd1 SEILQQIMNITDQSLDEAQARKHTLNCHRMKPALFVSVLCEIKEKTVLSLRNTQEDEPPDP
Ag_exd1 -----VLSLRNTQEDEPPDP
Pt_exd1 SEILQQIMNITDQSLDEAQARKHTLNCHRMKPALFVSVLCEIKEKTVLSLRNTQEDEPPDP
*****

```

```

comp102492_c1_seq1 QLMRLDNMLIAEGVAGPEKGGGAGAAAANATAATAAGG--PNQPENAI EHS DYRAKLAQIR
Cs_exd1 QLMRLDNMPIAEGVAGPEKGGGASAAAANASAAAAS---QGSQCENAIEHS DYRAKLAQIR
Ag_exd1 QLMRLDNMLIAEGVAGPEKGGGAGAAAANASAAA---SSGPGQAENAIEHS DYGAKLAQIR
Pt_exd1 QLMRLDNMLIAEGVAGPEKGGGAGAAAANASAAAASQSPGGPQAENAIEHS DYRAKLAQIR
***** .*****:*: * *****

```

```

comp102492_c1_seq1 QIYHQELEKYEQACNEFTTHVMNLLREQRSRTRPITPKEIERMVQIIHKKFN SIQVQLKQS
Cs_exd1 QIYHQELEKYEQACNEFTTHVMNLLREQRSRTRPITPKEIERMVQIIHKKFN SIQVQLKQS
Ag_exd1 QIYHQELEKYEQACNEFTTHVMNLLREQRSRTRPITPKEIERMVQIIHKKFN SIQVQLKQS
Pt_exd1 QIYHQELEKYEQACNEFTTHVMNLLREQRSRTRPITPKEIERMVQIIHKKFN SIQVQLKQS
*****

```

```

comp102492_c1_seq1 TCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITVS
Cs_exd1 TCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITVS
Ag_exd1 TCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYP-----
Pt_exd1 TCEAVMILRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITVS
*****

```

```

comp102492_c1_seq1  QVSNWFGNKRIRYKKNIGKAQEEANLYAAKKADYINDCYSPPSAAGSSPYSLGPPSSQGQM
Cs_exd1             QVSNWFGNKRIRYKKNIGKAQEEANLYAAKKADYINDCYSPPSAAGSSPYSLALSSQGQM
Ag_exd1             -----
Pt_exd1             QISNWFNGNKRIRYKKNIGKAQEEANLYAAKKADYINDCYSPPSAAGSSPYSLAPSSQGQM

comp102492_c1_seq1  MSPPPGPATPQDSMYNMSMNGSDSYSSMGANVQSQANALRHVISQTAGYGDGMPPPSSAS
Cs_exd1             ISPPPGAGTPQESMYNMSMNGGDSYSSMGANVQSQANALRHVISQTAGYADGLPPQSTAS
Ag_exd1             -----
Pt_exd1             ISPPPGSGTPQDSMYNMSMNGGDSYSSMGANVQSQANALRHVISQTAGYPDGLPPQSSAS

comp102492_c1_seq1  SIYDPGMHQTSELHP-----
Cs_exd1             MYDPAGMHQVGTDSIVSNAENGSVMALKAVSHYRHAKMRNYQ
Ag_exd1             -----
Pt_exd1             MYDPAGMHQVRI-----
    
```

Supplemental Table 3: Identities of spider *exd1*

	comp102...	Cs_exd1	Ag_exd1	Pt_exd1
comp102492_c1_seq1		87.33%	93.30%	88.89%
Cs_exd1	87.33%		93.82%	95.09%
Ag_exd1	93.30%	93.82%		96.09%
Pt_exd1	88.89%	95.09%	96.09%	

Alignment of *exd2* in spiders

CLUSTAL O(1.2.1) multiple sequence alignment

```

Locus_1_Transcript_58675/166847  MDDPQQRMHHPVSQHTSVSMAGHVVPQHGYGMPQQPHGVDPNSQPPQDQDVRKHDISEI
comp122137_c6_seq1              -MDDQQTMMHPVSQHPSVSM---VPQHGYGMPQQPPHG--VDQGPPTDQEPKHDISEI
Ag_exd2                          -----
Cs_exd2                          -MDDPQRMHHPVSQHTSVSMAGHVVPQHGYGMPQQPHDT--TGQAPPHDPDVRKHDISEI

Locus_1_Transcript_58675/166847  LQQIMNITDQSLDEAQARKHTLNCHRMKPALFVLCIIEKKTIVLSLRNTQEEPPDPQLM
comp122137_c6_seq1              LQQIMNITDQSLDEAQARKHTLNCHRMKTALEFVLCIIEKKTIVLSLRNTQEEPPDPQLM
Ag_exd2                          -----VLSLRNTQEEPPDPQLM
Cs_exd2                          LQQIMNITDQSLDEAQARKHTLNCHRMKPALFVLCIIEKKTIVLSLRNTQEEPPDPQLM
*****

Locus_1_Transcript_58675/166847  RLDNMLLAEGVAGPEKGGGAGAAANASAAAASSGPQSENAIEHSDYRAKLAQIRQIYHQEL
comp122137_c6_seq1              RLDNMLIAEGVAGPEKGGGAGAAANATAAAAAGPMSENAIEHSDYRAKLAQIRQIYHQEL
Ag_exd2                          RLDNMLIAEGVAGPEKGGGAGAAANASAAAASSGPQSENAIEHSDYRAKLAQIRQIYHQEL
Cs_exd2                          RLDNMLIAEGVAGPEKAGSAAAN---SAAAAGSPQSENAIEHSDYRAKLAQIRQIYHQEL
*****:*****.*:*      :***:..* *****

Locus_1_Transcript_58675/166847  EKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKLNSI-----
comp122137_c6_seq1              EKYEQACNEFTTHVMNLLREQSRTRPITGKEIERMVQIIHKKFNSIQVQLKQSTCEAVMI
Ag_exd2                          EKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIRVQLKQSTCEAVMI
Cs_exd2                          EKYEQACNEFTTHVMNLLREQSRTRPITPKEIERMVQIIHKKFNSIQVQLKQSTCEAVMI
*****:*****:***

Locus_1_Transcript_58675/166847  -----
comp122137_c6_seq1              LRSRFLDAR-----
Ag_exd2                          LRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYP-----
Cs_exd2                          LRSRFLDARRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCGITVSQVSNWFG

Locus_1_Transcript_58675/166847  -----
comp122137_c6_seq1              -----
Ag_exd2                          -----
Cs_exd2                          NKRIRYKKNIGKAQEEANLYAAKKAGKPDLCDDFF
    
```

Supplemental Table 4: Identities of spider *exd2*

	Locus_1_T...	comp122...	Ag_exd2	Cs_exd2
Locus_1_Transcript_5...		87.56%	98.39%	88.89%
comp122137_c6_seq1	87.56%		95.24%	88.62%
Ag_exd2	98.39%	95.24%		93.71%
Cs_exd2	88.89%	88.62%	93.71%	

11.4.2 homothorax

11.4.2.1 *hth1* paralog sequences

>comp102262_c0_seq1

```

TTTTTTTTTAACTTTCACTTTTATTTTTGAATATTTTTTTTTAAATTTAACAGGTTACAACAATACAATCATAAGCAGTCAATACAAATGCAG
AGTCACCTATATGAGAAAAGTTCATACAGATTACCACCGTTATATATATAATCAAACGGACAGAAGAAATAATAATAAATTTCCAATTCAA
ATATTTCTAATTTATCCCATCAAACGTGAAAAAAAATATACTGTATAACATTAATATAATAATTAATAAGTGCCATGATTTAAACAATAACGG
GATGTAACAAAAAAAATCATATATCCATCGCTAATAATGTCTCCATAAAGAAATAAACGGAAATCAATCACAGCACCATCATAGATAAAAATAATA
TGAGCATTTATTTAAATTTAAATCATTAATTTAAAAATATATTGCAGCAGCAGTAAGTGCCAATAATCATCAATATCTAATCACTATTGTAA
AAATCGAGAAAAAATGTAACACGACCGTTTTTTTTTTTTTTGTTCTTGGTAAATTTGCAAAAAACCAATAAAAAATAACACGCCGTAACACAT
GATTAATTTTATAAATATTACAACAACATTGAAAAATGTGCAGCTCCAAAATAAGAACAGAAAAAGCGAGCGAGAAGATACAGTATTGTATA
TTATATATGCAGTATTGTCTATTCTTTTTAAATTTCAACATTTTCAAATAATGTCCATACCGACATGACCACAATGAATTTCTCGGGGTAAGG
GGAAGTAAAAATTACACCGAATGAAATATAATGTGTATTATATTACATATAAAGAAGTGGCCAACAGTCAATTTCTCAAAAACCAACAGCAAAC
AATTGTCTATATAATAATAATAATAATAATAATAATAACCGGAGCATCATCATTTCAAGAGTTCGTTTTTTCATCTGATTGGACTGGACAATCA
ATTATTTTATCATGATCAAATAGAGATAAGAGATAAAAATACATTTATCCATTATACACACACAGTGGCATCGACATGTAGAGTAGAGATAATA
ATTATCATTACACATTTATTTATCATTATTATTACAACGGCAGCATCTATAAAGAGAAGGGTTATATATAAATTATGGCCATTTAGAAAACAAAAAT
AATAATTCCAATGTTTCAATCAAGCACAAAAAATTTTTTTGGGAGATTCATGTTTAAATTTATAGTTATAAATAATTTTTCGGCAACATGT
AAGTATTTTTTTGTTCTATAGTTTTGTCATTTCCCATGAAATTCAGAGACATAAATTTTTATGGGAATTTTTTTGGGAAGAGGGTTAAAAATTTTT
GCGGTAGATGAGGTGGGAAAAATCTAAAAAATATAGTTTAAATTTTTTTTTTCTGTTGGAACAATGAATTTTTTTCTCTAAAAAATTTTT
TTTTCCCTAAATTTTTGTATTTTCAATTTTTTAAAAAGGGACTTGTGTGTGATATAAGCTCCGAGGAAAAAATGGCTACTTGGCCGATTAT
TTTCATTTTTTTTTTTAGCAATAGCCCATTTTTGAGGAGAAAAGGGTTTTTTGTTGTTATTGTGCGTATAAATGTTTGTCTTGTGAAGTTGA
AATGTTTTAACTAGCGTGGATATCCATGATGTGCCACCGGAACCGTCTGACGGAGATCCTTGGCGTGGCAATCCGGGATGCCATGGGCGGT
GCGCCATCATCATAGCGTGGGATGTCCCGAAGAAGCATTTGCTTGCAGTGAACGGGTGAACGCAGCTGAGACATCTGAGAATATCCGCCATT
CCTCCCATGTGGCCATGCCATGGCTCCGTCACTGCATGAAAGATTTCGCATACCTGTGGGCGGATGTGCATCTGTGGCTTCCATCCATCATG
TATCCCATACCTGCGCCATCGGGTCCGTACGCAGCTGTAGCTCCACCAGCTCTGTTGGATTGGTCAATCATAGGCTGAACTATTCGCCGCTTGTGA
TTGATAAACCAATTTACTTGCAGAATCGTCAGACCAGTATCTTGAGCAAGTTGCTTCTTCTGATCTTCAGAAGGGTATGGATGCGTAAGATG
CTGGAAGACCCAGGCTCTCATGATATTCGTGGCGACCTTCGGGAAGATTCTCTCTTCTTCTGGCGTTTCTTACTTCTGTGCGTCTGCTCTTC
GCGGTTCCATCTCCGGATCCAATACTAGCATCGCCGTAAGTCGAGACATCGGCTTCACTGTTGTTGTGCGGTGCTCATCTGCGAACTCGGTTGTT
GGGATATAGTCCCGGTGTCCTGTGCAATCAGCGGGCGACCTTGCATCTCTCTCTCTGTGCTGTACGAATTCAGGGACTGCGACGGGGGCTCT
GGTCAGGCGTGCAGCAATTTGCCAAGTGTGACCAGTCTGATCGGGATGCTCTGCCACCATTTGCCACCAGCACCACCGCTTCCACCACCGGCG
CACCACCATTTGCCACTGCGCATTTGCTGCTGTTGTTGTTGTCACCGAGATCAACCGGTTTGGAAATCTTTTCGTGATGACCAAAATCGATGGGA
TCTTACCTTTAGGCAACTGATATATCGTTGACAGAAGTTGTCGCATAAAGTCTGTCACCTTTTCCAGTTTCAACAGATGGAATCTCAAGACCTGA
ATAGCTTGTACCATAAGGCTGTCTAATTCGGGATTTGGGGCATAATATGGTTTTCTTGCCTTATCTGTTTGGCAAATACTGTAATGTCTTCATT
GAAAGATTCGGAAGAGCAAACATCTCCACCAGCAATTCAGGCTCTGTTGGAGTAGACGTAGCTAATTCGCATTTCTCAAAAATAGAGCCAACA
ATGGAACAATGGATGTCCTGATATAGCGTCTTTGTGCGCTTGTGTGAACATCGGGTACGGAACCCATAATATGATTGGACATGACACCCGAC
ACATGATTACCCTGGTATTGATGCATAGATGGTGTGATTCATATGAGGAGCGTCCACAGCGATTGCATGGCCCTACTGTGGCCATGCGGATC
GTAGAGCGACCCAGTCCGTCCATGGACCCGTAATGCGACATACTTCTGCTACTGCATGGTGGCGAACGGGTGTTCCACAGGGTAAGGCTCAC
TCTTCATAAAGAGCTGCACACATTCCTACTTGAACAACCTCTATAAATTTAAATATTATCTATTTATTTTTTGTAGAAATTTCTCAAAA
AATTGATCTTTTCCCGTGAATAAATACGAAAAATCTTCTGGAATCACTTTTTTTTTAGATATATATATATCGATCGGAAGAGGTTAATGCCTGC
AATATTAATAATCTTTTATTTATTTATTTATTTTTCACAAGAAACCAATATTAGAGAAGTTGTCACGCAATTTATAGATTCAATGGCTTCC
TTTTACTTTTTCGCAAAATTTCTATCTAGTTCTTCAATCGAAGTTATTACTTCAACCGTAACAACAACCAACCTGTTGACAATAAATTTTCT
GTGAAGAAAAAATCCTTCAAGATGGTAGTAGTACCCTTATGTGTATAATAATAACACAGAAACACACAAAACACACTCACACACAAG
AAACACCACTAACCGGCTTCAAACCTGTTATCACACTTTTTCTCTCTCTTCTTGAAGAAAAAATTTATATATAATAATAATAAATAAATTTT
TTGAGCAGCAACACCAGGTCAGTCAATCGATGATGATGATGATAATAATAATGATAACAATAATATCATATATGTCAGTCCGAT
ATAAACACAGGAAATTTCCAATAAATTTTTCTTACCAGATCAATTTTTTTTTTTTTTCTCTTGGAGGAAAAAGGGATGAAATAAATTTTTTTT
TTTTAAATAATAAATAACGTGATCGATTAAATTAATTTTCAAATGCGAGGAATTTTAAATTAATTAATTTATTTTTTTTTTTGCAAGATCT
AAACGCGCGGAAGTTAAGAAGGAGCCGAGATGGAGAAGAAGTGCAAAGTGGCAATTTGAATGGGAAATGCGCTTTTTGCGCTTTTTTCCCTC
CCCGTGGGTGTTTTACCGCAGTTGGCGACGGCAGCGCTCCCGTGGCGAGCAAACTCAACTTGGAGATTATCTTCTATCTTTCCCACTT
TCTCTTTTCCCAAGGATAGGAAATAATGTGGAAAAAGAAATAACAATTTCACTTAAATTTTTTTTTTTACTAATTTATTTAAAAATTTATC
ATCCGCTTAAAAAATTTCTATAGATGATTTATTTCAATTAATTTCTTCTTACCGCTTTTTCTTCTCCCGTGTGAAAAATTTCCGATATTT
TTAGTAAAAAGAAGGAAAAAGTTTGAAGAGCTGATTGGTTAG

```

>comp102262_c0_seq1_-ORF_16_(frame_2)_translation

MKSEYPVVEHPFATMQYDEGMSHYGSMGPGSLYDPHGHSRAMQSLGHAPHMNHTPSMHQYHGNHVSGVMSNHIMGSVPDVHNKRDKDAIY
GHPLFLLALIFEKCELATCTPREPGIAGGDVCSSESNEDITVFAKQIRQEKPYYAPNPELDSLMLVQAIQVLRFHLLLEKVELCDNFCQRYISCLK
GKMPIDLVIDERDSKPGDLGDNNSNGSGNGGGAGGGGNGGRNPDTTGHSSDNSSTPDQRPPSLSNSYSTGGEDARSPADSTGTPG
PISQPPSSQMSTDNNEADVSTYGDASIGSGDGTGEDDDDDRSKQRKQKRGIFPKVATNIMRAWLFLQHLTHPYPSDQKKQLAQDTGLTILQVNN
WFINARRRIVQPMIDQSNRAGGATAAYGPDGAGMGYMMDGSQMHIRPPGMQNLSCSDGAMGMGHMGMGGYSQMSQLRSPVHSQAMLLPG
HPHAMMMAHGPMGHPGLPPQGSYPDGSGGHIMDIHAS

>Pp_hth1_RACE_Nt374_Nt267

TGTTTTCAAGTAGTAATGTGTGCAGCTTCTTTATGAAGAGTGAGCCTTACCCTGTGGAACACCCGTTGCCACCATGCAGTACGACGAAGGTAT
GTCGCATTACGGGTCCATGACGACGACCTGGGTGCTCTACGATCCGCATGGCCACAGTAGGGCCATGCAATCGCTGGGACAGCTCCACATATGA
ATCACACACCATCTATGCATCAATACCAGGTAATCATGTGTCGGGTATCATGTCCAATCATATTTGGGTTCCGTACCCGATGTTCAACAAG
CGGCACAAAGACCGCTATATACGGACACCCATTGTTCCATTGTTGGCTCTAAATTTTGAGAAAATGCGAATTAGCTACGTGTACTCCACGAGAGCC
TGGAAATTGCTGGTGGAGATGTTTGTCTTCCGAATCTTTCAATGAAGACATTACAGTATTTGCCAAACAGATAAGGCAAGAAAAACCATATTAT
GCCCAAATCCCGAATTAGACAGCCTTATGGTACAAGTATTAGGTCTTGGATTCCATCTGTTGGAACGGAAAAGGTGCACGAGTTATGCGA
CAACTTCTGTCAACGATATATCAGTTGCTGAAAGGTAAGATGCCATCGATTTGGTCATCGACGAAAGAGATTCCAAACCGGTGATCTCGGTG
ACAACAACAACAGCAGCAATGCGCGCAGTGGCAATGGTGGTGGCGCCGTTGGTGAAGCGGTGCTGGTGGCAATGGTGGCAGAGGCAATCC
CGATACGACTGGTACAGTTCCGACAATTCGTGCGCGCTGACCAAGTCTTTCGTTCCCTTACCAGAGGCCCCCGTCGAGTCCCTGAATTCGTACAG
CACAGGAGGAGGATGCAAGGTCGCCGCTGATTTCGACAGGGACACCCGGACCTATATCCCAACAACCGAGTTCGACAGTACGACCCGACAACA
ACAGTGAAGCCGATGTCGACTTACGTAAGAATATATTTCAATTTATTCATTTATTTAAAAACCAATATGTTTCTGTATTGTTGCTCAGTCT
TTCATGGTTTGTGGTGAATACAGTCAACATACACACACACACTCACTCACACTCTGTTATAATGACATTGCAGATTCCACGAAATTTGGT
CGTTATAACCGAGGCTGACATTTTCAAGACCTTTTACGAAAAATGTAATCAGGCGAAAATGAAAAATATACTACTGAGCATGAAGTGTATTTC
TTTCATCAAATTTATGTTGACCATATTTGATCGAGAAAAATGCGAGAAATGCGTGTGTGGAGATAATTCATACGCGACATACTTGATTCTCACC
AACATGTGGCAATAGTCGGTGAATCGCCAAATATTGCTTATGAGTACCCATGGCGAGAAAATACATTGCTACAGCAAGTCTGCTAAAAATGCA
TTTTACGGTGGAAAAATTGAACATGATCCAAATATGTGAGATGAAAAATTTATAATACAGTGTCTCAACCCCATGCACCACGATTTCAAATTAAT
TTGTAATAAATGGACTAAATAGCAGATCCAGATCACTTGAAGTGTCTGCGGCACAGTCCAAATCCCTCAAGCGGCAACGACATATTTTCA
GAACACTTGAGGGACCTTTCAGATTCCACATCACTGAAGCGGAGCACAACACGTTTTCAAGATAACTTTAGTATCTCACGGCACATTTACAGATCT
CAAGTAACTTGTGGTGTCTTCCAGATCACTTAGTGGCCCGGACAGCTTTTCAAATCAATGAAATGGCCCGGCTACTCTGCGTTGATACCACTG
CTTTC

>Pp_hth1_RACE_Nt374_Nt267

MKSEYPVVEHPFATMQYDEGMSHYGSMGPGSLYDPHGHSRAMQSLGHAPHMNHTPSMHQYHGNHVSGVMSNHIMGSVPDVHNKRDKDAIY
HPLFLLALIFEKCELATCTPREPGIAGGDVCSSESNEDITVFAKQIRQEKPYYAPNPELDSLMLVQAIQVLRFHLLLEKVELCDNFCQRYISCLKG
KMPIDLVIDERDSKPGDLGDNNSNGSGNGGGAGGGGNGGRNPDTTGHSSDNSSTPDQSFVYQRPSSQLNSYSTGGEDARSPADST
GTPGPISQPPSSQMSTDNNEADVSTYGNIFQFIHLEKFNQYVSCIVAQSFMV

>Locus_81_Transcript_57/63_Confidence_0.052_Length_1558

TCAGAAATCTTGAAGGATTGCGCCGGACAATAAGGGACGGTGCAGGAAATTTTCTTTAGCTTGTAGCCGGTGTGTGTTTGTGTATGTGTTG
TCATTTGCAAGTGATTTAAGAGGTCAAAGTATACTTATAAAAAATATAAAAAAGCACCATTGATACCGTCTTAAAGGAAAACCTGGATTT
TTCGTAAGGAAGTTAATAAACAACCTTCCGGAAGAAAAAAGTTTTCGAAAGTTTACTGTTTACCTGTTTCGGCGCAATCCTTCAA
GATTTCTGATTTGTATATTTATTTAATAATGCTGTGATCGATTTATTTTGTGATCCTTGAAGGATAAAAAATCGTCTGCCAAAAGGATAAATCG
TTTGAATATTAATTTAAATATCTTTTGAAGTGGTACGTTTGGTAAGTTAATCCTTTCGTCGACAGCTATCTTTGACAGATTCTGTAATTTG
AATGGCCATAACAGACCGTGTCTTATGAAGAGTGAACCATACCCAATTGACCATTCCTTGTCTATGAATACGAAAGATGGTATGCCGCACTACGG
CGGATGGAGCGTCAACGTCCTTATACGACCCGACCGGCATCGGCAATGCAACCCCTATCTCATGCCCCACATATGAATCACACACCCCTCAAT
GCACAGTATCACAGCAATCATGTATCCATGTCTAACCATATTATGGAACTATGACACAGATGTTTATAAGAGAGATAAGGATGCCATATATGG
CATCTCTGTTCCCTCTGCTAGCACTGATATTTGAAAAGTGTGAATGCGCCAGTGTACCCCGAGGAGCCCGGTATAGTGGCGGAGATGCTG
CTGCTCTGAATCTTTCAATGAAGACATAGCTGTCTTCCGCAACAGATTAGACAAGAAAAACCATATTATTTCTCCAAATCCGAGTTAGACAGTC
TTATGGTTCAAGCGATTCAAGTATTAAGGTTCCATTTATAGAGTTAGAAAAGGTACACGAATTTGCGGATAACTTCTGCCAGCGTTACATTAG
CTGTTTAAAGGCAAAATGCAATTTGATTTAGTATCGATGAAAGGACAGCAAACTGGTACTTGGGAGACAATAATAAACAGCAGTAAC
GGTGGTGGTGGTGGTGGTAAACAGTGGGCGAGGAGGTGGAAGGGTAACCCCTGACACCACTGGACACAGCTCAGACAACCTACCACTCC
AGACCAATCTTTTATACCTACCAGAGACCACCTTCAATCTCTAACTCTATAGTACCGGACCTGATGATGCCAGATACCCCGCGGATCTAC
AGAACTCCAGGCCATTTTCAACAACCTAGTTCAAACTTAGTACTGACAACAACAGTGAAGCAGGTGATGCGAGTATCGGGTCTGGTATG
GAACGGGTGAGGATGATGATGATGATAGGAGTAAAAACCTCAAAAGAAAAGAGGCATATTTCCAAAAGTAGCCACAACATAATGAGGGCAT
GGCTTTTCAACATCTAACGCATCCCTATCCATCCGAAGATCAAAA

>Locus_81_Transcript_57/63_Confidence_0.052_Length_1558_-ORF_1_(frame_3)_translation

MTDSCNLNGHNRPLLMKSEYPIDHSLAMQYEDGMPHYGMDGPTSLYDPHGHRAMQPLSHAPHMNHTPSMHQYHNSHVMSNHIMGTVPD
VHKRDKDAIYGHPLFLLALIFEKCELATCTPREPGIAGGDVCSSESNEDIIVFAKQIRQEKPYYSPNPELDSLMLVQAIQVLRFHLLLEKVELCD
NFCQRYISCLKGKMPIDLVIDERDSKPGDLGDNNSNGSGNGGGAGGGGNGGRNPDTTGHSSDNSSTPDQSFVYQRPSSQLNSYSTGPD
RSPAGSTGTPGPISQPPSSQLSTDNNEADVSTYGDASIGSGDGTGEDDDDDRSKQRKQKRGIFPKVATNIMRAWLFLQHLTHPYPSDQ

11.4.2.2 hth2 paralogs sequences

>comp120146_c0_seq1

TAAAAAGTTAACAGAAACCAAACTCTGTTATTACACGCACTACACGGAAAAACTTGGTGGGTGAAGTCGTTTACCGCGGAGCCCATCTGT
GAGAATTCGCGCGGCGCACCTTCCACAAGTTTATCCGCGGTTATCGGGTGGACTTCGGTGGTACCATTTCTCTCGTGTATTATTGTGCTTTT
TTTTCAATCGTTGAATAAATTTGGTTTCCGGCTTTCCTTTGGAATAATAAAAAACGACGTACGATAATGCAGTATAACGAAGACGGCATCCCTC

Supplemental Material

ACCTTACGGAGTGGACGGTGGCGGGCCGCGTCCCTTTACGACCCGCACGGCCGATGCCGAACCTGTCCCATCACATGAACCACGGACCGTCCA
ATAATCTCCATCAGTATGGCAACAGCTCATGTCAACATCGCCAACACCGTATGGGGTCCATGCCGGACGTTCAAAAACGGGACAAAAGCGCTATA
TACGGCGATCCATGTTTCCATGTTGGCGCTAATCTTCGAAAAATGTGAACCTGCCACTTGTACGCCGAGGGAACCCAGGCATACCCGGCAACGA
TGTCTGTTCATCGGAGTCAATCAACGAAGACATCGCAGTTTGGCCAAACAGATAGGCAGGAGAGACCTTACTACAGCCCGCAGGAACTCG
ACAGCATCATGGTACAAGCAATCCAAGTCTTGAGATTCATCTATTGGAATGGAGAAAAGTCCACGAGCTGTGCGACAATTTCTGCCAGAGATA
ATCAGCTGCCTGAAGGCAAGATGCCCATAGATTTGGTCATCGAGGAGAGGGACACCAAACCCGAACCTCGCGGACACCAACAACAACAGCAACGG
CAGCAGTACTGGGGGGCCCGGTGTGTCCCAAGGGTATGATGGACACTTCGGGACACAGCACGGACAGCGCGTCCACTCTGATCAGAGAC
CACCATCGCAGAGTTTGAATTACGGTCCGGTCAGCGATGACGTCCGGTCCGGCGGCTCCGGGGCACACCCTGCCATTTGCCAAACAGCCCGGT
CCCAGCAGAGCACGGACAACAACAGCGAAAGTCCGGTGAAGTGGGACGCCAGCATCGGATCTGGAGAAGGGACCGGCGACGAAGACGACGACGAG
AGCAAGAAGAACCAGAAGAAACGTGGGATATCCCAAGGTCTACTAACATCATGAGGGCTTGGCTCTTCCAGCATCTCACGCATCCATATC
CATCCAGGACGAGAAGAAGCAGCTCGCTCAGGATACGGGTCTCACGATCTTCAAGTCAACAATTTGGTTCATCAATGCCAGGCGTAGAATTTGTA
CAGCCGATGATGATCAGTCAACAGAGCAGGTGGCAGTACGGTCCACAGGAGCGCTTACAGCCCGGAGTCTTCTATGGGATACCTGATGGA
TGGTGTGCCGAGATGCATATACAGCAGGTCTCCAGGGATTTGCCGAGTTCGTTCGATGGGTCACATGGGCTACTCCAGCTCCGTTCTCTCTGTTCCA
TTCCAGGCCATGCTGATCCCGGACACCAGCCATGATGATGTCTCACCTGGGTTGCCGCTCCACCTCTCACGGATCGCTTATGACAGTTC
GCTCCGAACATCATGGACCTTACTCGAGCTGAGACTGTTCACTTCAAAAATATTATAAATTCACAATATATATAAATATATGATCTCGATG
TGTGAAAGAGATCGTTATGCGTGACTTAATATAACATAGATTTGTGGCGCCATCTATATATATGCTACCAGTTTTTTTTTTGTCTTCTCTTT
CCCATTTGCATCACTCACTCATCTGATGATAATATACGTATTTGTATGTACATAAACGAGAATTTATCGGTCGGTAAATGTGTAATAAG
TAATTTAGTGGCTCTTGCAGGCCCCGGCCGTAACAGCTCGAAGTTCTTAATTTCTTCAATATCTCAATTTTGTATTGCAATTT
GATTTCTTGTCTCTCATTGCGATTAGATATCTTTATAAATGTTGAAAAATGCGGGCCACAAAGTTTTTGGACCCATGTACATGTTGTGTTCAA
AAGTCTCGCTATAATGTGGGATTGATTTCCGATTTCTCCCATGTTGTCCAGCATGTGGCAATATTTGGCACATTCAGATGTAATACTAGAT
GCATTCCTAAATGAAGGCCGTTTTAAACACATGGACTCACTGGGGGAGGGGTGCACAAAGGGCCGGTCACTTGGCATCATTTGGCCGGGTA
GTAAGGCAAAATTTGTCCGGCATATGCGTGGTGAAGTGTATGTCGACTGTCGACACACATTTGAAATTTGATAGGATTTACTCACCCGCTGTG
TTTGTGTATATTAGAACTGCAGCAGGCTAGTAAGGGCCACAGTTCACATGAATCGGCATCAGTTGCAGCAGTCTATAAGGGCCACAGTCTCT
GCAGATTTGGCATTAAATAGCAGTAGGGTTTTCGGTGCAATTTGAAATGCTTGTGGTAGGTTTGAACAGTCTCTATCTATACGCATTTGCATTAA
TCGTGACAGGATTGTAAGACATGCCATTGTGTGCCATTGTGCCGTACATTTAGCACCAGTTATGACTGAATGTATGCGTCACTGTTACATA
CGTTAGCACAATAGTCGCCGGTGGAAAGAGGCCACATATCCACATGAGTCGTGATCAGTTGTAGCATGGTGAAGAGCCACAGTTTTCTGCAT
ATTTGCATTAATATCGGCAGGATTACAGGCGCATTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
GGGCCACTGCATTATACATAGTGGCACAACCTCTGGGAGTATTTATGACCACCTGTGTTACACACACACACACATAAGAAGCAATACGG
CCACAATCTCCACACATCGGCATCGCTTTTGACAGGTTGCAAGAAACATAGTTCTCTACAAAATTAATATTAATTTTCAGCAAGACTTGAACAGC
CCTATCTTTTGGCATTAGCATTGTTGGCAGAATTTCAACCACCGCTGCCATCAATTTATGACAAGCTAGTAATTTGCCACAGTCTCTACACATT
GTTGCCGGTAGGTTTGGGGCATATATGATATGATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
TTTGTCCGCATGCTATGCACATTCGCATCAAATGATGCGGGAAAAATTTGATTTGATGACATGCACATATTCATAGGCATCAATTTGTGTAAC
CAGTTAGGGTCAAAGTCTCTACATAATTAGAATCGATTGTGGAAGGGTGCACCTGTCACACAAATCGACTCCAGTTGTGCACATGCTTGAAGA
GCGAGCTTAAGTGGTACAACATAAATGGCTGTGCCACATCAGCCAAAACCCCGGAATGCCAGGAAGAGAGCCCGTAGTCCCCAGAATTTCTG
CGATCAATTTCCAGCTGAACACGTAAGTGTGGAAGCCACATCTTGTGCATTTCCCGGAAATAGCTTTGAATTTACTTAATTTTAACTTTT
GGAGTATAAAAAATGAGTATAAATTTGTAATTTTCCGTTTTTAAATGATTTTATATAGTTAAGGACTATAAATTTATCTTTTCATGTTGATGAAATTC
GAGGCACGAACGAAAAGGAATAATTTCTCAGAAATGTAACCTTGAATAATGTTGGAAATACATCGAGGGTATTTCTTATCTCTATACATTTTTTG
ATAAATATCAAAGTAGAAATTTATGTATAGTCCAATATTTGATGTAATAAAATCTTCTCTCTCTCTCAAAAAA

>comp120146_c0_seq1

MQYNEDGIPHPYGVDDGGPPSLYDHRPMPNLSSHMHGSPNNLHQYGNSHVNIANHVMGSMPPDVHKRDKDAIYGHPLFPLALIFEKCELATC
TPREPIGPNDCVSSSEFNEDIAVFAKQIRQERPYSPDEELDSIMVQAIQVLRHFLLELEKVELCDNFCQRYISLKGKMPIDLVEIRDTPKELGD
TNNNSNGSSYCGPPCVRPMMMDTSGHSTDSASTPDQRPPSLSLNYGVPVSDVRSASAGTPCPLSQPASQSQSTDNNEVGEWDSIGSGEGTG
DEDDDDRAKKNQKRRGIFPKVATNIMRAWLFLQHLTHPYPSQKQKLAQDGLTILQVNNWFINARRRIVQPMIDQSNRAGGSGIPPGASYSPE
SMGYLMDGVPMHIRPGLQLPLDSSMGHMYSQRLSPVHSQLMPLPGHHAMMMSHPGLPPPPHGGSPYDSSPPNIMDLHSS

>Locus_10999_Transcript_20/21_Confidence_0.221_Length_2489

CTCCAACTTCCAAAAGCGAAACAAATATGCAGGAAGAAAGTTTGTGATTTTCGCCAACCTCGCCAAGCATTCACAGATTATTTATCATCTCC
GTTCAAGTTGACGGTGGTGCAAAGAAAGGGTGGGTATAAAGCCCGCTACAGAAACCGGCTGAGCGAGGGCCGAAATTTCTCGTGT
TGTTTTGTATCGTCAGATCAACACCTGGTTGGAAGAGTAACGTTGTGCACATGCAATATCCTGAAGATGGGATGCCCTATTACGGCCATGGGGAT
GGATCGGCGGGCTATATGACCCCATAGGCAAAATCTTATGAACCACCATGGTGTATCATGCCAATCATGTGTCTATTGCTAATCATGT
CATGGGATCTACACCTGATGTGCAAAAAGGATAAAGATGCCATTTTGGACACCGCTTTTCTCTGTAGCCTGATCTTTGAAAAATGTG
AGCTGGCACCTGCACCCTAGGGAACAGGCATCGCTGGGGGAGAGCTGTGTTCTCAGAGTCTTTCAATGAGGATATCGCTGCTTTGAAAA
CAGATTAAGGAAGAAAGACCATTGATGATGCAATCCTGAACTAGATAGCCTTATGGTGAAGCAATCCAAGTGCCTTAGATTTCCATTTGTTAG
AATTGAAAAAGTTATGAGCTCTGTGATAACTTTTGTGACGTTACATCAGCTGCCTAAAGGGTAAGATGCCATTTGATTTGGTTATCGAAGA
GAGAGATACTAAACCTGAGTTAGGGGACACCAATAACAACAGCAACGGTAGCAGTTTCTGTGGTGGTCTCCTCTGTCTCAGGGGTATGCTGG
ATACTTCCGGCGACATAGTACTGATAGTGGCTTACACCAGATCAGGGCCACTATGAAGACATGTCTGTGGGAAATATGGAAGGGCTAGGGAG
ACCACCTTCAAACTCGAATATGTTTCCGTAGGCGATGATGTTCCGTCCCAACCGGTTCAACTGGCACTCCATGTCTCTGTCAACAACAGCC
TTCTTCCCAACAAAGTACAGCAATAACAGCTGAAGCTGGTGTGATGTCAGATATTTTTCAGGAGAAGGCTCAGGTGATAGGATGATGATGAAAGG
GGTGTGAAGAAAGGACAGGAAGAGGGGATTTTTTCTAAAGTAGGCACAAACATACTGAGGGCATGGCTATTTTCAGCATCTTACGCACCCGT
ATCCATCTGAAGATCAAAAAAGCAGCTGTCTAAGACACAGTCTTACGATTTCTACAGGTAACAATTTGGTTTATCAATGCCCGTAGAAGAA
AGAGCTTTTAAATTTCTGGCTCCAATTTGACTCATATTTTGGAGGATTTAAATTTTAAATCCGGCAGGAAAGCATAACCATGTTTTTGGAGGTA
ATTGATCTTAACCAACCAAGTAAATGCTTAAATTTACTCTTCAAAACCATAGTTTAAATAGTTGAAATTTGTGTCATTTACAGGTTTATCAATGCC
CGTAGAAGAATAGTTCAACCTATGATTTGATCAGTCAACAGAGCTGGAGGACTATGGGAACACCAGGACCGACATACAGCCCTGAATCTCTG
GTATGGGCTATATGATGGATGGAGTGTGCCATGCATATCAGATCATCCAGCCTGCAGAATTTGCCTTGGCCAGAGAACATGACCATTGGCCAT
ATGGCAGGATATTTCTAGCTCAGGTACCGGTCCACTCAAAAGGAATGCTTCTTCTGGACACCTATGATGATGCATGTTGGTGGTCCACCC
ACCTCCGATGGATCGCTTATGATAGTTCTCCCCCAGTAAATGACCTTGCACAATAGTTAGTTTATAAACAATATTTGATATGGATAAT
TTTGTGTGTGAGAAATTTGTCATTTTGTATACCTCCCCCAAAAACCTTAAATCATGTTTCATTTTGTAGTGCATAAATAAATTTTATCAAA
GTGTCATAAATTTATTTACTGTGGCTTCTCCACTGCAATAATGCTCGTCAATTTACAGGTTTTAAATTTAAGTTAAAGAAAGAAATTTT

TTTAGCATAACTTTAGTTCTTATCCTTTATTGTTAATAAAGTTCGTCTGAAATTAACATATACGTATTAACAATATAACTTAAAAACAATATT
 TTTAAATGCTACAAATTTAAACCACTTACTTAAAAAGAAATCAATTATTAAATACATTTTAGCTGATATATGTTGTTAATTTAATATAGTTTT
 TATCATTATAAATTTGTGATTTTATGTTTTAAAAAATAGTTTGTGTTCAAATAATTGTCGGAATGTTTAGTGATTTTTGTACAAATGAATG
 GTCATGTTAAGTTAAAGTGTGATCTTCTTAGAAAAATGTTTAAAAATGTGATGATGATTGCTGTAAAACGCTTTGTATATAAATATAACACT
 GAATAAATTAATCAGTATTAATTAATAAAAAA

>Locus_10999_Transcript_20/21_Confidence_0.221_Length_2489_-_ORF_2_(frame_3)_translation

MQYPEDGMPHYGHGDSAGGLYDPRQNLNMHHGVYHANHVSIANHVMGSTPDVVGKRDKDAIFGHPLFLLALIFEKCELATCTPREPGIAGGD
 VCSSESFNEDIAACFAKQIKERPLYDANPELDSLMVQAIQVLRFHLELEKVELCDNFCQRYISCLKGMKPIDLVIEERDTKPELGDNTNNSNGSSF
 CGGPPCVSRGMLDTSGGHSTDSGSTPDQGHYEDMSVGNMERLGRPPSLSNYGVSVDVRSPTGSGTGPCPLSQPSSQSTDNNEAGDASICSG
 EGSGEDDDDERGGKQRKRGRIFPKVATNLRWLFQHLTHPYPSDQKKQLAQDTGLTILQVNNWFNARRRIESLNSGNSCTHILRIKIFNPAG
 KA

>Pt_hth2_sk4_complete_extraction

GACAAGGATGCAATATACGGTCACCCGCTTTTTCTCTGTTAGCCTTGATCTTTGAAAAATGTGAGCTGGCCACCTGCACCCTAGGGAACCAGG
 CATCGCTGGGGGAGACGCTGTTCTCAGAGTCTTTCAATGAGGATATCGCCTGCTTTGCAAAACAGATTAAGGAAGAAAGACCATTGTATGAT
 GCCAATCCTGAACCTAGATAGCCTTATGGTGCAAGCAATCCAAGTGCCTTAGATCCATTGTTAGAAATGGAAAAAGTTCATGAGCTCTGTGATA
 ACTTTTGTCAACGTACATCAGCTGCCAAAAGGTAAGATGCCCATGATTGTTGTTATCGAAGAGAGAGATACTAAACCTGAGTTAGGGGACAC
 AAATAACAACAGCAACGGTAGCAGTTTCTGTGGTGGTCTCCCTGTGCTCCAGGGGATGCTGGATACTTCCGGCGGACATAGTACTGATAGTG
 GCTCTACACCAGATCAGGGCCACTATGAAGACATGTCTGTGGGAAATATGAAAAGGCTAGGGAGACCACCTTCCAAATCCCTGAATATGGTTCC
 GTAGGCGATGATGTTCCGTCACCGGTTCAACTGGCACTCCATGTCTCTGTGCACAACAGCCTTCTTCCCAACAAAGTACAGACAATAACAG
 TGAAGCTGGTGTGATGTTGTTTGTTCAGGAGAAGGCTCAGGTGATGAGGATGATGATGAAAAGGGTGGTAAGAAAAAGGAGAAAGAGGGG
 TATTTTTCTAAAGTAGCAACAACATAATGAGGGCATGGCTATTTACGATCTTACGCCACCGTATCCATCTGAAGATCAAAAAAGCAGCTTG
 CTCAGACACAGGCTTACGATTCTACAGGTCAACAACCTGGTTTCATCAATGC

>Pt_hth2_sk4_complete_extraction_translation

DKDAIYGHPLFLLALIFEKCELATCTPREPGIAGGDACSSSEFNEDIAACFAKQIKERPLYDANPELDSLMVQAIQVLRFHLELEKVELCDNFCQ
 RYISCLKGMKPIDLVIEERDTKPELGDNTNNSNGSSFCCGPPCVSRGMLDTSGGHSTDSGSTPDQGHYEDMSVGNMERLGRPPSLSNYGVSVDV
 RSPGTGTPCPLSQPSSQSTDNNEAGDASICSGEGSDEDDEGGKQRKRGRIFPKVATNLRWLFQHLTHPYPSDQKKQLAQDTGL
 TILQVNNWFN

Supplemental Table 5: Accession number, gene names and organism used for the phylogenetic tree of *hth*

Name	Accession	Description
Ag_hth1	CAS89757	Acanthoscurria geniculata hth1 partial
Ag_hth2	CAS89758	Acanthoscurria geniculata hth2 partial
Aga_hth	XM_308010	Anopheles gambiae str. PEST AGAP002178-PA
Al_hth	AGP03153	homothorax, partial [Archezogetes longisetosus]
Am_hth	XM_006564734	PREDICTED: Apis mellifera homeobox protein homothorax-like
Cs_hth1	CAD57739	Cupiennius salei-hth1
Cs_hth2	CAD57729	Cupiennius salei-hth2
Cscu_hth	AKN63488	homothorax, partial [Centruroides sculpturatus]
Dm_hth	AAN13474	Drosophila melanogaster hth isoform c
Dr_meis1b	NP_571968	Danio rerio-meis1b
Dr_meis2	NP_571971	Danio rerio-meis2
Dr_meis3	NP_571853	Danio rerio-meis3
Dr-meis4.1	AF376049_1	Danio rerio-meis4.1
Ek_hth	CAX63044	Euperipatoides kanangrensis partial mRNA for homothorax protein (hth gene), isoform A

Lp_hth	AKN63487	homothorax, partial [<i>Limulus polyphemus</i>]
Mm_meis1	NP_034919	<i>Mus musculus</i> -meis1
Mm_meis2	NP_001153040	<i>Mus musculus</i> -meis2
Mm_meis3	AAI17533	<i>Mus musculus</i> -meis3
Of_hth	AAS93633	<i>Oncopeltus faciatu</i> s hth
Ph_hth	CAO98908	<i>Parhyale hawaiiensis</i> hth
Po_hth	AKN63489	homothorax, partial [<i>Phalangium opilio</i>]
Tc_hth	NP_001034489	<i>Tribolium castaneum</i> hth

Supplemental Table 6: Identity of *hth1* paralogs to *D. melanogaster hth*

	Dm_hth
Dm_hth	
Pp_hth1_comp102262	50.9%
Cs_hth1	51.7%
Ag_hth1	53.0%
Pt_hth1_Locus_81	51.2%

Supplemental Table 7: Identity of *hth2* paralogs to *D. melanogaster hth*

	Dm_hth
Dm_hth	
Pp_hth2_comp120146	49.9%
Pt_hth2_Locus10999	47.3%
Ag_hth2	41.1%
Cs_hth2	45.9%

Spider *hth1* alignment:

	1	10	20	30	40	50	60
Pp_hth1_comp102262	MKSE	PYPVEH	PFATM	QYDE	GSM	SHYGS	M
Cs_hth1	MKSE	PYPVEH	PFATM	QYDE	GSM	SHYGS	M
Ag_hth1	MKSE	PYPVEH	PFATM	QYDE	GSM	SHYGS	M
Pt_hth1_Locus_81	MKSE	PYPVEH	PFATM	QYDE	GSM	SHYGS	M
Pp_hth1_comp102262	HQYH	GNHVS	GVMS	NHIMS	VSP	VDVH	N
Cs_hth1	HQYH	GNHVS	GVMS	NHIMS	VSP	VDVH	N
Ag_hth1	HQYH	GNHVS	GVMS	NHIMS	VSP	VDVH	N
Pt_hth1_Locus_81	HQYH	GNHVS	GVMS	NHIMS	VSP	VDVH	N
Pp_hth1_comp102262	AGGD	VCSSES	FNED	ITV	FAKQ	IRQE	K
Cs_hth1	AGGD	VCSSES	FNED	ITV	FAKQ	IRQE	K
Ag_hth1	AGGD	VCSSES	FNED	ITV	FAKQ	IRQE	K
Pt_hth1_Locus_81	AGGD	VCSSES	FNED	ITV	FAKQ	IRQE	K
Pp_hth1_comp102262	DNFC	QRYI	ISCL	KG	KMP	IDL	V
Cs_hth1	DNFC	QRYI	ISCL	KG	KMP	IDL	V
Ag_hth1	DNFC	QRYI	ISCL	KG	KMP	IDL	V
Pt_hth1_Locus_81	DNFC	QRYI	ISCL	KG	KMP	IDL	V
Pp_hth1_comp102262	RGNP	DTT	GHSS	DN	S	STP	D
Cs_hth1	RGNP	DTT	GHSS	DN	S	STP	D
Ag_hth1	RGNP	DTT	GHSS	DN	S	STP	D
Pt_hth1_Locus_81	RGNP	DTT	GHSS	DN	S	STP	D
Pp_hth1_comp102262	SQM	STD	NN	SE	A	D	V
Cs_hth1	SQM	STD	NN	SE	A	D	V
Ag_hth1	SQM	STD	NN	SE	A	D	V
Pt_hth1_Locus_81	SQM	STD	NN	SE	A	D	V

```

Pp_hth1_comp102262 HLTHPYPSEDQKKQLAQTGLTILQVNNWFINARRRIVQPMIDQSNRAGGATAAYGPDGA
Cs_hth1 HLTHPYPSEDQKKQLAQTGLTILQVNNWFINARRRIVQPMIDQSNRAGGASAAAYGPEGA
Ag_hth1 HLTHPYPSEDQKKTVGSRYRSHNS-----
Pt_hth1_Locus_81 HLTHPYPSEDQ-----

Pp_hth1_comp102262 GMGYMMDGSQQMHIRPPGMQNLSCSDGAMGMGHMGGMGGSQMSQLRSPVHSQAMLLPGH
Cs_hth1 GMGYMMDGAQQMHIRPPGMQNLSCSEGAMGMGHMGGMGGSQMSQLRSPVHSQAMLLPGH
Ag_hth1 -----
Pt_hth1_Locus_81 -----

Pp_hth1_comp102262 PHAMMMAHGPMGHPGLPPQGSFYDGSGGHIMDIHAS
Cs_hth1 PHAMMMAHGPMGHPGLPPQGSFYDASGGHIMDIHAS
Ag_hth1 -----
Pt_hth1_Locus_81 -----

```

Supplemental Table 8: Identity of *hth1* spider paralogs

	Pp_hth1_c...	Cs_hth1	Ag_hth1	Pt_hth1_Lo...
Pp_hth1_comp102262		88.372%	85.417%	84.054%
Cs_hth1	88.372%		89.324%	88.950%
Ag_hth1	85.417%	89.324%		95.472%
Pt_hth1_Locus_81	84.054%	88.950%	95.472%	

Spider *hth2* alignment:

```

          1          10          20          30          40          50          60
          |          |          |          |          |          |          |
Pp_hth2_comp120146 MQYNEDGIPHPYGVDDGGGPPSLYDPHRPMPLNSHHMNHGPSNNLHQYGNSHVNIANHVVMG
Pt_hth2_Locus10999 MQYPEDGMPHYGHGD-GSAGGLYDPHR-QNLMNH-----HGVIHANHVSIANHVVMG
Ag_hth2 -----
Cs_hth2 MQYGDDGIPHYAPPQDGPGLYDAHRHQGIPNH-----HSVYHPNHVAVANHVVMG

Pp_hth2_comp120146 SMPDVHKRDKDAIYGHPLFPLLALIFEKCELATCTPREPGIPGNDVCSSESFNEDIAVFA
Pt_hth2_Locus10999 STPDVHKRDKDAIFGHPLFPLLALIFEKCELATCTPREPGIAGGDVCSSESFNEDIAVFA
Ag_hth2 -----
Cs_hth2 SHPDAHKRDKDAIFGHPLFPLLALIFEKCELATCTPREPGIAGGDVCSSESFNEDITVFA

Pp_hth2_comp120146 KQIRQERFYSPDEELDSIMVQAIQVLRFHLELEKVVHELCDNFCQRYISCLKGMKPIDL
Pt_hth2_Locus10999 KQIKERFLYDANPELDSLMVQAIQVLRFHLELEKVVHELCDNFCQRYISCLKGMKPIDL
Ag_hth2 -----
Cs_hth2 KQVKEEKPFYVANQELDSIMVQAIQVLRFHLELEKVVHELCDNFCQRYISCLKGMKPIDL

Pp_hth2_comp120146 VIEERDTKPELGDTNNNSNGSSYCGGPPCVPRGMDTSG-HSTDSASTPDQ-----
Pt_hth2_Locus10999 VIEERDTKPELGDTNNNSNGSSFCCGGPPCVSRGMLDTSGGHSTDSGSTPDQGHYEDMSVG
Ag_hth2 -----
Cs_hth2 VIDERDTKPELGDTNNNSNGSSFCCGGPPCVSRGLDTSGGHSTDSASTPDQGHYEDIPVG

Pp_hth2_comp120146 -----RPPSQSLNYGVPVDDVRSAP-SAGTPCPLSQQPASQQSTDNNSEVGEWDASIGS
Pt_hth2_Locus10999 NMERLGRPPSQSLNYGVPVDDVRSPTGTGTGTPCPLSQQPSSQQSTDNNSEAGD--ASICS
Ag_hth2 -----RPPSQSLNYGVPVDDVRSAPAGTGTGTPCPLSQQPSSQQSTDNNSEAGD--ASIGS
Cs_hth2 SMERLGRPPSQSLNYGVPVDDVRSPPNGSSGTGTPCPLSQQPSSQQSTDNNSEAGD--ASVCS

Pp_hth2_comp120146 GEGTGEDDDDDRA-KKQKQKRGIFPKVATNIMRAWLQHLTHPYPSEDQKKQLAQTGLT
Pt_hth2_Locus10999 GEGSGDEDDDDERGGKKRQKRGIFPKVATNILRAWLQHLTHPYPSEDQKKQLAQTGLT
Ag_hth2 GEGSGDDDDDDARD-KKRQKTRGIFPKVATNIMRAWLQHLTHPYPSEDQKK-----
Cs_hth2 GDGSGDEDDDDERGG-KKRQKQKRGIFPKVATNIMRAWLQHLTHPYPSEEQKKQLAQTGLT

Pp_hth2_comp120146 ILQVNNWFINARRRIVQPMIDQSNRAGGSIGPPGASYPSSSMGYLMDGVPQMHIRPGLQ
Pt_hth2_Locus10999 ILQVNNWFINARRRISLNSGSNCT---HIL-RIKIFNPAGKA-----
Ag_hth2 -----
Cs_hth2 ILQVNNWFINARRRIVQPMIDQSNRAGGSIA-PGAAYSPESSMGYIMDGSSQMHIRSSSL

Pp_hth2_comp120146 G-L---PDSSMGH-MGYSQLRSPVHSQAMLI PGHAMMMSHPLPPPPHGSFYDSSFPN

```

Supplemental Material

```

Pt_hth2_Locus10999 -----
Ag_hth2 -----
Cs_hth2 QNLSCFENMAIGHMAGYSQLRSPVHSQAMLLPGHP-MMMSHPSLP--PPHSSPYDSSPPS

Pp_hth2_comp120146 IMDLHSS
Pt_hth2_Locus10999 -----
Ag_hth2 -----
Cs_hth2 IMDLHSS
  
```

Supplemental Table 9: Identity of *hth2* spider paralogs

	Pp_hth2_c...	Pt_hth2_Lo...	Ag_hth2	Cs_hth2
Pp_hth2_comp120146		72.457%	84.615%	75.309%
Pt_hth2_Locus10999	72.457%		90.291%	81.586%
Ag_hth2	84.615%	90.291%		88.235%
Cs_hth2	75.309%	81.586%	88.235%	

General *hth* alignment:

CLUSTAL O(1.2.1) multiple sequence alignment

```

Dm_hth -----MAQPRYDDGLHGYGMDSGAAAAAMYDPHAGHRP
Pt-hth1 MTDSCNLSNGHNRPLLMLKSEYPYIDHSL-AMQY-EDGMPHYGGM--DGPTSLYDPHGHRAM
Cs_hth1 -----MKSEFYPLEHSL-AMQY-EDGMPHYGAM--DG-GPMYDPHGPHRA
Pp_hth1 -----MKSEFYPVEHPFATMQY-DEGMSHYGSM--DGPGLYDPHGHSRA
Pp-hth2 -----MQYNEDGIPHPYGVDDGGPPSLYDPHRPMNP
Cs_hth2 -----MQYGGDDGIPHYPPQDGGPSLYDAHRHQGI
Pt-hth2 -----MQYPEDGMPHYGHGD-GSAGGLYDPHR-QNL
                          :::*      .      : ** *
  
```

```

Dm_hth PGLQGLPSHHSPTHAAAAAATVGMHGYHSGAGGHGTPSHVSPVGNHLMGAIPEVH-KR
Pt-hth1 -QP---L-SHAPHMNHTP-----SMHQYHSN-----HVS--MSNHIMGTVPDVH-KR
Cs_hth1 -QMQLSLG-GHGPHMNHTP-----LHQYHGN-----HVSVMGNHIMGSVPDVH-KR
Pp_hth1 --MQSL--GHAPHMNHTP-----SMHQYHGN-----HVSVMNSHIMGSVPDVH-KR
Pp-hth2 -----L-SH--HMNHGSPN---NLHQYGN-----HVN--IANHVMGSMPDVH-KR
Cs_hth2 -----P-NH-----HSVYHPN-----HVA--VANHVMGSHPDVH-KR
Pt-hth2 -----M-NH-----HGVYHAN-----HVS--IANHVMGSTPDVH-KR
                          *          *          :. ** : ** : * . **
  
```

```

Dm_hth DKDAIYEHPLFPLLALIFEKCELATCTPREPGVQGGDVCSSSEFNEDIAMFSKQIRSQKF
Pt-hth1 DKDAIYGHPLFPLLALIFEKCELATCTPREPGIAGGDVCSSEFNEDIAVFAKQIRQEKF
Cs_hth1 DKDAIYRHLPLFPLLALIFEKCELATCTPREPGIAGGDVCSSEFNEDIAVFAKQIRQEKF
Pp_hth1 DKDAIYGHPLFPLLALIFEKCELATCTPREPGIAGGDVCSSEFNEDITVFAKQIRQEKF
Pp-hth2 DKDAIYGHPLFPLLALIFEKCELATCTPREPGIPGNDVCSSEFNEDIAVFAKQIRQERF
Cs_hth2 DKDAIFGHPLFPLLALIFEKCELATCTPREPGIAGGDVCSSEFNEDITVFAKQVKEEKF
Pt-hth2 DKDAIFGHPLFPLLALIFEKCELATCTPREPGIAGGDVCSSEFNEDIAACFAKQIKEERF
*****: *****: * *****: ***: .: : *
  
```

```

Dm_hth YYTADPEVDSLMLVQAIQVLRFHLELEKVELCDNFCQRYISCLKGKMPIDLVIDERD TT
Pt-hth1 YYSNPPELDSLMLVQAIQVLRFHLELEKVELCDNFCQRYISCLKGKMPIDLVIDERDSK
Cs_hth1 YYSNPPELDSLMLVQAIQVLRFHLELEKVELCDNFCQRYINCLKGKMPIDLVIDERDSK
Pp_hth1 YYAPNPELDSLMLVQAIQVLRFHLELEKVELCDNFCQRYISCLKGKMPIDLVIDERDSK
Pp-hth2 YYSPEELDSIMVQAIQVLRFHLELEKVELCDNFCQRYISCLKGKMPIDLVEERDTK
Cs_hth2 FYYVANQELDSIMVQAIQVLRFHLELEKVELCDNFCQRYISCLKGKMPIDLVIDERDTK
Pt-hth2 LYDANPELDSLMLVQAIQVLRFHLELEKVELCDNFCQRYISCLKGKMPIDLVEERDTK
* : * : ** : ***** : ** : ***** : ** :
  
```

```

Dm_hth KPPELGSANGE-----GRSNAD--STSHTDGASTPD-----
Pt-hth1 -PGDLGDNNSNGGGGAGG-----GNSGAGGGRGNPDT-TGHSSDNSSTPDQS-----
Cs_hth1 -PGDLGDNNSNGGGGAGGSGNG--GNPGRGNPDT-TGHSSDNSSTPDQS-----
Pp_hth1 -PGDLGDNNSNGGGGAGGSGGAGGNGGRGNPDT-TGHSSDNSSTPD-----
Pp-hth2 -PE-LGDTNNSNGSSYCGPPCV-----PRGMDTSG-HSTDSASTPDQ-----
Cs_hth2 -PE-LGDTNNSNGSSFCGPPCV-----SRGLLDTSGGHSTDSASTPDQGHYEDI
Pt-hth2 -PE-LGDTNNSNGSSFCGPPCV-----SRGMLDTSGGHSTDSGSTPDQGHYEDM
* ** . * .          * . *          : * . ****
  
```

```
Dm_hth -----VRPPSSLSYGGAMNDDARSP-GAGSTPGPLSQPPALDTSDDPGKFLSSLN
Pt-hth1 ----FIPYQRPPSQSLNSYSTGPDDARSPAGSTGTPGPISQQPSSQLSTDNN-----
Cs_hth1 ----FLPYQRPPSQSLNSYSTGPDDARSPAGSTGTPGPISQQPSSQLSTDNN-----
Pp_hth1 -----QRPPSQSLNSYSTGGEDARSPADSTGTPGPISQQPSSQMSTDNN-----
Pp-hth2 -----RPPSQSLN-YGVPVSDDVRSAP-SAGTFCPLSQPPASQQSTDNN-----
Cs_hth2 PVGSMERLGRPPSQSLN-YGSVGDVRSVSPNGSSGTFCPMSQQPSSQQSTDNN-----
Pt-hth2 SVGNMERLGRPPSQSLN-YGSVGDVRSPTGSTGTPCPLSQPPSSQQSTDNN-----
          ****.**. . :*.*** :.* **:* : :.* :
```

```
Dm_hth PSELYDGRWCRREWSPPADARNADARRLYSSVFLGSPDNFGTSASGDASNASIGSGEG
Pt-hth1 -----SEA-----GDASIGSGDG
Cs_hth1 -----SEA-----GDASIGSGDG
Pp_hth1 -----SEADVSTYGDASIGSGDG
Pp-hth2 -----SEVGE---WDASIGSGEG
Cs_hth2 -----SEAGD-----ASVCSGDG
Pt-hth2 -----SEAGD-----ASICSGEG
          . **:* **:*
```

```
Dm_hth TGEEDDDA-SGKKNQKKRGIFFKVATNIMRAWLQHLTHPYPSEDQKKQLAQDTGLTILQ
Pt-hth1 TGEDDDDDRS-KKRQKKRGIFFKVATNIMRAWLQHLTHPYPSEDQ-----
Cs_hth1 TGEDDDDDRS-KKRQKKRGIFFKVATNIMRAWLQHLTHPYPSEDQKKQLAQDTGLTILQ
Pp_hth1 TGEDDDDDRS-KKRQKKRGIFFKVATNIMRAWLQHLTHPYPSEDQKKQLAQDTGLTILQ
Pp-hth2 TGEDDDDDRA-KKNQKKRGIFFKVATNIMRAWLQHLTHPYPSEDQKKQLAQDTGLTILQ
Cs_hth2 SGDEDDDERG-KKRQKKRGIFFKVATNIMRAWLQHLTHPYPSEEQKKQLAQDTGLTILQ
Pt-hth2 SGDEDDDERGGKKRQKKRGIFFKVATNIMRAWLQHLTHPYPSEDQKKQLAQDTGLTILQ
          :*:* **.* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*
```

```
Dm_hth VNNWF INARRRIVQPM IDQSNRAVYTPHPGSPGYG--HDAMGYMMDSQAHMMHRPPGDPG
Pt-hth1 -----
Cs_hth1 VNNWF INARRRIVQPM IDQSNRAGGASA---AYGPEGAGMGYMDGAQQMHIRPPGMQN
Pp_hth1 VNNWF INARRRIVQPM IDQSNRAGGATA---AYGPDGAGMGYMDGSSQMHIRPPGMQN
Pp-hth2 VNNWF INARRRIVQPM IDQSNRAGGSIGPPGASYSY-ESSMGYLMGVPMHIRPGLQG-
Cs_hth2 VNNWF INARRRIVQPM IDQSNRAGGSIA-PGAAYS-ESSMGYMDGSSQMHIRSSSLQN
Pt-hth2 VNNWF INARRRIESLN SGSNCT---HIL-RIKIFNP-AGKA-----
```

```
Dm_hth FH-----QGYPHYPPAEYY--GQHL-----
Pt-hth1 -----
Cs_hth1 LSCSEGAMGMGHMGGMGYSQMSQLRSPVHSQAMLLPGHPHAMMAHGPMPGHPGLPPQGS
Pp_hth1 LSCSDGAMGMGHMGGMGYSQMSQLRSPVHSQAMLLPGHPHAMMAHGPMPGHPGLPPQGS
Pp-hth2 L---PDS-SMGH---MG---YSQLRSPVHSQAMLLPGHPHAMMMSH---PGLPPPPHGS
Cs_hth2 LSCPENM-AIGHM---AG---YSQLRSPVHSQAMLLPGHP-MMMSH---PSLP--PPHS
Pt-hth2 -----
```

```
Dm_hth -----
Pt-hth1 -----
Cs_hth1 PYDASGGHIMDIHAS
Pp_hth1 PYDGGGGHIMDIHAS
Pp-hth2 PYDSSPPNIMDLHSS
Cs_hth2 PYDSSPPSIMDLHSS
Pt-hth2 -----
```

11.4.3 dachshund

1. Species abbreviations

Af = *Artemia franciscana* (brine shrimp)
Ag = *Acanthoscurria geniculata* (bird spider)
Cs = *Cupiennius salei* (American wandering spider)
Dm = *Drosophila melanogaster* (fruit fly)
Gg = *Gallus gallus* (chicken)
Gm = *Glomeris marginata* (pill millipede)
Hs = *Homo sapiens* (human)
Mm = *Mus musculus* (house mouse)
MMa = *Mesobuthus martensii* (bark scorpion)
Na = *Neanthes arenaceodentata* (polchaete worm)
Of = *Oncopeltus fasciatus* (milkweed bug)
Ot = *Onthophagus taurus* (horned beetle)
Pd = *Platynereis dumerilii* (polychaete worm)
Pp = *Pholcus phalangioides* (cellar spider)
Pt = *Parasteatoda tepidariorum* (common house spider)
Sk = *Saccoglossus kowalevskii* (acorn worm)
Sp = *Strongylocentrotus purpuratus* (sea urchin)

Tc = *Tribolium castaneum* (red flour beetle) Td = *Thermobia domestica* (firebrat)
Tl = *Triops longicaudatus* (fairy shrimp)

2. Primer sequences

Initial amplification of *dac2* fragments from *Pholcus phalangioides* have been performed with these degenerate primers:

dacPCR1-Fw: 5'-TGY YTN CCN CAR GCN TTY GAN YTN TT-3'

dacPCR1-Rev: 5'-ARN AMR TCC ATY TTN ARY TCN GYY TTY TC-3'

An aliquot of this PCR has been used in a second round of PCR with the following nested primers:

dacNest-Fw: 5'- CCN GTN GTN TGY AAY GTN GAR CAR GT-3'

dacNest-Rev: 5'-CKN GCR TTR TCN GCN GCN ACY TT-3'

Additional sequence information was acquired via RACE PCR using the following primers:

Pp-dac1-3'RACE-1: 5'-GGT GCT GCC GTG ACC TTG CCC TTG-3'

Pp-dac1-5'RACE-1: 5'-CCG GGC TGT GTC CAT TAG CAG AAG C-3' Pp-dac2-3'RACE-1:

5'-GCC CAT CTT CTC CAA AAC GGT TAC GGT C-3' Pp-dac2-5'RACE-1: 5'-CAT TAT AAG CAC CGT CAC TGC CGC TCA G-3'

3. Sequences and sequence accession numbers

>dac1_Parasteatoda_tepidariorum| Accession number LN877883
 MEGSSGEGSTSPSADMPSPSSLQNSPNPDSSVMVSDCITSVSNNGGLMCGPPMVSLSPKTIINNPVSMHGMGVSS
 PSYQQRTSSASPPQMVSKPYTPPMSTAGAMNALVVKQPQLEHMPAHMGMLAMRNNGGVHHHGYLEGAPPRLERPMHP
 ISPPGITTDATANDCRLIDYRGAKVAAFLVNGDYLLCLPQAFELFLKHLVGGGLHTVYTKLKRLDITPIVCNVEQV
 RILRGLGAIQPGVNRCKLLSCKDFDTLYKDCTTARPPGRPPKRATMVGINHNGTNPSSMLLKKSRMDGEYPGYENGH
 IGDRVDKSHLLANGYSHHVAAAQAATAATGGAPPHLNPLPFMALNHAAAAAAHNSMLSGTLPLAATGAHPSSA
 QGLNSSSHGHLNSSSSSGRGPESSSVIKERTSHNNDVINSTRLRDDRDIIVDSKDRLYGYDNHRMKDQAFNLGYF
 WLLAGAQAANGHSVPLNLSQHSSRPSNSNNNPITNGPGSGGGGTGGEHSGSENDGPDNDSDVDCDDDDSERDVS
 DNPVDSSTANTDRLTSSQQQLGYPAMGGDLSSLQAQAASSMETLLRYIQGLLKVAADNAKQOERQISLEKAELKM
 ELLREREVREGIEKQLMDEQORTRILYQKRLKKEKRTRRRVQEQLEAEVKKRAQYEEALRSNSAETLRLLNESLAQ
 ELERNARAEAEHKMQDCPMSV

>dac2_Parasteatoda_tepidariorum| Accession number KP725072
 MDVQGQMSPTTGMTSSPRSLQTPPNLDNMTLGGGDGNPSSNGGMVSPKQHVTVGLSQMGSPGSSSLGRSSSVSPPG
 PALSLVTTTKPYQASSSQPNGGLPPSLPRSCGQNLMPHHQVSHQPOMGHPPLGYHRSHPSSSPPGIAADMTTNDNR
 LIDYRGAKVAAFLVHGDYLLCLPQAFELFLKHLVGGGLHTVYTKLKRLDITPIVCNVEQVRILRGLGAIQPGVNR
 KLLSCKDFDALYKDCTTARRCFSSAGRDIPHHELAATSSIYGPIDNSSRPPGRPPKRVSMVSMTSGVTTTGMVKKSR
 LDDSYHSYENGHITGDRLDKHMPLNGYGHGPGTSHLSPLPFMALNSSHHNSVITGGVSMATTATHSSSHASSRGDV
 TRERQMGSDVIASTRLRDDRVDASDGKDRLYDVSQKDSYVNGTGNNGHSPVLNLSQNSRGGGGGGGGVGLH
 SAEGSGSEGPYNDGADDDYSDDDDDDRDQDFSDHPDVSSTANGERLAPSQMFPSHLAVDGGSPGTQOVVSSIETL
 LRNIQGLLKVAADNARQOERQINIEKAELKMEVLRERELRETLEKQLLDEQORTRILYQKRLKKEKRSRRRFQEQ
 ESEVKKRAQYEEALRTTSADTIRLLNDTLTQELERERNSLAESENNAHDCQMTA

>dac1_Pholcus_phalangioides| Accession number LN877884
 MEGSSAEGSTSPSGDMGSPPSIQNSPNPDPSGMVNDCATSSAQVGMCPMMVSLSPKTNQASAMQNLVSSPNST
 GEMRHQQHPSARTSPNSPQISVMHAVSSKPYTPPMSTAGVMSALASASAGLHKHDTPPTHIPSHHATAMGMLS
 MRNGVHHHGYSAPRLERPMQPLSPPGITTDATANDCRLIDYRGAKVAAFLVNGDYLLCLPQAFELFLKHLVGGGL
 HTVYTKLKRLDITPIVCNVEQVRILRGLGAIQPGVNRCELLSCKDFDTLYKDCTTARPPGRPPKRATMVGMSPNGA
 NHSILLKTKIEGDYPVYENGYISGDRVDKSHLLANGYSHHVAAAHAATAATGAAPPHMNLPLPFMALNHAAAAA
 AHNSMLSAASGAATPLAATGPHANAALNSSHGGLNSSSTSSGRNAESSSTVIKERPSLGNEVNSTRARDDR
 PDAPDSKDRLYGFDSHRMKDQAFNLGYFWLLAGASANGHSVPLNLSQHSSRPSGNGTNGPTSAVVTTGGDHSGSEN
 AYNDAAADDDNDSEDDDDDERDQDLSDNPVDSSTANTDRLISSQQALAYPNVGPGESGALAAQTASSMETLLRNI
 QGLLKVAADNARQOERQINIEKAELKMELLREREVREGIEKQLLDEQORTRILYQKRLKKEKRNRRRVQEQLEAEV
 KKRAQYEEALRNSSAETLRLLNDCPMSV

>dac2_Pholcus_phalangioides| Accession number LN877885
 MEVPGIISASSPAAITSPRSLQTPPNLDNMTLTNENNSINGPSMGVPVGMGSPKLSAGHMQNMSSSPGSSAGRSSSV
 SPPGPALSLVSKSSYHTSQSHSHLQAAVNGVTTSQLSRTCPPGVLTSHPSHQLHPPPLAYHRAMHPSSPPGIAAD
 MTTNDRCRLIDYRGAKVAAFLVSGEYLLCLPQAFELFLKHLVGGGLHTVYTKLKRLDITPIVCNVEQVRILRGLGAI
 QPGVNRCKLLSCKDFDALYKDCTTARRCFSSAGRDLPHEHELAATAATSLYGPIDSSSRPPGRPPKRVPSMVGIGGS
 QVVVSSASLLSKKARLDQYHYHYENGHITSRLEKAHLQNGYGGQITTSHVIVSSALSFPMALSSHHGITNPSSI
 TTAANTHGGNVPTSRQDTRVSHITSDSLSSNRVREDRLDGSSTKERLFYDVSRTKDAYINGASTNNGHSPVLNLS
 SQNSRNVVVGSGTSNNLNHPDQLSGSDGAYNDNGDDDDYTDGDDDDRDQDFSDHPDVTSTANGGSSTERIVPSAS
 GQQIFPFVGGHLALDTGAVSPTHQVAISSIETLLRNIQGLLKVAADNARQOERQINIEKDAFNIIPITSEVERLDT
 DSVLNRA

>gi|313661346|ref|NP_001186375.1| dachshund homolog 1 [Gallus gallus]
 MAVPAALIPAPLVPVPPVSTSACTTTTTSSAAATSSPAPAIAPPPASGTNLFREPIAAPAAAAA
 TVTSTATTTTTGGSGGGGGSGGGGSPSLGSGTGGGTSGTSTPNAGAAGSLPGKPVYSTPSPVENTPQNN
 ECKMVDLRGAKVASFTVEGCELIPLQAFDLFLKHLVGGGLHTVYTKLKRLDITPVVCNVEQVRILRGLGAI
 IQPGVNRCKLISRKDFETLYNDCTNASSRPPGRPPKRTQSVTSPENSHIMPHSVPLMSPGIIPTGLTAA
 AAAAAATNAIAEAMKVKKIKLEAMSNYHANNNQHAGDSENGDLNSSVGLLELPMMPHPLIPVSLPPA
 SVTMAMSQMNLSTIANMAAAAQVQSPPSRVETSVIKERVDPSPSPAPSLEEGRRPGSHSPSSHRSSSVSS

Supplemental Material

SPARTESSSDRIPVHQNGLSMNQMLMGLSPNVLPGPKEGDLAGHDMGHESKRIHIEKDETPPLSTPTARDS
LDKLSLTGHGQPLPPGFPSFPLFPDGLSSIETLLTNIQGLLKVADNARAQEKQVQLEKTELKMEFRER
ELRETLEKQLAVEQKNRAIQKRLKKEKKAKRKLQEALEFETKRREQAEQTLKQAASDLSRVLNDSLTP
EIEADRSGGRTDAERTIQDGRLYLKTVMY

>gi|45382177|ref|NP_990126.1| dachshund homolog 2 [Gallus gallus]
MAVSAPPVIAATSSGGGAGGSAGLFRADPLYSSPAESPRLTNSLVNSFLSAGGGGTGAGGGGGGGNECKM
VDLHGKVASFLVEGQELICLPQVFDLFLKHLVGGGLHTVYTKLRDLISPVVCTVEQVRILRGLGAIQPG
VNRCKLITRKDFETLYNDCTNASSRPPGRPPKRSGLVAIQENARLLPHGVPGLLSPGLISPTGLTAAAMAE
AMKLOKMKLMAMNSLHSGSQNGTESENEELNSNAGGSESSWDKDKLQSPITTSQHSIGHPTLSGQSSL
GSAHPLSPLQQNHLLTNRIDLPMMPHPLLPVSLPPASVAMAMNQMNHLNTIANMAAAAQMHSPLSRAG
ASVIKერიQDSSPAPSLEDSQRPGSHASSHQSSSVSSPSQLDNTPDRIAMLTNSREGELIDQETGTSL
KKMQEKEEVQIAIPIPKPTLDKVQLAGQALPPGFPAFPLFADGLSSVETLLTNIQGLLKVADNARVQE
KQIQQEKELKMELCRERELRESLERQLTAELOSRATIQKRLKKEKKAKRKLQEALEFESKRREQVEQAL
KQATSGDGLRMLNDAGIPDMEIEHNGTQHDSAAQENRYYIKPTIMY

>gi|629266050|ref|NP_542937.3| dachshund homolog 1 isoform a [Homo sapiens]
MAVPAALIPPTQLVPPQPPISTSASSSGTTTSTSSATSSPAPSIGPPASSGPTLFRPEPIASAAAAAATV
TSTGGGGGGGGGGGGGGGSGNCGGSSNCNPNLAAASNGSGGGGGGI SAGGGVASSTPINASTGS
SSSSSSSSSSSSSSSSSSSSSSCGPLPGKPVYSTPSPVENTPQNECKMVDLRGAKVASFTVEGCELIC
LPQAFDLFLKHLVGGGLHTVYTKLRLEITPVVCNVEQVRILRGLGAIQPGVNRCKLISRKDFETLYNDCT
NASSRPPGRPPKRTQSVTSPENSHIMPHSVPLMSPGIIPPTGLTAAAAAAAATNAAIAEAMKVKKIKLE
AMSNYHASNNQHGADENGDMNSSVGLLELPFMMMPHPLIPVSLPPASVTMAMSQMNHLSTIANMAAAAQV
QSPPSRVETSVIKERVPDSSPAPSLEEGRRPGSHPSHRSSSVSSSPARTESSSDRIPVHQNGLSMNQ
LMGLSPNVLPGPKEGDLAGHDMGHESKRMHIEKDETPPLSTPTARDSLDKLSLTGHGQPLPPGFPSFPLFP
DGLSSIETLLTNIQGLLKVADNARAQEKQVQLEKTELKMDFLRERELRETLEKQLAMEQKNRAIVQKRL
KKEKKAKRKLQEALEFETKRREQAEQTLKQAASDLSRVLNDSLTP EIEADRSGGRTDAERTIQDGRLYL
KTTVMY

>gi|16876441|ref|NP_444511.1| dachshund homolog 2 isoform a [Homo sapiens]
MAVSASPVISATSSGAGVPGGLFRAEPLYSTPREPPRLTPNMINSFVNNHNSNSAGGGGRGNTTNTNECRM
VDMHGKVASFLMDGQELICLPQVFDLFLKHLVGGGLHTVYTKLRDLISPVVCTVEQVRILRGLGAIQPG
VNRCKLITRKDFETLFTDCTNARRKQMRKQAVNSSRPPGRPPKRSGLVQLQENARLLTHAVPGLLSPGLI
TPTGITAAAMAEAMKLOKMKLMAMNTLQNGSQNGTESEPDLLNSNTGGSESSWDKDKMQSPFAAPGPQH
GIAHAALAGQPGIGGAPTLNPLQQNHLLTNRDLPLFMMMPHPLLPVSLPPASVAMAMNQMNHLNTIANMA
AAAQIHSPLSRAGTSVIKERIPESPPAPSLEENHRPGSQTSSTSSSVSSPSQMDHHLERMEEVPVQI
PIMKSPLDKIQLTPGQALPAGFPFPFIFADSLSSVETLLTNIQGLLKVADNARIQEKQIQQEKELRLE
LYREREIRENLERQLAVELQSRITMQKRLKKEKTKRKLQEALEFESKRREQVEQALKQATTSDSGLRML
KDTGIPDIEIENNGTPHDSAAQQGNYCYCLEMAQQLYSA

>gi|6681129|ref|NP_031852.1| dachshund homolog 1 isoform 1 [Mus musculus]
MAVPAALIPPTQLVPPQPPISTSASSSGTTTSTSSATSSPAPSIGPPASSGPTLFRPEPIASSASSSAAA
TVTSPGGGGGGGGGGGGGGGSGNCGGSSNCNPSLAAGSSGGGVSAGGGGASSTPITASTGSSSSSSSSSS
SSSSSSSSSSSSSSSSSSCGPLPGKPVYSTPSPVENTPQNECKMVDLRGAKVASFTVEGCELICLPQAFDL
FLKHLVGGGLHTVYTKLRLEITPVVCNVEQVRILRGLGAIQPGVNRCKLISRKDFETLYNDCTNASSRPG
RPPKRTQSVTSPENSHIMPHSVPLMSPGIIPPTGLTAAAAAAAATNAAIAEAMKVKKIKLEAMSNYHA
SNNQHGADENGDMNSSVGSWDETLHSPSQSQAPVAHARMPAAFSLPVSHPLNHLQSHLPPN
GLELPFMMMPHPLIPVSLPPASVTMAMSQMNHLSTIANMAAAAQVQSPPSRVETSVIKERVPDSSPAPS
LEEGRRPGSHPSHRSSSVSSSPARTESSSDRIPVHQNGLSMNQMLMGLSPNVLPGPKEGDLAGHDMGHE
SKRIHIEKDETPPLSTPTARDSIDKLSLTGHGQPLPPGFPSFPLFPDGLSSIETLLTNIQGLLKVADNAR
AQEKQVQLEKTELKMDFLRERELRETLEKQLAMEQKNRAIVQKRLKKEKKAKRKLQEALEFETKRREQAE
QTLKQAASADLSRVLNDSLTP EIEADRSGGRADAERTIQDGRLYLKTVMY

>gi|217035136|ref|NP_001136042.1| dachshund homolog 2 isoform 1 [Mus
musculus]
MAVSAPPVIAATSSSAGVPGGLFRAEPLYSSGPEPPRLTPNMINSFMANHNGSVLGGGIGGGSGGSSNT

NTNECRMVDMHGKVASFLMDGQELICLPQVFDLFLKHLVGGHTVYTKLKRLDISPVVCTVEQVRIIRGLG
 LGAIQPGVNRCKLITRKDFETLFTDCTNASSRPPGRPPKRSGLVLDQDNARLLPHAVPGLLSPGLITPTGIT
 AAAMAEAMKLLQKMKLMAMNTLQNGSQNGTESEPDDLNSTTGGSESSWDKDKIQSPLAASGPPQHGHIAHAA
 LAGQPGLGGAPTLPNPLQNNHLLSNRLDLPFMMMPHLLPVSLPPASVAMAMNQMNHLNTIANMAAAAQIH
 SPLSRAGASVIKERIPESPAPSLSEESHRRPGSQTSSHPSSSVSSSSPQMDHHSERVMMPNNREELIVD
 QDNGQSIKKFQRDNKEEVPAQIPVMKSPLDKIQLAPGQALHPGFPFPFI FADSLSSVETLLTNIQGLLKV
 ALDNARIQEKQIQQEKELRIELFREREIRENLERQLAVELQSRSTMQRKLLKKEKAKRKLQEALEFESK
 RREQVEQALKQATSGDSGLRMLKDSGIPDIEIENSGTPHDSAAMQA

>gi|642919228|ref|XP_008191786.1| PREDICTED: dachshund homolog 1 [Tribolium
 castaneum]

MDNSGMDASSEHSSHASGSPAVAVKPSPPAANSNTHQVHSPRQASPLGPPMSMAHPLSPPPVSASL
 ARSMALGHLPPPGLLNSLGLVLSPLDLMAHHPRSYNSPPPISTSPTANECKLVDRGQKVAAFI I
 AGDTMLCLPQAFELFLKHLVGGHTVYTKLKRLDIVPLVCNVEQVRIIRGLGAIQPGVNRCKLLSCKDFD
 TLYRDC TTASSRPPGRPPKRAPVGLSLAASHLQQQLKQLRDLNGDYPYENGHMGAEFPLDIMRRDMSRL
 EKSPLLANGYNHPHLSHMQFMQLPHPAASHLQQLKQLRDLNGDYPYENGHMGAEFPLDIMRRDMSRL
 CRAAYEDIVKHLERLREERCDDPDRPLPLDQKARDLSPRNGSPTDHSVPLNLSKSGGGSAGSLGDGDSGS
 EADGPDVPPSPRSPRASAEDDDDDNISDPEDDDDDKQDMDEGDLPLPAAPGGDSQQAALNYSTLASAVAS
 ANGPSQDPSISSTETLLRNIQGLLKVAADNARQQERQINYEKAEKMDVLREREVVKDSLRLVDEQKMR
 ALVQKRLRRERRARRRLQDQLEAEVKRRTLMEEALKAAGAAEQRLRI INEKVAPSEQKLSNPKVTTSSSTPN
 QSISSESERVSEVERMERERVESPPPAYGQQTREPPPPPPENKWPWGYSGIDLMTGAAFWQNYSESLA
 QELELERKSRQQQVERDVKSPLQDRSGYYKNSMLFTSTAT

>gi|624296|gb|AAC46506.1| Dachshund [Drosophila melanogaster]

MDSVTSEQSENNNNKAGAINNNINSI ERAAGNMNTTAMLHMQQQQAAAAAAAAAAQHQHQHNGRDG
 SNSRRGANEDRPNRGSSTGRGSSCSPASSPSRHPASVSPVSSLNHSMMQMQQQQQQQQQQQQQHQ
 LSPPPHGMPSGNLPTGLPPRMPHGLPPLSLGLLNSLQMMHHSPLLEMAAHHHVPVRSYNSPPPISTS
 DPSANECKLVEYRGQKVAAFI ISNETMLCLPQAFELFLKHLVGGHTVYTKLKRLDIVPLVCNVEQVRI
 RGLGAIQPGVNRCKLLSCKDFDILYRDC TTARCLSIKPPESNSLQFRSSRPPGRPPKRGVGLSLPPTHLS
 QHPQLKKHRLDNGDYAYENGI SDMKSPLLANGYNPPP INHMAFMQNAHHPGAAALMSPGMPHGLHAR
 PESQMLKAAAQNAAGMSAANMDALARS GIWENCRAAYEDIVKHLERLREERTDERQQAMGVGGAVVDRVGN
 VMVAEHKPRDLSRNCSPTRQSPVLNLSKSGGNTDHGGSNCDAGSERSDCHSVAGSPARGGSRSLDEGSR
 SGVGGGVASHVGGGI IGVEDDEEEENLSDENQSEVDERCLAKDDEDLSDTERDNLSTGSAAAAAAAAA
 AAHLQHQAASHHHQTVGLSHLGVVPPHQRGSPSSAEAAAAALQHQRALNYSQLAAAAAVANGAAVGGGA
 VANGPTGGGALT PNEALLAANDAAALAGGLALGPLGIDAHAAVPASSTETLLRNIQSLKVAADNARQQ
 ERQISYEKAEKMDVLREREVVKDSLRLVDERKLRVLYQKFRFRERKIRIRYQQQLGGGSGAVKGPNA
 NGSGGGSGSACSESEACGGGGGDAKCNNNNNNNNNNNNSNSSSHGGDVETMKENDAGSEKSDKSLASNS
 SCDVTAGHSAGSVSGSISAGLNC PDSPTHFKREPSDHEGVSVERQPRSSAASLAHDEDNIKRCGSRDRD
 ERPPSVAASVAAEGSLAAAAAAAAAAAAAAAAASNGKGPWSYPGIDLMTGAFWQNYSESLAQELEMERS
 RAANAERDVKSPLSERPTAYYKNSVLFGSAN

>gi|38490436|emb|CAD57736.1| dachshund, partial [Cupiennius salei]

RLERPMLVSPPPGLTTDATANDCRLIDYRGAKVA AFLVNGDYLLCLPQAFELFLKHLVGGHTVYTKLKR
 LDITPIVCNVEQVRIIRGLGAIQPGVNRCKLLSCKDFD TLYKDC TTASGCCRPGRPPKRATMVGINHNGA
 TGHGMLLKKPRMDGEYPGYENGI IAGDRVDKSHLLANGYSHHVAAAQAAVATHLNLPLPFMALNHAAAAA
 AHHNSMLTGALPLAATGAHTPGGALNTSTSSHTLSASTPARGPESSSVIKERTSHGNDVINSTRRLRDER
 VDLVDSKERLYGFD SHRMKDQAFNLNGYFWLLAGGSANGHSVPLNLSQHSSRPSNSSGNNLNNTTNG
 PGGPNSGEHSGSENAYNVEGEDDVNDSEDDDDDDREEDLSDNPVSSSTANTDRLTSSQQQLAYPAMMGDM
 GPLPGQTASSIETLLRNIQGLLKVAADNARHQERQISLEKAEKMDVLREREVVKDSLRLVDERKLRVLY
 YQKRLKKEKRSRRRVQEQLEAEVKKRAQYEEALRSNSQETLRLNLSLAQELERERNARAEAEHKMQDCP MSV

>gi|229614606|emb|CAS89751.1| dachshund protein, partial [Acanthoscurria
 geniculata]

RILRGLGAIQPGVNRCKLLSCKDFD TLYKDC TTARPPGRPPKRATMVGMSPNGPTPAMLLKKSRLDGEYPA
 YENHIGSDRVDKSHMLANGYSHHVAAAQAAVAAATGAPPHINPLPFMALNHAAAAAHHNSMLSSGVSL
 PLAATAHGHGLSTSRSDTSVIKERQSHGSDVINSSRLRDERGEPAESKDRLYGFD SHRMKDQAFNLNGYFW

Supplemental Material

LLAGASAANGHSPVLNLSQHSRPNATCNGTATGDQSGSETAYNEGPDDDDSEDDDDREQDLSDNPDV
SSTANTDRILTSTQQUALTYPNVSGDANALAAQTASSMETLLRNIQGLL

>gi|91982995|gb|ABE68635.1| dachshund, partial [Triops longicaudatus]
RILRGLGAIQPGVNRCKLLSCKDFDILFKDCTTARPPGRPPKRSPLMGMTSPHCDALMKMKRPRLDNGDYM
DYGNPVDNRIEKSPLLANGYNPPPTHLNPMQFMALNQAAAAAAHSALLGSPGMPGSMGSLALGPA
ARESLSGNTGGNHSSAAGLPGANGSNPSGPNPVLSPFGPIPGSGGVPRPDGSNVAKAQPPPLALEHMTA
ARTGFWENCRAAYEDVVKHLERLKERSEMERNVESRGRDFTTHNGSGTSPILNLSKTPSAGGLNGTEEG
DRIERGPRDTSADSVSEGARTPHRDDDEEMDRGSDIEEDDKNDMSDAIERGSAHHSDKSVVSNPAGNNHA
AAAAAAAFNYSAALLGATVTDPTAMLTSTESLLRNIQSLLKVAADNARQQDRQANYEKAELKMEILRE
RELRESIEKQVAEEQIRIRLLYQKRLKERRARRKLEQMDPDGKRRAQLEEVVKNASSGEALKLISGER
DSHCLKEI

>gi|91982993|gb|ABE68634.1| dachshund, partial [Artemia franciscana]
RILRGLGAIQPGVNRCKLLSCKDFDALYRDCTTASSRPPGRPPKRSPLLGNSPSSSQTEGFKKPRIESDYS
SAFQNGRTNEGRYDKALLGTPPSHLHPQLMAFNAAAQAFLGNPGSLHHPGLVGMHMQPQITQTSLP
AHLESQRREQSPNEPNARTGFWENCRAAYEDAVKHFRKKEEKIDCSAPQQATSHSSKGERNAGEASSREN
SPALNLSKTCCESSIGGEDSNSIEERSPVHENESETSQEGDHDTKTVDAVAQDHLVLPVSEKSHVGA
GNSNPQQAMAAAAAYNYSTLLSAAAAADPSLVLASTETLLCNIQGLL

>gi|33352130|emb|CAD82906.1| dachshund protein, partial [Glomeris
marginata]
RILRGLGAIQPGVNRCKLLSCKDFDVLYKDCTTASSRPPGRPPKRASLLGMSPGHQETLLKLLKSRLNGD
YTSYENGHIAAHLKGLHRAIDPMMLFPLLTGETRMEKSPLLANGYNHPPTHINPLPFMALNHHGHSATA
ILNPATGVPISSSHSMSSRSDGSIKERAHVADVLAARLKEEKNDCDRSIMALDHRVRKLNQLNGSSNGH
SPVLNLSQKSTCAVAGDQSGSDAALNDPCDDDDDNISDVDDDDDEKDQDLSDGPVSCGANGDRSAANATA
AAAAAAAYSTLLGETSGGISSIIETLLRNIQGLL

>gi|530776472|emb|CDG15335.1| Dachshund, partial [Thermobia domestica]
PVVCNVEQVRILRGLGAIQPGVNRCKLLSCKDFDVLYRDCTTASSRPPGRPPKRATVGLSIAASQMNHHHL
KRHRENGEYPGFENGHLGDPPEMEKSPLLANGYNHPPTHNLHMQFMQLNHPAAAHTAILSPAGIPLSSS
HGLARADGSIKKNQGIPIGIDAIARSGIWENCRAAYEDIVKHLERLREERGEAERALALDQKARDLSSHNG
SSNGHSPVLNL

>gi|46405859|gb|AAS93632.1| dachshund, partial [Oncopeltus fasciatus]
VRILRGLGAIQPGVNRCKLLSCKDFDILYRDCTTARCLMTNKSTERPPKRAPVGLSLAASHLSQQIKKHR
MDNGDYSSGYENGHISDTPRMEKSPLLANGYNHPPTHNLHMQFMQLNHPGPGHTAILNPQLQHHLIKPP
PPMDALSRSGIWENCRAAYEDIVKHLERLREERGETDRALALDQKPRDLSSHNGSSTGQSPVLNLSKSN
EASGSETGGTGPEDDEDEDEGEPSDQEEASDKEQELSDREEGGSGPTVSPHALNYSALGAAGPNSAPTS
DPTISSSTETLLRNIQGLL

>gi|91718848|gb|ABE57142.1| dachshund, partial [Onthophagus taurus]
RILRGLSAIQPGVNRCKLLSCKDFDTLHKDCTTASSRPPGRPPKRAPVGLSLAASHLQHQQKQRMNDG
YPYENGHMGDISRMEKSPLPANGHNHPPHLGHMQLPHAAAHSALLSPAMPHNLGRPDGSVIKNQG
MPSMEAIARSGIWENCRAAYEDIVKHLERLREERGDPRPLDQKVRDFSPRNGSPTEHSPVLNLSKS
GGGSAGSMGDADHSGSGADGPDAPPTPRSPRSAAADDEDDNISEPEDDDEKDQDMDEGDLPLPVVPGNDS
QQAAINYSTLASAVASANGQNHQDPSISSSTETLLRNIQGLL

>gi|196123803|gb|ACG70190.1| dachshund protein [Saccoglossus kowalevskii]
MSLAGIATSLSPPTQVLFKLDKPAYSTPPPVSNDPANNECKMVEYRGAKVASFSIDAREMICLPQAFDLF
LKHLVGGGLHTVYTKLRDLITPVVCNVEQVRILRGLGAIQPGVNRCKLVTRDEFDILYNDCTTASSRPPGR
PPKRTLPPPGLSPSGHFLLKKPKIENGCLTNGIDFSGFSPGLGRDGKTVGIPNGYLSPYLLTTHPSMPT
SLAMVTSHSSMHASHRHDEPISLSVNTTAKSDGTSKSDKSPSSSDLSPPRSGIDRPKEQQCSPPRKEAE
KEHERHTDVKSHKSDNKRNSTDQRDHHHHYHQHSTESEKISSIGNGLNYNQTAAALYNIPIENTSSIETL
LTNIQGLLQVAADNARQQEKQINLEKAELKMELLREKELREALEKQLVAEQKTRAIQKRLKKEKAKR

LQEQLEVEQKTRSQTDSSVIKPNTEALRVINDAKIASKYSYPSTEALSQELKLERNARTDAERKLLDIR
GSLQSFTEENFLAKNGHMDYSQYRMECKDPRNISGDEIDDDVIDNDIEDNDDDMADIEDED

>gi|390362165|ref|XP_791836.3| PREDICTED: dachshund homolog 1-like
[Strongylocentrotus purpuratus]
MKMGSPPTQVPLSSVLYKLDRAVAYSTPPPVSNDPANNECKLIEYRGAKIASFNINGDCMICLPQAFELF
LKHLVGGGLHTVYTKLRKLSITPVVCNVEQVRILRGLGAIQPGVNRCKLLTRKQFDILYEDCTTASRPGRP
PKRNIPLSLTPTIHGLSRMGPTLDLLKRPKLENGHHRLPFNGYEHAAAAPYLLATHPAFLPSSLSMPTN
HLTASPAKAAEMHHLHHDLAGAAHLHNLHKLKLSVGI PRSSSQPSSSHASVDIDDHEEDLENGDMS PRSSD
GPSPRMGGSQGGLSPKSDSSSAMHEDEDEKESDRIHQEAMAAMEHSLANAREENKPSGMNGALSEGASR
GEEYEQLHGNSTSQSGMPLPGMTATDILSHQVGPSSMETLLTNIQGLLKVASDNARQRDKQINLEKAELK
LELLRERELRESMEKQLVAEQRRFAILVRRLKREKRSRRRLQEQAESTTQQQQQQREREREQQKMKGSL
DQEPSRSPPIHTIQPIQAVAQESIRALNGGDRRSPCPDVYEKGEWLRGSSLVVS

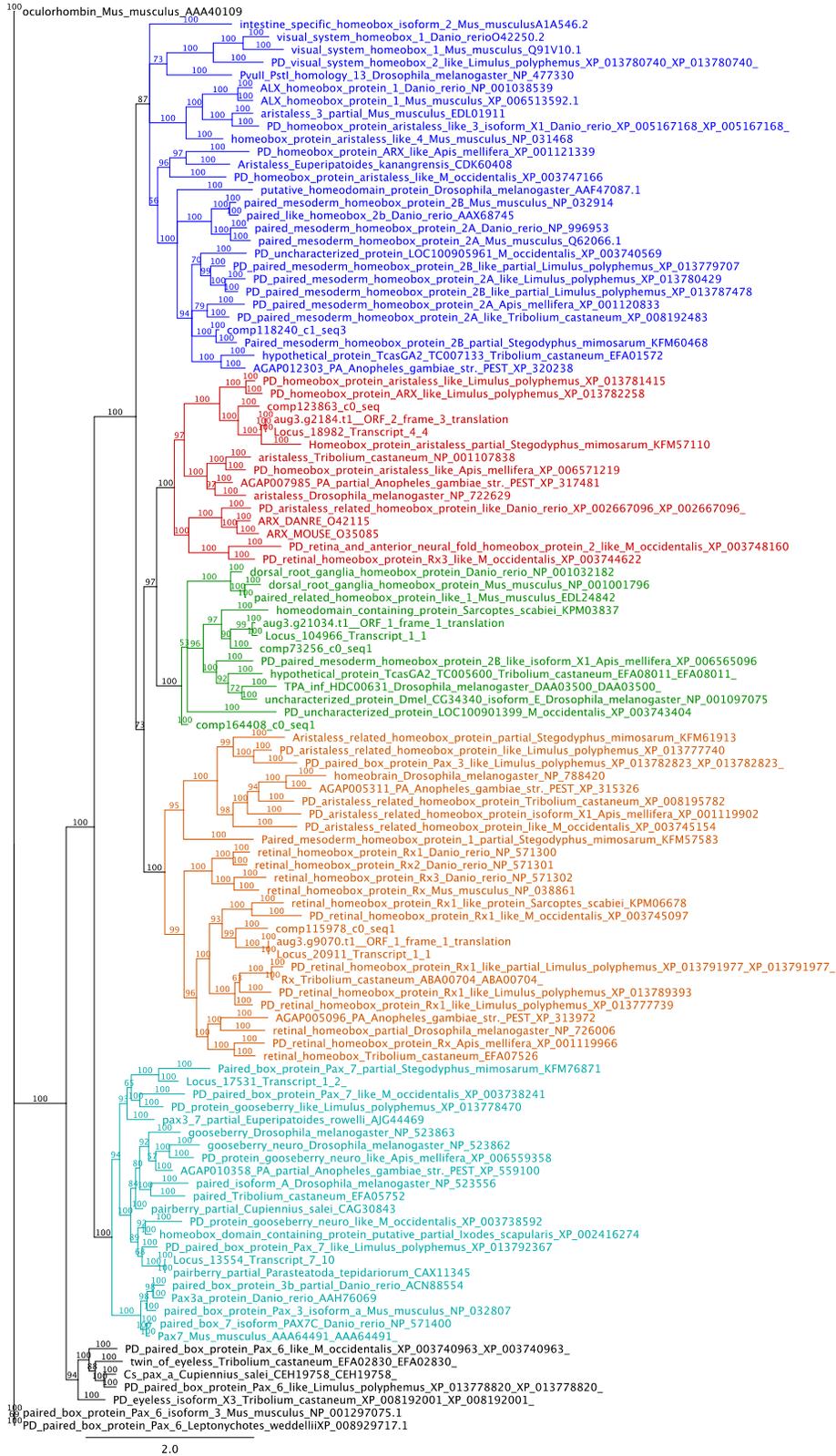
>gi|295656530|gb|ADG26728.1| dachshund protein, partial [Platynereis
dumerilii]
MMMEAVPPRPTSLPSPSRTPVLSPOPPPAPMHQGPPLPPPPSHSIVSMSSISTTTSSILPLKMEKPIYSS
PPPAASNENNTCKMIDYRGAKVAFAFKVDGRELICLPQAFELFLKHLVGGGLHTVYTKLRKLDITPIVCNV
EQVRILRGLGAIQPGVNRCKLISPGEFDVLYDDCTNSSARPPGRPPKRSPIAAHPETIEKLRKQRMGEY
YSPNRLGQFMFSDPKKSPFLFGNGYHYPPHILASMNVPFMPLGHPMNTMAMANHMGMRPDGSIIRERQAP
DLASPRPKDDGRYDENNMEHNNKPLDKPLNLQMDRPRRRPSLTDKDSGMSTPSDVMTNGQLDLKSMKHS
APSINEDLDDSEDDKEDDDMDSDAPPMSANSDAADKMAASSLPYQASQMLNTETTGISSVETLLMNI
QGLLKVASENARHRERQMNIEKAEKLMELMRERELRESLEKQMGDEQRTKVNLRRLK

>gi|224830721|gb|ACN66455.1| dachshund, partial [Neanthes arenaceodentata]
LKRLDITPIVCNVEQVRILRGLGAIQPGVNRCKLISPGEFDVLYDDCTNSSARPPGRPPKRSPIITANPETI
EKLRKQRMGEYYPSPNRLGQFMFSDPKKSPFLFGNGYHYPPHILASMNVPFMPLGHPMNTMAMANHMSMR
PDGSIIRERQAPEDLASPRPKDDGRYDENNMEHNNKPLDKPLNLQMERTRERRPSLTDKDSGMSTPSDV
MTNGQLDLKSMKHTAPSINEDLDDSEDDKEDDDMDSDAPPMSANSDQAEKMPQSSLYQASQMLNNEA
TAISSVETLLMNIQGLLKVASENARHRERQMNIEKAEKLMELMRERELRESLEKQMGEEQRTKINLRRL
KKEKKARRKLOETLEQMGAKTHPMEDLPSPKPLSTESHRSINDGMSSQDSERDRASCHSSPDRRLPGSAS
LYTYDAHSGLQNLMETFLGPKQKSCNGSTNGYLHPPVIKATESV

>MMA07092| Cao et al. 2013
MSAFYFVSRPGRPPKRASMI GMHVSTSGHTTLLKKTRESEYSAYENGHITGERIDKSHSLLANGYSHHSSTTAG
APPHISPIPFMALNHPSSHPNSMVPVSLPLSATSHSHSHQLTTSRTDASLNKDRSSHDVIVSSRVRDRVDNLDA
KERLYNLECHRVKDQYMNNGTSVNCHSPVLNLSQNSSRRGGGEHSGSETAYNDCQEDDYNDSEDDREDFSD
NPDVTSTGNSERLTNQALFYPHVQVDGMTATGQVVSSIETLLRNIQGLLKVAADNARQQDRQISVEKAEKMEVM
RERELRENLEKQLLEEQRIRILYQKRLKKEKRSRRRIQEQLESEMKKRVQYEEALRNSAETIRLLNDSLAQELE
RAERNVRNDTEHKVQGTVFKS

>MMA07100| Cao et al. 2013
MDDSRGHCTPSPTSGLASPHSLQTPPNLEPGVIVTESETLSPKHGQQTSTMLASPCSTLETTPRRSPVSPAGHGN
VLPQGSQHHPYPLGMAGMNGGIPKHCLDAYHPQTTVSLAARTPVSAAGLAYSNRPMHPASPPGVPADISVNECR
LIDYRGAKVAFLVNTTEYLLCLPQAFELFLKHLVGGGLHTVYTKLRKLDITPIVCNVEQVRILRGLGAIQPGVNR
KLLSCKDFDTLYKDCTTARSTH

>MMA52425| Cao et al. 2013
MDGSSGEGTISPSGDLSSPQSLQTSNPNDSSGLPGDSGQILAGGSPKPVTS MAGLPGTGSQVIGGNRSSPVGSPG
SAMPVSVAKSYAASIVSAGALSGCGPKHLEPLPLGLTAPRPGAPTQVNSYGLHSGRIHTASPPGLTTDATAND
CRLIDYRGAKVAFLVNGEYLLCLPQAFELFLKHLVGGGLHTVYTKLRKLDITPIVCNVEQVRILRGLGAIQPGVNR
RCKLLSCKDFDSLYKDCTTARHLWPGRYKSNSYWKQRPGLPGVASVESWPGLIFFELGLHTHTSRSEVTWERE
KRDRYPITTKLYELE



Supplemental Figure 8: Unrooted 50% majority rule consensus tree to identify true *al* homologs. The same phylogram as in Supplemental Figure 7 in a magnified view to display gene and species names.

Supplemental Material

>comp123863_c0_seq1

GTCCCCGCCAGCAGTACTAGTGTGTGGTTGGTTGTGTACCTCGAAGCATCACCGTTCTCTCCTTCGATATTGTTGTCTCTTTCAACAACCTGAAGG
GGAAAAGAGAGAGAACTGAATCATTGATGTGTGTGCGTTCGTGCATGTTGTGATACACAGACAAGTATCGTCTTCGGTTTCCAGTGC
AGTGATCCGGGACTTCCATCTGAAAATCTCTTTCTTTTTCGCGGACATTCATGTTGTTTTCAAGCTGTGGCCGGTGAATCAAGTAAA
GTTCCAAGACCAGTCTATCTTTCAGTGCCTTGTGCTTCAACAACGTCCATCTGCTGCCACCAATTCGAGGAGGACAGGACAAAGGAAGAGAAG
AGAAGCAGAGTCTCACCGTGCAGCCACCATCATCTGCCGGACTGAACGCCGTAACCTTCTCCTTGCCTCCACAATATCACGGTTACCACCA
GTTCCATCATTACTTCAACAGCTGTAATATGGGAATATTGGACAAAGTCGAGACCCACCCTGCTTAAGCGAATCCGCCGGAACACTGAACATGA
CCACCAAAACGGAATCAGTCAATCACCAAAAACAGGGACAGCTCACGGTTATGCGACTGATCGGTGACAAAATCTCCGCGGTGGCCAAACACCCCG
AGTCCAAAAGTCCATATCGCAGGAAGTGAAGGGCAATCCGTCATCGGCATACGGAACGAAAGGATAATCTTCGTACCTCCCGAGGACTTGACCAT
GACTTCTGTGGTGGAGCTGTCGCGGACAGTGACCACGCGGTGACGGAAGACGAGGACGAAACCGACGGTAAAATCTCGAAACGGCAACGAA
GTTCCGTCCAAGCAACTTACTTGAAGGTGAGGATTTCCCGAAGAGGAAGCAGAGAGGTTACAGAAGCAGTTTACAGAGCTTCCAGCTGGAGGA
ACTCGAAGAAAGCTTTGCGAGAACCATTATCCTGACGTGTTACTCGGGAAGAATAGCCATGAGAGTTGACCTCACAGAAGCTAGAGTACAGG
TTTGGTTCCAGAACCCGAGCAAAAATGGCGCAAAACAAGAGAAAGCGGCGGTGGACAGTCAAAAACAGGGATACAATCTTATTCGTATCA
ACAACCTGCGCGTCCATACCAGCGACAGTCCAGTATCGTCAAAGTCTTCAACCCCTTAGGATACCCTAGAAGCTACGACCTGAGTCTTTTCAAC
ACAGCTCATCACCATCATCTCAGTTCCTCCGCTTACTGCCACACCCGGGCTATCGCTGTTCCGCCCGGCTCCGAACTTCTCGCACCTGCTG
GTTACAGCCTCAGGACCTACCATACCCAGCTGACTGTTCTCGGCAGCTGCAGCGGCCACTTATCCTCACCATACAGTGCCTTCCCTCCGACAT
CGTCCAAAGTCTCTTGGTTAACTATCGAGTCAGAACCAGCTAACTCCCGTGCCTACTGACCTTACCACCGGGGAAGCTTACACCGGTTTGC
TGACGGTTCCTCCGTCCGAGCAACACCACCGCTGCTTCGGCTACGCTCTACAGCTCCGAACTTCCACGACTGGAACCTTACACACACTGATCCT
TACCAGGATCTGGTTCACCAATCCACCACCGTTCGCCCAACAACAACGTCTGCAACAGCAGCAGCGTCTGATCAGCAGCACCTCAGACAAGC
GTGAGCTGGACAGCAAGTTCAGCATCGCTGCTACGCCTCAAAGCGAGGGAACACGAGGTGAGGATGGAATGCTGAGGAAGTCCAACGG
CGAAGTCAAGTGAAGATTATATGTGGGGTCTTACCAGACAGCTGGGAAGGACCTGCTTAATTGAAATTTGTGAGATTCTCCAGTGAAAAAA
ACCGTCCAGGTAACGAGTCAAGAACAGGAAAAATATAAGACAAAACCTCTCTTATTTTGGCTTGAAAAACCTTTTCATACACACTATCCGTG
AAGATGCCTTACATGAGCATTGAAAAACGGTCAATTTTCATCATGTAGTGTCTGACTGGTCTATTTTCCGTCTAAATAACGTCCTATTTACAGCT
ATTCTTGATGTTCAACGTGAACATGAGACACTGAAGCGCTGAATCAGTGTCTATTTTTCCTCACTAAAAAGCCACATTCAATTAAGACTGTTT
AGAAAACAGTGAATATCCTATCATATGGACATCGATTTCCGTTTTGCACTTTGAATGATCATTTTACCAGGCGAGGATCCAGGGTTATGTGTCG
GGGAGGGGGTCACTCATGTGCATTAAGGTCAATAGTGTCCCGTATTTACCAGACAGGCCATCTCCCATATGATCCGGTCTTGAAGAAGA
GCATCGCCGACTCTGGATCAGCGCCTGTATCTTGATATAATAATACCATGAACCTTCTCATATCAATAACTGTTATGATTGTTGGTGTGAATATA
CCCATGGATAAGGCGCTAGAGAGCCAGGAGTATCCTCTATCATACTCTATGTTAGAGCGACTAACTAGGGTCTTAAAACTCAGGTTAGC
TTATCCGCCATCTTCGAGCGAAGTAGGGGACGTGATTTACTATAGCCAGAAATTTCCCAAAGTGAATAAAATCAATCCCTTGGGGTTGAT
GAAAAGGACCATGTGGACGATAGAATCGATGGGTTTAGTATAGGATCTATACGTTAGCTTTTGGCTGGATGAGACGATGAAGAAAAAAAT
GAAGAAGTGGGAGATGATGAAGCAACCAAAATTTGGGAACTATGCTACAGCTAGGAATGCATAAACTTTTCAGTCCCGCGCCAAATATGTGG
TGTTTTCTGCCTCATTTTTTGTTCTTTAAAAATGCGCGTTTTCTTATTTACAATGAGGTTTTACTTTAATTTTCCAAATAAAAAATAATGATTAT
TTTCATGACATAAATTA

>Locus_18982_Transcript_4/4_Confidence_0.600_Length_6071

CGTTAAGCCTCAATCCCTACACGTAACAGGGTTTCGTTCTGATAACGGTAAAGTAAATGCGGCGTGTATTATCTATTTGTGCTTTTATAAAT
AAAATGATTATCTTACCCTGATACAAAATGATAAAAATGATTGAAAGTAGCTTCTCCGCTTGGTAATTAACCTACCATATTTAATGTGCCTTT
CTCCAAGAAAAATACTATTTGCTTATTTGTACCCCGTCCAGTCTATTTGCGTTCCAGGTTGTCGGTTCCTCGAGGTCCTCAGACCCAGGGA
AGGTGGCATCGTCCGGATTTGACCGACCCACTCCAACACTTTAAAAATACTATGGGCATATCAGATAAAGATCGAACCTCAGGAGTGCATTA
CGGAAGCCGCGCCACTTAACTCATGGAAGCCACCAAAAGACTCAACAGCTCATCATCGAAAACCAATTCATCATACAAAATCTC
GTTGTGCACTTTGTAGGGGATAAATCCCTAGACCAAGTCAAGTCCAGAACAAATGAAACCATGCCAACAAAGATGCTGTGGCAGAGACTTGTG
AGTGACCGGTGGTAAATGATGAGTGGACAGTGGTGTGATAAATATCAGTGTGGCTGATGATGATGGTGAACAACAAGGATCAGCTTCTGCA
GAATCTTTGATGGAAGGAGATGATTTCCAAAGAGGAAGATACGTAGATACAGGACAACCTTTACCAGTAATCAACTGATGAATGGAAAGAG
CTTTTGCACGGACTCATTTCTGATGTTTTTACAAGAGAAGAACTTGTATGAGGGTGACTTGAAGCCAGAGTCAAGTATGAGTTTCA
AAATCCCGCAGCAAAAATGGCGAAGCAAGAAAAAGCAGCTGGTGGACAAATCACAAAGTCAAGGCTCAAACTCCCTATTACAGCTACAGCCCTA
CCGTCCTTACCAGCATCTGCACCAGGATCTGAAAAACCAATTCAGTGTCTGAGCTATCCTAGATCGTATGACCTTGGTCTTTTGAACGAGCAGC
AGTAGCTCACCATCACCATCATCCAAATTTCTCCACCTTATTTACCACCGCTGGCTTATCAGGACTGTTTCCGCCAGCTCCTAGTTCTTGAC
TCAAGCTGGCTATTTCCCTTCGAGATTTGCCCTTACTAGCTTATTTCCAGCAGCATTATCTTCACTTATTCCGTAACCTGTTTCTTCCCAAC
ATCATTCAAACACTTACTCGTAACTTATCAGCACAGAAGCCAAAGCTTAACTTCTCCTGAACCTGTTGCCACGGAACCATATC
TAGCCCTACTTGGCAGGACTGCAAGTGTAAATTCGTCTCCGCAACTGTTTTCAACAACTGTTTATGATGCTTATTAGAAAAGTAAATGAGTTG
GTTTCATCAACTGGTCTCCGATACACCACAGCTACAAGTCTGTGGATTTTGTGCGAAGGTTGAGCATAGCTGCTTAAGACTTAAAGCT
CGAGAACATGAAGTGCATGAGGATGTTGAGAAAAGCGAATGGGGAAGTGAAGTTGAGTTAAATGAGCGACAAAAAATATAGAATATTTA
TAATACTGTGGCTGAGAGAAAATGCAGCAGTGGATTTTCAACATGGCAACGTCCAAAGAATTTATCATGAACTTTAAAACTCGTTTCGTGAA
ATTACATTGAACGGACATTTCTATTGATTTTCAATTTGAATAACTGAGATTTGTTTTCATGACTTAAAACTTTTCCAAATTTCTGTCTTTTCATCATTT
AGTCAAGTTGACAGTAACAGTTCGGGTGGTGGGCAAAATGCCATCTTTGACAACAAAATATTTTGAACATATATGAACAGTTTTGCTTTGTTTGT
GCCTCAAATTTTATAAAACCAAGGTTGTGTAATTTGTGTTTATGTTATCATTATCTGATGACGCTCATGTGAGTTAGCTCAATAGAGTG
GAAAGATCCGACTTTACAAAATACCTTAAAAAGTCTTTCAATAAAGAAACTACATAAATGTCACTTATTATATCAAGTGCCTGGTGTGTTTGG
GAAGCTAATACTAAAAGTCGAGTTTCGAGGAAATAATGCTCTTACTTAAACAACCTACAGAAAATCAGCACTACGTTGAGCGGATTTGTT
AAATCCCTACTTTATGCTTTTACAGAGTTTTGCAATTTGAGTCAATATTTGATGCTTATTAGAAAAGTAAATGAGTTGAAATACCGCCGCTAAAGTTA
TGTTTTCTCATCAAAACATTTCAATATTTATGTTAACAGAATTTTAAATACGAGCTGAGATGATTTTGTCTAATTTTGTAAACAGAAAAT
AATGATATGTTTTGATTTGTGGTTCGCTGTTTTTATCAAGAAGCTAAAAAATAAGAAAATAGTTTTTATAATAATATTTTAAAAATCTAAAGA
AAGCTAGTTTAAAAAATACTTTGGATTTATCTATAATAATCTAAGCTGGTTTTAAAACTAAAGATCTTATTTATCAAAAAACAGTATTTTCT
TTTAAATCTCCAGTGTGCTTATGCTAATCAATCTTAAATCTCTACTAAGAAAATAAGAAAATTAATGAATTTATTTTATGATTCAAACTTAAAAATA
AATTCATTTAAATATTTAATGCTTTCTTTCAGCAGTCTATAAAGAAAGTACTCGTGCATTTTCTTCTCCCTCAATTTAAAAATTTTATCAAT
CTTTTAAAAACGGTTTTATGCAATTTAGTTTTAGAGAATTTAGGTTTCACTTCTGAAAATGGGAAAAACTTAAATGTAATGCAAACTTAA
GTTTTTGACTTGAATAATAAGCGAGCAATTTATCAGCTCTCTGAAATTTACGCCAGGTTTTATTTTACTTATCGATTAATGCGAAAAAAAATA
ATTAATAATAAATCTACCGGAAATAAATTTGCGAAAATTTGTTCCGATGATGATAATAATAATAAATCAAGGAAAAATTTTCAGTACCGAACATA
GATTTAAAAAATCTCGGAAATTTTTTTTGGGTGACTTAGTCTACTACGTTGGTGGATGAAGGTTAATCAGTTTAAAAAATTTGCTGGAATTA
TAGACAATTTAAAAATCTTAAATAGAAAATAAATAACATGTGACAATCTGAAAAAGTAAAGATATTTGGATTAAAAAACATATTTAGTCTGTTGTA

TGAATATAACTTTTTTCTGCCAGGTAAATAATCAAGACGGTGTGATAACGAATATGTGCTTTGTAACAATTCCTGCTGTGCCAGACTTTT
 ATTTTACGCGTGTATAAAATTCATAAATATCACAGCAAACTTCTCCCTTAAGTTTTCAATTCACCAGTAAAAAACAGCTCCGGGG
 GGGAGGGGCGAGTGTCTATTGAGTCACAAACAATGTATATCTAATTTCTCTAAAATATTTTCTACAAGTCAGACTTTACAGTGAGAGGGAA
 TAATTTTCGGAGTAAAAATAAAGTATTCTCTTTAAAAATCGTTAAAAATAAACACGCTCTCAAGCAACTGCAACGCGTCTATCCGGAGT
 CCTAAAATCGACCTGACTGTTCTAAAAACGCGTCACAAAATCAAAAAATAATCTCTTGTGCATAAATTTGACGTAATTAATTTTTAATTACCT
 TTATTCAAAATTTTACCAAGAAAAAACACGTAAGTCTTTGGTTGTTAATGAAGTAATCTTTGATTTTTAAGTTCAACTTAAAAGCAACTTTT
 CACAATCCCATAAAAAATTTAAACTTTTTACTAGTTTGCATTTTGAAGTATTTAGTTAGTATAAATAATCAACGACTTTTTTTAGAATG
 TTATATCACCTTGATTTTTTTTTTAATATTTTGGCTTACGTCCTAAAAATTAATTTGCTTCAACAAAAAGTTTTGTTATTGAGATTTAGCACTT
 TTATTGAAGTTTTTCGCAATATTTAAATTCATACCTAAAAATTTTTTAAAAATTTGGACTTCAGCCTTACTGATTTTATAAGGGAAGATCCAAT
 AAGTTCCCAACTGAGTAAATGATTATTTAAAATAAACTTTTGGCATTGAAGAAAAACAAAGTATCAATCCTGATTCTAATAAACTTTGAATTAT
 TTTTCAATTTTTAGTAAAATCATTGAATAATCTTTGTATCTAAAGACTAAAGGTATCACGTATAATTAATATAGAATTTAGAAAAGTAAAAATGAG
 AAAAAATCTCGAATTTAGCCATTACAATAAATATTGCTATAAAACAAAGTACTCATTGATGTATAGATGTAATGTTGTTACTATTTTACCA
 AAGTTGAATTTTATTATGATGATCATGATGATTCGACTTTCTCTGATTTTAAATAATAAATAAATGATATACCTTTTCACTCTTTAATTTACATTACT
 TTTTAAACATAAATGCAAAATGTAGTAATTTTCAGCAATATTTGTAATATATTATGAAATGATTCCTTTTCGAGACTATTTAACATTTTTTACG
 GTATAATCTGGAAAAATATAAGTCAAATAAAAAAATTGATTTCTGCTTAATAAATTAATTTTAAATTTGCACTCGAATCACATCTAACATCTA
 TCAATAACTTTAATTTTAAACAAACGTAATTTAACGCAAAATAATTAACAAGTCGGAAATATAAATTAACATTTGATGTCTCCTCGTAAAT
 ACATCTGTTTTAGAATTTTCAACGAATTTATTTTTCAATTTGAGAAATGGTTTTACTTTAAATTTAAAGTTATTATAACCCCTTTGAGAAGTT
 TGAAGTTCAAATCTTTGAGACATACAGGGTATTGATAATGATCACAATAAAAAATATAATTTAAAAAAAATTTGGCGAAGATAAAGTAAAAA
 TGATACATATTTGGAACGCATAAATATAATTTAAGCGAGGAAAAAGTAAACGCTGTAGAAATCGGTTGAATCACTACTGAGAAAGATGGAATA
 TTAAAAAATGAAGTTGCAATTTATTTAAACATAAAAAACATATTGAATTAACAGCTTTCAAATGCAATGTGATGTACAAATATTGTTGCAAAAT
 GATGTATGTTTAAAAAATACTATGGTATACATATAAATTTTATTTTTGAATTTATGTAACATGCATATAAATGTTCTGTTTTGAGAAAT
 ATGCACTGATACTTCAATGTTATTTTTTTGTATCAAAGTGTATTTGAAATGAAACGCTCAAAACAGTAAATTTAAATTTTAAATTTTAAATTTT
 TAAGCTTTAAAGTAAAGAAAAGTAAATTTTAAAGAAATGTATTCTTAGATCAGATATAATTTCTTATTTTATAAGTGTCAAAGTTTAT
 TAAAAATCCATAAATTTGCTGATTTTTAAAAATTTAGTGTAGGAAAATTTTGGTTTTCTTTAAATATCCTTATGTGCTCTGTAAATCAATGCAAAATA
 TATAAATGTTTATGATCCATAAAAAATTTGAAACATCCCAATAATTAACAATAAAGCCACTTTGCACTTGCAGTACTATTTTCCATCTATCG
 AATGGCTTATTTGATCGCGGTTGTGGATGGATCAAAGGATTCAGATTCATGTGGCTCTACGCTAAAATAAATTTATTTCTTTCTCAAAAAAC
 AAAACAAAAAATCAAAGAAATGAAAAACAATTCCTAGATATTGTGGAAAATTTTCAATTTGAAATATGCTCTGCCACTTTATCAACATTTAAT
 TTTTTCTTAAATTTTTTACAAAATTTGTACAGCTTGTGAGCATTGAGTTTAAAGAAATGGTCAACAACCTTTAACTATAAAAAATATTTCA
 CTTTGTAAATATAAATGATTTCCGGTGCATAAGTTCTCTGTACTTTGAAGGGAAGTTACCGTCCATTTAATCATTTATTTGTTGGAAATG
 TGATTTGCATATTTAATGGCATATTTAAGTTTACGTTTGCACCTTCAAATAAAGTGAAGTGTGTCAGTTCTTATGATTTGTAATTTAGGT
 GACTTTCTGTTAATAAAAAATGGGAAAAATAAAAA

>aug3.g2184.t1

ATACAAATGATAAATTTGATTGAAAGTAGCTTCTCGGTCTTGGTAATTAACGGCATATTTTAAATGTGCTTTCTCCAAGAAAAAATACTATT
 TGCTTATTTGTACCCCGTCCAGTCTATTGCGTTCACGTTTGTCCGTTCTCGGAGTCCAATCCAGACCACGGAAGGTGGCATCGTCCGGGATT
 GACCGACCACCTCCACACTTTTAAAAATACTATGGGCATATCAGATAAGATCGAACGTCAGGAGTGCATTAGCGAAGCCCGCCACCCCTTA
 ACCTCATGGAAGCCACCAAAAGGATCAACAGCCTCATCATCGAAAACCAATTCATCATCACAAAATCTGTGTGCACCTTGTAGGGGAT
 AAATCCCCTAGACCAGCTTCAAGTCCAGAACAATGAAACCATGCCAACAAGATGCTGTGGCAGAGGACTTGTGAGTACCCGTTGGTAAATGATG
 AGGTGGACAGTGGTGAATAATTCAGTGTGGCTGATGATGATGGTGAAGAACAACAAAGGATCAGCTTCTGAGAAATCTTTGATGGAAGGAGA
 TGATTTTCCAAAGGAGGAAATACGTAGATACAGGACAATTTACCAGTAATCAACTGATGAATTTGAAAGAGCTTTTGCAGGACTCATTTAT
 CCTGATGTTTTTACAAGAGAAGAACTTGTCTATGAGGGTTGACTTGAAGCCAGAGTTCAAGTATGGTTTCAAATCGCCGAGCAAAATGGC
 GGAAGCAAGAAAAGCAGCTGGTGGACAATCACAAGTCAAGGCTACAATCCCTATTCAGCATCTACAGCGCTACCGTCTTACCAGCATCTGCA
 CCAGGATCTGGAAAACCATTCAGTGTCTGTGAGCTATCTAGATCGTATGACCTTGGTCTTTGAACGCGAGCAGCAGTAGCTCACCATCACCATCA
 TCCAAATTTCTCCACCTTATTTACCACCCTGGCTTATCAGGACTGTTTCCGCCAGCTCCTAGTTTCTTACTCAAGCTGGCTATTTCCCTTCG
 AGATTTGCCTTACACTAGCTTATTTCCAGCAGCAATTTCTCACTTATTCGCTAACTGGTTCATTTCTCCAACATCATTTCAAACATTTACTCGC
 TAACCTATCAGCACAGAACAGGCCAAAGCTTACAACCTTCTCTGAACTAAGTCTGTGTCACGGAACCATATCCTAGCCTACTTGCAGCAGCA
 CTGCAAGTGTAAATCTGTCTCTGCAACTGTTTCAACAACCTCCCTACTGTAACCTCCACCGCAGCATTGCCAGTTCATCAACTGTTCTCCG
 TATCACCACCAGTCAAGTCTGTGGATTTGATCGAAGAAGTTCGAGCATAGCTGTCTAAGACTTAAAGCTCGAGAACATGAAGTGCATG
 GAGTTGTTGAGAAAAGCAATGGGGAAGTGCCAAAGTGGTTAGCGCGCTGACTCGGAAGCAATGGCTGCGGGTTCGAATCCCGCTCTGGATA
 TGGATGTTCTCTCTCTCTGTGCTGATTTCTGACGTGTGAATGTATGAATGTGGCCACCCTATGAACGGGATTTTGTGGCAGTGGGAGTGGGT
 AATGTTGCTCGCTCCGTGACATTTGATCACAGGTGCCACTGGGTAACGATAAATGAGCAACTTCCGGCATCTTCTAAGGCGAAGAACTAAA
 GTTCAAGTTCAGTCCAGCGGTTATTAATAAAAAAAGAAATACTCAAATTTTATGATTAGTTCACTATAAAGCTGTGATGTAAT

>aug3.g2184.t1_-_ORF_2_(frame_3)_translation

MRSTLSVLGGPIQTTEGGIVPDLTDHHSNTFKNTMGISDKIERQECISEAARTLNLMEAHQKQDQPHHRKPIHHHKLVSVHLVGDKSPRPASSPE
 QMKPCQDDAVAEEDLSVTGGNDEVDSGDENISVADDDGENNKGASAEESLMEGDDFPKRKRIRRYRTTFTSNQLDELERAFARTHYPDVFTREELA
 MRVDLTEARVQVWFQNRRAKWRKQEKAAAGGQSQSYNYPSASTALPSLPASAPGSGKPFSAISYPRSYDLGLLNAAAVAHHHHHPQFPPPYLP
 PPGLSGLFRPAPFLTQAGYSLRDLPYTSLFPAALSSPYSVTGSPPTSFQTLLANLSAQNRPKLTPSPVATPEYPSLLATSTASANSAPATVVSQ
 PSPVTTPPAALPRSTGSPVPPATSPVDFDRSSSIAALRLKAREHEVRMELLRKANGEVAKWLARLTAKPMAAGSNPALDMDVLSLCCGILTCE
 CMNVAHMPMNGYLWQVGVGNVARLRDIGSQVPTG

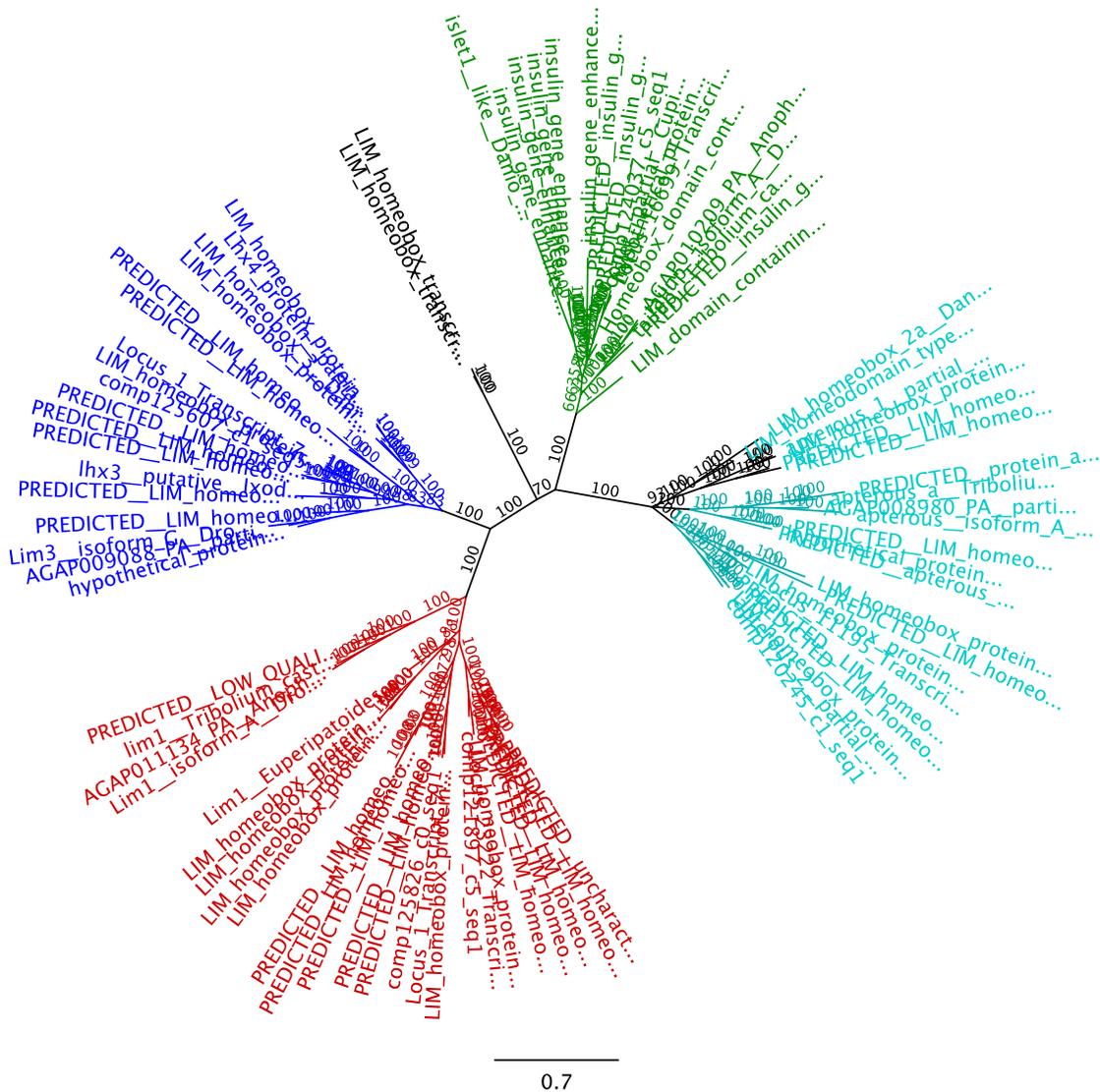
>comp123863_c0_seq1_-_ORF_1_(frame_3)_translation

MKISFFFADIHVFKLWVPESSKVPRAHLSSALSFNHLLPPIRGQDKGREGEAESPPSSAGLNAVNFLLPPQYHYHQFHYYFNSC
 NMGILDKVETHHCLSESARTLNMTTKHESVNHQNKGLTVMRLIGDKSPPLANTPSPKISQEVKGNPSSAYGTDKNPSSPPEDLTMSTVVDVVA
 DSDHAVTEDEDETDGKIPNGNEVPSNDNLLEGEDFPKRKRIRRYRTTFTSFQLELEKAFARTHYPDVFTREELAMRVDLTEARVQVWFQNRRA
 KWRKQEKAAAGGQSQNYPSSTTLPSPATSPVSSKSNFLGYPRTYDLSLNFNTAHHHHHPQFPPPYLPPGLSGLFRPAPNPLAPAGYSLRDLPY
 PAGLFSAAAAATLSSPYSAFPPTSFQSLLVNLSQNRPKLPVPTDLTTEAYTGLLTAASSVTTPTAASATSTTSQPSPTGTVPVGLSPGSGSPNPPPL
 PPTTTSATAAASSAAPQTSVELDRSSSIAALRLKAREHEVRMELRKNSEV*

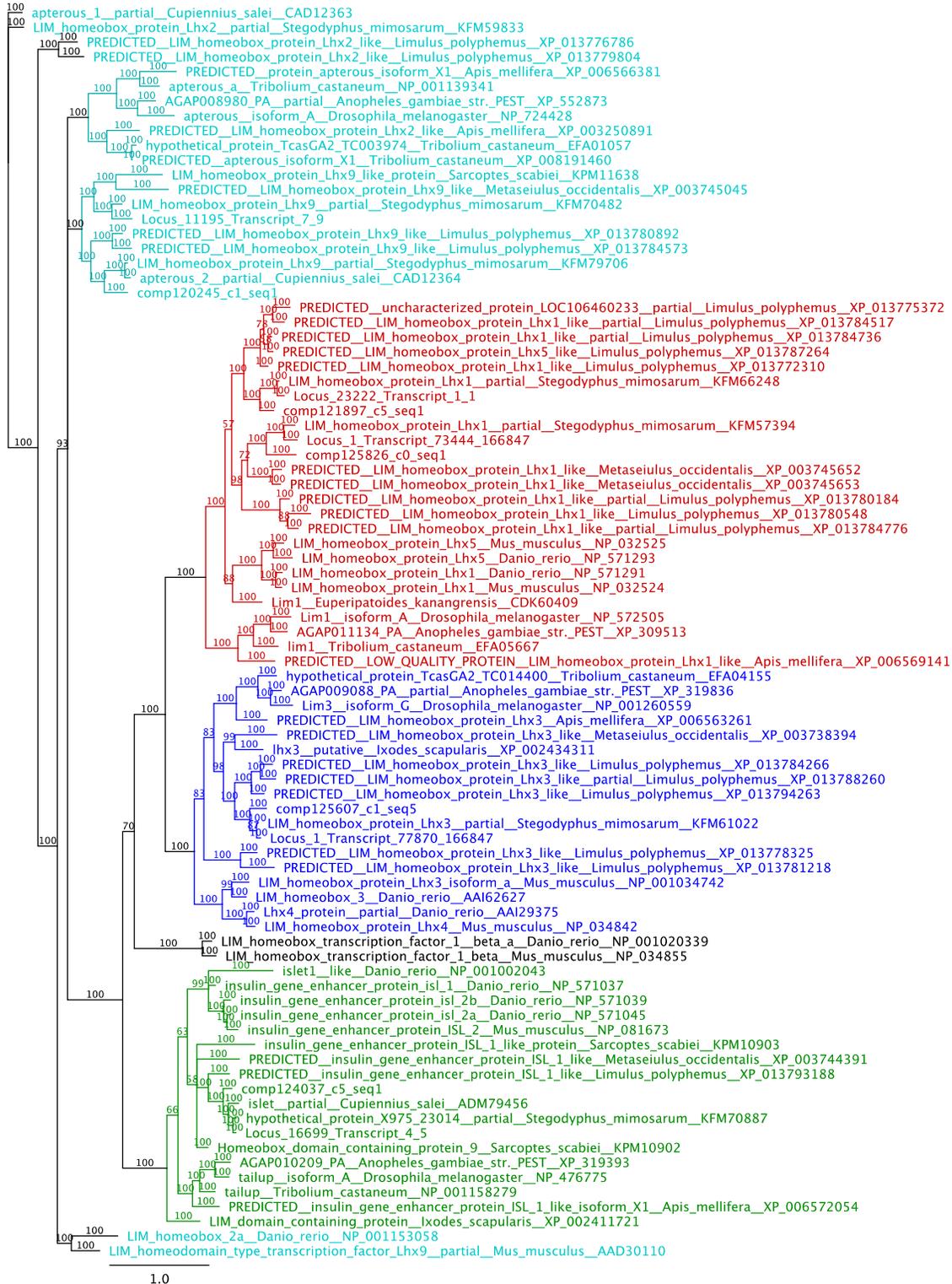
>Locus_18982_Transcript_4/4_Confidence_0.600_Length_6071_-_ORF_1_(frame_1)_translation
 MRSTLSVLGGPIQTTEGGIVPDLTDHHSNTFKNTMGISDKIERQECISEAARTLNMEAHQKDQQPHHRKPIHHHHKLSVVHLVGDKSPRPASSPE
 QMKPCQQDAVAEDLSVTGGNDEVDSDGENISVADDDGENNKGSAESAELMEGDDFPKRKIRRYRTTFTSNQLDELERAFARTHYPDVFTREELA
 MRVDLTARVQVWFQNRRAKWRKQKAAGGQSQSQGYNPYSASTALPSLPASAPGSGKPFALSYPYSDLGLLNAAVAHHHHHPQFPPYLP
 PPGLSGLFRPAPSLTQAGYSRLDLPYTLSPAALSSPYSVTGSFPPTSFQTLNLANLSAQNRPKLTTSPELSPVATEPYPSSLATSTASANSAPATVSQ
 PSPTVTPPAALPRSTGSPVSPATSPVDFDRRSSIAALRLKAREHEVRMELLRKANGEVS*

11.5.2 Lim

Lim1.2Locus_2322_Transcript_1/1 help by dawid Salamanca



Supplemental Figure 9: Unrooted 50% majority rule consensus tree to identify true *Lim* homologs. The Jones amino acid substitution model was chosen after mixed model testing. After reaching 1078000 generations and an average standard deviation of split frequencies below 0.01 a total number of 4314 trees were written in two files (each file contained 2157 trees of which 1618 were sampled).



Supplemental Figure 10: Unrooted 50% majority rule consensus tree to identify true *Lim* homologs. The same phylogram as in Supplemental Figure 9 in a magnified view to display gene and species names.

Supplemental Material

>comp121897_c5_seq1_len=2584

GTATATAATCACCTTGTTTTTTTTTTTTGTGTTTTTTTTTTTGTGTTTACCTGTGAGTTTTGGTGTCTGTGATGGTGGACATCAGTTGACAGTTTGAC
GTGTCACTTCCCCCATGACGTGTCTGTTTACTATCCGACGAGACGAACTGTCCATACCGTGCCGTTGAATCCGAAGATAAATTAGCCTGAATTC
AAGAAGGACAGGAGGACGAGGAAGAAGAGAGGACGAGGAGCGTAAGAGGAAGCGGAGGAGTCCAGCCGTGCTGTTCCCAACTGTGCCGTACG
GGGACGGCCGGTGTCCGGCCGTCGATACCCGCGCCGGTGCCGCTGAGGAGTTCGCCCATCAAGGTGGAAGCGTTCCGGGGTAAGAACTCGGGTA
CCCGGGGGCACTTACGGTGGCCTCTATCCCGCACTCCACCTGCCACAGCTTCGCTTCGCGACGGTGCCGGAACGGGAGTCAACCCGGAGCCGGAC
GCGGCTGCCGTCGACATGAAGATGGTGCAGTGCGCCGGCTGCGAACGGCCGATACTGGACCGTTTCTGCTGAACGTGCTGGACAGGTGCTGGCA
CGCGCGTTGCGTCCAGTGTGCGAGTGAAGTGCAATCTCACGGAGAAGTGCTTCACGCGGAGGGCAAGCTCTACTGCCGGACGGACTTCTTCA
AGCGTTACGGCACGAAGTGTCTCGGGGTGCTCGCAAGGGATATCTCCGACAGACCTCGTGAGGAGAGCCCGAAGCAAGGTGTTCACCTGAAGTGC
TTTACATGCCTGGTATGTGAAAAGCAGCTGTCCACGGCGAGGAACCTACGTCTTGACGAGAACCGATTCTGCTGTAAAAGAGGATTACCTCAA
CAGCCGGCATAGCCAAGGTGTACTAGTAGCAAACTCGCCTTCGGACACGAACGATCTGGACGACCTCGGCGACCCGTTAGGCGGCGACCCCGACA
GCGGCCACATGAGCGAGAGACTCTAGTGTGGAACGCGAAGGCACCCGGAATGAACGGGGCCCGCCGGTCCGTAGCCCGTCCGGGCCCCACCG
GGAGCTCACACGTTACCGCACGGTGCACACACCACGACCCGACCCGGCAACGGTGGCAATAACAGTGGACCGAAAACAGCGGGCCCGGACAAC
CGGTGGCCCCGGGTGGACCTCCGAACGACGAGAACACGCGCGGACGAAGAGGAGGGTCCCGAACCACCATCAAAGCCAAACAGCTGGA
CCCTCAAGGCAGCCTTCGCTGCGACACCTAAACCGACGAGGCACATTCGCGAACAGCTTCGCGCAGGAGACTGGACTCAACATGCGTGTATACAG
GTATGGTTTCAGAACCCGGCCTCTAAAGAAAGCGAATGAAACAGCTGAGTACGTTAGGAGCCAGGAGGCATTTCTCAGGAGTCCACGAAGAG
CCATGAGGCCGCTGAGGCCCGAATGTCTCCGGATGGTTTACGACGACGCCAGACATGGTAGCCGGGCCAAATTTCTGGATACGCGTATTTCTCG
GATTCTGCGAGCCCCGGTGTATTTGGTTACGGTGGTACGCTGGTTTTTACGACTTTTTCCCGGACAACCTCCACCGGATGGTATGGGTTACCC
CGGGATCTCGGCCCTGGACAGAATGTTCTCCGCCACCGACTAGCATGGAGCAACCTCTACCTTCGGCCATGGGACCTGGTCCGGTGAACCTTGG
AGCGGACGCGCGTACCTGACGCTCACGGTACCTCTGTGCGAGAGATCGAGTCCCGAGAGCGTCTCAGCTGCGCTCCGTCAACGACGCGTA
CGGCCCGCAGCGTCCGGGGTCCCGACGGTTCGCCGTGCCCTCGGTGGGCATCTGCCATAACCGAGACTCCAGTCTGGTGACAACACGCGTC
CTCGGTGACTCCCGGGCGGCAATTAACCTCCTGAAGCGCAACATTCCTCGGCTTGACGCTAATGTCTCAAAGAGAAGGACAACCTGGACGACA
ACTGAAAAAAGAAGAAGAAAAAACAACACGACGACATGAATCATCATCGTCATCATCAAACATCATCATGAACATTCATGTGACGA
ACGAAATGTAATTCGAAAGGAGCGGAAAAATCGTGCATTGTGTACAGTACAGGACGAATGCATCGAGCATTAATCGTGCCATTTGAAAAC
AAAAAATTATGTATGTCTCGGTAATGTGATGAGAATCCGGTTATTATGGATAATCACCGACCATTTTAAAGTTTGATAAATTTCCCTGAAT
TTTTATGACTTTTACCGTAATTATGCATGATATCTAAGCGTGGTGGTGAACATGGATAAGTCTTACGAAATTCGTACGTTTGTGCGTTATTA
CGCATAATCACCATTTCTGATTTGGTACAAATGGTAAATCGGTTCCGTTGCCATGATATTATTTTTATTGAGTCGATACCGGTTATAATGAACGC
TGATTTAACGAATTTTACACACATCCCTGAACCTAATCTTACGAATCTGAAATCGGACCTCGTTATAAGAAATCATATTGAACAAAT
TTTACGATTGCACCCCGATA

>comp125826_c0_seq1_len=2345

CCGCTGCTGAACGAGCGGAGCTGAAAGGCTCCGTCGCGGAGTTCTCTCTTCCCCCTCCCGTTTTGCGAGATATATTTCCAGTTTTTGTCCGT
TTTTTTTTTAAATTTCCGAAAACAGAAATTTCTCAAACGCGTGTATTTTTTTGAAGAAAAATAATAATTGTGTTTTTTGTACATAATATAT
ATATATTTGTTGGCGTTTGTGATGAAGACCTGGTGAATGCGCCGGTGCGAACGGCCTATACTGGATCGCTTCTTGCTGAGCGTCGTGACAG
AGCGTGGCAGTGAAATGCGTCAAATGTTGGGATTGTAAAAGCACTTTGGCTGACAAGTGTCTCTAGGGACGGTAAACTCTTTGACAGAAA
GACTTCTCAGGCGTTTCGGAACGAAATGCGGTGGCTGTCCGAAGGATATCACCGACGGACCTCGTTCGAAGAGCCGAGGTCAAGTGTTC
CTTGGCTTGCTTCACTTGCTGTGTCGCGAAGCAGTTGTCCACGGCGAAGAGCTTTATGTGTTAGACGAGAACCGTTTCATTTGCAAGGAGG
ACTACATCACAGCACCAAGTTCAGCATCACACCATCATCATCACCAAGCCTGTTCTTCGAGTTCGCCTGTGGATTCCGCGGAGTTCGATG
AATCACTCGAAATGCCAGTACTGACCCCGTTCTCCCGCAGAGCGGATAACGAGAGGGATTCCAGTATCGAACCGATCGGGATCTACACAAC
GATCCCGTTTCGGGTGGTTCGGCAGTGAACAGTCTACTGGGAACGGCAATGGTCAACAGTCCAACATAGTGGACGGCGTGCACCTGTCGGCCAC
GGATTCGTGTCTGAGGGCTCATCTGGACAAGCGCCAGTACGACGCGGCTCCCGCGGTACGAAACGACGGGGCCCGGACAACCATCAAGG
CCAAGCAACTAGAGACCTGAAAGCAGCCTTCGCCGCAACGCTAAGCCTACGAGGCATATCAGAGAACAACCTCGCTCAGGAACTGGACTCAAC
ATGCGAGTTATACAGGTTTGGTTCAGAAATAGAAGATCTAAGAGAGGGCGGATGAAACAGTTGAGTACTCTGGGTTCCAGGCGGCATTTCTTCA
GGAATCCCGGAGAGCTATGAGGCCTTACGAGCGGTCTCGGAGCGGATGGTCTCATGGAAGATGTACAGGGACATCATAACGGCCGTATCCA
TATTTTTCAGATTCAGCAGCCCTAATGATTTCCGGATACGGGCCACATCCAGGTTTTTATGAGTACTTCGCTAATCATTTCTCAAGAGATGACCGT
GTTTCCGGGTACCATGCCAATCGGACCAGCTAGCGTCGACGGACATCCGCTGGATTACAAAAACCACCGACCGTCCGTGGCATGCCATCCATGGC

GAATGGAGGCGTCAACCTGGCTTCAGACGCACCATACATGGCACCGAACGGAGATTACGTCCGTCACCACCAATCCAGCCGAAGAATGGTATGG
CCATCCAAATCCAGGCGTCCATGGACGGGTACGCTTCCAGGAGCGCTCAAGAAGCTTATCCGAACATGCCTCAGCAGCACAAACGCCACAGCTC
AGCTCCGACACTCCCCTGTGGTGACCAGAAACTGCTGGTGTAAATGGAAGTGGTGTGAACTAATAGTGAAGTGTCTCCGTGCCTCAAGAG
TTATGTACAGTAATAACTCGCATAATAATGCTGACACTTTTTCTTTTTCTCTTGGTTTTCTTCCACCTCATTCCGACAGCCGATTTCGTGCTG
CGGCGTTCACGAAAACATAATAATTTGCATAATTTGACGACAGGAAAACGGGTCTGAGACGCCCCCTGATGGATTTTTTTCTCAAGTCAATAG
ATATTGAAACTGTGATTCATTAATAGAAAAATTTATGTGTGAAATATAAAAAAATGGTCTACACTTTGCGTGGCTTCCACGAAGAGGATGCTG
ACAGGATCCAGGATCCTGACAGGCCAGGAGCCCGGCAGCCACATTTTTTAGTGCATAAATCACAAGGTCTCATAGAAGTCTTGCCAAAAAGC
CCACCGAAAATCCGGCTTAGGAGGTTTCGATGGAGGGATTTGGGAATCCTTGATACTCTATAGGAAAGTGGTTCCCAAAGTGTAGGTCGAGTCA
ATGAACTTTGGGGGGGGGAAAGATTCTCCAAAATTAACAACGTCCAAATAAGAAATTCATTTCTACTTTAAAAACGTATCAAACCTAAAAAATG
CCTTTACTCAGCTGAGCTAATTACAATGTATCATAAATGTAATTTATCAAATTTCTACTTGTACAAT

>Locus_1_Transcript_73444/166847_Confidence_1.000_Length_855

AGGAGGTTTGGTACAAAATGTGGTGGCTGTCCCAAGGCATATGTCCAACAGATCTGGTACGGAGAGCGAGGGGTGCGGTGTACCATGTGCCT
GTTTCACGTGCCCTTTGCCGAAGCAGCCTCTACCGGAGAAGAACTTTATGTACTTGACGAAACAGTTTCGTGTGCAAAGATGACTTCAAT
AGGCAACAGCAACATGTGTCAAGTTCACCGCTCGATTACCAGAAGAGTCCCTAGGTAGTCTCAGGGTATTTCGAGCATCGTTGTGACGGTGG
CTCAGGAGCTAATAGTCTCACAACAACACTCCCAACAGTGAAGAACCCGTCGATTGTTTGGTTGCTGGACCTTCTCGAATGACGAAGCAGGAG
GAGGGACTAAAAGAAGAGGTCTCTGACCACATCAAGGCTAACAGCTGGAGACATTGAAGGCTGCTTTCGCCGCAACGCCAAACCGACAAGG
CACATAAGAGAACAATTAGCACAGGAAACGGGACTTAATATGCGTGTACATACAGGTGTGGTTTAAAAACAGGCGTTGAAAGAGCGAAGAATGA
AGCAACTGAGCAACTTGGGAGCAAGAAGACATTTTTTTAGGAATCCAAGCGAGCGATGCGTCTCTTAGAGCCGGTCTCATGGACGAAGCTGTC
CCACACAACGCCCTTTGCCTATTTCTCAGATTGCAACAGTCCAAATGACTTCAACTATGGACCCCATCCAGCTTTCTACGACTATTTCCAGTCTC
CATCCTCAAGGTGTATAACAGATTTTCCGGAACTGTCCCACTAAGTGCAGGAGGAAATAACTCTCGAACATTGCAACACAATCAATAATGG
T

>Locus_23222_Transcript_1/1_Confidence_1.000_Length_2928

CATCGACTTCTTAACAAAATTTGTGAGTCAACGTTATTTTATATTAATTTATTTATAATATATATTAATAAATGATTTGTTAGCAACAAAATA
CTAATTATGAAACTAAAAAAAATTCATATTTTTTATATTTACTTAAAATGGTGAATTTATTTATTATACAAAATAAAGAAGTTTTCTCT
TTGCTTATACAAAAAATCTTTCTAGTAATTTCTAGAAATTTATTTCAATTTCCCATGCAAGAGAGCAAACATATACTTACATTTAAATTTT
TTACCAACATTAATAATTTTCTTTTTTAAATTAACCTGTTTAAAAAATAAATAAATTAATAAAGTGAATATAACTTCCCTTTCTTAATG
TACTCTCTGTCTTTTACATTTTACATTTTGTTCCTAATGATAGCAACATAAACAATAAAGAATCCCTTTTCTATATTTACAATAATGAT
TTCATTTTCGTATTTAGAAATTTGGCATCCATGTTATTTCACTTCAGTCTTTTCTCTCTCTATGCCGACCACAAGAAAGCTTAGGTGCAGAC
GATAAGTCTCTGCAAATATGAAATTTTCTGTACAGTTTCAGAAATGAAATACGAAGAATTCAGTTCACAAATCATCAGACGATTACAG
TTTCAATAGTCCAAATTTTATTCGGTGGATACTGACACTGGTGGTGCCAAAATGGCGTGACGTCACCAAACCTGGAGTCTCTGACATAGTTGGATG
ACCGGATAAGTTAGCCGGAATCCGGCAGATGATGATCGTGGTCCGTAAGCGTCTGATGCTAGGCTGAGGACACTGTCCGGGCTAGATCGCT
GTGACATGAGATCACCATGCGTCGCCAGATATGGTGCATCTGCCCTAAGTTGACAGGTCCTGGGCCAACACGGCGGAGAGAGGTTGCTCCATA
CTAGTTGGAGGAGGAGGAACATTCTGACCGGCTCTAGTACACCGTGGTATGCCATGCCATCTGGTGGTGGTTGTCCAGGAAAGAAGTCTGAGA
AGCCTGGCTGCCACCATAGCCAAAGTCTCCCGACTCCCAGAATCTGAGAAGTAAGAATAGCCAGAATTAGGTCCGGACACCATGTCGGGACTG
TCATCAAGGCCATCAGGAGACATTTCCCGCCTAAGTGGTGCATCGCTCTTCTGGGCTGCGAAAGAAATGCCTTCTGCACCCATAGAATTAG
CTGTTTCATTCGCGCTCTTTTCGATCTTCTATTCTGAAACCACACCTGTATAACCTCATGTTGAGGCCGTTTCTTGCGCCAACTGTTCTCGTAT
ATGTCTCGTTGGCTTAGCGGTTGACGGAATGTGCTCTTAGGTTTCCAGTTGTTTGGCTTTGATAGTGGTGGGAGTCCCTTCTCTTCTGAC
CTGGCGTATTTTCGTCATTTCTGTGCTGCAGAGTTGGACGAATTTGGTGGTTTCACTGCTGCTGTTGTTACCGTTAGGCCACAGGAAGCAGGATTA
GGCCCACTGGCTGGGCCACCAGCACTACCCGGCGTGAATGCAGATGCATTTGAGGCCCTGTAGGACTGCTAACAGTAGAGGTTCCACCATTAAG
GCCATTATGACCTCCGTCTCTCAACGCTGGAGTCTCTCTCACTTAAATGTCATCTGTTGTCAGACCCAAACCGTCTCAAGATCATCGAGGTC
GTTACCGTCTGAAGGGAACCTTGACGAAGGACACCTTACTATTTCTGCTGCTTAGGTAGTCACTGTAAAGATGAACCTATTTTCGTCAGTA
CATAAAGTCTTCTCAGTTGATAACTGCTTTCTACAGACGAGGCATGTAAGCATTTTCAGGTGAAATACTTTATTTCTCGCTCTTCTAACTAGA
TCTGTTGGAGATATGCCCTGCGCACACCCGAACATTTTGTGCCGAAACGCTTGA AAAAATCATTCCGGCAGTAGAGTTTACCTTCCCTGGAATA
ACACTTTTCTGGAGTGTACACTTACATTCGCAGCATAACAACGCACTTTGCATGCCATGATCTGTCTAGTACACTGCAAATGTATCTGTCTTGA
TAGGTCTTTCGCAACCGGCACATTGTACCATTTTCATGTCTGTGCGAATTAAGGGCAGGGTCCGGCCAGCCCCGCGCCACGCCAGGTCCCAA

Supplemental Material

GTGTGTCCACCTTGGACGCCCTCAGTCCCCTGGCTCATAGGACAGACCAGGGCTGATAGCTCGCTGACCAACGAGGCTTTTGAATCGACCTCCT
CCTGGTTCCTCAAGAGGCAAGAGGACTGGGTCGTCAGATAAGTCGAGCAGGTCGCCGTTTTGCTTGGTCAATAACTTATCAGTACCTATATAGAA
TTCTCGAACCAAAAACCTGACGGGAAAAAATTAGAATCCAAAACAGCTCAAATGAACTACTCTGGTTTATGTTGCTTTATGAAACAGAAATTTTG
ATCAGTCATTCATTACAAAACCTCTGTCTTGAACAGTTCTTACCATTAAACTGCGATTATTTAATTAGCTGATTTACTGCAAAACTTCGAAAGC
CGAATAACATTGTATGCTTCTCAAATACCGGTTTGGTAATATTATTGATGAGTATCATTCAATCTTATAAAAACTATAATAAAAAATTCATAA
AAAAAAGCGAACTAATCTGTTAACCGAAGAGCGGCTGAATCGAAAATGAAGAAAATAGACGATCTGAATCAAGAAGGTAACAACAGTTTCTC
GAACCGTCACCTGACCAGTTTACGGCATGAAATGGGAACGCTGGGCACTATGCGTGCGTCCGGTATAACAGTTTGCACCAGACTAGCTTTTCT
ACTGCGACTGCAGCTTGTTCGAGGAGCAGCGTGTTCACCGCCGTTCCCTTACCACACGCTTCTTGGCTTATCTCTGGCT

>comp121897_c5_seq1_len=2584_ORF_1_(frame_2)_translation

MNPKIISLNFKKDRRTRKKRGRGGVRSGGGPAVLFPYVYGDGRCPAVDTRAAVPLRSSPIKVEAFGVRTRVPGGTYGGLYPALHLPQLRFATVP
ERESNPEPDAAAADMKMQCAGCERPIILDRFLLNVLDERSWHARCVCCECKCNLTKCFTRREGKLYCRTDFFKRYGKCSGCSQGISPTDLVRRRA
RSKVFHLKCFCLVCRKQLSTGEELYVLNENRFVCKEDYLNRSRHSQGVLVANSFSDTNDLDDLGDPLGGDPDSGHMSERDSSVEREGTMNGGPP
VGSPPGPPGAHTVPHGANTTTGPGNGGNSGTENSGPGTTGGGGGPPNDENTPGTKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQE
TGLNMRVIQVWFQNRRSKERRMKQLSTLGARRHFFRPRRAMRPLRPGMSPDGLDDSPDMVAGPNSGYAFSDSGSPGDFGYGGQPGFYDFPFG
QPPPDGMGYGILGPGQNVPPPPTSMQPLPSAMGPGVNLGADAPYLTPHGDLLSQRSSPESVLSLPSVNDAYGPQRSRGGPDGFPALGGHPAIT
ETPVW

>comp125826_c0_seq1_len=2345_ORF_1_(frame_2)_translation

MKTLVQCAGCERPIILDRFLLSVVDRAWHVCKVCWDCSTLADKCFSRDGKLFCKRDFRRFGTKCGGCSEGISPTDLVRRRARGVFHLACFTCL
LCRKQLSTGEELYVLNENRFICKEDYITSTKQHHHHHHHHQACSSSPVDSPEFDESLEMPGTDGSPAERDNERDSSIERDRDLHNDPVSGGSA
VNSPTGNGNGQQSNIVDGVHSSATDSCLQGSSQGASDDGGGPGTKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWF
QNRRSKERRMKQLSTLGSRRHFFRPRRAMRPLRAGLADGLMEDVQGHNGPYFYSDSSSPNDFGYGPHPGFYEFANHSQEMTVFPGTMP
GPASVDGHPLDSQNPPTVAGMPSMANGGVNLASDAPYMAPNGDYVRHHQSSPKNGMAIQIQASMDGYASRSAQEAYPNMPQQHQTPLSSDTP
VW

>Locus_1_Transcript_73444/166847_Confidence_1.000_Length_855_-_ORF_1_(frame_1)_translation

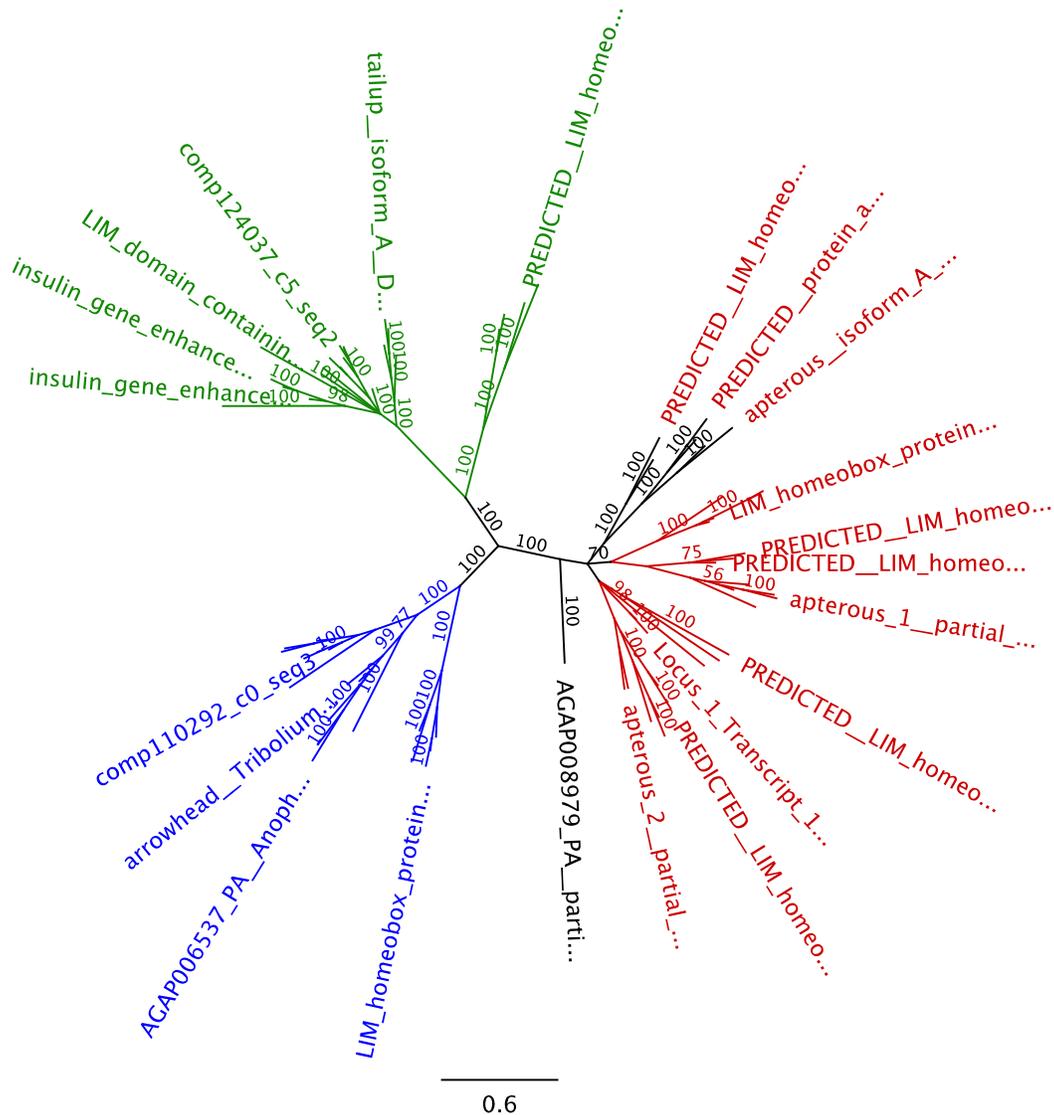
RRFGTKCGGCSQGICPTDLVRRRARGRVYHVACFTCLLCRKQPSTGEELYVLNENRFVCKDDFNRRQQHVSSPLDSPEESLSPGQDSSIVVSGGSG
ANSPHNNTPNSEEPVDCLVAGPSSNDEAGGGTKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQ
LSNLGARRHFFRPRRAMRPLRAGLMDAEPHNAFAYFSDSNSPNDFNYGPHPAFYDYFPSSHPQGGITDFPGTVPITAGGNNTLEHSNTINNG

>Locus_23222_Transcript_1/1_Confidence_1.000_Length_2928_-_ORF_1_(frame_2)_translation

MTTQSSCLLRNQEEDVSKASLVSELSALGLSYEPTGLRASKVDLTGTWRGRGAGRTLPLIRTDMMKMQCAGCERPIKDRYICSVLDRSWHAKCV
VCCECKCTLQEKYSREGKLYCRNDFFRFGTKCSGCAQGISPTDLVRRARNKVHFLKCFCLVCRKQLSTGEELYVLNENRFIFTDYLSRNSQG
VLAASSPSDGNLDDLGDGLGPDHGHLSERDSSVERDGGHNLNGGTSTVSSPTGPMHLHSTPGSAGGPASGPNPASCNGPNNSSSETTNS
NSAAQNDENTPGQKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLSSMGARRHFFRPRRAM
RPLRPGMSPDGLDDSPDMVSGPNSGYSYFSDSGSPGDFGYGGQPGFYDFPFGQPPPDGMAYHGVLPQNVPPPPTSMQPLSAVVGPGPVNLGA
DAPYLATHGDLMSQRSSPDSVLSLASDAYGPPRSSAGFPANLSGHPTMSETPVW

11.5.3 apterous

apterous2 comp118341 help by Julia Schneider



Supplemental Figure 11: Unrooted 50% majority rule consensus tree to identify true *ap* homologs. The Jones amino acid substitution model was chosen after mixed model testing. After reaching 3000000 generations a total number of 12002 trees were written in two files (each file contained 6001 trees of which 4501 were sampled). Colors indicate monophyletic groups in the tree.



Supplemental Figure 12: Unrooted 50% majority rule consensus tree to identify true *ap* homologs. The same phylogram as in Supplemental Figure 11 in a magnified view to display gene and species names.

```
>comp117589_c0_seq1
CCCCAATAAAAAATTTAATTTAGGAAAAAAAACAGTAAAGCCCATCTAGTGTCCCTCATTCTTATCGCGGGGAGCAGCCTCCCTTATCGG
GTGGCAGCAAAGGGCAGCAGGTGGGCAGCAAAGGGCAGCACCCATTGTCCCGAACCCTAATTTCCCCCTACTCATATGCTCTCATATAACTTG
GAAAAAAGACCACCAAACTTACAAGAATGTTTGAGAAATCGGGGGAGGAAAAAAGGAACATTTATTGGACTCACAAATGAATATCGGCTC
AAATAAGGTTCTGTAATGTTGTTGTTGTTCTTCTCGGTGATCTGAGTCGGCTGAAAGGGGAGTAAAGGGGAAAGACGCTGAAAAACGATTTGTGCTT
CAGAAGTGAGCCTGCGACGTCAGTTCGACAGAGGCTTCGAGTCTTGCCTATGTCCGGCCGTGTATACGCTGTCTGAAGATATTCGGCTCCA
TTTCGCTCTGGCGTTTGGAAACCATACCTGTAGCACTCTCTCGAAAGCCAGCTTGGAGGACAGCTGCTCAAGTCTTTGGCGTCTGGGTTAT
GGTTGATGGCAAATACGACTTCATGGCCCTCAGCTGGTGGTGTGTTGAAAGAAGTCTCATCTTTTAGGTCTCATGACTGCGCCGTGTTGTTCG
CTGGACGTGCTGCTGCTGCTACCCGTAGCCGTGACGCTACTGCTGACGTTGGATTGGTGGTGTGTCCCAACGAGGAGGAAAGTTGTGCGG
ACGCAAGGAACAACACCGTCCGCACAGTTCAGCAAAGACCCCTCGAAGGGGAGTAAAGGGGAAAGACGCTGAAAAACGATTTGTGCTTCTGGGA
CGACCTTTGAGTGGCTTTTCTGCTGGACGTTTGCCTTCTTTGCCACCACCGCCACCCTGTTGCAGCTGAGCGGCCGACTGCACGGCCG
```


Supplemental Material

AGGGCCAGGGACCTCGTCTACACCTCCACTGCTTACGTTGCGCCTGGTGCAACACGGCCCTGACTCAGGGCGATTACTTCGGTCTGAGGGACAAC
CTGGTCTATTGTAGGGGACATTTACGAACTCATGGTGCACGGTGATAATTCATGGCCCTCCGCTGCGGATAGTCCAGAGGGTGCACCTCGGACCCCT
CGACCACCAAGAACATTTCCCGGAGTCCGCTGATGCAATCGAACACGAGATCCGTAGCCAGTTGCTGCTACCTCTGCCTACTCGCCTCTACC
GGGTGCTCCCACTGGATCATCTGGTGTAAAGGAAAGGAGCCGAGAAAAAGTTCCGAAGCTGGCGGAGTGATACATCCGCAATCCACCC
GGTCTGACAATGATGGACGGGAAGGTTGTGCCCTTACCTGTCCGTTATGGACCAGAACGCGTCTGGTCTGACAAACGGGATCGGCACCTCTCC
GCCGCTCAAAGAACCAGAGGATGAGGACGTCGTTCAAGCACCATCAGTTAAGAACCATGAAGTCTATTTCTCGATCAACCAGAACCCGTGACG
CGAAGGACCTGAAGCAGCTGGCTCAAAAAACAGGGTTGTGCAAGCGGGTGTGCAAGGTGTGGTTTTCAGAATGCTCGGGCAAAATGGCGTAGAAA
CAATATGAAGCAAGGACTGACCTTTCGTCGAACGGTGGTGGTTCAGCCGACAGCAGCAGCAACCGTCTCAGCCACCGCATGTGAGCAGTC
CAGGAGCTACCAGCTCTTTCTCGGAGTTCGAGCCCGTCCGCTACCTGCGGTGGCGCTTCCACCACCGCAACAACAACAACACCCCGATCATCACC
CCCAGTCGATAGTGGCCGTCAGCCACGCAACATCATCATCATCACCAACAGACCAGCAGCTACTTCCACCGATCGACTACGAGCCAGCCG
GATCGAGCCTGAGCCCTGCCACCTCAAGCCCGACACATCGCGTGCACGGCAGCAGGAACCTTGGCGTTGACATCTTCCAGGAACCTTTCTAAA
TGATGAAATTCGAATCAGTGGGGGAAGGGGGCCCGCTCAGGCAATGTCAGTCACTTCACTGTACAAAAAAGAAGAAGAAAAAGAAATCCTA
ATGTTTCCCGAATGGTTGTCCACGGGATTTCCGAATCAGCTCCAGTATAGACATGACACACGCTGGAAAAAAGAAACTTTTGAAACATCC
TCTCAAATTTGATCAGTCATTTCTTCTATCATGTTGACCAGGGGTGCCAACCTCCAGCCCGGATCAAATGCAGCCCGCAGGAAGTTA
TAACTGGCCCGGAAATGTGTTCTGTAGAAAAGTACAATAATGTAGTATTAGTATAAAAAGTCAATCCACACAACACGATAAATGCCTATGC
TGTAAGCGTGGATCCAGGTAATGTATCCGGAAGGTAAGTTTATTGTGTAGGCT

>comp120245_c1_seq1_-_ORF_1_(frame_3)_translation

MPVISIGDYEPKPHLCAAGCGLKISDRYLLAVDRQWHARCLCKECHVPLDSELCFARDGNIYCKEDYRRLFVAVKRCARCQRGIFANEMVMR
DLVYHLHCFCTCAWCNTALTQGDYFLRDLNLYVCRAHYELMVHGDNCMASAADSEPALGPLDHQKPPGVALMQSNHEIRSPVAATSAYSPLP
APTGSSGVRKRGRKRKRSSEAGGVHPPHPLTMMDAEGCALHLSVMDQNASLTTGSAPPPPPQRTKRMRTSFKHHQLRMTKSYFSINQNPDA
KDLKLAQKTGLSKRVLQWFLQVFNARAKWRRNNMKQGTDPSSNNGGQPNQTTQQQPSQPPHVSSPGATSSFSESPVATCGGASTTGNNNNNPDH
HPQSIVAVTPTQHNNHHHQTQHVLPIDYEPAGSSLTPLPPQARHSHGSENLPDLSFQELF

>Locus_1_Transcript_14384/166847

AATTGGAGTCCAGTGGAGCAAGTGGGATGGCCCTCTCACAACTCTCATGATGGAGAGTGGTGAGGGAGTTGCAATCTCCACTTGTCTC
ACTGGACTCCAATTCCGGCGAGTCCACTGGCCTGAACCCATCCAGTGTCTCAGCGGACGAAGCGGATGCGGACATCCTTCAAGCATCATCAGT
TGCGGACGATGAAATCCTACTTTGCTATCAATCAGAATCCCGATGCCAAAGACCTGGAGCAACTTGGCCAGAACTGGGCTTTCCAAAAGGGT
CTCCAGGTGTGGTTTTCAGAACGCCGTGCCAAATGGCGCCGAAATATCTGAGGCAGGTGGACCTCCAGCAAGCGCGCCAGAACTCTCAAATCC
TCAGCCCTCTCGCAATCTACGCATCCGAACAGCCCGCGCCACAAGTTCTTCTCAGAGCCAGCCGTCGGGGGGGATCGGGCATGGGGC
AGAACTCTCAGCCCTCTTCACTTCCGCCCTTGGCCGACGCGGGTGTATGGAATACGAGGGCAACTCCGGTCTGACGTCACCTAGTCCCAAG
ACACGTCTCCCCAGTGGTGAACCTTTGCCCTTTCAGATCATTCCAGGAATGTTCTAAGTGTGAATTTTGTTTTAACTTTACCCGTAACCTTGA
GGAAGAAGGAAAAAATAATATCGAACGTCAGAATTAATATTCATTTTTCAGAGTAAATTTTTCAGTGAATTTTTCAGTGAATTTTTCAGTGA
GGATTGAAATATAGAGCACTACTATGACTTTCAATTTTGTGTTTTGTTTTGAACTTTCAGAAATCGAAGTTAAAGTTATTTTATGCACTCT
GTTGGAGTTGGCAATCTACGGCTAACGAGCCGAGTCATGTTCTGGAATAAATTTGAGGAGTTTGTAAAAATAGCTTATTGCAATATTTCAA
ATTTGTTTCAAACCTTATTTTATTTACTCGTAACGACCCGAATGATATTTTGAATCCTCATCAAAAAGAAAGTCAAAAATAGTAC
ACACATAGTGGTGGCAAAGTTTGTAGTGAATAAATAAACCGCTTTTGAATTTTAAACAAATGCTTCTGAGAAAAAGCGGAGATGAAAGT
ATCAAAACCTGTTTGAATGCCAGATA

>Locus_1_Transcript_14384/166847_Confidence_1.000_Length_1163_-_ORF_1_(frame_1)_translation

NWSPVRTSGDGLLSQTLMMESGEGCNLHLSLSDNSATSTGLNPSASQRTKRMRTSFKHHQLRMTKSYFAINQNPDAKDLQLAQTGLSKR
VLQVWFQVFNARAKWRRNNLRQVDLQGGQNSQNPQSSQSTHPNSPGATSSFSESPCGGGSGMGAELSHASSLPPAAQPGVMEYEGNSGLTSL
VPRHVSPSGEPLPLTSFQELF

>Locus_4700_Transcript_14/16

GCNNNNNNNNNNNNNNNNNNNNNNNACGCGTACCACACTTTGGGAAATACTGACCAATGTTTTAATGAAAATTACACCACCCGCTTTAATT
TGGGCTAATAAAAAAATATTGACTAAGAGCTTTTATGAAATTTAAAAAACCATAATATTAAC TCCAAAATAATTTCAACATTTAATTTAATT
TTTATAAAGATTTATAAATAAAGTAAACTCTTTTAAAGTTTTTTCTTCAAACCTTAATCATTCTAGAAATAAAAAAAGACATTTGCAAAAAATA
AATTTTCTGTGAATACTGGTCTTCAATACGCTTACATGTAACAACAGTCATAAATGAAAAACAAAAAATAACATTTGGCGTGGTCGATATTT
GAATACACAATTTGTAGCCAAATAGCAAAAAATACATAAAACGTTAAATCATATTTATGATCAGTCTTTAATATGCTTTGCAAAAAGAACCCCA
AAAATATGTAATAAATAAATCATAAAGTTAAATAAAAAAGGTAACCAACGTCGTGAAACATGAAAATAATCTTATGTTGATCAGTCTTTTGA
AGTCTTATGATTTGGACACCTACAAGCGACGTCATCGTGGGTATCGTATCGATCTTGATTGACGATTTGGTTTATGTTAGAACTTTTCTTTCCATT
CTATTTGCAATAAGCCTATCCCGTCAAACGTAATTCGATAAGCGATCTCGATAGGTTATCGTATCTTTCGATAAATTCGAAACTAATAAAT
CTTTTAAATTAATGAAATATGCAATTAATACTTAGAAAAAGAGATAAAAATAAGTTAATAATTTTCCCTTTCATAAGTAAAAGAAGCCTACAATAA
AACAAGCCATAATAGTTGTAGCGTTAACACGCCATCTATTTTATAAACATTCAGCTGGTGTGAGGTTATAGGTCAGGTGGTTGCGTCTCAGCTG
TCTGCTTCTCAGAGTTGAGAGTATCGCATGGAATAACAGTTAACATTTTAAATATACCTTTGCGTAGTCATGGTAACAAAAGTAAATAAAAGGTAG
AGGAATCAATTAATAATCGCTTTACAGTGGAAATGTCATTAATAGATGAATATATCCCGTAAATGGGCTATAGTTCACTGAATATGATAGCCTT
GTCAATGGGCTATATCTATGCATGCTCTGTAATCGAGCATAGTACGCTTGTAAATTCGTAGCGAAATCCATTTAATTTGGGTTTTAATTAAT
ATAAATGCTAATTTTCTCATATGCAATATATTTTGTGTAACATGAAATGAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT
TTTTTAGGCAATATATTTGTTATGAGTACCAATAGCTTTTTTCCCTTTTTTCTGTGTGATGTGATGTTTTCTAAAATAAATAATTTTGA
ACTGTAATAATAAATAAAGATATCACTTCAAGTATTTAAAAAATTTTTTTTGTGAGGATTTCTATAATTTAATATTTGAGAATAAATAC
CAAAAATAAGCTATACAATATATAGTTTATAAATAATCAATATTTCTAACTTAAAGGTTTTGTAATTCATTAATAAATAATTTAATGATACCA
ACTTTTAACTTTAACTAGTATTTAAATTTCTTAGATTTCTTAAAGTTTAGTATTTTCCAGTATCAACCTACAATAAAGCAACCTATTTG
TAAACAAAAATACTAGAATTAATAAGAGTGGGGAACCTCTTTCTATAATGTTAACTGTTAATTTGTGTTTCTTAAACAGTGATAATGAGGTT
TATGCACAGTTAAGAAAAACCTTAGTACTTATAGACTAACGAACTTTTTTAAACTTAATTTATCTGATATTAAGTAAATACTATCATATACT
ATTTATATATAACGTAATGTAATAAATCGACTGTTTCCATTAGTGTCTTTTTAAATCTACTTTTTTAAAGTATTTTTCTTAGTTTTCTT
TCTTTTAGTACTTCTGTTAACCAAAAAATAATGATAAATAAATAAAGCAATTTGCCTACAACAATAGAGGAACTCCACAGAAAGGTT
TCCCATCTCAACTCTAATTTTATTTCTAAATTTAAACAAAGTCTTAAAGTTAAGAAAGTATTCACTACTTCCCTAAATAAACAACCACTAATAA
AAATAAAAAATAACTTTATAATAAATAAATGATGAAAAAGAAATAAATTTGTTTTGATATCATTAGTATTTGTTTTTACTTCCGCTT

TTTTTAAACCTATTTGTTGACCAGTTAGTAAATGAAAACAAAGATTCTGAGATTCAGAGTTTCCAAAAAGGTTAGGATATATTTTGTGCAAC
TTAAACTAAAGAAAAGAAAACGTGAATAATGAGCAATATTTGAACATAAATGGACGCTACTATACTTTAAGCTCTATAAACGAAAATCTTTCTCA
ATTTTTAAAGAAAAGAAAATGGAGAGAATTAAGGAAGTTTTTTTTTCAACCGAAAACACAATATGTCAATTTCTGTACTTTTCAACGTTGTT
TTATTAATTTTTTTTTTTTTCATTTTATCATAACATTCCTTTTTTACATAAATTTTATAAGTTTGTATCTATATTTAAATAAAGGAAATTT
TGTGGAAATAAAGGCATTTGTGGACAATGACTGAACACAAAGTTTAAAGAAGACTGGCATTCTTTACTACCGAGGGTACGAATGATAACGGCG
GGTAAAATCATAATTTAATTC AATACGAGAAATCCAGCAGGTCCAACTCAAATTTAAAGCAATCACTTCCAGACTTTACGAACAACATTTTAT
TCCTATGAAACCAGCTGTTGAGCTCCACCAAACGATTTACATGGAATTC AATTCGGAGTTTAAATATCTGAACCGATACTTAAATTAATTTTA
TTTTAATTAAGTTGGACTTATAGTATTTCTTGATGATCCGGAGGGTACTGGCCATAAAAAGCCAGTTGCGTTCAATGCTTCTATTAAGGAAATTT
GAATTCAGATGATTTGTTGAAGTTAAGTCAGAAATGCCAGAAGTTCATTGTTAAAAATTTGTATTTTCTTAAATTTAAATTTTTAAAAAATTA
TCGGATAGTAAATATGTA AATGCTACTTTGTGAAAATGCATTTAACAGTTAACATAATTTATTTTGTTCGCATACTCTATTTGAATCAAAG
AAAAAAAAGTTACTTTTCTTTGAGGTC AATTTAATACACTAATTTATGTTTCATCTAGAAATGAATTAATTTGTAATTTATTTCTGATATA
TTTTAATTTAAAAAAAACCTCAAAAAGCCGTAATCAATTTTTTTTTTCCATTTGAAGTACTTTTTGAAATGTTCTTCCAGACATTTTTGTATTG
CAAAAACTTTTATTTAACTATTTATTTGATAAACACTATAGCAAAAAGGCACTTAAAAATTTTATACATAAAAAAATTTAAAAACTTTAATAACT
TAATACTTTCAAAAATCCCTCCCTAAATTTACAAAATAAACATACTAAATTTGGAGAAAAAAAATTTAAATATAAAAACCCCTAAAATATCTCAA
CTACTATAAAGGAAATTTGTGAGAAATTTGGTGA AATTTAATGTGGTGGGAATTTGGCATGCTTTTTAAACCTTTGACGCAGTTGGGAGATATCGC
AGTTGGGTGTCCCTTTTATATATTTTCCGGTGGGTGTACCCACTATTTAAAATACTTAAAACCTCAGTATTTGTA AAAAAAGTTAAAACTGTTTA
AAATTTACATGAGATTCAGAGAGTATCA AATGGATTAAAGTATTTCCATTTAAAAAAAATCAATTTACCAGCTCACAATATCATGCTCCGC
AATATCGACTCAATTTACTAGAACACAGATTTGCCCTACAAAACAGATTTAGAGTAAATTTAGAGTAAATTTAGTAAATTTAGTAAATTTCTGATGTT
CTTGTGTTATGGGTGTCCGCTGGACATCGAGGATCGAGGAACCCAGTCAAATGGATCGCAGAAATTTGAGGCTGAGACGCCAGTTTCCAGAGAGG
TTTTGACGCTCTGTGCGATTTCTGGACCTTGGCTTTGCTGACGGAGCAAATTTGCGCTCCACTTTGCTCTGCGCTTTTGGAAACCAACCTGTAGCA
CCTTTTTGCACAAGCCAGTCTTCTGGGAGAGCTGTTTCCAGATCTTTGGCATCAGGATTTGATTAATCGAGAAATATGATTTTCTGCGCTCAGC
TGATGATGCTTTGAAGGATTTGCGCATGCGCTTTGGACCGCTCAGGTGTTCCATTTGAAGTGTCTGCTGCTCACTTTCCACAGCGGTCAAGCTGTTT
ATTGTGCTGAAGCGGCGAGTGGTGTACTTTGACAAGGAACAGTGGTGTCTGGAGCTGGGACCCCTGTGGCCAGCAGTAATGAAGAGGAGAGA
AAGCTCTTTTGTCTTTCTTGGTCTTCCCTTCTGACCAAGCATTCTTCCGATTTGCATTTACAGTCTGATGGGAAGATCTTCCCACTATTTACTGTC
CTCTGCTGCTGTTGGAACAGTTCATACCGCGGCAGGTTAGGAATCTGTGCAATGTTTGGCACCTTCCGTTGAAGCACTTGCCTGCTGAATCTT
GCTTGCCTCTCTTCTTCAAACACTGCTGGAAACAGAGTGGGCTCTCATATGCAAGCATCTCATAGTACTCTGCAATAAATGATGCTC
TGCTTCAGTCCGAAGTATTCACCTTTTGTGAGGGTTTGTGCAAGTGGCGCAAGTGAACAGTGGAGGTTGATAGACGCTTCTCTGCTGCTGAT
CACCAGTTCTCTTGGAGAGATGGATAGGTGGCAGCGGTTGCAACGCTTGTATAGAAAACAACCTGTAATAATCTTCTTGAATAAATATGGCCA
TCTCTGAAAAGCAGGTC AATTCAGAATCCAGTTGCATTTGCACTGACAGCAGGTCAGGCAAGACATATGCCACTGTCTGTCAACAGCCAACAG
GTAGTACCTGTCAACGATGACACCCCGCATCCAGCACACCTTACAGGTGATGAAGATACAGACGATTTGCCCTGCTCCCACTGATCCACCCCT
GCTGATCTTCCGCTGATGACAGGATTCACTTGGCTTGTGCCACACACACCAGCATATCGTTCTTTCCCGAAGAAATTTCTGTTCTGGACCAAC
TCAGGAGGACGCGCAGCTCAGGAATCCATATTTCTGGAATAGTTTTTCTGTTTTCAGGAAGTGTGGATCTTCCAGACTGTTTCTGGAATTTCTGTGGA
ATGACGCTGAGGCTGGCATATCGATGATTTAACTCTCAGTTATACTTGAATCCATTTCCGATAAAGAAAAAGTCGTTTGAACAAAACAGT
CAATAAAATTAACATAAAAAAATAATTTATGTAATCACTGTA AAAAAATGTTTCTTCCACCTGATGACTATAGATTTATGAACAGAGGGACATTC
CACATACTATCTAGTCTCACTTCCATTTGAAGTGTGTTGCTGCAACGACTGTTTCCGAATTTCTGTGGAATGAC

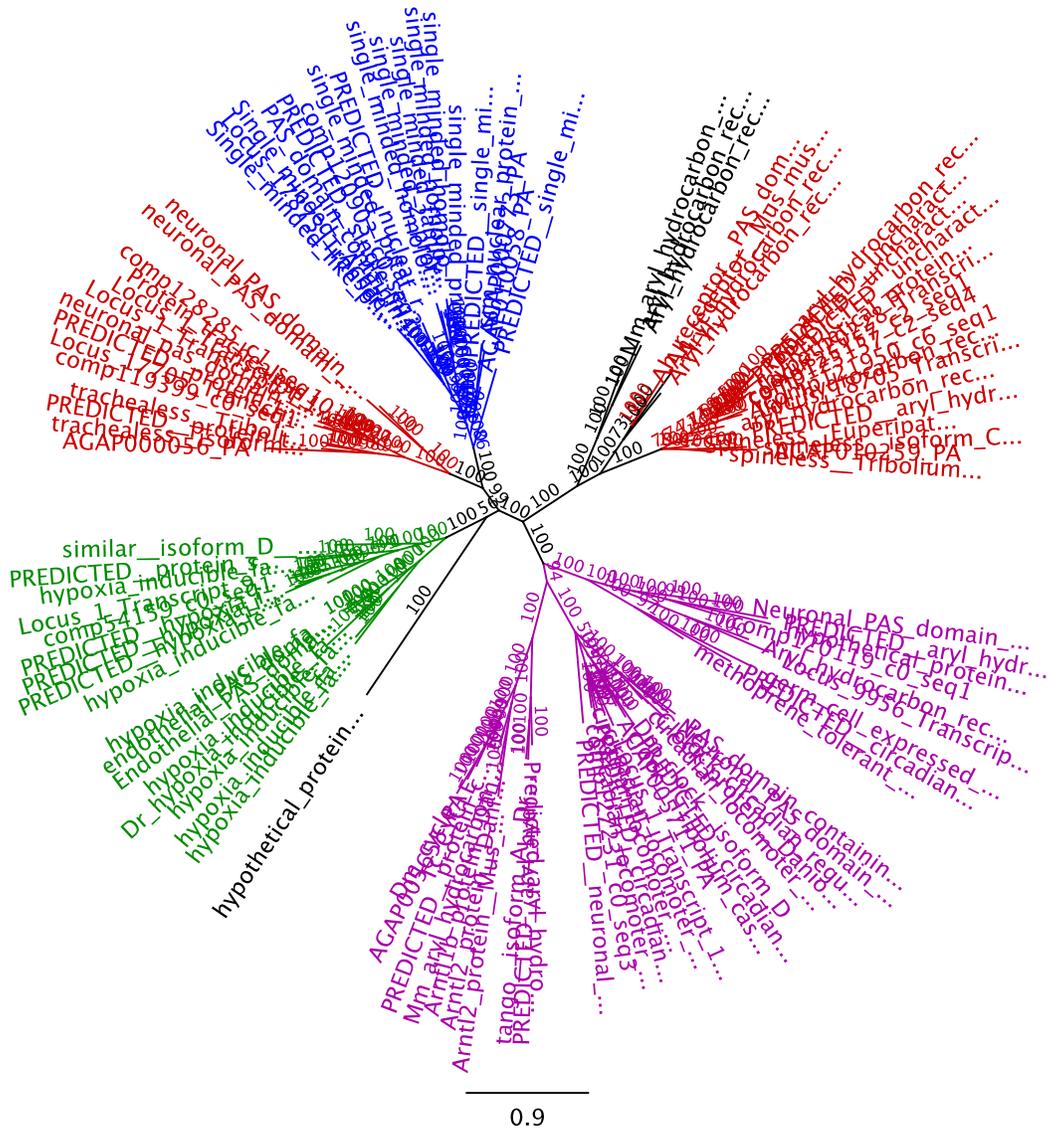
>Locus_4700_Transcript_14/16_Confidence_0.321_Length_5545_-ORF_7_(frame_2)_translation
MLVVCGSNASESCTEERISRGGSMGAEGKSSVSSSPVRCAGCGGVIIVDRYLLAVDRQWHMSCLTCCQCKVQLDSELTCFSRDGHYKEDYRFLFSI
KRCTRCHLSISPRELVMRAREHVYHLHCFCTCATCNKPLTKGEYFGLKQDIHCRSHYEMLLQYERPYLCSQQLKEEGKQSDGKCFNGRCQNIQI
PNLPRYGTVP TAAEDSK

>Locus_11195_Transcript_3/9
TGCCCAATAAACTGTACATAAACA AAAACGATGCATTTTCATCGTTTACTGTGTAACCATGAATTTATTACAACCGCAATTCACCAGAATGCG
TATCAGTGAGTATTTATTTCTATATAAAATATGTATGAATGCAATGTGATCAGCAACACAACAAGAGAAAAAATTTATATACAAAAAACAGAGA
TATGCATTTGTCGATGAAGTGTAGCGTCAGATGTTCTATACAAAATAATCGTAAATTTCTCTCGAAAGGAAGAACATTTCCGCAAAAACCAAACT
TTTTTTTTCTTAGTTATTTGAACGTTTACATTTATGAATTTGCGTTCGCTTTGCCAAAACACTTATTTAGAAGACTCCATCAACTTGTTTAGTT
TTTTTTTTGCTTTTGTACCTTACACAAATGTACGATGAGTGGCATTAAACGAATGTGAGCTAAGAAAAAAAATAC TGGAGAAACAGTTTC
AAAAGACATGATATTTGAAAGAAAGGGAAAAAGCATGCTATACAATTA AATATTTGATCTATCTTGTATGTA AAAAAATAATAGTTTGGCTA
TAAAGCTATTTTGGACGAAATTAATTTATTTCCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
ATGATTTAGAAAACGCGCACTAACATAACATAAACAACACTAGTATTTTGGGATTTATTTATTTAATTTCTGACAAAATATATAAATTTT
GGACAAAAGAGAGAATTTTTCATTCATACGAATTTCTTTTATAAGTACTTCAACATTTAAATTTTACTTGTCTTTTTTTGAGAATTAATA
ATGTTGGTATAATACTTTATAGACCTTTAAGCATTTATCTTGAATTTAGAAAAACTGATTTAAAAATTTAAAAAATTTATAAATTTCTAAAA
GAGCATAAATTAATTTCTTAAAGAGCATAAATTTATAGCTCTAGCTGCAGGAGTAAAATAAATAAATCAGCAAGCATTATAAATAAAGATCTT
CTTAAAAATACTGAAATTTTCTAAAAATCTCACAGAGAATTTTCATGTGATTTAGAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
ACAGAAAACGGTAAATTTTCTAAAATCCGCAAAAACAAAAGGAAACGTTTTCAGTTTTAAATTAATAAGAGACGTAGGAATAACTGTATGTAA
AATTAATAAATAAATAAATTTTAAAAATATATAAATTTAGGTTCTTAAAAATAAATTTAAATTTAAACATATGAACGAAGAAGTTTCAAAAAT
ACATTTCTGACACGTTGTTGACTTTACTTCTATAGATACTTGCAGCTGCATAAAAATGGTTAAATTTCCACCCGATTTTCTGATATAAAG
TTTTAAAAAATCTGTTAAGAATAGATCAAAAATAGCAGCTTTAGATCTTCCGAATGAACAAATTAACCAAAATACTTAATCTAATTAAGGA
TTAAGCTATTTACACAAAATTTGAATTTAGCAAAAGTAACTTTCTGAATTTAGATTTAAGATTTAAGATTTAAGATTTAAGATTTAAGATTTA
ATACAAAATAAATTTTAAAAAGGCTGATAAATAAATAAAGATTTTAAAAAATGCTTTTATATTTGGGTTGTAATAAGTAATAAATAAATTT
ATATGTTTAAATCCCTCGTTTACAATTTGTTTAAAAATCGCATTTAATTAAGCAATGAAATTAAGATTAACGAACGCATAAAAATAAAGGAAATTTAAG
TGATTTGAAAGAAAGTTAATCTTTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
AACGAAAATGAAAGATAAAGAAATTTATAAGATAAATAAATTTTAAAGAGGAAATTTAAAAAATTTTCTTAAATAAATAAATAAATAAATAA
CAACAAAATATTTTAAATCTTCAAAATTTATTTCAATAACTGGAATACTGTAATAAAGCTTATCTTATTATAAATAAGTTAAAAAGTTGTTATAAATG
TAAAATAGCTTATGAGATTTATTTTACGCTACGATTTTCATGATTA AACAATAAATAATCATTTTAAAGAAATGCTGCACTAATAATTTTTC
ATAAGACATTTGAGTTAGACTACAATTTAAAAATTTTATTTAGATAAATAAGCAAAATCAAACTCTGCGCTTTTCTTCTCATAACTAGTCTTTTT
CAGTATCTTTGCTCGCAATTTCTTTCTCAACTCTCTGAATGATTTTAAATGATTTAAATAAATTTAATAAATTTAAATATTTTATCATTTTT
TAATGTTTATATAAATTTCTACTAAGTATCAAAAAGAAAGTAGATTTAGATTTAAGCTAAGAAATTAAGCTAAAAAATAAATAAATAAATAA
TTATAGCTATTTAAAAAATAAAGTCATTTGATACAATTTAGAAAGTACTGCCAGAGATTTTTTAAAACTAACATTTACTTTTTCTTTGTATT

ATACTAACAGTCAAATTTAGTTAAAAATTGACTATTTAAACCATGCTATAAAAAAATCAGGATTGTACAAATAAGATTTAAAGCACAAAAC
ATTGTTGTGTCTAGATTTAGAGATATTTCTTTAAAAATTCTAAACATTAACCTAAGTTGATATTCGAATTGAACGCTCTCAATTTACACAACTCAA
AGTAAATTTAAGCTTTATCAAAACCATTTAAATTTGTTATCTTAGTTAAGAAAAGTTAATCGAGTGAAAATGGTTTACATAACCAAAAAAACAG
TTTGGTTTAGAGTTAAGCTTCAACAATAAACTTTTAAAAAGATTTTCTCTCCCGAAAACAAGATAAATTAGATATTTCCATTAAATTTCA
ATATATAAAGTTTCAACAAAAATAATACGATGATAATGGTTATTTCTCCACACTCCACTTTTAAACAATAATTTTATTATTAATAAAAAATAATT
AAATCTGCAATATGGTGTCTCAAAATAATTTATGAGAAAAAGGAGAGTTTGAACATTTAATTCATACTTATTTTATGAAACCGATAAACGAT
TTTAATAATTAATTTTAAAAATCGAGATTTCTTCAACAATTTAATAAAGTGCAAGTTTCAATGAAAGTTTCTATTAGATAATATTGATTTT
GAGCTAGTAGATTTATTTGTAGTTAAAAAATATGAGTTTGGCTATCCTTGGTGTCTTCCATTTAAAAAGTAATTAACCTGTTTTCGGTGTG
AATCAAAAATGACACAGGGTGTTCGTTATCATAACTCTCAATGCATAATAATATCATGAAACTTTTCAAAAAATAAGTAAAAATTACAGGCAC
AAATGAGGTCAGAAATTAATATTTGAGGAAGAAAAAGAAACGTTATAAAAAATGTAACAAATCATCATTAAAGACAAGCGAGAATATTAATGTA
TGTCTTCTAATGGACGAAATTTGTCATCTAATATTTGACGGATTCTAAAAGTAAATATTAAGGGGTAATTTGAAGAGCACCTTACTCAACAA
GAACGATGTTAGAAAACAACTGTTTATTTAACAGCAAAATATGACTTCTTGTGTATATCTTGTGATATAAAGGCATGTTAAAAA
AAAAAAGAAAAAAACTATATTTATGCAACGTTCTCAGCTTAAATGTTTTCAGAATTTCTAAGAGAGGCAAGGTGCATCTTATCAACTTAAAA
TCGCAGAGATGCAAGGTCAGAAAGGCTTTCCTAAAAAGGCAGATGATGTTGCAAGAAAATTTGATCGTTCCTTCACTGAATCGTCATTTCCGCAC
AAGGACACAGAATATTCAGGAATTTCCATGTTCCGTAGATATGAGATCTGAATTTGCAACTTCGAACACTTTCTTGGTCCGATTCCCTTCCGCTT
TGCACATTTATTTAAAAACAATATTTCAAGTACGGTGATTTTCTTATTTCTTATGCAATCACCGCTGCAGGAAAGTATTACCAGCCAGCA
TTGCTTGATCTCGAGTCTTCCAGGATACCTCAACCCACCTAGAAAATTTGAAACGTTTACCTGAGACCTTGAATAAGTTTAAATTTCAAAA
TAATATTAACCATTTGGTTACGGCTGTAACACAGGTGGACCAGAAATGACGGGCTTCACTACTAAGCGCGTAAGCGAAGTGTAAATAAT
AATAGACACATAATATTTGTTATGCAATGCAATATATAATATTAAGATGGTAATATTAATACATTAACCAAAAAAATTACTATTTTGTTCATTT
CCAAACTTATACGTGCAATCGAAAGAGTCCAAAGATGTTGCTGATGGATCTTTAAAAAATGCATGGGCATGATCGACTTTAGGAATATTCATG
AGAGTTTTTGTGAAATTTGTTCTATCAGAACAGTTTCAATGGAAGGATGTGACATGGAAGAGACATTATCTCCATTATGGACAGATCTCGGATT
AAGTCTTGGTATTGACCAAGTGGGGTCATAGGATCACTACTGTCAAAGTCCATGGATGCTCCAGGCCACGTTGGGACAGGAGCCATCTG
AAGCGACACCACAGGAATGCACTGGACTGGGCTCGGAATAGGAAGTGTAGCTCCGGGCTACAGATCCCTCCCCCATGTGGCCGCCCTGG
AGCAACTGAGGGTGGAGGGCTTTCGCTGGTAATCATAACAGAAACAGAATCTGGTTGTTTGTGTTGTTTTCGCTCCATTTCCGCTCGGGCTT
TTGGAACCATACTGAAAGCACTCTTTTGAAGACCTGTTTCTGGGCTAATTTGTTTAAATCTTTTGCATCTGGATTCTGATTGATGGCAAAAT
AGGATTTCATAGTCTTAAATTTGATGATGTTAAAAGATGTGCGCATGCGCTTTGTCCTCTGACTCCCTCCACCACCAGATGATAGTGGGGTGAA
GATTTGAGAATTACCAGTACCATCTAAATGGCCTAGGTGAAATTTGACAGGAGGAGGGTCCAGCGCGGATAAACTGTATTGTGCTGATAGTG
GATCTGATCCGGATAGATCCCATCTTTTCTTCTCCTCGGTCCTCTTCCGGCCGACGCCGACTCCGCTGAGTTATTTACACATCCGGCAG
TGTAGGGACCATAGAATCCGGAGTTGAATCCGGCGTGTGATTTCCGCCGAGTCAGAAGGTAGACCGGGAGGAGTGAAGTAGATGATGCGGAGG
ATGAAGGAGGGATGACGGATGATGATCAGGGCTGTAGTAGGAGGTGTCGTCGCGGGAAAAACATCTCGTAGTGTGTTCCGCAGTAGACGAGGTTG
TCCCGTAGACCGAAGTAATCACCTTGTGTGAGAGCGGTATTGACACAGGCACATGTAAAAACAATGGATGTGGAAGACGAGATCCCGCGCCGCAT
CACCAATCTTTGAGCATGGATACCTTGTGACATCGGGAACAGCATCTTTTGGATGAGAACAGCCTGTAATAATCTTCTTTACAATAAATATGGC
CATCTTTCAAAAACATGAAAGTTCTTCTATCTAGTTGTGCCTTGCCTTATGACACTTTAAACAATAGACATGCCATTGTAAGTCAACAGCATTT
AAATAATAACGGTCCACTATATGCGCGCCACAACCCGCACAGTCTGTTCCGAGGCGGGGAGTGTCAAAGTCCGAAGACGATATGACAGGCATGAT
CCGACTGTCCGGTGGATGGCTAGTTACCATCTGTGGACCATAACGCTTCCGAAGAAAGGGTCTCTGGTATGGTACCAGGGTGGCCGTTGGTGTGG
TTTCGTTAAAAAACAACAAAAGTTACAGTAATATTTACCAACGCACCATTAATTTACCAAAGCGCCACTATTTACAGATATTTTGAACATAAAA
CATCTATAATTTGTAACATCAAATAATATGAAATAACTATGCAACACGACTATGCAAATTAATTTGCATGAATTTGATTTCTATAACACGTGC
TGCTAGAACACAACCTGAATTTGTACAAATGAAAGAATAAGATCGGTCGCCCAACCCGCACAGTCTTCGGAGGCGG

>Locus_11195_Transcript_3/9_Confidence_0.733_Length_6107_-_ORF_7_(frame_2)_translation
MVNITVFLFFNETTPTATLVPYQRTLSSEAYGPQMVTSHPDSDRIMPVISSDFDTPPPNDGAGCGRHIVDRYYLNAVLDLQWHVYCLKCHKCK
AQLDEELSCFWKDGHIYCKEDYRFLSSKRCCSRQQGIIHAQELVMRARDLVFHIHCFTCAWCNTALTQGDYFGLRDLNVYCRTHYEMFSRDDTS
YYSPDHHPSLLHPPHLLTTPGLPSDCGGIPTDSTPGFYGPYTAGCVNNSDGVPSGRKRPRKKNKNDLSGSDPLSRHNTGLSPLDPASCNLHL
GHLDTGNSQSSPPLSSGGGQRTKRMRTSFKHHQLRMTKSYFAINQNPDAKDLKQLAQKTGLSKRVLQVWFQNAKWRNRNNKQPDVSVG
MITSEQPSTPQLLQGHMGGGICSPGATSSYSESPVHSCGGVSDGSLSTHVGPASMDFDSSDPMTPLVNHQGLNPGSVHNGDNVFSFHVTSFHE
LF

11.5.4 spineless



Supplemental Figure 13: Unrooted 50% majority rule consensus tree to identify true *trh* and *ss* homologs.

The Jones amino acid substitution model was chosen after mixed model testing. After reaching 3000000 generations a total number of 12002 trees were written in two files (each file contained 6001 trees of which 4501 were sampled). Colors indicate monophyletic groups in the tree.



Supplemental Figure 14: Unrooted 50% majority rule consensus tree to identify true *trh* and *ss* homologs.

The same phylogram as in Supplemental Figure 13 in a magnified view to display gene and species names.

>comp121950_c6_seq2

CGCAAAAATGTGATCGGTGTTACCCACCTCGAATGAAATGTCCTGTGCAAAATGGCGTACTGCTGAAAGTCACTTGAGGAGCCACTTTACCACTT
GCAGGACTGGTACTGTTGTGCGGTGAGGCTGATTTTGGAGTTGCTGCGCTACCACTCCATTTGCATGGTGAGCCCATCATATTACCACCACTGAC
TTGATCGATTTGCTGGTGCCCGGATTTTGACGACTGCTCCGAGTACCTGCCAGACAATTTCAAGTTCATAAGTGTATTGTTTCCGGGGCTACCGG
ACGTCGGGGTTGTGCGGCAACTTCTGTTAGAACTGGAGACACTCCAGGAAAGTCTCCTTCTCCAGCGTTCGAAGCGTCATTCTGATGATT
CCGTTGTTCCACATGACCACGCTCTGACGATCCGTCTGATAATCCAAAGTCTGTTGCTGATGATGATGATGATGCTGATGAATGATCGCTTATC
CTCTTTCCGAAAAAGAGTCAGATCCCTGGAATTCATCAATACACCGTTCTGCTGCTGCTACCGTTTCATCCCGGCTTGTAGTGTCTTTGCGTTTGTGCT
GAGACAATCGTTGGCGAACGTGAAGTTGATGCCACCGTTGTTGCTGCTGCACGTTAGTTTCGCGAAAATTTGAGATCTCTATCGATCGATGATGGCA
AAATAGTCTGAGAGTTGGCCGTGACCACCTGCTGCGGCAAGTTCTCATCACATCACCAATTAGTGGCCACGTAAGGTCGCAGCAAGAGTTGAC
GATTTAGTGGAGATTCGGTTCCACCGTTGATTTCTGGGCTCCATCTGACACTGGTAGAGCCCTTCGTCTTTACCCAATGGCGAAGTAGCGTT
GCTTTCGGCTGTAATCCAGCCGAAAAACAGTCTGGCAGTTGCCGCTGCAGTCGCCGCAACGCTAAGAAACCGTTGCTCGGATATTGGGCGA
CGGCATAGTCAGTATAGTATGTGCTATCGGCTGTAGTGTAGGAAGGCTTCTCAGCGCTGAGGAACCGGTTGTTGCTATGGTGCCTGAAAT
AGGTTCTGTTGATGCTGGCACTGGCGTAAGGTGCAGCGATTCCGTTAACGTTCTCAGCGGCTGATGCGTACGGGTAACATTATTATAGCCCGC
GACTGCAGCCGCTGCCGTTGCTGCCGCTGCCACGGCCGCCGCTCCTCCATGTACGATGTGTCCGCGAGGGCACCTCCCATCATTGCCGGTGACGT
ACCTCCCGAGGAAGAGGACGACTCTTCTTTTTCGACAGGTGCGAGTATGTCCTGATCTGAGATTTGTAGCGGCGGTTCTCTGACACCTTA
GCAGGAAGTCACTGTCCGAGAGGAGTCCGCTTCTGTCCGCAAGGGAGTTCAGCCACCGCTAGATAAGTCACTTTGAAGTCCATAGTTCTCTTG
CCGAGAAGATCTCGCCCTTCTTCTTCCATGAGAGGCTGTGCGTGCACAAGATGAAATCAGGTTTACTGTTCTTGTAAACGAGCCTGGCACTAGT
TTGTAACCACTGCCAGTGCCTTCTTTCGACAGGTGCGAGTATGTCCTGATCTGAGATTTGTAGCGGCGGTTCTCTGACACCTTA
CGTAAGCGAGGTCAATCGAATGGACTAGATCGTATCCTCCTTTGCTAGCGAGTTCGGCGTCGGAATAACCCAAAACTTGCTTTGCCCTTTGGTCC
AATGAGACTAGAGATAAGTCGAGCTTGTGTTTGTCTTGAACATGGCGTCTTTTGGGCATCTCGAGTAAAGATGGTGGACCGAAAGGGTGC
ACAGAGCGAACAGTGCTAGTGGCGGCTTTCGGTCTTGTGATTCTGGCCATGCAGAACCTTGATTCTTCTCCGATGTCCAACCTCAAAAAACCA
GATGTGTTATCCAGCAGACCCGAAACCTAACAGTAAATTCGCTCCAAGACTTGACTGTTGTCTGACTGTAAGATTCCCTGAGCGTGAGGGC
GCTACGGTCAAGGTAACCTGGGCAATCCACATCAGGTGGCGTTGTAGCTTTCCTATCTTCAGAGTGGACCAGCTCATAGACACTCTGGTGAC
AAATATCCGACTGATGGAACCCAGGTAAGTTTCTACGGTGTGCGAGGTGTAGAAAATTTCCCTTCACAGGTGAGGATCAACAGAAACCCGTTT
AGGGCCTGCAGTATCATGTCACCATCTAACAGATGCTCCTGGTAGAGATCTCTCGCCTTCATCTGACACTGGGATTGCTGCTGCTGCTGTTG
CTGTTCTTGGACTCCTTGTGATCTGAGGTCGTTTATTGTGCAATGCTGCTTGAATAAGCTCTTGGTTCGTAGATAGCTCACGGATAACCTGA
GAATGGACAGTTTATCCAGTTTACAGACAACGTTCTGGTGAAGGGAAGAAGGCTGGCTAAGGTGTCCAATTCGGCGTTGAGTCGCTCTCGATG
TCTTTTGTCTGGATTGGATTGGTTTGGCCTTCTTTTTCGGGACACGGACGTTTGTCTGTTCTCCTTCTCCTTCTAGTAGCGTACATAGTACC
AAGATGACTCATGTTGTCGCGGAGGACCGTCCATGTAACAGGACGGCCACTGGAACGGCATGCGCCGGGGCTGGATGTGCTTCACTGTACA
GCAGCAGGTGCAATCTCGATCTTCGAGTCAAATAAGTCGGTAAATCGACCAAATAATATAAACAAGTAGACAACGTCACCGTGTGGTCTCTT
GAAGAAGGAACCGTTCCATTTTTTCTGTAATCCTGGAGATATAACCAATCACTTCTCGGATTATACGATATCTTTTTCTTTTTTTAAGCTAAT
GTCATACAGTGCATCAATTTTTTCTCTGAAAATTACCTCATTTCTCGAAGTGGAGAAACATCAGTCGATACCTGAAAATCAACGCACTGATCT
TCCACTTTCAGTTTTCTCACATTATGAAGAGAAATGATTATTACTTCTGACGATTTTCCAAGCGTTTTTCTCATGTAATATCGGCTCATAGCGA
TTTGTATCTTTCGCGACAACATGATGCACATTCACGACAGCGCACCTCGGCAACCCGAAAAATTCACGATTTCCACAAATCCAGCATGCTTAT
CAATTTTATC

>comp125157_c2_seq1

CACGAAGGTAATAAATCACCAATTGGTTGGTTGTTTTCAGGTGAGAGGCGGGTTAGTTGAAAGCTGGCCGGCCGCGCACTCCAAGCATCAAT
CAAACACTAGCAGCTGGTGTCTGCTCTCACTAGAAAGCCTGGATTTCACTCTTGGACTATTTAATATTATCATCATTTAATTATTCTCTCGCA
CGTTTATTTCAATTTTTATATAGAAAATATTGAAACGTAATTTTTATTCTGTTTCTCCTCTCGTGAATGAATCAGCTCAGTCGATTCTCT
CATCATGAATTCAGTGTGAATTTATGACTCGGATCAGTGACAAGGAAAAAGAACAAAGTGTCTTAGAATGTCTACGACAGGCAAGTGG
TGTAAGTGATTACGTCGGTGGCTGCCTTCTATGTAAGTGTGCAATAGGTGTGAAATCATACAACCAGAAGATATGAATCATCTGGTACTATGT
ACGCCACTAAAAGGAGGCTAGGAACAGTAAAAGCGTTTCGAGTGCCACCAAAGACGGCCAAACGAAATCGAATCCAAGCAAGAGGCACAGAGA
GAGGCTCAACGCCAACTGGACACGCTAGCCAGCCTGCTCCCTTCGAGCAAAGCGTTCTCTCTAAACTCGACAACTGTCCATACTGAGGCTCA
CGCTCAGTTACTCCGGACTAAAAGTTACTTCCAGGCTGTATCAGACGGGAAAAAGGACATAGAGTTGAAGCCCGAGAAATGCTTTACCCCGAG

Supplemental Material

ATGTTAATGGAAGGTGATCTCATACTTCAAGCTCTGAACGGCTTTCTCATGATATTAACGTGCGACGGAGAAGTATTTTACACATCACACACAGT
GGAAACGTATCTAGGCTTCCATCAGAGCGACATCTGCCACCAAAGTGTCTACGAGCTTGTGCATTCGGAAGACAGGGAGGAGCTCCAGAGGCATC
TCATGTGGAATTCACACCTCACTTCAGACCGCAGCGATTGCACCAATTCACGAGGCTCTTTCTCCAGAAAATGGTCATCTTCTAGAAAAGGAATTC
ACTGTGAGGTTTCGGTGTATTATAGACAACACATCTGGGTTTCTGAGGCTAGACATCAGAGCAGAATCAAAATACTTCACGGGCAAAATCACA
AGACGGAAGAGCCTCCCTTGGCTCTGTTCCGCGTGTGACTCCTTTCCGGCCACCGTCGCTGGAGATGCCGCAAAAAGGATTTCGTTGTTCAAG
AGCAAGCACAACTCGACCTATCCCTGGTCTCCATGGACCAGAGGGGTAAGCAGTTACTAGGCTATGCAGACACGGACCTATCAAGCAAAGGAGG
TTACGATCTAGTCCATTCGATGATCTCGTTATGTAGCAAGCGCCATCAAGAATATTGAAGACGGGGCGTCTGGCCTCATAGCTTACCGAT
TATTGACCAAAGATGGAATAATCCAATGGCTCCAAACAAGCGCAAGATTGGTGTACAAGAACAGCAAACCAGATTTTCATCCTCTGCACCCACAG
CCACTCATGGAGGAAGAAGGGCGTGACCTTTGCGGCAAAAAGGACGATGGATTTCAAAGTTACGTACCTCGATTCTGGATTGGCCAGCATGAACG
ATCGCAGCAGCATTATGTCGGATAGTGACTTTATAACGAGGTGTCAAAGAAACAGCGGATACAAGTCACAAATCCGAGATTTGTTGTCAGCGTC
GTGCCGGAAGAGGAAGTCCCAGGTACCGATCCGTATATCGAGGAGACCACCTATAACAACATGTATCCATACACAACTCTGAAAATATAAAAC
GGCTTTACCTTACGTGGCGACTGCCATGCCCGAGAATATATTCACGGCAGCCATTGACAATCCAGGTTCCCTGTCCGCGGAGAACTTCTCCATT
ATCGTCACCTGGGCGACTACTACCCGACTACGCTACCATCCAGCGGCTCAGTACACGAGCAACGGATTCCCTGGACATGACGTCAGCTGCCGCA
GAACAACGTACATCCCGACTGCGTATCCGGATACCAGCCAGTTGGCTAAGGAGGAGAAATACTACCCTGCCAACTCGAGTCGACGAAGTTCAAC
GGCGACGCACAACAGAGACTAGTACAGCAAGTAGGAGTCGCTAACACGAAGCCTTCGTGCTGCGAATCCTACGTGACAGTGAGTTCCGATTCCAG
CGCTTCGTTGAGAGTTGCCCTCATGAGCGCTACCAGCGCGGGATCCAACCCCTGTTGACGACGTTGGACAGAGACCCGAAACTGGGTGAGCTCG
GTTGCCAATCTGAGCTTCACTTTGCTCCGGTTAGTTCGGAGTGCTAAACGTGATCAAACCAAAGACTTCACCCGCTGCCACTCATCTCAGG
GATTCCGCCCATCAACACCACGTGCCGAAAGACGTCAAAGTCCCCTACCCGCTCGGACTATGTTCCGGCTTTTCCAACGACACCAGGCAGAG
CGTACTAATGTGGGTAACCTCTTGGCAGACCCGACCGCAGGTCATCCAATGACCCGACCATATTCGACAGCAAATGTGCAAAATACAGCGGAC
ATTTGGACATGGGAGAAAATCTCCGCTTTAGAACCCTCGGCGAGTGGTGTGAGTCCATATGCAAGTGGAGCAGCAACGCGACCACTCCGAAA
TCTTCATCTCCCACAATAGTACCAGTCCAGCTAGTGGGAAAATGACCCGGCGTCAATTACAATGCTACGACCCTTACCATCCGGACATATC
TCACGAGGTGTATCTGGCGAGGAACGATCTTGGCTACAACCTGGCTGAAGCAAACCCGTAATCCCCTGCTGTGATATCGGAAGTCACAAAATA
CACTACTCAATCCAGACTGACATCATTCTACAAAGTCCATTGCAGAGTCTAATGTTCTACTCCGAAATAAAGTAAATTACTCCACAAAAGCTGCA
TGTGGAATAAAGTACAATTGTAATCAGAAGTCAGTCAGACAAAAGTGGGAACAGTGAATATTTGGAAGCAAATCTGGCGAAGAAAAGTCATG
ACTCAGAACAAAATAAAAAAAGAAAAGTCACATTCCATTGCTCAATTCTGGATGTATGTAATATAATGTATACATATTTTTGAAAGTACGAA
AACTATTTCAAAGCATACGGTAGGGAACCTTTAAAATACGCACTGTCCACAACTGTTAGAATACTGATAGAAAATAGTTTAGGGTTGCCAAC
GAAAGTCTGTACCGAAAGTAATACTGCATTTGCTACAACAGTCACGACACGATCACAATGAATGACTTTTCACACCAGATCGTTCTGCAATGTT
TGCTTTGCTGATATGAAGCCCTCGAAAGTTTCAGATAAAAAATTACATCGATACAGCACCTGACTTGAAGTTCACAAAATTTTTGTTTTACTCTAT
ACAGTATTTAATGCATAAAAAATTAATAGAAGGAATCGTTAAGTTCCCCACGAGCAATGAAGATCAAATACATTAGTTATGGTAGAAAAAAA
AAAGAAATCCTACCCTCTCTTTCTCTCAGATCACACACATACACACAATACAGATTAATAAAGTTATAATTTATCCTGAAAGGAATAAAAAAT
TGTATGCTGTTCTGTTCTCAACTACAATTTGAATTTCAAGCAGAAATATCCCTGTATCTCTGCTGCAACCGGAACTATCCACTTTGCATCATGA
AGAAAATGGAATCTGTGAAAACGAGTGGTGTGTTGTTGTGACATAATATTATGAACTATATTTTCAGGAGAGACAAAACAAGACTACTGATAATAT
ACAGTGATCCCGCTTATAACGTTTTCAA

>Locus_18705_Transcript_3/6_Confidence_0.762_Length_7353

AATTATGGGAGAAAGAGAGACCGCCGTGCAGGGGGAGGGCCAGCATATGACCCCCATTTTCGAACGAAAATGAAGAGAACTCCGCGCAGT
AGCATGTCAGTTTAGAGTTTGTGTCCACAGAATAGGAACTATGTGATGTGCGCCCTGAGAGGGGAATTTGTGTAATAAACGCCCTGGGAAAA
GCCCATCAATTTTTTTGTTGTGATTTTTTTGGATTGTTTCTTTTTCGCTCCAATTCACATTATCTATAAAGGATTAATTTGTGACACAAGGTG
ATGGTTTTGTTTATGCTGAGGCTATGGTGGCCGAGTGTCTGTAAGCGCGCGGCTAGGCTTGGCCTGGGGCCGCCCTTACTTTAATGTG
CAGCCCGTTGGAGAGCGCCCTGTGCATCATGAACCCTGAGTACTATGTACGCTACCAAAAGAAGAAGACGAAAATAGCAAAAGTGTTCGTG
TACCACCCAAAGAAGGCCAAACCAATCAAACCCAGCAAAAGACATCGTGAACGTCTAAACGCAGAGTTGGACACCTTGGCTAGTCTACTTCCC
TTTGACAAAACGTCCTGTCTAAGCTAGACAAACTCTCCATTCTCCGATTAAGTGTAGCTATCTCAGGACTAAAAGCTACTTCCAAGTGCCT
CTATCGGGATCAGGAGAGATCACAGGAGCATAGCATGAAATCTCGAGATCATATGTACAGTGACCTTCTACAGAAGGAGATCTATTCTTCAA
GCACTCAGCGGTTTCTTCTGATCTTAACGTGTGATGGAGAAGTGTCTACACATCGCACACCGTAGAACTTATTTGGGCTTTCATCAGTCTGA
TATTTGTCACCAAAGTGTATACGAGCTTGTCCATTCGAAGACAGAGAAGAACTGCAGAGGCATCTCATGTGGAATTTCCAGTTGTCCACGAAA
GAAGCAATCTTTCTCTGCAGGAAGCTCTACAGTCGGAAAACAGTCAACTTTTGAAAGGAATTTTACTGTCCGATTTCAGGTGCTTGGTGAACAAC

ACGTC TGGTTTCTTGAGGTTAGACATTCGAGGTGCGATCAAGGTGCTGCACGGCCAGAATCACAAAGACGGAAGACCCCTCTCTGCCCTCTTCGC
TGTGTGCACACCTTTTCGGTCTCTCTCCCTCTTGGAGATGCCCCACAAGGATACCATGTTCAAGAGCAAGCACAAACTTGACCTTTCTCTAGTCTC
TTTCGACCAAAGGGCGAAGCAAGTGTAGGTTACTCTGACACAGAACTCGCCGATAAGGGCTGTATGACCTAGTCCATTATGATGATCTCGCTT
ATGTAGCTAGTGGCACCAGAATTACTTAAAACAGGAGCCTCTGGTTTGATAGCATATAGGCTAGCAACAAAAACAGGATCCTGGCAATGGCT
TCAAACGAGTGCAAGACTTGTATATAAAAAATAGCAAACAGACTTCATTTCTTGCACACACAGACCTCTCATGGAAGAAGAAGGTCGAGATCTC
CTTGGAAAAAGAACAATGGACTTCAAAGTCACCTACTTAGATGGTGGCTTAAATTCCTTGAACGATCGGAGTGGTCTTCTCTGACAGTGACTT
TGTCTCCGCTGCCAGAGGACAAAGAGATACAAGTCACAAATTAGAGACATCATTACCACGTGCAGAAAAAGAAAAGCCAGTGAATGGACCA
CCTTCTACAAGCTCTATTGTAGGAGGCGCATTCGCTGATACATCGACTTATATGGACGATGCTGCTGCAGCTGCAGCTACTGCAGCCGTAGTAGC
GGCAGGCTATAACAATGTTTATCCTTATGGCACAACAGACAATATGAATGGTTTAAAGTGTCTTACGGAGCAACGGCTTCAATGCCACAGAACC
TATTCTCAGCTGCCATAGACAATACGAGGTTTATTTCTGCTGAGAACTTTCTTCACTATAGACCTATAGGAACTTATTATCCGGACTATGCAGTC
GCGCAGTATCCATCTAATGGTTTCTGGATGTGGCTGCCTCTGCAACGGCAGCCAGAACGTGTTTCGGCCCGAGCGCTTTTCAAACAGATTCTGC
TAGCCTTACAGTAAGCAACACCAAAGAAGATAAAGTGTACCAATGTCAGATTGAAGCGGCAAAGTTTGGCGTCAATTCACCATCTCCTCCTCAGC
AACACATAGTGAACAATAAATCCCACCATCGTCCAAGTGTGCGACAGTGTACCTTCTACATGAATGGGGTGGATCCCCTAAGGGGGTGTGTTG
GTGAATAATAATGGAGTGGGTGTCCAATCGTCTTCCATCCATCTTACCGGTTACGTTAGATCACCATGATATAAAAAGTTCGAGATCTCACATGTAC
AGGAAATTTGGCATTCCCTTTACATACCCTCACCACGAACAGCGAGTGTATCAACAATGGACTCAGTGGAACTAAATCCAAAAACACTTTTAAATG
GAGGTCCGAATAGCATCTTAATGACTATTTCGAGATCTACCACATTTGGCTAAAGAAGAAAAGATCTCATCACCTTACGACAGTAGGCATAATAT
GCTTGGTGGAAATAGCAGAGATGCACTCTCTGTGGACTGTCCAGCAACAAGCTCCTCCTTTTACCACCTTATGACATGAAATCGTCCCGATACG
GGGACGGTAAAGCAAGATGGTGAATAGGTAATGTGACCAGTTCACCTTGAAGTGGAGTGGGAGTGCAGCTGCACCCAAATCTGTTTACCTCAT
AACAGTACCAGTCCAGCCAGTGGTAAAGTGGCTCCCAAGTCACTTACAGCAGCAGCCATATACTCAGGACATGCGGCTAGAGCTTCCACTGGT
GAGGTCCGATATCATACGACCATCTCCGTCAAACAAGAAGGAAACCCACTCTTATCCATATCTGAAGTAACGAACACTTTATTGGACCATCACT
AGATGGCGCTCTCTACACCAGTGAACAGCACCAAAAAAGTACCTTCTACAGTCAATTTGGAACCTTTCATCATTTGTAAATTACATTGAAG
AATATTATTATAGCAATGAAAAGTAGTGAACGGATTGAAAATACTACTACTTTCTCGGTGAAGAAAATACTGAAAATCAGAAGACCGTATT
TCTAAAATAAAAAAAGTTAAACAAATGTAGTTAAGTTTTTAAACGTAGTAAACTAAGAAAATTCAGGCTTTATTGTGCCGTAGAATCTGATTAA
TTCTCTTAATTCAGTCAATTATAATAATTTAAAATTTTTATAAAATTTTGTGAGTTAAAAAATTTATATTGTGTCATGGTGAATTTCAAATTT
TTTTTACTAACAAATCTATCAATATATTGTGGTATAAATTAGGTATAAATCATAGAGAACAGAGAAGAATCATAGAGAATCATAGAGTATAAAA
ACCAGTTCGGAACAAAGCCAGCTTAAATTTAAAATTTGTCTTTCTTAATTTATATAAAAACTAAAATGTTTTTCTATAAAAAATAGCTTAAA
GCTTTTTTATCCCTTTTAAAAAATTTGCCAAAATTAATTTAAAATGAAAAAATCATCGTTTTTGTATAAGTTTTAAAAATGTATTGAGCAAG
CATTTAAAGGTACCGTTTGATTCAATACTTTTAACTTATGATGAAACCGTTGTAGCTTATAATGAGAAAGGATCTTTGTGCAATGTCTCTAGTT
TTTATGTTTTCTTCTAACAAACGAAATATGGTAAAATGTGAAAATGAATTTGATAAAAAATATTTTTTTTTTGTTCTTTAAAAAAAAGCTCAA
ATTAACCTGAAGTGTGGTGTGTTGTGTAACCTTTTAAATATGTTTTGAGCAAGCTATACATCAGCCGTTTCATTCAATACTCTTAAAAATAA
ATTAGCTTTTCTTCTTAAAGCCATAAAAAATTAATAAATTTTAACTTTTAAATTTAGCCTTTCATAACAAAATATTTTAAGACACTATATC
CATATTTTATAATGAATGAAGATATTACCCTAATCATTACAATGTGTCTATTAATAATAATAGTATATAAAAAACTTTTACAAGTAAAAGCT
TTATATAGATATAAAATATTAAGTAAGGGCAAAATATATGTATACATTGAATAAAAAATCAAATGTTTTTTTTTAAATTTGCAATTTTTTTAAAACCG
TTGTAACCGATAATGAAAAGGATCGTTGGGCAAGGTCTTTCTTTTCAATATTTTCTTCAATGGAGAAAATGTGGTAAAATATTTAAAATGAA
ATATATAAAAAATGGATTTGATAAGAATCAATGAAATCAAAGAAAATATTGAGCAATAGCAACTGTAAAAATTTTTTAAAAAGAAGCGAAAAAC
ATTATTATTGTACAGAGAATGTGTTCAAATTTGAACTATAGCATTTCATGACACAAAACCAGTTTTTTTTGTCTGAATTTCTCACCTTCTCTTT
TTTTCTATACATATCCTATATGTTATGTTATCGAGTTAGCATTCTATCCTCTATCACACAGGCACATTTTAAAAATCGCTTAGAACTTAACTTGCT
TATAAATTCACAAATTTTGA AAAAATTTGCAAAAATTTAAAATTTGAAAAAATATTTTTTCCCAAATATATTTAACTTATACAATAACAT
CGCCATTCATCGGTTTATCAAAGACTTGTACAAAAGAATTTTTGAAACATTACATTTTAACTATTTTATCTATTTTGTATTTTTTGTATTAT
ATACATACAGGTAGGTGAAAAATTAATAATAAATTAATCTTTGTTTGACATTTGAATATTGAATATTAATCAGAGCATTGGAATCAGAT
TTATTAACAGTTTAACTACTAAACAAAACAATGTCTAGGGCAAAAATTTGTACCTTATAAATTTGAAAAATGACTCTTAACTTAGGCTTTCCAG
CGCATTAAAAATAAACCATTATAGATTAAGATGAAAAATCGATATAACTGAAAAATAAATTTGAAAAATATAAATAGTGTCTTACTGTGTTGAA
AATGAATCAATGTAATCTAAATTTGTTGTGTAAGATGAAATTTATATGTGAGTACACAAAATATAAGTTTCTGATATTACTCAAATAGATACATT
TTGTAACCCGTTATTATAAGTTTTTTGAACATTTTCGTAGAATTTCAATATGGTTTTCTAATGTATTATAAATTTATTTAAAATAAGGTTTTT
TCGTGATGTTAGTTATTTCTTAGATTAATATGCAATTTCTATTTGCTTGCAATTTTAAAAATGANNTAATGATGAAAATAATGAATGAATGATGA
AATAATTTAAAATAATTTTAAATAATGATTCATTCATAATGAATGCATTTATTTAAAATAATCAGTCATGATTGATAAAAATTAATGAAAATA

Supplemental Material

AATAAAAAATTTAACTAAGCTCTTAAACCTACGGTCTACAAGTAGTGTATTATTAATTTGAAAAATCAATCATCTGGTAAATATATGACTCTGA
GAACTTAAGACTTATTTTAGATAATCTCATCCTTTAATGTTAGATTATTTTATCATCTTTTCAGTTAATTTTCATCATTACAAGCCAGGTTTT
TTATCAAAAAGTTTATTATTTATCAAAAATTAATAATCATATGCTAAGATATTCAAATAACTTGATCCACCAATGTTTCAGTTGAGTTAATTATT
TATTATTACAGGTAATTAGTTTCTAATTCAAAGTTAATCATCACATTAATAAATTAGATCCCTAAAAATTTGTGCCTCCTCTGAGTTATCAATTCT
TTAATAAECTTAATCAGTTCTTTTAATGGCTTCATTCTTTTAATTAATTGCATAACAGTATAATATTAATACTTCAAAAATTAATATTAATATTG
GAAAAATAATAATAAATTGTCTACGAAAAAGACGCGACAAACATGTACTTCGATTAATATCGATTAATCGTATATATCGATTAATTAATCTTGA
GATTCAATTTATAAGGAACCTCAGCTTATTGTTTTAATACTACAAGTAATGAGCTATTGTATTCAAAAATCAATCATTAGATGTTAAAATTAGACGT
GGAATACCTTGAAAAATCGATTCATGAGATTGAATTCATAAACATGTACTCTTTTGTAAAGAAAAAGCATATACTTTTTTGGTTTAGTTAA
ATCCGAGATTCCTTCATGGTAAGTTCAACTAATTATTTAGTATTGAATAATGTGACATAGGTTTTAAAAATTAATCATCATATGCTAGAATTA
GACGTGTTAAACTTAGATGCTGAAAAATTAATCAATATGAATTTTTTTCATTGTTATAAACATCATATTATTCGGTCAAAACTAATTGTAAT
AATTTTTAAAAAATCATAACTGCTTTGAATAGGAGACGAAAAAATTTGACTTCTCTGTTATTTATCGATTAAAGATGATTTTGATAAAT
TTAATGAAGATTTATTGATTAACAATTAATTGTTAGCTTTCTAAGCATTCAAGTAATATATCAGATATAATCATCATATGTTTGAATAAGACG
CGTAAAAATAAGGCATAATAATCGATTCAATTAATTTATTCATTTTCATGCATGGAATTATATCTGATGCTTTTATAAGAACAACCTTCTC
TAATTATTTTATTCATTTTCAGGATACGTGTTATTAATATTAATATTCAAATATTTGGAATAAGACATAAAAAAGACTTTTCTGATGCTAATA
GTAATTTGTATAAAAAATCTATAATGCTAGAGATTAATTTAAAAAGAATCTTAGATTTAAAAAGTTAAAAAATTTCAACATTCATCACTGCAAA
TAAATTTGTAATAAGAAAAAACAATAAATTAACACGTAATTAAGCTTTTTATATTATCAGGTATTGTCCAAGTAAAAATTCACAGAGACT
AATCAAGTCGATTGTTACAACGATTGGCATCCGACTTGGCGAGATTTTCGTTTTCTGACAGATGAACCTGCTTTATGAATTTGCACATTTTTG
AAGTTAATTTTACCAAAGTAGCTCCATCTATCGGAGAAGTATAAATTTGTGAGAAAGTCGAGTCGGCTGCTTCACAAATACAGTCTCTGGTGT
GACATAATGCATTTGTAGACTTGTAAAAATTTGTTGTTGAATAGTGAAGTTTTTGTGAATATATATTTTCTGCTGAGCTCTAAAAATTTTC
ATTTTCATGAGATAAATATTTGAGTAAAGACTGGTGAATATTCATTTGTAAATATAAAAAATTTGAAAGTTTCAACACTTTGCTAATGCAAAACTAG
TGTGAAATTTGCACTTATTTTCATCGATTTAATTTATTTTGTGTCATATAAATAAAAAATTTGCAAATTTGTTGTAAGAAAAA

>Locus_20228_Transcript_1/2_Confidence_0.600_Length_2760

GGATAGTATGAGATAGATAAATATGAATCATTTAGGAACATATGTATGCAACAAAAAGAAGGAGACGGAACAGTAAAAGTGTCCGTGCCCTCCA
AAAGATGGCCAACTAAATCAATCCTAGTAAAAGGCATCGAGAAAGATTGAATGCAGAAATGGATACACTCGCAAACCTTCTTCCATTTGAGC
AGAGTGTCTTTCAAAATTAGACAAGTTATCTATACTTCGTCTCAGTGTAGTTATCTCAGAAACAAAAAGTTATTTTCAAGTTGCACTTCCAAG
GGAAAAAGACTCAACAGATTTGAAACCCAGAGATTTATTTATTTCTGAAATGCTTATGGAAGGTGATCTGATATTGCAGGCATTAATGGATT
CTAATGATCCTCACCTGTGATGGCGAAGTTTTTACACATCTCATAAGTGGAGACATATCTGGGATTTTCATCAGAGTGATATATGTCATCAGAG
TGTTTATGAACCTGTGCATTCAGAAGATAGAGAAGAAGCTTCAAAGACATTTATTTGGAATTTCAATTACCTCCAGAAAGAAGACATCTCAAC
TTACAACAACACTTGAAACTGTCAATATCTGGAGCGAAATTTACCCTGTTGTTTTGCTGCTTTTTGGACAATACATCCGGATTTTTGAGATT
AGACATTCGAGGGCGAATCAAAATCTTTCATGGGCAAAATCAGAAAGAACCACCTTTAGCATTGTTTGTCTGTCACACCCGTTTGGTCTCCAT
CCTTACTTGAATGCCTCAAAAAGATTTCCATGTTTAAAAAGTAAACATAAGTTGGATTTGTCTCTAGTCTCAATGGACACAAGGGGAAATGTTG
CCAGTGCATCAGGAATTTGCTGAAAACAGGTGCCTCGGGACTTATAGCTTACAGACTTCTTCAAAGGATCTCAAAATATCAATGGTTACAGACT
AGTGCTAGACTAGTTTACAAAAACAGCAAACAGATTTTATACTTTGCACCACAGACCTTTAATGGAAAGAAGAAGGAAGGATTTGCTTGGTA
AGCGTACAATGGATTTCAAAGTAACATATTTAGACAGCGGTTTGGCAACAATGAATGATCGCAGTGGAAATGTTATCTGATAGTGACTTTATTAT
GCGTTGTCAACGTAACAGACGGTATAAATCTCAAATACGAGACTTGTGTCATGTCGTAAGAGAAAAGTACCGGATACCTATATCGAAGATACT
CCTTATAAATAATATGTATACGTACTTCAACGAAAAATGGACTTTACCCTATGCAACGGTACCTCAAAATATATTCACAGCTGCAATAGATAA
TAGTAGATTCTTATCTGCTGAAAATTTTTTACATTATAGACATCTGGGTGGATATTATCTGATTACAGTCAATATGGCTAGGTTTTCTACCTAC
CCAGATGGTTCCAGTTTGTAAAGGAGAAAAATACTATCCTTGCCAGTTAAATGGAGAACAAAATAATAAACCATCTTGTGTAAGATGTAC
GCTTAATGACAAGCGCTGATTTAAACTGTTTTCTTCGACAGATTGATTAATGTTTTAAACCACATCTCAGGGATATTTCTAAGGATAATAA
ATCANNNNNNNNNNNNNNNNNNNAAATGCATGTAATGGAGTGCCGCTACAACGCCAAAATCTTCTTCCCAAGAAATAGTACAAGTCCCTG
CTAGTGGAAAAGTAGCTCCTTATACGACTACTGCCTATCATCCGGATCCAGAATGGCATGCAGGCAGCCTGGCTACATGGAGACCTCATGTCAGT
GAGGAGCGAAGTTGGACTCCAGAGAGTCACTCGTAAACCTTTACTCTCATATCAGAAGTAACAAACACACTTCTTAATCATGACTAAGACT
TTATAACTCCTTAAAGAAGCATCTTAAATAGAGTTGGACACTTCCAAGGAGTAACTGTTGCTCGACACTACATTTCTTATTACAGGAGCCTTC
AGTTAATTAACCCTCAGTAACGCTTCTTGTGCAAGTAACAAATACCAGCCAAATATAACTAAGACTTAGTAAATCCTTTCCGAAATACCT
TAAGGAAATAGGACATCATACCACACGTAACAAACACCTTACTCAATCTTAACTGATACTAAGCAAATTCATTAGGAAGAACCTTAAGGGCTC

TATAGACAGTCAGCGGTAGCCATTTAATCTCCCAAAAAAAGTAACAAATACCTCACTCATTCTTAACCTAAGGCTGAGTATATTTCTCAAGGAGA
 ACCTTAAGGATATCGGGATCTCATAGACATTGAGCTCTCCCTATAAGAAAACAAAAGACTCAATCCTAAGCTGGACTAAGCAAATTTCTTTGACCT
 CGATATTGGTCTCTATAGACAATCATCACTTTATCTCTCAAATTTGGAAGTAACAACCTTGACTAAGAAAAACCTTCACCTTCATTAGAAAAAT
 GATGTCAGAACTTTTAAATTTGATTGCAAAGAGTTAATCAAAGTAAATTTAAAACGCTGCATTACAGTCTAATAATTTTGAATAATAAACGT
 ATTTTTTAAATTTGGCTTGTGTATATTTGACAATTCATAAAATTACTTTTCTTAATAAGATTTATCAGTTGTCTCATTATTTTCTTTTCGTAAT
 TGCCAGACACCCAAAAAAAATTCAAAACCTTATTCGCAATAACCACGTTTCATAAATTCCTAAGTTTAAATAAAAAATTTAAAAAATGAATG
 AAAAAAAAAGAAAAAGAAAG

>comp121950_c6_seq2_-_ORF_1_(frame_2)_translation

MLYSETTSSPGACRSGRPGYMDGPRGTTMSHLGTMYATKRRRRNSKTVRVPQKEGQTKSNPSKRHRERLNAELDTLASLLPFDQNVLSKLDKLS
 ILRLSVSYLRTKSYFQAALHNKPTSDRQGVQEQQQQQQSQCMKARDLYQEHLDDGDMILQALNGFLLILTCEGEVFTSHTVETYLGFHQSDI
 CHQSVYELVHSEDREELQRHLMWNAQLPADRSALTLRESLQSDNSQVLERNFVFRCLLDNTSGFLRLDIRGRIKVLHGQNHKTEDPPLALFAL
 CTPFGPPSLEMPQKAMFVSKHKLDSLVSVDQRAKQVLYGSDAELASKGGYDLVHYDDLAYVASAHQELKKTGASGLIAYRLVAKNGSWQWL
 QTSARLVYKNSKPDFILCTHRPLMEEEGRDLLGKRTMDFKVTYLDGGLNSLADRSGLLSDSDFVLRQCQRTRRYKQIRDMISTCRKRKSPSSSSGGT
 SPAMMGALADTSYMEDAAVAAAATAAAAVAGYNNVYPIYASAAENVNGIAAPYASASMQQNLFTAIDNTRFLSAENAFHYRPIGTYTYDYA
 VAQYPSNGFLDVAATAAASARTCFASAGFTAESNATSPGKDEGLYQCQMEAQKYNNGTESPPKSSNSCCDPYVATNGDVMRNLPPQVVTANSQTI
 LPSSIDRDLKFRLETCSSNNGGINFTFANDCLTTNAKTLKPGMNGTTTNGVLMNSRDLTFLAKEDKAIHQHHHHHQQTLDYQTDQRQSVVMWN
 NGIITNDASNAGERRPFGVSPVSNRSCPTTPTSGSPGNNTYDLKLSGRYSEQSSKSGHQIDQVSGGNMMGSPCKWGSAAATPKSASPHNSTSP
 ASGKVAPQVTFSSTPFAQDISFEVGEHRSHFC*

>comp125157_c2_seq1_-_ORF_1_(frame_1)_translation

MPFYVLCNRCEIHPEDMNLHGTMYATKRRRRNSKSVRVPKDGQTKSNPSKRHRERLNAELDTLASLLPFEQSVLSKLDKLSILRLSVSYLRTKSY
 FQAVIRREKDIELKPREMLYPEMLMEGLDILQALNGFLMILTCDGEVFTSHTVETYLGFHQSDICHQSVYELVHSEDREELQRHLMWNSHLTSDR
 SDCTIHEALSPENHLLERNFTVFRCLLDNTSGFLRLDIRGRIKILHGQNHKTEEPPLALFAVCTPFGPPSLEMPQKDSLFSKHKLDLVSMD
 QRGKQLLYADTDLSSKGGYDLVHFDDLAYVASAHQELKKTGASGLIAYRLTKDGKQWLQTSARLVYKNSKPDFILCTHRPLMEEEGRDLLGK
 RTMDFKVTYLDGGLNSLNDRSSIMSDSDFITRCQRNRYKQIRDLASCRKRKVPKGTDPYIEETTYNNMYPYTPPENINGLSPYVATAMPQNI
 TAAIDNSRFLSAENFLHYRHLGSIYPDYAHHHPAAQYTSNGFLDMTSAARTTYIPTAYPDTSQLAKEEKYPCQLESTKFNQDAQQRLVQVGV
 NTKPSCCESYVTVSSDSSASLRGLMSATSAGIQPLLTLLDRPKLGEGLGCHNLSFTLPPVSSECLNVKPKTSPPLTHLRDSPHQHHVPKDVKGP
 ASDYVSAFNSDTRQSVLMWGNLSRPTAGHPMTGTIFDSKCKYSGHLDMEKESPPLEPLGSGVSSICKWSSNATTPKSSSPHNSTSPASGKIAPAV
 NYNATTTYHPDISHEVYLARNDLGLQPWLKQTNPLLSISEVTNTLLNPD*

>Locus_18705_Transcript_3/6_Confidence_0.762_Length_7353_-_ORF_1_(frame_2)_-_ORF_1_(frame_2)_translation

MCSPLERPPLCIMNHLSTMYATKRRRRNSKSVRVPKDGQTKSNPSKRHRERLNAELDTLASLLPFDQNVLSKLDKLSILRLSVSYLRTKSYFQAAL
 YRDQERSQEHSMKSRDHMYSDLPTEGDLILQALSGFLLILTCDGEVFTSHTVETYLGFHQSDICHQSVYELVHSEDREELQRHLMWNSQLSPERS
 NLSLQEQALQSENSQLLERNFTVFRCLLDNTSGFLRLDIRGRIKVLHGQNHKTEDPPLALFAVCTPFGPPSLEMPHKDTMFKSKHKLDLVSF
 QRAKQVLYGSDTEADKGCYDLVHYDDLAYVASAHQELKKTGASGLIAYRLATKTGSWQWLQTSARLVYKNSKPDFILCTHRPLMEEEGRDLLGK
 RTMDFKVTYLDGGLNSLNDRSSIMSDSDFVLRQCQRTRRYKQIRDIITCRKRKSPVNGPPSTSSIVGGALPDTSTYMDAAAAAATAAVVAAGYN
 NVYYPYGTDDNMNGLSASYGATASMPQNLFSAAIDNTRFISAENFLHYRPIGTYTYDYAVAQYPSNGFLDVAASATAARTCFGPSAFQTSASLQLS
 NTKEDKLYQCQIEAAKFGVNSPSPQQHIVNNKSPSSNCCDSVPSYMGVDPLRGCLVNNNGVGVQSSSILPVTLDHHDIKVRDLTCTGNLAFPL
 HTLTTNSECINNGLSGTSKNTFNGGPNLSILMTIRDLPHPKEEKISSPYDSRHNMLAWNSRDALSDCPATSSSFTTYDMKSSRYGDGKQDGEI
 GNVTSPPCKWGSAAAPKSASPHNSTSPASGKVAPQVTSSTPYTQDMRLELPLVRSDIIRPSSVKQEGNPLLSISEVTNTLLDHH*

>Locus_20228_Transcript_1/2_Confidence_0.600_Length_2760_-_ORF_1_(frame_2)_translation

MNHLGTMYATKRRRRNSKSVRVPKDGQTKSNPSKRHRERLNAELDTLANLLPFEQSVLSKLDKLSILRLSVSYLRTKSYFQVALRREKSDTLK
 RDLLYSEMLMEGLDILQALNGFLMILTCDGEVFTSHTVETYLGFHQSDICHQSVYELVHSEDREELQRHLLWNSQLPPERTDLNLQQTLENCQYL

ERNFTVFRCLLDNTSGFLRLDIRGRIKILHGQNKQEPPLALFAVCTPFPPSLEMPQKDSMFKSKHKLDLSLVSMDTRGKCCQCTSGIAENRCLG
TYSLQTSCKGSQISMVTD*

11.5.5 tracheless

>comp119399_c0_seq1

AATAAACACAGTATGGAAAAAATAATAATAAAAAAATAAATAATAATTCGGTTACAGCACGCCCTCGTTATATACGGACAACCATA
ACATGATATCATAAATTCGAGGATTTTCTTACCTCACGCCGGTCGAAAAACAAGGATAAAATCTTCAATTCGATATCTGATATTGAATAAAAT
GTGATCTGAAATTTGATCTGATTAATCAATTACCATGGATTATATTCGATGATATTATCAAGATAACTGACTGATAACAGTGGGTTTATATAGGT
ATCGAATATCGTCGGGGAAAAATTCGGTGACTCTGGATAATGAAAGTGTGACGATGGTGATGGAATGAATGAATACGCCGTTGCGCGTGTGA
CTGTGCTGAAATTTCCAGTTGACCGTGCCGACGTCGTTTCTCGTGCGGATTGACAGGACAAGAGTGGCGATTTTCAGTGCCTAAAAGTGGATAT
TGGTACACAAATGTGCTTGACGGTGAACGTGGTATTGTTAAAAAAGAAAAAGAAAGAGTGCATGCAGCTGATACTTATGGATCGTT
ACCAGTGGATCTACATATACCGCAGCAAGGCTATCCAGCTTATTACAGGAAAGGGCATGACGACAGTCCGGGTCCACCGGTAGTTCAAACACGAA
TTTTGAAATGCGTAAAGAAAAGTCTCGGGACGCGCAAGATCTCGGGAGGAAAAGAGAATTATGAGTCTACGAGTTGGCGAAGATGCTACC
ACTGCCAGGGGCGATTACGACGACGCTTGACAAGGCTTCGATCATTGACTGACCAATAAGCTTCTCAAGCTGAAGGACTTTGTGCGACATGGAG
ATCTCCGTTGGAGGAGATGGCCATCACTGAAAGGAGGGAGCTTACGGTACGATCGGTGATGATTGATCTTTTCGAGCAGCATCAAGCAAT
CAGCTATTACAGTCACTTGACGGATTGCGTTGCACTTGGAGCTGACGGAAGATTCTGTACGTATCGGAAACGGTGTCTATCTACCTTGGTCT
ATCGCAGGTCGAAATGACTGGTAGTAGCGTTTTCGATTATGTCATCAGCAGGACCAATCGGAATTTGGCCGATATGATTGGCATTAACTGTGTC
ATTCGCCATCCGCGACGCTCTCGGCATCCGTGGCATCGGACGAAGGGTCTCTTCGACACCGGGGACTTCCACGCCGACGGGATTCGATAGGCTG
CTCCCTTGTGACGCTCAGCAATCAAAGCAACGGCCAAACCTTTCAGAGGTCATTTCTGCATTCCGATGAAGTCCACTTTCAAAAGAGAGGGTGC
CATTTCAGTCACTCCGGTTACAGAGTTGTATTGATACTTCCATCTTAGACCACAGTATACATTTTCATCATCAAAGAAATCTCCTCAACGAT
AATGGGCTCGTTGGTGTAGCGATTGCACTGCCTCCACCGTGAACAGAGCTGAGGTTAGAACAAGACATGTTTGTACGAGGCTGACGTTCCG
ATTTACAGATATCACACTGTGAACCGAAGTTGCAGATCTCTCGACTATACAGCTGACGAAATAACTGGCCGCAATATGTACAGTCTTTGCCAC
GCACAAGATGTGAGAATCTTCGAAAGTGCATATGGATCTCATTCAAAGGACAAGTGTCTGTTATTACAGGTTAATGAACAAGACCG
GTGGTTACAGTGGTTCCAGAGCTGTGCTACAGTCAATGCAACTCGAAGAACGGGACGAAGAACGATCATATGTGAAATTTATGATTGAG
TAACGTTGAATACGAACACTGCATCATGGACTACAGCCAATTACCATCTCGAGGGAATTGAAAAGTGAAGGAGCCTAGTAGTTCCGGAAGGGGT
ACGACACTGATAAGGAAGATGGAAGAGCAGATAGTGGTTCCGCTAGGGATATCCATCCAGATGATCTCCAGCATCTAGACCAGAGACTGATG
AAGGCAAGGAATCCGAACATTACGATCAAGTACTGAGATCGACGCCAATAAAGGGCAGCAACACATTCAAACCTCGGCTTCGCCAACATGTGA
GGATTCAGAGTCAACGAATATCCAGTGTCAATTACTTTGATCGGAGTGAATGAACCGCCTTGGACCAACAAATTCGATGGATCAACAGAAC
GAAGATCAGCTTACCTTCTGTGGTCCACCACGAGAAACGAACAAAGCCGACTCACTGTCAAGTTTCGGCCGAAAGCGAGTCAAGAGTCC
GAGATCCCGTCCAGAACACTCATCTAGTCTCGAAGTCTTCTTGGATAACCGCCATTACATACAGAGCGAGTCCGACCGGACCAAGGACG
CCATACTGACTCCGTCGGTACATGTGTCAGCTCAGAGTCAATTTGAAACAACACCGAAGCCAGTTCACCCGACCGCTCGGACAAAACTCA
GTTTCCAGGCCCTGGACCCGTTCCCTGTGTCCATCACGAGTCCGGTCCGTTTCAAGGTGATGGGAACTCGAACAAGGAGACAGCTAG
TGCTAGCATGTCCGTCAGAGAACTGGAGGATGCCATGAACAGACATTTGCCCGCAGCAGTCACTTCAAACCTGCAGTTCTCCGGACTCTCTTA
TGCTCAAGACGTACGCTTCGGGTCCCGCTGTGTACAGCATCTCACTCGGCTGCTCCAGGCCCATCCAATGGATAGGTGGGACACCGGACTC
TACCAGCTACGACCTTCTACGGCAGATCTATGTGAGCCGCAATCAGTGTGATCCGATCCGGTGCACACGCCAGTAAACACGGCTGCTACGGGGAT
GTCCAAGGAATTTACCGACTCTCCGGTGGCCAGAATGAACCGGTTACGCGGACACGGCACAAATTTATGTTACAGTCCAAAATATCGCCAGA
CGGTGGTATGGCGTCAATCCGGTTACGCTCGGCGCCGGCTCAGTGGTCAAGTGGCGGGTACGCGGATTCGTGCCACGGGATGACCCCGCC
GAGTCCGTTCTCCGAGGACAAATCAACACGGAGACGTCAGCTTCAATGAGGCGGGCCGTTGGCCGAGCTGCCCAATGCCTCACATGA
GGCATTAGCTACAGCGGACTCCGGCCACTGACGACCTGCCCTGAAACCGGATTTGTTGTTACATCCAGGTAATCTGGACCGGCTGCCTACG
GTATCCCAACAGACACTCTCGCCGATCAGCATCAGTTGTATCATCATCCGAGCAGCTTTCATCTATACCATCTACACCAGGAGGCTCCAAAC
CACCATTACATGCTGTAATGGCACGGCTTCGTGGTTTTGCAATCTCATTTCGTGATCACTCGGTTGAACTTGATTGAATGTTGTACATTTCCG
AGGAGAGAGGAGTCCCTTTTGTCCCGAATCTTCTCATAGCTTCTCAAAGAGGGATTGGTTCAATTCGCGGACCGCGTTCTCTCGAATTTCCG
TAAGTGATGATAGGATATAGAAAAAATGAAAAGAAATATATGTTTATATTTGTTGTGATGCACATCGCTTTCAGAAATAAATATC
ACGGTGTGATCATTACGCGCATCTGTGTTTATAAAGCAACCTGAAACGCAAGTAAAGTATCAACTTCATTGTTTACGGAGAGCGATTTCGTT
TGCGTAAAATATAATGCATATGTTTATGAAAATTTCTACATGAAATATAACACCTACGTTTTAATTTTATTTTGTATTACAAAATTTGT
TGATTTCTCTCTTCTCTGTTTTGCTGTCCGTTTTGAGATATGTGAAAGTCTAAAACAATGTACAAAATACAATGTTAGCTATTATTCA
GAAATGGCAAAAAAATACCGACGAAAATGTGATTTGATGCATTAATAATTTAATGGGAAAAAACAATTTATTATTCGACTGACTCTTCT
CTGAATAGTGTTTTACACAAATTTTAAAGTGCATCATTCAAAGTTTGAATTTTATAGATGACTTTCACA

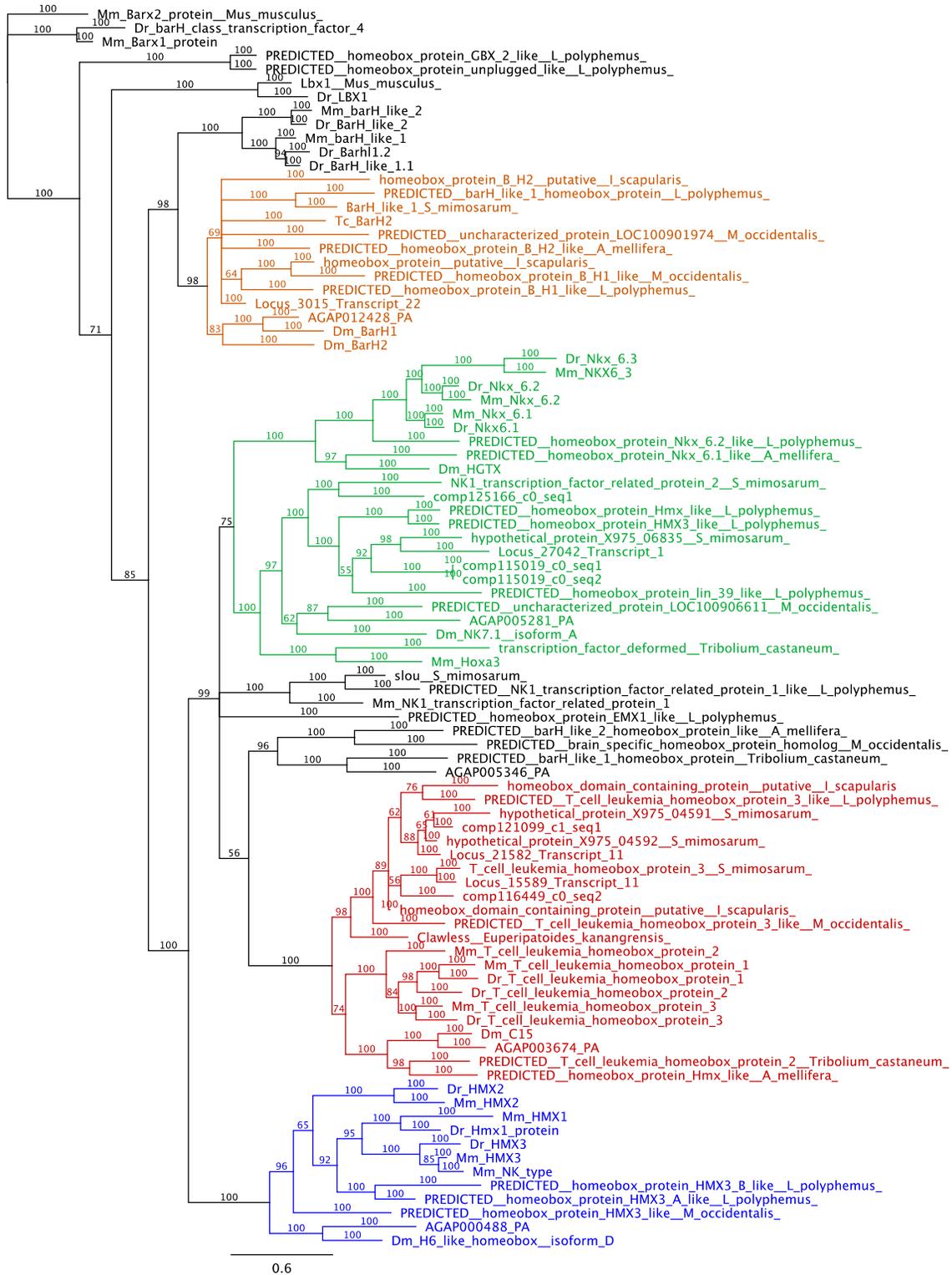
>comp128285_c1_seq10

AAATCTCAAAAAAATAAAAAACCGAAAAACAACAGTTTTTTTGTAAGAACGGAAATCAAAAAATTAATGTTACGAATTTAGTTTTTTGCTAAA
AAGTCATATAATATTACTTTATACTTTCTGCTACTCGGAAATAAATACGGTTTTTTTTTCAGCGTCAATTGACAATACTTTTGTGACCAGTAGAAT
CCAATGATGAGATCATTGTGTGAGGATTTGATTTTATGTTTTTCTCGCCATGATTTCTGGGATTAATAATGTTCACTACATGTGCGAAGCGAGA
AAACTGACACGAGAACGATGGTATTACTACGACTGTTGTTAATTAAGTATCATTGGTGGATGAACCGGGCAGTGCCTTTTAAATGCATGAAAC
CACCCAGCCAGCGCTCGAGGAGTCTCAAGAGAAGATGGGCTTCGTAAGAAAGCGGATCTGGATGAAATTTTCTACAATTTACAGGATAGGGTATG
CAAAACCAAGTCCCGGGGCCCCATGATACTTTTGAGACGTCCTTGAACCAAGATGTATGATACCACAAATAAGAAACATTTTATTGGACC
AGTACAACAAGGTCAAGCTAGTAAGGCCAGACTAGCAAAATCTGATTTGGCGTATGGGATATCGCTGAAACTAACATTTCTTTCTGTTCAAGTGA
GTCCCTGTGCTCCATTTCTCTGAAGCATTATGACTTTTTTGAATAATGATAGTGGCGGTACGCCCTCACTCATTTCGAAACTCATAACGAG
TACCGTAGAATCATTCCGTGAGTTCGAAGCAGAAGCTGAGACTTTTCAATCCCGTTAAAATGCGAAGAGGGTGAATTTGGCGAGTTCAAGAAC
CAACGAGTACTGACCTCATCCAGTAAATTTGGCTTTCAGTGGTGAGCAGAGGCACTTTTATAATTAACAAAGCTCTCAGGGGTGTGCTGTTGTT
TTGTTGTTTTCAATTTGGCTGATTTAATGGCGCGTAAAAAGGCTGTGTTGTTTTAATGGTAGAAGAGTTTCGAGCTTGCAGTAAAGCAGTTGT
GTGTTGTTTTTCAACACAACACTGACAAGTTACGACGAAGTTAAAAAAGAGTACGGGTTGTGTTTTTGGACTTGGCTGACGATTCAGAGCGAA
GAAGCTGTGTTGCTGTTTTTTTGTAGTGGGCAACGGTTTTTCGAGTACCAGTAAAAAAGTGTGCATGTTGTTTTTCACTGTGCTGACGACTTC
CGAGCAGGAGTGCAGACTTTTTATTGTTGTCAAAAATACGAAAAAGTACTGCGTCCGAGGAGCTTAGACACTAGATTTCTCCCATCCGGTGAAT

Supplemental Material

>Locus_17705_Transcript_4/7_Confidence_0.682_Length_6507

ATTTCCCCCTCTTTTTTTTACATCCACCAGTCCGAGTTGCGTCTTCCCGGGCAACGTCACTGACGATCTAGGCCTAACGGGGACGAAAA
AGATTTCCAATCGAAAGCCACTAGTCACTTCTCGGCCAACGAAGCAGAGAGGTGTCACGGCAAAGTCCGACTAGAAGTCCGGCGCGTTA
CGCACGCTTTCGCTGTTTTACGCATTTGTTCTTTGTTTACATTAAGGCGAGTAGGGGACATTGCCTGACGTATGTTATTTGGTCAAGAAATGA
TTGTGATATATTGACGAGAAGACTTATGATTATTTTTTTTATTGATTGATAGACTAGAGATTTAAATCACATCTGATTTTGATTTAACTATCGG
CTGTACCCGCTGTGCCATAAATAAAGTTGCTGAGCTAGTTTTTGTGTGGATAATTGACGTAAATCAATGGATTTTGAAATAGGTAGTTTACG
TAAACAATTGGAGACGTTTTTATCCACGGTAAACAAGTGATTAAGGAAAAAATTAATACTACAAAAACATTTAATTCTAATAATGACGTAAC
AAATATTTAAATGAGTGAATAGTAAAGAGGAGGTGATTTTTTTTTTACCCCGAAAAACAAAGTGGATTTTCATGTTTGTATGATTTTGAA
TAACAGTTTTTGCACACGATTTTGGTGAATTTCTCATGACAGAAAGTGAAGTTAAAGAACACTTTTTGGAGGTTATTCCTTTTTGTTCTAATAGT
GTGTTTTTTCGTTTTGGGTTATGTCTTAAAAACAATTGATCTGATGATGATGCAACCTGAAGGATATGGGTCCATACCCATGGACCTGCACATTT
ACAGCAAGGATACCATCATATTACAGAAAAGGGCATGATGATACACTGGACCACCAATAGTACAAACGAGAATCCTTGAAATGCGAAAAAGAG
AAGTCCAGAGATGCTGCTCGTTCAAGGCGAGGCAAAGAAAACAGAAATTTACGAACTCGCCAAATGTTGCTTTACCGGGAGCCATCACAA
TCAGTTAGATAAAGACTCCATCATTAGACTACCATCAGTTTCTTAAAGCTCAAGGAATTCACGGCACATGGCGACTCTCTCTGGCGAAAAAGAG
GTCCATCTCTCAAAAGTGGTAGTCTAAGATCTAGATCGATGACATCGATTGCAATGGATCTCTTTGAGCAGCATCAAGCAACGCACATTTCTCAG
TCATTGGACGGTTTTGCTTTTTGCCCTTGGTGCAGATGGACGGTTTTTGTATGCTCTCAGAAAACAGTCTCTATTTATTTAGGTCTTTCACAGGTGCGA
AATGACAGGAAGTAGTGTTCGATTACATCCATCAACAAGACCAATCAGAAATTTAGCAGATCTAATAGGAATCAACATGTGCCCTTCCCATCTG
CAACATCACTCCAGCCTCTGTAGCTTCGGATGACGGTTCGCTTTCGAACCCGGGACCATCTACACCCACTGGGTTCCGATAGACCCCTCCAA
TGAGTATAAAGACGCAAGTCCAGGTCAAAGTCTACAAAGATCTTTCTGATTAAGAATGAAATCTACATTGACCAAAACCGGTTGCCATTTTAA
ATCCTCTGGCTACAGAGTTGACTAGTTTTATCACATCTTAGCCCTCAATATAATTTTTCTTTCGTTCTCGAGGAAACAGGCACCTACAGTCATGG
GACTTGTGGGTTTTAGCCATAGCCCTTCTCTCCTTCACTCAACGAGTACGGCTAGAACCCTGATGTTTGTAAACAAGGCTGACATCTGATTT
CGGATATCTCACTGTGACGCAAGGATCTGAACTCTGGACTACACTGCAGAAGAAATAACCGGAACAAGTATGTACACACTGTGCCACGCACA
AGACATCGAAAACTCGAAAGTGCCATATAGATTTGATCCAAAAAGTCAAGTAATGAGTGGTTACTATAGTTAAATGAATAAAAAAATGGCGGT
TACACTTGGCTGCAGAGCTGCGCAACAGTCAATTTGCAACACCAAGAATGCCGAAGAAGAAACCATTATCTGTGTTAATTACGTCATCAGTCATGT
AGAATATCCCCATTGCGTCATGGATTATTGCCAACTACCAGCATCAGTGACTTAAAGAAATGAAGATCCCAGCAGTTCTGAGAGGGGTACAAGTC
CAGACAAAAGAGGGAGGCAGAGACAATACGGCTACAGGGGACGTTGAAGCGGAGCAGAGGTCACATCGAGGCCAGGAAACCGCGGCGACGTCCA
TGACGATGGTCAAGTATTGAAGGAAGTAGTCAATTTGGATGAGGAGCAACAAAAGAGATTTGCAGGAATTACACAGGACATCGGAGAAAACCG
GGCAATTTCCCTAAAAGAGTTTCCAGGAGTTTTCGTTTTACGACCACTCAAAATGCAAGACGTAACCTCGGACATCCGCTTCCGAGTCTGCGCGGAGC
AGCCGATAACAGACAACGAAAAAGAAACTCGCAGATTCCTTATCGGTGATGAATGAAAACAACGACAACATCAAAGCCCTCGTCCAGGAAGC
TCTTCGTCGACTAGTTGAAGTTCTGTGCTGGACAATGGTACCACGTATAGTGAAGTGACCAACAATGACCTAGTTGTGGACGACATGTCTCTGTA
GCCAATCCGGTTAAATTTCTCAAAATGGGAGGAGGGAAGGACACTCTGAAAACCTCCCCGACAAAAGTGATAAGGTTACTAGACCATGGACTAG
ATCTCCGCGCCGACCGCAATGCATACGACACAACACTGTAGTAAAAATTCATCGGCAATGTCTGTTAGAGATTTAGAAGATGTCATGAACAGGCCT
TGCCACAGCACAATGGATCTCTGCGGAGGCTCGCCAGATGTAATAATGCTCAAACTATTTCCCCACATCGGGTTACCAACACAGCGCCCAA
TACAATGGATTGGTGGTCCACCAGCACCAGCATCTTTGCCAGCAACCACTGTTACGACAAAATACGTCGAGTAGAGAGTCTGTCTATCCGTTCC
GCTGGTGGCATGCGACGCGTCATGGTTGCTACGGTGATACCATTCAAGGCACACTTCTACGCCCTGGTGGACCTGATCCGCCATACAGTGA
TCTCATGTTGCAATCTGCCGCCAACTTTCCAGTTCATCGCCAACTACGGTTCAACTGGATCATCATATTTGGACAATTGCAATGCCATGACGC
CGCAAGCTCAGTATCTCCGAGAGCAAAATTTCAACGAGGACGACGCTGTCGCCGCTGCTGCGGCAATACCGCACATGAGACATCTGCAACTGCG
ACAGACCCGAATTTCTCCAGCACTTACCGTTAAAACCACATCAATTTCTGTACATCCAAGTAATTTAGACCATCCGGGTGCACCATATGGTCA
TCACCACAGCAACAACACTTTCTCCAGATCAGCATCAATCTTTGTATCATCACCAAGTAGCTTTTCAATTTATACCATCCACGCCAGGTGGCTC
TAAAACAGCCCTACATAACAACAAACGGCACTTCATGGTCTGTGACGGCATCATAGATAGTTTCAATGAATTTTTTTGTTGTACAAAAA
AAAAATATGGTGGAACTTTTAAACGCGTTTTTAACTCAGAAATTTGCGGTCACCTTTACTTAGTCGATGGTTGTGAATTAAGCGCTT
GAAAACGCAATGCAAAAGTCTAGCTTGCACAACCTTTATTTGATACAAAATTTATATATAAAAAAAGATTTTAAACAAAGAAAGTGAAGT
GAAAATCAAAATTTGGTGTATGAAGACAAGAGTTATCGGATTTTTGAAGGAGATGGAACTATTCATTTCTGATAAATGCAATTTGTGAAT
CAAAATGTTATTTGTTTATAAAAAAATATTTTTATCATTCAAATCAAATCATGTAATGAAAATCAAGTTAAAATTTGATTTGCGTGCAT
AAAAATATACACATGCTTTCAATGTTATTTTTATTACTATAACGTTATGGATGGTACACAAAAATTTATTTAATGATTTTTTTAGATAATG
CAATTTATTGTTTTTACATATAATGCTTACAACATTTGAAATTTGTTTTCAGATAGGTAACAAATTTGCAAAAAAGCTTTTAAACATCCCAA
AAACGACATAGATATATTGCCGAAAACATTGAAATATCAACTATTCTACGATCATGAATCAAACGTTCTGAAATGTTTCTAATATGTTTGA
TATTTCTGAAACACTCCAGAAATATTTCATTCTGCCAAAATGAAACTATTTATCAAAATTTAATTTATCTGATACGAACTTGGCATGTTTTG
AGTTGTATTTGCATAGGAAAAACCTTAAATGAAAATATCCAACCTTAAGTTGAATTTAAATTAATATCATTTGATTTAAGTTATTACTTAA
TTTATCTGATACATAATTCGATGTTTTGAGTGGTATAATGATATAGGAAGAATTTTCAATTTCTGTTACATTAATGAAATCTTTTCAAAG
AGAATTCGGCTAAATATTTCTTTTTAGCATTAACCAATCTCAATGACCAATTAATTTTTCAAATTTCCCATTTTGTGCTCAAACAACATTC
AATTAATTTTTTTAAAAATCCAAAATGTTTATTTGAAAAAATTTGAAACATTTAATAATGACTATAGCCTAAATTTACTAATTTTTAAAT
GTTATGTTTAGTTATAGGCATTTTTTGGTAAACGGAATTATAGTTCAAAGCGAATTCGTCATTTTTTTGTTACAAAATTCATTTAATTTAAGT
ATTTATTTCCAATATTTTTTAAATGCTCGGTAACGGAATAAGTTTCTGTAATTCCTTACCGAACATTTAAAAAACAATCTTTAAAAA
AATAGTTTTGTTTAAATTTGTATATTTGTATAATTTGATAAACAATTTAAATTTGTATATCTCTGGCTGATCTAAATGTATATCAAAAAA
AAATTAATAATTTAATTTATGAATTTTAAACAATTTTACAATTTTGTGTTGACCTGCAATTTCTGTTTTGTAGAATTTGCTTTATTTGAAT
TCACAAAATCTATTCAATGCAAGATAAAGAATATTCACGTGTAAGAATATTTTAAATGAAAAATATAAATAAAAAGAAATATATTGATAAT
AGAATAAGAACAGCCAAATACAGATCCATAAAAACAATATTTAAATATCGGAATAATAATTAATGCAAAACATTTTACGTAATAAATGAAATTT
AAAACTGAATAATTTGAAATCTTCAAAAAAATAAATATGAAAAATTTAGCACTTTTTTTTTCTGAAAGCAAGATCAAGTTTGTGAT
TAAGCTCTTCTTTGCAAGTATACAAATTTTCAAAATTTTACAATTTTACAATTTTCACTTTGCACTTTGCTCAAACAACATTC
AAGATAAAAAAGAAAAAGGGGTTAAAAGCACAATGAAATGATAAGCCACATTTTCACTCAAGTTGCACGAAAGTTATTGATCAAAAACA
CCACGATTTTTATGATAATGTTTGTGTTTATATACTAATATCAAGTAAAGTAAAGTCAAGTTTGGCTTTGTGATGCTAATATCAATCAAATTT
TCCGAAAAATTTAAAGTATGTACCAGAAAACAATACTAAGAAACAGTTTGTATCAGAAAAATTTAAAAAAGAGCTTCGATAAAAAATGGG
CTTATAATAATTTTATGGAATAAATGCTTTTTACATACATTTATAGATTTACTTTACTTACATTTTAAATTTATATACAACATTTTCTAATAAAGCA
TATATAAAATTAATATACATAATTTTTTTTAACTTTTTTAGCTTAAAAAATTTTCTAGTGTGTTTTATTAAGTATTTAAATGCTAATTTATA
AAATATTCGAATAATTTTAACTACTATATTTTAAAGACTAGTTTTGATCAAATAAATTTAAGACATTAGAATTAAGAAAAATGCTTTTT
ATACATTTACTTTGTTAATTTTACTTCTGTTGCTATTTACTATTTGAAACTGTTTATTTGACTGATTTCTAAATACTGTTTTGAAATTTTTTT
TTTTTACTTTTTGAAATTTTATTCATTTTTTAAAAACAGAAGTGTGTTATTTTCAATAATTTAAATAAATGAGCCCTTTTTTGGCCATTT
TAAAAAGATTTTAAACAAGTTACATTTAAATCATTCTTAACTGCTAGAAAATAAATGAAAATAAATAAATCATTCCATTTACTTTCCCGGTGCT



Supplemental Table 10: Accession numbers of *cfl* homologs from different species

Name	Accession	Description	Organism
Aedes-C15	EAT43307	AAEL005250-PA [Aedes aegypti]	Aedes aegypti
Aga-AGAP003674	EAA08958	AGAP003674-PA [Anopheles gambiae str. PEST]	Anopheles gambiae str. PEST
Am-hmx-like	XP_001119904	PREDICTED: homeobox protein Hmx-like [Apis mellifera]	Apis mellifera
Dm-C15	AAF55898	C15 [Drosophila melanogaster]	Drosophila melanogaster
Dr-Tlx1	AAI62115	T-cell leukemia, homeobox 1 [Danio rerio]	Danio rerio
Dr-Tlx2	NP_705937	T-cell leukemia homeobox protein 2 [Danio rerio]	Danio rerio
Dr-Tlx3	NP_739572	T-cell leukemia homeobox protein 3 [Danio rerio]	Danio rerio
Ek-cll	CDK60407	Clawless [Euperipatoides kanangrensis]	Euperipatoides kanangrensis
Mm-Tlx1	AAH18246	T-cell leukemia, homeobox 1 [Mus musculus]	Mus musculus
Mm-Tlx2	AAI38241	T-cell leukemia, homeobox 2 [Mus musculus]	Mus musculus
Mm-Tlx3	AAI45632	T-cell leukemia, homeobox 3 [Mus musculus]	Mus musculus
STRPU-W4YBJ6	W4YBJ6	Uncharacterized protein OS=Strongylocentrotus purpuratus PE=4 SV=1	Strongylocentrotus purpuratus
STRPU-W4Z6X5	W4Z6X5	Uncharacterized protein OS= GN=Sp-Tlx3 PE=4 SV=1	Strongylocentrotus purpuratus
Tc-TC011749	EFA10692	hypothetical protein TcasGA2_TC011749	Tribolium castaneum

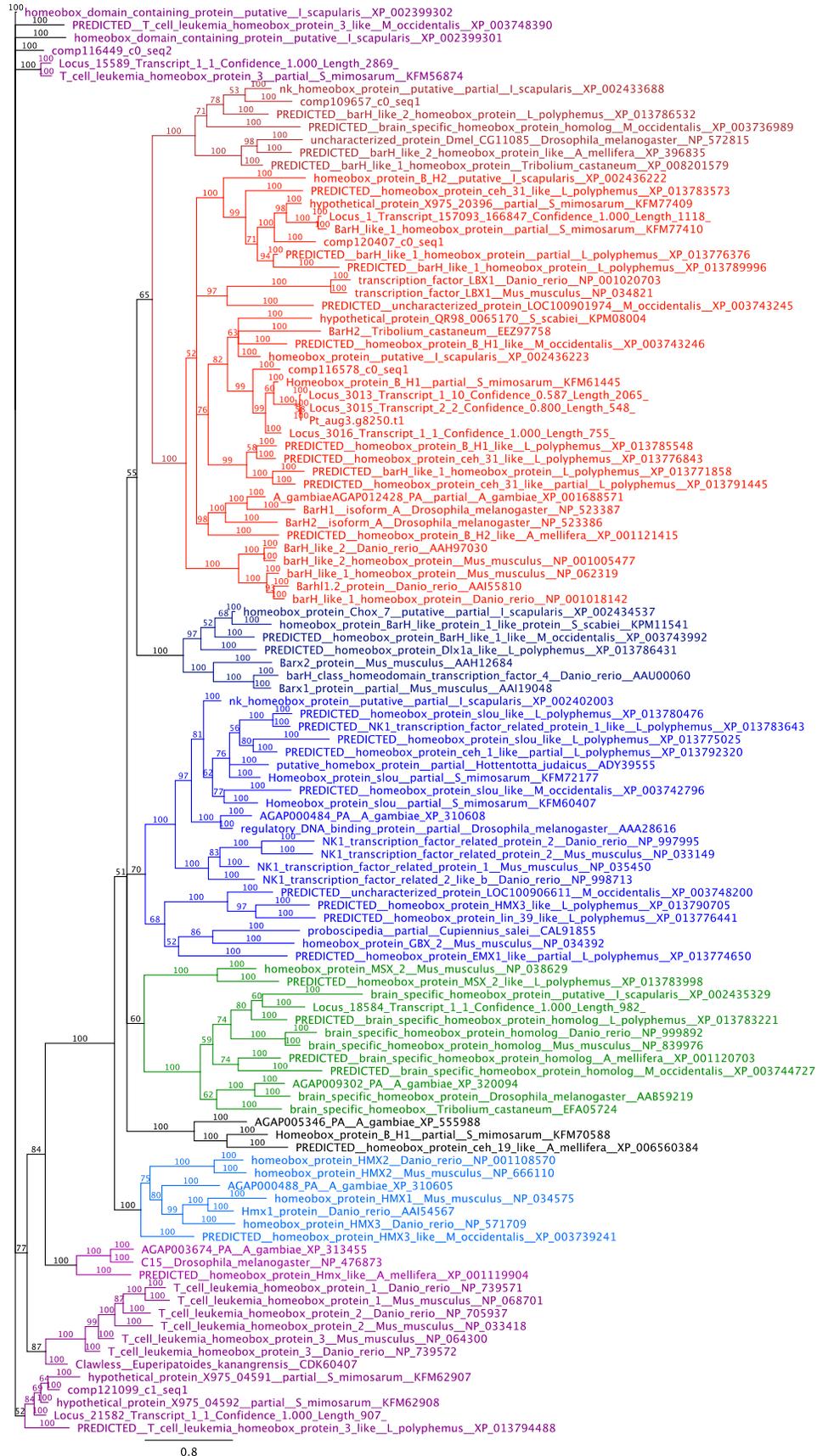
>comp116449_c0_seq2

GTAGAATTCAGCTCGGAGCGTGTGTCAACCGGTTCAACGTTGTAGGAAAATATCTCCGTGTTTTATTGAACATCACCAACATTCGGGAGGTGC
 GTGAGCACATGCTCGGAGGTAATAAAAAATTGAGCTCAGTCGGATGTGTAAAGGATAACGATAATCTCTTTGAAAAATACATGTCTGAAAATTC
 GGACTTGAATGCTTAAAAAGACATTCGTTTTTTCAGCTTCTCTCTCGCAAATTCAGAATTTCCGACTCTGATAAATTTTCAGATAGGAATTT
 TGAGTAAGAAGAAAATGGATGTGATTTGGACTAAACGGAAACGGTAGATGTGGCAGAAAAAAGACATTTGGGAAACAGTAAATACACTAGTCT
 AATAATACTTTGACAGCTGTTGTTTTGTCTCAAGATGTTATGTGATAGGATATCGGAGAGTCGACATAGGCTCACCGAGTAGCGTGGTCGACGAC
 GACGAAGAGCAGTGGACGTCATGGACCAGGTTGATACGGACGACGGTGGGATCCATTCAGGCGTCCGGACAGTCGCGCTCACCGCTGTCCAC
 GTCATCGAATTCACCGTGAACAACAGGTTGCTGTACCAAACAACCAAACTATCTTTCAGCATCAGCCGACTACTCGGAGATGACCATCACA
 ACCAGCATCAGCAATATCGCCAGTCATCGTCTCTCAGGAGTACCGGACTCGGAATCGCCACCCGGAAGACACCGCCGCCACCAACGTCGAGA
 ACAGCAGCGTCCAGATGACTGCCAGGAGAGTTCGTACAGCTCGGGCTCATCGCCACAGATGTCATCTCACCGGTGAACTCTGGTACTTTCCCG
 TCTAGGAGCACCATAACGGTACGCTGATCAACAGCAGCAGCGGTGCTTGTGTGCTACTGGACTCGGTACCCACAGTACATTCATTCGATCT
 CAAATCTGGTACCATTCCCGTGTATGCTCATATGATCCTCTACAGAGTGGTCCAAACGGTACCGTATCAGGGTACCAGTCCAGGCGCCGAC
 CGCCTTTACCCTACCATCATCCGGATCAAGTTTTCTTGGGTTGGAGCTCAAACCGTATGCAAAGACCGATTAGCAGTGCCTTTTTACTCCCAC
 CCGGTATGCCACCGTGCCTCCCAGAGTATCGGGCATCCCTACCAGAACCGAACCCACCAAAGAGGAAGAAACCCCGGACATCTTCCAGGAGGA
 TGCAGATATGTAACTGGAGAAGAGTTCCACAAAACAGAAATACCTGGCTTCTACCGAGAGGGCCACGTTGGCCAAAACAGCTCAAAATGACCGA
 CGCTCAAGTCAAGACCTGTTTCAAACAGGCGGACGAAATGGAGGAGGCAAACAGCCGAAGAAAGGGAAGCAGAACGACAGGCGCCAAAACCGA
 CTGATGATGCTCTCCAAGCCGAAGTCTGTCGAAGACGATATACGACCAGAACAGGGACTCGTTATGCATAGGCAACCGCTCGTTGACGCACT
 GCAGACTCTGCAACCGTGGGCGCGGAATCAGCAGCGGCTGCCACGCGCGCGGTGGCCGCTGCCGACGCGGACACCACAGGCATCAGGACGG
 TAGGCCGAGTTCTCGTCCGCCACCTCCACCACCGCCCCACCACATTCCTTAGCATCACCGATGTGTGATTAAAGTTATGCGCGTATTAAC
 AGACTATAACACCATTATCATCGACTACGAAACCCCCCTTTCTTTTCAGGACTTGACGAAATATATAGTCTTGTGATATTTCCGATATAAAA
 AATTTTCATTTTTGTGACCACCGTATTATGTTTAGTTTTCCCGTGTACAAAGAAAAAATTTGAATAATGTTATATTAATAATTTGAATAGACAGA
 TAATACAAATTTAGATTTTTTTATTAATTTTTGTAGAAAAGTTGCTGGTGCCTTGACCTATGAGTTTTCCCTATTGGGAGTGTATTTACGCTACA
 TACGAAAATATTCGCTGAAAATGCAGAAATTCATGAAGAATTTTTGATTAATCCTTTGATGACACAGTGTATTTATGTTGCAACGGCAT
 CGGAAGCAAACAGCTGAGAATCGACAGCGCCGCAACCTATAAGATGTAACTTTACAAGATGCGCTCTGGTGTAAATGTCTTTTAGCTGTAAA
 AGCTGCTTTCTGGATAACAATTTGAGTTCTGGTTACTTACATTTGCGCTCTACTGCAGAGGTTTCAACCTT

>comp121099_c1_seq1

ATATTTAAAAAATTTACATGTTCTCACAAAAATCATTTCCAATGCAACTGGTATTGATACTTACTTTTATATTAAGAGTAAATGCTA
 TTAAAGATGTATCCAGCAAGAAAATAGCTTACAATAGTAACATAAAGATGTTCTGATCTCGCTATACATGTCCAAAAAATAATTTTTTTTT
 TACCTTGAATAATATTAGAATTAAGCAAGACCGGAACAATAAGTTGTTTTCAATCTATGACGGTAAATGAAATGAATGCTGAATCCAGTTTAT
 AATACTTTTATAAAGTTTCATTTACTGGAAAAGGAATTTGAAAAAATAACACAAAAACATTTGATTTTATGACGACGAAAACAAAGAATCCATCACT
 CCATACGTTTTGAGAAGCAATGAAAATCGTTACATTTTTCTGTGAGTTTCCATCTAGTTATAAAGAAATGAAAAAAGTGAATTTAGTCTT
 TTTTTTCTTAGTTTTAATTAACCGCTATATATCAGTCCATCTCACTATAAGTGAAGTTCTAGTTTTGGCACTAATAGGCGAAAATAAAAAA
 CAAAAGGAAACGAGAAGAAACAAGATTACAATGCACCTTCTAGGCTTTGACGTGTGGACACACATTTGATGAACTAAACGATTTTTCTATC

Supplemental Material



Supplemental Material

TCAGCGGACTATGCCATGGTAAGATTGACTGGATGATTATACATGCTGTGGTGCAGAAATTAATAAAATGATTTTCTTTAATGTGAAGTGATTT
TTTGGTTGAACATCACTGCTATAATTAATTGGCTTAAACAATTGAAATGAATAATAGAAACGAAAAATTAGTACTTAA

>Locus_3013_Transcript_10/10_Confidence_0.587_Length_2075

ACGATCATTCTGGTCTGGAAGAAACAAACAGTAATGACCCATCCAGTAGTCCACCTTACAATCTAAAAGTTAAAAACAGAGAAAAGCAAGGAC
GGCATTACTGACCATCAGCTTCAAACGCTGAAAAAGAGCTTCGAGCGGACAGAAATACCTCAGCGTCCAGGACCGAATGGAGCTAGCGTCCAAAC
TCAATCTCACAGATACTCAAGTCAAAACATGGTATCAAAAATAGAAGGACGAAATGGAAGCGTCAAACAGCTGTTGGTTTGAAGCTTTTAGCCGA
AGCCGGTAACTATGCCGAGTTCAGCGTATGCTTCACTACTCTACTGCTGGAATCAATATGCCGCTAATTTACCTTTTACCCTGGATTAC
CAGGAATGGACATGTATTACCAGCAAGGACTTGGTCTTTCAGAAACCAATGGCTTACAGGATGTATCCATCTGCTCCAACGACAGCCATGGGATTG
CCAGTCAATTTCTCTCATCTACTGTGCCACCAATGGCTGCGGCAAGTGCACCTTACCACCTCATCTTCCAGCGTCACTTTACCTTCTTATTTT
GCTAGGGAACCCCTTGTGAAAATAGATTGATAGCTATTTAAAAAAATTTAATTTATTTGTGATCCTCGCTGAAAAATGACCGAAGACTT
GTAATAACTCTTGGACTTCTCAAACGCTTCCAGAACTTGTACATATTTGTATAGAACTCTTGGACTTCTCTTTGTCAACTATGAAAAGAAAAGG
GTGCTTCTAATATGAGTTGATGAATTTGGCTGACTACTAGTCTATGGCGCAGGTTTGGCGAAAAATGATAGAAATGGCGATAAAAAATCCAGGA
TTGAAGTTAAAAATAGGAGCTGCCTCAGTGATAAGACAAGAGTAAAAGAATGCCTGCGCGCCAGGTTATTCACCTTTTTATGGACTGTAACAGAA
AATGAAAAACCATTTTGCATTTTTTTTTTAAATTTCTTTTGTAGGATTGCATATGTACATTTAAAGTGTACATTTTGTATGATACATTAAGTGT
ACAATTTCTGTACATTTCAATTTATGGCGCGCTTTATGAAAAAATAGAAAATATAGCATCGCTTAGAATGTGAAAAATGAATATACTTAA
ATGATTTCTAATATAAATTTTTTTATGAGTGTTTTTGTAGTGCATAAAAATTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAA
TTGTTTTTGAACACATCTTAAATTTTAAATTTTAAATTTCTTAAACAGATTTAAGAAATTAATTAATTTTACAGTAACTTATTAACAGACCTA
AAAAGAATTAGCTTGCATTTCTTCAATTTTAACTCATATAGTTTTTAATTTATTTTACAGCTAACTTATTAACAGACCTAATAAATTCATGA
TTTGATTTTAAAGGGTTGGACGATTAAGGATTAGTGTAAAGTTAAACTTTTCAAAGTATTAAATTTCTCTCTAAAATAGTCAATTTTTTTT
CATCGAAGAAATAAGTTTTTTCTGTGATCTCAGTTAATAACTAATTTAATAATCGAACAGTGAATTTAATAATCCAATACTACTCGCTAATACG
TAGCTATTTACAGTCCATCCGCAAGAAACAGACAGTAAATTTTCAATTAATCGAGATATGCTTGAATAACTACTTACCTTTTCGCTGCAAAAA
GTTTTCAATTTGCGCACTACCTGTTATTTAATAAATTTTTTCAAATTAGGCATGCACATTTTCTGTTTTCTTTCTTTCCATTTTGTATTTAGCA
TTTTCAATATAAATTTGCAAAATTTATATTTCTTTTGTGTTTTCTTTCAATTAGATGCATCGTTTTTATTTATCCATTATGGTGTAGGTATA
GTTTTTATGTTTTAAAAATTTCTTTTGTACGCTTCTATACAATTTTCGCTTCCGTTTTTGGTGTCTTTGATTTCCAATTTTTTTGTGTCTTCGT
CTTTTTCTCTCCCTTATTTTGTGGCGTTTTCAGTTTATAGAGTAATGATTAATTTTCTTTTGTCTTTTAAAGTTAATTTTTTGTATTAC

>aug3.g8250.t1

AATCATTCCATGTTCAAAGGTTATTGTTTTATCCCGGCTTGCAGCAATATCTATCCGAGCAAAATGTTTCAATTTCCGAAAGGGTGGTCCACCACC
ACGCTGCCCTCACAGAATTCACCTCCTCCCATTCGTGTCAAAGGTTGACCATCCAGAAGCAAAATAGTGACGATGATGATTTTTGGATTGACGT
CGAAAAAGCAAGCAGATGATGAATCAAACGATCATTCTGGTTCGGAAGAAACAAACAGTAATGACCCATCCAGTAGTCCACTTAAACTTA
AAAGTTAAAAACAGAGAAAAGCAAGGACGGCATTTACTGACCATCAGCTTCAAACGCTGAAAAAGAGCTTCGAGCGACAGAAATACCTCAGCG
TCCAGGACCGAATGGAAGTCCAAACTCAATCTCACAGATACTCAAGTCAAAACATGGTATCAAAAATAGAAGGACGAAATGGAAGCGTCA
AACAGCTGTTGGTTTGAAGCTTTTAGCCGAAGCCGGTAACTATGCTGCAGTTTCCAGGATGCTTCACTACTCTTACTGCGTGAATCAATATG
CCGCTAATTTACCTTTTACCCTGGATTACCAGGAATGGACATGTATTACCAGCAAGGACTTGGTCTTCCAGAAACCAATGGCTTACAGGATGTAT
CCATCTGCTCCAACGACAGCCATGGGATTGGCAGTCAATTTCTCTTCACTACTGTCACCAATGGCTGCGCAAGTCCACTTACCACCTCA
TCTTCCAGCTCATCTTACCTTCTTATTTTCGCTAGGGAACCCCTTGTGAAAATAGATTGATAGCTATTTAAAAAAATTTAATTTATTTATTTG
GATCCTCCGTCGAAAAATGACAGAACTTGTAAATACTTGGACTTCTCAAACGTTCCAGAATTTGTACATATTTGTATAGAAGTCTTGGAC
TTTTCTTTTCAACTATGAAAGAAAAGGTGGCTTCTAATATGAGTTGATGAATTTGGCTGACTACTAGTCTATGGCGCAGTTTTTGGCGAAAA
ATGATAGAAATGGCGATGAAAATCCAGGATTGAAGTTAAAAATAGGAGTGCCTCAGTGAAGACAAGAGTAAAAGAAATGCCTGACCGCCAGG
TTATTCATTTTCTATGGACAGTTACAGAAAATGAAAAACATTTTGCACCTTTTTTTTTTAAATTTCTTTTGTAGTGAAGTTGCATATGTAATTA
AGTGTACATTTTGTATGATACATTTAAAGTGTACAATTTCTGTACATTTTCAATTTATGGCGCGCTTTATGAAAAAATAGAAAATATAGCATCGCT
TAGAATGCTGAAAAATGAATATACTTAAATGATTTCTAATATAAATTTTTTTTATGAGTGTTTTTGTTACGCGACATAAAATTTAAAAA
AATAATAAATAATTTACATGAAAATATACATTTGTTTTGAAGGCACATTTCAATTTTAAATTTTAAATTTTAAACAGATTTAAGAATTA
ATTTTACATGTAATTTTATGCAATTTTCAAAGAAATTTCTTGTGATTTTCAATTTTAACTCATATAGTTTTAATTTATTTTACAG
TAAACTTTAATACAGACCTAATAATTCATGATTTGATTTTAAAGGTTAGACGATTAAGGATTAGTGAAGTTATAACTTTTCTAAAAGTATT
AATTTTCTCTAAAATAGTCAATTTTTTTCTTCAAAGAAATAAGTTTTTTCTGTGATCTCAGTTAATAACTAATTTAATATCGAACAGTGA
ATTTAATAATCCAATACTACTCGCTAATACGTAGCTATTTACATCCATGGCAAGAAACGAGACAGTAAATTTGCAATTAATCCAGATATTC
TTGCAATAACTACTTACCTTTTCCGTCGAAAAGTTTTCAATTTGCGCACTACCTGTTATTTAATAATTTCTTTCAAATTTAGGCATGCATTTT
CTGTTTTCTTCCATTTTGTATTTATGCAATTTTCAATATAAAAAATTTGCAAAATTTATATTTCTTTTGTGTTTTCTTTCAAATTAGATGCATC
GTTTTTATTTATCCATTTATGGTGTAGGTATAGTTTTTATGTTTTAAAAATTTCTTTTGTACGCTTCTATACAATTTCCGCTTCCGTTTTTGGT
CTTTGATTTCCAATTTTTTTGTGTGCTTCGCTCTTTTCTGCTCCCTTATTTTGTGGCGTTTCAGTTTATAGAGTAATGATTAATTTCTTT
TGCTTTTAAAGTTAATTTTTGTATTACAATGGAGCGACCGTAGCTGCTCAACAATTTGTAATTTCAATACCTAATTTTTTGTTTTTTCTTAAAGT
AAGCGTAAATATACAGAGCCAGTGAAGAGATAAAAAAATTTTGTGCTGTATTGACCTTTCCAAGTATTGTTTCGCCATCAAATTTTTAAAT
TTTTTTTTTTTTTTTAAATTTATAGAAGCAGCATTGATATTTGCAAAGACCGTTTCAGATACAACACTACAAAAGTAGTTGATTTGATCATTTTAAAG
ATTTATTTATTTATTTATTTTTTTTTTTTGTGCGATCGTTAAAAAAATCAGTTTTTATCAAGCATGCACCGTATTGGAAATTTGCGCAATGATTCAC
TTAGTTTGAACACAAAGAAAAATGAATTTATGCAAACTGTTTTAACTATATAAATCTTATTTGTTTCAAACAGTTAGTTACACTTCTTCAA
GAAAGTCAAATCTATTTATTTCAAACCTAATTAAGCAAAAATTTGCAAAAATGAGTGAATAAACTGTAATAAATGATTTATATAATTCAGTAA
TTGTACAACTATAATAATTTGTAATTTGTAAGGAATTTTCTCTTTTAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA
TAGGCTGTAGCTATACAAAAGTTTGAAGAAATAAGAATAAAAAAGAAATGCAACTAATGTAATTTAATAAATTTATTTATCTTTATATATTAG
CTTTTTTTTTTTTTTAAATTTTACAGTTTTAAGTTGAATTTTCAAGAGTCTGCTAAACAATAATTTCAAATTTTTTTTTTTTTTTTTTAA
ATTTTAAATATTTTACAGTTTTAAGTTGAATTTTCAAGAGTCTGCTAAACAATAATTTCAAATTTTGTACATAAAAATTTTGAACAATTTTAT
TGCAATTTAATAATTTTATTTGCTTCAAACCTTAAAAAAGAAAAAATTTGCTATATGCTATATTTTCTTTCAGTTTAAATTTTCTCTTC
ATGTCGCCGGAATGTAATAAGTGGTAGCATGTTTTTAAATTTGAAAAATTTAGTTTTTACAGTAAACGACTAGCATACGCTTAAAGAAAGGATAA
CTACTTTTGCCTTGTATTTTTTAAAAAATAACTGGCTATTTGAATTTATGAAAACATGATTTTATTTGATTTATAATACTATTTCTGTCAACCA
ATTTTTAATTTTCAATAACATTTGGCGCAGCTGTACATTTCACTTATAAATGTGATTTTAACT

>aug3.g8250.t1

MFYPLGQQYLSEQMFEFKGGPPPTLPSQNSPPSHSCQKVTIQKTNSSDDDFWIDVENETDDESNDHSGLEETNSNDPSSSPYLNKVKKQRKARTAFTHQLQTLKESFERQKYLVSQDRMELASKLNLTDTQVKTWYQNRRTKWKQRQTAVGLELLAEAGNYAAVQRMLQSTPYWLNQYAAANLPTAGLPGMDMYRQGLGLQKPMAYRMYPSAPTTAMGLPVNFSSTVPPMAAASATSPSSASSLPSYFAREPPC

>comp116578_c0_seq1_-_ORF_1_(frame_2)_translation

MTVDLPQKMPADRDRLRVVDDDDGSRDDSPPPQSTTAFRVAASPPSVVAGGGTTAPSSFLIRDILSGTEQQQQQRQSHIRGSGFSPVGYTASPS TTNSNPTSPLSASDPGHHHHHHPPPPHPLMHMPLAHHHHALQHHPAISGQFFSPLDIRRPNGEVSPLRDDEIDEDGVSSDDDCSRSGKNDPQK LNSSNSPDPSKSPYLNKVKKQRKARTAFTHQLQTLKESFERQKYLVSQDRMELAAKLNLTDTQVKTWYQNRRTKWKQRQTAVGLELLAEAGN YAAVQRMLQSSPYWLNHYAANLSFSSVPGMDMYRQAAAALVQKPMAYRMYPTTTATPMGMPVNFSSSTVPPPPPPMNSNSAPSPTSPLSST TSSSTLNSFYGREPPPC

>comp120407_c0_seq1_-_ORF_1_(frame_2)_translation

MTVTMSADTINGRTSAEPTVEELVEQPTDLSTKKESETESGSDNSPRSSPSTLADVRDSNPAPTSGPSFLIKNLIDHNSHHHQQSHQDFAGLIPSTF HRYRPSFVSPNPFYFLPHFAANIPSGKFCFDQDGSNRRMDSSESEGRPPSPVYQRTLPQELADYKQICDSRDVLTTRPNDAAETVSSSKDRISY TAVIDSDTDSISNHDGESFDAQPPREDSPLKKNKPRKARTAFTHQLRTEKESFERQKYLVSQDRMELAAKLNLTDTQVKTWYQNRRTKWKQRQ AMIGFEVFPDACLARFPRWLWGAQAPYWPYSYSSYLSTISSLSGLNTNPLVGGTSSSPASPPSLEGYGRQAGSAVSNVSRPLLPRAVYPPSPSNF SSVSPFYQQAGP

>Locus_1_Transcript_157093/166847_Confidence_1.0_Length_1118_-_ORF_1_(frame_3)_translation

MKNKKPRKARTAFTHQLRTEKESFERQKYLVSQDRMELAAKLNLTDTQVKTWYQNRRTKWKQRQNMTCFYAADSATLARFPRWLWGPSPAS YFPSYSSYSSYLSTMSSLSTLGAALHTPASSTSVPTLDLYQRQSPVSSSTIGRPVLPRAVYPTPSTYPSLQPFHGRPTQ*

>Locus_3013_Transcript_10/10_Confidence_0.587_Length_2075_-_ORF_1_(frame_3)_translation

MEETNSNDPSSSPYLNKVKKQRKARTAFTHQLQTLKESFERQKYLVSQDRMELASKLNLTDTQVKTWYQNRRTKWKQRQTAVGLELLAEAGN YAAVQRMLQSTPYWLNQYAAANLPTAGLPGMDMYRQGLGLQKPMAYRMYPSAPTTAMGLPVNFSSTVPPMAAASATSPSSASSLPSYFAR EPPC*

Curriculum vitae

Personal Data:

Name	Natascha Zhang neé Turetzek
Date of birth	08th November 1986
Place of birth	Holzminden, Germany
Home adress	Stumpfe Eiche 12 37077 Göttingen, Germany
Work adress	Georg-August-Universität Göttingen Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie Abteilung für Entwicklungsbiologie GZMB Ernst-Caspari-Haus Justus-von-Liebig-Weg 11 37077 Göttingen, Germany
Email	nturetz@gwdg.de

Education:

2011- 2016	PhD student of the GGNB (Göttingen Graduate School for Neurosciences and Molecular Biosciences), PhD program: Genes and Development, Georg August University, Department of Developmental Biology Doctoral Thesis Project “Developmental studies of appendage formation and patterning in spiders”
2009 – 2011	Master program Developmental, Neural and Behavioral Biology at the Georg-August- University Göttingen, Final grade “very good” 1.2 Master thesis “On the role of hox genes and co-factors in body plan specification in spiders” at the Institute for Zoology and Anthropology, Department of Developmental Biology (Prof. Dr. Wimmer) in the laboratory for Evolutionary Developmental Biology of Arthropods (Dr. Prpic Schäper) Thesis evaluation: “very good” (1.0)
2006 – 2009	Bachelor of Science in biology at the Georg-August-University of Göttingen. Final grade “good” (1.6).

Major field molecular biology with 6 focused modules (developmental biology, biochemistry, plant physiology, bioinformatics, genetic, microbiology).
Bachelor Thesis "Function of Mars in Neuroblasts of *Drosophila melanogaster* and its Interactions with other Mitosis Relevant Proteins" at the GZMB), in the Department of Stem Cell Biology (Prof. Dr. Wodarz) Thesis evaluation: "good" (1.7)

1999 - 2006

Secondary school "Campe-Gymnasium-Holzminden"

1993 – 1999

Elementary school "Grundschule Stadtoldendorf" and orientation stage of the secondary school at the "Orientierungsstufe Stadtoldendorf"

Stipends and Awards:

Christiane-Nüsslein-Volhard-Foundation fellowship and a "Women in Science" Award by L'Oréal Deutschland and the Deutsche UNESCO-Kommission.

Work Experience:

Dec. 2010 – Nov. 2011

Student Assistant at the GZMB, department developmental biology (Dr. Prof. Wimmer) in the laboratory for evolutionary developmental biology of arthropods (Dr. Prpic Schäper)

Aug. 2010

Behavioral Observations of Malagasy *verreauxi* Sifakas (*Propithecus verreauxi verreauxi*) in Kirindy Forest, Madagascar, for the German Primate Centre (DPZ), department of Behavior biology and sociobiology/anthropology (Prof. Dr. Kappeler),

Sep. 2008– Aug. 2009

Student Assistant at the Max Planck Institute for Biophysical Chemistry Department of Protein crystallization, Structural Biochemistry Group (Prof. Dr. Wahl)

Teaching activities:

2014

Master thesis supervision Tatiana Königsmann-Erichsen

2012

Supervision of Tatiana Königsmann-Erichsen (Lab rotation 04.07-05.09 2013)

2013	Master thesis supervision of Monique Weidner
2012	Supervision of Monique Weidner (Lab rotation 04.02-30.03.2013)
Since 2011	Teaching assistant of the key competence module evolutionary developmental biology of the Master program Developmental, Neural and Behavior biology
2011	Bachelor thesis supervision of Tobias Klös
2011	Supervision of Tobias Klös (Lab rotation 18.04 – 27.05.2011)
Apr. – Jul. 2008	Teaching assistant for the lecture of basic biology “Ringvorlesung” (for undergraduate students) at the Georg-August- University Göttingen

Languages:

German (native)

English (fluent, writing, reading and speaking).

List of Publications:

Turetzek, N., Prpic, N.–M. Observations on germ band development in the cellar spider *Pholcus phalangioides*. In preparation.

Turetzek, N., Pechmann, M., Schomburg, C., Schneider, J., and Prpic, N.-M. (2016). Neofunctionalization of a Duplicate *dachshund* Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. *Mol Biol Evol* 33, 109–121.

Schomburg, C., **Turetzek, N.**, Schacht, M.I., Schneider, J., Kirfel, P., Prpic, N.-M., and Posnien, N. (2015). Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *EvoDevo* 6, 15.

Janssen, R., Schönauer, A., Weber, M., **Turetzek, N.**, Hogvall, M., Goss, G.E., Patel, N.H., McGregor, A.P., and Hilbrant, M. (2015). The evolution and expression of panarthropod *frizzled* genes. *Frizzled* 96.

Khadjeh S, **Turetzek N**, Pechmann M, Prpic NM.
Warum Spinnen (nur) acht Beine haben.
Biologie in unserer Zeit 43, 112-117 (2013).

Khadjeh S, **Turetzek N**, Pechmann M, Schwager EE, Wimmer EA, Damen WGM, Prpic NM
Divergent role of the Hox gene *Antennapedia* in spiders is responsible for the convergent evolution of abdominal limb repression.

Proceedings of the National Academy of Sciences of the United States of America 109, 4921-4926 (2012).

Pechmann M, Khadjeh S, **Turetzek** N, Prpic NM
Beinentwicklung und Gliedmaßen-Evolution bei Spinnen.
BIOspektrum 18, 365-368 (2012).

Pechmann, M., Khadjeh, S., **Turetzek**, N., McGregor, A.P., Damen, W.G.M., and Prpic, N.-M.
(2011). Novel Function of *Distal-less* as a Gap Gene during Spider Segmentation. PLoS Genetics
7, e1002342.