Phylum-wide transcriptome analysis of oogenesis and early embryogenesis in selected nematode species

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Christopher Kraus

aus Waldbröl

Hundt Druck GmbH, Köln

2016

Berichterstatter:	Prof. Dr. E. Schierenberg
(Gutachter)	Prof. Dr. S. Roth
Dissertationsgesuch:	20.04.2015
Disputation:	23.06.2015

Doctoral Thesis

Phylum-wide transcriptome analysis of oogenesis and early embryogenesis in selected nematode species

Author:
Christopher Kraus

Supervisor:
Prof. Dr. E. Schierenberg

Thesis Committee:

Prof. Dr. E. Schierenberg

Prof. Dr. S. Roth

Prof. Dr. T. Wiehe

Dedicated to

my angel Jessica

my parents Annemarie and Ewald

and my little brother Kevin

Abstract

Oogenesis is a prerequisite for embryogenesis in Metazoa. During both biological processes important decisions must be made to form the embryo and hence ensure the next generation: (1) Maternal gene products (mRNAs, proteins and nutrients) must be supplied to the embryo. (2) Polarity must be established and axes must be specified. While incorporation of maternal gene products occurs during oogenesis, the time point of polarity establishment and axis specification varies among species, as it is accomplished either prior, during, or after fertilisation. But not only the time point when these events take place varies among species but also the underlying mechanisms by which they are triggered. For the nematode model *Caenorhabditis elegans* the underlying pathways and gene regulatory networks (GRNs) are well understood. It is known that there the sperm entry point initiates a primary polarity in the 1-celled egg and with it the establishment of the anteroposterior axis. However, studies of other nematodes demonstrated that polarity establishment can be independent of sperm entry (Goldstein et al. 1998, Lahl et al. 2006) and that cleavage patterns, symmetry formation and cell specification also differ from *C. elegans*. In contrast to the studied Chromadorea (more derived nematodes including *C. elegans*), embryos of some marine Enoplea (more basal representatives) even show no discernible early polarity and blastomeres can adopt variable cell fates (Voronov and Panchin 1998). The underlying pathways controlling the obviously variant embryonic processes in non-*Caenorhabditis* nematodes are essentially unknown. In this thesis I addressed this issue by performing a detailed unbiased comparative transcriptome analysis based on microarrays and RNA sequencing of selected developmental stages in a variety of nematodes from different phylogenetic branches with *C. elegans* as a reference system and a nematomorph as an outgroup representative. In addition, I made use of available genomic data to determine the presence or absence of genes for which no expression had been detected. In particular, I focussed on components of selected pathways or GRNs which are known to play essential roles during *C. elegans* development and/or other invertebrate or vertebrate model systems. Oogenesis must be regulated differently in non-*Caenorhabditis* nematodes, as crucial controlling components of Wnt and sex determination signaling are absent in these species. In this respect, I identified female-specific expression of potential polarity associated genes during gonad development and oogenesis in the Enoplean nematode *Romanomermis culicivorax*. I could show that known downstream components of the polarity complexes PAR-3/-6/PKC-3 and PAR-1/-2 are absent in non-*Caenorhabditis* species. Even PAR-2 as part of the polarity complex does not exist in these nematodes. Instead, transcriptomes of nematodes (including *C. elegans*), show expression of other polarity-associated complexes such as the Lgl (Lethal giant larvae) complex. This result could pose an alternative route for nematodes and nematomorphs to initiate polarity during early embryogenesis. I could show

that crucial pathways of axis specification, such as Wnt and BMP are very different in *C. elegans* compared to other nematodes. In the former, Wnt signaling, for instance, is mediated by four paralogous β-catenins, while other Chromadorea have fewer and Enoplea only one β-catenin. The transcriptomes of *R. culicivorax* and the nematomorph show that regulators of BMP (e.g. Chordin), are specifically expressed during early embryogenesis only in Enoplea and the close outgroup of nematomorphs.

In conclusion, my results demonstrate that the molecular machinery controlling oogenesis and embryogenesis in nematodes is unexpectedly variable and *C. elegans* cannot be taken as a general model for nematode development. Under this perspective, Enoplean nematodes show more similarities with outgroups than with *C. elegans*. It appears that certain pathway components were lost or gained during evolution and others adopted new functions. Based on my findings I can conjecture, which pathway components may be ancestral and which were newly acquired in the course of nematode evolution.

Zusammenfassung

Der biologische Prozess der Oogenese ist eine Vorraussetzung für die Embryogenese in Metazoa. In beiden Prozessen müssen wichtige entwicklungsbiologische Entscheidungen für die nächste Generation getroffen werden: (1) Maternale Vorprodukte (z.B. mRNAs und Proteine) und Nährstoffe müssen dem Embryo zugeführt werden. (2) Polarität muss etabliert und Achsen müssen festgelegt werden. Maternale Vorprodukte und Nährstoffe werden während der Oogenese eingelagert. Der Zeitpunkt der Polaritätsetablierung und Achsenfestlegung variiert zwischen einzelnen Spezies und kann vor, während, oder nach der Befruchtung erfolgen. Aber nicht nur der Zeitpunkt der Aktivierung dieser biologischen Prozesse, sondern auch die molekularen Mechanismen, welche sie steuern sind unterschiedlich. Für den Modellorganismus *Caenorhabditis elegans* sind die Signalwege und genregulatorischen Netzwerke (GRN) gut untersucht. Es ist bekannt, dass der Eintrittspunkt des Spermiums die Polarität im einzelligen *C. elegans* Embryo festlegt, und somit die anterior-posterior Achse determiniert. Andere Studien haben hingegen gezeigt, dass die Polarität in Nematoden auch unabhängig vom Eintrittspunkt des Spermiums festgelegt werden kann (Goldstein et al. 1998, Lahl et al. 2006) und dass Teilungsmuster, Symmetriebildung und Zellspezifikation unterschiedlich zu *C. elegans* sein müssen. Im Gegensatz zu den untersuchten Chromadorea (eher abgeleitete Nematoden, die auch *C. elegans* beinhalten) existieren auch einige marine Enoplea (eher basale Vertreter der Nematoden) deren Embryos keine klar ersichtliche Polarität zeigen und Blastomere können unterschiedliche Zellschicksale annehmen (Voronov and Panchin 1998). Die unterliegenden Signalwege, die diese offensichtlich unterschiedlichen embryologischen Prozesse kontrollieren, sind in Nematoden außerhalb des Genus *Caenorhabditis* weitestgehend unbekannt. In dieser Doktorarbeit befasse ich mich mit diesem Problem, indem ich einen Transkriptom-Vergleich mit Hilfe von Microarrays und RNA Sequenzierungen (RNAseq) für ausgewählte Entwicklungsstadien von Nematoden unterschiedlicher phylogenetischer Gruppen durchgeführt habe. Hierzu nutzte ich das Modell *C. elegans* als Referenzsystem und einen Nematomorphen als repräsentativen Außengruppenvertreter. Zusätzlich habe ich Genominformation genutzt, um das Vorhandensein und die Abwesenheit von Genen, für die ich keine Expression nachweisen konnte, festzustellen. Hierbei habe ich mich auf Komponenten von ausgewählten Signalwegen und GRNs fokussiert, von denen bekannt ist, dass sie essentiell für die Entwicklung von *C. elegans* und Invertebraten- und Vertebratenmodelle sind. Die Oogenese in *C. elegans* muss anders reguliert werden als in Nematoden anderer Gattungen, da essentielle regulative Komponenten des Wnt und des "Sex-Determinations" Signalweges in diesen Spezies nicht existieren. In diesem Zusammenhang habe ich spezifische Expression von Genen, die mit Gonadenentwicklung und

Oogenese assoziiert werden können, in dem Enoplea *Romanomermis culicivorax* detektiert. Zusätzlich konnte ich zeigen, dass bekannte Komponenten der Polaritätskomplexe PAR-3/-6/PKC-3 und PAR-1/-2 in Spezies außerhalb des Genus *Caenorhabditis* nicht existieren. Selbst PAR-2, als Bestandteil eines dieser Komplexes, existiert nicht in diesen Nematoden. Hingegen zeigen Analysen der Nematodentranskriptome (einschließlich *C. elegans*), dass andere polaritätsassoziierte Komplexe, wie z.B. Lgl (Lethal giant larvae), exprimiert werden. Diese Ergebnisse zeigen einen alternativen Weg, um Polarität während der Frühentwicklung in Nematoden festzulegen. Ich konnte auch zeigen, dass wichtige Signalwege der Achsenfestlegung, wie z.B. Wnt und BMP, sehr unterschiedlich zu *C. elegans* sein müssen. Zum Beispiel existieren für den Wnt Signalweg in *C. elegans* vier β-Catenin Paraloge, während andere Chromadorea weniger und Enoplea nur ein β-Catenin besitzen. Die Transkriptome von *R. culicivorax* und dem Nematomorphen zeigen, dass Regulatoren des BMP Signalweges (z.B. Chordin) spezifisch in Enoplea und Nematomorphen während der Frühentwicklung exprimiert werden.

Zusammenfassend zeigen meine Ergebnisse, dass die molekulare Maschinerie, die Oogenese und Embryogenese in Nematoden reguliert, unerwartet variabel ist und somit kann *C. elegans* nicht als repräsentatives Modell für Nematoden-Entwicklung betrachtet werden. Es scheint so, als ob Komponenten bestimmer Signalwege während der Evolution verloren gegangen sind oder neue Funktionen angenommen haben müssen. Basierend auf meinen Ergebnissen kann ich vermuten, welche Signalweg-Komponenten anzestral sind und welche während der Nematodenevolution neu erworben wurden.

Acknowledgements

First of all I have to thank my girlfriend Jessica for her love, her support and time she had to waste for me during the days of writing this thesis. I love you my angel. Without you this would not have been possible.

I also like to thank my parents Annemarie and Ewald Kraus and my little brother Kevin for their support during this period of my life.

I like to thank all the members of the Schierenberg Lab, Ndifon, Julia, Theresa, Lydia. Thanks guys! Especially, I like to thank Michael, Philipp, Dom and Peter for long and deep scientific discussions, advice and help.

I have to thank the Kohara lab, Yuji Kohara, Hiroshi Kagoshima and Hideaki Hiraki for the colaboration on the nematode *Diploscapter coronatus* and the opportunity to work with their genome and EST sequencing data of this parthenogenote.

I like to thank Prof. Edward G. Platzer and Prof. Ben Hanelt for supporting me with samples of the Enoplean nematode *Romanomermis culicivorax* and the nematomorph *Grodius sp.*

I like to thank Itai Yanai for interesting suggestions and providing me with information for extraction of infinitesimal amounts of RNA.

I like to thank Guenter Plickert for his support and very interesting suggestions for my thesis.

I like to thank all Edinburgh people I had the (short) honor to met: Mark Blaxter, Sujai Kumar, Georgios Koutsovoulos and Dominik.

Of course, I like thank my supervisor Prof. Einhard Schierenberg for giving me such an interesting project and supporting me from the start to the end.
THANKS TO YOU ALL!

Contents

Abstract **iv**

Zusammenfassung **vii**

Acknowledgements **ix**

Contents **x**

List of Figures **xiii**

List of Tables **xv**

Abbreviations **xvi**

1 Introduction **1**
1.1 How is oogenesis and early embryogenesis molecularly controlled in Enoplean and Chromadorean nematodes? . . . 1
1.2 Evo-Devo and Next Generation Sequencing: The study of development and evolution in the times of quantitative high-throughput sequencing technologies . . . 3
1.3 *Caenorhabditis elegans*: An established model for embryogenesis . . . 6
1.4 Nematodes: A phylum with considerable wealth of developmental strategies 7
1.5 Oocyte-to-embryo transition is essential for early embryogenesis in all Metazoa . . . 9
1.6 Oocyte-to-embryo transition in *C. elegans* depends on MAP-kinase signaling 11
1.7 Anteroposterior axis formation and soma-germline separation take place very early in *C. elegans* . . . 14
1.8 Non-Caenorhabditis nematodes show considerable variances with respect to polarity establishment and axis formation . . . 15
1.9 RNAseq allows molecular analysis of development in non-model nematodes 16

2 Materials and Methods **18**
2.1 Materials and media . . . 18
2.2 Worm cultures and egg collection . . . 21
2.3 RNA extraction . . . 23
2.4 RNA extraction from samples with low RNA content . . . 24
2.5 Amplification of low amounts of total RNA by *in-vitro* transcription . . . 26

2.6 RNA sequencing . . . 29
2.7 Preparation of RNA interference experiments . . . 31
2.8 RNAi experiments of identified candidate genes involved in oocyte maturation in *C. elegans* . . . 31
2.9 Sanger sequencing . . . 32
2.10 Microarray analysis . . . 33
2.11 *In-situ* hybridisations . . . 35
2.12 Reference Genomes and transcriptomes . . . 38
2.13 OrthoMCL of nematode and outgroup species . . . 40
2.14 Programmes and tools for post-sequencing analysis . . . 41
2.15 OrthoMCL for early transcriptomes . . . 43
2.16 Genome predictions, allele analysis and single copy gene analysis in *Diploscapter coronatus* . . . 44

3 Results **47**
3.1 Microarray analysis in combination with RNAi finds new candidate genes for the *C. elegans* oocyte-to-embryo transition and the oocyte-maturation pathway . . . 47
3.2 RNAi of potential oocyte maturation regulators reveal candidate genes for the gene regulatory network . . . 54
3.3 *C. elegans* gonad formation, spermatogenesis and oogenesis indicate clade V-specific signaling events . . . 59
3.4 RNAseq of *R. culicivorax* virgins and males finds orthologues for oogenesis from *Drosophila* and *vertebrates* . . . 63
3.5 *Ascaris suum* germline-specific genes allow identification of new candidate genes relevant for the germline of *C. elegans* and other Chromadorean nematodes . . . 65
3.6 Establishment of RNAseq post-sequencing data analysis - a comprehensive workflow to generate transcriptomes of satisfying quality . . . 67
3.7 Sharing pathways with other Metazoa: Basal nematodes express signaling pathway components of invertebrates and vertebrates during early development . . . 72
3.8 Early BMP signaling exists in the clade I nematode *R. culicivorax* and the nematomorph *Gordius sp.* . . . 74
3.9 Wnt signaling via β-catenin paralogues WRM-1 and SYS-1 is specific for the genus *Caenorhabditits* . . . 77
3.10 Wnt pathway components exist in all studied nematodes and are expressed during early embryogenesis . . . 82
3.11 Notch signaling acts during early development in all nematodes, but specific core components seem to vary between Enoplean and Chromadorean nematodes . . . 84
3.12 NF-κB signaling may play a role during early development of *Gordius sp.* and *R. culicivorax* . . . 87
3.13 Crucial polarity-inducing factors are specific for the genus *Caenorhabditis* 89
3.14 Early polarity establishment involves Lethal giant larvae, Scribbled and Crumbs signaling in nematodes . . . 91
3.15 Hunchback, a highly conserved regulator and its role during Enoplean early development . . . 93

3.16 The parthenogenotes *Diploscapter coronatus* and *Panagrolaimus sp. PS1159* reveal unique transcriptomic features 95
3.17 *D. coronatus*, a product of interspecies hybridisation? 96

4 Discussion **103**
4.1 A Phylum-wide transcriptome comparison demonstrates expression of developmental regulators so far not known from nematodes 103
4.2 Gamete-to-embryo transition: a consequence of gonad formation, oogenesis, oocyte maturation and maternally expressed gene products 105
4.2.1 *C. elegans* oocyte maturation: Recruitment of genes originally involved in neuronal development? 106
4.2.2 Expression and genome analysis of the Enoplea *R. culicivorax* and the Chromadorea *A. suum* indicate new candidate genes associated with germline formation and gametogenesis in all nematodes 109
4.3 Polarity establishment and axis specification 113
4.3.1 Conserved planar cell polarity complexes may play a central role in polarity establishment and axis formation in nematodes 114
4.3.2 Wnt, BMP/DPP and NF-κB signaling may specify axes in Enoplean nematodes 116
4.3.3 The gap gene *hunchback* as first insight into anterior patterning of Enoplean embryos 120
4.4 Do modifiers of mTOR signaling facilitate polarity establishment in parthenogenetic nematodes? 121
4.5 Conclusion: Enoplean oogenesis and early embryogenesis appear to be more similar to outgroups than to Chromadorea 122

A Supplementary figures **125**

B Custom programmes **128**

C Supplementary tables **139**

Bibliography **158**

Declaration of Authorship **180**

Lebenslauf **182**

List of Figures

1.1 Phylogenetic tree of selected nematode species. 7
1.2 Schematic representation of the adult *C. elegans* gonad. 11
1.3 Schematic representation of the gene regulatory network regulating oocyte maturation. 12

2.1 Gel-photo visualising the quality of *Acrobeloides sp. PS1146* total RNA. . 28

3.1 Microarray QC plot for the knock-down experiments of *par-5(RNAi)* and *itr-1(RNAi) goa-1(RNAi)*. 48
3.2 Microarray QC plot for the knock-down experiments of *wt*, *lip-1(RNAi)*, *mek-2(RNAi)*, *mbk-2(RNAi)*, *mpk-1(RNAi)*. 49
3.3 MA plot for *mpk-1(RNAi)* versus wild-type microarray experiments. . . . 50
3.4 Scatter and volcano plot for gene expression of *mpk-1(RNAi) versus wild-type*. 51
3.5 RNAi results: Larvae hatching phenotypes for 8 candidates involved in oocyte-to-embryo transition after 72 hours. 56
3.6 Time courses of hatched larvae for different RNAi approaches. 57
3.7 Differential Interference Contrast images of selected RNAi phenotypes . . 58
3.8 Gonad devlopment from L_1 to adult, oogenesis and oocyte maturation in *C. elegans* 62
3.9 Heatmap visualising *Romanomermis culicivorax* virgin versus male differential expression. 63
3.10 Schematic representation of identified orthologues of genes involved in gonad formation, oogenesis, or spermatogenesis in *R. culicivorax*. 64
3.11 Comprehensive workflow of RNAseq analysis. 67
3.12 Sequential re-assembly of *D. coronatus* transciptomes 71
3.13 Expressed clusters and genes during early development for seven species. . 72
3.14 Representation of present and absent orthologues for components of the BMP signaling pathway. 75
3.15 Identification of early expressed orthologues for Drosphila BMP/DPP signaling components. 76
3.16 *In-situ* hybridizations for *R. culicivorax chordin*. 77
3.17 Existing orthologues of the *C. elegans* Wnt network in investigated nematodes. 78
3.18 Expressed orthologues for the *C. elegans* Wnt signaling network among transcriptomes of early embryogenesis of selected nematodes. 79
3.19 Phylogenetic tree for identification of orthologues for the β-catenins HMP-2, BAR-1, WRM-1, SYS-1. 81
3.20 Orthologues of the Human Wnt, RYK network in nematodes. 82

3.21 Early expression for orthologues of Human Wnt/RYK signaling components for selected nematodes. 83
3.22 Screening for absent and present orthologues of the *C. elegans* Notch signaling network among nematodes. 84
3.23 Early expression of orthologues of components of the *C. elegans* Notch signaling network for selected nematodes. 85
3.24 Screening for orthologues of the Human NF-κB signaling network among nematodes. 87
3.25 Early expressed orthologues of the Human NF-κB signaling network for selected nematodes. 88
3.26 Screening for orthologues of crucial *C. elegans* polarity establishment complexes among nematode species. 89
3.27 Early expression of orthologues of the *C. elegans* polarity establishment network for selected nematodes. 90
3.28 Screening for orthologues of Drosophila basoapical polarity complexes among nematodes. 91
3.29 Early expressed orthologues for the Drosophila Lgl (Lethal giant larvae), Scrib (Scribble) and Crb(Crumbs) signaling network for selected nematodes. 92
3.30 Screening for orthologues of Drosophila *hunchback* and interacting proteins among nematode species. 93
3.31 Early expression of orthologues for the Drosophila Hunchback network for selected nematodes. 93
3.32 *R. culicivorax in-situ* hybridisations showing *hunchback* expression during early embryogenesis. 94
3.33 Protein alignments of LAMTOR4/5 for *D. coronatus* and *Panagrlaimus sp. PS1159*. 96
3.34 Orthology clustering of clade III, IV, and V nematode species. 98
3.35 Analysis of *D. coronatus* ITS sequences. 100
3.36 Analysis of *D. coronatus* allele frequency and single copy genes. 101
3.37 Alignment of the genomic loci for the single copy gene F53G2.3 in *D. coroantus* 102

4.1 Potential function of identified genes involved in the gene regulatory network of oocyte maturation in *C. elegans*. 108

A.1 Venn-diagram representing the orthology clustering of early transcriptomes. 126
A.2 Phylogentic tree for Human RYK orthologs. 127

List of Tables

2.1 Amplification efficiency of the Ambion Message AMP II kit for 20 ng and 2,000 ng of total RNA. 28
2.2 Quantity of created TrueSeq libraries. 29
2.3 Results of sequencing by Illumina platforms. 30
2.4 Sanger sequencing primers used in this thesis. 32
2.5 Programme of the sequencing PCR. 32
2.6 Testing of correlation between $\log_2$ expression foldchange (M) and mean of $\log_2$ expression (A). 35
2.7 PCR programme for amplification of *R. culicivorax chordin* and *hunchback*. 36
2.8 Primers used for amplification of the orthologues for *chrodin* and *hunchback*. 38
2.9 Nematode genomes used in this thesis. 39
2.10 Outgroup genomes used in this thesis. 39
2.11 Reference transcriptomes used in this thesis. 40
2.12 Nematode genomes used for clustering orthologues between *D. coronatus* 6 *Caenorhabditis* species and outgroups. 45

3.1 Foldchange correlation data for all RNAi knock-downs. 52
3.2 Candidate genes found by RNAi knock-downs and microarray analysis . . 53
3.3 Chromadorean-specific germline regulators found in *A. suum* 66
3.4 Number of retrieved transcripts for each transcriptome for 5 nematode species. 70
3.5 Screened pathways and processes for early development in Enoplean and Chromadorean nematodes. 74

C.1 RNAi and microarray candidates screened for known RNAi phenotypes . 139
C.2 Significantly enriched GO-terms for *Gordius sp.* 1-cell stage. 146
C.3 Significantly enriched GO-terms for *R. culicivorax* 1- to 4-cell stage. . . . 154

Abbreviations

°C	**degree celsius**
μ**l**	**m**icro **l**iter
μ**g**	**m**icro **g**ramm
a-p axis	**a**nterio**p**osterior **axis**
BCIP	5-**b**romo-4-**c**hloro-3-**i**ndonyl **p**hosphate
CaCl	**ca**lcium **c**h**l**oride
cm	**c**enti **m**eter
cDNA	**c**omplementary **DNA**
DAPI	4'6-**dia**idino-2-**p**henyl**i**ndole
DIC	**D**ifferential **I**nterference **C**ontrast
DEPC	**die**thyl**p**yro**c**arbonate
DEPC-H_2O	**DEPC**-treated water
Drosophila	***Drosophila*** *melanogaster*
dsRNA	**d**ouble**s**tranded **RNA**
d-v axis	**d**sorso**v**entral **axis**
DIG	**dig**oxygenine
dH_2O	**d**estilled **H_2O**
DNA	**d**eoxyribo**n**ucleic **a**cid
dNTP	**d**eoxyribo**n**ucleotid**t**ri**p**hosphate
edgeR	**e**mpirical analysis of **DGE** in **R**
EDTA	**e**thyen**di**a**t**etra**a**cetic acid
e.g.	**e**xempli **g**ratia; for example
EtOH	**et**han**o**l
Evo-Devo	**evo**lution and **dev**lopment
fc	**f**old**c**hange

Fig.	**fig**ure
g	**g**ramm
G	**g**-force
GFP	**g**reen **f**luorescent **p**rotein
GO	**g**ene **o**ntology
GRN	**g**ene **r**egulatory **n**etwork
Gu-Mix	**gu**adinium **mix**
GuSCN	**gu**adiniumthiocyanate
IPTG	**i**so**p**ropyl-β-D-**t**hio**g**alactopyranoside
ISH	***i**n-**s**itu* **h**ybridisation
ITS	**i**nternal **t**ranscribed **s**pacer rRNA
IVT	***i**n-**v**itro* **t**ranscription
KH_2PO_4	potassiumdihydrogene phosphate
KPO_4	potassium phosphate buffer
LSU	**l**arge **su**bunit of ribosomal RNA
LPA	**l**inear **p**olyacryl**a**mide
LB	**L**uria **B**ertani medium
M	**m**olar
MAP kinase	**m**itogen **a**ctivated **p**rotein kinase
MeOH	**m**ethanol
mg	**m**illi **g**ramm
min	**min**ute
$MgCl_2$	**m**a**g**nesium **c**h**l**oride
$MgSO_4$	**m**a**g**nesium **s**ulfate
ml	**m**illi **l**iter
mM	**m**illi **m**olar
MOC	**m**icrotuble **o**rganising **c**enter
mpk-1	**m**ap **k**inase 1
mRNA	**m**essenger **RNA**
MSP	**m**ajor **s**perm **p**rotein
NaOCl	sodium hypochloride solution
Na-acetate	sodium acetate
NaCl	sodium chloride

Na_2HPO_4	disodiumhydrogene phosphate
NBT	**n**itro **b**lue **t**etrazolim chloride
NCBI	**N**ational **C**enter for **B**iotechnology **I**nformation
NGS	**n**ext **g**eneration **s**equencing
NGM	**n**ematode **g**rowth **m**edium
PBS	**p**hosphate **b**uffered **s**aline
PcG	**p**oly**c**comb **g**roup genes
PCP	**p**lanar **c**ell **p**olarity pathway
PGC	**p**rimordial **g**erm **c**ell
QC	**q**uality**c**ontrol
RMA	**r**obust **m**ultiarray **a**varage
RNA	**r**ibo**n**ucleic **a**cid
rpm	**r**ounds **p**er **m**inutes
rRNA	**r**ibosomal **RNA**
RNAi	**RNA i**nterference
RNAseq	second generation **RNA seq**uencing
RSEM	**R**NAseq by **e**xpection **m**aximation
RSF1	**R**emodeling and **s**pacing **f**actor 1
RT	**r**oom **t**emperature
sec	**sec**ond
sd	**s**tandard **d**eviation
SSC	**s**aline **s**odium **c**itrate
SSU	**s**mall **su**bunit of ribosomal RNA
Tab.	**tab**le
TMM	**t**rimmed **m**ean of **m** values normalisation method
TN	**t**ris **N**aCl buffer
tRNA	**t**ransfer **RNA**
Tri-mix	mix consisting of guadiniumthiocyanate, acid phenol and β-mecaptoethanol
Xenopus	***Xenopus*** *laevis*
zebrafish	*Danio rerio*

Chapter 1

Introduction

1.1 How is oogenesis and early embryogenesis molecularly controlled in Enoplean and Chromadorean nematodes?

Oogenesis and early embryogenesis are crucial biological processes for formation of the embryo and propagation of the next generation in most Metazoa. During both processes important prerequisites must be met: The embryo must be supplied with maternal gene products (e.g. mRNAs and proteins) and nutrients, polarity needs to be established and the prospective axes need to be specified. In order to understand how this is managed, it is necessary to understand oogenesis and early embryogenesis. Oogenesis can be subdivided into the biological processes of oogonia and oocyte formation, oocyte maturation (Kimble and Crittenden 2007) and oocyte-to-embryo transition (Govindan et al. 2006). Oogenesis ends with fertilisation by sperm (where sperm is present) and is followed by early embryogenesis. All these biological processes are spatially and temporally regulated, but the exact time points and the molecular mechanisms controlling these processes vary among different taxa (McCarter et al. 1999, Govindan et al. 2006, Kimble and Crittenden 2007, Roth and Lynch 2009, Clift and Schuh 2013, Li and Albertini 2013).

For the nematode model *C. elegans* oogenesis and early embryogenesis are very well understood (Miller et al. 2003, Gönczy and Rose 2005, Govindan et al. 2006, Kimble and Crittenden 2007). It has also been elucidated how important decisions, such as polarity establishment and axis specification, are controlled in *C. elegans*: The entry of

the sperm initiates a gene regulatory network (GRN) to induce polarity in the 1-celled egg (for a review refer to Gönczy and Rose 2005). This is a prerequisite for axis specification, soma-germline separation during the first cleavage (Gönczy and Rose 2005) and its invariant embryogenesis (Sulston et al. 1983). However, for other nematodes it was shown that polarity establishment can be independent of the sperm-entry point (Goldstein et al. 1998) and sperm may not even be required (as in parthenogenetic nematodes; Lahl et al. 2006). In this respect cleavage patterns and cell specification during early embryogenesis can be very different in comparison to *C. elegans* (Skiba and Schierenberg 1992, Wiegner and Schierenberg 1998, Wiegner and Schierenberg 1999, Laugsch and Schierenberg 2004, Schulze and Schierenberg 2008, Lahl et al. 2009, Schulze and Schierenberg 2009, Schulze and Schierenberg 2011, Schulze et al. 2012). Some representatives of the Enoplea (phylogenetically rather basal nematodes; Fig. 1.1; De Ley 2006) show no polarity at all and blastomeres can adopt variable cell fates (Voronov and Panchin 1998). This shows that early embryogenesis is very different among nematodes and raises several questions: How and when are important developmental decisions like polarity establishment and axis specification made in different nematodes? When does polarity establishment occur? Prior, during or after fertilisation? Which pathways and GRNs drive oogenesis and early embryogenesis?

In order to address these questions in this thesis, I used an unbiased transcriptome analysis approach for oogenesis and early embryogenesis in selected nematode species (Fig. 1.1; Blaxter et al. 1998). This approach is unbiased in the sense that specimen were selected (here embryos) from wilde-type populations without experimental influences (neither physical, e.g. mechanical stress, nor chemical, e.g. application of drugs or RNA interference). My main objective in this thesis is a transcriptome comparison of Chromadorea (phylogenetically rather derived nematodes; Fig. 1.1) and Enoplea (phylogenetically more basal), with respect to oogenesis and early embryogenesis. For a proper comparison I used the model *C. elegans* as a reference and I added the nematomorph *Gordius sp.* as an outgroup. Additionally, I used available nematode genomes from all clades of the phylum to confirm genomic presence and absence of orthologous transcripts. My comparative transcriptome approach allows searches for expressed orthologues by two strategies: (1) I can search for expressed orthologues of known pathways from the model *C. elegans*, or (2) I can use outgroup models such as fruit fly (*Drosophila melanogaster*), claw frog (*Xenopus laevis*), zebrafish (*Danio rerio*), mouse (*Mus musculus*) and humans (*Homo sapiens*) to search for commonalities to Enoplea

and Chromadorea (including *C. elegans*).

For this unbiased transcriptome comparison I used RNA analysis techniques such as microarrays and RNA sequencing (RNAseq) to identify orthologous transcripts which are expressed during oogenesis and early embryogenesis. What are these techniques and how are these applied to get information which helps with the objective of this thesis? In the introduction I will give background information on what is known concerning oogenesis and early embryogenesis in nematodes. I will explain how this mainly developmental biological topic can be put into an evolutionary perspective and how RNA analysis techniques may help to understand the molecular underpinnings of oogenesis and early embryogenesis. Furthermore, I will explain how my transcriptome comparison may help to identify Enoplea- and Chromadorea-specific pathways and regulators during oogenesis and early embryogenesis. I will also explain how such insights may add to the understanding of the molecular mechanisms of polarity establishment and axis formation in these nematode groups.

1.2 Evo-Devo and Next Generation Sequencing: The study of development and evolution in the times of quantitative high-throughput sequencing technologies

In order to deal with developmental questions for inter-species comparison it is necessary to put them into a evolutionary perspective. Contributing to the main question in this thesis (how is oogenesis and early embryogenesis controlled in nematodes) needs understanding of (1) what we know about evolution (2) how can I combine it with the concept of developmental biology and (3) what methods can we use to generate data to answer this question. Here, I give a short overview about the study of "Evolution and Development" (abbreviated as "Evo-Devo") and how it benefited in the past from crucial findings and how the underlying theoretical concept and advanced technologies can be used to understand polarity establishment.

In the past the biological discipline of Evo-Devo dealt with the understanding of developmental cues which emerged in the course of evolution (for a review on Evo-Devo and current scientific challenges refer to Müller 2007). The first pioneer studies in nematodes

were based on classical approaches such as mechanical manipulation of early embryos (Boveri 1899, 1910), UV-radiation (Nelson et al. 1989, Stevens 1909) or laser ablations (Kimble and White 1981, Schierenberg 1984).

These pioneer experiments unravelled fundamental principles of development, such as the identifying the centriole as the "microtuble organising center" (MOC) (Boveri 1887, Wilson 2008), the chromosomes as stable entities and the process of chromatin diminution in somatic but not germline cells in *Ascaris* (Boveri 1887). However, these experiments lacked the knowledge of genes, DNA and their function.

Only via mutagenesis and screening for loss of function (lof) alleles, scientists were able to pinpoint how genes interact with one another and form gene regulatory networks (GRN), which drive development in model organisms such as *Drosophila melanogaster* (Nüsslein-Volhard and Wieschaus 1980) or *C. elegans* (Brenner 1974). These experiments allowed a broad understanding of genes, pathways and GRNs.

It did not took long before it became clear that certain developmental pathways, for example Wnt, Hedgehog, and Hox signaling, were conserved among invertebrate and vertebrate models (Gerhart 1999, Aboobaker and Blaxter 2003a,b). Nonetheless, a comparison of genes involved in development between different species could only be achieved by putting the newly gained information into an evolutionary perspective. Hence, the discipline of Evo-Devo was born.

Mutagenesis studies brought a unprecedented knowledge of developmental control and showed which developmental key regulators are conserved in model organisms (Gerhart 1999). A serious cutback was, however, that these studies had been restricted to a small number of model systems. Hence, the understanding of evolution of GRNs among Metazoans still lacked important information due to insufficient species sampling in the different animal phyla.

Recently, the methodology of Sanger, which dominated the market since the 1970s, was replaced by sophisticated multiplexing sequencing platforms which allowed sequencing of millions of reads in parallel and thus analysis of whole genomes of newly sampled species even in laboratories with tight budget. Among the first of such techniques were the Roche 454, the SOLiD and the Illumina platform (so-called second generation sequencing platforms; Strausberg et al. 2008). Recently, third generation sequencing platforms such as PacBio, Biotorrent and Oxford Nanopore promise higher numbers and longer sequencing reads (Niedringhaus et al. 2011). These changes in sequencing techniques led to the new era of Next Generation Sequencing (NGS), which allows sequencing and

assembling of whole genomes of genetically completely undescribed species.

Today, the wealth of genomic data and an extended spectrum of species which can be used for analysis opens the door to better understand evolutionary aspects of taxon-specific differences. Nonetheless, the unprecedented wealth of data comes with a very basic shortcoming: While the amount and complexity of data increased dramatically, it was impossible to compare whole genomes and relevant GRNs among species by pure sequence alignments. Today, Genbank includes more than 360 million gene sequences from more than 300,000 species (Benson et al. 2015). The combination of phylogenetics, bioinformatics and data mining techniques allows now a reliable approach to analyse these huge and complex data sets.

What can we learn from genomic sequencing for Evo-Devo? NGS allowed a first view into the genomic background of a board diversity of animals to understand phyologentic characteristics of whole phyla (for example in the case of the avian phylogenomics project; Zhang et al. 2014). But especially Developmental Biology (and Evo-Devo) is interested in crucial developmental stages. In this respect, RNA sequencing by Next Generation Sequencing platforms (RNAseq) is a very advantageous method, as it allows complete sequencing of various developmental stages on the messenger RNA (mRNA) level. So, it is possible to correlate transcriptomes directly to specific developmental stage of the studied species.

In my thesis I sought to better understand similarities and differences of GRNs, which drive the early development of selected nematode species, by applying RNAseq.

1.3 *Caenorhabditis elegans*: An established model for embryogenesis

As explained in the previous section, Developmental Biology and Evo-Devo benefited from model organisms in the past, as forward and reverse genetics could be established in these species and strongly contributed to the understanding of gene regulatory networks (GRNs; Nüsslein-Volhard and Wieschaus 1980, St Johnston 2002). In the case of nematodes, Sydney Brenner introduced *Caenorhabditis elegans* initially as a model for genetic analysis of the nervous system (Brenner 1974). Later, reverse genetics was also established in form of RNA interference (RNAi; Fire et al. 1998, Tabara et al. 1998, Timmons and Fire 1998), different techniques for transformation (Boulin and Bessereau 2007, Robert and Bessereau 2007) and the CRISPR/Cas9 system for gene and genome editing (Chen et al. 2013). *C. elegans* became the first Metazoan organism with a completely sequenced genome (*Caenorhabditis elegans* Sequencing Consortium 1998).
This broad set of methods and techniques not only led to the understanding of many fundamental cellular, genetic and developmental aspects of the biology in general, it also led to the nobelprizes for the genetic analysis of organ development and programmed cell death (Brenner 1974, Sulston 1974, 1976, Sulston and Horvitz 1977, Sulston et al. 1983), establishment of the RNA interference (RNAi) technique to selectively knock-down gene function (Fire et al. 1998, Tabara et al. 1998, Timmons and Fire 1998) and the use of green fluorescent protein (GFP) to visualise gene expression *in vivo* (Shimomura et al. 1962, Chalfie et al. 1994, Heim et al. 1995, Shaner et al. 2005).
Hence, *C. elegans* became one of the most famous model organisms studied by more than 1000 laboratories around the globe (wormbase.org/resources/laboratory; version WS246). Just in the year 2014 there were more than 1,500 papers published which used information gathered by experiments using *C. elegans* (pubmed.org).
Eventhough, *C. elegans* is a very important model for numerous biological aspects, especially for Evo-Devo, several studies indicated doubt whether it is representative for the whole phylum of nematodes. Many differences between *C. elegans* and other nematodes have been found concerning cellular aspects of embryogenesis (Skiba and Schierenberg 1992, Goldstein et al. 1998, Voronov and Panchin 1998, Wiegner and Schierenberg 1998, Wiegner and Schierenberg 1999, Laugsch and Schierenberg 2004, Schulze and Schierenberg 2008, Schulze and Schierenberg 2009, Lahl et al. 2009, Schulze and Schierenberg

2011, Schulze et al. 2012) and loss of gene pathways such as Hedgehog signaling (personal communication P. H. Schiffer, M. Kroiher; Hao et al. 2006) and Hox-gene clusters (Aboobaker and Blaxter 2003a,b).

Hence, in this thesis I want to make a contribution to the understanding of how embryogenesis is controlled in these nematodes and what pathways may play a role.

1.4 Nematodes: A phylum with considerable wealth of developmental strategies

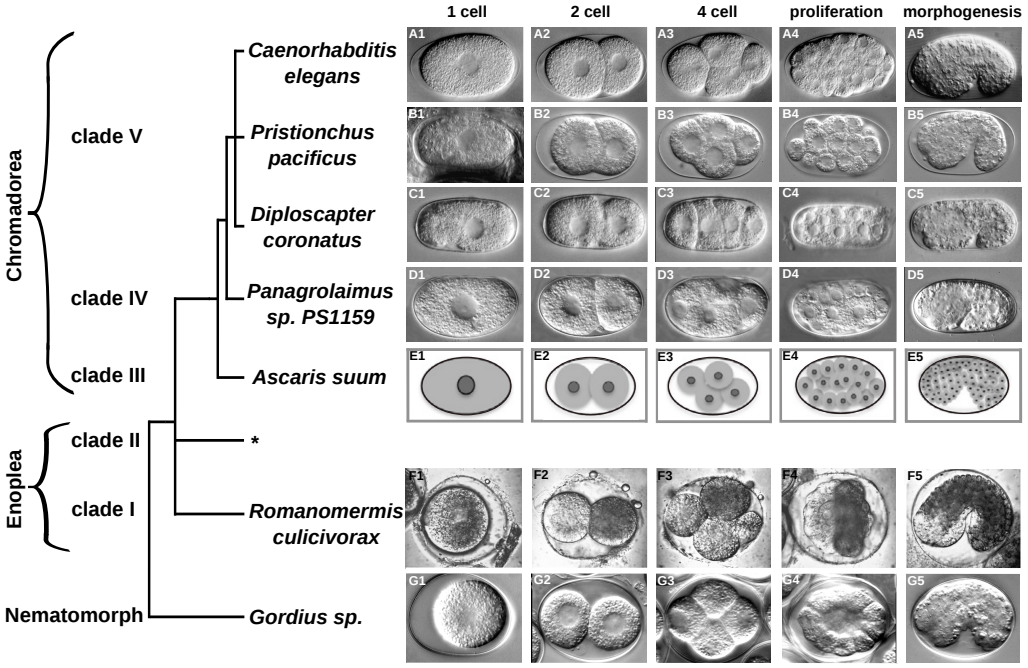

Figure 1.1: Phylogenetic tree of selected nematode species. In this thesis I focus on the early embryogenesis of the Chromadorean species *Caenorhabditits elegans* (clade V), *Diploscapter coronatus* (clade V), *Pristionchus pacificus* (clade V), *Ascaris suum* (clade III), the Enoplean species *Romanomeris culicivorax* (clade I) and the nematomorph *Gordius sp.* (outgroup). The phylogenetic tree was modified from Blaxter et al. 1998. Due to insufficient species from clade II, which cannot be cultured under laborartory conditions, I could not include clade II species (indicated by *). *A. suum* embryonic pictures were taken from Wang et al. 2014. The differential interference contrast (DIC) pictures for all other species were taken in the Schierenberg laboratory (J. Schulze, E. Schierenberg personal communication). Anterior is always left and dorsal at the top.

The phylum of nematodes is one of the most species-rich phyla in the kingdom of animals. There are estimations of several million to 100 million species (Meldal et al. 2007). Nematodes occupy all conceivable biological niches, they exist in extreme habitats such

as Antarctica (Wharton and Ferns 1995, Wharton 2003), African deep mines (Borgonie et al. 2011), in any soil, fresh and salt water and on decaying fruits, as for example the genus *Caenorhabditis* (Kiontke et al. 2011). Besides from that nematodes have an enormous impact on our lives, as they are known to be crop-pests (Trudgill and Blok 2001), parasites of domestic animals and introduce health threatening diseases to the human population, such as lymphatic filariasis or river-blindness (Fenwick 2012, Ottesen et al. 2008).

Nematodes occupy different ecological niches and follow different modes of early development (Fig. 1.1; Goldstein et al. 1998, Sommer 2001, Laugsch and Schierenberg 2004, Schulze and Schierenberg 2008, Lahl et al. 2009, Schulze and Schierenberg 2009, Schulze and Schierenberg 2011, Schulze et al. 2012, Sommer 2012). Depending on the positioning of these species in the phylum of nematodes, early development can be rather different from *C. elegans*. Yet, nearly all nematodes look very much alike during morphogenesis and when they hatch as larvae (Fig. 1.1). Remarkably, the clade II nematode *Enoplus brevis* passes through, equal division during early development (Voronov and Panchin 1998). Furthermore, it was shown previously that another clade II nematode, *Tobriulus stefanskii* shows atypical gastrulation in comparison to the model *C. elegans* and instead shows similarity to gastrulation known from vertebrates (Schulze and Schierenberg 2008; Schierenberg personal communication; Fig. 1.1). Only recently scientists started to investigate the genetic basis of these striking developmental differences (Schiffer et al. 2014; personal communication M. Kroiher, J. Camps, N. Nsah).

So, nematodes are species rich and show certain unique characteristics with respect to early embryogenesis, but the molecular underpinnings still need to be tapped. Here, I chose a set of seven species which cover 4 of 5 clades (Fig. 1.1) which I intend to analyse during early embryogenesis. Transcriptomic data from these species shall give insights into early embryogenesis and may help to understand which contributing pathways exist specifically in Enoplean and/or Chromadorean nematodes.

1.5 Oocyte-to-embryo transition is essential for early embryogenesis in all Metazoa

As mentioned in section 1.1, the time point of polarity establishment and primary (a-p) axis formation might be prior, during or after fertilisation in certain species, hence it is important to understand oogenesis.

Oogenesis includes oocyte maturation and oocyte-to-embryo transition and the oocyte is subsequently fertilised by sperm (for reviews on this topic refer to Sardet et al. 2007, Clift and Schuh 2013, Li and Albertini 2013 for chordates and vertebrates; for invertebrates refer to Roth 2001, Roth and Lynch 2009, Costache et al. 2014).

All three processes are important for the complex differentiation of the oocyte. This egg cell must pass through specific "milestones" prior and after fertilisation to guarantee the development and survival of the next generation: (1) The oocyte incorporates maternal gene products to allow early cleavage prior to activation of zygotic transcription. (2) The oocyte forms a specific chromatin structure, the synaptonemal complex to allow continuity of transcription during meiotic arrest. (3) In most animals, except mammals, the oocyte incorporates nutrients for embryonic metabolism. (4) Incorporated maternal information modifies chromatin conformation (e.g. via histone-methylation) to establish a genomic imprinting (epigenome) for the embryo. (5) The oocyte establishes an mechanism to prevent polyspermy (Li and Albertini 2013).

The described milestones possess general similarity in all investigated organisms, but the details of accomplishing the milestones of oogenesis, oocyte maturation and fertilisation are very different among organisms: Taxon-specific differences include the stage of meiotic arrest, the amount and types of incorporated maternal RNA, incorporation of yolk and nutrients and composition of the oocyte surface.

Furthermore, the specific structure of the gonad may have important implications for the oocyte, e.g. establishment of polarity in *Drosophila* (Roth and Lynch 2009). Here, as well as in *C. elegans* the oocyte interacts with the surrounding somatic cells (Govindan et al. 2006, Roth and Lynch 2009). Similarly, mammalian oocytes also form cell-cell contacts with the surrounding somatic cells (Li and Albertini 2013). Hence, oocyte-soma interactions may have an effect on early embryogenesis and polarity establishment and need to be investigated.

Despite the very general conformities among all animals, the mechanisms, of how these

different processes are achieved, are different among organisms. All the explained processes were investigated to a rather high level in model organisms, hence it is possible to use these models as references to compare nematodes with. In the following sections, I will first emphasize what we know from the model *C. elegans* about oocyte-to-embryo transition. Second, I will address the question: Which GRNs regulate oogenesis and oocyte-to-embryo transition *C. elegans*?

1.6 Oocyte-to-embryo transition in *C. elegans* depends on MAP-kinase signaling

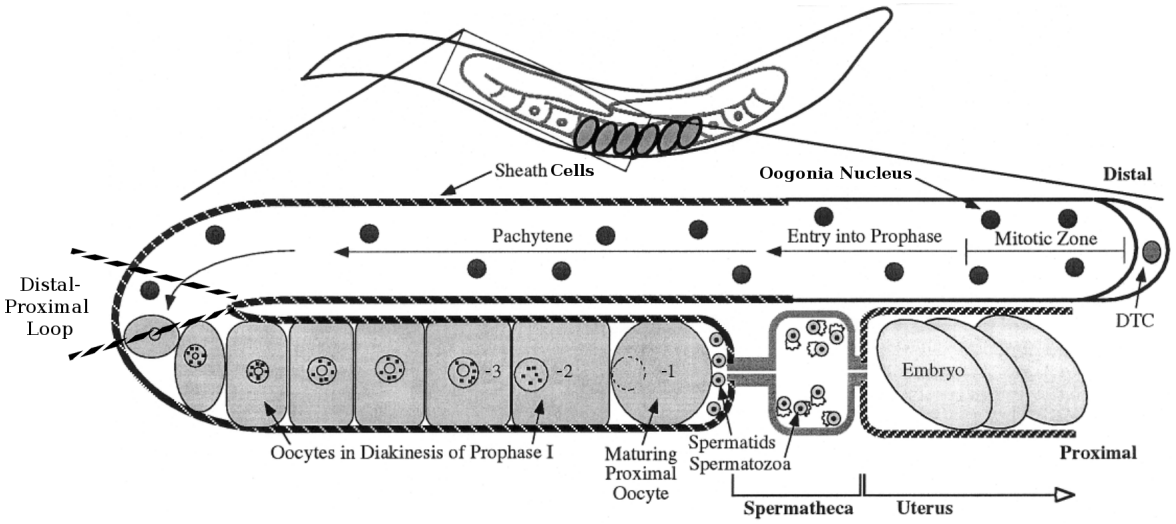

Figure 1.2: Schematic representation of a gonad of an adult *C. elegans* hermaphrodite. Oogonia (located in the distal arm of the gonad) are indicated by black dots. Box-shaped oocytes are localised in the proximal gonad arm and are fully surrounded by cell membranes (grey). The most proximal oocyte (indicated as "−1 oocyte") undergoes oocyte maturation (grey elipsoid). The nuclear envelopment breakdown (NEBD) of the nucleus is indicated by a dashed lines in the "−1 oocyte". Modified after McCarter et al. (1999).

The *C. elegans* gonad resembles in its morphology and its function an assembly line (McCarter et al. 1999, Greenstein 2005, Hubbard and Greenstein 2005): Oocyte precursors, so-called oogonia, are located at the very tip of the distal end of the gonad arms (Fig. 1.2). In this area oogonia interact via cell-cell interactions with the distal tip cell (DTC). The DTC forms cell-cell contacts with the mitotically dividing oogonia (Kimble and Crittenden 2007). The oogonia initiate meiosis by moving towards the proximal end of the gonad and away from the DTC (Fig. 1.2).

For the oogonia meiosis starts with prophase and they undergo pachytene arrest (Fig. 1.2). After migrating through the loop connecting the proximal with the distal gonad arm (Fig. 1.2), oogonia start to form cell membranes and become completely encapsulated. Furthermore, this transition is also marked by moving from pachytene to diplotene arrest and the former oogonia become oocytes. The arrested oocytes are stacked as cuboid cells in the proximal gonad arm (McCarter et al. 1999; Fig. 1.2). All oocyte resemble this shape with exception of the most proximal "−1 oocyte" (Fig. 1.2). This oocyte starts maturing, meaning that the diplotene arrest is lifted and the oocyte continues

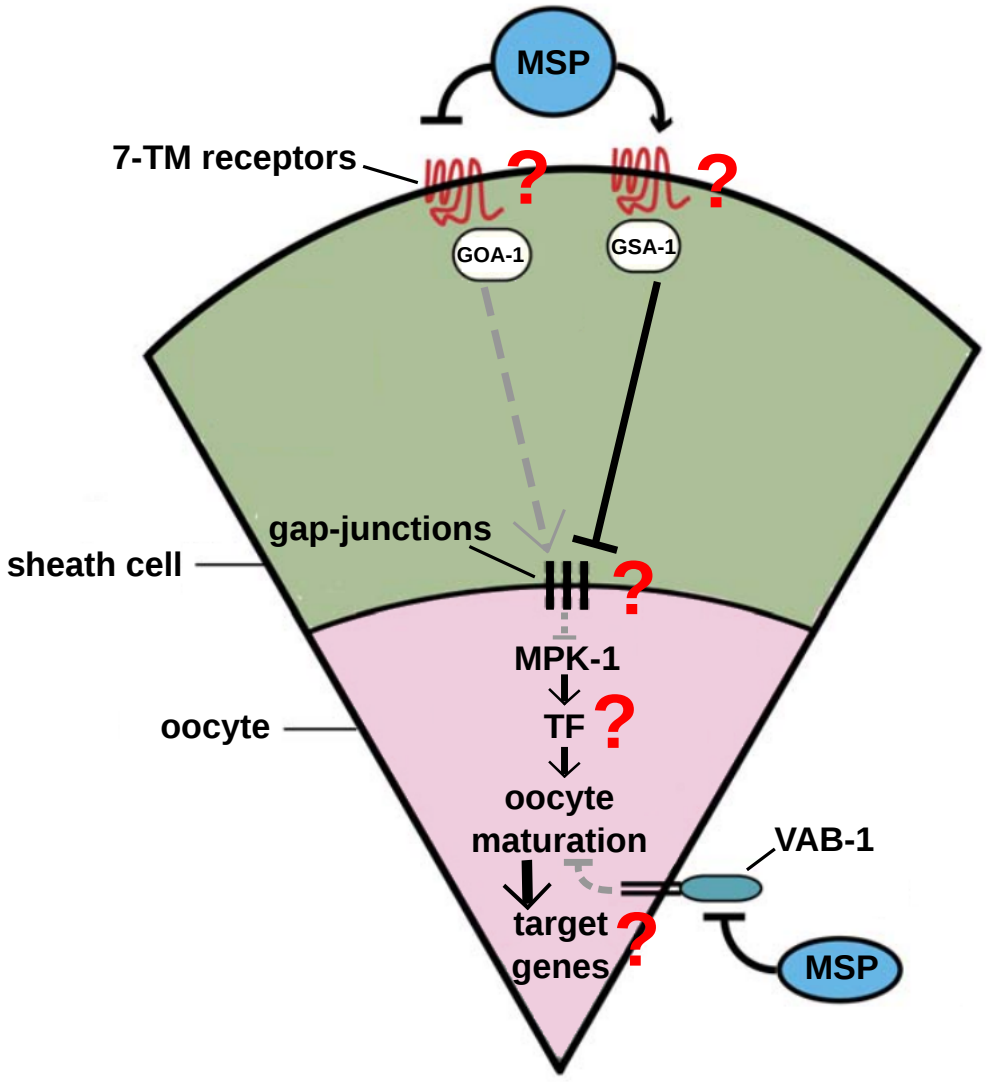

Figure 1.3: Schematic representation of the cell-cell interaction between somatic sheath cells and the "-1 oocyte" and the underlying GRN which regulates the process of oocyte maturation. Grey dashed lines indicate activating (arrow head) and inhibiting (blunt end) signal of oocyte maturation under the condition of absent sperm and MSP. However, under the condition of present sperm, MSP binds to seven-transmembrane receptors and the ephrin receptor VAB-1 and activates the MAP-kinase MPK-1. MPK-1 induces the process of oocyte maturation via an unknown transcription factors (TF). Unknown receptors, gap-junction molecules, transcription factors and downstream target genes are indicated by red question marks. Modified after (Govindan et al., 2006).

with meiosis. Morphologically, this process is characterized by the nuclear envelope breakdown (NEBD; McCarter et al. 1999).

The process of oocyte maturation is regulated by (1.) mitogen-activated protein (MAP) kinase pathway (with MPK-1 as the *C. elegans* orthologue) and (2.) by the cell-cell interactions between the somatic sheath cells of the gonad and the maturing oocyte at the "−1" position (Fig. 1.2; Miller et al. 2001, 2003, Govindan et al. 2006). The necessary cell-cell interactions take place via innexin gap junctions between sheath cells and

the oocyte (Fig 1.3). The sperm induces oocyte maturation via secreted Major Sperm Protein (MSP). MSP binds to the ephrin receptor VAB-1 of the oocyte as well as to presently unknown seven-transmembrane receptors which are coupled to the G-protein subunits GOA-1 and GSA-1 (Fig 1.3; Govindan et al. 2006). This interaction between sperm and oocyte ensures that the diplotene arrest is lifted as long as sperm is available. Thus, it is known that two cues are responsible for the maturation of the oocyte: 1.) Cell-cell interaction via gap-junctions between the sheath cells of the somatic gonad and 2.) the excretion of the the MSP by the sperm.

Oocyte maturation is intrinsically suppressed by the G-protein *alpha* subunit GOA-1. In the case of sperm being present MSP is secreted and binds to unknown seven-transmembrane receptors and activates the G-protein *alpha* subunit GSA-1 which activates MAP kinase signaling, while GOA-1 is simultaneously deactivated by the binding of MSP. In parallel to that MSP binds to the ephrin receptor VAB-1 (Fig 1.3; Miller et al. 2003). This part of the pathway activates the oocyte maturation process without interacting with the MAP-kinase pathway. Besides from this GAP-junction interactions with the somatic gonad are inhibited by the binding of MSP, avoiding inhibition of oocyte maturation (Fig 1.3; Miller et al. 2003, Govindan et al. 2006).

Even though, the GRN regulating oocyte maturation is elucidated to a very high degree, certain key regulators of the pathways are still unknown (Fig. 1.3; marked by red question marks): Possibly, not all of cell-cell interaction proteins and/or gap-junction components which interact with the somatic gonad are elucidated, yet. Furthermore, the seven-transmembrane receptors, as well as the acting transcription factor and the downstream target genes of the GRN have not been identified (Fig. 1.3; red question marks).

In order to identify potential upstream and downstream candidate genes for the MAP-kinase pathway involved in oocyte maturation and oocyte-to-embryo transition we used RNAi of known pathway components to find differentially over- and/or underexpressed genes using microarrays (see Materials and Methods).

1.7 Anteroposterior axis formation and soma-germline separation take place very early in *C. elegans*

How is polarity established in *C. elegans*? In *C. elegans* the sperm entry point determines the posterior pole and thus anteroposterior polarity of the embryo. It also initiates the asymmetric division of the zygote into a larger somatic AB and a smaller germline cell P_1 (Deppe et al. 1978, Gönczy and Rose 2005).

This process induces a directed cortical flow, leading to an anterior positioning of the PAR-3/-6/PKC-3 complex and a posterior positioning of the PAR-2/-5 complex. Especially, the first complex is conserved in nearly all Metazoa and an essential part of the so-called planar cell polarity (PCP) pathway (Knust and Bossinger 2002). The differential positioning of both complexes results in the posterior positioning of the LET-99 protein at the cortex of the zygote (Gönczy and Rose 2005). This positioning of the regulator LET-99 leads to the tethering of the Gα subunit to the posterior membrane of the zygote. Gα binds GPR-1/-2 and the microtubule network (Nguyen-Ngoc et al. 2007). The latter generates an unequal pulling force onto both poles of the mitotic spindle, thus moving the separating chromosomes from the center of the egg towards a more posterior position. This acentric position marks the future division plane of the zygote and thus leads to asymmetric cell division, giving rise to the larger anterior AB cell (soma) and the smaller posterior P cell (germline). Hence, in contrast to most other animal taxa both, establishment of the a-p axis and soma-germline separation, are accomplished prior to the cleavage of the *C. elegans* zygote.

As indicated previously, early development in other nematodes deviates from *C. elegans* with respect to polarity establishment and early embryogenesis. Here, I use *C. elegans* as a reference system to better understand molecularly what are the differences to other nematodes.

1.8 Non-Caenorhabditis nematodes show considerable variances with respect to polarity establishment and axis formation

Different studies in the past showed that unlike *C. elegans* the sperm-entry point is not responsible for asymmetric cell division and anteroposterior axis formation in certain other nematode species (Goldstein et al. 1998, Lahl et al. 2006, 2009). Such species appear to exist in all clades of the nematode phylum: The best examples are representatives of clade III and IV, like *Arobeloides* and *Panagrolaimus* species (Goldstein et al. 1998). There is even a close relative of *Caenorhabditis*, the parthenogenote *Diploscapter coronatus* with a very different mechanism of establishing of anteroposterior polarity without the presence of sperm (Lahl et al. 2006).

In such species asymmetry within the zygote has to be controlled by a different molecular mechanism compared to *C. elegans*. Furthermore, the above mentioned examples also show variances with respect to early development. In the case of *D. coronatus* the 4-cell stage is not rhomboid as in *C. elegans*, instead the cells are set-up in a linear sequence without direct cell-cell contact between ABp and the P_2 cell (Fig. 1.1 B3; Lahl et al. 2006). In *C. elegans* this contact is crucial to induce the fate of the ABp cell via Notch signaling from the germline. In *D. coronatus* such an induction by Notch signaling is impossible due to the lack of cell-cell contact (Fig. 1.1 C3).

Other nematodes such as *R. culicivorax* (clade I) rearrange cells into a rhomboid 4-cell stage as in *C. elegans*, although spindles in the equal sized 2-cell blastomeres are both oriented longitudinally. In the 4-cell stage, fates of the blastomeres are reversed along the dorsoventral axis compared to *C. elegans* (Schulze and Schierenberg 2008, 2009), i. e. the dorsal cell makes mesoderm and the ventral one ectoderm.

This shows that crucial differences between *C. elegans* and other nematodes exist and *C. elegans* is an appropriate reference to compare other nematodes with.

1.9 RNAseq allows molecular analysis of development in non-model nematodes

By analysing development in detail on the cellular level, our laboratory revealed significant differences among selected nematodes, which include (1) differences in the establishment of polarity, (2) a shift from maternally supplied fate determinants towards zygotically expressed determinants, (3) considerable variations of gastrulation (Schulze and Schierenberg 2011). In addition, we showed that spatio-temporal expression of genes involved in cell fate specification is surprisingly variable among nematodes, indicating that observed developmental diversity correlates with differences in molecular signaling (Schiffer et al. 2014; personal communication Ndifon Nsah, Julia Camps and Michael Kroiher, Köln). All these findings support our view that development in *C. elegans* is not typical for nematodes in general.

Although it is known for many developmental stages in *C. elegans* which molecular pathways determine certain cell fates, the knowledge of such pathways remains largely elusive in other nematodes. Thus, in this thesis I will analyse the expression patterns during early developmental stages among selected representatives of nematodes by employing RNAseq.

This approach allows analysis of expressed genes during early development in an unbiased fashion. Combining RNAseq with global genome and transcriptome clusterings shall give insights into the distribution of orthologues among the selected nematodes species and outgroups (Fig. 1.1). One of the obstacles of following such an approach is the small amount of total RNA that can be extracted from nematode embryos. To circumvent this problem, in the past sampled RNA from *C. elegans* was amplified by *in-vitro* transcription (IVT)-based methods prior to RNA sequencing (Hashimshony et al. 2012). Here, I will establish a similar method for other nematode embryos to facilitate RNA sequencing. Concerning the selected nematodes of interest, I intend to get first insights into early transcriptomes and underlying GRNs of Chromadorean and Enoplean nematodes, as well as nematomorphs (Fig. 1.1). With this approach I intend to address different questions such as:

(1) Which are the conserved pathways and genes found in nematode species of different clades in comparison to other animal phyla? How do they control early embryogenesis? Are they involved in polarity establishment?

(2) Are there so far undescribed genes which are expressed in all investigated nematodes, including *C. elegans*?

(3) In what respect is *C. elegans* different from other nematodes concerning gene expression during early development?

(4) Are there conserved GRNs among basal Enoplean nematodes and other model organisms?

(5) To what extent are there molecular differences among non-*Caenorhabditis* species with respect to polarity establishment and axis specification?

Chapter 2

Materials and Methods

2.1 Materials and media

Most chemicals and solutions used for the methods presented here, were obtained from the companies AppliChem GmbH (Darmstadt, Germany), Roche Diagnostics Inc. (Indianapolis, USA; Mannheim, Germany), Sigma Aldrich GmbH (Steinheim, Germany), Merck KGaA (Darmstadt, Germany), Roth GmbH (Karlsruhe, Germany). All chemicals and solutions which were not produced by these companies will be listed separately.

DEPC-H_2O:

- add 500 μl DEPC to 500 ml dH_2O
- shake at RT for several hours
- autoclave at 130°C for 20 min at a pressure of 2 bar

Low salt plates for nematode culturing (500 ml):

- 10 g agar
- 500 ml H_2O
- autoclave at 130°C for 20 min at a pressure of 2 bar
- 1 ml cholesterol (5 $^{mg}/_{ml}$) in EtOH

Nematode growth medium (NGM) (1000 ml):

- 2.5 g bacto-peptone
- 10 g NaCl
- 17 g agar
- 925 ml H_2O
- autoclave at 130°C for 20 min at a pressure of 2 bar
- 1 ml 1M $MgSO_4$
- 1 ml 1M $CaCl_2$
- 1 ml cholesterol (5 $^{mg}/_{ml}$) in EtOH
- 25 ml KPO_4 buffer (pH 6.0)
- 2 ml ampicillin (50 $^{mg}/_{ml}$)
- 1 μl 1 M IPTG

GU-mix (25 ml):

- 416.5 μl sarcrosyl (35%)
- 833.5 μl sodium citrate (0.75M)
- 11.8 g guadinium thiocyanate (118.16 $^{g}/_{mol}$)
- 3.6 μl β-mecaptoethanol
- ad 25 ml DEPC-H_2O

Tri-mix (1 ml):

- 500 μl GU-mix
- 500 μl acid phenol
- 50 μl 2M Na-acetate

- 7.2 μl β-mecaptoethanol

4% (v/v) Sodium hypochloride (NaOCl) solution (1.7 ml):

- 600 μl 12% (v/v) sodium hypochlorid
- 250 μl 5M KOH
- 850 μl dH_2O

20 x SSC (1000 ml):

- 175.3 g NaCL
- 82.2 g sodium citrate
- add 1000 ml dH_2O
- adjust pH to 7.0

Hybmix (10 ml):

- 5 ml 50% (v/v) formamide
- 2.5 ml 20 X SSC
- 10 mg DNA
- 10 μl heparine (100 $^{mg}/_{ml}$)
- 50 μl 20% (v/v) Tween20
- 2 ml 10% (v/w) destransulfate solution

2x SSC-formamide (75 ml):

- 7.5 ml 2x SSC
- 45 ml 60% (v/v) Formamid
- add H_2O to a total volume of 75 ml

2x TN (250 ml):

- 25 ml 100 mM Tris (pH 7.5)
- 7.5 ml 150 mM NaCl

Luria Bertani (LB) medium (1000 ml):

- 10 g Pepton
- 5 g yeast extract
- 10 g NaCl
- 17 g Agar
- add dH_2O to a total volume of 1000 ml
- autoclave at 130°C for 20 min at a pressure of 2 bar
- 2 ml ampicillin (50 $^{mg}/_{ml}$)
- 2.5 ml tetracycline (5 $^{mg}/_{ml}$)

2.2 Worm cultures and egg collection

I cultured the species *Caenorhabditis elegans*, *Pristionchus pacificus*, *Diploscapter coronatus* and *Panagrolaimus sp. PS1159* on low salt agar plates with *Escherichia coli* (strain OP50) as food source (Lahl et al. 2003).
Eggs of the nematodes *Pristionchus pacificus*, *Diploscapter coronatus* and *Panagrolaimus sp. PS1159* were directly collected from agar plates. The collected embryos were checked by using dissection microscopes. For each species several biological replicates were created by independent collection of eggs. Eggs were collected in Eppendorf tubes, filled with 25 μl of dH_2O. While collecting the eggs, Eppendorf tubes were kept on ice to slow down embryogenesis. After collecting 100 embryos in the range of 1- to 4-cell stage (in the case of *D. coronatus* 1- to 8-cell stage), collected embryos were shock-frozen by treatment with liquid nitrogen and then kept at −80°C.

The Enoplean nematode *Romanomermis culicivorax* was retrieved as live cultures from Prof. Dr. E. Platzer (Department of Nematology, University of Riverside, USA). A volume of approximately 1 ml of *R. culicivorax* L_4 larvae were put in 35 ml of spring water (in 9 cm gamma-irradiated plastic petri-dishes; Sarstedt, Germany). L_4 larvae were kept at constant 24°C until they reached adulthood (roughly for 4 weeks and with approximately 8 hours of natural day light per day). Adult *R. culivorax* virgins and males were collected separately prior to fertilisation and put into 25 μl of dH_2O. Samples of adult virgins and males were shock-frozen using liquid nitrogen and afterwards stored at −80°C. *R. culicivorax* eggs were collected directly from the liquid cultures after successful egg laying of fertilised females. The stage of the eggs was determined by using dissection microscope. After confirming that embryos were at stages of 1- to 4-cell stage, embryos were immediately moved into Eppendorf tubes filled with 25 μl of dH_2O. Eppendorf tubes were kept on ice to slow down embryogenesis of collected eggs. After collecting 100 eggs, eppendorf tubes were shock-frozen by using liquid nitrogen. Collected eggs were stored at −80°C.

The nematomoph *Gordius. sp.* was retrieved as live cultures from Prof. Dr. B. Hanelt (Department of Biology, University of New Mexico, Albuquerque, USA). *Gordius sp.* male and female adults were kept in water consisting of 75% demineralised and 25% spring water. Uteri were collected after successful mating of the females with males. Uteri were cut at both ends and few eggs from each end were immediately mounted onto slides to determine the stage and verify synchronous development of the embryos. Only such uteri which consisted of synchronised embryos at 1-cell stage were immediately put into 25 μl of dH_2O and afterwards shock-frozen by treatment with liquid nitrogen to preserve mRNA content. For each independent biological replicate of *Gordius sp.* 1-cell stages I retrieved approximately 10 mg of tissue, hence it was possible to extract total RNA by conventional RNA precipitation methods (see section 2.5). Collected uteri were stored at −80°C.

2.3 RNA extraction

In the cases of *R. culicivorax* adults and *Gordius sp.* uteri, I had enough tissue (more than 10 mg) to successfully precipitate and extact total RNA without using RNA carrier molecules (compare to section 2.4). Hence, I could use a modified version of the protocol established by Chomczynski and Sacchi (1987).

For this purpose I produced Tri-mix (see section 2.1). I quickly transferred tissue pellets of approximately 10 mg into 1 ml Tri-mix without letting the tissue thaw and then destroyed the tissue using a homogeniser (Ultra-Turrax, IKA Werke GmbH). I homogenised the sample for 10 min while cooling the tube with ice. Then, I added 200 μl chloroform and vortexed the sample. I incubated the sample at room temperature (RT) for 5 min and then centrifuged it for 10 min at 15,000 G. I transferred the transparent aquaneous phase (without the inter-phase) into a new Eppendorf tube and thus I separated the total RNA from protein and genomic DNA. I added 0.025 volumes of 1M acidic acid. I vortext the sample and added 0.5 volumes of pre-cooled 100% EtOH (-20°C). I vortexed again and precipitated the sample over night at -20°C. Next day, I centrifuged the sample at 15,000 G for 20 min. Afterwards, I removed the supernatant and dried the RNA pellet for 10 min I resuspended the RNA pellet in 125 μl of GU-mix (see section 2.1) and added 3.125 μl 1M acidic acid, vortexed the sample and added 70 μl 100% EtOH . I precipitated the RNA over night at -20°C. The following day, I centrifuged the sample at 15,000 G for 20 min and afterwards washed the sample by removing the supernantant and adding 500 μl EtOH (70%). I repeated the last step once more. I removed most of the supernatant and let the remaining EtOH evaporate, by incubating the eppendorf tube with an open lid at RT. After approximately 15 min I dissolved the RNA pellet in 20 μl DEPC-H_2O. In order to avoid formation of secondary structures of the RNA in the aquaneous medium, I incubated the sample in a pre-heated waterbath (65°C) for 5 min and vortexed the sample. I repeated the last step once more. Then, I immediately put the tube onto ice. The quality of the total RNA was checked by using degenerative agarose-gel electrophoresis and I quantified the amount of the total RNA using the nanodrop 1000 photometer (Agilent Inc.). All used total RNAs were not degenerated and had $^{230nm}/_{260nm}$- and $^{230nm}/_{260nm}$-absorbtion ratios of greater than 1.9. Extracted RNA was stored at -80°C.

2.4 RNA extraction from samples with low RNA content

In order to have enough mRNA for RNA sequencing I first had to extract total RNA without losing too much mRNA content, then I had to amplify the mRNA from the extracted total RNA.

For each sample I retrieved from collected nematode embryos, I needed to precipitate the RNA which was available from 100 embryos (which is roughly 1 to 3% of an estimated total RNA content of 10 ng; section 2.5). I needed a method to quickly transfer the samples into a medium which would inactivate RNAses while maintaining stable mRNA. Hence, I used a modified version of the protocol after Baugh et al. (2001) and Hashimshony et al. (2012) (personal communication Itai Yanai).

For the whole procedure I used DEPC-treated eppendorf tubes and RNAse-free pipette tips. As shown in section 2.2 nematode embryo samples were stored in 25 μl dH_2O at $-80°C$. I intended to precipitate the total RNA of the samples by destroying the tissue and then dissolving mRNA, DNA and protein in a mix consisting of guandinium thiocyanate, phenol and β-mecaptoethanol, the so-called Tri-mix (Chomczynski and Sacchi 1987). In order to have the necessary concentration of 4M guandinium thiocyanate, I had to increase the guandinium concentration in the Tri-Mix. For creating a proper guandinium thiocyanate solution (or 6M-GU-mix) I used the following protocol:

6M-GU-mix (2.5 ml):

- 53.57 μl sarcrosyl (35%)
- 125.025 μl sodium citrate (0.75M)
- 1.77 g guandinium thiocyanate (118.16 $^g/_mol$)
- ad 2.5 ml DEPC-H_2O

I prepared for each sample the Tri-Mix by the following protocol:

Tri-mix (200 μl):

- 75 μl 6M-GU-mix
- 100 μl acid phenol

- 10 μl 2M Na acetate
- 1.44 μl β-mecaptoethanol

To the Tri-mix I added 1 μl (5 $^{\mu g}/_{\mu l}$) LPA (linear polyacylamide; Life Technologies Inc.) and 1 μl (10 $^{\mu g}/_{\mu l}$) tRNA as carrier molecules for RNA precipitation. Now, I could start to extract total RNA from the samples. By shock-freezing the samples I retrieved tissue pellets which contained the frozen embryos. I transferred the pellets quickly into cryo-tubes and kept them in liquid nitrogen. I quickly transferred the cryo-tube (containg the tissue pellet) onto ice. I pipetted 175 μl of the prepared Tri-mix onto the tissue pellet and used a homogeniser (Xenox, Proxxon GmbH, Föhren, Germany) to destroy the tissue and dissolve RNA in the Tri-mix. I homogenised the disolved pellet for 10 min at 20,000 rpm (rounds per minutes) using the homogeniser (Xenox, Proxxon GmbH, Föhren, Germany) while cooling the cryo-tube simultaneously with ice. Afterwards, I transferred the Tri-mix with the homogenised tissue into a DEPC-treated eppendorf tube. I vortexed the sample for 10 sec at the highest setting and then I added 40 μl chloroform. I vortexed the sample again for 30 sec and then I centrifuged the sample at 18,000 rpm for 5 min. I transferred the transperant aquaenous phase into a new DEPC-treated eppendorf tube and added 120 μl isopropanol (100%). I mixed the sample by shaking and vortexing for 10 sec and then I incubated the sample for 2 min at RT. I centrifuged for 10 min at 18,000 rpm with the eppendorf tube's hinge aligned with the rotor. I carefully removed the complete supernatant. In most cases I could spot a transperant pellet at the bottom of the tube. I carefully added 500 μl of 70% EtOH, vortexed for 10 sec and inverted the Eppendorf tube several times. I centrifuged the sample again for 2 min at 18,000 rpm, but this time with the hinge alinged 180° relative to the previous centrifugation step. I carefully removed the supernatant and centrifuged again at 18,000 rpm for 10 sec. I removed the remaining supernantant and dried the pellet by opening the lid of the tube and waiting for 2 min. I dissolved the RNA pellet in 2 μl DEPC-H_2O and immediately transferred the eppendorf tube into liquid nitrogen and stored the sample at -80°C to avoid degradation of the extracted RNA.

2.5 Amplification of low amounts of total RNA by *in-vitro* transcription

The total RNA, and messenger RNA (mRNA) content of collected embryo samples for *P. pacificus*, *D. coronatus*, *Panagrolaimus sp. PS1159* and *R. culicivorax* could be roughly estimated by comparing with *C. elegans* embryo samples. The total RNA content of *C. elegans* embryo is approximately 200 ng per embryo (personal communication Itai Yanai). Hence, an estimation of a total RNA content for embryonic samples for the 100 embryos of other species was approximately 20 ng and the mRNA content (which correlates to 1 to 3% of the total RNA content) was estimated to be approximately 0.2 ng. This is by far too low for RNA sequencing by Illumina platforms (see section 2.6). Hence, amplification of the collected embryo samples was necessary for selected nematode species (see above). Therefore, I established a technique in the Schierenberg laboratory to amplify the low amount of mRNA so that RNA sequencing by Illumina platforms was feasible. For this purpose I used the "Ambion Message AMP II aRNA ^{TM}" amplification kit (AM1751; Life Technologies Inc.). This kit was previously used to amplify the RNA content of single blastomeres of *C. elegans* (Hashimshony et al. 2012). The RNA amplification of this kit is based on 4 successive steps (Life-Technologies 2011): (1) Reverse transcription of mRNA into cDNA (complementary DNA) by usage of the "ArrayScript ^{TM}" reverse transcriptase (Life Technologies Inc.) directed by primers incorporating T7 promotor sequences and targeting the 3'-poly-A tail of mRNAs. (2) Generation of double-stranded cDNA. (3) *In-vitro* transcription (IVT) by usage of the incorporated T7 promotor of the double-stranded cDNA. (4) Purification of the synthesised mRNA. The amplified and purified mRNA could then be used to create sequencing libraries for the Illumina HiSeq and MiSeq platforms.

For the preparation of each sample, I used extracted total RNA of approximately 100 embryos, which was dissolved in 2 μl DEPC-H_2O (see section 2.4). For the amplification by the "Message AMP II aRNA ^{TM}" kit, I modified a few steps: I performed one round of amplification by IVT (the kit allows up to two rounds of IVT) and for all pipetting schemes to prepare solutions for each of the aforementioned steps, I only used $^1/_5$ of the recommended volumes (similar to single blastomere RNA sequencing by Hashimshony et al. 2012). For most purification steps I used also $^1/_5$ of the recommended volumes.

This includes usage of the "cDNA binding buffer" and the "aRNA binding buffer" (Life-Technologies 2011). For the "wash buffer" I used the recommended volumes. Instead of using the "aRNA filter cartridges" for purification of amplified mRNA, I used "cDNA filter cartridges" (AM10066G; Life Technologies Inc.). Before starting with the whole procedure recommended by the handbook of the "Message AMP II aRNA ^{TM}" amplification kit, I mixed 2 μl of my previously extracted total RNA (see section 2.4) with 2 μl of the "reverse transcription master mix" (Life-Technologies 2011) and incubated the reaction for 10 min at 70°C to denaturate the total RNA in the samples and avoid formation of secondary RNA structures.

Besides from the mentioned alterations to the protocol of the handbook of the "Message AMP II aRNA ^{TM}" amplification kit, I followed all recommended steps as described in the handbook (Life-Technologies 2011). All incubation steps were performed on a "PeqSTAR 2X Universal Gradient ^{TM}" thermocycler (PEQLAB Biotechnolgie GmbH, Erlangen, Germany). Successfully amplified mRNA was dissolved in 15μl DEPC-H_2O and stored at -80°C before RNA sequencing.

For each amplified nematode total RNA sample I added two additional samples of purified total RNA of high quality as positive controls for successfull amplification of mRNA (Tab. 2.1; Fig. 2.1) to test whether all reagents of the kit were functional and amplification of low amounts of total RNA was possible. I used 20 ng total RNA of the nematode *Acrobeloides sp. PS1146* and 2,000 ng total RNA from human HeLa cells (which was provided by the "Message AMP II aRNA ^{TM}" amplification kit). The amount of amplified mRNA was quantified using the "Qubit 2.0 ^{TM}" fluorometer (Life Technologies Inc.) with the "Qubit RNA HS Assay Kit" (Q32852; Life Technologies Inc.). By usage of these two postive controls I could quantify the level of amplification for both positive controls to assure that the amplification by IVT in general was successful (Tab. 2.1). I did not quantify the amount of amplified nematode samples to avoid loss of precious amplified mRNA prior to RNA sequencing. In average I observed a 461-fold amplification of the inital mRNA content (Tab. 2.1). This should correlate with approximately 20 ng of pure mRNA, enough material for construction of sequencing libraries and RNA sequencing.

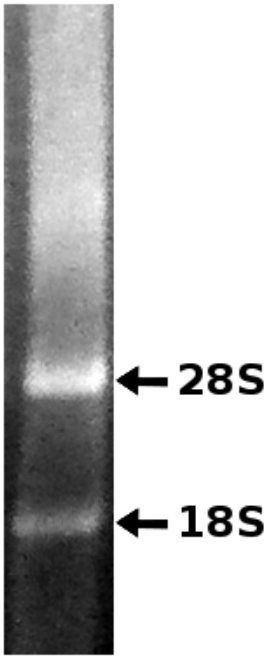

Figure 2.1: Gel-photo visualising the quality of *Acrobeloides sp. PS1146* total RNA used as positive control (see text) after denaturating agarose-gel electrophoresis. The band of the 28S rRNA is much brighter than the band of the 18S rRNA, indicating high RNA quality.

Table 2.1: Amplification efficiency of the Ambion Message AMP II kit for 20 ng and 2,000 ng of total RNA. Abbreviations: *A. s.* - *Acrobeloides sp. PS1146*; HeLa - human HeLa cells.

sample	**total RNA amount before amplification [ng]**	**mRNA amount after amplification [ng]**	**amplification factor**
A. s. total RNA	20	92.2 (±27.3)	461-fold (±136)
HeLa cell total RNA	2,000	33,300 (±6,011)	1,664-fold (±300)

2.6 RNA sequencing

RNA sequencing was done on Illumina HiSeq and MiSeq platforms (Illumina Inc.; Tab. 2.3) by the Cologne Center for Genomics (CCG, Cologne, Germany). For the preparation of the sequencing libraries the "TrueSeq RNA Sample Prep kit" version 2 (Illumina Inc.) was used. For library construction of total RNA extracted from *R. culicivorax* adult virgins and males and for *Gordius sp.* embryos the Illumina "TrueSeq" kit was used without modifications after standard procedure (see section 2.4).

Table 2.2: Quantity and average length of sequences of created TrueSeq libraries. (*) these libraries were sequenced twice and technical replicates were generated. Abbreviations: *D. c.* - *Diploscapter coronatus*; *G. sp.* - *Gordius sp.*; *P. p.* - *Pristionchus pacificus*; *P. sp.* - *Panagrolaimus sp. PS1159*; *R. c.* - *Romanomermis culicivorax*.

species	**biologcial sample**	**library**	**concentration** [$^{ng}/_{\mu l}$]	**average sequence length [bp]**
P. p.	1- to 4-cell	Pp	20.7	275
D. c.	1- to 8-cell	Dc#1	0.415	285
D. c.	1- to 8-cell	Dc#2	1.91	281
D. c.	1- to 8-cell	Dc#4*	25.9	290
D. c.	1- to 8-cell	Dc#5	23.2	267
P. s.	1- to 4-cell	Psp	1.37	264
R. c.	1- to 4-cell	Rc#1*	53.3	287
R. c.	1- to 4-cell	Rc#2	61.7	276
R. c.	1- to 4-cell	Rc#3	42.5	286
R. c.	virgins	RcV	19.7	264
R. c.	males	RcM	26.5	269
G. sp.	1-cell	Gsp#1	24.5	277
G. sp.	1-cell	Gsp#2	26.4	280

For the amplified mRNA samples of *P. pacificus*, *D. coronatus*, *Panagrolaimus sp. PS1159* and *R. culicivorax* RNA libraries were created by using "TrueSeq RNA Sample Prep kit" without the pre-liminary mRNA selection step. Instead the amplified mRNA was directly used for the library preparation.

In total 13 libraries were created. The quantity and the quality of the sequencing results

was either tested by using the Bioanalyzer 2100 (in the case of *R. culicivorax* adults; Agilent Technologies Inc.) or by using the Tapestation 2200 (Agilent Technologies Inc.). Most of the created libraries had a concentration of greater than 19 $^{ng}/_{\mu l}$ and an average DNA length of 278 bp (Tab. 2.2). This allowed sequencing by the Illumina platform, as it was possible to retrieve an abundant number of reads from sequencing all created libraries (Tab. 2.3).

Table 2.3: Results of sequencing by Illumina platforms. (*) these libraries were sequenced twice and technical replicates were generated. For abbreviations refer to Tab 2.2

species	**biological sample**	**library library**	**platform platform**	**read length [bp]**	**number of sequenced reads**
P. p.	1- to 4-cell	Pp	MiSeq	75 bp	21,983,252
D. c.	1- to 8-cell	Dc#1	MiSeq	75 bp	6,990,001
D. c.	1- to 8-cell	Dc#2	MiSeq	75 bp	8,680,901
D. c.	1- to 8-cell	Dc#3	MiSeq	75 bp	8,898,677
D. c.	1- to 8-cell	Dc#4 *	HiSeq 3000	100 bp	31,199,716
D. c.	1- to 8-cell	Dc#5	HiSeq 3000	100 bp	28,416,086
P. s.	1- to 4-cell	Psp	HiSeq 3000	100 bp	26,004,158
R. c.	1- to 4-cell	Rc#1 *	HiSeq 3000	100 bp	44,716,067
R. c.	1- to 4-cell	Rc#2	HiSeq 3000	100 bp	27,224,273
R. c.	1- to 4-cell	Rc#3	HiSeq 3000	100 bp	24,005,760
R. c.	1- to 4-cell	Rc#4	HiSeq 3000	100 bp	56,550,222
R. c.	virgins	RcV	HiSeq 3000	100 bp	25,882,167
R. c.	males	RcM	HiSeq 3000	100 bp	21,454,265
G. sp.	1-cell	Gsp#1	HiSeq 3000	100 bp	31,629,341
G. sp.	1-cell	Gsp#2	HiSeq 3000	100 bp	37,168,752

2.7 Preparation of RNA interference experiments

RNA interference (RNAi) experiments were performed on 12-well NGM-plates. NGM medium was prepared after the described protocol (section 2.1). *Escherichia coli* (strain HT115) feeding bacteria (Timmons and Fire 1998) were prepared by inoculation of an over night culture of 5 ml in Luria Bertani (LB) medium containing 100 $^{\mu g}/_{ml}$ ampicillin and kept at 37°C (shaken at 200 rpm). The next day, the over-night cultures were diluted (1:30) in new LB medium (containing 100 $^{\mu g}/_{ml}$ ampicillin) and shaken at 200 rpm at 37°C until the *E. coli* HT115 feeding bacteria reached an optical density of OD_{600} = 0.4 (this took approximately 90 min). Now I added 1 M IPTG, which equals an concentration of 1 mM in solution. I incubated the *E. coli* HT115 feeding bacteria for 4 hours at 37°C (shaken at 200 rpm). dsRNA expressing bacteria were put on NGM-plates. The prepared NGM-plates were stored at 4°.

2.8 RNAi experiments of identified candidate genes involved in oocyte maturation in *C. elegans*

RNA interference (RNAi) experiments for *C. elegans* adults were performed on NGM (nematode growth medium) plates (see section 2.1) which were coated with the *Escherichia coli* strain HT115 for RNAi by feeding (Timmons and Fire 1998). The used constructs for the genes *inx-19*, *unc-7*, *plx-2*, *str-102*, *cng-1*, *cnd-1*, *aptf-1*, *aptf-4*, *nhr-81* and *ztf-11* were retrieved from a RNAi feeding construct library (Kamath et al. 2003). We used squencing PCR (section 2.9) and confirmed by aligments the identity for these constructs.

I repeated each RNAi experiment in 6 replicates on 12-well plates. One L_4 larva was placed onto the surface of a single well and raised to adulthood. *C. elegans* adults were removed after the adult hermaphrodites had laid 5 eggs of the next generation. The F_1 generation wwas raised to adulthood under the condition of dsRNA expressing bacteria. The time point of first laid eggs was used as the starting point to score for phenotypes. Adults of the F_1 generation were transferred to the next well after 24 hours of egg laying. After another 24 hours the eggs were checked for hatched larvae. Hatched larvae of the F_2 generation were raised to adulthood and afterwards analysed for RNAi phenotypes

by usage of differential interference contrast (DIC) microscopy. It was tested if these RNAi-treated adult hermaphrodites showed prominent phenotypes which were associated with oocyte maturation. For this purpose, I investigated the morphology of the gonad, the usterus and the vulva (Fig. 3.7).

2.9 Sanger sequencing

RNAi constructs retrieved from the RNAi feeding library (Kamath et al. 2003) were sequenced to confirm the identiy of the targeted *C. elegans* genes. For this purpose I used the Big Dye Kit version 3.1 (Life Technologies Inc.). As all constructs were cloned into the L4440 plasmid (Kamath et al. 2003) which incorporates T7 and T3 promotor sequences, I used a primer against the T7 promotor sequence (Tab. 2.4) to amplify the incorporated insert in the L4440 plasmid by PCR (Tab. 2.7). The PCR product was sequenced by a ABI3730 (Life Technologies Inc.) capillary sequencer in the Cologne Center for Genomics.

Table 2.4: Sanger sequencing primers used in this thesis.

primer	**target sequence**	**sequence**
dT7-seq_sense	T3 promtor site of the L4440 plasmid	5'-AAC-CTG-GCT-TAT-CGA-AAT-3'
dT7-seq_antisense	T7 promtor site of the L4440 plasmid	5'-GTA-AAA-CGA-CGG-CCA-GTG-AG-3'

Table 2.5: Programme of the sequencing PCR.

	step	**temperature**	**duration**
	1. inital denaturation	95°C	1 min 20 sec
	2. inital annealing	51°C	20 sec
	3. inital elongation	60°C	4 min
	4. denaturation	95°C	20 sec
28x	5. annealing	51°C	20 sec
	6. elongation	60°C	4 min
	7. end elongation	60°C	10 min
	8. storrage	8°C	∞

2.10 Microarray analysis

RNAi experiments to knock-down the genes *par-5*, *itr-1* and *goa-1* were done previously in the Schierenberg laboratory (personal communication Peter Heger). I did RNAi experiments for *mpk-1*, *mek-2*, *lip-1* and *mbk-2* in a previous study in the Schierenberg laboratory (Kraus 2009). Each RNAi experiment was done twice to generate independent biologcial replicates.

Extracted total RNA of RNAi-treated *C. elegans* adults for knock-down of the genes *par-5*, *itr-1*, *goa-1*, *mpk-1*, *mek-2*, *lip-1* and *mbk-2* were shipped to the company Atlas Biolabs GmbH (Berlin, Germany). This company performed library preparations for hybridisation of cDNA on *C. elegans* specific Affymetrix microarrys ("*C. elegans* Genome Array", catalogoue No. 900383, Affymetrix Inc.). This was done in two independent rounds of experiments: In the first experiment cDNA libraries for *par-5(RNAi)*, *itr-1(RNAi)*, *goa-1(RNAi)* and wild-type were hybridised to the microarrys. In the second round cDNA libraries of *mpk-1(RNAi)*, *mek-2(RNAi)*, *lip-1(RNAi)*, *mbk-2(RNAi)* and wild-type were hybridised to the microarrays.

Raw data retrieved from both experiments was provided by Atlas Biolabs GmbH (Berlin, Germany). I used this raw data from both rounds of experiments and prepared it for analysis using the programming languages "R" and "Bioconductor" (bioconductor.org). For this purpose I used the following packages of Bioconductor: "affy" (version 1.32.0; Gautier et al. 2004), "simpleaffy" (version 2.30.0; Wilson and Miller 2005) and the *C. elegans* specific affymetrix microarray annotation database "c.elegans.db" (version 2.8.0; Carlson 2012). By this approach I could read the raw data and normalise it using the algorithms MAS5 (Hubbell et al. 2002). I did pairwise comparisons between expression data of the RNAi experiments and the wild-type control. Such genes (the more precise term is "features" for the hybridising oligomeres on the surface of the microarray) which showed a significant foldchange (fc) of at least $\pm$ 1 (two-tail t-test of $<10^{-3}$; a fc of $\pm$ 1 equals a 2-fold over- or underexpression due to binary logarithmic nature of the unit) were extracted (section 3.1).

For annotation of significantly over- and underexpressed genes the aforementioned *C. elegans* database package was used (Carlson 2012). Quality assessment of the microarray data was done by the quality control (QC) plots of Wilson and Miller (2005) and MA-plots (Heber and Sick 2006).

The QC plot by Wilson and Miller (2005) uses 4 criteria to determine the quality of

raw data provided by Affymetrix microarrys (see section 3.1; Fig. 3.1 and 3.2). A detailed description of each of these quality criteria can be found in the manufacturer's handbook (Wilson and Miller 2005, Affymetrix 2004). Reference values for all 4 quality criteria are also defined by the manufacturer (Affymetrix 2004) and if the margins of reference values were met, the QC plots marked the quality criteria in blue. If, however, any reference value for the quality criteria was not within the margins of the reference values, the specific quality criterion is marked red. In the case of our RNAi experiments all quality criteria had values which were in the margins of the reference values set up by Affymetrix Inc. (Fig. 3.1, 3.2).

After confirming the quality of the microarray raw data, I needed to verify that normalisation was successful. For this purpose I used so-called "MA-plots". The MA-plot is a method to identify a systematic bias in expression data (Heber and Sick 2006) between two experimental setups (here a comparison between RNAi experiments and wild-type negative control). The $\log_2$ expression foldchange ($M = log_2 \frac{RNAi}{wt}$) is plotted against the mean $\log_2$ expression ($A = \frac{1}{2} \cdot (log_2(RNAi) + log_2(wt))$) of both experiments (Fig. 3.3). If the expression data were normalised there should be (1) no correlation between M and A, showing that expression values were not biased by the experimental setup between RNAi experiment and wild-type control. (2) All foldchange values should tend to "cluster" around a minimal value of zero (Fig. 3.3), indicating that expression foldchange is indepdent of average expression between RNAi experiment and wild-type control. Here, I could show by linear regression that the foldchange between RNAi experiments and wild-type control was indeed independent of the average expression, showing that all expression data were properly normalised (Tab. 2.6; Fig. 3.3). For testing for significant independence of the expression values of RNAi experiments and the wild-type control I used "Pearson product-moment correlation coefficient" (or short Pearson's r; Pearson 1895; Tab 2.6).

After quality assessment I tested whether expression values between RNAi experiments and wilde-type control were correlated, following the assumption that most expression values should be similar between RNAi experiment and wild-type control. Only few expression differences should exist and can be understood as a consequence of the experimental setup where RNAi treated adults are compared with wild type. I plotted $\log_2$ expression values of RNAi experiments against wild-type control and used linear

Table 2.6: Testing of correlation between $\log_2$ expression foldchange (M) and mean of $\log_2$ expression (A). Abbreviations: M = $\log_2$ expression foldchange between RNAi approach and wild-type; $M = log_2 \frac{RNAi}{wt}$; A = average $\log_2$ foldchange between RNAi approach and wild-type; $A = \frac{1}{2} \cdot (log_2(RNAi) + log_2(wt))$.

RNAi knock-down	**Pearson's r**	**p-value**
par-5(RNAi)	-0.081	$< 2.2 \cdot 10^{-16}$
itr-1(RNAi)	-0.060	$< 2.2 \cdot 10^{-16}$
goa-1(RNAi)	-0.116	$< 2.2 \cdot 10^{-16}$
lip-1(RNAi)	-0.096	$< 2.2 \cdot 10^{-16}$
mbk-1(RNAi)	-0.087	$< 2.2 \cdot 10^{-16}$
mek-2(RNAi)	-0.027	$< 6.1 \cdot 10^{-5}$
mpk-1(RNAi)	0.207	< 0.01

regression analysis to test for possible correlation. Pearson's r values were positive and always greater than 0.97 (the maximum value for positive correlation is 1.00), indicating a strong positive correlation between RNAi experiments and wild-type control (Tab. 3.1). Hence, neither the RNAi nor the wild-type $\log_2$ expression data showed any indication of a systematic bias so that both would not (Pearson's r = 0) or even negatively (Pearson's r <0) be correlated (expression in RNAi would have an negative effect on expression of wild-type or vice versa).

Thus, Affymetrix microarray-based expression data did not only meet all recommended quality criteria set by the manufacturer, I also successfully normalised the data and could show that neither the RNAi experiments, nor the wild-type controls showed biased data. Therefore, our microarray expression data is reliable and can be used for identification of significantly over- and/or underexpressed genes (see Results; section 3.1).

In order to identify expression of genes which do not follow the correlation between RNAi experiments and wild-type and may be a consequence of RNAi effects, I identified candidate genes by significant foldchange of $\pm$ 1 (which correlates with twice as high expression under the condition of RNAi knock-down). I tested significance by two-tailed t-tests (Results; Fig. 3.4 B).

2.11 *In-situ* hybridisations

R. culicivorax orthologues of *hunchback* and *chordin* were identified by orthology clustering (section 2.13). Identified sequences were confirmed to be orthologues of the aforementioned genes by phylogenetic analysis (section 2.14)

Prior to *in-situ* hybridisation (ISH), it was necessary to clone the target sequence of mRNAs which will be analysed and generate anti-sense RNA probes. For this purpose *R. culicivorax* sequences orthologous to human *Chordin* and *Drosophila huchback* were amplified from cDNA via polymerase chain reaction (PCR; Tab. 2.7) using mRNA specific primers (Tab. 2.8). Due to usage of a DNA polymerase which adds adenine residues at the ends of amplicons (TAQ from *Thermus aquaticus*) it was possible to clone the amplified DNA fragements via T/A cloning (Zhou et al. 1995, Hadjeb and Berkowitz 1996) into the expression plasmid pBluescript KS.

Anti-sense RNA probes for *R. culicivorax* orthologous of *chordin* and *hunchback* were synthesised from linearised plasmid DNA by *in-vitro* transcription using the "Dig-RNA labeling kit" from Roche (Mannheim, Germany). The synthesised probes were purified by successive usage of precipitation using 4M LiCl and 12M LiCl. Successfully purified probes were put into hybmix (see section 2.1) and stored at −20°C.

Table 2.7: PCR programme for amplification of *R. culicivorax chordin* and *hunchback*.

	step	**temperature**	**duration**
	1. inital denaturation	95°C	4 min
40x	4. denaturation	95°C	30 sec
	5. annealing	55°C	30 sec
	6. elongation	60°C	4 min
	7. end elongation	68°C	10 min
	8. addition of adenine residues	72°C	10 min
	9. storage	4°C	∞

In total 3 probes were created: 2 anti-sense probes against the mRNAs of the orthologous sequences of *chrodin* and *hunchback* and a single sense mRNA for the cloned sequence of *chrodin*. The latter sense mRNA was used as a negative control to distinguish stainings created by binding of the anti-sense probes against target mRNAs from tissue intrinsic background stainings.

The procedure of hybridising the anti-sense probes against the target mRNAs in *R. culicivorax* embryos is described in detail below:

Prior to the *in-situ* hybridisation it was necessary to sample enough *R. culicivorax* embryos (section 2.2). Collected embryos were pooled by centrifuging at 500 g for 2 min and then treated with hyprochloride solution for 2 min (NaOCl; section 2.1). The reaction was stopped by adding 1x PBS. Embryos were centrifuged at 500 g and again

washed with 1x PBS. The last two steps were repeated 3 times. Afterwards embryos were washed with dH_2O 4 times to remove residual PBS. Prepared *R. culcivorax* eggs were put onto slides coated with poly-L-lysine. After putting coverslips onto the slides, the slides were shock-frozen in liquid nitrogen. The coverslips were quickly removed using razor blades to crack the egg-shells of the embryo. The prepared slides were immediately transferred into ice-cold MeOH ($-20°C$). These slides could be stored for serveral months at $-20°C$.

MeOH was removed from prepared slides by a decreasing MeOH dilution series with decreasing concentration: Slides were washed for 10 min in 90% (v/v) MeOH, then in 70% (v/v) MeOH and finally in 50% (v/v) MeOH. 50 μl of pre-heated (37°C) fixating solution (Streck Tissue Fixative; Streck Laboratories Ltd., Philadelphia, USA) was put onto the slides. Slides were transferred into a prepared wet-chamber and incubated for 60 min at 37°C to fix the collected embryos. Slides were washed 10 min in dH_2O and then 10 min in 2x SSC buffer (pH 7.0) and then 50 μL of pre-heated Hyb Mix (42°C) was added. The slides were incubated at 42°C for 90 min in the wet chamber.

In order to prepare the anti-sense- and sense probes for ISH, it was necessary to degenerate the probes (to circumvent formation of secondary structures which would impair hybridisation to target mRNA). 0.3 to 1 $^{ng}/_{\mu l}$ of each probe was incubated for 10 min at 65°C. After removal of the hybmix from slides, the RNA probes (cooled to 42°C) were put onto the slides and incubated at 42°C overnight. During this step the anti-sense probes hybridised to the target mRNAs.

The next day, the RNA probes were removed and the slides were washed in the following solutions: (1) In 2x SSC at 42°C for 20 min. (2) In SSC-formamide mix at 42°C for 30 min. (3) Twice in 2x SSC for 10 min at RT. (4) In TN for 10 min at RT. The samples were blocked at least for 30 min using 50 μl blocksolution (in wet chamber at RT) to enhance specificity of the anti-digoxigenine (DIG) antibody.

Block solution:

- 5% (v/w) milk powder in dH_2O
- TN buffer

After blocking, slides were washed in TN buffer for 10 min at RT. Now, 50 μl of anti DIG antibody was added (1:2000 diluted in block solution) to the sample and incubated at 37°C for 1 hour in the wet-chamber. Afterwards the slides were washed twice in TN buffer for 10 min and then twice in TMN buffer (10 min each time). Now, 50 μl developing mix (consisting of NBT/BCIP and TMN buffer and alkaline phosphatase; "anti-digoxgenin AP NBT/BCIP kit", Roche Diagnostics Inc.) were added. In order to reduce the background staining by endogenous alcalic phosphatase, 500 mM levamisole was added to the developing mix. The slides were incubated in a dark wet chamber during the staining. The staining was constantly monitored. After successful staining, the reaction was stopped by neutralising the pH-value (50 mM EDTA in 1x PBS were added). Slides were finalised by removing residual stop mix and adding 50 μl mounting buffer (stop mix + 50% (v/v) glycerol) and putting a cover-slip onto the slides.

Table 2.8: Primers used for amplification of the orthologues for *R. culicivorax chrodin* and *hunchback*.

gene	primer	sequence
chrodin	RCsca10161for1 chor	5'-GGC-GTT-TGG-CAA-AAT-ATT-CC-3'
chrodin	RCsca10161rev1 chor	5'-TAT-TCA-CGT-GGT-GTT-TCG-ACC-3'
hunchback	RCTm.4618 for1	5'-GAA-ATT-CGC-CCA-ATT-CGG-'3
hunchback	RCTm.4618 rev1	5'-AAC-TTC-CAA-TTC-CAC-CGA-GG-'3

2.12 Reference Genomes and transcriptomes

Genomes and transcriptomes of nematode and outgroup species were retrieved from different sources. Tables 2.9, 2.10 and 2.11 explain the origin of each omics data set.
For the reference transcriptome of *C. elegans* I used published data from Hashimshony et al. (2012). In order to have a complete set of early expressed transcripts, I cumulatively combined the expressed transcripts of each blastomere of the *C. elegans* 1- to 4-cell stage (Hashimshony et al. 2012). All transcripts which had an average TPM-value (transcripts per million reads) value of >5 were considered as expressed (this is opposed

to the original paper where a TPM-value of 100 was regarded as clearly expressed transcript; Hashimshony et al. 2012). By this step I ensured that the 1- to 4-cell stage should include all expressed transcripts.

In order to have a similar reference transcriptome for early embryogenesis (1- to 4-cell stage) I retrieved the data from the published transcriptome set(s) of *A. suum* (Wang et al. 2014). Here I cumulatively pooled the expressed transcripts of the 1- to 4-cell stages and subtracted expressed transcripts of the 10- to 26-cell stages to ensure that expression during 1- to 4-cell stages would be part of the transcriptomic analysis for early embryogenesis in the case of *A. suum*.

Table 2.9: Genomes of nematode species used in this thesis. (*) The genome of *D. coronatus* was sequenced in the Kohara laboratory and could be accessed and used for further analysis due to cooperation with the Kohara laboratory (Yuji Kohara, Mishima, Japan). (**) *Panagrolaimus sp. PS1159* was sequenced by Philipp H. Schiffer (personal communication). (***) Genomes sequenced in a colaboration between Blaxter (Mark Blaxter, Edinburgh, UK) and Schierenberg laboratory.

species	**source**
Ascaris suum	Wang et al. (2012)
Brugia malayi	Scott and Ghedin (2009)
Bursaphelenchus xylophilus	Kikuchi et al. (2011)
Caenorhabditis angaria	ftp://ftp.wormbase.org/pub/wormbase/species/c_angaria/
Caenorhabditis briggsae	Gupta and Sternberg (2003), Stein et al. (2003)
Caenorhabditis elegans	*Caenorhabditis elegans* Sequencing Consortium (1998)
Caenorhabditis remanei	Gupta and Sternberg (2003)
Diploscapter coronatus	Kohara laboratory*
Dirofilaria immitis	Godel et al. (2012)
Enoplus brevis	Blaxter, Schierenberg laboratory (***)
Meloidogyne hapla	Opperman et al. (2008)
Pristionchus pacificus	Dieterich et al. (2008)
Panagrolaimus sp. PS1159	Schierenberg Laboratory (**)
Romanomermis culicivorax	Schiffer et al. (2013)
Trichinella spiralis	Mitreva et al. (2011)

Table 2.10: Genomes of outgroup species used in this thesis. (*) The transcriptome of the nematomorph *Gordius sp.* was created from male, female and embryos in a collaborative effort between the Blaxter (Mark Blaxter, Edinburgh, UK) and the Schierenberg laboratories.

species	**source**
Gordius sp.	Blaxter, Schierenberg laboratory*
Homo sapiens	uniprot.org
Drosophila melanogaster	flybase.org

Table 2.11: Reference transcriptomes used in this thesis to confirm that identified transcripts were indeed from transcriptomic data and not due to genomic or bacterial contaminations (see section 3.6). (*) The transcriptome of pooled mixed stage embryos, larvae and adults *D. coronatus* was sequenced in the Kohara laboratory and could be accessed and used for further analysis due to cooperation with the Kohara laboratory (Yuji Kohara, Mishima, Japan). (**) Transcriptomes sequenced in a colaboration between Blaxter (Mark Blaxter, Edinburgh, UK) and Schierenberg laboratory. (***) The *Panagrolaimus sp. PS1159* transcriptome was also sequenced by Philipp H. Schiffer (personal comunication).

species	**source**
Ascaris suum	Wang et al. (2014)
Caenorhabditis elegans	Hashimshony et al. (2012)
Diploscapter coronatus	Kohara laboratory*
Gordius sp.	Blaxter, Schierenberg laboratory **
Panagrolaimus sp. PS1159	Schierenberg laboratory ***
Pristionchus pacificus	ftp.wormbase.org/pub/ wormbase/species/ p_pacificus/sequence/transcripts (version WS247) Sommer laboratory
Romanomermis culicivorax	Schierenberg laboratory

2.13 OrthoMCL of nematode and outgroup species

For identifying orthologous proteins and protein families shared among 16 nematode-species and 12 outgroup species (see section 2.12) the OrthoMCL algorithm (version 2.0.7; Li et al. 2003) was used and established as pipeline for orthologue detection in the Schierenberg laboratory (this was done by Philipp H. Schiffer and Georgios Koutsovoulos; both granted me permission to use data from the retrieved OrthoMCL clustering to analyse my own data sets; personal communication Philipp H. Schiffer). Prior to clustering protein families and orthologues, retrieved whole proteome sets for all species involved (section 2.12), were filtered by the programme "Cd-hit" (Li and Godzik 2006) for redundancy at a threshold of 99% identity. The procedure of OrthoMCL was done after Fischer et al. (2011) with standard parameters. For the clustering step a inflation-parameter of 1.5 was set (Li et al. 2003).

2.14 Programmes and tools for post-sequencing analysis

For post-sequencing analysis (see Results; section 3.6) I used several methods (Fig. 3.11): (1) I used Blast2GO (Conesa et al. 2005) to predict functions of early expressed transcripts of seven species (Fig. 1.1). (2) I used Fisher's Exact Test, which is integrated in Blast2GO to test whether significantly over- and/or underrepresented gene-ontology (GO) terms exist in transcriptomes of early embryogenesis in comparison to reference transcriptomes (section 2.12; Blüthgen et al. 2005). (3) I accessed the provided orthologous clustering for whole genomes of 27 species (section 2.12 and 2.13) by custom programmes which I wrote in the programming language Perl (see Appendix B). (4) I used the GeneMANIA data base (Mostafavi et al. 2008, Warde-Farley et al. 2010) to extend my search for proteins which physically, or genetically interact with identified *Caenorhabditis*, Chromadorea, or Enoplea specific genes/proteins. (5) In special cases I used phylogenetic analysis to confirm the prediction by the different OrthoMCL clusterings that proteins are indeed orthologues.

Interproscan pipeline: For prediction of protein domains, the programme "Interproscan" (version 5 release RC6; Quevillon et al. 2005, Hunter et al. 2012) was used. This programme was established as a pipeline in the Schierenberg laboratory by Philipp H. Schiffer and Georgios Koutsovoulos (personal communication Philipp H. Schiffer and Georgios Koutovoulos) to predict protein domains across all proteomes of 27 species (section 2.12).

Blast2GO: Gene functions of sequenced transcripts were predicted by usage of the programme Blast2GO (version 2.8; Conesa et al. 2005). For this purpose whole reference transcriptomes (section 2.12) were compared with Genbank (Benson et al. 2015) via the gateway of NCBI (National Center for Biotechnology Information). Transcripts were translated by Transdecoder (Haas et al. 2013) and then compared with Genbank using the programme Blastp at a threshold for the expection value of E $<10^{-3}$ (Altschul et al. 1990). Afterwards, the integrated mapping and annotation procedures of Blast2GO were used to infer predicted gene ontology (GO) terms for each transcript of whole reference transcriptomes. I used the integrated procedure of Fisher's Exact Test to test whether specific GO-terms were over- and/or underrepresented in transcriptomes retrieved from

early embryogeneis of seven species (section 2.6 and 2.12) in comparison to the reference transcriptomes (Blüthgen et al. 2005).

Analysis of OrthoMCL clustering via custom Perl scripts: I used several custom Perl programmes (more precisely scripts) which I wrote. I did the main part of the analysis with three custom Perl scripts (Appendix B): These programmes allowed analysis of the OrthoMCL clustering for 27 species and the orthologous clustering of the transcriptomes retrived from early embryos of seven species.
By establishing these programmes I analysed the broad data sets for different aspects: (1) I could access the orthologous clustering for 27 species by using a single gene (or more preciesely gene identifier retrieved from public databases such as uniprot.org flybase.org or wormbase.org) and looked in which species orthologues to this specific gene were present or absent. (2) I could specifically identify transcripts of the transcriptomes of early embryos which have orthologues during early embryogenesis in other species. (3) I could extract whole protein sets in a batch approach for orthologous clusters expressed during early embryogenesis of the selected species.

Search for physically or genetically interacting proteins: I used the data base GeneMANIA (Mostafavi et al. 2008, Warde-Farley et al. 2010), to search for proteins interacting with orthologues, which I previously identified by quering the orthology clustering for 27 species (see above). GeneMANIA incorporates studies proving interactions between proteins for model organisms such as *C. elegans*, *Drosophila*, zebrafish, mouse and humans. This way I could search for absence and presence of orthologues components of whole pathways and GRNs in our genome and in my transcriptome data.
Phylogenetic analysis: Retrieved protein sequences were aligned using the programme "Clustal ω" (version 1.2.0; Sievers et al. 2011) using standard parameters. Afterwards, appropriate models for correct amino acid replacement were identified by the programme "Prottest3" (Abascal et al. 2005, Darriba et al. 2011). Search for appropriate models was done under consideration of invariant amino acid sites. Found models were confirmed by Bayesian information criterion (BIC) and Akaike information criterion (AIC) (Abascal et al. 2005). Based on the alignments provided by Clustal ω and the amino acid replacement model suggested by Prottest3, phylogenetic trees were created using

the programme "RAxML" for 100 bootstraps (Stamatakis 2006).

2.15 OrthoMCL for early transcriptomes

Transcriptome data retrieved from libraries for all sequenced species during early embryogenesis (Tab. 2.3) was retrived from the Cologne Center for Genomics (section 2.7). Illumina TrueSeq (version 2) adaptors (section 2.7) were eliminated by the programme "Trimmomatic" using stardard parameters for paired-end reads (Bolger et al. 2014). I filtered RNAseq sequencing reads, which were prone to false positve and false negative nucleotides at the 3'- and 5'-ends of each read, by using the software tool "Sickle" (https://github.com/najoshi/sickle). Afterwards, I used the *de-novo* sequence assembler for transcriptome data "Trinity" (version r20140717; Grabherr et al. 2011) to assemble the reads to a transcriptome. I used Trinity for paired-end reads for each replicate of raw data retrieved from the CCG for a single species and early developmental stage (Tab. 2.3). I reduced necessary memory resources required by Trinity by setting the "min_kmer_cov" to the value 2. Generation of transcriptomes by Trinity was done on the local CHEOPS computer cluster at the University of Cologne (http://rrzk.uni-koeln.de/cheops.html?&L=0).

Generated transcriptomes were mapped against reference transcriptomes (section 2.12) by usage of the mapping tools "Bowtie2" (Langmead and Salzberg 2012) and "BLAT" (Kent 2002). I used the programme "Samtools" (version 0.2.0-rc12-1-gbbe85a9; Li et al. 2009) to extract only sequences which mapped to the reference transcriptomes. All other sequences which were either the result of bacterial or genomic contamination were excluded by this step. Hence, only transcriptomic data which was in concordance with reference transcriptomic data was used for the postassembly analysis.

In order to identify early expressed orthologues among seven species (section 1.4; Fig. 1.1) expressed transcripts were translated into protein using the programme "Transdecoder" (section 3.6; Haas et al. 2013). For Transdecoder I used standard parameters, I only changed the minimum translated protein length from 100 amino acids to 50 amino acids. Translated amino-acid sequences from the transcriptomes of each species were used as input for the algorithm OrthoMCL (Li et al. 2003). For setting up the necessary database, aligning all incorporated protein sequences from all species (the so-call "all-vs-all blast

step") and the clustering of orthologous sequences I followed the instructions of Fischer et al. (2011). As database served the open-source database MariaDB (version 10.0; mariadb.com). For the "all-vs-all blast step" I used "Legacy BLAST" (version 2.2.26; Altschul et al. 1990) and for the clustering step I used "Mcl" (version 14.137; Enright et al. 2002).

For analysis of the identified orthologue clusters I wrote custom Perl scripts (Appendix B). These scripts allowed custom searches for clusters shared or excluded from certain species or the custom search for orthologues of a single protein of interest in the clustering.

2.16 Genome predictions, allele analysis and single copy gene analysis in *Diploscapter coronatus*

Genome sequencing, assembly and gene predictions:

DNA of *Diploscapter coronatus* (strain PDL 0010; originally provided by P. De Ley, University of California, Riverside) was collected and pooled from mixed-stage individuals. A genome assembly was constructed in a hybrid assay by the Kohara laboratory (personal communication Yuji Kohara, Hiroshi Kagoshima, Hideaki Haruki). This genome was available to me due to a cooperation between the Kohara (Yuji Kohara, Mishima, Japan) and the Schierenberg laboratory.

Augustus protein prediction and identification of allelic gene variants:

We used Augustus (version 2.5.5; http://bioinf.uni-greifswald.de/augustus/; Stanke and Morgenstern 2005, Stanke and Waack 2003) to *de-novo* predict genes in the *D. coronatus* draft genome. We trained Augustus using a gene set derived from the CEGMA pipeline and for additional hints *D. coronatus* EST libraries were used (retrieved from the Kohara laboratory; personal communication Yuji Kohara, Hiroshi Kagoshima, Hideaki Haruki). This way, we determined the intron-exon structure and the positions on the genome for all of the predicted genes. By aligning and clustering all EST libraries making use of Cd-hit (Li and Godzik 2006, Fu et al. 2012) at a threshold of 90% identity, we identified ESTs belonging to the same gene. By re-mapping clustered ESTs against the genome

using Blat (Kent 2002), we confirmed the exact positions of each predicted gene on the genome. Augustus gene predictions, which were confirmed on EST-level and re-confirmed on protein level by an all-versus-all protein blast approach (at a threshold of 98% site-specific identity; Altschul et al., 1990) to belong to each other, were considered alleles of the same gene under the condition that these alleles are positioned on different contigs. Taking into account the relative position on the genome, we deduced the number of alleles by counting positions and contigs.

OrthoMCL clustering and identification of the presence and absence of orthologues for early expressed *C. elegans genes*:

To reliably compare orthologues we applied the OrthoMCL clustering pipeline including proteomes of 6 *Caenorhabditis* species, *D. coronatus*, *P. pacificus*, *Panagrellus redivivus* and *A. suum* (Tab. 2.12).

Gene expression data for the blastomeres AB and P1 from the *C. elegans* 2-cell stage, were retrieved from Hashimshony et al. (2012). The OrthoMCL clustering was used to identify orthologous genes in other species. The absence of *C. elegans* genes in the *D. coronatus* genome was confirmed by reciprocal BLAST search.

Table 2.12: Nematode genomes used for clustering orthologues between *D. coronatus* 6 *Caenorhabditis* species and outgroups. (*) The genome of *D. coronatus* was sequenced in the Kohara laboratory and could be accessed and used for further analysis due to cooperation with the Kohara laboratory (Yuji Kohara, Mishima, Japan).

species	**source**
Ascaris suum	Wang et al. (2012)
Caenorhabditis angaria	Mortazavi et al. (2010)
Caenorhabditis briggsae	Gupta and Sternberg (2003), Stein et al. (2003)
Canorhabditis elegans	*Caenorhabditis elegans* Sequencing Consortium (1998)
Canorhabditis japonica	www.wormbase.org/db/gb2/gbrowse/c_japonica
Canorhabditis sp. 11	genome.wustl.edu/pub/organism/ Invertebrates/Caenorhabditis_sp11_JU1373
Diploscapter coronatus	Kohara laboratory*
Panagrellus redivivus	Srinivasan et al. (2013)
Pristionchus pacificus	Dieterich et al. (2008)

Single-copy gene analysis:

Single-copy genes present in nematodes of clades I, III, IV and V, as well as in *Drosophila melanogaster* and *Saccharomyces cerevisiae* were retrieved from Mitreva et al. (2011).

Eleven single-copy genes were arbitrarily selected. *D. coronatus* and *C. remanei* orthologues were identified by OrthoMCL. For *C. remanei*, candidates were identified using the data from Mitreva et al. (2011) and for *D. coronatus* by re-blasting the *C. elegans* orthologues against the *D. coronatus* genome. We compared orthologues by pairwise alignments of *D. coronatus* alleles against each other and the *C. elegans* alleles against the *C. remanei* sequence in clustalW (Thompson et al. 1994).

Orthologues of single-copy genes were scanned for conserved protein domains with InterProScan (http://www.ebi.ac.uk/interpro/; Quevillon et al. 2005, Hunter et al. 2012). We counted the number of synonymous and non-synonymous mutations within the respective single-copy genes by pairwise alignment of the sequences on the nucleotide level, taking into account the appropriate reading frame by using the KaKs Calculator (version 1.2) with standard parameters (http://evolution.genomics.org.cn/software.htm; Zhang et al. 2006). For statistics of average nucleotide exchange rates, the one-tailed Welch t-test for non-equal variances was applied. We tested specifically the null hypothesis, i.e. whether the portion of non-synonymous mutations is equal or greater than the portion of synonymous mutations. The null hypothesis was rejected at a confidence level of α >0.01.

D. coronatus ITS, SSU and LSU rDNA analysis:

For each *D. coronatus* rDNA gene, two individuals were picked and lysed. This procedure was done by Theresa Vogt in the Schierenberg laboratory (Theresa Vogt granted me permission to use this data here; Theresa Vogt personal communication; Vogt 2012). Using single-worm PCR Williams et al. 1996) she cloned sequences from each rDNA gene and individual into separate pBluescript KS cloning-vectors. For amplification of the ribosomal small subunit (SSU) we used primers described in Holterman et al. (2006) and Floyd et al. (2002), for the ribosomal large subunit (LSU) primers from Sonnenberg et al. (2007) and for the ribosomal interspacer (ITS) from Vrain (1993).

Chapter 3

Results

3.1 Microarray analysis in combination with RNAi finds new candidate genes for the *C. elegans* oocyte-to-embryo transition and the oocyte-maturation pathway

The oocyte-to-embryo transition in *C. elegans* is mainly coordinated by the MAP-kinase pathway (see introduction; Miller et al. 2003). Even though, much effort was put into the elucidation of the underling gene regulatory network, the molecular regulators upstream and the transcription factors, as well as the downstream targets still need to be identified (see introduction; Fig. 1.3). Hence, we sought to elucidate the biological process of oocyte maturation in *C. elegans* further by using a combination of RNA interference (RNAi) and microarray analysis. RNAi allows knock-downs of specific target genes in the already elucidated pathway. Microarrays were used to better understand the changes in gene expression after RNAi knock-down compared to wild-type (see introduction and material and methods).

We designed in total 7 RNAi experiments for known genes of the pathway regulating oocyte-to-embryo transition: *par-5*, *itr-1*, *goa-1*, *lip-1*, *mek-2*, *mpk-1* and *mbk-2*. We performed two successive microarray experiments on RNAi-treated *C. elegans* mixed cultures in two replicates (Fig. 3.1 and 3.2; Materials and Methods).

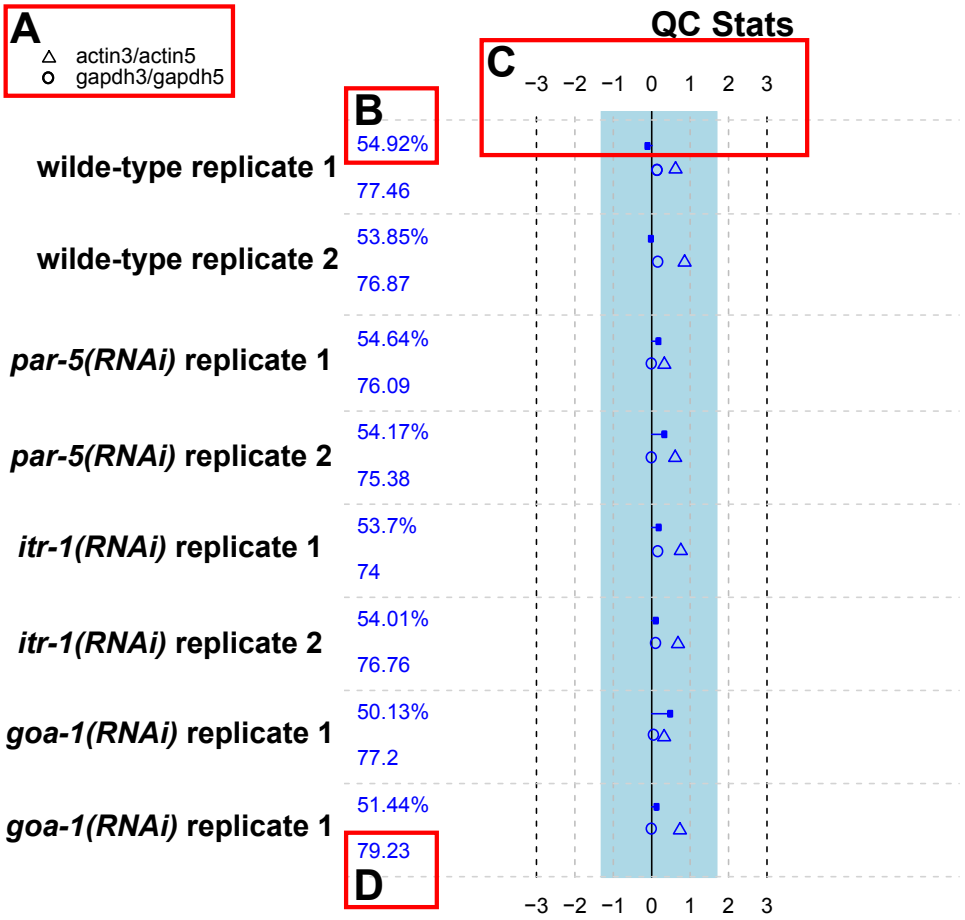

Figure 3.1: Quality control (QC) plot after Wilson and Miller 2005 was used to visualise the quality of the RNAi knock-downs of *par-5(RNAi)* and *itr-1(RNAi) goa-1(RNAi)*. Each row represents one used experimental replicate. This plot includes four quality criteria: (**A**) The expression ratio of 5'- to 3'-probes of the house-keeping genes *actin* and *gapdh* (*Glyceraldehyde-3-phosphate dehydrogenase*) for RNA quality assessment, (**B**) the percentage of identified genes of the complete microarray, (**C**) the scale factor of the Mas5 algorithm (see Materials and Methods), indicating the average expression differences between all microarrays, and (**D**) the signal background noise of each microarray (see Materials and Methods). All values of the quality criteria are within reference margins (highlighted in blue).

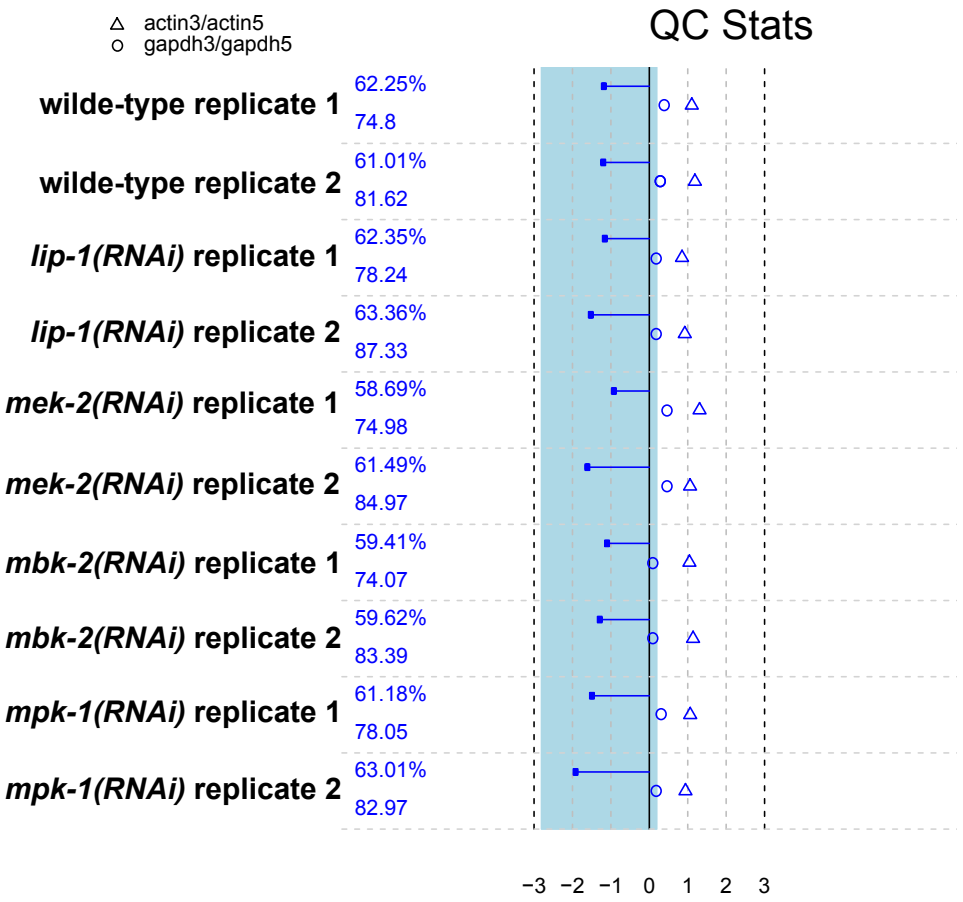

Figure 3.2: Quality control (QC) plot after Wilson and Miller 2005 to visualise the quality of microarray results for RNAi experiments for *lip-1(RNAi)*, *mek-2(RNAi)*, *mbk-2(RNAi)*, *mpk-1(RNAi)*. Even though, there seemed to be a shift for the Mas5 scale factors (see Materials and Methods), this phenomenon was observed similarly for all microarrays and hence all of them are comparable to each other. Therefore, all quality control criteria were met (highlighted in blue).

In order to understand the expression changes of *C. elegans*, we used Mas5 (Hubbell et al. 2002) algorithm to normalise the expression data (Materials and Methods). Furthermore, I normalised the expression values of all RNAi experiments to compare the expression levels of the different RNAi experiment versus wild-type control. By plotting the relative expression differences between our RNAi experiments and the wild-type control, against the "rank"of the genes (see above) on the different microarrays, we found that expression follows a regression line (Fig. 3.3 B). There was no correlation between foldchange of the different experiments and the average expression values (Fig. 3.3 B), indicating that these values were (1) not biased by the experimental setup and (2) average fold-change values ranged around zero (Fig. 3.3 B) and Tab. 3.1). This indicates that all expression values were properly normalised, thus the conditions for a fair comparison between RNAi knock-downs and wild-type control were given. Since, most of

the differences between our RNAi approaches and the wild-type control were correlated (Fig. 3.4 A and Tab. 3.1), I am confident that the underlying microarray data are conclusive.

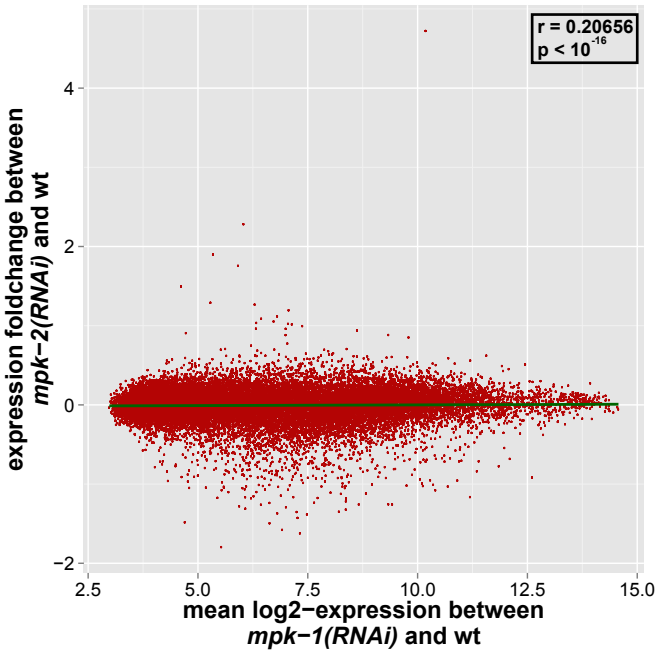

Figure 3.3: Additional qualitiy control testing for the microarray experiments of *lip-1(RNAi)*, *mek-2(RNAi)*, *mbk-2(RNAi)* and *mpk-1(RNAi)*. MA-plot (Materials and Methods) for *mpk-1(RNAi)* versus wild-type expression. The average expression level ratio of *mpk-1(RNAi)* and wild-type (X-coordinate) is plotted against the foldchange ($\log_2$ expression ratio). The regression line (marked in green) shows no correlation between foldchange and average expression levels (Pearson's r equals 0.207; p-value $<10^{-16}$; Tab. 2.6; see Materials and Methods).

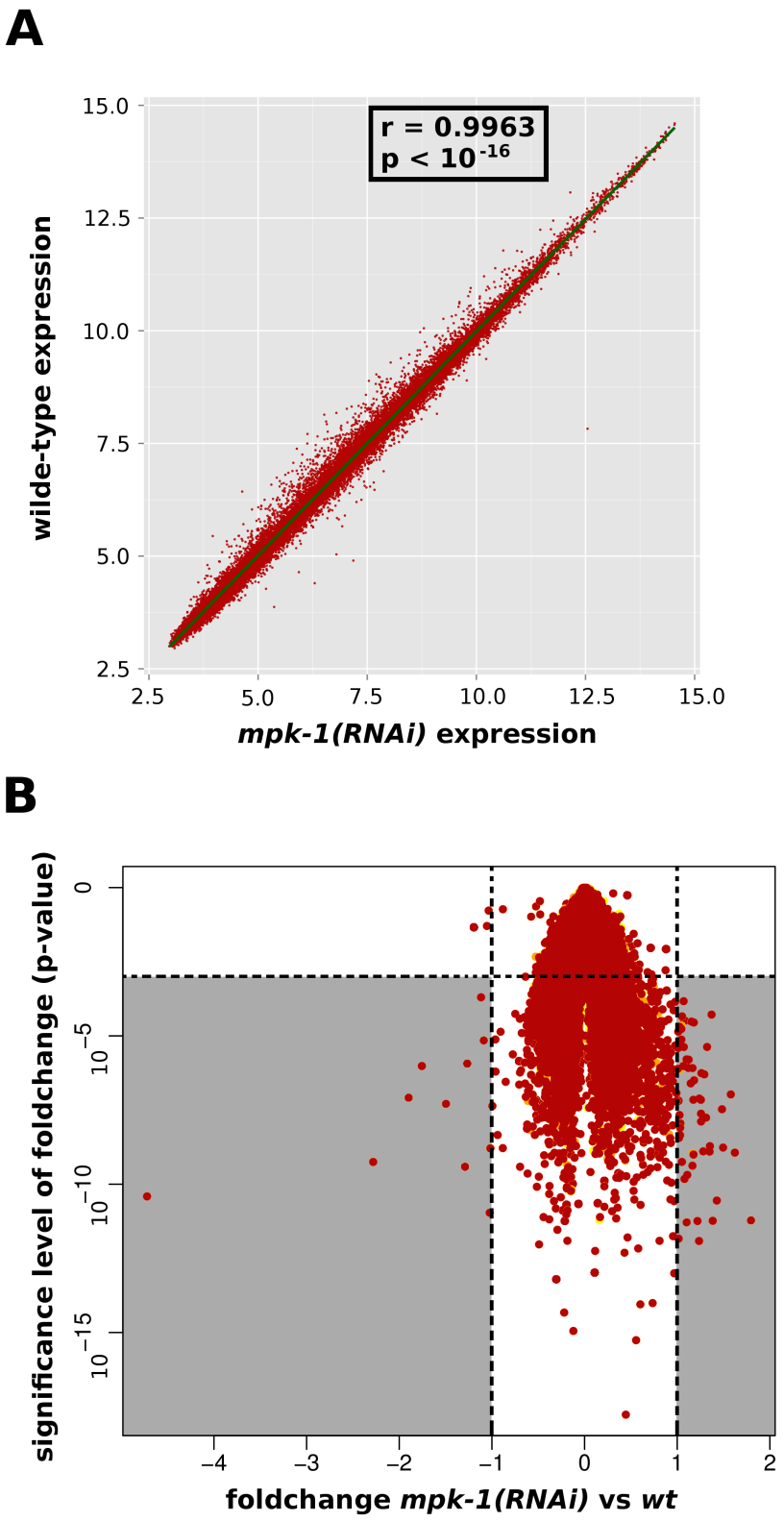

Figure 3.4: (**A**) Expression of *mpk-1(RNAi)* versus wild-type. Both expression values are positively correlated to each other (represented by green regression line; Pearson's r = 0.0996; p-value $<10^{-16}$). (**B**) Volcano plot representing the relative expression values as foldchange (binary logarithmic representation; $\log_2$) between *mpk-1(RNAi)* and wild-type in comparison to the p-value of the underlying significance test for expression values (two-tailed t-test; Materials and Methods). Two criteria were used to identify significantly over- and/or underexpressed genes (marked by grey background): (1.) A foldchange of greater than 1.0 (indicated by vertical dashed lines) and (2.) a significance level of p $<10^{-3}$ (horizontal dashed line).

Table 3.1: Results for the correlation and the foldchange data of RNAi knockdowns vs wild-type and their respective significance values (p-values). Correlations were tested by Pearson's r (Materials and Methods).

RNAi knock-down	**foldchange correlation**	
	Pearson's r	**p-value**
par-5(RNAi)	0.989	$< 2.2 \cdot 10^{-16}$
itr-1(RNAi)	0.975	$< 2.2 \cdot 10^{-16}$
goa-1(RNAi)	0.995	$< 2.2 \cdot 10^{-16}$
lip-1(RNAi)	0.996	$< 2.2 \cdot 10^{-16}$
mbk-1(RNAi)	0.991	$< 2.2 \cdot 10^{-16}$
mek-2(RNAi)	0.995	$< 2.2 \cdot 10^{-16}$
mpk-1(RNAi)	0.996	$< 2.2 \cdot 10^{-16}$

I screened for candidate genes for two criteria: (1.) An expression foldchange of at least 1.0 (equalling a 2-fold over-/underexpression; see Materials and Methods; Fig. 3.4 B horizontal lines). (2.) A significant change in expression (Fig. 3.4 B, horizontal dashed lines).

I found in total 1,458 genes which conformed both criteria (Tab. 3.2). These genes might be potential upstream and/or downstream targets of the MAP-kinase pathway (see introduction; 1.3). I searched for relevant candidate genes of oocyte-to-embryo transition, using the wormbase database (wormbase.org). Here I screened for RNAi phenotypes related to oocyte-to-embryo transition (see Materials and Methods). I found in total 152 potential candidates (Tab. C.1).

Promising candidate genes were then selected to test their role in oocyte-to-embryo transition by performing RNAi experiments (Tab. 3.2). For this purpose, I used RNAi constructs specific for the candidate genes (Kamath et al. 2003) to test if these genes are relevant for the MAP-kinase GRN in *C. elegans* (see section 3.2).

Table 3.2: Candidate genes correlated with oocyte-to-embryo transition. Selected candidates are presented in this table. (*) RNAi experiment(s) which identified potential upstream and/or downstream genes for the MAP-kinase signaling during oocyte maturation in *C. elegans*.

RNAi knock-down(*)	**affected candidate gene**	**foldchange (log_2)**	**expression change**	**protein class**	**function**
par-5(RNAi)	*inx-19*	−2.8	−6.9	innexin	cell-cell interaction
par-5(RNAi)	*unc-7*	3.1	8.6	innexin	cell-cell interaction
itr-1(RNAi)	*plx-2*	−1.3	−2.5	plexin	cell
goa-1(RNAi)		-1.5		receptor	membrane
par-5(RNAi)	*str-102*	2.4	5.3	serpentine	cell
par-5(RNAi)		2.4	5.3	receptor	membrane
itr-1(RNAi)	*cng-1*	4.2	18.4	cyclic	ion channel
goa-1(RNAi)		4.0	16.0	nucleotide	activity
goa-1(RNAi)		4.0	16.0	gated channel	
par-5(RNAi)	*cnd-1*	−1.4	−2.6	helix-loop-helix	DNA
goa-1(RNAi)		-3.3	9.9	transcription factor	binding
itr-1(RNAi)	*aptf-1*	−2.1	−4.3	AP-2 like transcription factor	DNA binding
itr-1(RNAi)	*aptf-4*	−1.6	−3.1	AP-2 like transcription factor	DNA binding
par-5(RNAi)	*nhr-81*	2.6	6.1	nuclear	DNA
itr-1(RNAi)		3.0	8.0	hormone	binding
goa-1(RNAi)		3.4	10.6	receptor	
itr-1(RNAi)	*ztf-11*	−1.3	−2.5	orthologue of	DNA
itr-1(RNAi)		−1.9	−3.7	human ZT18	binding

3.2 RNAi of potential oocyte maturation regulators reveal candidate genes for the gene regulatory network

In order to identify new target genes which are involved in oocyte maturation and oocyte-to-embryo transition in *C. elegans* (see Introduction and section 3.1), I used RNA interference (RNAi) against known target genes of the gene regulatory network (GRN) of the MAP-kinase pathway which controls oocyte-to-embryo transition (Introduction; section 3.1; Miller et al. 2003). My target genes were *par-5*, *goa-1*, *itr-1*, *lip-1*, *mek-2*, *mbk-2* and *mpk-1* (*mpk-1* encodes the MAP-kinase which initiates maturation of the "-1" oocyte; McCarter et al. 1999, Miller et al. 2003).

By extracting the RNA from whole *C. elegans* N2 populations and by loading these samples onto *C. elegans*-specific microarrys (Materials and Methods), I identified 1,458 potential candidate genes for the MAP-kinase signaling pathway (see section 3.1). I screened the nematode database (wormbase.org) for potential genes with RNAi phenotypes correlated with oocyte maturation (Materials and Methods). In total, I found 152 genes with potential functions in this GRN (Tab. C.1). We identifed candidate genes with potential functions upstream of the MAP-kinase MPK-1 (Fig. 1.3), such as innexins (Tab. 3.2) involved in the communication between somatic sheath cell with the oocyte (see Introduction). I also found transcription factors potentially involved in oocyte maturation (Tab. 3.2; section 3.1).

In order to verify the function of these candidate genes in oocyte maturation I used RNAi. For this purpose, I brought parental *C. elegans* adults onto plates covered by double stranded RNA (dsRNA) expressing bacteria (Timmons and Fire 1998; Materials and Methods). The first generation of laid eggs was raised to adulthood and at the time-point of the first egg-laying by the F_1-generation I started scoring phenotypes (Materials and Methods). As a suitable positive control, I applied RNAi against the known target gene *mpk-1* of the Map-kinase pathway (Miller et al. 2003). This verified the specificity of our RNAi assay (Fig. 3.5, 3.6). To exclude unspecific RNAi effects, we applied RNAi against the GFP sequences, absent in nematodes (Fig. 3.5, 3.6).

In my RNAi experiments against the selected candidate genes I scored for absent or reduced hatched larvae and gonad aberrances. This way, I could narrow down the list of interesting candidates to 8 genes (Fig. 3.5). For these 8 candidate genes I found a significantly reduced number of hatched larvae for each time-point within 72 hours

(Fig. 3.5, 3.6) in comparison to our negative control (*GFP(RNAi)*).

By using Differential Interference Contrast microscopy (DIC), I was able to determine the reason for these observed phenotypes (Fig. 3.7). In the cases of *inx-19(RNAi)*, *nhr-81(RNAi)*, *plx-2(RNAi)* and *str-102(RNAi)* the amount of box-shaped oocytes (Fig. 3.7 D, F, H, J; arrows) was strongly increased in comparison to the *GFP(RNAi)* control (introduction; McCarter et al. 1999). These oocytes were much smaller and in the cases of *inx-19(RNAi)* and *nhr-81(RNAi)* were even visible in the distal part of the gonad, which is normally filled with small oogonia only (Fig. 3.7 D, F; open arrows).

Generally, the oocytes seemed to be densely packed in contrast to the *GFP(RNAi)* control. There was no ovoid shaped oocyte adjacent to the spermatheca, as known from wild-type (McCarter et al. 1999) and the GFP negative control (Fig. 3.7 B; marked by o′). Other markers of oocyte maturation, such as the nuclear envelope breakdown (see Introduction; Fig. 1.2), could not be recognized, too (Fig. 3.7 D, F, H, J). This means that there were no maturing oocytes present in the RNAi phenotypes of *inx-19(RNAi)*, *nhr-81(RNAi)*, *plx-2(RNAi)* and *str-102(RNAi)*. Hence, it seems that maturation did not occur in these RNAi-treated animals. This explains the significantly decreased number of hatched larvae (Fig. 3.5 and 3.6). By comparing the uterus with the *GFP(RNAi)* control, it became clear that embryos were either absent, as in the cases of *inx-19(RNAi)* and *str-102(RNAi)* (Fig. 3.7 E, K), or most of the embryos in the uterus were degenerated (Fig. 3.7 G, I; arrows).

Henceforth, I could confirm 8 candidate genes to be involved in the process of oocyte-to-embryo transition in *C. elegans*. Thus, a broad RNAi screening approach for my set of 1,458 candidate genes (see section 3.1) has the potential to identify many more regulators for the process of oocyte maturation and the underlying GRN in *C. elegans*.

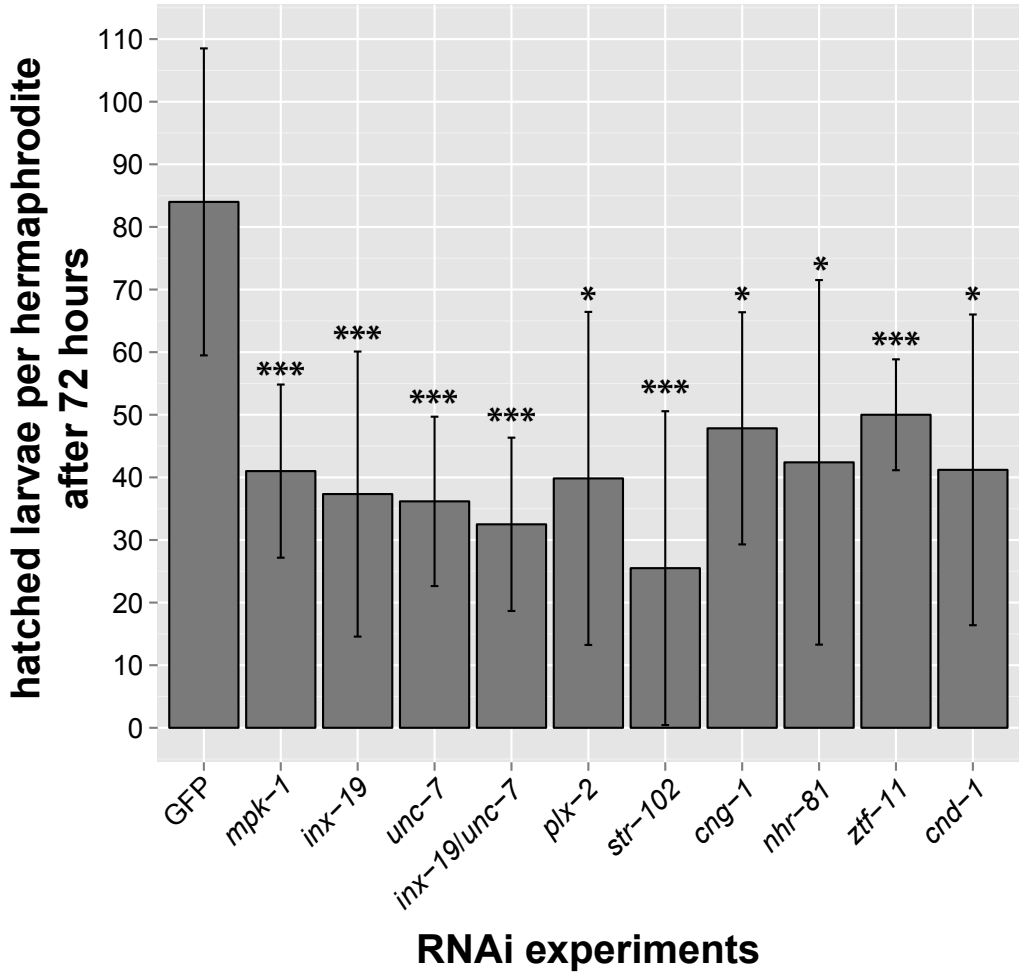

Figure 3.5: Bar plot representing the total amount of hatched larvae after 72 hours in different RNAi experiments. *GFP(RNAi)* was used as negative control and *mpk-1(RNAi)* as positive control. Differences concerning the number hatched larvae between both RNAi approaches were highly significant (one-tailed Welch test for variable variances at a significance level of α <0.001; marked by three asterisks). RNAi approaches for *inx-19*, *unc-7*, *str-102* and *ztf-11* also showed a highly significant reduction of hatched larvae in comparison to the negative control (*GFP(RNAi)*). In the case of the double RNAi knock-down of *inx-19* and *unc-7* the number of hatched larvae was even further reduced. In the case of the RNAi approaches for *plx-2*, *cng-1*, *nhr-81*, *ztf-11* and *cnd-1* the number of hatched larvae was still significantly reduced (one-tailed Welch test at a significance level of α <0.01; marked by a single asterisk).

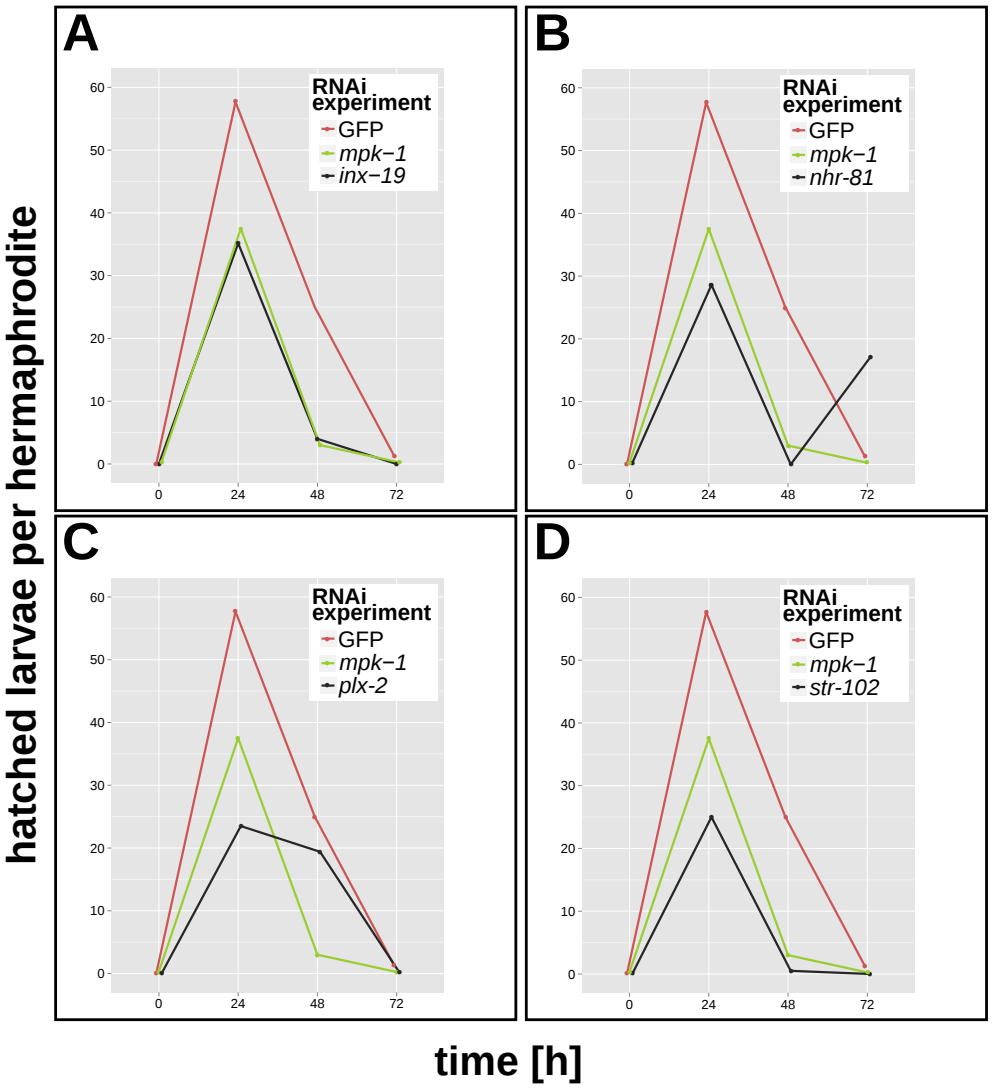

Figure 3.6: Average number of hatched larvae for (**A**) *inx-19(RNAi)*, (**B**) *nhr-81(RNAi)*, (**C**) *plx-2(RNAi)* and (**D**) *str-102(RNAi)* over a time course of 72 hours (marked as black curves). The negative control (*GFP(RNAi)*) is marked in red and the positive control (*mpk-1(RNAi)*) is marked in green.

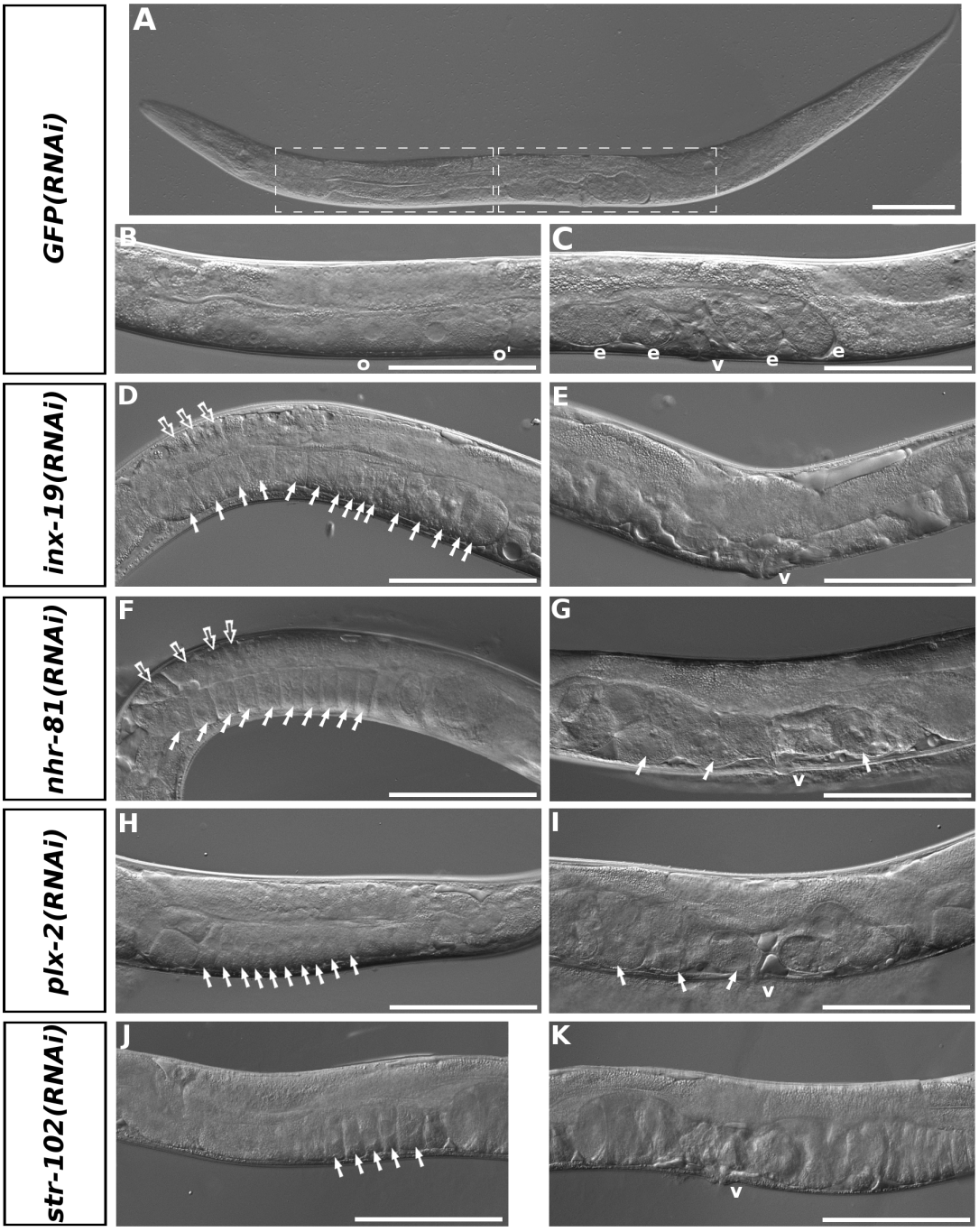

Figure 3.7: Differential Interference Contrast (DIC) images of selected RNAi phenotypes. **(A, B, C)** *GFP(RNAi)* was used as negative control to exclude unspecific RNAi effects. **(B, C)** Magnified images of the gonad (B) and the uterus (C) of an adult hermaphrodite after treatment with *GFP(RNAi)* (A; dashed squares). **(D, F, H, J)** Gonads of adult hermaphrodites after RNAi treatments. Aberrant oocytes in proximal gonad are marked by arrows. Aberrant oocytes in distal gonad are marked by open arrows. **(E, G, I, K)** Uteri of adult hermaphrodites after different RNAi treatments. Abberant embryos are marked by arrows. All animals were screened for RNAi effects about 24 hours after reaching adulthood. Anterior is to the left and dorsal to the top. Abbreviations: o = oocyte; o′ = maturing "-1" oocyte; e = embryo; v = vulva. Scale bars: 100 μm.

3.3 *C. elegans* gonad formation, spermatogenesis and oogenesis indicate clade V-specific signaling events

The *C. elegans* gonad is an essential and versatile organ. The gonad develops postembryonically, guided by the distal tip cell (DTC) and harbours later the maturing oocytes. The DTC generates the niche for the germline stem cells (GSCs; introduction; Kimble and Crittenden 2007). The whole process from initial differentiation of the DTCs, through the formation of the male and female gametes, up to the final maturing oocytes and their fertilisation, is precisely controlled (for a review on germline development refer to Byrd and Kimble 2009, Kimble and Crittenden 2007).

The question arises whether the known *C. elegans* genes controlling tissue development and functionality are found in other nematode species as well. To answer this question I used several approaches: (1) OrthoMCL clustering (Materials and Methods) to identify orthologues of the *C. elegans* germline set (Reinke et al. 2004, 2000) to identify whether orthologues were present or absent in other nematodes. (2) I screened known *C. elegans* genes for key processes such as gonad (Kimble and Crittenden 2007), stem-cell (Byrd and Kimble 2009, Nadarajan et al. 2009), P granule formation (Voronina 2013) and oocyte maturation (Miller et al. 2001, 2003, Govindan et al. 2006) for present or absent orthologues in other nematodes. In total I screened 620 genes known to be important for either of the aforementioned processes affecting the gonad for orthologues in other clades and I found that the majority (84%) of Enoplean and Chromadorean species carry homologues of these genes. 98 genes (16 %) of this set did not show any homologous sequences. In order to get a deeper insight into the functionality of this gene set, I compared 50 genes which play a role in gonad-associated processes (Fig. 3.8). Interestingly, for each of the aforementioned gonad-associated processes I detected genes for which orthologues are absent in non-*Caenorhabditis* species (Fig. 3.8).

In order to correlate gonad-associated processes with underlying GRNs, it is necessary to give detailed background information for the model *C. elegans*. For easier reference, this background information is given here and not in the introduction.

Background information: The gonad formation initially starts with the extensive postembryonic cleavage of gonad precursor cells, (Z1 to Z4; Fig. 3.8 A), where the somatic gonad forms from Z1 and Z4, while the inner pair (Z_2 and Z_3) generate all germ cells (Friedman et al. 2000). Lateral signaling between the different Z cells determines the fate of Z1 and Z4, including formation of the DTCs which depends on GON-4 activity (Fig. 3.8 A). The germ cell precursors Z2 and Z3 are controlled via Wnt signaling through the frizzled receptors LIN-17 and MOM-5, the disheveled orthologues DSH-2 and MIG-5, the β-catenin SYS-1 and the LEF/TCF transcription factor POP-1 (Fig. 3.8 A; Kimble and Crittenden 2007). The hermaphroditic gonad of *C. elegans* produces first sperm and later shifts to oocyte production (Fig. 3.8 B). The so-called "global sex determination" pathway determines the switch from spermatogenesis to oogenesis (Fig. 3.8 B; Kimble and Crittenden 2007). This pathway depends on key-regulator proteins as for example FEM-3, FOG-1/-3, TRA-1 and DAZ-1 (Fig. 3.8 B; Kimble and Crittenden 2007). Furthermore, oogenesis also depends onto different signaling pathways which regulate the stem cell pool of mitotically dividing oogonia (Introduction; Kimble and Crittenden 2007) the entry of oogonia into meiosis, pachytene and diplotene silencing during prophase I, oocyte maturation and P granule formation (Fig. 3.8 C; Kimble and Crittenden 2007). The pool of mitotically dividing oogonia depends on the interaction with the DTC (Introduction). Here, Notch signaling via the DSL ligand LAG-2 and the Notch receptor GLP-1, is crucial (Kimble and Crittenden 2007). The Notch signal suppresses meiosis in the vicinity of the DTC (Fig. 3.8 C; Kimble and Crittenden 2007). Meiosis staring with prophase I is arrested in pachytene and then diplotene (Introduction; Fig. 1.2). Both prophase I arrests are controlled by MES-2/-3/-4 and -6 (Fig. 3.8 C; Holdeman et al. 1998, Kelly and Fire 1998, Fong et al. 2002). During diplotene arrest FLH-1 and -3 inhibit microRNAs which are supplied as maternal gene products (Walhout et al. 2002, Ow et al. 2008).

During oocyte maturation the diplotene arrest is lifted from the most mature (-1) oocyte and meiosis is resumed (Fig. 3.8 C grey circle, 1.2). During this process bivalents and chromatids are separated. In the case of *C. elegans* this is facilitated by the DNA binding proteins ZIM-1/-2/-3 and HIM-8 (Fig. 3.8 C; Phillips and Dernburg 2006). During oocyte maturation Anaphase-Promoting Complex (APC) activity is necessary, in the case of *C. elegans* the APC5 subunit orthologue GFI-3 is required for this step (Fig. 3.8 C; Stein et al. 2010). Also during oocyte maturation P granules form, which are protein and mRNA aggregates (including e.g. GLH-1/-4, NOS-2; Fig. 3.8 C, small

black dots; Voronina 2013) required for embryogenesis.

Screening the above mentioned regulatory genes for orthologues in clade I to V nematdoe genomes showed that there are key regulators during all above-mentioned processes of gonad formation and germline activity which are only available in clade V nematodes (Fig 3.8): This includes DTC cell fate determination which is regulated by a cascade including Wnt signaling and the β-catenine SYS-1 (Fig 3.8 A) (Kimble and Crittenden 2007), as well as the switch from sperm to oocyte production which includes the sex-determination genes FEM-3 and FOG-1 (Kimble and Crittenden 2007) (Fig 3.8 B). Intriguingly, oogonia and stemcell maintenance is also involved, as the DSL ligand LAG-2 is only present in clade V nematodes (Fig 3.8 C, Kimble and Crittenden 2007). Furthermore, the process of oocyte formation shows key regulators present only in clade V, such as the MES-3 gene of the zygotene and diplotene silencing complex (Holdeman et al. 1998, Kelly and Fire 1998, Fong et al. 2002). Even crucial steps of chromosome pairing and separation during meiosis I and II involve clade V-specific chromosome adaptor genes such as ZIM-1/-2/-3/ and HIM-8 (Phillips and Dernburg 2006).
My results show that even though the majority of genes associated with gonad development, including spermatogenesis, oogenesis and oocyte maturation are conserved among clade I to V nematodes, important key regulators in each of the described processes are specific for clade V. Hence, crucial steps during these processes must be controlled by different proteins or even pathways outside of clade V.

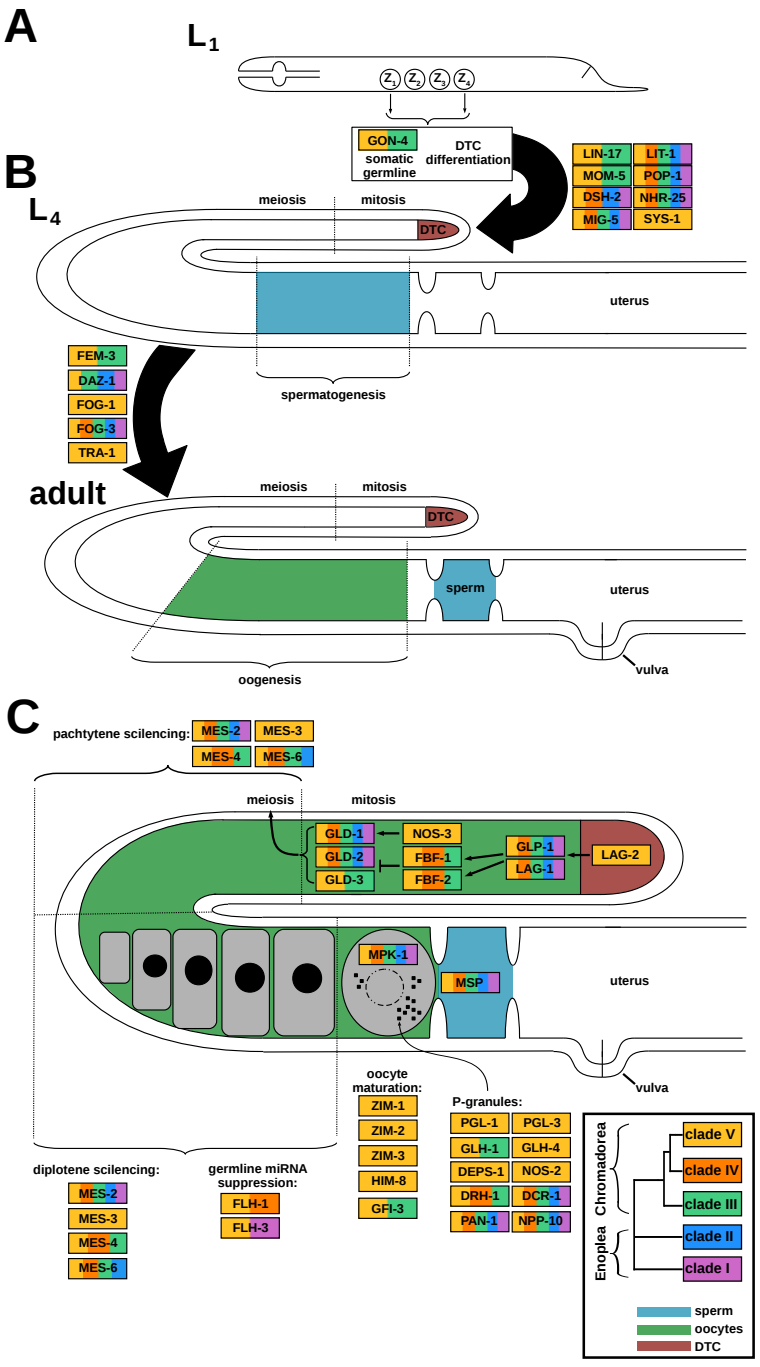

Figure 3.8: Schematic representation of the processes of (**A**) Z-cell primordium and transition to L_4 larval stage,(**B**) shift from spermatogenesis (during L_4 larval stage) to oogenesis (during young adulthood) and (**C**) oogenesis in *C. elegans*. All four processes are controlled by different interacting genes (see text). Identified orthologues of controlling genes in clade I to V nematodes are shown in boxes with clade-specific colours.

3.4 RNAseq of *R. culicivorax* virgins and males finds orthologues for oogenesis from *Drosophila* and *vertebrates*

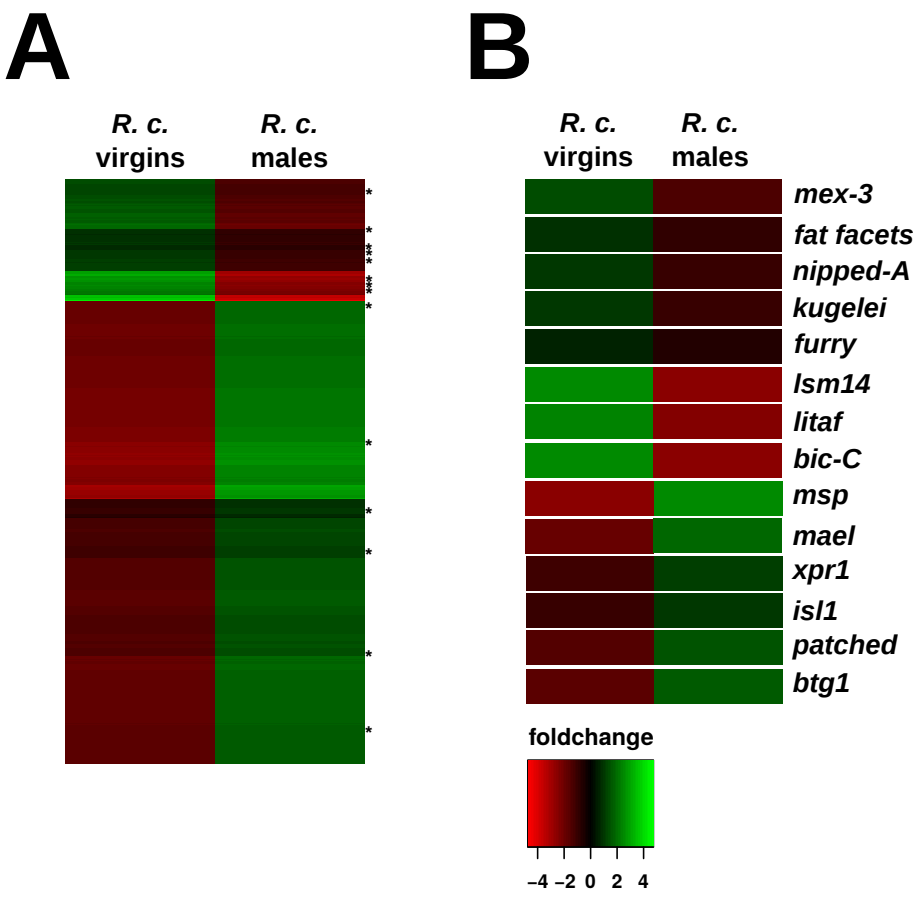

Figure 3.9: Heatmaps visualising the differential expression between adult virgins and males in *R. culicivorax*. (**A**) Significant differential expression i. e. over- and underexpression of 1,378 transcripts ($p<10^{-3}$, FDR <0.05). A foldchange of -2 < $\log_2$, or > +2 equals an over/-underexpression of at least 4-fold. By applying our Interproscan pipeline and Blast2GO (Materials and Methods), I identified several orthologues known to be involved in gonad-associated processes in other species (marked by asterisks) (**B**) Identified expressed orthologues of known regulators for oogenesis, spermatogenesis and gonad formation in *R. culicivorax*. These genes represent only a sub-set of the 1,378 significantly over- and/or underexpressed genes.

After analysing the *Caenorhabditis* gonad, its formation and oogenesis and finding a set of crucial genes which are specific for clade V nematodes (section 3.3), I concentrated on the clade I nematode *R. culivorax* (Fig. 1.2) to find candidate genes involved in Enoplean gonad development. I sequenced the transcriptome of unfertilized adult females (virgins) and compared them to the transcriptome of adult males (Materials and Methods). With this approach I intended to identify expressed transcripts relevant for gonad formation,

	genomes																
	N	I		II	III				IV			V					
	G. s. [a]	*R. c.* [a]	*T. s.*	*E. b.*	*A. s.* [a]	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* [a]	*D. c.* [a]	*P. p.* [a]	*C. a.*	*C. b.*	*C. r.*	*C. e.* [a]
MEX-3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Faf	+**	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Nipped -A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kugelei	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fry	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
LSM14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
LITAF	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bic-C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MSP	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MAEL	+**	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
XPR1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ISL1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ptc	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BTG1	+**	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 3.10: Phylum-wide comparison of orthologues involved in gonad development, oogenesis and spermatogenesis of nematode and nematomorph species. Identified orthologues are symbolised by "+" and absent orthologues are symbolised by "−". For genes where orthologues were expected, but were not detected by OrthoMCL clustering, genome-wide searches were repeated by using Blastp and orthologues were identified by phylogenetic analysis (marked by two asterisks; see Materials and Methods). I-V, clades (Fig 1.1; Blaxter et al. 1998). Abbreviations: *A. s.* - *Ascaris suum*; *B. m.* - *Brugia malayi*; *B. x.* - *Bursaphelenchus xylophilus*; *C. a.* - *Caenorhabditis angaria*; *C. b.* - *Caenorhabditis briggsae*; *C. e.* - *Caenorhabditis elegans*; *C. r.* - *Caenorhabditis remanei*; *D. c.* - *Diploscapter coronatus*; *D. i.* - *Dirofilaria immitis*; *E. b.* - *Enoplus brevis*; *G. s.* - *Gordius sp.*; *L. l.* - *Loa loa*; *M. h.* - *Meloidogyne hapla*; N - nematomorph; *P. p.* - *Pristionchus pacificus*; *P. s.* - *Panagrolaimus sp. PS1159*; *R. c.* - *Romanomermis culicivorax*; *T. s.* - *Trichinella spiralis.*

spermatogenesis, oogenesis and possibly incorporated maternal mRNAs in oocytes. By differential expression analysis I sought to identify target mRNAs restricted to the gonad and female specific reproduction. For this approach, I applied RSEM (RNAseq by expectation maximisation) transcript abundance estimation (Li and Dewey 2011, Li et al. 2010), TMM (trimmed mean of m values) based normalisation (Robinson and Oshlack 2010) and edgeR (empirical analysis of DGE in R) expression comparison (Robinson et al. 2010, Robinson and Smyth 2007, 2008) (see Materials and Methods) to identify significantly over- and underexpressed transcripts between virgins and males (Materials and Methods). In total I found 1,378 differentially expressed genes (Fig. 3.9 A). 225 of these genes were significantly upregulated in virgins, but not in males (FDR < 0.05).

By clustering protein sequences of 15 nematodes, one nematomorph genome and 12 outgroup genomes (Materials and Methods), I identified target genes which are known to be expressed during gametogenesis in model organisms such as *Drosophila*, *Xenopus* and mouse, verifying the efficacy of my RNAseq approach (Fig. 3.9 B). Among known

germline regulators were genes such as *fat facetes*, *furry*, *lsm14* and *bic-C* (Fig. 3.9 B). By using an established pipeline for Interproscan from our laboratory (Materials and Methods) and by combining these data with Blast2GO (Materials and Methods), I identified candidate genes with potential involvement in gonad formation and gametogenesis in *R. culicivorax* (Fig. 3.9 B, 3.10).

These results suggest that most genes involved in gonad formation and gametogenesis in *R. culicivorax* are conserved among the phylum of nematodes. Only for Faf and Mael I identified Enoplean-restricted orthologues (Fig. 3.10). It would be interesting to further investigate their role in Enoplean species in the aforementioned processes of gonad formation and gametogenesis. Additionally, knock-outs of these genes via e. g. the CRISPR/Cas9 system or RNAi, may reveal their potential impact on early embryogenesis in Enoplean species.

3.5 *Ascaris suum* germline-specific genes allow identification of new candidate genes relevant for the germline of *C. elegans* and other Chromadorean nematodes

I screened 686 *Ascaris suum* genes which are eliminated from the soma by a process called "chromatin diminution" during the early asymmetric cleavages of the zygote (Wang et al. 2012). As these genes persist in the germline, but not in the soma, and therefore are germline-associated, I used these genes to screen for potential regulators of gonad formation, oogenesis and spermatogenesis. By using a combination of our established Interproscan pipeline and Blast2GO (Materials and Methods), I was able to identify 32 homologues specific for the germline. Furthermore, I identified on basis of this gene set 10 orthologues of known germline-specific regulators for nematodes of clade III, IV and V (Tab. 3.3). It would be highly interesting to explore whether the remaining undescribed candidate genes play a role in *C. elegans* during gonad formation, oogenesis and/or spermatogenesis.

Table 3.3: Chromadorean-specific germline regulators found in the *A. suum* gene set (Wang et al. 2012). The functions of the respective *A. suum* genes were inferred from blast to gene ontology* (Blast2GO) and protein domain analysis ** (Interproscan; Quevillon et al. 2005, Hunter et al. 2012; Materials and Methods). For a detailed description for GO terms refer to Carbon et al. 2009.

gene name	**clade specific orthologues**	**best blast hit**	**proposed gene function** *	**proposed protein domain function** **
ASU_08876	III	caesin kinase II subunit α (*Xenopus laevis*; *Mus musculus*)	spermatogenesis (GO:0007283)	Serine/threonine- dual specificity protein kinase, catalytic domain (IPR002290)
ASU_10038	III	FEM-3 (*C. elegans*) PUMILIO (*D. melanogaster*)	reproductive process (GO:0022414)	Pumilio RNA-binding repeat (IPR001313)
ASU_10497	III and IV	CREB3L1 (*Mus musculus*; *Homo sapiens*)	spermatogenesis (GO:0007283)	Basic-leucine zipper domain (IPR004827) Eukaryotic transcription factor, Skn-1-like, DNA-binding (IPR008917)
ASU_11448	III	HIM-14 (*C. elegans*); MSH4 (*Homo sapiens*)	pachytene (GO:0000239); synaptonemal complex (GO:0000795); female gamete generation (GO:0007292); spermatogenesis (GO:0007283); ovarian follicle development (GO:0001541);	DNA mismatch repair protein MutS, core (IPR007696) DNA mismatch repair protein MutS, C-terminal (IPR000432)
ASU_11803	III, IV and V	BTF-3 (*C. elegans*) BTF3 (*Mus musculus*)	pachytene (GO:0000239) synaptonemal complex (GO:0000795) female gamete generation (GO:0007292)	DNA mismatch repair protein MutS, core (IPR007696) DNA mismatch repair protein MutS, C-terminal (IPR000432)
ASU_11845	III and V	MSH-5 (*C. elegans*) MSH5 (*Homo sapiens*)	meiosis I (GO:0007127)	DNA mismatch repair protein MutS, core (IPR007696); DNA mismatch repair protein MutS, C-terminal (IPR000432)
ASU_11963	III	PIM1 (*Homo sapiens*)	male meiosis (GO:0007140)	Protein kinase, catalytic domain (IPR000719); Protein kinase, ATP binding site (IPR017441)
ASU_12478	III	GLP-1 (*C. elegans*); Notch 2 (*Homo sapiens*)	developmental process involved in reproduction (GO:0003006)	Ankyrin repeat-containing domain (IPR020683); EGF-like calcium-binding (IPR001881); Epidermal growth factor-like domain (IPR000742)
ASU_13406	III	SMC3 (*Xenopus laevis*; *Homo sapiens*)	meiosis I (GO:0007127); meiotic cohesion complex (GO:0030892); chromosome organization involved in meiosis (GO:0070192)	P-loop containing nucleoside triphosphate hydrolase (IPR027417); RecF/RecN/SMC (IPR003395)
ASU_14618	III	DAZ-1 (*C. elegans*); BOLL (*Homo sapiens*)	single fertilization (GO:0007338); germ cell migration (GO:0008354); germ cell development (GO:0007281); positive regulation of meiosis (GO:0045836)	Nucleotide-binding, alpha-beta plait (IPR012677); RNA recognition motif domain (IPR000504)

3.6 Establishment of RNAseq post-sequencing data analysis - a comprehensive workflow to generate transcriptomes of satisfying quality

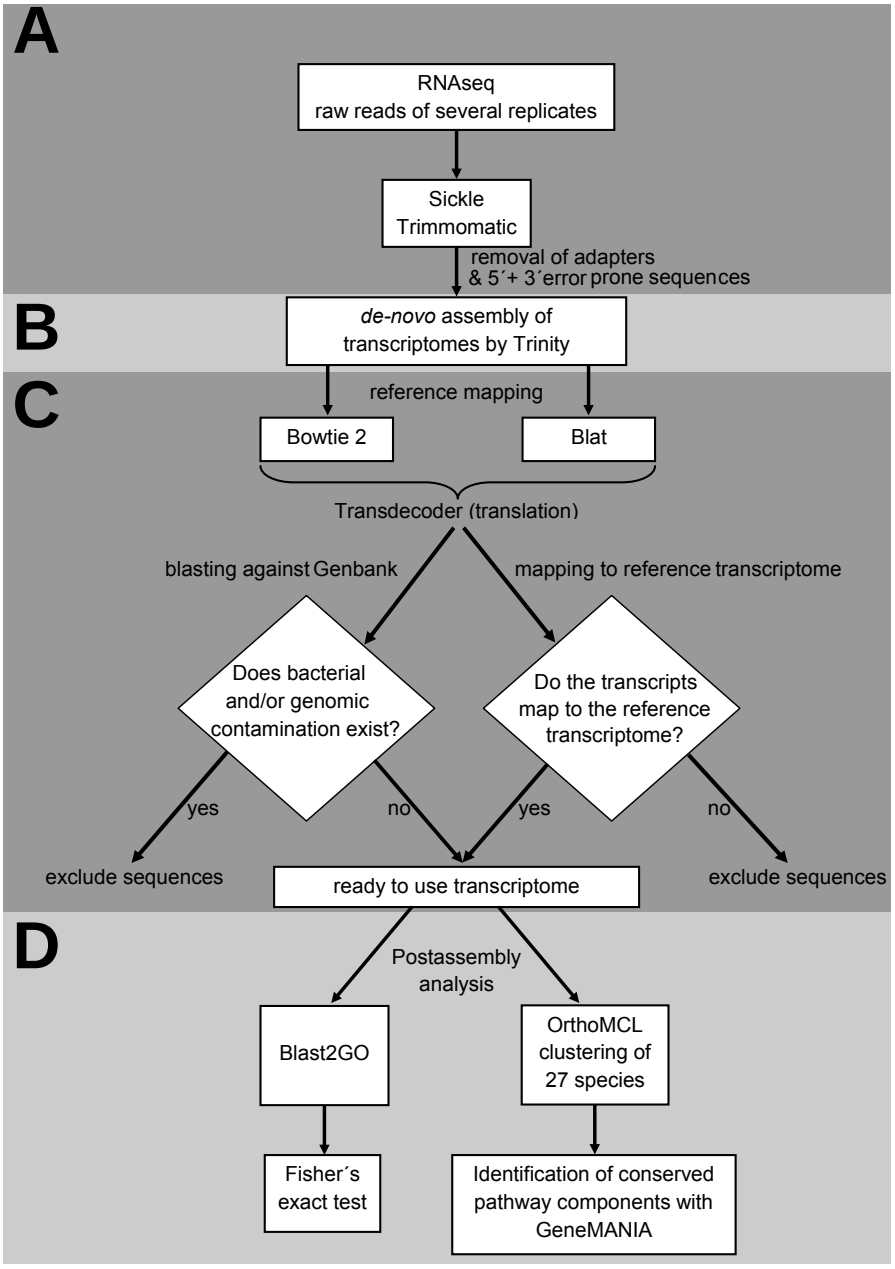

Figure 3.11: Flow diagram depicting the comprehensive workflow of post-sequencing steps to generate transcriptomes of satisfying quality. The procedure includes the following steps: (**A**) Sequencing adaptor removal and elimination of error-prone nucleotide sequences at the 5'- and 3'-ends of each read. (**B**) *De-novo* assembly of the first transcriptome by the programme Trinity. (**C**) Assessment of bacterial and genomic contaminations by reference mapping and blasting against Genbank. (**D**) Post-assembly analysis of interesting transcripts.

RNAseq in general and specifically for amplified mRNAs (see section 2.4 and Introduction), needs a proper quality assessment before analysing the data. Especially, the circumstances under which I had to extract mRNA by acid phenol/chloroform phase separation and amplify by *in-vitro* transcription (IVT) (see Materials and Methods), led to the problem that there was a certain amount of bacterial and genomic contamination in my data. Henceforth, I sought to identify the important transcripts for my analysis. For this purpose, I established a workflow to guarantee proper quality of my RNAseq data (Fig. 3.11).

My workflow consists of four successive steps which can be re-iterated in case of missing replicates to improve already analysed sequencing information (Fig. 3.11):

(1) First of all, I eliminated the incorporated Illumina adaptors which are part of the PCR step on the Illumina sequencing platforms (Materials and Methods). Next, I eliminated RNAseq sequencing reads which were prone to false positve and false negative nucleotides at the 3'- and 5'-ends of each read by using the software tool "Sickle" (https://github.com/najoshi/sickle; Fig. 3.11, A).

(2) I used the *de-novo* assembler Trinity to assemble the different transcriptomes from the RNAseq raw data (Grabherr et al. 2011; Fig. 3.11 B). I purposely used this assembler instead of a reference based assembler, as this allowed me to identify transcripts which are not part of draft reference genomes. Especially, NGS-based genomes are prone to gaps in sequences due to the loss of DNA regions which are difficult to sequence (e.g. due to inaccessible conformation of DNA as in the case of heterochromatin) and/or show insufficient coverage. Besides from this, assemblers used for the assembly of whole genomes are also prone to errors, especially in the case of repetitive sequences (Grabherr et al. 2011). Henceforth, instead of risking loss of scarce and/or not assembled transcripts due to an error-prone reference genome, I used the *de-novo* assembler Trinity instead and, if necessary, excluded sequences which were obviously not part of the reference genome (or reference transcriptome) by post-assembly analysis (see below).

(3) In the third step, I used different approaches in parallel to filter out bacterial and genomic contamination from my transcriptomes. For this, I used the short-sequence mappers "Bowtie2" (Langmead and Salzberg 2012) and "Blat" (Kent 2002) to map the transcripts to appropriate reference transcriptomes. Furthermore, I used "Transdecoder" (Haas et al. 2013) to predict the amino-acid sequences for the identified transcripts (Fig. 3.11, C). By blasting the assembled transcripts against Genbank (Benson

et al. 2015) and by using only mapped sequences, I could successfully eliminate bacterial transcripts from my transcriptome data. I mapped the transcripts against reference transcriptomes for all developmental stages of the studied species (Fig. 1.1). The successfully filtered transcriptomes were then used for further analysis (Fig. 3.11 C).
(4) For the comparative analysis of the complete transcriptomes, I used different means in parallel to gather information about the found transcripts: I used Blast2GO to find homologues of functionally described model organisms by sequence comparison with Genbank (Benson et al. 2015). I also used Fisher's exact test to identify over- or underrepresented gene ontology (GO) terms for transcripts of each transcriptome in comparison to the reference transcriptomes. I used our established OrthoMCL clustering pipeline for 27 species, including different invertebrate and vertebrate genomes and 13 nematode genomes (Material and Methods; Li et al. 2003), to identify orthologues among the mentioned animal groups (Fig. 3.11 D). After successfully identifying existing and expressed orthologues, I used the data base "GeneMANIA" (see Materials and Methods) to identify interacting proteins which belong to the same pathway or complex. This way I could increase my searches and could cover whole pathways which may be relevant to my scientific objective.
In total I sequenced early transcriptomes of 5 species (introduction; Fig. 1.1) and by using my work-flow, I successfully retrieved data from 13 RNA-libraries which are the result of independent sampling of more than 100 eggs per replicate (Materials and Methods; Tab. 3.4).
For one single nematode (*D. coronatus*) I tested to what extent the use of several replicates increases the quality assessed transcriptome. For this purpose I re-assembled different combinations of the 5 libraries (Fig. 3.12). This sequential re-assembling of transcriptomes from different libraries demonstrates that even though there was initial bacterial and genomic contamination in my RNA-sequencing raw data, I was able to reach a plateau of transcripts (Fig. 3.12). In the case of *D. coronatus* I could retrieve nearly 89% of the number of transcripts in comparison to *C. elegans* (Tab. 3.5). In the case of *R. culicivorax* I was even able to reach 99% of the number of expressed transcripts in comparison to *C. elegans* (Tab. 3.4). These results suggest that I reached a saturation for sequenced transcripts under the condition of amplificiation of mRNA via IVT prior to RNAseq (Materials and Methods). Therefore, the presented transcriptomes appear to be complete to a relatively high degree in the cases of *R. culicivorax* and *D. coronatus*. *Gordius sp.*, in contrast to the other species, could be sequenced without

amplification of mRNAs, due to the huge amount of isolated embryos (see Materials and Methods), thus its transcriptome is considered to be complete. In case of *Panagrolaimus sp. PS1159* the single RNAseq library which I could retrieve, yielded considerably more transcripts than are known to be expressed in *C. elegans*. For this reason, I cannot exclude that the transcriptome of *Panagrolaimus sp. PS1159* is still fragmented due to insufficient sequenced replicates and this high number of retrieved transcripts is due to incomplete assembly of the sequencing reads. This may have led to an overestimation of transcripts. In case of *P. pacificus* the transcriptome suffered from too little sequencing information as it shows the lowest number of transcripts probably due to the fact that only a single RNAseq library was generated (Tab. 3.4). In order to improve the quality of the *P. pacificus* transcriptome, I needed to identify which transcripts of the published reference transcriptome (see Materials and Methods) were actually expressed. For this purpose I mapped my *P. pacificus* transcripts to the reference transcriptome and retrieved the complete sequences, yielding a much better median transcript length (Materials and Methods; Tab. 3.5).

My RNAseq approach may not have detected all rare transcripts for the transcriptomes of *R. culicivorax*, *Panagrolaimus sp. PS1159* and *D. coronatus*, hence inference of absent transcripts in my transcriptomes is impossible. The transcriptomes of *C. elegans* and *A. suum* can be considered as complete (Hashimshony et al. 2012, Wang et al. 2012) and will serve as a reliable reference to compare all other transcriptomes of early embryonic stages with. Hence identification of present and absent orthologues during early development of *C. elegans* and *A. suum* is feasible.

Table 3.4: Number of the number of retrieved transcripts and mean and median length per sequences species after applying my workflow (Fig. 3.11) to the raw data. *Gordius sp.* was sequenced without RNA amplification (marked by single asterisk). Additionally, I included transcriptomes from the published transcriptomes for *A. suum* (Wang et al. 2014) and *C. elegans* (Hashimshony et al. 2012). The number of independent (biological) replicates used for the published transcriptomes was previously published (see text) and is not part of this analysis (marked by two asterisks).

species	number of biological replicates	number of transcripts	median transcript length [bp]
*Gordius sp.**	2	12,070	872.5
R. culicivorax	4	7,236	522
P. sp. PS1159	1	15,568	420
D. coronatus	5	6,546	381
P. pacificus	1	1,608	1,152
A. suum	**	3,093	1,062
C. elegans	**	7,304	1,431

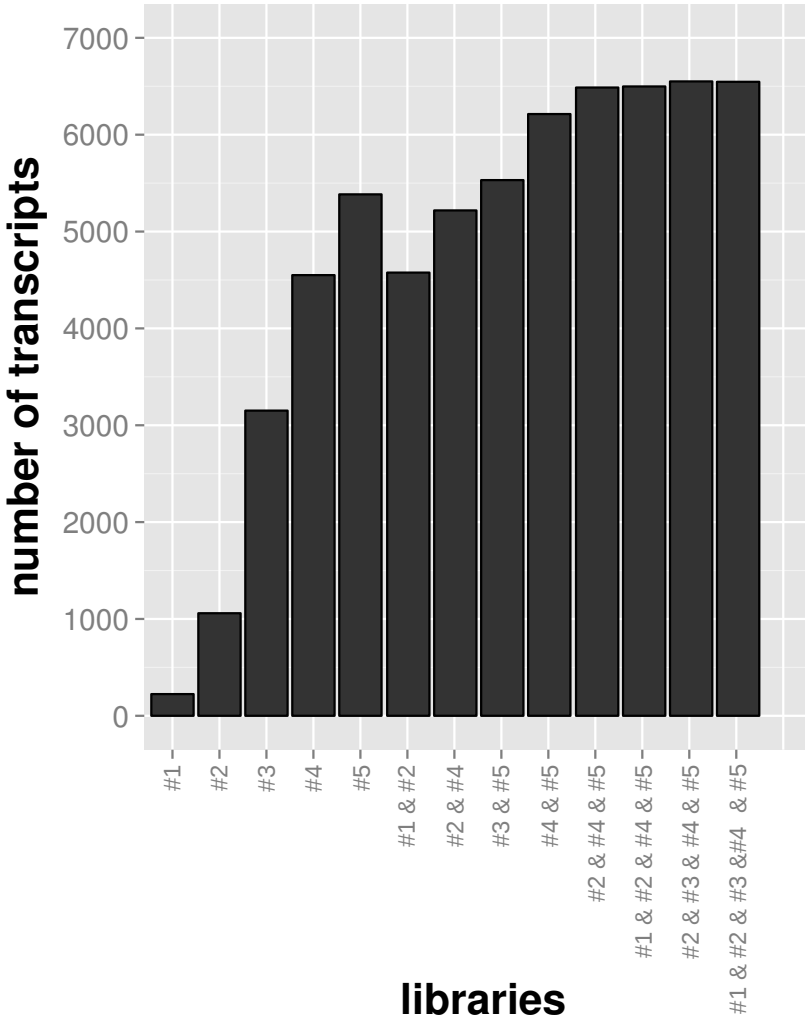

Figure 3.12: Number of retrieved transcripts by random re-assembly of the different libraries for early development in the nematode *D. coronatus*. By using the original RNAseq libraries (#1 to #5) as permutated input for the transcriptome assembly, the number of retrieved transcripts appears to saturate at arround 6,500 transcripts.

3.7 Sharing pathways with other Metazoa: Basal nematodes express signaling pathway components of invertebrates and vertebrates during early development

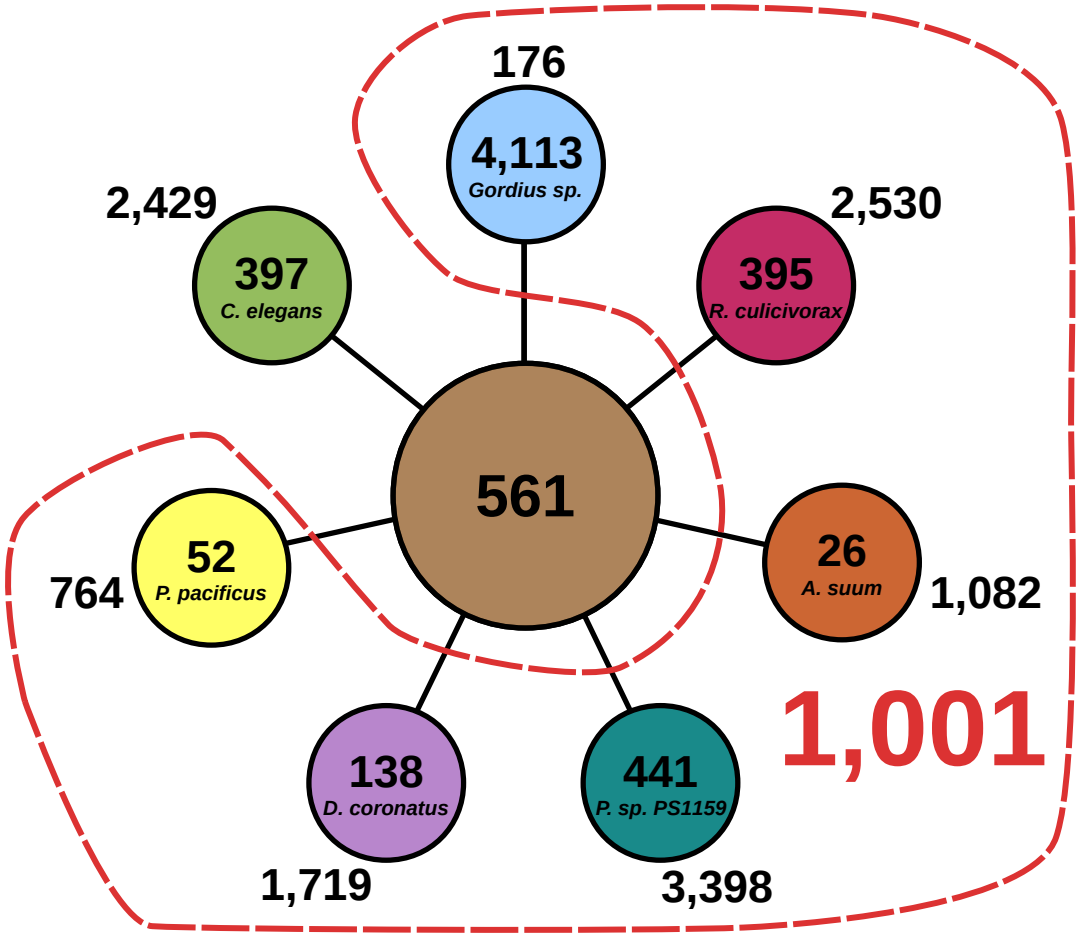

Figure 3.13: Expressed clusters and genes during early development for seven species. Shared clusters among all species (central circle). Numbers of early expressed genes with no orthologues to other species during early embryogenesis are depicted next to the circles. The number of orthologous clusters which are cumulatively shared among species after exclusion of *C. elegans* is represented in red (dashed line).

I clustered early expressed transcripts in all available species via OrthoMCL (Fig. A.1 and 3.13). In total I found 561 clusters which were shared among the transcriptomes of the Chromadorean species *C. elegans*, *D. coronatus*, *P. pacificus*, *P. sp. PS1159* and *A. suum*, as well as the Enoplean species *R. culicivorax* and the nematomorph outgroup species *Gordius sp.*. I found 674 clusters shared by non-*C. elegans* species. There were 173 clusters which are specific to all Chromadorean species (i.e. excluding *R. culcivorax* and *Gordius sp*). 1,944 clusters were shared among all nematodes (i.e. excluding *Gordius*

sp.; Fig. A.1), suggesting that most protein families expressed during early development are shared among all investigated nematodes. By subtracting all *C. elegans*-specific clusters from the shared clusters of the remaining species I retrieved 1,001 cumulatively shared ones among the remaining species (Fig. 3.13 dashed line). This indicates that the majority of shared orthologues is not expressed in *C. elegans* during early embryogenesis. In order to what may be the potential function of orthologues which are not expressed in the model *C. elegans* I focused on the Enoplean *R. culicivorax* and the nematomorph *Gordius sp.*

To investigate to what extent specific early expressed transcripts in *R. culicivorax* and *Gordius sp.* are shared by animal models from other taxa, I used these data to find orthologues in models such as *Drosophila*, *Xenopus*, zebrafish, mouse, rat and humans. By using Fisher's exact test for enriched gene ontology (GO) terms in the Enoplean *R. culicivorax* and the nematomorph *Gordius sp.*, I found enrichment of 52 GO-terms for *R. culcivorax* and enrichment of 98 GO-terms in case of *Gordius sp.* (Tab. C.2 and C.3). The most striking GO-terms involved genes which are associated with NF-κb, the apoptosis signaling pathway and double strand break repair systems (Tab. C.2 and C.3).

I extended the search to find additional pathway components involved in early development of aforementioned models in *Gordius sp.* and *R. culicivorax*. Screening the genomes, and early transcriptomes for 15 pathways or biological processes known to be involved in development, I found pathway components which are only in present in Enoplean nematodes or nematomorphs (Tab. 3.5).

Most strikingly, I found orthologues specifically expressed during early development of *R. culicivorax* and *Gordius sp.*, such as Hunchback, NFKB1, Noggin, Chordin/Short gastrulation, RSF1 (remodelling and spacing factor 1) and YY1 (Ying and Yang1; Pleiohomeotic). This means that several genes in the pathway of BMP, Wnt, NF-κB and SWI/SNF signaling which are known from model organisms such as *Drosophila*, mouse and humans, are expressed during the early development of *Gordius sp.* and *R. culicivorax*, but not in the other analysed species. These pathways were further investigated (Tab. 3.5; section 3.8 to 3.17).

In total I screened 10,454 early expressed *Gordius sp.* and 7,273 early expressed *R. culicivorax* transcripts and found orthologues in 16 different signaling pathways and/or biological processes (Tab. 3.5). Based on these results, I screened which orthologues of these parthways are expressed during early development in *Gordius sp.*, *R. culicivorax*, *A. suum*, *P. sp. PS1159*, *D. coronatus*, *P. pacificus*, and *C. elegans* (section 3.8 to

3.17). Furthermore, I also looked for essential pathways required for early development in *C. elegans*, such as PAR-3/-6/PKC-3, SKN-1/MED, MOM-5/Wnt and GLP-1/Notch signaling. Here, I found the unexpected absence of orthologues in all nematodes except *Caenorhabditis* (section 3.9, 3.11, 3.15, 3.16).

Table 3.5: 10,454 *Gordius sp.* (*G. sp.*) and 7,273 *R. culicivorax* (*R. c.*) early expressed transcripts were used as queries to search for homologues in *Drosophila*, *Xenopus*, zebrafish, mouse, rat and humans by using Blastp against Genbank (Materials and Methods). By applying Blast2GO onto this pre-screening approach, *Gordius sp.* and *R. culcivorax* transcripts were selected for homologues to known regulators of signaling pathways or biological processes of aforementioned model organisms.

signaling pathway or biological process	**pathway components found in *G. sp.* and/or *R. c.***
BMP/TGF-β signaling	50
Wnt signaling	44
notch signaling	29
SWI/SNF RSF1 complex	35
Toll/NF-κB signaling	38
LGL/DGL/Scrib and PCP signaling	34
mTORC signaling	25
C.elegans polarity establishment	33
early developmental genes	72
insulin/FOXO signaling	6
JNK signaling	8
Jak-Stat/Interleukin signaling	7
hedgehog signaling	7
gonad development	3
asymmetric cell division	14
stemcell factors	8

3.8 Early BMP signaling exists in the clade I nematode *R. culicivorax* and the nematomorph *Gordius sp.*

One of the aims of my thesis is the detection of orthologues involved in axis formation and polarity establishment in Enoplean and Chromadorean nematodes (Introduction), hence I investigated the BMP/DPP (Bone morphogenetic protein; Decapentaplegic) and the Wnt signaling pathways. In many invertebrates and vertebrates anterioposterior (a-p) and dorsoventral (d-v) axis patterning is strongly depends on the interplay of Wnt and BMP signaling pathways. While Wnt mainly drives a-p axis formation, BMP signaling

	genomes																
	N	I		II	III				IV			V					
	G. s. (*)	*R. c.* (*)	*T. s.*	*E. b.*	*A. s.* (*)	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* (*)	*D. c.* (*)	*P. p.* (*)	*C. a.*	*C. b.*	*C. r.*	*C. e.* (*)
CHRD Sog	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BMP1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BMP5/6/7/8A Scw/ Gbb	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BMP2/4 Dpp	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
TWSG1 Tsg	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
ZDH17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
TLL1 Tdl	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NOG	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Cv	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
Cv-2	+	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-
Mad	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
LanA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Figure 3.14: Present ("+") and absent ("−") orthologues of the BMP signaling pathway components among 16 nematode species and the nematomorph *Gordius sp.* Species for which not only genomic but also transcriptomic data for early expressed transcripts were available are marked by a single asterisk. I - V, clades (Fig 1.1; Blaxter et al. 1998). Abbreviations: *A. s.* - *Ascaris suum*; *B. m.* - *Brugia malayi*; *B. x.* - *Bursaphelencuhs xylophilus*; *C. a.* - *Caenorhabditis angaria*; *C. b.* - *Caenorhabditis briggsae*; *C. e.* - *Caenorhabditis elegans*; *C. r.* - *Caenorhabditis remanei*; *D. c.* - *Diploscapter coronatus*; *D. i.* - *Dirofilaria immitis*; *E. b.* - *Enoplus brevis*; *G. s.* - *Gordius sp.*; *L. l.* - *Loa loa*; *M. h.* - *Meloidogyne hapla*; N - nematomorph; *P. p.* - *Pristionchus pacificus*; *P. s.* - *Panagrolaimus sp. PS1159*; *R. c.* - *Romanomermis culicivorax*; *T. s.* - *Trichinella spiralis.*

is involved in d-v axis formation (Hikasa and Sokol 2013). Both signaling pathways act via secreted ligands (e.g. BMP2/4/DPP) and inhibitors (Chordin/Short gasturlation, or Noggin; Yamamoto and Oelgeschläger 2004, Hikasa and Sokol 2013) which form gradients along the prospective axes within the embryo (Introduction; Hikasa and Sokol 2013).

Using our OrthoMCL clustering I screened for genes involved in BMP and Wnt signaling (see 3.9 and 3.10) in the studies nematodes. For BMP/DPP signaling, I found not only orthologues in the genomes of Enplean and Chromadorean nematodes, but also early expression of the BMP/DPP inhibitors Sog (Short gastrulation) or Chordin, Tsg (Twisted gastrulation), Cv (Crossveinless) and Cv-2 (Crossveinless−2) in *R. culicivorax* and *Gordius sp.* (Fig. 3.14 and 3.15). I even found orthologues for the vertebrate specific BMP/DPP inhibitor Noggin, which was expressed in both species. In addition to BMP/DPP inhibitors I found also orthologues for the Tld (Tolloid) and BMP1 metallo proteases, orthologues for the receptor BMPR2, orthologues for the BMP/DPP ligand

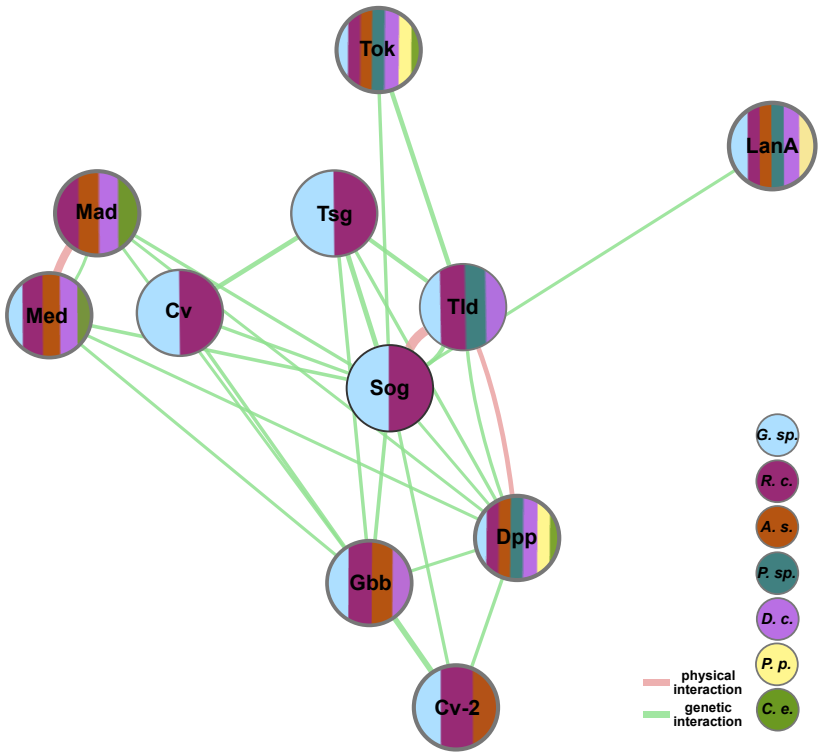

Figure 3.15: Schematic representation of the *Drosophila* signaling network for BMP/DPP pathway components. Expressed orthologues of the pathway components during early transcriptomes of seven species are shown (inferred from transcriptomic data and orthologous clustering; compare Fig. 3.14). Expressed orthologues are visualised by a species-specific colour code. *A. s.* - *Ascaris suum*; *C. e.* - *Caenorhabditis elegans*; *D. c.* - *Diploscapter coronatus*; *G. sp.* - *Gordius sp.*; *P. p.* - *Pristionchus pacificus*; *P. sp.* - *Panagrolaimus sp. PS1159*; *R. c.* - *Romanomermis culicivorax*.

and SMAD transcription factors Med (Medea) and Mad (Mothers against Decapentaplegic) to be expressed in nearly all examined early transcriptomes (Fig. 3.15). This result suggests that all necessary BMP signaling components exist in nematomorphs, Enoplean and clade III nematodes to allow proper BMP/DPP signaling (Fig. 3.15). Especially, BMP/DPP inhibitors, such as Sog/Chordin, Noggin and Tsg are a prerequisite for proper BMP signaling in invertebrate and vertebrate species (Smith and Harland 1992, Sasai et al. 1994, Chang et al. 2001). The presented results suggest that (1) BMP/DPP is fully functional in *Gordius sp.* and *R. culicivorax* and (2) may therefore play a role in axis formation in both species. Taking into account that BMP signaling (and dorsoventral patterning) *C. elegans* is only found during post-embryonic development (e.g. postembryonic mesoderm patterning and dauer larva formation; Estevez et al. 1993, Foehr and Liu 2008, Tian et al. 2010, Gumienny and Savage-Dunn 2013), opens the possibility that in Enoplean nematodes this signaling pathway is an altenative way of polarity establishment and axis formation like in other invertebrates and vertebtates.

In order to test whether BMP/DPP signaling against Chordin may have a role in axis formation during early embryogenesis of *R. culicivorax*, we performed *in-situ* hybridisations against the Sog/Chordin orthologue (Fig. 3.16). We found ubiquitous expression during the 1-4 cell stages of *R. culcivorax*. Afterwards the expression of *chordin* gradually decreased (Fig. 3.16 C, D, E, F). As *chordin* was ubiquitously expressed in *R. culcivorax* the question arises whether BMP/DPP signaling can play the assumed role as antagonist of BMP/DPP signaling. These results strongly suggests that components of BMP/DPP signaling are expressed during early development and therefore may play a role in controlling axis formation. In addition, I could show that important BMP/DPP signaling components are conserved among Enoplean nematodes, nematomorphs, other invertebrates and vertebrates.

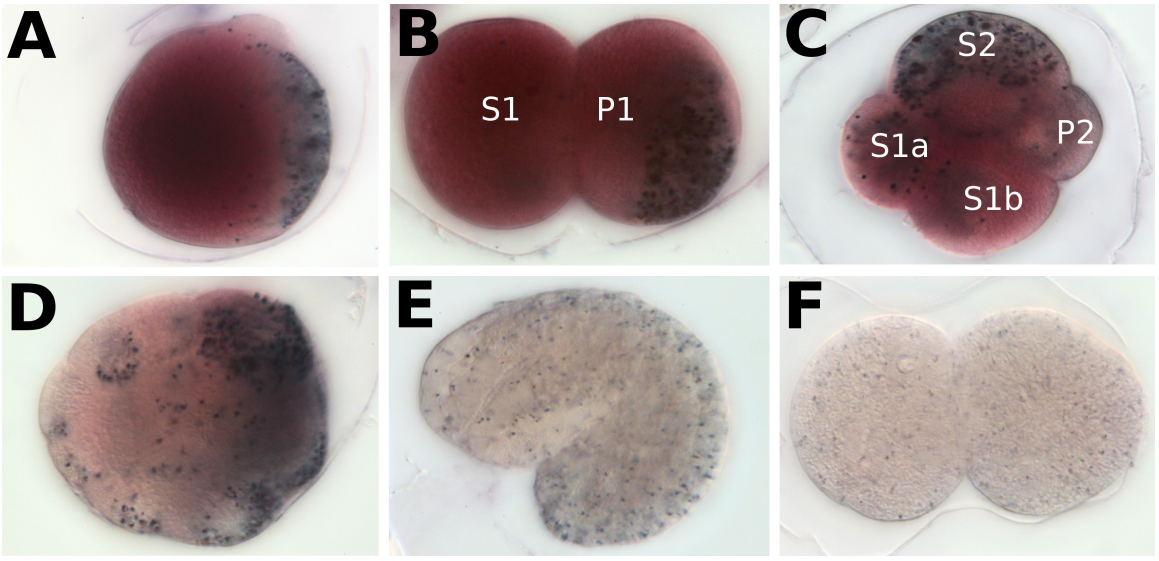

Figure 3.16: *In-situ* hybridizations show *chordin* expression (purple) in *R. culicivorax* during (**A**) 1-cell, (**B**) 2-cell, (**C**) 4-cell, (**D**) 16-cell, (**E**) morphogenesis and (**F**) sense RNA negative control. Differential seggregation of natural dark pigment was previously described Schulze and Schierenberg 2009.

3.9 Wnt signaling via β-catenin paralogues WRM-1 and SYS-1 is specific for the genus *Caenorhabditits*

Wnt signaling is another major pathway involved in axis formation in vertebrates and many invertebrates (Hikasa and Sokol 2013). Here, I investigated the potential role of Wnt signaling for early development in nematodes by searching for early expressed

	genomes																
	N	I			II	III				IV			V				
	G. s. (*)	*R. c.* (*)	*T. s.*	*E. b.*	*A. s.* (*)	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* (*)	*D. c.* (*)	*P. p.* (*)	*C. a.*	*C. b.*	*C. r.*	*C. e.* (*)
MOM-2	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
MOM-5	+	+	-	+	+	+	+	+	+	+	+	+	+**	+	+	+	+
POP-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SYS-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
WRM-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
HMP-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BAR-1	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
LIT-1	+	+	+	+	+	+	+	+	+	+	+	+	+**	+	+	+	+
SDZ-26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
MOM-4	+	+	-	+	+	+	+	+	-	+	+	+	+	+	+	+	+
ZFP-1	+	+	-	+	+	+	+	+	-	-	+	+	+	+	+	+	+
TSN-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PLP-1	+	+	+	+	+	+	+	+	+	+	+	+	+**	+	+	+	+
MUT-7	+	-	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+
CAM-1	+	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+
LIN-37	+	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+
EFN-4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
JUN-1	+	+**	-	+**	+	+	+	+	+	+	+	+	+	-	+	+	+
GOA-1	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+
EGL-8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MES-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+

Figure 3.17: Present ("+") and absent ("−") orthologues of the Wnt signaling pathway components among 16 nematode species and the nematomorph *Gordius sp.* For further details see caption of figure 3.14.

orthologues of *C. elegans* genes participating in Wnt signaling (Fig. 3.17, 3.18) and for respective vertebrate orthologues (see section 3.10). I looked whether the Wnt ligand MOM-2, the frizzled receptor MOM-5, the transcriptional TCF/LEF regulator POP-1 and other interacting components of the Wnt pathway also exist in genomes of the studied nematodes (Fig. 3.17) and if yes, whether they are expressed during early development (Fig. 3.18). I found that highly conserved components of Wnt signaling, such as MOM-2, MOM-5 and POP-1 have orthologues among all investigated nematodes (Fig. 3.17). In concordance with this I found Wnt components acting in early embryogenesis in nearly all studied species (Fig. 3.18).

I also found components of the Wnt signaling network in *C. elegans* for which no orthologues exist outside the genus *Caenorhabditis*, such as EFN-4, MES-1 and the β-catenins SYS-1 and WRM-1 (Fig. 3.17). Interestingly, I also found orthologues of certain Wnt signaling components only expressed in *C. elegans* (Fig. 3.18). In *C. elegans* Wnt signaling differs from other invertebrates and vertebrates in that there exist several β-catenin

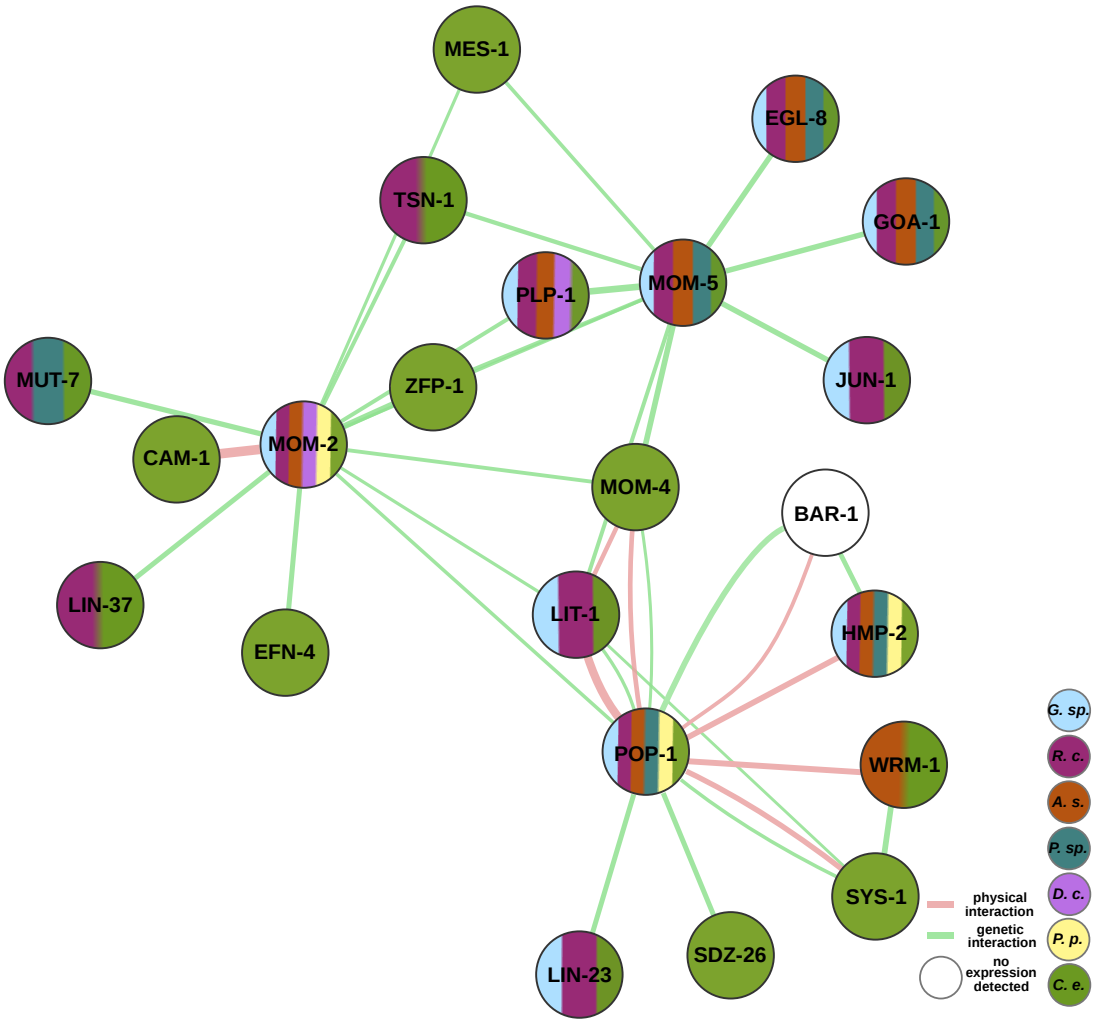

Figure 3.18: Schematic representation of the *C. elegans* signaling network for Wnt pathway components. Expressed orthologues of the pathway components during early embryogenesis of seven species are shown (inferred from transcriptomic data and orthologous clustering; compare Fig. 3.17). For further details see caption of figure 3.14.

transcription factor paralogues with distinct functions during early development (Eisenmann 2005, Mizumoto and Sawa 2007). I investigated this particularity and found that the crucial β-catenins WRM-1 and SYS-1 of the so-called "Wnt asymmetry pathway" (Eisenmann 2005) only exists in the genus *Caenorhabditis* (Fig. 3.18, 3.17). Concerning the four *C. elegans* β-catenin paralogues I found for Enoplean nematodes and the nematomorph *Gordius sp.* only one orthologue, while most Chromadorean nematode genomes possess 3 β-catenins (HMP-2, BAR-1, WRM-1). Only the genus *Caenorhabditis* has the additional SYS-1 (Fig. 3.19).

For orthologues of the β-catenin BAR-1 seems not to be expressed in all investigated nematodes during early development (Fig. 3.18). As Wnt signaling is mediated in most animals via a single β-catenin, I looked for the expression of the β-catenin orthologue

HMP-2 which I found to be the closest orthologue to β-catenin/armadillo of the models *Drosophila*, mouse and humans for Chromadorean nematodes (Fig. 3.19). I found that HMP-2 was expressed during early development in all Chromadorean species (Fig. 3.18). Here, I showed that most Wnt signaling components known from *C. elegans* have orthologues in most nematodes of clade I to V (Fig. 3.17). Nevertheless, the early expression of these Wnt signaling components seems to be in many cases exclusive in *C. elegans* (Fig. 3.18). Adding to that, the number of essential β-catenins also differs among the phylum of nematode (Fig. 3.19), showing that Wnt signaling may function differentially between Enoplea and Chromadorea. Thus, I sought to investigate whether other Enoplean-specific Wnt signaling components exist (see section 3.10).

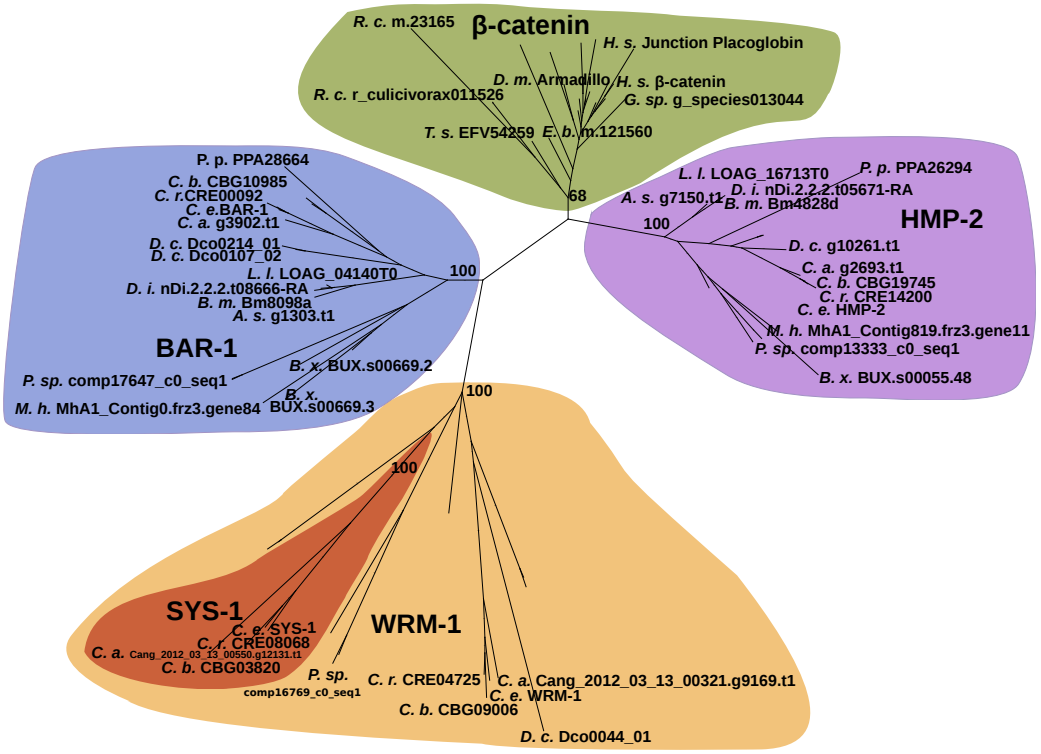

Figure 3.19: Unrooted phylogenetic tree for the β-catenin orthologues HMP-2, BAR-1, WRM-1, SYS-1 and β-catenin/Armadillo. The phylogenetic tree is based on clustalω alignment (Material and Methods; Sievers et al. 2011) for 84 homologous armadillo repeat domain protein sequences from 37 species. The tree was constructed based on Maximum Likelihood statistics by the programme RAxML (Material and Methods; Stamatakis 2006) for 100 bootstraps. β-catenin clusters are colour coded. Each cluster, besides from one for the outgroup β-catenin/Armadillo proteins, was supported by a bootstrap value of 100%. Intriguingly, nematomorph and Enoplean β-catenin orthologues exclusively cluster with outgroup β-catenins of *Drosophila* and humans, while Chromadorean nematodes partially possess all four β-catenins known from *C. elegans*. Only species of the *Caenorhabditis* genus have orthologues of the SYS-1 *beta*-catenin. Abbreviations: *A. s.* - *Ascaris suum*; *B. m.* - *Brugia malayi*; *B. x.* - *Bursaphelenchus xylophilus*; *C. a.* - *Caenorhabditis angaria*; *C. b.* - *Caenorhabditis briggsae*; *C. e.* - *Caenorhabditis elegans*; *C. r.* - *Caenorhabditis remanei*; *D. c.* - *Diploscapter coronatus*; *D. i.* - *Dirofilaria immitis*; *D. m.* - *Drosophila melanogaster*; *E. b.* - *Enoplus brevis*; *G. s.* - *Gordius sp.*; *H. s.* - *Homo sapiens*; *L. l.* - *Loa loa*; *M. h.* - *Meloidogyne hapla*; N - nematomorph; *P. p.* - *Pristionchus pacificus*; *P. s.* - *Panagrolaimus sp. PS1159*; *R. c.* - *Romanomermis culicivorax*; *T. s.* - *Trichinella spiralis*.

3.10 Wnt pathway components exist in all studied nematodes and are expressed during early embryogenesis

	genomes																
	N	I		II	III				IV			V					
	G. s. (*)	*R. c.* (*)	*T. s.*	*E. b.*	*A. s.* (*)	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* (*)	*D. c.* (*)	*P. p.* (*)	*C. a.*	*C. b.*	*C. r.*	*C. e.* (*)
RYK	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
TRABD 2B	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
WNT1	+	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**
WNT3	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**
WNT3A	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**
WNT8A	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**	+**
FZD1	+	+	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+
FZD8	+	+**	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
FAT3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FAT4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NPEPPS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DCHS1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CNNM2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EPHB3	+	+	+	+	+**	+	+	+	+	+	+	+	+**	+	+	+	+
LRP6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PORCN	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CELSR2	+**	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
AGK	+**	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+

Figure 3.20: Present ("+") and absent ("−") orthologues of the human Wnt signaling pathway components with emphasis onto the Wnt receptor RYK among 16 nematode species and the nematomorph *Gordius sp.* For further details see caption of figure 3.14.

In order to identify Enoplean-specific upstream components of Wnt signaling, I screened more than 44 genes known to be involved in this pathway (Tab. 3.5) for orthologues existing specifically in Enoplean and nematomorph genomes and to be expressed in *Gordius sp.* and/or *R. culicivorax* (Fig. 3.20, 3.21). I could not identify any Wnt signaling component, besides from β-catenin/Armadillo (Fig. 3.19) to have specific orthologues in Enoplean nematodes and the nematomorph *Gordius sp.* (Fig. 3.20). By investigating the early expression of existing Wnt signaling orthologs, I detected Wnt signaling components to be expressed in all investigated species (Fig. 3.21).

I found that most Wnt signaling components known from humans have orthologues in all (or nearly all) nematodes (Fig. 3.20). In concordance with this I find early expression of most orthologues in nearly all investigated species (Fig. 3.21). Interestingly, for the Wnt co-receptor RYK and the Wnt inhibitor TRABD2B I did not detect any expression in other nematodes except *Gordius sp.* and *R. culicivorax* and *D. coronatus* (Fig.. 3.21). This result shows that most upstream components of Wnt signaling (such as ligands

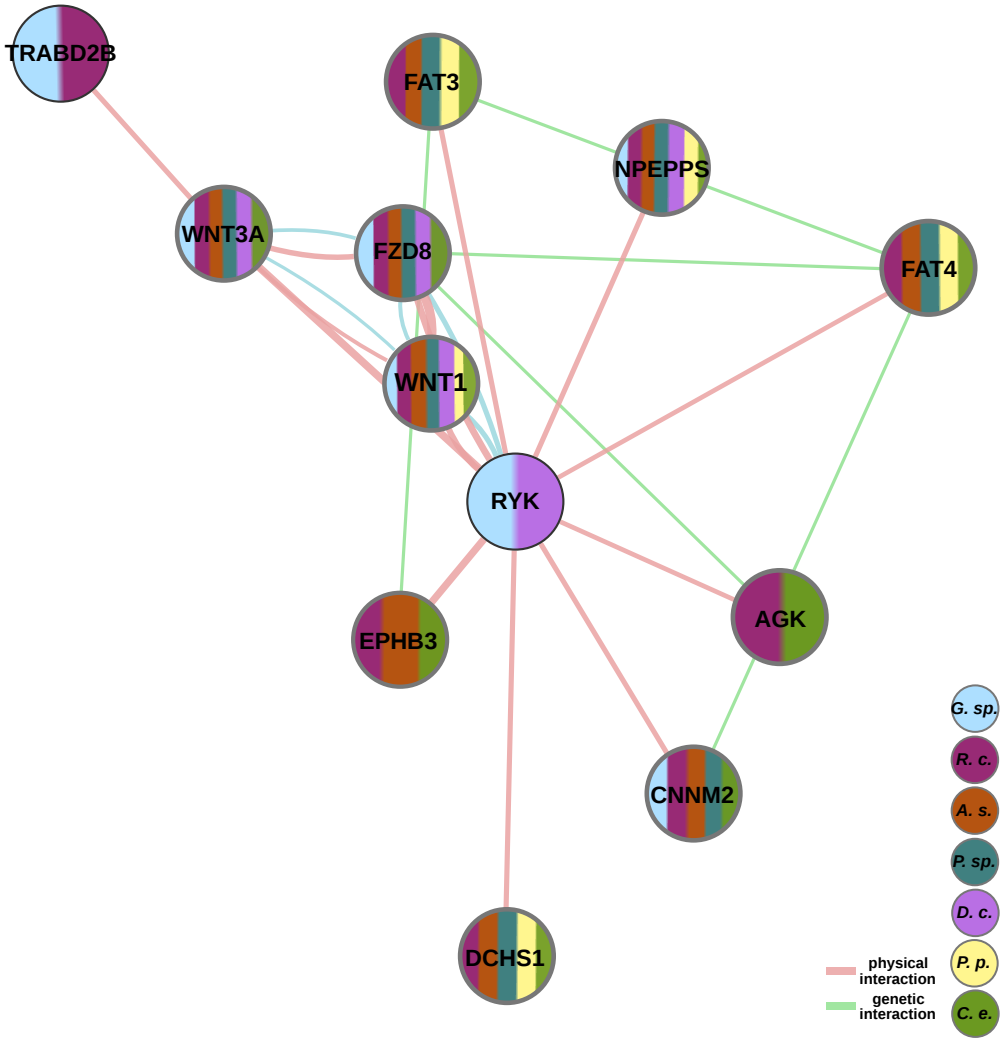

Figure 3.21: Schematic representation of the human signaling network for Wnt pathway components. Expressed orthologues of the pathway components during early embryogenesis of seven species are shown (inferred from transcriptomic data and orthologous clustering; compare Fig. 3.17). For further details see caption of figure 3.14.

and receptors) and downstream target genes are (1.) conserved among nematodes and (2.) are also expressed during early development. The only components found to be expressed only in a fraction of the studied nematodes seem to be the Wnt inhibitor TRABD2B and the Wnt co-receptor RYK.

3.11 Notch signaling acts during early development in all nematodes, but specific core components seem to vary between Enoplean and Chromadorean nematodes

	genomes																
	N	I		II	III				IV			V					
	G. s. (*)	*R. c.* (*)	*T. s.*	*E. b.*	*A. s.* (*)	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* (*)	*D. c.* (*)	*P. p.* (*)	*C. a.*	*C. b.*	*C. r.*	*C. e.* (*)
APX-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
GLP-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
LAG-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PES-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
POS-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
SUR-2	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+
EOR-2	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+	+	+
KIN-15	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+
KIN-19	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SUP-17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PRO-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GFL-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ZFP-1	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+
VPS-22	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+
EPN-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EGO-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SEL-12	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
TEN-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GSKA-3	-	-	-	-	+	+	-	-	-	-	-	-	-	+	+	+	+
REF-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
CIR-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+
LIN-11	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+

Figure 3.22: Present ("+") and absent ("−") orthologues of the *C. elegans* Notch signaling pathway for 16 nematode species and the nematomorph *Gordius sp.* For further details see caption of figure 3.14.

From *C. elegans* it is known that Notch signaling is important for early development, as it determines the fate fo the AB cells (Priess 2005). I found that Notch signaling is also active in Enoplea, like in Chromadorean nematodes, as orthologues of the CSL protein LAG-2 and the Notch receptor GLP-1 exist in their genomes (Fig. 3.22) plus they are expressed during early development (Fig. 3.23). I searched therefore for orthologues of the *Drosophila* and human DSL ligands and further identified orthologues in *R. culicivorax* and *Gordius. sp.* I found that many known *C. elegans* Notch signaling components and target genes exist exclusively in the genomes of *Caenorhabditis* species (Fig. 3.22), such as REF-1, GSKA-3, KIN-15, POS-1 and APX-1. Especially, the finding that the DSL ligand APX-1 is only present in the genus *Caenorhabditis* and plays an important

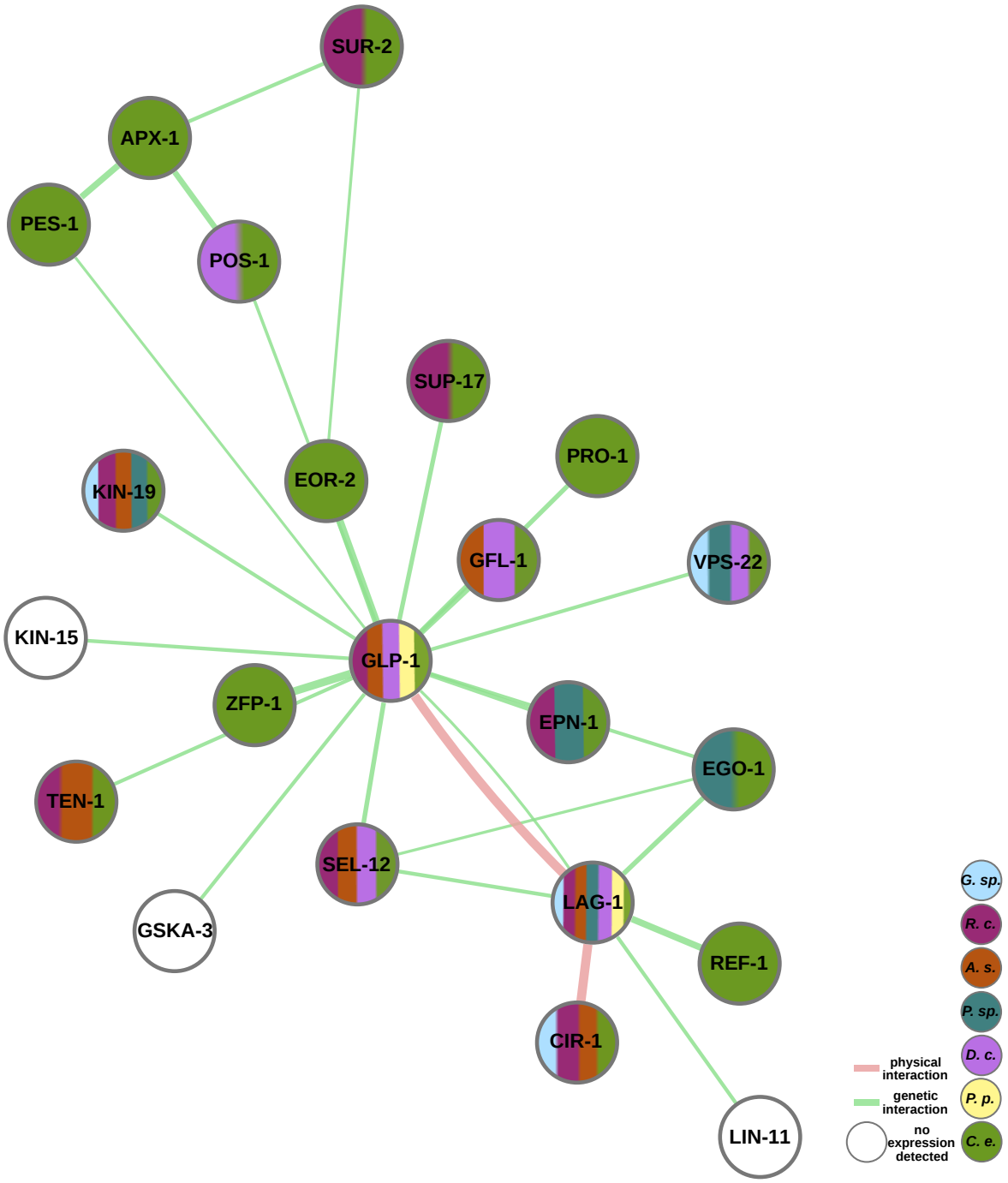

Figure 3.23: Schematic representation of the *C. elegans* signaling network for Notch pathway components. Expressed orthologues of the pathway components during early embryogenesis of seven species are shown (inferred from transcriptomic data and orthologous clustering; compare Fig. 3.17). For further details see caption of figure 3.14.

role during the AB blastomere fate determination in *C. elegans* (Mickey et al. 1996) indicates that Notch signaling during early embryogenesis must be different in non-*C. elegans* nematodes. Nonetheless, my findings also suggest that Notch signaling per se exists in all nematodes and is probably crucial for embryogenesis as conserved Notch

signaling components such as GLP-1 and LAG-1 were found to be expressed during early embryogenesis in all (or nearly all) investigated species.

The observed differences suggest that in more distant relatives of *C. elegans* Notch signaling may be involved in other processes than the induction of AB fate, e.g. axis determination.

3.12 NF-κB signaling may play a role during early development of *Gordius sp.* and *R. culicivorax*

	genomes																
	N	I		II	III				IV			V					
	G. s. (*)	*R. c.* (*)	*T. s.*	*E. b.*	*A. s.* (*)	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* (*)	*D. c.* (*)	*P. p.* (*)	*C. a.*	*C. b.*	*C. r.*	*C. e.* (*)
NFKB1	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BOK	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BCL2	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
RBX1	+**	+	+**	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BCL2L1	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
VDAC1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
AIFM1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MAP3K8	-	-	-	+	+	+	+	+	+	+	+**	-	-	-	-	-	-
KPNA3	+	+	+	+	+	+	+	+	-	+	+	+	+**	+	+	+	+

Figure 3.24: Present ("+") and absent ("−") orthologues of the NF-κB signaling pathway components among 16 nematode species and the nematomorph *Gordius sp.* For further details see caption of figure 3.14.

Toll signaling, a variant of NF-κB signaling, is crucial in *Drosophila* for d-v axis formation (Minakhina and Steward 2006). Here, I investigated whether known NF-κB components have orthologues in nematodes and whether such orthologues are also expressed during early embryogenesis.

I found orthologues and expression of the most important component of the NF-κB signaling pathway only in *R. culicivorax* and *Gordius sp*, namely the transcription factor NF-κB1 (Fig. 3.24, 3.25). These results suggest that in this respect the Enoplean nematode *R. culicivorax* and the nematomorph *Gordius sp.* show more similarity to *Drosophila* than to Chromadorean nematodes. Both species, *R. culicivorax* and *Gordius sp.*, may therefore use an alternative pathway for axis specification compared to *C. elegans* (see section 3.8, 3.9 and 3.10).

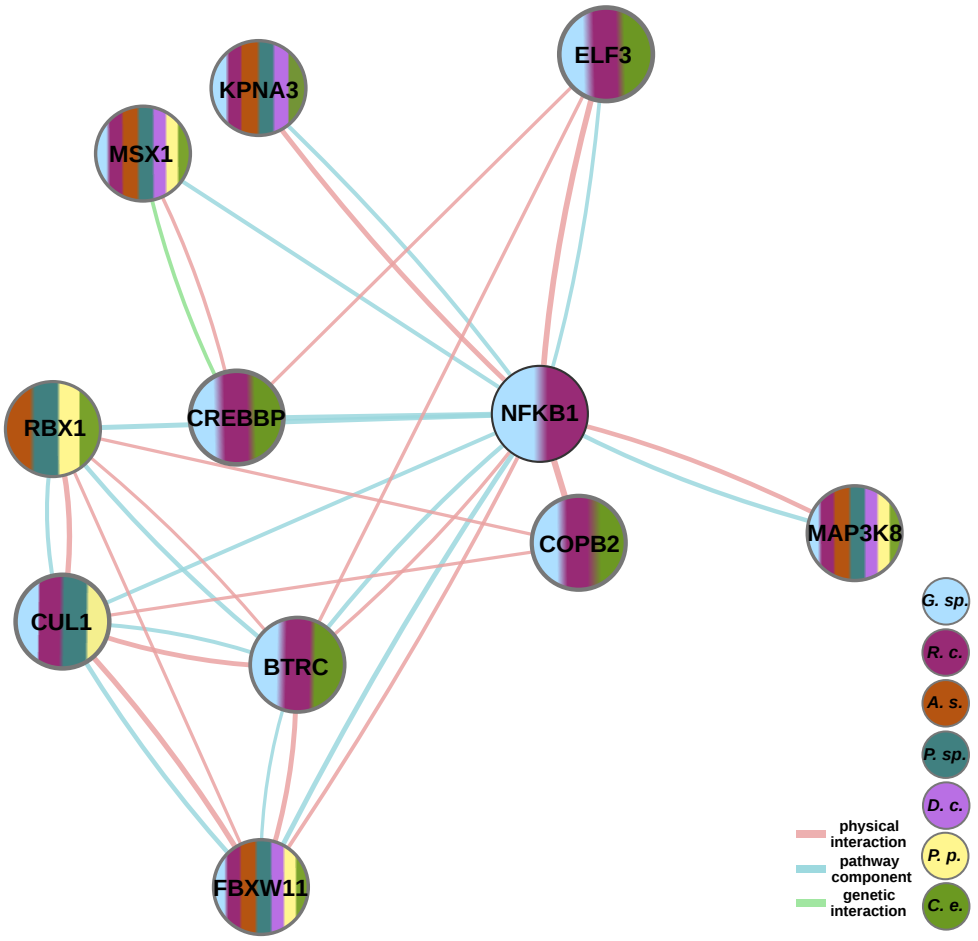

Figure 3.25: Schematic representation of the human signaling network for NF-κB pathway components. Expressed orthologues of the pathway components during early embryogenesis of seven species are shown (inferred from transcriptomic data and orthologous clustering; compare Fig. 3.17). For further details see caption of figure 3.14.

3.13 Crucial polarity-inducing factors are specific for the genus *Caenorhabditis*

	genomes																
	N	I		II	III				IV			V					
	G. s. (*)	*R. c.* (*)	*T. s.*	*E. b.*	*A. s.* (*)	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* (*)	*D. c.* (*)	*P. p.* (*)	*C. a.*	*C. b.*	*C. r.*	*C. e.* (*)
LET-99	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
PAR-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PAR-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
PAR-3	+	+**	-	+**	+	+	+	+	-	+	+	+	+	+	+	+	+
PAR-4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PAR-5	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	+	+
PAR-6	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+
GPR-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
GPR-2	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
LIN-5	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
LIS-1	+	+	+	+	+	+	+	+	+	+	+	+	+**	+	+	+	+
SPD-2	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
SPD-5	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
AIR-1	-	-	-	-	-	+	+	+	-	-	-	-	-	-	+	+	+
SIR-2.1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FEM-2	+	+	+	+	+	+	+	+	+	+	+	+	-**	-	+	+	+
COPB-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GPA-16	+	+	+	+	+	+	+	+	+	+	+	+	+**	-	+	+	+
OOC-3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NOCA-1	+	-**	-	-	+	+	+	+	-	-	+	+	+	+	+	+	+
PKC-3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CDC-42	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
UBC-18	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+
DHC-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ZYG-1	-	-	-	-	+	+	+	+	-	-	-	+	+	+	+	+	+
COL-139	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
T07C4.10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
F59E12.11	+	+	+	+	+	+	+	+	+	+	+	+	-**	+	+	+	+
TPXL-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
ICP-1	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+

Figure 3.26: Present ("+") and absent ("−") orthologues of essential components for polarity establishment in *C. elegans* among 16 nematode species and the nematomorph *Gordius sp.* For further details see caption of figure 3.14.

I also searched for orthologues of early cell polarity-inducing factors known from the model *C. elegans* (Introduction; Gönczy and Rose 2005). I found orthologues for essential proteins such as LET-99, PAR-2, PAR-5, GPR-1/-2, LIN-5, ZYG-1, SPD-2/-5

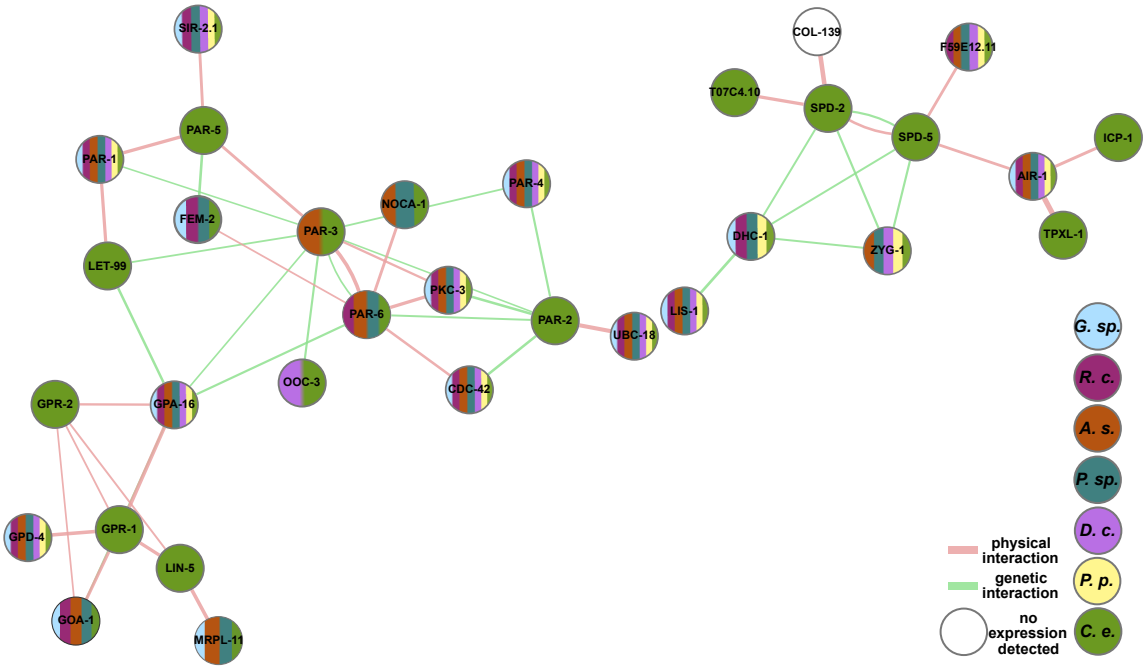

Figure 3.27: Schematic representation of the *C. elegans* signaling network for proteins involved in early polarity establishment. Only expressed orthologues of the pathway components in the early transcriptomes of seven species are shown (Fig. 3.26). For further details see caption of figure 3.14.

only in species of the genus *Caenorhabditis* (Fig. 3.26). Furthermore, I could verify the expression of these proteins only for *C. elegans* during early development (Fig. 3.27). In contrast to this, the core components of the PAR-3/-6/PKC-3 complex are generally conserved among nematodes, as I could identify orthologues in all investigated genomes (Fig. 3.26). All components of this protein complex are also expressed during early development in all (or nearly all) investigated nematode species and in the nematomorph *Gordius sp.* (Fig. 3.27). Nonetheless, the force-generating proteins which act in asymmetric cell division in *C. elegans* (GPR-1/2, LIN-5) and the upstream regulators, SPD-2/-5, PAR-2/-5 and LET-99 do not seem to be conserved in other nematodes (Fig. 3.26). Hence, this result suggests that polarity establishment in these species must be achieved by a modified signaling mechanism.

3.14 Early polarity establishment involves Lethal giant larvae, Scribbled and Crumbs signaling in nematodes

	genomes																
	N	I		II	III				IV			V					
	G. s. (*)	*R. c.* (*)	*T. s.*	*E. b.*	*A. s.* (*)	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* (*)	*D. c.* (*)	*P. p.* (*)	*C. a.*	*C. b.*	*C. r.*	*C. e.* (*)
Lgl	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DLG-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PAR-6	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+
Sema-5 c	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Zip	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Sdt	+	+	+	+	+	+	+	-	-	-	+**	+**	+**	+	+	+	+
Numb	+	+	+	+	+	+	+	+	+	+	+	+	+**	+	+	+	+
Crb	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Scrib	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Nuf	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
aPKC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tomos yn	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
Ssu72	+	+	+	+	+	+	+	+	+	+	+**	+	+	+	+	+	+
Pnt	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-
CycE	+	+	-	+	+	-	-	-	+	+	+	+	+	-	+	+	+
Sec13	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Pak	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Figure 3.28: Present ("+") and absent ("-") orthologues of basolateral polarity complex components among 16 nematode species and the nematomorph *Gordius sp.* For further details see caption of figure 3.14.

The Lethal giant larvae (Lgl) orthologue LGL-1 was recently found to be important for early embryogenesis in *C. elegans* (Beatty et al. 2010, 2013), as it interacts with the PAR-3/-6/PKC-3 complex during the 1-cell stage to induce polarity. In concordance with this I found orthologues of Lgl in other nematodes and also its expression during early development (Fig. 3.28, 3.29). I extended my search and detected orthologues of other interacting basoapical polarity genes to be expressed in all (or nearly all) investigated nematode species (Fig. 3.28, 3.29). Among these are Scribbled and Crumbs (Fig. 3.29). These proteins are members of the Lgl/Dlg/Scrib and Crumbs complexes involved in basoapical polarity establishment in epithelia, e.g. during early development of *Drosophila* (Knust and Bossinger 2002)

I found orthologues in all (or nearly all) nematodes. It would be interesting to explore whether *scrib* and *crumbs* is also important for polarity establishment in the model *C.*

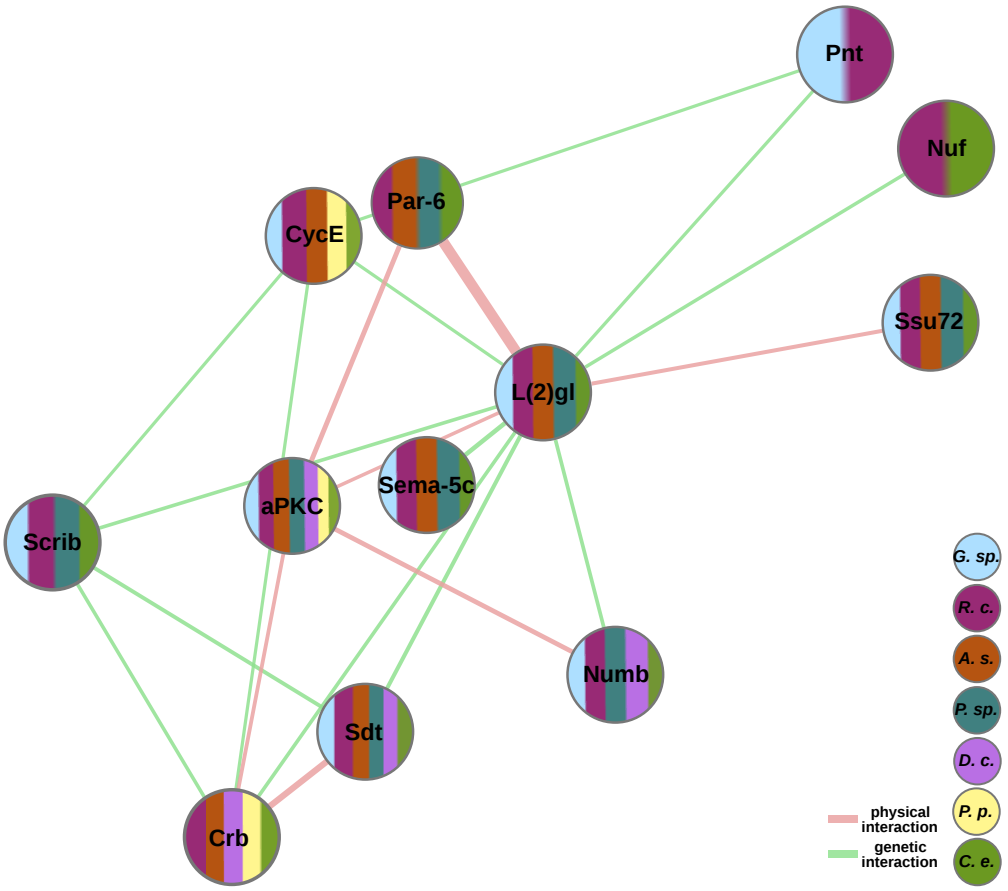

Figure 3.29: Schematic representation of the *Drosophila* network for basoapical polarity pathway components. Only expressed orthologues in the early transcriptomes of seven species are shown (Fig. 3.28). For further details see caption of figure 3.14.

elegans, a question not addressed so far.

My transcriptome analysis suggests that expression of the complexes PAR-3/-6/PKC-3, Lgl/Dlg/Scrib and Crumbs takes place generally in nematodes during early development (Fig. 3.29).

3.15 Hunchback, a highly conserved regulator and its role during Enoplean early development

	genomes																
	N	I		II	III				IV			V					
	G. s. (*)	*R. c.* (*)	*T. s.*	*E. b.*	*A. s.* (*)	*L. l.*	*B. m.*	*D. i.*	*M. h.*	*B. x.*	*P. s.* (*)	*D. c.* (*)	*P. p.* (*)	*C. a.*	*C. b.*	*C. r.*	*C. e.* (*)
Hb	+**	+	+	+**	-	-	-	-	-	-	-	-	-	-	-	-	-
Ftz	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ftz-f1	+	+	+	+**	+	+	+	+	+**	+	+	+	+	+	+	+	+
Mi-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Pho	+	+	+	+	+	-	-	-	+	-	-	-	-	-	-	-	-
Kr	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Kni	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nanos	+**	+**	+**	+**	+**	-	+**	-	+**	+**	-	-	-	+**	-	-	+**
Brat	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tsl	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bcd	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 3.30: Present ("+") and absent ("-") orthologues of the *Drosophila* signaling components among 16 nematode species and the nematomorph *Gordius sp.* For further details see caption of figure 3.14.

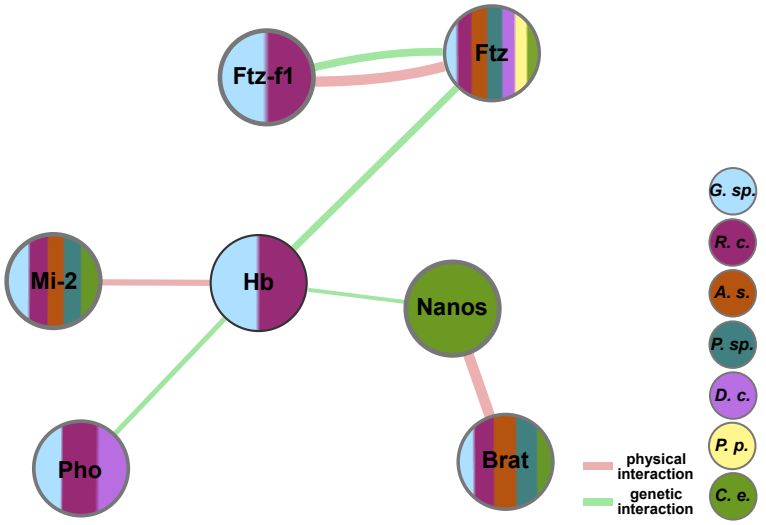

Figure 3.31: Schematic representation of the *Drosophila* Hunchback signaling components. Expressed orthologues of the pathway components during early embryogenesis of seven species are shown (inferred from transcriptomic data; Fig. 3.30). For further details see caption of figure 3.14.

I was interested whether early developmental genes from other model organisms may play a role during early development in nematodes, hence I screened 72 genes known to be expressed during early embryogenesis in other species such as *Drosphila*, *Xenopus*,

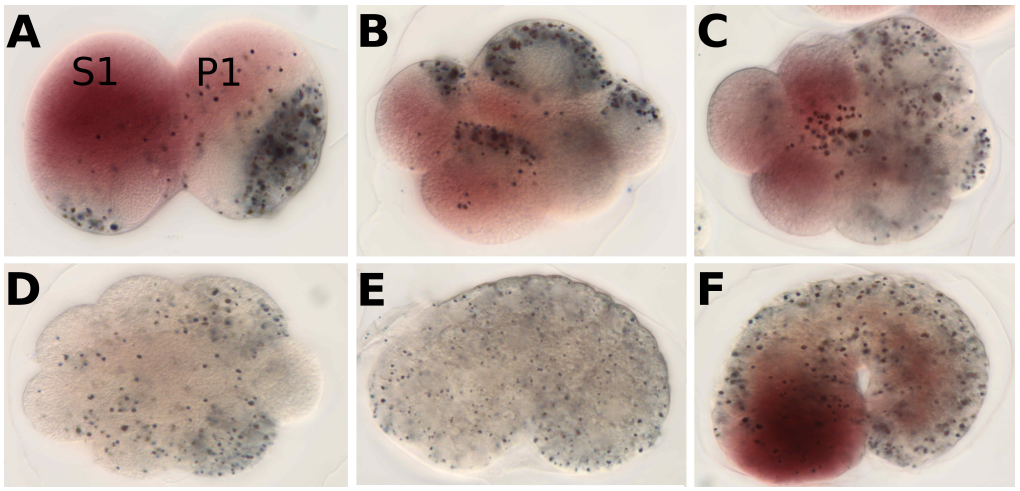

Figure 3.32: *R. culicivorax in-situ* hybridization showing the *hunchback* expression during **(A)** 2-cell, **(B)** 7-cell, **(C)** ca. 12-cell, **(D)** ca. 24-cell, **(E)** early morphogenesis and **(F)** advanced morphogenesis stage.

zebrafish, mouse and humans (Tab. 3.5). I found orthologues of the *Drosophila* gene *hunchback* in the genomes of Enoplean but not Chromadorean species (Fig. 3.30). Extending my search to interacting proteins, I found Fushi tarazu (Ftz), and its physically interacting protein Ftz-f1. In addition, I detected orthologues for Pleiohomeotic (Pho) in Enoplean nematodes and in the Chromadorean nematode *D. coronatus* (Fig. 3.30). All these interacting orthologues are expressed during early development (Fig. 3.31). Together with my collegues from the Schierenberg laboratory, I further investigated the expression pattern of *hunchback* in *R. culicivorax* via *in-situ* hybridisation. We found the *R. culicivorax* orthologue of *hunchback* to be stronger expressed in the somatic precursor cell S1 (named AB in *C. elegans*) and its descendants (Fig.3.32 A, B, C). Much later, during morphogenesis, we found strong expression in the head of the embryo (Fig. 3.32 F). Temporal and spatial expression is in concordance with the *Drosophila* orthologue, indicating functional conservation between *Drosophila* and the Enoplean nematode *R. culicivorax*. The late anterior expression seems to be in concordance with the *Drosophila* orthologue as well, as Hunchback is known to be involved in head development in short and long germ insects (Wolff et al. 1995, Perry et al. 2012). These results suggest that Hunchback signaling and its expression is conserved between *R. culicivorax* and *Drosophila* and thus may be conserved in all Enoplean nematodes and nematomorphs.

3.16 The parthenogenotes *Diploscapter coronatus* and *Panagrolaimus sp. PS1159* reveal unique transcriptomic features

In my analysis of early expression among selected nematodes (Fig. 3.13) I incorporated two parthenogenetic species of two different clades: *Diploscapter coronatus* (clade V) and *Panagrolaimus sp. PS1159*(clade IV). These parthenogenotes lack sperm which is known to be the initial trigger for polarity establishment, subsequent asymmetric cell division and soma/germline separation in *C. elegans* (Introduction; Goldstein et al. 1998, Lahl et al. 2006). Thus, these processes must be controlled by a different mechanism in parthenogenetic species. Nevertheless, all three nematodes show similar characteristics during early embryogenesis, even though there are some differences concerning intracellular events and behaviour of blastomeres during the first divisions (Introduction; Goldstein et al. 1998, Lahl et al. 2006). Therefore, the question arises, what the mechanism is, which drives the essential process of polarity establishment, subsequent asymmetric cell division and soma/germline separation in the absence of sperm.

Through sequencing of early transcriptomes of both species and clustering of protein families via OrthoMCL (section 3.7; Materials and Methods), I found 77 clusters which are unique for these parthenogenotes. These clusters include 100 *D. coronatus* and 97 *Panagrolaimus sp. PS1159* genes. By searching for homologues in other animal models such as humans, mouse, *Drosophila* and *C. elegans* via Blastp, I identified 17 homologous sequences for humans, 16 for mouse, 8 for *Drosophila* and 260 for *C. elegans*.

Among the potentially interesting ones are the homologous sequences to the mTORC signaling effectors LAMTOR4 and LAMTOR5. Aligning these sequences with outgroup sequences, I confirmed that these are indeed orthologues (Fig. 3.33).

LAMTOR4 and LAMTOR5 are expressed early and exclusively in both parthenogenotes. As both proteins are important components of the mTORC signaling pathway (Bar-Peled et al. 2012), it would be fascinating to explore whether LAMTOR4/5 play a role in polarity induction during early embryogenesis in both parthenogenotes as substitution for the absent sperm. In addition, it would be interesting to overexpress LAMTOR4/5 orthologues from the parthenogenotes in *C. elegans* and check whether this induces parthenogenetic development or at least certain steps towards it.

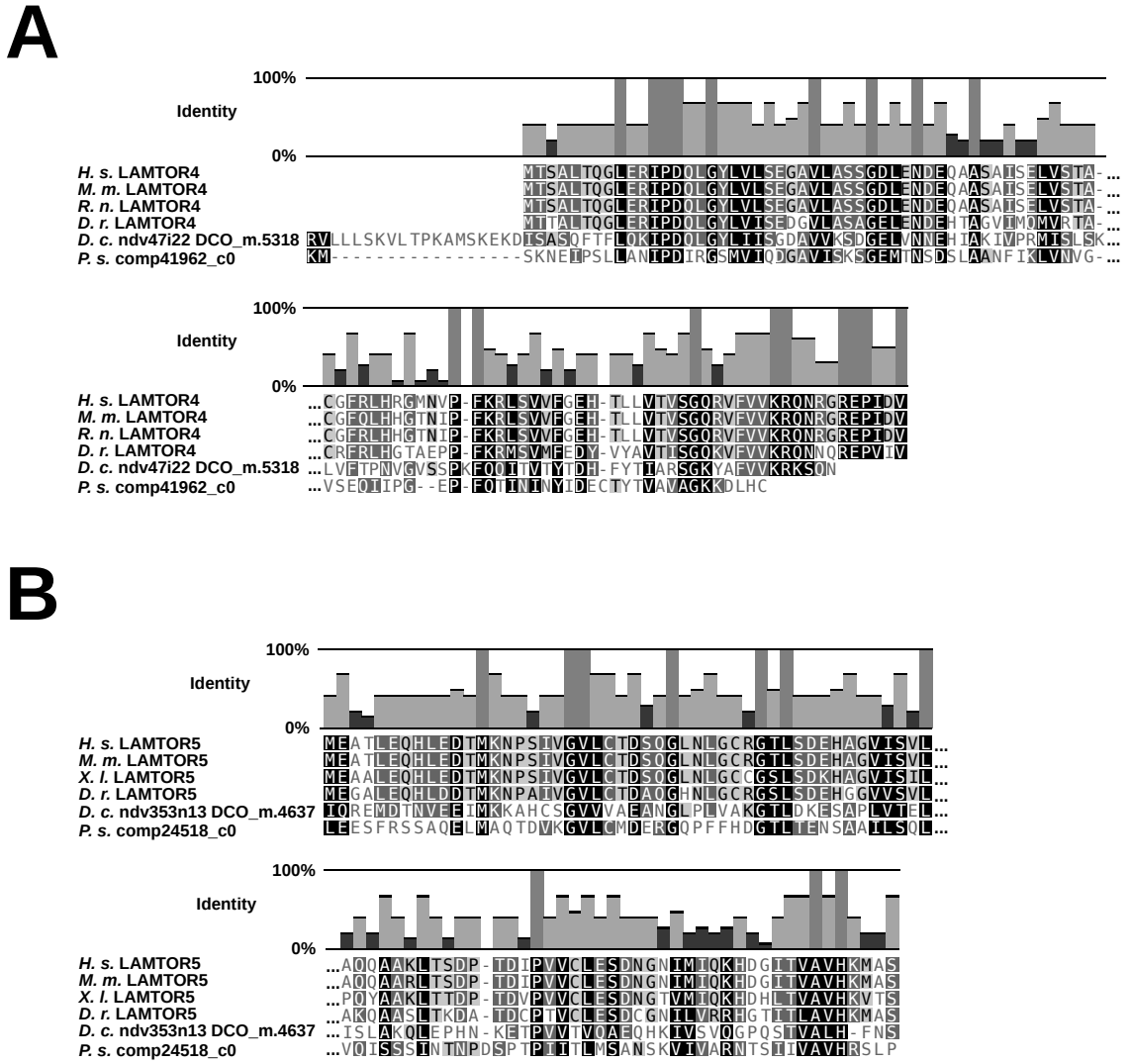

Figure 3.33: Protein alignments of the mTOR modulators LAMTOR4 (**A**) and LAMTOR5 (**B**) of vertebrate species and the identified homologous protein sequences of the parthenogenotes *D. coronatus* and *Panagrolaimus sp. PS1159*.

3.17 *D. coronatus*, a product of interspecies hybridisation?

The parthenogenetic nematode *D. coronatus* is a rather close relative to *C. elegans* (Kiontke and Fitch 2005). This particularity leads to the question of what is required for parthenogenesis to arise, especially with regard to the genus *Caenorhabditis* where no parthenogenetic nematode has been found. Therefore, I compare the *D. coronatus* genome (which I could access due to a cooperation with the laboratory of Yuji Kohara, Mishima, Japan; Materials and Methods) with other sexual species of clade V (Fig. 3.34). In order to do so, I used OrthoMCL (Li et al. 2003) and increased

the resolution of the clustering process by incorporating in addition to *C. elegans* four other previously sequenced *Caenorhabditis* species, as well as the hermaphroditic satellite model *Pristionchus pacificus*, and as outgroups the dioecious *Panagrellus redivivus* (clade IV) and *Ascaris suum* (clade III). In total I found 8,135 orthologous clusters, which includes 1,734 *P. redivivus*+*A. suum* specific clusters and shared clusters among all investigated species (Fig. 3.34). 80% of these were present in clade V-species, indicating a nematode-specific core set of protein families. 11,589 orthologue clusters were not shared with *P. redivivus*+*A. suum* (Fig. 3.34). This indicates that the majority of clusters are clade V specific.

I compared proteomes and their respective orthologues shared between the studied species. By looking at the 5 *Caenorhabditis* species on the one hand and at *D. coronatus*, *P. pacificus* and *P. redivivus* on the other hand, I found that 5,699 orthologue clusters, or nearly 50% of all clade V-specific clusters were restricted to the genus *Caenorhabditis*. This suggests that during evolution a considerable number of new genes must have arisen in the lineage leading to the genus *Caenorhabditis*.

I recognized a high number of *D. coronatus*-specific orthologue clusters (3,378; Fig. 3.34). Roughly half of them (1,835) were found to consist of two proteins while the number of unique species-specific proteins, that therefore do not cluster, was relatively small (2,727). In contrast, in other nematode species also analysed by second generation sequencing, e.g. *C. japonica* or *C. angaria* (Materials and Methods; Mortazavi et al. 2010) the number of clusters which consist exclusively of two proteins is much smaller (649 in the case of *C. japonica* and 699 in the case of *C. angaria*). This result means that an unusually high percentage of *D. coronatus*-specific clusters contains exactly two proteins. This led to the hypothesis that *D. coronatus* may have arisen from inter-species hybridisation and hence its genome contains two alleles for each gene.

I noticed that most *D. coronatus*-specific clusters consisted of just two distinct protein variants. In order to determine whether in general the *D. coronatus* genome contains two distinguishable alleles for each gene, I performed a gene sequence analysis of the whole genome and in contrast to *C. elegans* found for the vast majority of genes (93%) two clearly distinguishable alleles (Fig. 3.36 A).

It has been claimed that parthenogenesis commonly arises via interspecies hybridisation (Simon et al. 2003). To explore this possibility for *D. coronatus* together with other

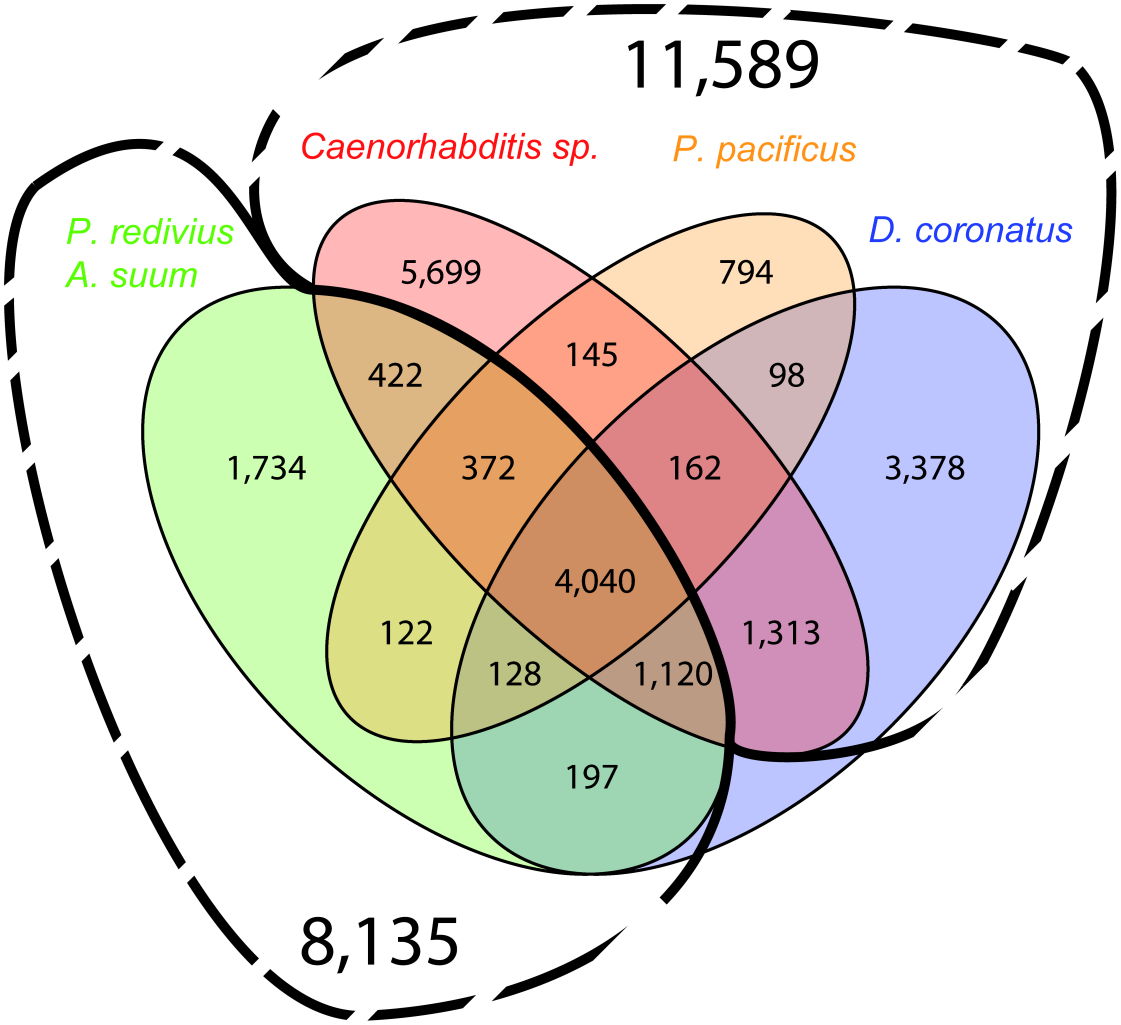

Figure 3.34: Distribution of shared and specific orthologous clusters for the clade V species *D. coronatus* (blue), *P. pacificus* (orange) and the genus *Caenorhabditis* (red), as well as for the respective outgroups *Panagrellus redivivus* (clade IV) and *Ascaris suum* (clade III) (green). Numbers in the Venn diagramm indicate cumulative numbers of unique or shared orthologous clusters, or protein families. 8,135 represents the sum of outgroup-specific (green) clusters plus clusters shared with clade V species. 11,589 represents the sum of clade V-specific unique and shared clusters.

members of the Schierenberg laboratory we performed a single-worm amplification of highly conserved rDNA genes and retrieved several clones of poly-copy rDNA genes (Fig. 3.35 A). We separately cloned sequences of the ITS, SSU and LSU regions. When aligning these and applying Maximum Likelihood and Bayesian statistics, all exhibited single nucleotide polymorphisms (SNPs) could be allocated to one of two distinct variants (Fig. 3.35 B). This indicates the presence of two distinct alleles for each of three tested rDNA genes (Fig. 3.35 B) a phenomenon not found in *C. elegans*.

These findings could either be explained with the coalescence of genomes from two

closely related dioecious species. Alternatively, they may be the result of endoduplication accompanied by accumulation of mutations as a consequence of parthenogenetic reproduction ("Meselson effect"; Mark Welch and Meselson 2000, Birky 2004).

If these were in fact originally derived from two different species one should expect differences in synteny, i.e. in the positions of corresponding gene sequences in the genome. To investigate this, we arbitrarily selected 11 single-copy genes conserved in the nematode phylum (Mitreva et al. 2011) and found that they not only are present in exactly two distinct alleles but are positioned on different contigs of the *D. coronatus* genome, too (Fig. 3.36 D, 3.37).

In search for further evidence for a hybridisation event we studied single-copy genes. While a Meselson effect should lead to a random distribution of synonymous and non-synonymous exchanges, interspecies hybridisation should result in an enrichment of synonymous mutations preserving protein function. Using Interproscan (Quevillon et al. 2005, Hunter et al. 2012) we identified conserved regions in *D. coronatus* single copy genes and checked the type of amino acid exchanges. We found the number of synonymous exchanges to be significantly higher in conserved domains than in less conserved regions (Fig 3.35 C). The same result was obtained when we compared *C. elegans* with *C. remanei* (Fig 3.35 C). Thus, amino acid sequences of the selected single-copy genes appear to be conserved to a similar extent between two closely related species as between two alleles in *D. coronatus*.

In summary, for the case of *D. coronatus* our data favor the assumption of an interspecies hybridisation. This does, however, not preclude an additional Meselson effect.

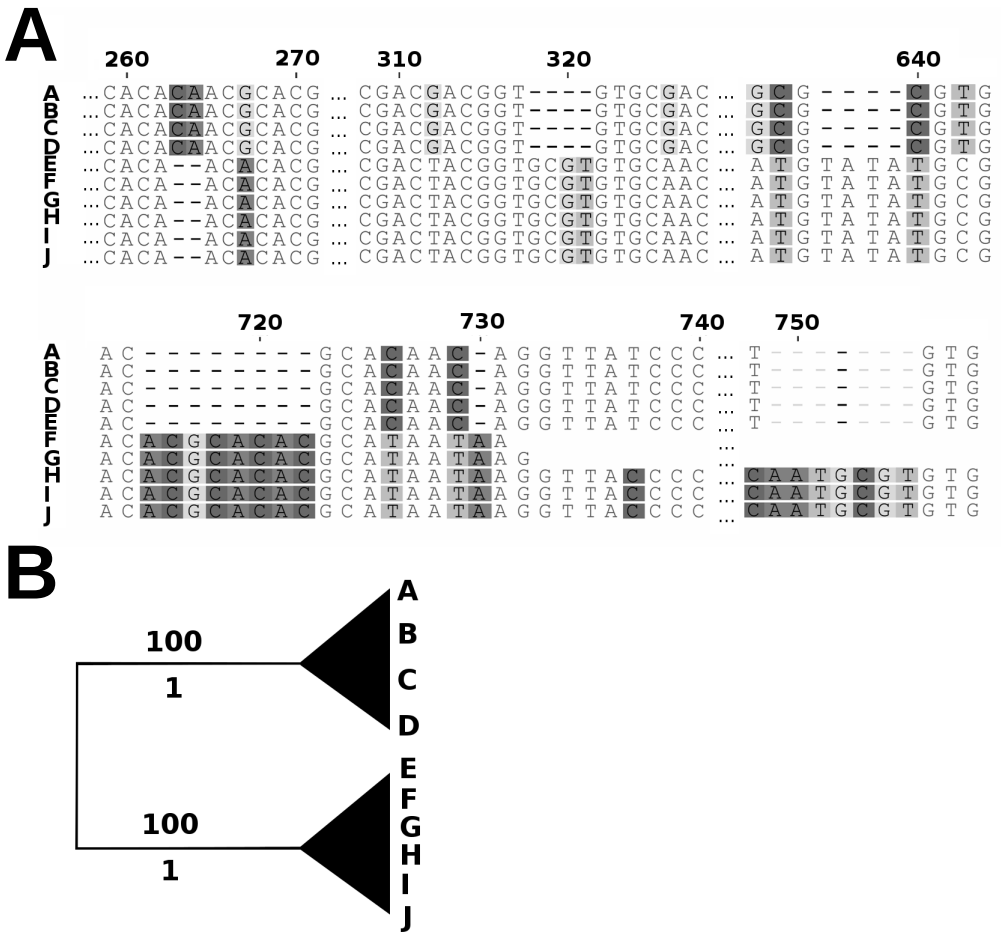

Figure 3.35: Sequence comparison of the small interspacing sequence (ITS) rDNA gene of *D. coronatus*. (**A**) Sequence alignment of individual clones (A-J) shows selected regions with distinct single nucleotide polymorphisms (SNPs). (**B**) Collapsed Maximum Likelihood tree visualizes two distinct clusters of sequenced clones. Bootstrap values are shown above and posterior probability beneath branches. Figure adopted from Theresa Vogt (2012; master thesis, University of Cologne; Vogt 2012)

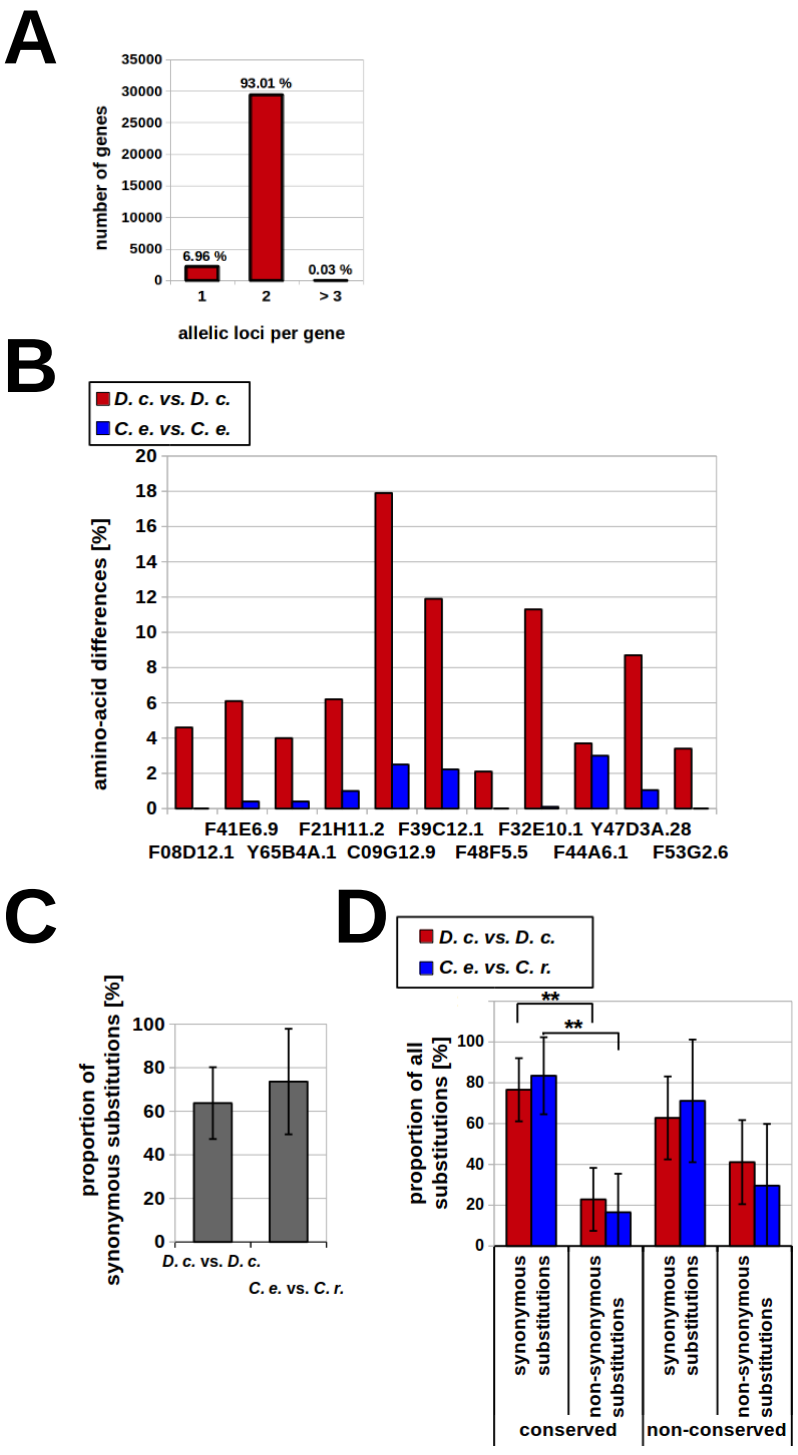

Figure 3.36: Analysis of *D. coronatus* allele frequency over the whole genome (**A**) and 11 arbitrarily selected highly conserved single copy genes (**B, C, D**). (**A**) Number of of genes which were found to map to 1, 2, 3 or more positions on the genome. (**B**) Amino-acid differences among variants of *C. elegans* (blue) and the two alleles of *D. coronatus* orthologues (red). (**C**) Comparison of synonymous substitutions between *D. coronatus* alleles and between *C. elegans* and *C. remanei* orthologues. No significant increase of synonymous subsitutions were found in the parthenogenote *D. coronatus* in comparison to the sexual relatives *C. elegans* and *C. remanei*. (**D**) Percentage of synonymous and non-synonymous substitutions in non-conserved and conserved protein regions. Significant differences are indicated by asterisks (α <0.01; n = 11; bars, standard-deviation). Abbreviations: *C. e.*, *C. elegans*; *C. r.*, *C. remanei*; *D. c.*, *D. coronatus*.

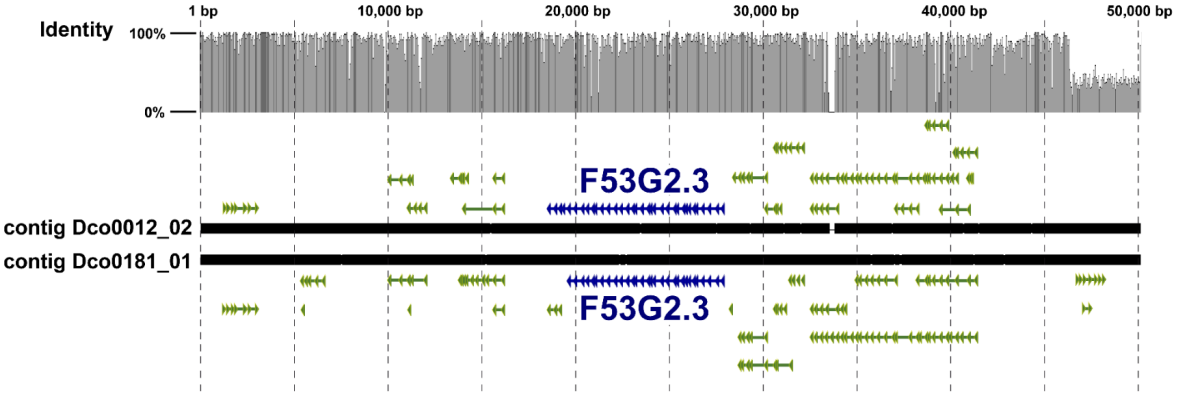

Figure 3.37: Genomic alignment of *D. coronatus* contigs representing alleles of the single copy gene F53G2.3 (blue). Alignment shows the most proximal upstream and downstream approximately 20,000 bp on both contigs, including all gene predictions in the vicinity of F53G2.3 (marked green). Arrowheads indicate exons and bars connecting arrowheads indicate introns.

Chapter 4

Discussion

4.1 A Phylum-wide transcriptome comparison demonstrates expression of developmental regulators so far not known from nematodes

In my thesis I addressed questions dealing with the transition from gamete to embryo and the early embryogenesis among Enoplean and Chromadorean nematodes (see Introduction; section 1.9). My main focus is a phylum-wide comparison of Chromadorean and Enoplean nematodes with the model *C. elegans* as reference system. Therfore, I compared expressed orthologues during germline formation and early embryogenesis in selected species among the phylum of nematodes and the nematomorph *Gordius sp.* This allowed an relatively unbiased investigation (see Introduction) of underlying expression patterns and inference of expressed pathways. Furthermore, I also wanted to contribute to a better understanding of oogenesis in *C. elegans* and therefore investigated the gene regulatory network (GRN) involved in oocyte maturation by knock-down of key regulators and subsequent RNA analysis techniques (see Results).

In order to analyse expression patterns, I used two different strategies: (1) I searched for orthologues involved in gonad formation, gametogenesis (see Results; sections 3.3 to 3.5) and early embryogenesis (see Results; sections 3.7 to 3.15), which do not exist on the genomic level in non-*Caenorhabditis* nematodes (and nematomorphs). (2) I searched for nematomorph-, Enoplea- and Chromadorea-specific genes which either have no orthologues in the genus *Caenorhabditis*, or are expressed exclusively in at least one

of the three above-mentioned taxa. This way, I identified important regulators known from other model organisms such as *Drosophila*, *Xenopus*, zebrafish, mouse and humans which are specific for Chromadorea, Enoplea and nematomorphs with respect to gonad formation, gametogenesis and/or early embryogenesis.

I identified crucial proteins of Wnt, Notch signaling, the global sex determination pathway, oogenesis silencing complexes and constitutive P-granule proteins to be specific for either the genus *Caenorhabditis*, or Chromadorea (section 3.3). These results suggest that signaling during these developmental processes must be regulated differently in Enoplea and nematomophs. From differentially expressed transcripts (analysed by RNA sequencing; see Results) of *R. culicivorax* virgins and males, I found first candidates which either are specific for Enolea and nematomorphs, or are at least not known to be important for gonad formation and gametogenesis in *C. elegans*.

Investigating early embryogenesis of selected Chromadorea and Enoplea, I found pathway components relevant to polarity establishment, such as the PAR-3/-6/PKC-3, Lgl/Scrib/Dlg and the Crumbs complexes, to be conserved among all nematodes and to be expressed during early embryogenesis in all (or nearly all) species. Strikingly, I found orthologues of proteins expressed early, which do not exist in the model *C. elegans* and other *Caenorhabditits* species. Such proteins are BMP/DPP inhibitors Chordin/Sog (Short gastrulation), Noggin, Crossveinless-2 and the NF-κB transcription factor (see Results). These findings point towards completely different signaling pathways which were previously not known to act during early embryogenesis in nematodes.

As I wanted to get even deeper insights of the potential role of these pathways during early embryogenesis, I screened for expressed genetically and/or physically interacting proteins and found that all of the aforementioned Enoplean-specific components of signaling pathways and complexes are expressed during early embryogenesis. Thus, at least in theory, these pathways can be fully functional during early embryogenesis and may have important roles in Enoplean early embryogenesis. Here, I want to discuss their putative function and I want to explore whether they may play a role in polarity establishment, axis specification and patterning of the early embryo. In this context I also want to explore what might be the function of PCP complexes such as Lgl/Scrib/Dlg and Crumbs during early embryogenesis. I will propose a set of "core" pathways which are conserved among all nematodes and I will show which pathways exist specifically in Enoplea and what function these pathways may play in these nematodes. I will speculate about the conservation of such pathways between "basal" Enoplea other animal

phyla. Furthermore, I want to speculate which set of pathways are ancestral and might have been lost in the course of Chromadorean and *Caenorhabditis* evolution.

4.2 Gamete-to-embryo transition: a consequence of gonad formation, oogenesis, oocyte maturation and maternally expressed gene products

A prerequisite for early embryogenesis is the transition of the oocyte to the embryo, a process which depends on several developmental processes such as oogenesis, and spermatogenesis. The special case of parthenogenetic nematodes will be discussed later. Even though, the underlying genes and pathways have been already described for these processes in the model *C. elegans*, we sought to better understand oocyte-to-embryo transition by analysing the GRN of oocyte maturation in *C. elegans* (see Results; sections 3.1, 3.2). Here, I will discuss how our findings identified new candidate genes for this GRN. Althrough, much is known about oocyte-to-embryo transition from *C. elegans* (see Introduction; Miller et al. 2003, Govindan et al. 2006) it is unclear to what extent the same genes also control these processes in other nematode species. Hence, I analysed gonad formation and gametogenesis in other nematodes such as the clade I nematode *Romanomermis culicivorax* (section 3.4) and the clade III nematode *Ascaris suum* (section 3.5). This allows an investigation of present and absent orthologues for each *C. elegans* genes involved in the described processes from three independent experimental sets (see Results; sections 3.3, 3.4 and 3.5). I identified similarities with respect to most genes relevant for these biological processes, but for the described key regulators known from *C. elegans*, I found many genes to be absent in other nematodes. These results indicate fundamental differences in the developmental control. Here, I will discuss what the differences are between Enoplea, Chromadorea and in particular species of the genus *Caenorhabditis*. In the end, I will discuss which regulators can be considered as the "original" gene set of the last common ancestor of all nematodes for the processes looked at.

4.2.1 *C. elegans* oocyte maturation: Recruitment of genes originally involved in neuronal development?

Oocyte-to-embryo transition is a prerequisite for propagation in nearly all Metazoa. This process includes oogenesis including oocyte maturation (see Introduction). In *C. elegans* it is known that a MAP-kinase pathway controls this process (Miller et al. 2003, Govindan et al. 2006). Nonetheless, knowledge of crucial components of signal transduction, such as involved serpentine receptors, gap-junction molecules, transcription factors and target genes remains rather elusive (Fig. 1.3). In this thesis, I presented our RNAi and microarray-based approach to identify candidate genes involved in this GRN (see Results; section 3.1 and 3.2). By knocking out genes known to be involved in the underlying GRN we found in total 1,458 candidate genes which may contribute to this GRN (section 3.1). For 8 of these candidate genes (Tab. 3.2), I could demonstrate that they are involved in oocyte maturation in *C. elegans*, as RNAi knock-down experiments showed impaired oocyte maturation (Fig. 3.5, 3.6, 3.7). Interestingly the knock-down by RNAi of these candidate genes not only decreased the number of hatched larvae (Fig. 3.5, 3.6), it also showed prominent abnormal phenotypes (Fig. 3.7); i.e. the gonads contained a high number of small box-shaped oocytes (Fig. 3.7 D, F, H, J). What could be the explanation for such a phenotype? The continuous production of oogonia and oocytes by the germ stem cells (Kimble and Crittenden 2007) seems to cause a continuous influx of oocytes and their stacking in the gonad due to the absense of proper maturation and subsequent fertilisation. This is in concordance with knock-down experiments, where uteri either contain no embryos (Fig. 3.7 E, K) or only degenerated ones (Fig. 3.7 G, I). Interestingly the knock-down experiments phenocopy defects in genes already known to be involved in this GRN (Miller et al. 2003, Whitten and Miller 2007). Therefore, my RNAi experiments caused impaired oocyte maturation and thus the knocked-down genes have roles in oocyte maturation. Taking into account that all of the mentioned genes are either innexins, transcription factors, or trans-membrane receptors (Tab. 3.2), they may function upstream of MPK-1 or are directly involved in activation of oocyte maturation (Fig. 4.1).

As, I did not characterise the role of each gene within this GRN. I can only speculate what the potential role of the identified genes may be: It is known that STR-102 and PLX-2 are transmembrane receptors (the former is a serpentine receptor and latter a semaphorin receptor involved in axon guidance; Ikegami et al. 2004; Tab 3.2), so

these genes are candidates for the unknown sepertine receptors of the GRN (Fig. 4.1). However, it remains to be determine whether MSP binds both receptors. The innexins INX-19 and UNC-7 may be candidates for cell-cell interactions between the somatic sheath cells and the oocyte (Fig. 4.1). Maybe both even interact in the formation of innexin heterohexamers (innexons; gap junctions). Double knock-down of both genes seem to have a synergistic effect, as the number of hatched larvae slightly decreased in comparison to separate knock-downs of both genes (Fig. 3.5). Surprisingly, I also found a cyclic nucleotide-gated channel protein (CNG-1) to be involved in oocyte maturation (Fig. 3.5, 3.6). CNG-1, a protein known to be involved in thermo- and chemosensing (Cho et al. 2005), opens the possibility of an alternative cell-cell interaction mechanism, perhaps via ions or small molecules secreted by the somatic sheath cells, inducing oocyte maturation (Fig. 4.1). Furthermore, I identified several DNA-binding proteins (CND-1, APTF-1, APTF-4, NHR-81 and ZTF-11) which might play roles in initiating oocyte maturation. These genes may be activated by phosphorylation of MPK-1. Especially in the case of APTF-1 and APTF-4 this seems plausible as these transcription factors are predicted to be homologous to AP-2 (Turek et al. 2013), a transcription factor of *Drosophila* involved in c-Jun N-terminal kinase (JNK) signaling (Kockel et al. 2001). It was shown for *Drosophila* that AP-1 acts downstream of Erk (the *Drosophila* orthologue of MPK-1; Kockel et al. 2001). If the AP-1 homologues APTF-1/-4 indeed act downstream to MPK-1, it would be interesting to investigate whether JNK signaling is involved in the process of oocyte maturation in *C. elegans*.

Analysing the 8 genes acting in oocyte maturation in *C. elegans*, I noticed that these genes are also involved in neuron development and function (Kockel et al. 2001, Ikegami et al. 2004, Cho et al. 2005, Chuang et al. 2007, Starich et al. 2009, Turek et al. 2013). In addition, VAB-1, an ephrin receptor tyrosine kinase (RTK), has also dual functions in neuron development (Mohamed and Chin-Sang 2006, Grossman et al. 2013) and oocyte maturation (Fig. 4.1; Miller et al. 2003). Hence, several genes important for neuron development and function may have been recruited during the evolution of *C. elegans* to mediate oocyte maturation and the oocyte-to-embryo transition. It would be fascinating to investigate this particularity in other Chromadorea and in Enoplea to reveal whether this dual function of certain genes is a conserved feature or emerged in one phylogenetic branch secondarily.

Furthermore, it would be attractive to screen more of the remaining 1,450 candidate genes by RNAi to find additional genes involved in oocyte maturation: In my analysis

of the presented 8 candidate genes (Results; section 3.2), phenotypes with an elevated production of defective embryos and simultaneously a depleted gonad were not found. Such a phenotype would have uncovered genes which ectopically activate oocyte maturation (downstream targets; Fig. 1.3) after knock-down of upstream regulators. In conclusion the expression data documented in this thesis encourage further screening for candidates to elucidate the GRN of oocyte maturation and of oocyte-to-embryo in *C. elegans*.

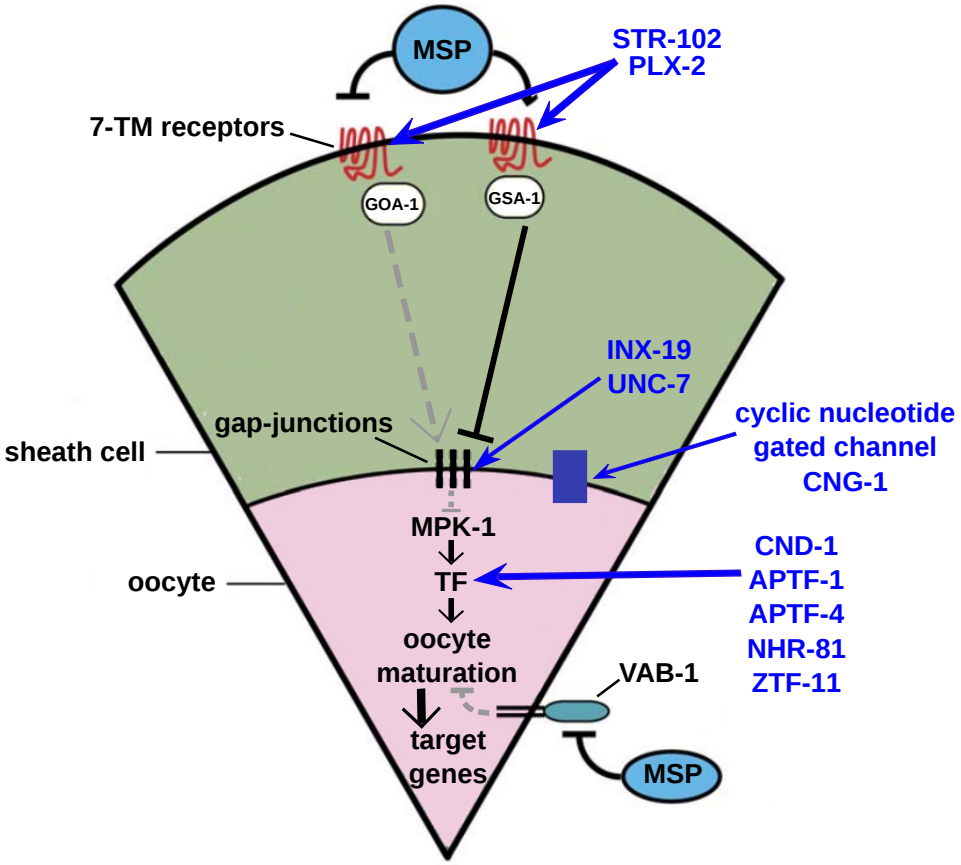

Figure 4.1: Schematic representation of cell-cell interactions between somatic sheath cells and the "−1 oocyte" and the underlying GRN which regulates the process of oocyte maturation. Grey dashed lines indicate activating (arrow head) and inhibiting (blunt end) signal of oocyte maturation when sperm and MSP are absent. When sperm is present (black arrows) MSP binds to seven-transmembrane receptors and the ephrin receptor VAB-1 and activates the MAP-kinase MPK-1. New candidates involved in oocyte maturation (see Results; section 3.2) are visualised in blue. Potential roles and positions of new candidate genes in the regulatory network are indicated by blue arrows. Modified after Govindan et al. (2006).

4.2.2 Expression and genome analysis of the Enoplea *R. culicivorax* and the Chromadorea *A. suum* indicate new candidate genes associated with germline formation and gametogenesis in all nematodes

Germline development and gamete formation include the processes of gonad formation, gametogenesis, the shift from spermatogenesis to oogenesis (in case of hermaphrodites) and oocyte maturation. These processes are prerequisites for fertilisation and embryogenesis (see Introduction). I showed that regulators, known from the model *C. elegans*, may be abesent in other species (see section 3.3). On the other hand, I found conserved regulators for all nematodes and regulators specific for Enoplea and/or Chromdorea (see sections 3.4 and 3.5).

What are fundamental differences between *C. elegans* (and additional representatives of the genus *Caenorhabditis*) and other nematode species? Here, I am going to discuss Chromadorea and/or *Caenorhabditis*-specific regulators for each mentioned process. I also discuss which other orthologues, found by differential expression analysis between adult *R. culivorax* virigns vs males (see Results; section 3.4) and by screening for germline specific genes from the *A. suum* genome (see section 3.5; Wang et al. 2012), may be involved in these processes and may even replace *Caenorhabditis*-specific regulators:

Gonad formation: I found that important Wnt signaling components involved in gonad formation in *C. elegans* were absent in other nematodes outside clade V (Fig. 3.8). These include the LIN-17 frizzled receptor and the β-catenin SYS-1. For the latter I could show that this particular β-catenin exclusively exists in the genus *Caenorhabditis* (Fig. 3.19). Hence, Wnt signaling as known from *C. elegans* cannot be mediated via these proteins in other nematodes, where alternative regulators must exist which mediate gonad formation. In adult *R. culicivorax* virgins I found orthologues for the *Drosophila* proteins Kugelei/Fat2 and Furry to be overexpressed (Fig. 3.19 and 3.10) that are important for the control of gonad elongation in *Drosophila*. In *Drosophila* they act in the so-called Fat/Dachsous planar cell polarity (PCP) pathway to control gonad elongation (Viktorinová et al. 2009, Horne-Badovinac et al. 2012). Surprisingly, in *C. elegans* orthologues of these genes cannot be correlated with gonad formation. Presently, it cannot be excluded that the expression in *R. culicivorax* virgins takes place

in some somatic tissues not important for germline-associated processes. Therefore, it would be intriguing to test their potential involvement during gonad formation by gene knock-out or mutagenesis. As Kugelei/Fat2 and Furry are essential for a-p axis formation in *Drosophila* (Horne-Badovinac et al. 2012) orthologues of these proteins may be important for axis formation and polarity establishment in non-*C. elegans* nematodes. Whether orthologues to Kugelei/Fat2 and Furry are an alternative to Wnt signaling during gonad formation needs to be investigated by functional assays, such as gene knock-down via RNAi or genome editing. In addtion, ectopically overexpressed *C. elegans* orthologues of Kugelei/Fat2 and Furry may show their possible function during *C. elegans* gonad formation.

Shift from spermatogenesis to oogenesis: I found that certain orthologues for regulators of the spermatogenesis-to-oogenesis shift were specific to Chromadorean species (Fig. 3.8 B). These involve components of the "global sex determination" pathway (Kimble and Crittenden 2007), such as FEM-3, FOG-1 and TRA-1. The question arises whether alternative proteins exist in other Chromadorean and Enoplean species with similar functions in sex determination. In that respect I found several candidates in *A. suum* which may have redundant function to the clade V-specific genes FOG-1 and TRA-1 (Tab. 3.3): *ASU_11803* and *ASU_14618*. These genes are predicted to play a role in germ cell or gamete development, as homologous sequences were found in *C. elegans*, *Drosophila* and humans by Blast2GO analysis (Tab. 3.3). Nonetheless, for both genes I could only detect orthologues in nematodes from clade III (Tab. 3.3). Hence, these genes may have taken up the role of the "missing" regulators FOG-1 and TRA-1 at least for clade III nematodes. A redundant function could be tested by rescue experiments in a *tra-1* and/or *fog-1* loss-of-function mutant background in *C. elegans*. In addition, I found the *A. suum* gene *ASU_10038* which is predicted to be homologous to *C. elegans* FEM-3. But *A. suum* has an orthologue for FEM-3 (here called aFEM-3). aFEM-3 did not cluster with ASU_10038, indicating that there might be two functionally redundant genes which take up the role of aFEM-3 in the *A. suum* germline.

For Enoplean nematodes, and in particular for *R. culicivorax*, I could not identify orthologues in other vertebrate and invertebrate model organisms which are involved in the shift from spermatogenesis to oogenesis. But this is not surprising, due to two circumstances: (1.) The shift of oocyte production to spermatogenesis is specific for hermaphroditic nematodes, such as *C. elegans* (*A. suum* and *R. culicivorax*

are bisexual species). Previously, it could be shown that the trait of hermaphroditism evolved independently several times in nematodes (Kiontke et al. 2011) and in *C. elegans* hermaphroditism is reversible after e.g. *fog-2* knock-out (Nelson et al. 1978, Barton et al. 1987). (2.) Especially, sex determination pathways tend to be extremly susceptible to evolutionary change (True and Haag 2001). Hence, it must be doubted that the molecular underpinnings which are part of *C. elegans* "global sex determination" pathway exist in *R. culicivorax* and *A. suum*.

Germline stem cell maintenance and entry into meiosis: From *C. elegans* it is known that mitotically dividing oogonia (or germline stem cells) are controlled by Notch signaling between the Distal Tip Cell (DTC) and adjacent oogonia (Fig. 3.8 B; Kimble and Crittenden 2007). I could show that Notch signaling and the subsequent pathway inducing meiosis involve genes for which orthologues only exist in Chromadorean nematodes. These encode the proteins LAG-2, FBF-1/-2, NOS-3 and GLD-3 (Fig. 3.8 B). However, the molecular signaling maintaning of a pool of mitotically dividing oogonia and the entrance of meiosis must be different in Enoplean nematodes, as no orthologs could be detected for LAG-2, FBF-1/-2, NOS-3 and GLD-3. I could not retrieve proteins which may have the same function in *R. culicivorax*. Nonetheless, it is hard to believe that Notch signaling is not responsible for gametic stem cell maintenance in Enoplean nematodes, as this is the case in the outgroup model *Drosophila* (Song et al. 2007, Xie et al. 2008). But the downstream signaling which involves FBF-1/-2, NOS-3 and GLD-3, appears to be the same in all Chromadorea. In contrast, GLD-3, a Bicaudal-C homologue which is essential for the shift from mitosis to meiosis in *C. elegans* (Eckmann et al. 2002, Kimble and Crittenden 2007) seems to be absent in Enoplean nematodes.
Intriguingly, in *C. elegans* exists BCC-1, an orthologue of Bicaudal-C from *Drosophila*. In *Drosophila* Bicaudal-C is essential for entry into meiosis and the posterior patterning of the embryo (Chicoine et al. 2007). But BCC-1 is not known to play a role in this process in *C. elegans*, nor was BCC-1 mRNA, or protein found to be expressed in the gonad at all. Differential expression analysis in *R. culicivorax* virgins and males demonstrated that the *Drosophila* Bicaudal-C orthologue not only exists in this Enoplean nematode, but is also expressed during this process. This suggests in contrast to *C. elegans* that the *R. culicivorax* Bicaudal-C orthologue may play a role during meiosis and axis specification during oogenesis. Especially, the latter aspect would be very interesting to

test experimentally in *Drosophila*: Maybe expression of *R. culicivorax* Bicaudal-C could rescue *bicaudal-C* loss of function (lof) alleles in *Drosophila*. Such an experiment would proof conserved functions of both orthologues and would pose an alternative for axis specification in Enoplean nematodes.

Oogenesis and oocyte maturation: The processes of oogenesis, including oocyte maturation, is essential for oocyte-to-embryo transition (see Introduction). For both processes I found genes which have orthologues only in Chromadorean species, or are even limited to clade V nematodes (Fig. 3.8 C). These involve the microRNA suppressors FLH-1/-3 (Walhout et al. 2002, Ow et al. 2008), the chromsome specific DNA binding proteins ZIM-1/-2/-3 and HIM-8, required for bivalent and chromatid separation (Phillips and Dernburg 2006) and several mRNAs and/or proteins incorporated in P granules (a form of germ granule composed of ribonucleotide complexes exclusively localised in germline blastomeres; for a review see Updike and Strome 2010) (Voronina 2013; see section 3.4). While the suppression of microRNAs occurs during diplotene silencing to inhibit microRNAs required for proper embryogenesis, the activity of the DNA adaptor proteins and the formation of P granules are part of oocyte maturation (Fig. 3.8 C). Especially, mRNAs and proteins incorporated into P granules represent a set of maternal supplied gene products for embryogenesis (Voronina 2013). I showed that crucial constitutive components of the P granules, such as GLH-4, PGL-1, -3 and DPES-1 (Voronina 2013) are specific for clade V nematodes (Fig. 3.8 C). GLH-4, PGL-1, -3 and DPES-1 are required for formation of P granules and the germline during embryogenesis of *C. elegans* (Spike et al. 2008a,b), but are absent in other Chromadorean and Enoplean nematodes. This shows that P-granule composition has to be different in Enoplea and maybe other mechanisms are at work to assure proper germcell differentiation. The dead-box helicase GLH-4, which is homologous to the *Drosophila* germline gene Vasa, is specifc for clade V nematodes. Interestingly, another Vasa homologue, GLH-1 was detected to be absent in Enoplean nematodes (Fig. 3.8). In total, *C. elegans* has four Vasa helicase homologues, where GLH-2/-3 (also components of P granules) have orthologous sequences in all nematodes (data not shown). It might be possible that a true *Drosophila* Vasa orthologue exists in Enoplean species, but I could not proof its exsistence. I found that the protein Faf (Fat facets) which determines the Vasa expression and germline formation in *Drosophila* (Liu et al. 2003), is expressed more than 2-fold higher in *R. culicivorax* virgins than in males (Fig. 3.9). Orthologues of Faf exist

exclusively in Enoplean nematodes (Fig. 3.10). *faf* mRNA could be maternally supplied in *R. culicivorax* oocytes and be relevant for germline formation during early embryogenesis. It would be interesting to analyse expression of the Vasa orthologue during *R. culicivorax* embryogenesis. This could show that germline determination via Faf and Vasa is conserved between Enoplea and insects. In conclusion, important regulators for P-granule formation during oocyte maturation are absent in other Chromadorean species and in Enoplean nematodes, hence other genes must act instead.

In summary, I found several orthologues which are specific to Enoplea and may have been lost in the course of evolution being replaced either in a Chromadorea-specific or in a clade V-specific manner. As Enoplea are considered to be closer to the root of the nematode phylum than Chromadorea, it seems that a different set of regulators exists for the discussed developmental processes in Enoplea. This leaves the exciting question what could be the reason for such dramatic modifications.

4.3 Polarity establishment and axis specification

Polarity establishment during oogenesis or embryogenesis is essential for body axis formation and proper development in all animals. In the case of the nematode model *C. elegans* this step is initiated by the sperm entry establishing the posterior pole of the embryo. This event activates a signaling cascade determining asymmetric cell division and the formation of a larger anterior soma and a smaller posterior germline cell (see Introduction; Gönczy and Rose 2005). How is polarity and axis formation and patterning managed in other nematdoes? Depending on the different clades it has been previously shown that axis formation can be independent of the sperm entry point (Goldstein et al. 1998) and the situation is even more intriguing in parthenogenetic species which lack sperm (Lahl et al. 2006). In order to address these questions, I used two strategies: (1) I searched for orthologues and their expression of *C. elegans* genes active in the early polarity pathway and downstream genes acting in pathways such as Notch and Wnt signaling. (2) I searched for orthologues of genes affecting development in model organisms such as *Drosophila*, *Xenopus*, zebrafish, mouse and humans, which are expressed

exclusively in the Enoplea *R. culicivorax* and the nematomoprh *Gordius sp.* (see Results; section 3.7). By this unbiased approach (see Introduction) I found orthologues for developmental regulators such as Hunchback, NFKB1, Noggin and Chordin/Short gastrulation to be absent in *C. elegans* (section 3.7). These orthologues are known from many species to be involved in axis formation, pattering and/or polarity establishment (Yamamoto and Oelgeschläger 2004, Minakhina and Steward 2006, Hikasa and Sokol 2013). Focusing on these pathways I tried to identify interacting components which are specifically expressed in *R. culicivorax* and *Gordius sp* (see Results; section 3.7). Here, I will discuss my findings and show which pathways are expressed in (1) all nematodes (2) specific for Chromadorea and (3) specific for Enoplea and/or nematomorphs. Then, I will discuss if and how these pathways are involved in polarity establishment and axis formation/patterning. In the end, I will conjecture which pathways may represent parts of the ancestral gene regulatory network involved in the two considered processes.

4.3.1 Conserved planar cell polarity complexes may play a central role in polarity establishment and axis formation in nematodes

From the model *C. elegans* it was shown that asymmetric cell division during the first cleavage of the zygote highly depends on the PAR-3/PAR-6/PKC-3 and the PAR-1/-2 complexes (Gönczy and Rose 2005). While the former is localised at the anterior cortex, the latter one is localised at the posterior cortex of the 1-cell embryo (Gönczy and Rose 2005). Consequently, both complexes inhibit each other and hence form an antero-posterior (a-p) polarity which determines the primary a-p axis and allows the typical soma-germline separation (Gönczy and Rose 2005). But this classical image of polarity formation is challenged by an increasing complexity as recently the Lethal giant larvae (Lgl) orthologue LGL-1 was identified and was shown to act in a redundant fashion to PAR-2 establishing polarity in *C. elegans* (Beatty et al. 2010, 2013). The orthologous tumor supressor Lgl is also known from other model organisms to initiate polarity establishment, especially in the planar cell polarity pathway (Knust and Bossinger 2002). Here it physically interacts with the proteins Dlg (Discs large) and Scrib (Scribbled), but also with other planar cell polarity (PCP) complexes, such as the Crumbs complex and the mentioned PAR-3/PAR-6/PKC-3 complex. Hence, I investigated for the *C. elegans* orthologues of the PAR-3/PAR-6/PKC-3 complex (see Results; section 3.14), the

Lgl/Dlg/Scrib and Crumbs complexes (section 3.15) whether they are expressed early in all six investigated nematodes and a single nematomorph species.

As expected the PAR-3/PAR-6/PKC-3 is conserved among all nematodes (Fig. 3.26) and I could identify expressed orthologues during early embryogenesis (Fig. 3.27). Interestingly, for PAR-2 I could not detect any orthologues outside the genus *Caenorhabditis* (Fig. 3.26) which is in concordance with earlier studies (Schiffer et al. 2013). Furthermore, I could identify several downstream proteins for polarity establishment in *C. elegans* which were also restricted to the genus *Caenorhabditis* (Fig. 3.27). Examples involve proteins required for asymmetric cell division during the 1-cell stage in *C. elegans* such as LET-99, GPR-1/2, LIN-5, SPD-2, SPD-5 and AIR-1 (Gönczy and Rose 2005; Nguyen-Ngoc et al. 2007; Fig. 3.26, 3.27). Consequently, it seems as if the machinery of inital polarity establishment is not only conserved among nematodes and is also expressed during early embryogenesis (Fig. 3.27), but that the downstream components which act in axis formation and asymmetric cell division are different. Yet, it was shown for most studied nematodes (with few exception such as *Enoplus brevis*; Voronov and Panchin 1998) that the first cleavage is asymmetric. Hence, the way how polarity is established in nematodes outside the genus *Caenorhabditis* and the machinery controlling asymmetric cell division must be different to a certain degree. In that respect I found conservation of Lgl and its expression in concordance with previous findings in *C. elegans* (Beatty et al. 2010, 2013), but I also found that Lgl and all interacting proteins of the same complex (1) have orthologues in all examined genomes (Fig. 3.28) and (2) are expressed during early embryogenesis (Fig. 3.29).

I could also detect orthologues of Crumbs and Sdt (Stardust), the core components of the Crumbs complex (Knust and Bossinger 2002), which are expressed during early embryogenesis in all, or nearly all nematodes (in a few species there was no expression detected which however is probably a false-negative result rather than actual absence of these orthologues). All three complexes (PAR-3/PAR-6/PKC-3, Lgl/Dlg/Scrib and Crumbs/Sdt) are known to interact especially in the PCP pathway to form e.g. basoapical polarity in epithelia of diverse animals (Knust and Bossinger 2002). Early expression of these complexes is fascinating, as they may theoretically constitute an alternative way of establishing polarity during early embryogenesis in nematodes. This is would be different from other animals where these PCP complexes are only known to play a critical role in epithelia formation during later embryogenesis or in the case of *Drosophila* during gastrulation (Bachmann et al. 2001, Hong et al. 2001). I also identified orthologues for

Numb which is known from *Drosophila* to be asymmetrically localised by the PAR-3/-6/PKC-3 and Lgl/Dlg/Scrib complexes to induce asymmetric cell divisions (Rhyu et al. 1994, Betschinger et al. 2003, Haenfler et al. 2012). Numb acts as a fate determinant downstream of both complexes in *Drosophila* (Betschinger et al. 2003, Haenfler et al. 2012) and may act instead of other, *Caenorhabditis* specific, downstream components in all other nematodes. Furthermore, I also found orthologues for other genes known to act downstream to the PCP complexes such as Ssu72 and Sema5c (Fig. 3.29). At this point the function of any of these genes during early embryogenesis is not clear in nematodes (including *C. elegans*). It needs to be investigated whether they are part of an alternative way to establish polarity and fate determination during the 1-cell stage.
The presented results show that complexes known from PCP in *Drosophila* are expressed in representatives of Chromadorea and Enoplea (and the nematomorph *Gordius sp.*) during early embryogenesis. Whether and how these complexes interact during early embryogenesis and to what extent they are involved in the establishment of polarity still needs to be determined. Experiments in *C. elegans* showed that in addition to the PAR-3/-6/PKC-3 complex also Lgl is involved in polarity establishment (Beatty et al. 2010, 2013), indicating that PCP complexes may contribute to polarity establishment also in all other nematodes. In contrast to that many downstream components of PCP complexes are *Caenorhabditis*-specific (see Results; section 3.14). The presented expression analysis indicates alternative candidates for downstream components in other Chromadorea and Enoplea (section 3.15; Fig. 3.29).
In conclusion, the presented expression analysis indicates that PCP complexes are expressed in Chromadorea and Enoplea during early embryogenesis. Downstream signaling instead must be different in other Chromadorea and Enoplea compared to the genus *Caenorhabditis*. Nonetheless, PCP complexes as polarity-inducing cues during early embryogenesis seem to be specific to all nematodes and the nearest outgroup, the nematomorphs, and may thus be the ancestral way of generating polarity in nematodes.

4.3.2 Wnt, BMP/DPP and NF-κB signaling may specify axes in Enoplean nematodes

Polarity establishment is a prerequisite for axis specification (Roth and Lynch 2009, Gönczy and Rose 2005). I stated in the previous section that polarity establishment

differs among animals and needs a triggering event prior, during or after fertilisation (see previous section). Does axis specificiation also differ among representatives of nematodes? From animal (model) organisms it is known that axis specification also varies with respect to different pathways which coordinate this process. The best elucidated pathways which act in vertebrate and invertebrate systems are Wnt, BMP/DPP and NF-κB signaling (Yamamoto and Oelgeschläger 2004, Minakhina and Steward 2006, Hikasa and Sokol 2013). I addressed the question whether (1) any of these pathways may be involved in polarity establishment and axis formation in nematodes and (2) differences exist between Enoplea, Chromadorea and the *Caenorhabditis* genus (see Results).

For Wnt and BMP/DPP signaling it is known that these two interacting pathways are involved in dorsoventral axis specification in invertebrates and vertebrates (Yamamoto and Oelgeschläger 2004, Minakhina and Steward 2006, Hikasa and Sokol 2013) and also influence the posterior patterning of the embryo (Merrill et al. 2004, Satoh et al. 2004, Martin and Kimelman 2009, Petersen and Reddien 2009). Is it possible that a similar control takes place in Enoplea? Why would the genus *Caenorhabditis* be different from other Chromadorea and Enoplea? First of all, BMP/DPP signaling is only required post-embryonically in *C. elegans* (Estevez et al. 1993, Foehr and Liu 2008, Tian et al. 2010, Gumienny and Savage-Dunn 2013), while Wnt signaling is used by the P2-cell in *C. elgans* to induce EMS fate (Mizumoto and Sawa 2007). But is Wnt signaling in all other nematodes the same as in *C. elegans*? In the case of *C. elegans* a special kind of Wnt signaling pathway acts to induce fates during early embryogenesis. The so-called "Wnt asymmetry pathway" acts via competition of the β-catenin paralogues WRM-1 and SYS-1 on the TCF/LEF transcription factor POP-1 (Mizumoto and Sawa 2007). Such signaling cannot work in Enoplean nematodes; I showed that these nematodes do not have the β-catenin paralogues SYS-1 and WRM-1. Instead, my data indicate that Enoplean nematodes have β-catenin orthologues which cluster together with human β-catenin and the orthologous Armadillo from *Drosophila* (Fig. 3.19). Hence, Wnt, as well as BMP/DPP signaling, must be different in Enoplean nematodes.

How is signaling via these pathways different in other Chromadorea and Enoplea? For the Chromadorea *D. coronatus* and the nematomorph *Gordius sp.* I could show that the alternative Wnt receptor RYK is expressed. As orthologues of this receptor exist in other nematodes as well, even in the genus *Caenorhabditis*, I cannot preclude that expression also occurs early in other nematodes. However, RYK is definitely not early expressed in *C. elegans*, since its transcriptome is considered to be complete (see Materials and

Methods; Results section 3.6). From vertebrate models it is known that RYK mediates non-canonical Wnt signaling (Lin et al. 2010, Hikasa and Sokol 2013). In concordance with this I identified expression of Dachsous (DCHS1), Fat3 and Fat4 all of which are known to act in a PCP pathway (Fat/Dachsous pathway) in epithelia of vertebrates and invertebrates (Viktorinová et al. 2009). This Fat/Dachsous pathway is associated with Wnt signaling (Rodríguez 2004, Zecca and Struhl 2010) and binds directly to RYK (Fig. 3.21). Especially for *Drosophila* it is known that these genes act in epithelia of the somatic gonad during oogenesis (Viktorinová et al. 2009). In the case of *Drosophila* it known that oogenesis determines the prospective anteroposterior and dorsoventral axis before these axes are formed during embryogenesis (Roth and Lynch 2009). Maybe Wnt and Fat/Dachsous signaling contributes to axis specification in non-*Caenrhabditis* species. Taken together, the findings that several β-catenin paralogues exist in *C. elegans* (see above), strongly suggest that even though Wnt signaling is pivotal for early embryogenesis in nematodes, it must be definitely different in nematodes outside the *Caenorhabditis* genus.

What about BMP/DPP signaling? From vertebrate models it is known that Wnt signaling and the interplay between Wnt and BMP/DPP signaling controls dorsoventral axis formation and even posterior cell fate determination (Hikasa and Sokol 2013). It was shown that certain BMP/DPP antagonists are important targets of Wnt signaling in vertebrate systems (Smith and Harland 1992, Sasai et al. 1994). Here, I showed that the *R. culicivorax* orthologue of Chordin (one of the BMP/DPP inhibitors) is expressed during early embryogenesis (Fig. 3.18, 3.19). Even though the expression domain of *chordin* was ubiquitous during early embryogenesis of *R. culicivorax* it confirms the data of my transcriptome analysis. One has to bear in mind, however, that in vertebrates Chordin and other inhibitors such as Noggin are secreted from organising centers such as the Mangold-Spemann organiser in *Xenopus* or the shield in zebrafish (Piccolo et al. 1996, Srinivasan et al. 2002). If anything analogous to an "organiser" would exists in nematodes, the best candidate is the P-cell. The *R. culicivorax* expression of *Chordin* does not fit to this view, as expression is not limited to the P-cell linage. However, it must be taken into account that the detected expression might be a consequence of the maternally expressed mRNA and the protein distribution may be spatially limited in the embryo to a certain area. Furthermore, I could show in *R. culicivorax* that Wnt and BMP/DPP signaling exist. Furthermore, the expressed key factors of BMP/DPP signaling show that formation of metalloprotease-inhibitor-DPP complexes as known from

invertebrate and vertebrate systems are possible (Piccolo et al. 1997, Marqués et al. 1997, Scott et al. 1999, Shimmi and O'Connor 2003), as different BMP/DPP inhibitors (Chordin, Noggin, Crossveinless-2, Twisted gastruation) and metalloproteases (BMP1 and Tolloid) are expressed in Enoplean nematodes (see Results; section 3.8; Fig. 3.15). My new findings open the possibility that interaction between Wnt and BMP/DPP signaling may be important for axis specification and posterior patterning in Enoplea during embryogenesis in contrast to Chromadorea.

At which embryonic stage would such a potential interaction between Wnt and BMP/DPP signaling be important for embryogenesis? Well, from vertebrate system it is known that the interaction of both pathways is pivotal for the dorsoventral axis specification prior to gastrulation and for posterior cell fate determination after gastrulation (Hikasa and Sokol 2013). Why would I detect then the expression of Wnt and BMP/DPP signaling components during early embryogenesis? It is possible that the detected mRNAs are of maternal origin which are important only during later stages, such as gastrulation. Hence, my transcriptome data not only reveals first insight into the early polarity establishment in the investigated species (see section 4.3.1) but also into maternally supplied gene products which may play a role during subsequent critical embryonic processes such as dorsoventral axis specification and posterior patterning at later stages.

How does NF-κB signaling fit into this picture? In many aspects NF-κB signaling is correlated with processes such as proliferation, migration of cells and immunity responses (Gilmore 2006). Only in insects, such as *Drosophila* it was shown that this particular pathway is important for dorsoventral axis specification (Minakhina and Steward 2006). Could that mean that, if I detect expression of NF-κB and the obligatory inhibitor IκB in Enoplean nematodes that this signaling pathway has a similar function as in *Drosophila*? Obviously not, as NF-κB signaling in this specific form is thought to be unique to insects, and in the taxon of insects representatives exist which depend less on NF-κB signaling than on BMP signaling for dorsoventral axis specification (Özüak et al. 2014). Still, I find components of NF-κB signaling to be expressed during early embryogenesis in the Enoplea *R. culicivorax* and in the nematomorph *Gordius sp.*. In contrast to this I could not detect expression of orthologues of the accompaning inhibtor IκB in any nematode and nematomorph genomes, indicating its absence in both phyla. However, in *C. elegans* exists an homologue of IκB, named IKB-1, which has no role during embryogenesis in *C. elegans* (Pujol et al. 2001). For all other nematodes and the nematomorph *Gordius*

sp. I could not detect IKB-1 expression. This makes it unlikely that canonical NF-κB plays a role during early embryogenesis in any investigated species. This does, however, not exclude the possibility that non-canonical NF-κB signaling may play a role (for a review on non-canonical NF-κB signaling see Gilmore 2006 and Brasier 2006). Whether and how NF-κB signaling is important to early embryogenesis in nematodes remains to be investigated.

The presented examples show that the molecular circuity must be different in non-*Caenorhabditis* nematodes and may even resemble fate determination via morphogen secretion (Wolpert 1996), rather than cell-fate inductions via direct cell-cell interaction, as known from the model *C. elegans*. It is difficult to imagine a morphogen gradient in embryos which consist of little as 4 cells during early embryogenesis, but at later stages, e.g. gastrulation such a scenario seems more plausible. Maybe secreted factors (or morphogens) may even be secreted by organising centers at certain stages. In this respect the Enoplean seawater nematodes such as *Enoplus brevis* is very interesting as it shows equal divisions up to gastrulation rather than formation of a somatic AB and a germline P cell (Voronov and Panchin 1998). Maybe this nematode resembles the ur-nematode which I imagine to have late cell-fate determination and organising centers similar to vertebrates. But why are all other nematodes I investigated so different from *E. brevis*, showing asymmetric cell divisions during early embryogenesis? Maybe colonisation of land by nematodes generated an evolutionary bottle-neck which led consequently to the loss of certain particularities, such as essential components of Wnt, BMP/DPP and NF-κB signaling.

4.3.3 The gap gene *hunchback* as first insight into anterior patterning of Enoplean embryos

For *Drosophila* it was shown that Hunchback signaling is essential for anterior patterning, head development and neurogenesis during embryogenesis (for review of all mentioned processes refer to Niessing et al. 1997 and Li et al. 2013). By determining the expression of orthologues for Hunchback and interacting proteins, I found only few to be present in nematodes (Fig. 3.30, 3.31). Important gap genes, such as Kr (Krüppel) or Kni (Knirps), or upstream acting genes like Bcd (Bicoid), which are known to interact with Hunchback in *Drosophila* (Jaeger 2011), could not be identfied in any nematode genome (Fig. 3.30).

Nevertheless, Hunchback itself was found to be expressed specifically in the Enoplean nematode *R. culicivorax* and the nematomorph *Gordius sp.* (Fig. 3.31). This is also supported by the expression of *hunchback* mRNA during early embryogenesis in *R. culiciorax* by *in-situ* hybridisation (Fig. 3.32). These results show that (1) Hunchback exists and is expressed during early development exclusively in Enoplea and nematomorphs and (2) Hunchback signaling cannot be mediated as known from *Drosophila* in Enoplea and nematomorphs.

What could then be the role of Hunchback signaling during early embryogenesis of *R. culicivorax*, *Gordius sp.* and possibly all Enoplean nematodes? *In-situ* hybridisation of *R. culicivorax hunchback* showed stronger expression in the somatic precursor cell S1 and its descendants during 2- to 24-cell stage (Fig. 3.32 A, B, C). At later stages the expression decreases dramatically (Fig. 3.32 D), but a much broader expression domain in the head region of the embryo becomes visible during late morphogenesis (Fig. 3.32 F). As Hunchback is known to be involved in head development and neurogenesis (Niessing et al. 1997, Li et al. 2013), these results idicate a similar function in *R. culicivorax*, even though the controlling GRN must be different.

As Hunchback is absent in Chromadrea, Hunchback signaling might indicate an alternative form of anterior patterning and neurogenesis in Enoplea. This particularity is interesting, as only descendants of the somatic cell S1 show *Chordin* expression. Previously, it was shown that descendants of S1 contributed as one linage to neurons (monoclonal; Schulze and Schierenberg 2009). This is different to *C. elegans* where neurogenesis is based on several cell linages (polyclonal; Sulston et al. 1983). As *hunchback* expression coincides with the S1 cell linage in *R. culicivorax* descendants (Fig. 3.32), this suggests functional conservation between the *Drosophila* and *R. culicivorax* orthologues, contrasting neurogenesis in *C. elegans*. This may represent the ancestral form of anterior patterning and neurogenesis in nematodes.

4.4 Do modifiers of mTOR signaling facilitate polarity establishment in parthenogenetic nematodes?

In my analysis of early expression I included the pathenogenotes *Panagrolaimus sp. PS1159* and *D. coronatus*. By searching for orthologues which are exclusively expressed

in both parthenogenotes during early development, I identified the genes LAMTOR4 and LAMTOR5 (see Results; section 3.17). The respective proteins were recently identified as modulators of the conserved mTOR pathway (Bar-Peled et al. 2012), which is known to control primarily growth and proliferation of cells in tissues. Why are these genes expressed exclusively in parthenogenotes in my analysis?

LAMTOR4 and LAMTOR5 were found to be a part of a higher molecular protein complex the so-called "Regulator" which additionally contains LAMTOR1, -2, and -3 (Bar-Peled et al. 2012). This protein complex has a critical role in binding the mTOR complex (mTORC) to the lysosome in *Drosophila* and mammalian cell cultures to facilitate mTOR signaling (Bar-Peled et al. 2012). In this respect it is important to mention that Regulator has a function as a scaffold to the membrane and modifier of downstream targets. If LAMTOR4 and LAMTOR5 were indeed involved in initiation of proliferation and polarity establishment in the investigated parthenogenotes, this would mean that the latter would be mediated via a so far undescribed pathway. Presently, it is impossible to correlate LAMTOR4/5 to any specific function in nematode embryos. However, as these genes are exclusively expressed in the two only distantly related parthogenetic species, it appears likely that these are somehow related to the mode of reproduction. It would be attractive to ectopically overexpress either LAMTOR4, or LAMTOR5 or even the whole protein complex Regulator in *C. elegans* in order to test whether this way first changes towards a parthenogenetic species can be initiated.

4.5 Conclusion: Enoplean oogenesis and early embryogenesis appear to be more similar to outgroups than to Chromadorea

In summary, my unbiased comparison of existing and expressed orthologues during oogenesis and early embryogenesis (see Introduction) revealed that many key-regulators known from *C. elegans* are absent in non-*Caenorhabditis* species. In contrast to that, I find many examples of regulators known from outgroup models to be expressed during both processes exclusively in Enoplea and nematomorphs. The presented analysis suggests that Enoplea share a core set of regulators with outgroup species. Here I propose a

core-set of developmental regulators for both processes which have functions in Enoplea and/or Chromadorea.

For oogenesis I could show that crucial signaling components known from *C. elegans*, are absent in other Chromadorea and Enoplean. Instead orthologues known from *Drosophila*, such as Furry, Fat2/Kugelei and Faf (Fat facetes) are specifically expressed during oogenesis in the Enoplean *R. culicivorax* and may represent parts of a GRN which originally controlled oogenesis in nematodes. As Furry and Fat2/Kugelei were shown to be part of Fat/Dachsous PCP pathway (Viktorinová et al. 2009, Horne-Badovinac et al. 2012) it appears as if PCP has a role in the germline of Enoplean species. The exact function of the Fat/Dachsous pathway in oogenesis of Enoplea remains to be determined.

For early embryogenesis my transcriptome analysis of six nematode species and a single nematomorph (Fig. 1.1), showed new findings with respect to the biological processes of polarity establishment, axis specification in Enoplean nematodes (see section 4.3). My transcriptome data suggest that polarity establishment in all nematodes may rely on PCP complexes such as PAR-3/-6/PKC-3, Lgl/Dlg/Scrib and Crumbs/Stardust. This view is further supported by the fact that the *C. elegans* Lgl orthologue was previously shown to be involved in polarity establishment (Beatty et al. 2010, 2013). If it is true that conserved PCP complexes are required for polarity establishment in nematodes, why is there another *Caenorhabdtitis*-specific PAR-2 complex involved in polarity establishment in *C. elegans* (Gönczy and Rose 2005)? Maybe polarity establishment during early embryogenesis originally depended on e.g. the Lgl/Dlg/Scrib complex in other Chromadorea and Enoplea and in the special case of the *Caenorhabditis* genus was replaced by the function of *par-2* and other genes. Nonetheless, these findings support the idea that the mentioned PCP complexes may represent the ancestral form of polarity establishment in the last common ancestor of all nematodes.

I also investigated axis specification during early embryogenesis by analysing the Wnt, BMP/DPP and NF-κB signaling pathways. I found evidence that all three pathways are fundamentally different in Enoplea in comparison to Chromadorea and specifically the genus *Caenorhabditis* (see section 4.3.2). Furthermore, it seems like Wnt and BMP/DPP signaling are more similar in Enoplea to vertebrate species than to Chromadorea, as Wnt signaling can only be mediated via an orthologue of vertebrate β-catenin (section 4.3.2; Fig. 3.19) and BMP signaling can be controlled by antagonists such as Chordin or Noggin which do not even exist in Chromadorea (see section 3.8).

Taken together these findings suggest that Enoplea show many aspects of oogenesis and early embryogenesis which are more similar to outgroups than to Chromadorea, giving strong support for the assumption that Enoplea are more basal in the phylum of nematodes than Chromadorea. This is in concordance with previous considerations where Enoplea were stated to be an "ancestral" taxon within the phylum of nematodes (De Ley 2006).

For future studies it would be interesting to further investigate oogenesis, polarity establishment and axis formation in Enoplea and Chromadorea. In particular, it would be fascinating to knock-out components of the aforementioned pathways and complexes to see if this would have an impact on any of these biological processes. This could be done by RNA interference, or genome editing via the CRISPR/Cas9 system in the Enoplean *R. culicivorax*. Alternatively, it would be interesting to do the same experiments for orthologs in the Chromadorean satellite model *P. pacificus*, where the CRISPR/Cas9 system has been successfully established (Witte et al. 2015).

Another approach would be to clone the sequences from Enoplea-specific genes, such as Chordin, to transform *Drosophila* strains and possibly rescue Sog (Short gasturation) loss-of-function alleles. Such experiments would demonstrate the functional conservation between Enoplean-specific orthologues and outgroup proteins. This would give even further evidence that Enoplea are more similar to outgroup species with respect to oogenesis and early development than to Chromadorea.

Appendix A

Supplementary figures

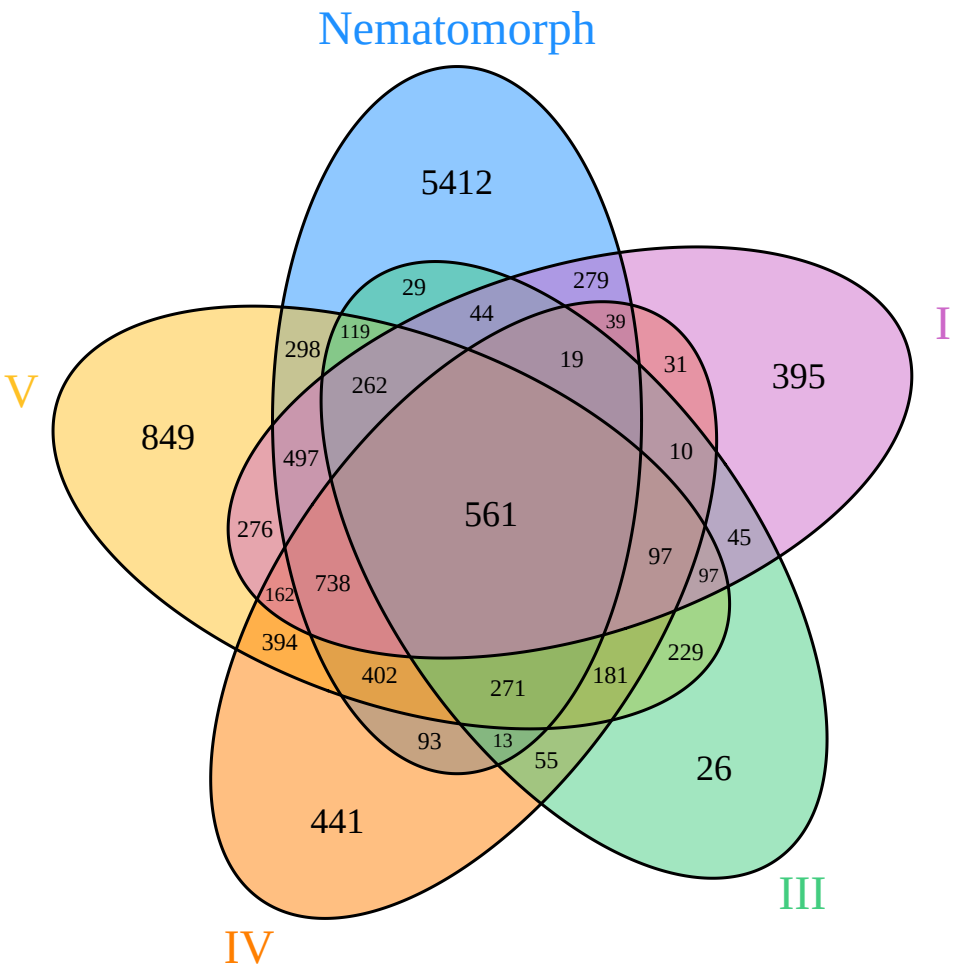

Figure A.1: Venn-diagram representing the orthology clustering of early transcriptomes of the nematodes *R. culicivorax* (clade I), *A. suum* (clade III), *P. sp. PS1159*, *D. coronatus* (clade V), *P. pacificus* (clade V) and *C. elegans*. As an outgroup I inculded the nematomorph *Gordius sp.* Transcripts for each species wre translated into protein sequences via the programme Transdecoder and orthologs were identified via OrthoMCL clustering (materials and methods).

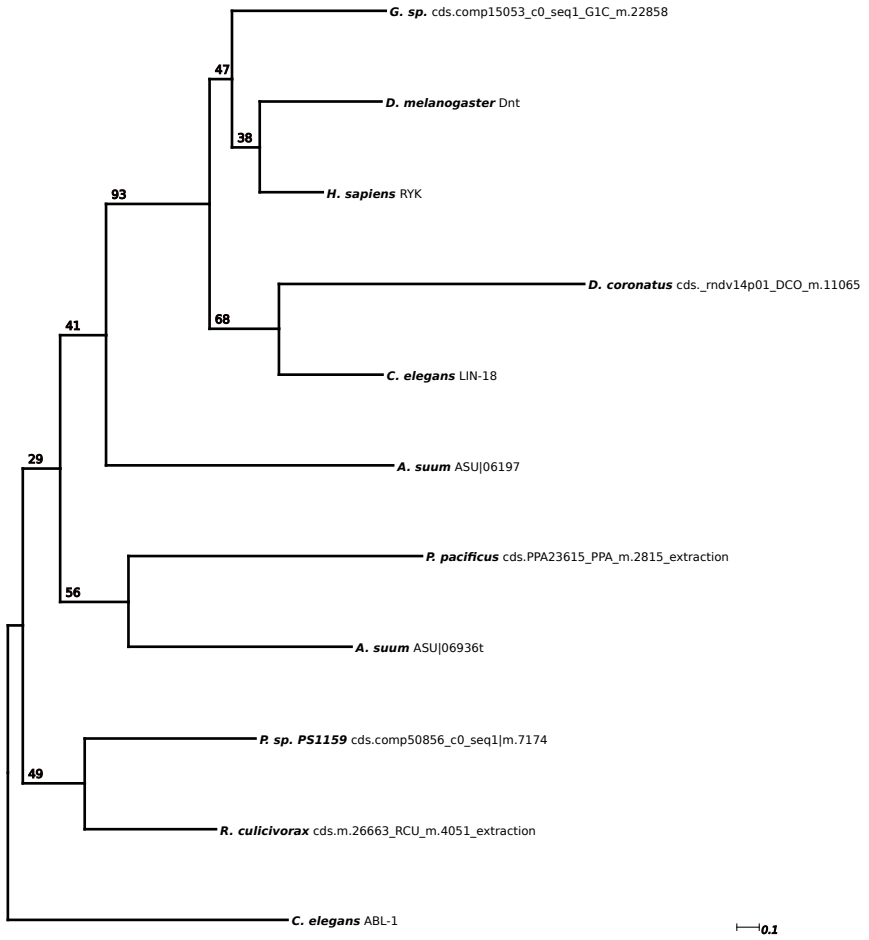

Figure A.2: Phylogentic tree for Human RYK orthologs. The tree was created by maximum likelihood statistics with a subsitution matrix of LG, optimised for Γ and invariable sites with prottest3 after BIC and AIC.

Appendix B

Custom programmes

"search_genes_all.pl" a custom Perl script to access orthologous clusters of 27 species via specific queries:

```
#! /usr/bin/perl

use strict;
use utf8;
use 5.010;

die "USAGE: ./search_genes_all.pl <WB-ID, Flybase-ID, UNIprot-ID> <output file name>\n" unless (@ARGV
== 2);

sub check_qual {
        my $val = shift;
        if ($val =~ />/ || $val =~ /</ || $val =~ /\|/ || $val =~ /\$/ || $val =~ /\^/ || $val =~ / /)
{
                die "error: $val: Symbols as \">, <, \|, \$, \^\" and SPACE are not allowed!\n";
        }
}

for (@ARGV) {
        &check_qual ($_);
}

my @error;
@error = map {
        if (($_ =~ /\AWBGene[0-9]*\b/) || ($_ =~ /\AFBgn[0-9]*\b/) || ($_ =~ /\A\w*\b/)) {
                        0;
                } else {
                        $_;
                }
        } @ARGV;

for (@error) {
        die "error: \"$_\" is not a valid ID!\n" if ($_);
}

my $ID_type = shift @ARGV;
my $file_name = shift @ARGV;

if ($ID_type =~ /\AUNI/) {
        $ID_type =~ s/\AUNI/sp\\\|/;
}

#open headers.txt to get the cluster names
open my $fh, "<", "./headers.txt";

my $ortholog = $ID_type;
while (<$fh>) {
        chomp;
        my $line = $_;
        my @line = split /\t/;
        if ($line =~ /$ortholog/) {
                $ID_type = $line[0];
                $ID_type =~ s/\A>//g;
                last;
        }
}

close $fh;
$ID_type =~ s/\|/\\\|/g;

#open OrthoMCL groups file to get the clusters
open $fh, "<", "./27proteomes.groups.txt" or die "./27proteomes.groups.txt: $!\n";

my %orthologs;

while (<$fh>) {
        chomp;
        my @line = split / /;
        shift @line; #getting rid of the first cluster header
        shift @line; #getting rid of the empty first array element
        foreach (@line) {
                if (/$ID_type/) {
                        if ($orthologs{$_}) {
                                die "error: \"$ID_type\" exists multiple times in the clustering!
\"$ID_type\" might be ambiguous or not a valid ID!\n";
```

```
                        } else {
                                $orthologs{$_} = [ (@line) ];
                        }
                }
        }
}

close $fh;

#get the identifier of the fasta files.
open my $fh, "<", "./headers_index.txt";

my %fasta_names;
while (<$fh>) {
        chomp;
        my @line = split /\t/;
        $line[0] =~ s/\A>//;
        foreach my $query (sort keys %orthologs) {
                foreach my $ortholog (@{ $orthologs{$query} }) {
                        if ($line[0] eq $ortholog) {
                                $fasta_names{$ortholog}=$line[1];
                                last;
                        }
                }

        }
}

close $fh;

my $current_ortholog;
my %ortholog_fh;
open my $ofh, ">", "$file_name.cluster.tsv" or die "$file_name.cluster.tsv: $!\n";
foreach my $cluster (sort keys %orthologs) {
        my %counter;

        #initializing %counter
        $counter{'HSA'}= 0;#outgroups
        $counter{'CCA'}= 0;
        $counter{'SPU'}= 0;
        $counter{'BFL'}= 0;
        $counter{'DME'}= 0;
        $counter{'TCA'}= 0;
        $counter{'DPU'}= 0;
        $counter{'API'}= 0;
        $counter{'SMA'}= 0;
        $counter{'HDU'}= 0;
        $counter{'TUR'}= 0;

        $counter{'GOR'}= 0;#nematomorphs

        $counter{'RCT'}= 0;#cladeI
        $counter{'RCU'}= 0;
        $counter{'TSP'}= 0;

        $counter{'EBR'}= 0;#cladeII

        $counter{'ASU'}= 0;#cladeIII
        $counter{'LOA'}= 0;
        $counter{'BMA'}= 0;
        $counter{'DIM'}= 0;


        $counter{'MHA'}= 0;#cladeIV
        $counter{'BUX'}= 0;


        $counter{'PPA'}= 0;#cladeV
        $counter{'CAN'}= 0;
        $counter{'CBR'}= 0;
        $counter{'CEL'}= 0;
        $counter{'CRE'}= 0;

        foreach my $ortholog (@{ $orthologs{$cluster} }) {
```

```
                my @tag = split /\|/, $ortholog;
                $current_ortholog = $fasta_names{$ortholog};
                unless ($current_ortholog eq ">" ) {
                        if ($ortholog_fh{$tag[0]}) {
                                $ortholog_fh{$tag[0]}->{$current_ortholog} = [ () ];
                        } else {
                                $ortholog_fh{$tag[0]} = { ( $current_ortholog,  [ ] ) };
                        }
                }
                $counter{$tag[0]} = $counter{$tag[0]} + 1;
        }

        print $ofh "HSA\t$counter{'HSA'}\n",
                        "CCA\t$counter{'CCA'}\n",
                        "SPU\t$counter{'SPU'}\n",
                        "BFL\t$counter{'BFL'}\n",
                        "DME\t$counter{'DME'}\n",
                        "TCA\t$counter{'TCA'}\n",
                        "DPU\t$counter{'DPU'}\n",
                        "API\t$counter{'API'}\n",
                        "SMA\t$counter{'SMA'}\n",
                        "HDU\t$counter{'HDU'}\n",
                        "TUR\t$counter{'TUR'}\n",
                        "GOR\t$counter{'GOR'}\n",
                        "RCT\t$counter{'RCT'}\n",
                        "RCU\t$counter{'RCU'}\n",
                        "TSP\t$counter{'TSP'}\n",
                        "EBR\t$counter{'EBR'}\n",
                        "ASU\t$counter{'ASU'}\n",
                        "LOA\t$counter{'LOA'}\n",
                        "BMA\t$counter{'BMA'}\n",
                        "DIM\t$counter{'DIM'}\n",
                        "MHA\t$counter{'MHA'}\n",
                        "BUX\t$counter{'BUX'}\n",
                        "PPA\t$counter{'PPA'}\n",
                        "CAN\t$counter{'CAN'}\n",
                        "CBR\t$counter{'CBR'}\n",
                        "CEL\t$counter{'CEL'}\n",
                        "CRE\t$counter{'CRE'}\n";

        print "HSA\t$counter{'HSA'}\n",
                        "CCA\t$counter{'CCA'}\n",
                        "SPU\t$counter{'SPU'}\n",
                        "BFL\t$counter{'BFL'}\n",
                        "DME\t$counter{'DME'}\n",
                        "TCA\t$counter{'TCA'}\n",
                        "DPU\t$counter{'DPU'}\n",
                        "API\t$counter{'API'}\n",
                        "SMA\t$counter{'SMA'}\n",
                        "HDU\t$counter{'HDU'}\n",
                        "TUR\t$counter{'TUR'}\n",
                        "GOR\t$counter{'GOR'}\n",
                        "RCT\t$counter{'RCT'}\n",
                        "RCU\t$counter{'RCU'}\n",
                        "TSP\t$counter{'TSP'}\n",
                        "EBR\t$counter{'EBR'}\n",
                        "ASU\t$counter{'ASU'}\n",
                        "LOA\t$counter{'LOA'}\n",
                        "BMA\t$counter{'BMA'}\n",
                        "DIM\t$counter{'DIM'}\n",
                        "MHA\t$counter{'MHA'}\n",
                        "BUX\t$counter{'BUX'}\n",
                        "PPA\t$counter{'PPA'}\n",
                        "CAN\t$counter{'CAN'}\n",
                        "CBR\t$counter{'CBR'}\n",
                        "CEL\t$counter{'CEL'}\n",
                        "CRE\t$counter{'CRE'}\n";
}

close $ofh;


foreach my $tag (sort keys %ortholog_fh) {
        open $fh, "<", "./fasta_files/$tag.fasta" or die "./fasta_files/$tag.fasta: $!\n";
```

```
        my $current_seq = "";

        while (<$fh>) {
                chomp;
                if (/\A>/) {
                        my $header;
                        $current_seq = "";
                        if ($tag) {
                                $header = substr($_, 0, 100); #Sequence-header in OrthoMCL clusters
are truncated to 20 symbols!
                                foreach my $seq (sort keys %{ $ortholog_fh{$tag} }) {
                                        if ($header eq $seq) {
                                                $current_seq = $header;
                                        }
                                }
                                #} elsif ($tag eq "CEL") {
                                #$header = substr($_, 0, 61);
                                #foreach my $seq (sort keys %{ $ortholog_fh{$tag} }) {
                                #if ($header =~ /\A"$seq"/) { #Random names in CEL orthoMCL
clustering! REGEX machine necessary!
                                #       $current_seq = $seq;
                                #}
                                #}
                        }

                } elsif (/\A\w/ && $current_seq) {
                        push @{ $ortholog_fh{$tag}->{$current_seq} }, $_;
                }
        }
        close $fh;
}


open $ofh, ">", "$file_name.fasta" or die "$file_name.fasta: $!\n";

foreach my $species (sort keys %ortholog_fh) {
        foreach my $header (sort keys %{ $ortholog_fh{$species} }) {
                print $ofh "\n$header\n";
                foreach my $seq ( @{ $ortholog_fh{$species}->{$header} } ) {
                        print $ofh "$seq\n";
                }
        }
}

close $ofh;
```

"search_transcriptomes.pl" a custom Perl script to access orthologous clusters of transcriptomes of early embryos of seven species:

```
#! /usr/bin/perl

use strict;
use utf8;
use 5.010;

die "USAGE: ./search_transcriptomes.pl <GOR|RCT> <GOR-ID|RCT-ID> <output name>\n" unless (@ARGV == 3);

sub check_qual {
        my $val = shift;
        if ($val =~ />/ || $val =~ /</ || $val =~ /\|/ || $val =~ /\$/ || $val =~ /\^/ || $val =~ / /)
{
                die "error: $val: Symbols as \", >, <, \|, \$, \^\" and SPACE are not allowed!\n";
        }
}

for (@ARGV) {
        &check_qual ($_);
}

my @error;
@error = map {
        if (($_ =~ /\AGOR\b/) || ($_ =~ /\ARCT\b/) || ($_ =~ /\A\w*\b/)) {
                        0;
                } else {
                        $_;
                }
        } @ARGV;

for (@error) {
        die "error: \"$_\" is not a valid species TAG, ID or output name!\n" if ($_);
}

my $species_TAG = shift @ARGV;
my $ID_type = shift @ARGV;
my $file_name = shift @ARGV;


open my $fh, "<", "./headers.txt" or die "./headers.txt: $!\n";
;

my $init_ortholog = "$species_TAG" . "\|" ."$ID_type";
while (<$fh>) {
        chomp;
        my $line = $_;
        my @line = split /\t/;
        if ($line =~ /\Q$init_ortholog\E/) {
                my @tmp = split / /, $line[1];
                $tmp[9] =~ s/:\S*\z//;
                $ID_type = $tmp[9];
                last;
        }
}

close $fh;

#if its GOR open GOR_official_VS_GORC.index.txt to get the cluster names
if ($species_TAG eq "GOR") {
        open $fh, "<", "./GOR_official_VS_GORC.index.txt" or die "./GOR_official_VS_GORC.index.txt: $!
\n";

        my $hit = 0;
        while (<$fh>) {
                chomp;
                my @line = split /\t/;
                if ($ID_type eq $line[1]) {
                        $ID_type = $line[0];
                        $hit = 1;
                        last;
                }

        }
        unless ($hit) {
                die "No early transcript found for the sequence \"$init_ortholog\" with the ID
```

```
\"$ID_type\" !\n";
        }
        $ID_type =~ s/\A/GORC\|cds\./;
        $ID_type =~ s/\z/_GORC_m\./;

} elsif( $species_TAG eq "RCT") {
        open $fh, "<", "./R_culi_1-4cell_VS_r_culicivorat_official.index.txt" or die "./
R_culi_1-4cell_VS_r_culicivorat_official.index.txt: $!\n";

        my $hit = 0;
        while (<$fh>) {
                chomp;
                my @line = split /\t/;
                if ($ID_type eq $line[1]) {
                        $ID_type = $line[0];
                        $hit = 1;
                        last;
                }
        }
        unless ($hit) {
                die "No early transcript found for the sequence \"$init_ortholog\" with the ID
\"$ID_type\" !\n";
        }
        $ID_type =~ s/\A/RCU|cds\./;
        $ID_type =~ s/\z/_RCU_m\./;
}

#open OrthoMCL groups file to get the clusters
open $fh, "<", "./GORC_PS11_CEL_PPA_ASU_RCU_G1C_DCO_goodProteins.groups.txt" or die "./
GORC_PS11_CEL_PPA_ASU_RCU_G1C_DCO_goodProteins.groups.txt: $!\n";

my %orthologs;

while (<$fh>) {
        chomp;
        my @line = split / /;
        shift @line; #getting rid of the first cluster header
        foreach (@line) {
                if (/\Q$ID_type\E/) {
                        if ($orthologs{$_}) {
                                die "error: \"$ID_type\" exists multiple times in the clustering!
\"$ID_type\" might be ambiguous or not a valid ID!\n";
                        } else {
                                $orthologs{$_} = [ (@line) ];
                        }
                }
        }
}

close $fh;

#get the identifier of the fasta files.

my %fasta_names;
foreach my $query (sort keys %orthologs) {
        foreach my $ortholog (@{ $orthologs{$query} }) {
                $fasta_names{$ortholog}=$ortholog;
        }
}

close $fh;

my $current_ortholog;
my %ortholog_fh;
open my $ofh, ">", "$file_name.cluster.tsv" or die "$file_name.cluster.tsv: $!\n";
foreach my $cluster (sort keys %orthologs) {
        my %counter;

        #initializing %counter
        $counter{'GORC'}= 0;
        $counter{'G1C'}= 0;
        $counter{'PS11'}= 0;
        $counter{'CEL'}= 0;
        $counter{'PPA'}= 0;
```

```
        $counter{'ASU'}= 0;
        $counter{'RCU'}= 0;
        $counter{'DCO'}= 0;

        foreach my $ortholog (@{ $orthologs{$cluster} }) {
                my @tag = split /\|/, $ortholog;
                $current_ortholog = $fasta_names{$ortholog};
                unless ($current_ortholog eq ">" ) {
                        if ($ortholog_fh{$tag[0]}) {
                                $ortholog_fh{$tag[0]}->{$current_ortholog} = [ () ];
                        } else {
                                $ortholog_fh{$tag[0]} = { ( $current_ortholog,  [ ] ) };
                        }
                }
                $counter{$tag[0]} = $counter{$tag[0]} + 1;
        }

        print $ofh "GORC\t$counter{'GORC'}\n",
                        "G1C\t$counter{'G1C'}\n",
                        "RCU\t$counter{'RCU'}\n",
                        "ASU\t$counter{'ASU'}\n",
                        "PS11\t$counter{'PS11'}\n",
                        "DCO\t$counter{'DCO'}\n",
                        "PPA\t$counter{'PPA'}\n",
                        "CEL\t$counter{'CEL'}\n";

        print "GORC\t$counter{'GORC'}\n",
                        "G1C\t$counter{'G1C'}\n",
                        "RCU\t$counter{'RCU'}\n",
                        "ASU\t$counter{'ASU'}\n",
                        "PS11\t$counter{'PS11'}\n",
                        "DCO\t$counter{'DCO'}\n",
                        "PPA\t$counter{'PPA'}\n",
                        "CEL\t$counter{'CEL'}\n";

}

close $ofh;


foreach my $tag (sort keys %ortholog_fh) {
        open $fh, "<", "./compliantFasta/$tag.fasta" or die "./compilantFasta/$tag.fasta: $!\n";
        my $current_seq = "";

        while (<$fh>) {
                chomp;
                if (/\A>/) {
                        my $header;
                        $current_seq = "";
                        if ($tag) {
                                $header = substr($_, 0, 100); #Sequence-header in OrthoMCL clusters
are truncated to 20 symbols!
                                foreach my $seq (sort keys %{ $ortholog_fh{$tag} }) {
                                        my $complete_seq = ">" . "$seq";
                                        if ($header eq $complete_seq) {
                                                $current_seq = $seq;
                                        }
                                }
                                #} elsif ($tag eq "CEL") {
                                #$header = substr($_, 0, 61);
                                #foreach my $seq (sort keys %{ $ortholog_fh{$tag} }) {
                                #if ($header =~ /\A"$seq"/) { #Random names in CEL orthoMCL
clustering! REGEX machine necessary!
                                #       $current_seq = $seq;
                                #}
                                #}
                        }

                } elsif (/\A\w/ && $current_seq) {
                        push @{ $ortholog_fh{$tag}->{$current_seq} }, $_;
                }
        }
        close $fh;
}
```

```
open $ofh, ">", "$file_name.fasta" or die "$file_name.fasta: $!\n";

foreach my $species (sort keys %ortholog_fh) {
        foreach my $header (sort keys %{ $ortholog_fh{$species} }) {
                print $ofh "\n>$header\n";
                foreach my $seq ( @{ $ortholog_fh{$species}->{$header} } ) {
                        print $ofh "$seq\n";
                }
        }
}

close $ofh;
```

"extract_clusters.pl" a custom Perl script to extract protein sequences from orthologous clusters shared and/or excluded from transcriptomes of early embryos:

```
#! /usr/bin/perl -w

use utf8;
use 5.010;
use strict;

die "Please insert extract_clusters.pl <species TAGs, separated by \"-\"> <exclude TAGs, separated by
\"-\"> <groups-file> <clusteroutput-file>\n" unless (@ARGV == 4);

my $species_TAG = shift @ARGV;

my @species;
unless ($species_TAG eq "-") {
        @species = split /-/, $species_TAG;
} else {
        $species[0] = $species_TAG;
}

my $species_hits = (@species);
my $exclude_TAGs = shift @ARGV;

my @exclude;
unless ($exclude_TAGs eq "-") {
        @exclude = split /-/, $exclude_TAGs;
} else {
        @exclude = ();
}


my $orthomclgroupsname = shift @ARGV;
my $output_spec = shift @ARGV;
open my $OUTPUT_cluster, ">", $output_spec or die $!;
open my $oMCL_groups, "<", $orthomclgroupsname or die $!;

my $cluster_no = 0;


#read in the data line by line
while(<$oMCL_groups>) {
        chomp;
        my $loop_exit = 0;
        my $hit = 0;
#stealing this from georgios intending to split cluster tag from proteins
        my @proteins = split(/\s+/, $_,2);
        foreach my $spec (@species) {
                if ($proteins[1]=~ m/($spec\|)+/) {
                        $hit ++;
                }
        }
        if ($hit == $species_hits) {
                if(@exclude) {
                        foreach my $exclude (@exclude) {
                                if ($proteins[1]=~ m/($exclude\|)+/) {
                                        $loop_exit = 1;
                                        last;
                                }
                        }
                }
                if ($loop_exit) {
                        next;
                } else {
                        ++$cluster_no;
                        my @genes = split / /, $proteins[1];
                        foreach (@genes) {
                                foreach my $spec (@species) {
                                        if ($_ =~ /\A$spec/) {
                                                print $OUTPUT_cluster "$_\n";
                                        }
                                }
                        }
                }
        }
}
print "Number of $species_TAG clusters: $cluster_no\n\n";
```

```
close $oMCL_groups;
close $OUTPUT_cluster;
```

Appendix C

Supplementary tables

Table C.1: Results of RNAi experiments in combination with affymetrix microarray analysis. In total there were 152 signficantly over- and/or underexpressed genes with wormbase entries for RNAi phenotypes such as "egg-laying defect", "vulva protrusion", "larvae hatching defect", or "early embryogenesis defect".

RNAi knock-down	**affymetrix microarray feature**	**target gene**
mek-2	179631_at	srz-66
mek-2	180107_at	srz-60
mek-2	181581_at	srz-32
mek-2	182245_at	grl-8
mek-2	188812_at	srb-11
mek-2	188970_at	cfi-1
mek-2	192765_at	mab-3
mbk-2	177755_at	srg-39
mbk-2	177807_at	srg-40
mbk-2	178976_at	str-263
mbk-2	179261_at	srh-178
mbk-2	182245_at	grl-8
mbk-2	182582_at	srh-173
mbk-2	183047_at	srab-2
mbk-2	184350_s_at	sta-1
mbk-2	185751_at	srbc-33

mbk-2	186055_at	grd-10
mbk-2	187415_at	sdz-1
mbk-2	187894_at	lim-6
mbk-2	188140_at	sra-7
mbk-2	188213_at	ptr-19
mbk-2	189928_at	nhr-219
mbk-2	190664_s_at	cfz-2
mbk-2	191316_at	srj-57
mbk-2	191439_at	srj-38
mbk-2	191476_at	srh-245
mbk-2	191498_at	str-143
mbk-2	191681_at	sox-2
mbk-2	192431_at	nhr-111
mbk-2	193511_at	ceh-24
mbk-2	193759_at	hlh-1
mbk-2	193833_s_at	ceh-43
lip-1	179072_at	srh-118
lip-1	179631_at	srz-66
lip-1	180564_at	grl-13
lip-1	191439_at	srj-38
goa-1	73854_s_at	dop-1
goa-1	174019_s_at	nhr-92
goa-1	174461_at	gei-13
goa-1	175782_s_at	hbl-1
goa-1	175965_at	unc-34
goa-1	176603_at	srg-6
goa-1	177036_at	lad-2
goa-1	177509_at	srh-277
goa-1	178868_at	eak-4
goa-1	180163_at	srw-115
goa-1	180185_at	str-9
goa-1	181042_at	srh-295
goa-1	181049_at	srv-7

goa-1	182610_at	srh-171
goa-1	183032_at	srt-28
goa-1	184474_at	srx-110
goa-1	186331_at	jip-1
goa-1	187594_at	cnd-1
goa-1	188084_s_at	plx-2
goa-1	188093_at	sra-2
goa-1	188187_at	ptr-13
goa-1	188255_at	srd-1
goa-1	189216_at	str-41
goa-1	189293_s_at	dsl-6
goa-1	189669_at	ceh-48
goa-1	190803_at	srh-7
goa-1	191152_at	str-67
goa-1	191279_at	srh-201
goa-1	193507_at	nhr-81
goa-1	194043_at	syd-1
goa-1	194074_at	nhr-255
par-1	174493_at	rab-37
par-1	175757_s_at	fax-1
par-1	181774_at	str-111
par-1	182198_at	inx-19
par-1	183032_at	srt-28
par-1	183968_at	vab-19
par-1	184081_at	srx-16
par-1	184621_at	srx-6
par-1	186284_at	srv-30
par-1	191055_at	str-20
par-1	191132_at	sri-63
par-1	191279_at	srh-201
par-1	191558_at	srj-15
par-1	191586_at	srh-185
par-1	192339_at	pik-1

par-1	192655_s_at	elt-1
par-1	192738_at	unc-7
par-1	193328_at	scl-20
par-1	193507_at	nhr-81
par-1	194086_at	mlt-11
par-1	188166_s_at	lin-40
par-1	188907_at	str-102
par-1	189601_at	srsx-30
par-1	190427_s_at	ftn-1
par-1	190493_at	clec-153
par-1	190586_at	nas-33
par-1	178937_at	srd-45
par-1	179191_at	srh-261
par-1	180164_at	fbxc-16
par-1	182037_at	fbxa-46
par-1	183568_at	srz-56
par-1	184526_at	trp-4
itr-1	171921_x_at	hil-7
itr-1	172530_x_at	srw-141
itr-1	173777_s_at	cng-1
itr-1	174159_at	mnp-1
itr-1	174266_at	nipi-3
itr-1	174406_at	tir-1
itr-1	174461_at	gei-13
itr-1	174603_at	ets-4
itr-1	175782_s_at	hbl-1
itr-1	176778_s_at	rrf-2
itr-1	177054_at	srv-24
itr-1	177258_at	srt-45
itr-1	177336_at	crn-2
itr-1	179445_at	srm-4
itr-1	179677_at	srh-257
itr-1	179837_s_at	cdh-3

itr-1	179865_at	sdz-21
itr-1	181070_at	max-2
itr-1	181230_at	zip-2
itr-1	181780_at	ckk-1
itr-1	182198_at	inx-19
itr-1	182257_at	lam-3
itr-1	182750_at	srbc-17
itr-1	183109_at	srg-25
itr-1	183143_at	srg-67
itr-1	183305_at	srx-7
itr-1	184503_at	srw-66
itr-1	184523_at	srw-97
itr-1	184621_at	srx-6
itr-1	184642_at	spe-38
itr-1	185994_at	srt-8
itr-1	186821_at	srw-38
itr-1	187082_s_at	mlk-1
itr-1	187415_at	sdz-1
itr-1	187962_at	ins-35
itr-1	187985_at	skr-3
itr-1	188042_at	sra-4
itr-1	188228_at	skr-4
itr-1	188340_at	str-32
itr-1	188538_at	ins-7
itr-1	188972_at	str-47
itr-1	189293_s_at	dsl-6
itr-1	190874_at	sprr-1
itr-1	191132_at	sri-63
itr-1	191145_at	aptf-1
itr-1	191249_at	str-220
itr-1	191399_s_at	ztf-11
itr-1	191721_at	ceh-17
itr-1	192272_at	egl-46

itr-1	193148_at	vhp-1
itr-1	193507_at	nhr-81
itr-1	193765_at	eps-8
mpk-1	177397_at	srg-32
mpk-1	178468_at	srd-70
mpk-1	179242_at	srx-44
mpk-1	180585_s_at	srbc-76
mpk-1	182245_at	grl-8
goa-1	172991_at	ptr-4
mpk-1	183032_at	srt-28
mpk-1	187869_s_at	lgc-33
mpk-1	188213_at	ptr-19
mpk-1	189136_at	cog-1
mpk-1	190124_at	sra-17
mpk-1	191481_at	srh-192
itr-1	174266_at	nipi-3
itr-1	175782_s_at	hbl-1
itr-1	184503_at	srw-66
itr-1	188084_s_at	plx-2
itr-1	189293_s_at	dsl-6
itr-1	192254_at	acr-9
itr-1	192254_at	acr-9
par-1	175965_at	unc-34
par-1	175965_at	unc-34
par-1	183032_at	srt-28
par-1	187594_at	cnd-1
par-1	193507_at	nhr-81
par-1	176778_s_at	rrf-2
par-1	182198_at	inx-19
par-1	184621_at	srx-6
par-1	187153_s_at	clp-4
par-1	190141_s_at	tba-7
par-1	191132_at	sri-63

par-1	193507_at	nhr-81
mbk-2	174159_at	mnp-1
mbk-2	179846_at	srg-20
mbk-2	187415_at	sdz-1
mbk-2	193833_s_at	ceh-43
mek-2	182245_at	grl-8
mbk-2	182245_at	grl-8
mbk-2	188213_at	ptr-19
mbk-2	190124_at	sra-17
lip-1	191439_at	srj-38

Table C.2: Significantly enriched GO-terms for *Gordius sp.* 1–cell stage.

GO-ID	GO-term	category	FDR	p-value
GO:0005829	cytosol	C	5.018066E-7	4.072692E-10
GO:0097190	apoptotic signaling pathway	P	1.143993E-3	7.564642E-6
GO:0043066	negative regulation of apoptotic process	P	1.261968E-3	8.705876E-6
GO:0005739	mitochondrion	C	1.38367E-3	1.010695E-5
GO:0051052	regulation of DNA metabolic process	P	1.65268E-3	1.324558E-5
GO:0051049	regulation of transport	P	2.033671E-3	1.733067E-5
GO:0006397	mRNA processing	P	2.175984E-3	1.898495E-5
GO:0000075	cell cycle checkpoint	P	2.284225E-3	2.061994E-5
GO:0001700	embryonic development via the syncytial blastoderm	P	2.911528E-3	2.806078E-5
GO:0009887	organ morphogenesis	P	2.919436E-3	2.843318E-5
GO:0019003	GDP binding	F	4.649092E-3	5.707011E-5
GO:0031328	positive regulation of cellular biosynthetic process	P	5.154205E-3	6.745383E-5
GO:0019901	protein kinase binding	F	5.504285E-3	7.371062E-5

GO:0045859	regulation of protein kinase activity	P	5.553645E-3	7.549847E-5
GO:1901701	cellular response to oxygen-containing compound	P	5.850997E-3	8.139329E-5
GO:0060341	regulation of cellular localization	P	6.387832E-3	9.012267E-5
GO:0051219	phosphoprotein binding	F	7.928794E-3	1.211373E-4
GO:0010948	negative regulation of cell cycle process	P	9.173606E-3	1.43323E-4
GO:0048167	regulation of synaptic plasticity	P	9.373166E-3	1.483427E-4
GO:0008285	negative regulation of cell proliferation	P	9.620389E-3	1.547038E-4
GO:0008015	blood circulation	P	9.620389E-3	1.573769E-4
GO:0050770	regulation of axonogenesis	P	9.716347E-3	1.617638E-4
GO:0048749	compound eye development	P	9.716347E-3	1.626456E-4
GO:0044765	single-organism transport	P	1.047432E-2	1.785215E-4
GO:0044463	cell projection part	C	1.048623E-2	1.797883E-4
GO:0048589	developmental growth	P	1.13428E-2	2.006683E-4
GO:0003729	mRNA binding	F	1.13428E-2	2.030165E-4

GO:0031400	negative regulation of protein modification process	P	1.13428E-2	2.036801E-4
GO:0010648	negative regulation of cell communication	P	1.138E-2	2.055026E-4
GO:0080134	regulation of response to stress	P	1.257845E-2	2.32249E-4
GO:0045087	innate immune response	P	1.262302E-2	2.351116E-4
GO:0014902	myotube differentiation	P	1.354414E-2	2.570833E-4
GO:0001653	peptide receptor activity	F	1.383455E-2	2.666697E-4
GO:0005730	nucleolus	C	1.436057E-2	2.797229E-4
GO:0034702	ion channel complex	C	1.439445E-2	2.868821E-4
GO:0060491	regulation of cell projection assembly	P	1.496156E-2	3.005366E-4
GO:0001558	regulation of cell growth	P	1.578926E-2	3.235701E-4
GO:0004930	G-protein coupled receptor activity	F	1.630554E-2	3.421953E-4
GO:0001817	regulation of cytokine production	P	1.630554E-2	3.434199E-4
GO:0048011	neurotrophin TRK receptor signaling pathway	P	1.630554E-2	3.434199E-4

GO:0023057	negative regulation of signaling	P	1.776377E-2	3.802531E-4
GO:0032868	response to insulin stimulus	P	1.806155E-2	3.884597E-4
GO:0000082	G1/S transition of mitotic cell cycle	P	1.870821E-2	4.061639E-4
GO:0051291	protein heterooligomerization	P	2.023182E-2	4.495119E-4
GO:0045937	positive regulation of phosphate metabolic process	P	2.023182E-2	4.515573E-4
GO:0006302	double-strand break repair	P	2.277007E-2	5.197591E-4
GO:0019221	cytokine-mediated signaling pathway	P	2.327713E-2	5.3843E-4
GO:0021766	hippocampus development	P	2.327713E-2	5.455023E-4
GO:0033120	positive regulation of RNA splicing	P	2.327713E-2	5.455023E-4
GO:0032319	regulation of Rho GTPase activity	P	2.327713E-2	5.455023E-4
GO:0004702	receptor signaling protein serine/threonine kinase activity	F	2.327713E-2	5.455023E-4
GO:0090066	regulation of anatomical structure size	P	2.346323E-2	5.52244E-4

GO:0051147	regulation of muscle cell differentiation	P	2.428197E-2	5.764412E-4
GO:0009605	response to external stimulus	P	2.607724E-2	6.244752E-4
GO:0048634	regulation of muscle organ development	P	2.607724E-2	6.322878E-4
GO:0005811	lipid particle	C	2.772547E-2	6.891285E-4
GO:0006261	DNA-dependent DNA replication	P	2.772547E-2	6.891285E-4
GO:0042326	negative regulation of phosphorylation	P	2.772547E-2	6.891285E-4
GO:0007605	sensory perception of sound	P	2.772547E-2	6.891285E-4
GO:0030425	dendrite	C	2.881658E-2	7.331663E-4
GO:0045597	positive regulation of cell differentiation	P	2.881658E-2	7.337893E-4
GO:0009628	response to abiotic stimulus	P	2.891036E-2	7.391104E-4
GO:0045202	synapse	C	2.976301E-2	7.690505E-4
GO:0045216	cell-cell junction organization	P	2.999982E-2	7.821805E-4
GO:0071417	cellular response to organonitrogen compound	P	3.047228E-2	7.999331E-4
GO:0038093	Fc receptor signaling pathway	P	3.047228E-2	8.080581E-4
GO:0005249	voltage-gated potassium channel activity	F	3.047228E-2	8.080581E-4

GO:1901990	regulation of mitotic cell cycle phase transition	P	3.047228E-2	8.130475E-4
GO:0015291	secondary active transmembrane transporter activity	F	3.058288E-2	8.191011E-4
GO:0035020	regulation of Rac protein signal transduction	P	3.297811E-2	8.899439E-4
GO:0000793	condensed chromosome	C	3.387217E-2	9.209435E-4
GO:0010035	response to inorganic substance	P	3.465751E-2	9.458123E-4
GO:0051260	protein homooligomerization	P	3.546035E-2	9.713193E-4
GO:0014074	response to purine-containing compound	P	3.64536E-2	1.015891E-3
GO:0008023	transcription elongation factor complex	C	3.64536E-2	1.015891E-3
GO:0046943	carboxylic acid transmembrane transporter activity	F	3.64536E-2	1.017017E-3
GO:0042493	response to drug	P	3.784497E-2	1.063514E-3
GO:0051648	vesicle localization	P	3.832294E-2	1.080834E-3

GO:0010628	positive regulation of gene expression	P	3.881158E-2	1.104541E-3
GO:0010557	positive regulation of macromolecule biosynthetic process	P	3.881158E-2	1.106427E-3
GO:0034654	nucleobase-containing compound biosynthetic process	P	3.912425E-2	1.11931E-3
GO:0071705	nitrogen compound transport	P	3.915869E-2	1.124268E-3
GO:0070374	positive regulation of ERK1 and ERK2 cascade	P	4.002634E-2	1.1573E-3
GO:0043596	nuclear replication fork	C	4.002634E-2	1.1573E-3
GO:0005887	integral to plasma membrane	C	4.120947E-2	1.195689E-3
GO:0046982	protein heterodimerization activity	F	4.188171E-2	1.223693E-3
GO:0030855	epithelial cell differentiation	P	4.188171E-2	1.227941E-3
GO:0008544	epidermis development	P	4.224976E-2	1.243018E-3
GO:0045017	glycerolipid biosynthetic process	P	4.341869E-2	1.290624E-3

GO:0051272	positive regulation of cellular component movement	P	4.341869E-2	1.290624E-3
GO:0044212	transcription regulatory region DNA binding	F	4.371136E-2	1.312627E-3
GO:0051347	positive regulation of transferase activity	P	4.48427E-2	1.373896E-3
GO:0048512	circadian behavior	P	4.518468E-2	1.388958E-3
GO:0050778	positive regulation of immune response	P	4.827545E-2	1.489648E-3
GO:0048585	negative regulation of response to stimulus	P	4.827545E-2	1.493762E-3
GO:0072358	cardiovascular system development	P	4.851367E-2	1.510976E-3
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	P	4.873565E-2	1.522835E-3
GO:0009890	negative regulation of biosynthetic process	P	4.98041E-2	1.561273E-3

Table C.3: Significantly enriched GO-terms for *R. culicivorax* 1- to 4-cell stage.

GO-ID	GO-term	category	FDR	p-value
GO:0003735	structural constituent of ribosome	F	5.916601E-12	6.939355E-15
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	P	3.254666E-8	1.086455E-10
GO:0006413	translational initiation	P	7.188691E-6	3.567106E-8
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	P	8.255783E-6	4.171092E-8
GO:0022625	cytosolic large ribosomal subunit	C	4.334837E-5	2.659409E-7
GO:0044430	cytoskeletal part	C	5.113338E-5	3.229283E-7
GO:0006415	translational termination	P	5.522951E-5	3.577907E-7
GO:0006414	translational elongation	P	7.746521E-5	5.591138E-7
GO:0006898	receptor-mediated endocytosis	P	2.021053E-4	1.641057E-6
GO:0006457	protein folding	P	2.225521E-4	1.82716E-6
GO:0022627	cytosolic small ribosomal subunit	C	5.054828E-4	4.697287E-6
GO:0005525	GTP binding	F	6.611893E-4	6.382827E-6

GO:0019083	viral transcription	P	7.087416E-4	6.969761E-6
GO:0043234	protein complex	C	1.827836E-3	2.044855E-5
GO:0006096	glycolysis	P	2.767979E-3	3.296402E-5
GO:0007067	mitosis	P	3.746738E-3	4.698634E-5
GO:0007052	mitotic spindle organization	P	5.03855E-3	6.682307E-5
GO:0030660	Golgi-associated vesicle membrane	C	5.629915E-3	7.771355E-5
GO:0005743	mitochondrial inner membrane	C	5.699245E-3	7.918475E-5
GO:0007098	centrosome cycle	P	8.072573E-3	1.201709E-4
GO:0051491	positive regulation of filopodium assembly	P	1.014146E-2	1.564588E-4
GO:0065004	protein-DNA complex assembly	P	1.148816E-2	1.804126E-4
GO:0003743	translation initiation factor activity	F	1.148816E-2	1.804126E-4
GO:0003779	actin binding	F	1.148816E-2	1.813811E-4
GO:0006626	protein targeting to mitochondrion	P	1.414172E-2	2.283803E-4
GO:0048731	system development	P	1.684012E-2	2.780351E-4
GO:0008565	protein transporter activity	F	1.791862E-2	2.974581E-4
GO:0043254	regulation of protein complex assembly	P	1.870178E-2	3.138335E-4
GO:0008380	RNA splicing	P	1.896967E-2	3.200405E-4

GO:0022618	ribonucleoprotein complex assembly	P	1.912525E-2	3.243908E-4
GO:0015630	microtubule cytoskeleton	C	2.060259E-2	3.568836E-4
GO:0045737	positive regulation of cyclin-dependent protein kinase activity	P	2.338331E-2	4.177098E-4
GO:0006301	postreplication repair	P	2.338331E-2	4.177098E-4
GO:0016538	cyclin-dependent protein serine/threonine kinase regulator activity	F	2.338331E-2	4.177098E-4
GO:0030422	production of siRNA involved in RNA interference	P	2.338331E-2	4.177098E-4
GO:0008312	7S RNA binding	F	2.338331E-2	4.177098E-4
GO:0042645	mitochondrial nucleoid	C	2.602504E-2	4.716851E-4
GO:0005811	lipid particle	C	2.675434E-2	4.899973E-4
GO:0031123	RNA 3'-end processing	P	2.675434E-2	4.899973E-4
GO:0009792	embryo development ending in birth or egg hatching	P	2.871033E-2	5.335914E-4
GO:0008340	determination of adult lifespan	P	2.979121E-2	5.563677E-4

GO:0042254	ribosome biogenesis	P	3.815584E-2	7.263517E-4
GO:0030552	cAMP binding	F	4.129595E-2	7.935796E-4
GO:0016604	nuclear body	C	4.303943E-2	8.3485E-4
GO:0006397	mRNA processing	P	4.539838E-2	9.287963E-4
GO:0071480	cellular response to gamma radiation	P	4.539838E-2	9.297575E-4
GO:0048205	COPI coating of Golgi vesicle	P	4.539838E-2	9.297575E-4
GO:0004360	glutamine-fructose-6-phosphate transaminase (isomerizing) activity	F	4.539838E-2	9.297575E-4
GO:0007249	I-kappaB kinase/NF-kappaB cascade	P	4.563452E-2	9.412116E-4
GO:0006184	GTP catabolic process	P	4.563452E-2	9.522424E-4
GO:0015629	actin cytoskeleton	C	4.597776E-2	9.66512E-4
GO:0034660	ncRNA metabolic process	P	4.87695E-2	1.033998E-3

Bibliography

Abascal, F., Zardoya, R., and Posada, D. (2005). Prottest: selection of best-fit models of protein evolution. *Bioinformatics*, 21(9):2104–2105.

Aboobaker, A. and Blaxter, M. (2003a). Hox gene evolution in nematodes: novelty conserved. *Curr Opin Genet Dev*, 13(6):593–598.

Aboobaker, A. A. and Blaxter, M. L. (2003b). Hox gene loss during dynamic evolution of the nematode cluster. *Curr Biol*, 13(1):37–40.

Affymetrix (2004). *Expression Analysis Technical Manual*. Affymetrix Inc.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3):403–410.

Bachmann, A., Schneider, M., Theilenberg, E., Grawe, F., and Knust, E. (2001). Drosophila Stardust is a partner of Crumbs in the control of epithelial cell polarity. *Nature*, 414(6864):638–643.

Bar-Peled, L., Schweitzer, L. D., Zoncu, R., and Sabatini, D. M. (2012). Ragulator is a GEF for the rag GTPases that signal amino acid levels to mTORC1. *Cell*, 150(6):1196–1208.

Barton, M. K., Schedl, T. B., and Kimble, J. (1987). Gain-of-function mutations of *fem-3*, a sex-determination gene in *Caenorhabditis elegans*. *Genetics*, 115(1):107–119.

Baugh, L. R., Hill, A. A., Brown, E. L., and Hunter, C. P. (2001). Quantitative analysis of mrna amplification by in vitro transcription. *Nucleic Acids Res*, 29(5):E29.

Beatty, A., Morton, D., and Kemphues, K. (2010). The *C. elegans* homolog of Drosophila Lethal giant larvae functions redundantly with PAR-2 to maintain polarity in the early embryo. *Development*, 137(23):3995–4004.

Beatty, A., Morton, D. G., and Kemphues, K. (2013). PAR-2, LGL-1 and the CDC-42 GAP CHIN-1 act in distinct pathways to maintain polarity in the *C. elegans* embryo. *Development*, 140(9):2005–2014.

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2015). Genbank. *Nucleic Acids Res*, 43(Database issue):D30–D35.

Betschinger, J., Mechtler, K., and Knoblich, J. A. (2003). The par complex directs asymmetric cell division by phosphorylating the cytoskeletal protein lgl. *Nature*, 422(6929):326–330.

Birky, Jr, C. W. (2004). Bdelloid rotifers revisited. *Proc Natl Acad Sci U S A*, 101(9):2651–2652.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T., and Thomas, W. K. (1998). A molecular evolutionary framework for the phylum nematoda. *Nature*, 392(6671):71–75.

Blüthgen, N., Brand, K., Cajavec, B., Swat, M., Herzel, H., and Beule, D. (2005). Biological profiling of gene groups utilizing gene ontology. *Genome Inform*, 16(1):106–115.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

Borgonie, G., García-Moyano, A., Litthauer, D., Bert, W., Bester, A., van Heerden, E., Möller, C., Erasmus, M., and Onstott, T. C. (2011). Nematoda from the terrestrial deep subsurface of south africa. *Nature*, 474(7349):79–82.

Boulin, T. and Bessereau, J.-L. (2007). Mos1-mediated insertional mutagenesis in *Caenorhabditis elegans*. *Nat Protoc*, 2(5):1276–1287.

Boveri, T. (1887). Ueber Differenzierung der Zellkerne wahrend der Furchung des Eies von *Ascaris megalocephala*. *Anat. Anz.*, 2:688–693.

Boveri, T. (1899). *Festschrift für C. v. Kupffer*, chapter Die Entwickelung von *Ascaris megalocephala* mit besonderer Rücksicht auf die Kernverhältnisse., pages 383–430. Gustav Fischer Verlag, Jena.

Boveri, T. (1910). *Festschrift für R. Hertwig*, chapter Die Potenzen der Ascaris-Blastomeren bei abgeänderter Furchung, pages 133—-214. Gustav Fischer Verlag, Jena.

Brasier, A. R. (2006). The NF-kappaB regulatory network. *Cardiovasc Toxicol*, 6(2):111–130.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics*.

Byrd, D. T. and Kimble, J. (2009). Scratching the niche that controls *Caenorhabditis elegans* germline stem cells. *Semin Cell Dev Biol*, 20(9):1107–1113.

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., , A. O. H., and , W. P. W. G. (2009). Amigo: online access to ontology and annotation data. *Bioinformatics*, 25(2):288–289.

Carlson, M. (2012). celegans.db: Affymetrix *C. elegans* annotation data (chip celegans). *R package*, bioconductor.org/packages/release/data/annotation/html/celegans.db.html.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805.

Chang, C., Holtzman, D. A., Chau, S., Chickering, T., Woolf, E. A., Holmgren, L. M., Bodorova, J., Gearing, D. P., Holmes, W. E., and Brivanlou, A. H. (2001). Twisted gastrulation can function as a bmp antagonist. *Nature*, 410(6827):483–487.

Chen, C., Fenk, L. A., and de Bono, M. (2013). Efficient genome editing in *Caenorhabditis elegans* by CRISPR-targeted homologous recombination. *Nucleic Acids Res*, 41(20):e193.

Chicoine, J., Benoit, P., Gamberi, C., Paliouras, M., Simonelig, M., and Lasko, P. (2007). Bicaudal-c recruits ccr4-not deadenylase to target mrnas and regulates oogenesis, cytoskeletal organization, and its own expression. *Dev Cell*, 13(5):691–704.

Cho, S.-W., Cho, J.-H., Song, H.-O., and Park, C.-S. (2005). Identification and characterization of a putative cyclic nucleotide-gated channel, CNG-1, in C. elegans. *Mol Cells*, 19(1):149–154.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of rna isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem*, 162(1):156–159.

Chuang, C.-F., Vanhoven, M. K., Fetter, R. D., Verselis, V. K., and Bargmann, C. I. (2007). An innexin-dependent cell network establishes left-right neuronal asymmetry in *C. elegans*. *Cell*, 129(4):787–799.

Clift, D. and Schuh, M. (2013). Restarting life: fertilization and the transition from meiosis to mitosis. *Nat Rev Mol Cell Biol*, 14(9):549–562.

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(18):3674–3676.

Costache, V., McDougall, A., and Dumollard, R. (2014). Cell cycle arrest and activation of development in marine invertebrate deuterostomes. *Biochem Biophys Res Commun*, 450(3):1175–1181.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). Prottest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, 27(8):1164–1165.

De Ley, P. (2006). A quick tour of nematode diversity and the backbone of nematode phylogeny. *WormBook*, pages 1–8.

Deppe, U., Schierenberg, E., Cole, T., Krieg, C., Schmitt, D., Yoder, B., and von Ehrenstein, G. (1978). Cell lineages of the embryo of the nematode caenorhabditis elegans. *Proc Natl Acad Sci U S A*, 75(1):376–380.

Dieterich, C., Clifton, S. W., Schuster, L. N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R., Godfrey, J., Minx, P., Mitreva, M., Roeseler, W., Tian, H., Witte, H., Yang, S.-P., Wilson, R. K., and Sommer, R. J. (2008). The pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet*, 40(10):1193–1198.

Eckmann, C. R., Kraemer, B., Wickens, M., and Kimble, J. (2002). Gld-3, a bicaudal-c homolog that inhibits fbf to control germline sex determination in c. elegans. *Dev Cell*, 3(5):697–710.

Eisenmann, D. M. (2005). Wnt signaling. *WormBook*, pages 1–17.

Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*, 30(7):1575–1584.

Estevez, M., Attisano, L., Wrana, J. L., Albert, P. S., Massagué, J., and Riddle, D. L. (1993). The daf-4 gene encodes a bone morphogenetic protein receptor controlling c. elegans dauer larva development. *Nature*, 365(6447):644–649.

Fenwick, A. (2012). The global burden of neglected tropical diseases. *Public Health*, 126(3):233–236.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded rna in caenorhabditis elegans. *Nature*, 391(6669):806–811.

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., and Stoeckert, Jr, C. J. (2011). Using orthomcl to assign proteins to orthomcl-db groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics*, Chapter 6:Unit 6.12.1–Unit 6.1219.

Floyd, R., Abebe, E., Papert, A., and Blaxter, M. (2002). Molecular barcodes for soil nematode identification. *Mol Ecol*, 11(4):839–850.

Foehr, M. L. and Liu, J. (2008). Dorsoventral patterning of the c. elegans postembryonic mesoderm requires both LIN-12/Notch and TGFbeta signaling. *Dev Biol*, 313(1):256–266.

Fong, Y., Bender, L., Wang, W., and Strome, S. (2002). Regulation of the different chromatin states of autosomes and x chromosomes in the germ line of *C. elegans*. *Science*, 296(5576):2235–2238.

Friedman, L., Santa Anna-Arriola, S., Hodgkin, J., and Kimble, J. (2000). *gon-4*, a cell lineage regulator required for gonadogenesis in *Caenorhabditis elegans*. *Dev Biol*, 228(2):350–362.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20(3):307–315.

Gerhart, J. (1999). 1998 warkany lecture: signaling pathways in development. *Teratology*, 60(4):226–239.

Gilmore, T. D. (2006). Introduction to NF-kappaB: players, pathways, perspectives. *Oncogene*, 25(51):6680–6684.

Godel, C., Kumar, S., Koutsovoulos, G., Ludin, P., Nilsson, D., Comandatore, F., Wrobel, N., Thompson, M., Schmid, C. D., Goto, S., Bringaud, F., Wolstenholme, A., Bandi, C., Epe, C., Kaminsky, R., Blaxter, M., and Mäser, P. (2012). The genome of the heartworm, *Dirofilaria immitis*, reveals drug and vaccine targets. *FASEB J*, 26(11):4650–4661.

Goldstein, B., Frisse, L. M., and Thomas, W. K. (1998). Embryonic axis specification in nematodes: evolution of the first step in development. *Curr Biol*, 8(3):157–160.

Gönczy, P. and Rose, L. S. (2005). Asymmetric cell division and axis formation in the embryo. *WormBook*, pages 1–20.

Govindan, J. A., Cheng, H., Harris, J. E., and Greenstein, D. (2006). Galphao/i and Galphas signaling function in parallel with the MSP/Eph receptor to control meiotic diapause in *C. elegans*. *Curr Biol*, 16(13):1257–1268.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from rna-seq data without a reference genome. *Nat Biotechnol*, 29(7):644–652.

Greenstein, D. (2005). Control of oocyte meiotic maturation and fertilization. *WormBook*, pages 1–12.

Grossman, E. N., Giurumescu, C. A., and Chisholm, A. D. (2013). Mechanisms of ephrin receptor protein kinase-independent signaling in amphid axon guidance in *Caenorhabditis elegans*. *Genetics*, 195(3):899–913.

Gumienny, T. L. and Savage-Dunn, C. (2013). Tgf-β signaling in *C. elegans*. *WormBook*, pages 1–34.

Gupta, B. P. and Sternberg, P. W. (2003). The draft genome sequence of the nematode *Caenorhabditis briggsae*, a companion to *C. elegans*. *Genome Biol*, 4(12):238.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N., and Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, 8(8):1494–1512.

Hadjeb, N. and Berkowitz, G. A. (1996). Preparation of T-over-hang vectors with high PCR product cloning efficiency. *Biotechniques*, 20(1):20–22.

Haenfler, J. M., Kuang, C., and Lee, C.-Y. (2012). Cortical aPKC kinase activity distinguishes neural stem cells from progenitor cells by ensuring asymmetric segregation of Numb. *Dev Biol*, 365(1):219–228.

Hao, L., Johnsen, R., Lauter, G., Baillie, D., and Bürglin, T. R. (2006). Comprehensive analysis of gene expression patterns of hedgehog-related genes. *BMC Genomics*, 7:280.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*, 2(3):666–673.

Heber, S. and Sick, B. (2006). Quality assessment of affymetrix genechip data. *OMICS*, 10(3):358–368.

Heim, R., Cubitt, A. B., and Tsien, R. Y. (1995). Improved green fluorescence. *Nature*.

Hikasa, H. and Sokol, S. Y. (2013). Wnt signaling in vertebrate axis specification. *Cold Spring Harb Perspect Biol*, 5(1):a007955.

Holdeman, R., Nehrt, S., and Strome, S. (1998). MES-2, a maternal protein essential for viability of the germline in *Caenorhabditis elegans*, is homologous to a Drosophila Polycomb group protein. *Development*, 125(13):2457–2467.

Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., Bongers, T., Holovachov, O., Bakker, J., and Helder, J. (2006). Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. *Mol Biol Evol*, 23(9):1792–1800.

Hong, Y., Stronach, B., Perrimon, N., Jan, L. Y., and Jan, Y. N. (2001). Drosophila Stardust interacts with Crumbs to control polarity of epithelia but not neuroblasts. *Nature*, 414(6864):634–638.

Horne-Badovinac, S., Hill, J., Gerlach, 2nd, G., Menegas, W., and Bilder, D. (2012). A screen for round egg mutants in Drosophila identifies tricornered, furry, and misshapen as regulators of egg chamber elongation. *G3 (Bethesda)*, 2(3):371–378.

Hubbard, E. J. A. and Greenstein, D. (2005). Introduction to the germ line. *WormBook*, pages 1–4.

Hubbell, E., Liu, W.-M., and Mei, R. (2002). Robust estimators for expression analysis. *Bioinformatics*, 18(12):1585–1592.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A. F., Rivoire, C., Sangrador-Vegas, A., Selengut, J. D., Sigrist, C. J. A., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P. D., Wu, C. H., Yeats, C., and Yong, S.-Y. (2012). Interpro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*, 40(Database issue):D306–D312.

Ikegami, R., Zheng, H., Ong, S.-H., and Culotti, J. (2004). Integration of semaphorin-2A/MAB-20, ephrin-4, and UNC-129 tgf-beta signaling pathways regulates sorting of distinct sensory rays in *C. elegans*. *Dev Cell*, 6(3):383–395.

Jaeger, J. (2011). The gap gene network. *Cell Mol Life Sci*, 68(2):243–274.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D. P., Zipperlen, P., and Ahringer, J. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using rnai. *Nature*, 421(6920):231–237.

Kelly, W. G. and Fire, A. (1998). Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. *Development*, 125(13):2451–2456.

Kent, W. J. (2002). Blat–the blast-like alignment tool. *Genome Res*, 12(4):656–664.

Kikuchi, T., Cotton, J. A., Dalzell, J. J., Hasegawa, K., Kanzaki, N., McVeigh, P., Takanashi, T., Tsai, I. J., Assefa, S. A., Cock, P. J. A., Otto, T. D., Hunt, M., Reid,

A. J., Sanchez-Flores, A., Tsuchihara, K., Yokoi, T., Larsson, M. C., Miwa, J., Maule, A. G., Sahashi, N., Jones, J. T., and Berriman, M. (2011). Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog*, 7(9):e1002219.

Kimble, J. and Crittenden, S. L. (2007). Controls of germline stem cells, entry into meiosis, and the sperm/oocyte decision in *Caenorhabditis elegans*. *Annu Rev Cell Dev Biol*, 23:405–433.

Kimble, J. E. and White, J. G. (1981). On the control of germ cell development in *Caenorhabditis elegans*. *Dev Biol*, 81(2):208–219.

Kiontke, K. and Fitch, D. H. A. (2005). The phylogenetic relationships of *Caenorhabditis* and other rhabditids. *WormBook*, pages 1–11.

Kiontke, K. C., Félix, M.-A., Ailion, M., Rockman, M. V., Braendle, C., Pénigault, J.-B., and Fitch, D. H. A. (2011). A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol Biol*, 11:339.

Knust, E. and Bossinger, O. (2002). Composition and formation of intercellular junctions in epithelial cells. *Science*, 298(5600):1955–1959.

Kockel, L., Homsy, J. G., and Bohmann, D. (2001). Drosophila AP-1: lessons from an invertebrate. *Oncogene*, 20(19):2347–2364.

Kraus, C. (2009). Charakterisierung von RNAi-induzierten Phänotypen im Oozyten-Maturations-Signalweg von *Caenorhabditis elegans*. *Bacherlor Thesis*.

Lahl, V., Halama, C., and Schierenberg, E. (2003). Comparative and experimental embryogenesis of plectidae (nematoda). *Dev Genes Evol*, 213(1):18–27.

Lahl, V., Sadler, B., and Schierenberg, E. (2006). Egg development in parthenogenetic nematodes: variations in meiosis and axis formation. *Int J Dev Biol*, 50(4):393–398.

Lahl, V., Schulze, J., and Schierenberg, E. (2009). Differences in embryonic pattern formation between caenorhabditis elegans and its close parthenogenetic relative diploscapter coronatus. *Int J Dev Biol*, 53(4):507–515.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat Methods*, 9(4):357–359.

Laugsch, M. and Schierenberg, E. (2004). Differences in maternal supply and early development of closely related nematode species. *Int J Dev Biol*, 48(7):655–662.

Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from rna-seq data with or without a reference genome. *BMC Bioinformatics*, 12:323.

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., and Dewey, C. N. (2010). Rna-seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, 26(4):493–500.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and , . G. P. D. P. S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Li, L., Stoeckert, Jr, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*, 13(9):2178–2189.

Li, R. and Albertini, D. F. (2013). The road to maturation: somatic cell interaction and self-organization of the mammalian oocyte. *Nat Rev Mol Cell Biol*, 14(3):141–152.

Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659.

Li, X., Chen, Z., and Desplan, C. (2013). Temporal patterning of neural progenitors in Drosophila. *Curr Top Dev Biol*, 105:69–96.

Life-Technologies (2011). *MessageAmp ^{TM} II aRNA Amplification Kit*. Handbook. Life Technologies Inc.

Lin, S., Baye, L. M., Westfall, T. A., and Slusarski, D. C. (2010). Wnt5b-RYK pathway provides directional signals to regulate gastrulation movement. *J Cell Biol*, 190(2):263–278.

Liu, N., Dansereau, D. A., and Lasko, P. (2003). Fat facets interacts with vasa in the Drosophila pole plasm and protects it from degradation. *Curr Biol*, 13(21):1905–1909.

Mark Welch, D. and Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, 288(5469):1211–1215.

Marqués, G., Musacchio, M., Shimell, M. J., Wünnenberg-Stapleton, K., Cho, K. W., and O'Connor, M. B. (1997). Production of a dpp activity gradient in the early

drosophila embryo through the opposing actions of the sog and tld proteins. *Cell*, 91(3):417–426.

Martin, B. L. and Kimelman, D. (2009). Wnt signaling and the evolution of embryonic posterior development. *Curr Biol*, 19(5):R215–R219.

McCarter, J., Bartlett, B., Dang, T., and Schedl, T. (1999). On the control of oocyte meiotic maturation and ovulation in *Caenorhabditis elegans*. *Dev Biol*, 205(1):111–128.

Meldal, B. H. M., Debenham, N. J., De Ley, P., De Ley, I. T., Vanfleteren, J. R., Vierstraete, A. R., Bert, W., Borgonie, G., Moens, T., Tyler, P. A., Austen, M. C., Blaxter, M. L., Rogers, A. D., and Lambshead, P. J. D. (2007). An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa. *Mol Phylogenet Evol*, 42(3):622–636.

Merrill, B. J., Pasolli, H. A., Polak, L., Rendl, M., García-García, M. J., Anderson, K. V., and Fuchs, E. (2004). Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development*, 131(2):263–274.

Mickey, K. M., Mello, C. C., Montgomery, M. K., Fire, A., and Priess, J. R. (1996). An inductive interaction in 4-cell stage *C. elegans* embryos involves APX-1 expression in the signalling cell. *Development*, 122(6):1791–1798.

Miller, M. A., Nguyen, V. Q., Lee, M. H., Kosinski, M., Schedl, T., Caprioli, R. M., and Greenstein, D. (2001). A sperm cytoskeletal protein that signals oocyte meiotic maturation and ovulation. *Science*, 291(5511):2144–2147.

Miller, M. A., Ruest, P. J., Kosinski, M., Hanks, S. K., and Greenstein, D. (2003). An Eph receptor sperm-sensing control mechanism for oocyte meiotic maturation in *Caenorhabditis elegans*. *Genes Dev*, 17(2):187–200.

Minakhina, S. and Steward, R. (2006). Nuclear factor-kappa b pathways in Drosophila. *Oncogene*, 25(51):6749–6757.

Mitreva, M., Jasmer, D. P., Zarlenga, D. S., Wang, Z., Abubucker, S., Martin, J., Taylor, C. M., Yin, Y., Fulton, L., Minx, P., Yang, S.-P., Warren, W. C., Fulton, R. S., Bhonagiri, V., Zhang, X., Hallsworth-Pepin, K., Clifton, S. W., McCarter,

J. P., Appleton, J., Mardis, E. R., and Wilson, R. K. (2011). The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat Genet*, 43(3):228–235.

Mizumoto, K. and Sawa, H. (2007). Two betas or not two betas: regulation of asymmetric division by beta-catenin. *Trends Cell Biol*, 17(10):465–473.

Mohamed, A. M. and Chin-Sang, I. D. (2006). Characterization of loss-of-function and gain-of-function eph receptor tyrosine kinase signaling in *C. elegans* axon targeting and cell migration. *Dev Biol*, 290(1):164–176.

Mortazavi, A., Schwarz, E. M., Williams, B., Schaeffer, L., Antoshechkin, I., Wold, B. J., and Sternberg, P. W. (2010). Scaffolding a *Caenorhabditis* nematode genome with rna-seq. *Genome Res*, 20(12):1740–1747.

Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., and Morris, Q. (2008). Genemania: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*, 9 Suppl 1:S4.

Müller, G. B. (2007). Evo-devo: extending the evolutionary synthesis. *Nat Rev Genet*, 8(12):943–949.

Nadarajan, S., Govindan, J. A., McGovern, M., Hubbard, E. J. A., and Greenstein, D. (2009). MSP and GLP-1/Notch signaling coordinately regulate actomyosin-dependent cytoplasmic streaming and oocyte growth in *C. elegans*. *Development*, 136(13):2223–2234.

Nelson, G. A., Lew, K. K., and Ward, S. (1978). Intersex, a temperature-sensitive mutant of the nematode *Caenorhabditis elegans*. *Dev Biol*, 66(2):386–409.

Nelson, G. A., Schubert, W. W., Marshall, T. M., Benton, E. R., and Benton, E. V. (1989). Radiation effects in *Caenorhabditis elegans*, mutagenesis by high and low LET ionizing radiation. *Mutat Res*, 212(2):181–192.

Nguyen-Ngoc, T., Afshar, K., and Gönczy, P. (2007). Coupling of cortical dynein and G alpha proteins mediates spindle positioning in *Caenorhabditis elegans*. *Nat Cell Biol*, 9(11):1294–1302.

Niedringhaus, T. P., Milanova, D., Kerby, M. B., Snyder, M. P., and Barron, A. E. (2011). Landscape of next-generation sequencing technologies. *Anal Chem*, 83(12):4327–4341.

Niessing, D., Rivera-Pomar, R., La Rosée, A., Häder, T., Schöck, F., Purnell, B. A., and Jäckle, H. (1997). A cascade of transcriptional control leading to axis determination in Drosophila. *J Cell Physiol*, 173(2):162–167.

Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature*, 287(5785):795–801.

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., and Windham, E. (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proc Natl Acad Sci U S A*, 105(39):14802–14807.

Ottesen, E. A., Hooper, P. J., Bradley, M., and Biswas, G. (2008). The global programme to eliminate lymphatic filariasis: health impact after 8 years. *PLoS Negl Trop Dis*, 2(10):e317.

Ow, M. C., Martinez, N. J., Olsen, P. H., Silverman, H. S., Barrasa, M. I., Conradt, B., Walhout, A. J. M., and Ambros, V. (2008). The FLYWCH transcription factors FLH-1, FLH-2, and FLH-3 repress embryonic expression of microrna genes in *C. elegans*. *Genes Dev*, 22(18):2520–2534.

Özüak, O., Buchta, T., Roth, S., and Lynch, J. A. (2014). Dorsoventral polarity of the nasonia embryo primarily relies on a bmp gradient formed without input from toll. *Curr Biol*, 24(20):2393–2398.

Pearson, K. (1895). Note on Regression and Inheritance in the Case of two Parents. *Proceedings of the Royal Society of London*, 58:240–242.

Perry, M. W., Bothma, J. P., Luu, R. D., and Levine, M. (2012). Precision of hunchback expression in the Drosophila embryo. *Curr Biol*, 22(23):2247–2252.

Petersen, C. P. and Reddien, P. W. (2009). Wnt signaling and the polarity of the primary body axis. *Cell*, 139(6):1056–1068.

Phillips, C. M. and Dernburg, A. F. (2006). A family of zinc-finger proteins is required for chromosome-specific pairing and synapsis during meiosis in *C. elegans*. *Dev Cell*, 11(6):817–829.

Piccolo, S., Agius, E., Lu, B., Goodman, S., Dale, L., and De Robertis, E. M. (1997). Cleavage of chordin by xolloid metalloprotease suggests a role for proteolytic processing in the regulation of spemann organizer activity. *Cell*, 91(3):407–416.

Piccolo, S., Sasai, Y., Lu, B., and De Robertis, E. M. (1996). Dorsoventral patterning in Xenopus: inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell*, 86(4):589–598.

Priess, J. R. (2005). Notch signaling in the *C. elegans* embryo. *WormBook*, pages 1–16.

Pujol, N., Link, E. M., Liu, L. X., Kurz, C. L., Alloing, G., Tan, M. W., Ray, K. P., Solari, R., Johnson, C. D., and Ewbank, J. J. (2001). A reverse genetic analysis of components of the toll signaling pathway in caenorhabditis elegans. *Curr Biol*, 11(11):809–821.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). Interproscan: protein domains identifier. *Nucleic Acids Res*, 33(Web Server issue):W116–W120.

Reinke, V., Gil, I. S., Ward, S., and Kazmer, K. (2004). Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development*, 131(2):311–323.

Reinke, V., Smith, H. E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S. J., Davis, E. B., Scherer, S., Ward, S., and Kim, S. K. (2000). A global profile of germline gene expression in *C. elegans*. *Mol Cell*, 6(3):605–616.

Rhyu, M. S., Jan, L. Y., and Jan, Y. N. (1994). Asymmetric distribution of numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. *Cell*, 76(3):477–491.

Robert, V. and Bessereau, J.-L. (2007). Targeted engineering of the *Caenorhabditis elegans* genome following Mos1-triggered chromosomal breaks. *EMBO J*, 26(1):170–183.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140.

Robinson, M. D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 11(3):R25.

Robinson, M. D. and Smyth, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, 23(21):2881–2887.

Robinson, M. D. and Smyth, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, 9(2):321–332.

Rodríguez, I. (2004). The dachsous gene, a member of the cadherin family, is required for Wg-dependent pattern formation in the Drosophila wing disc. *Development*, 131(13):3195–3206.

Roth, S. (2001). Drosophila oogenesis: coordinating germ line and soma. *Curr Biol*, 11(19):R779–R781.

Roth, S. and Lynch, J. A. (2009). Symmetry breaking during Drosophila oogenesis. *Cold Spring Harb Perspect Biol*, 1(2):a001891.

Sardet, C., Paix, A., Prodon, F., Dru, P., and Chenevert, J. (2007). From oocyte to 16-cell stage: cytoplasmic and cortical reorganizations that pattern the ascidian embryo. *Dev Dyn*, 236(7):1716–1731.

Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L. K., and De Robertis, E. M. (1994). Xenopus chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell*, 79(5):779–790.

Satoh, K., Kasai, M., Ishidao, T., Tago, K., Ohwada, S., Hasegawa, Y., Senda, T., Takada, S., Nada, S., Nakamura, T., and Akiyama, T. (2004). Anteriorization of neural fate by inhibitor of beta-catenin and t cell factor (icat), a negative regulator of wnt signaling. *Proc Natl Acad Sci U S A*, 101(21):8017–8021.

Schierenberg, E. (1984). Altered cell-division rates after laser-induced cell fusion in nematode embryos. *Dev Biol*, 101(1):240–245.

Schiffer, P. H., Kroiher, M., Kraus, C., Koutsovoulos, G. D., Kumar, S., Camps, J. I. R., Nsah, N. A., Stappert, D., Morris, K., Heger, P., Altmüller, J., Frommolt, P., Nürnberg, P., Thomas, W. K., Blaxter, M. L., and Schierenberg, E. (2013). The genome of *Romanomermis culicivorax*: revealing fundamental changes in the core developmental genetic toolkit in Nematoda. *BMC Genomics*, 14:923.

Schiffer, P. H., Nsah, N. A., Grotehusmann, H., Kroiher, M., Loer, C., and Schierenberg, E. (2014). Developmental variations among panagrolaimid nematodes indicate developmental system drift within a small taxonomic unit. *Dev Genes Evol*, 224(3):183–188.

Schulze, J., Houthoofd, W., Uenk, J., Vangestel, S., and Schierenberg, E. (2012). Plectus - a stepping stone in embryonic cell lineage evolution of nematodes. *Evodevo*, 3(1):13.

Schulze, J. and Schierenberg, E. (2008). Cellular pattern formation, establishment of polarity and segregation of colored cytoplasm in embryos of the nematode *Romanomermis culicivorax*. *Dev Biol*, 315(2):426–436.

Schulze, J. and Schierenberg, E. (2009). Embryogenesis of *Romanomermis culicivorax*: an alternative way to construct a nematode. *Dev Biol*, 334(1):10–21.

Schulze, J. and Schierenberg, E. (2011). Evolution of embryonic development in nematodes. *Evodevo*, 2(1):18.

Scott, A. L. and Ghedin, E. (2009). The genome of brugia malayi - all worms are not created equal. *Parasitol Int*, 58(1):6–11.

Scott, I. C., Blitz, I. L., Pappano, W. N., Imamura, Y., Clark, T. G., Steiglitz, B. M., Thomas, C. L., Maas, S. A., Takahara, K., Cho, K. W., and Greenspan, D. S. (1999). Mammalian BMP-1/Tolloid-related metalloproteinases, including novel family member mammalian Tolloid-like 2, have differential enzymatic activities and distributions of expression relevant to patterning and skeletogenesis. *Dev Biol*, 213(2):283–300.

Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nat Methods*, 2(12):905–909.

Shimmi, O. and O'Connor, M. B. (2003). Physical properties of tld, sog, tsg and dpp protein interactions are predicted to help create a sharp boundary in bmp signals during dorsoventral patterning of the drosophila embryo. *Development*, 130(19):4673–4682.

Shimomura, O., Johnson, F. H. ., and Saiga, Y. (1962). Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, *Aequorea*. *J. of Cell. Physiol.*

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G.

(2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol*, 7:539.

Simon, J.-C., Delmotte, F., Rispe, C., and Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. *Biol. J. Linnean. Soc.*

Skiba, F. and Schierenberg, E. (1992). Cell lineages, developmental timing, and spatial pattern formation in embryos of free-living soil nematodes. *Dev Biol*, 151(2):597–610.

Smith, W. C. and Harland, R. M. (1992). Expression cloning of noggin, a new dorsalizing factor localized to the spemann organizer in Xenopus embryos. *Cell*, 70(5):829–840.

Sommer, R. J. (2001). As good as they get: cells in nematode vulva development and evolution. *Curr Opin Cell Biol*, 13(6):715–720.

Sommer, R. J. (2012). Evolution of regulatory networks: nematode vulva induction as an example of developmental systems drift. *Adv Exp Med Biol*, 751:79–91.

Song, X., Call, G. B., Kirilly, D., and Xie, T. (2007). Notch signaling controls germline stem cell niche formation in the Drosophila ovary. *Development*, 134(6):1071–1080.

Sonnenberg, R., Nolte, A. W., and Tautz, D. (2007). An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Front Zool*, 4:6.

Spike, C., Meyer, N., Racen, E., Orsborn, A., Kirchner, J., Kuznicki, K., Yee, C., Bennett, K., and Strome, S. (2008a). Genetic analysis of the *Caenorhabditis elegans* GLH family of P-granule proteins. *Genetics*, 178(4):1973–1987.

Spike, C. A., Bader, J., Reinke, V., and Strome, S. (2008b). DEPS-1 promotes P-granule assembly and RNA interference in *C. elegans* germ cells. *Development*, 135(5):983–993.

Srinivasan, J., Dillman, A. R., Macchietto, M. G., Heikkinen, L., Lakso, M., Fracchia, K. M., Antoshechkin, I., Mortazavi, A., Wong, G., and Sternberg, P. W. (2013). The draft genome and transcriptome of *Panagrellus redivivus* are shaped by the harsh demands of a free-living lifestyle. *Genetics*, 193(4):1279–1295.

Srinivasan, S., Rashka, K. E., and Bier, E. (2002). Creation of a Sog morphogen gradient in the Drosophila embryo. *Dev Cell*, 2(1):91–101.

St Johnston, D. (2002). The art and design of genetic screens: *Drosophila melanogaster*. *Nat Rev Genet*, 3(3):176–188.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21):2688–2690.

Stanke, M. and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*, 33(Web Server issue):W465–W467.

Stanke, M. and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, 19 Suppl 2:ii215–ii225.

Starich, T. A., Xu, J., Skerrett, I. M., Nicholson, B. J., and Shaw, J. E. (2009). Interactions between innexins UNC-7 and UNC-9 mediate electrical synapse specificity in the *Caenorhabditis elegans* locomotory nervous system. *Neural Dev*, 4:16.

Stein, K. K., Nesmith, J. E., Ross, B. D., and Golden, A. (2010). Functional redundancy of paralogs of an anaphase promoting complex/cyclosome subunit in *Caenorhabditis elegans* meiosis. *Genetics*, 186(4):1285–1293.

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D. H. A., Fulton, L. A., Fulton, R. E., Griffiths-Jones, S., Harris, T. W., Hillier, L. W., Kamath, R., Kuwabara, P. E., Mardis, E. R., Marra, M. A., Miner, T. L., Minx, P., Mullikin, J. C., Plumb, R. W., Rogers, J., Schein, J. E., Sohrmann, M., Spieth, J., Stajich, J. E., Wei, C., Willey, D., Wilson, R. K., Durbin, R., and Waterston, R. H. (2003). The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol*, 1(2):E45.

Stevens, N. (1909). The effect of ultraviolet light upon the developing eggs of *Ascaris megalocephala*. *Arch. Entw. Mech.*, 27:622–639.

Strausberg, R. L., Levy, S., and Rogers, Y.-H. (2008). Emerging DNA sequencing technologies for human genomic medicine. *Drug Discov Today*, 13(13-14):569–577.

Sulston, J. E., B. S. (1974). The DNA of *Caenorhabditis elegans*. *Genetics*.

Sulston, J. E. (1976). Post-embryonic development in the ventral cord of *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci*, 275(938):287–297.

Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, caenorhabditis elegans. *Dev Biol*, 56(1):110–156.

Sulston, J. E., Schierenberg, E., White, J. G., and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev Biol*, 100(1):64–119.

Tabara, H., Grishok, A., and Mello, C. C. (1998). Rnai in *C. elegans*: soaking in the genome sequence. *Science*, 282(5388):430–431.

Caenorhabditis elegans Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 282(5396):2012–2018.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–4680.

Tian, C., Sen, D., Shi, H., Foehr, M. L., Plavskin, Y., Vatamaniuk, O. K., and Liu, J. (2010). The RGM protein DRAG-1 positively regulates a BMP-like signaling pathway in *Caenorhabditis elegans*. *Development*, 137(14):2375–2384.

Timmons, L. and Fire, A. (1998). Specific interference by ingested dsRNA. *Nature*, 395(6705):854.

Trudgill, D. L. and Blok, V. C. (2001). Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. *Annu Rev Phytopathol*, 39:53–77.

True, J. R. and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol Dev*, 3(2):109–119.

Turek, M., Lewandrowski, I., and Bringmann, H. (2013). An AP2 transcription factor is required for a sleep-active neuron to induce sleep-like quiescence in *C. elegans*. *Curr Biol*, 23(22):2215–2223.

Updike, D. and Strome, S. (2010). P granule assembly and function in *Caenorhabditis elegans* germ cells. *J Androl*, 31(1):53–60.

Viktorinová, I., König, T., Schlichting, K., and Dahmann, C. (2009). The cadherin Fat2 is required for planar cell polarity in the Drosophila ovary. *Development*, 136(24):4123–4132.

Vogt, T. (2012). Phylogenetic Analysis of parthenogenetic Nematodes by Next Generation Sequencing. *Master Thesis*.

Voronina, E. (2013). The diverse functions of germline P-granules in *Caenorhabditis elegans*. *Mol Reprod Dev*, 80(8):624–631.

Voronov, D. A. and Panchin, Y. V. (1998). Cell lineage in marine nematode *Enoplus brevis*. *Development*, 125(1):143–150.

Vrain, T. C. (1993). Restriction fragment length polymorphism separates species of the *Xiphinema americanum* group. *J Nematol*, 25(3):361–364.

Walhout, A. J. M., Reboul, J., Shtanko, O., Bertin, N., Vaglio, P., Ge, H., Lee, H., Doucette-Stamm, L., Gunsalus, K. C., Schetter, A. J., Morton, D. G., Kemphues, K. J., Reinke, V., Kim, S. K., Piano, F., and Vidal, M. (2002). Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr Biol*, 12(22):1952–1958.

Wang, J., Garrey, J., and Davis, R. E. (2014). Transcription in pronuclei and one- to four-cell embryos drives early development in a nematode. *Curr Biol*, 24(2):124–133.

Wang, J., Mitreva, M., Berriman, M., Thorne, A., Magrini, V., Koutsovoulos, G., Kumar, S., Blaxter, M. L., and Davis, R. E. (2012). Silencing of germline-expressed genes by dna elimination in somatic cells. *Dev Cell*, 23(5):1072–1080.

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., and Morris, Q. (2010). The geneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*, 38(Web Server issue):W214–W220.

Wharton and Ferns (1995). Survival of intracellular freezing by the Antarctic nematode *Panagrolaimus davidi*. *J Exp Biol*, 198(Pt 6):1381–1387.

Wharton, D. A. (2003). The environmental physiology of antarctic terrestrial nematodes: a review. *J Comp Physiol B*, 173(8):621–628.

Whitten, S. J. and Miller, M. A. (2007). The role of gap junctions in caenorhabditis elegans oocyte maturation and fertilization. *Dev Biol*, 301(2):432–446.

Wiegner, O. and Schierenberg, E. (1998). Specification of gut cell fate differs significantly between the nematodes acrobeloides nanus and caenorhabditis elegans. *Dev Biol*, 204(1):3–14.

Wiegner, O. and Schierenberg, E. (1999). Regulative development in a nematode embryo: a hierarchy of cell fate transformations. *Dev Biol*, 215(1):1–12.

Williams, K. J., Fisher, J. M., and Langridge, P. (1996). Development of a PCR-based allele-specific assay from an RFLP probe linked to resistance to cereal cyst nematode in wheat. *Genome*, 39(4):798–801.

Wilson, C. L. and Miller, C. J. (2005). Simpleaffy: a bioconductor package for affymetrix quality control and data analysis. *Bioinformatics*, 21(18):3683–3685.

Wilson, P. G. (2008). Centriole inheritance. *Prion*, 2(1):9–16.

Witte, H., Moreno, E., Rödelsperger, C., Kim, J., Kim, J.-S., Streit, A., and Sommer, R. J. (2015). Gene inactivation using the crispr/cas9 system in the nematode pristionchus pacificus. *Dev Genes Evol*, 225(1):55–62.

Wolff, C., Sommer, R., Schröder, R., Glaser, G., and Tautz, D. (1995). Conserved and divergent expression aspects of the Drosophila segmentation gene *hunchback* in the short germ band embryo of the flour beetle Tribolium. *Development*, 121(12):4227–4236.

Wolpert, L. (1996). One hundred years of positional information. *Trends Genet*, 12(9):359–364.

Xie, T., Song, X., Jin, Z., Pan, L., Weng, C., Chen, S., and Zhang, N. (2008). Interactions between stem cells and their niche in the Drosophila ovary. *Cold Spring Harb Symp Quant Biol*, 73:39–47.

Yamamoto, Y. and Oelgeschläger, M. (2004). Regulation of bone morphogenetic proteins in early embryonic development. *Naturwissenschaften*, 91(11):519–534.

Zecca, M. and Struhl, G. (2010). A feed-forward circuit linking wingless, fat-dachsous signaling, and the warts-hippo pathway to drosophila wing growth. *PLoS Biol*, 8(6):e1000386.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., Yang, W., Hu, J., Xiao, J., Yang, Z., Liu, Y., Xie, Q., Yu, H., Lian, J., Wen, P., Zhang, F., Li, H., Zeng, Y., Xiong, Z., Liu, S., Zhou, L., Huang, Z., An, N., Wang, J., Zheng, Q., Xiong, Y., Wang, G., Wang, B., Wang, J., Fan, Y., da Fonseca, R. R., Alfaro-Núñez, A., Schubert, M., Orlando, L., Mourier, T., Howard, J. T., Ganapathy, G., Pfenning, A., Whitney, O., Rivas, M. V., Hara, E., Smith, J., Farré, M., Narayan, J., Slavov, G., Romanov, M. N., Borges, R., Machado, J. P., Khan, I., Springer, M. S., Gatesy, J., Hoffmann, F. G., Opazo, J. C., Håstad, O., Sawyer, R. H., Kim, H., Kim, K.-W., Kim, H. J., Cho, S., Li, N., Huang, Y., Bruford, M. W., Zhan, X., Dixon, A., Bertelsen, M. F., Derryberry, E., Warren, W., Wilson, R. K., Li, S., Ray, D. A., Green, R. E., O'Brien, S. J., Griffin, D., Johnson, W. E., Haussler, D., Ryder, O. A., Willerslev, E., Graves, G. R., Alström, P., Fjeldså, J., Mindell, D. P., Edwards, S. V., Braun, E. L., Rahbek, C., Burt, D. W., Houde, P., Zhang, Y., Yang, H., Wang, J., , A. G. C., Jarvis, E. D., Gilbert, M. T. P., and Wang, J. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215):1311–1320.

Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G. K.-S., and Yu, J. (2006). Kaks_calculator: calculating ka and ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*, 4(4):259–263.

Zhou, M. Y., Clark, S. E., and Gomez-Sanchez, C. E. (1995). Universal cloning method by TA strategy. *Biotechniques*, 19(1):34–35.

Erklärung

Erklärung (entsprechend §4 Abs. 1 Nr. 9 der Promotionsordnung vom 02. Februar 2006, mit Änderungen vom 10. Mai 2012):

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit einschließlich Tabellen, Karten und Abbildungen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie abgesehen von unten angegebenen Teilpublikationen noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Einhard Schierenberg betreut worden.

Unterschrift:

Datum:

Teilpublikationen:

Peter Sarkies, Murray E. Selkirk, John T. Jones, Vivian Blok, Thomas Boothby, Bob Goldstein, Ben Hanelt, Alex Ardila-Garcia, Naomi M. Fast, Phillip M. Schiffer, Christopher Kraus, Mark J. Taylor, Georgios Koutsovoulos, Mark L. Blaxter, Eric A. Miska, (2015) “Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages.”, PLoS Biology 13(2):e1002061

Philipp H Schiffer, Michael Kroiher, Christopher Kraus, Georgios D Koutsovoulos, Sujai Kumar, Julia I R Camps, Ndifon A Nsah, Dominik Stappert, Krystalynne Morris, Peter Heger, Janine Altmüller, Peter Frommolt, Peter Nürnberg, W Kelley Thomas, Mark L Blaxter and Einhard Schierenberg, (2013) “The genome of Romanomermis culicivorax: revealing fundamental changes in the core developmental genetic toolkit in Nematoda.”, BMC Genomics 14:923

Vorträge:

Vortrag am “Evolutionary Biology of Caenorhabditis and other Nematodes" Meeting
Welcome Trust Sanger Institute, European Institute for Bioinformatics (EMBL), Hinxton, Cambridge (GB)

- Titel des Vortrags:
  “Establishment of embryonic polarity in nematodes: How to access non-model nematodes”

Eingeladener Vortragender bei der “Euro Evo Devo” Konferenz 2014
Universtität zu Wien

- Titel des Vortrags:
  “Insights into the evolution of early development of parthenogenetic nematodes by second generation sequencing”

Poster:

Poster Präsentation bei der "3rd summer school of Evolutionary Developmental Biology"
Instituto di Veneto di Scienze Lettre ed Arti (Venedig)

- Titel des Posters:
  “Dramatic differences on the level of genome and development between two closely related nematode species with different modes of reproduction”

Lebenslauf

Christopher Kraus

Geburtstag:	17. Februar 1986 in Waldbröl
Adresse:	Am Weiherhof 7, 52382 Niederzier
E-Mail:	c.kraus@uni-koeln.de
Mobiltelefon:	0160-6508478
Telefon (Büro):	0221-4701564
Familienstand:	ledig

Hochschulausbildung

Februar 2012 bis Juni 2015
Doktor rer. nat. in Zoologie
Universität zu Köln
Zoologisches Institut
- Titel der Doktorarbeit:
  „Phylum-wide transcriptome analysis of oogenesis and early embryogenesis in selected nematode species“
- Spezialisierung in Bioinformatik und Next Generation Sequencing

Oktober 2009 bis Dezember 2011
Master of Science in Biologie
Universität zu Köln
- Titel der Masterarbeit:
  „Roles of Growth Factor Signaling during Cutaneous Wound Healing of Adult Zebrafish (*Danio rerio*)“
- Spezialisierung in Entwicklungsbiologie und Genetik

April 2007 bis September 2009
Bachelor of Science in Biologie
Universität zu Köln
- Wiederaufnahme des Studium nach Ableisten des Zivildienstes
- Titel der Bachelorarbeit:
  „Charakterisierung von RNAi-induzierten Phänotypen im Oozyten-Maturations Signalweg in *Caenorhabditis elegans*“

Oktober 2005 bis Februar 2006
Aufnahme des Biologiestudiums an der Universität zu Köln
- Absolvierung des ersten Semesters

Zivildienst

Mai 2006 bis Januar 2007
Behindertenwerkstätte Oberberg (BWO)
Bereich Gartenbau

Schulausbildung

bis Juni 2005
Abitur
Gymnasium Nümbrecht