Chromatin diminution in *Mesocyclops edax* (Crustacea, Copepoda): similarity of the pre- and post-diminution euchromatic genomes.

Christian McKinnon

Thesis presented to the

Faculty of Graduate and Postdoctoral Studies

University of Ottawa,

Ottawa, Ontario, Canada

As a partial fulfilment to the M.Sc. in Biology degree

Abstract:

Chromatin diminution is defined as the elimination of DNA during the differentiation of early embryonic cells into pre-somatic cells. While it was first observed in the nematode *Parascaris equorum*, it also been identified in other parasitic nematodes, hagfish and copepods. In the copepod *Mesocyclops edax*, up to 90% of genomic DNA is eliminated during chromatin diminution. It was previously shown that the eliminated DNA contained highly repetitive heterochromatic sequences. Here, we digested pre- and post-diminution DNA with *BamHI* and produced small libraries of clones from each. Analyses revealed no decrease in low copy numbered sequences, such as transposable elements. Rather, both libraries are found to be surprisingly similar in all aspects analysed. Further comparison also demonstrated similarity of our libraries with the DNA sequences eliminated from *Cyclops kolensis*. Consequently, we suggest that *M. edax* eliminates portions of euchromatic DNA, in addition to the previously characterized satellite sequences.

Résumé :

La diminution de chromatine est définie comme l'élimination d'ADN durant la différentiation de cellules embryonnaires en cellules pré-somatiques. Découverte chez le nématode *Parascaris equorum*, elle a aussi été observée chez d'autres nématodes parasites, des myxines et des copépodes. Chez *Mesocyclops edax* (copepoda), jusqu'à 90% d'ADN génomique est éliminé lors de la diminution de chromatine. Il a été démontré qu'il y a réduction d'ADN hétérochromatique hautement répété. Dans l'étude présente, nous avons coupé de l'ADN pré- et post-diminution avec *BamHI* et nous en avons produit des librairies de clones. Les analyses menées n'ont démontré aucune réduction de séquences à faible répétition, telles que les éléments transposables. Plutôt, nos deux librairies se retrouvent à être similaires dans tout les aspects étudiés, de même qu'avec les séquences d'ADN éliminés chez *Cyclops kolensis*. Conséquemment, nous suggérons que *M. edax* élimine une portion d'ADN euchromatique, en plus de l'ADN satellite répété, tel que démontré auparavant.

Acknowledgements:

First and foremost, I would like to thank Dr. Guy Drouin for giving me the opportunity to pursue this study and guiding me throughout its length. His expertise and contributions are greatly appreciated and will remain a considerable influence though my upcoming career.

Thanks also go to Dr. Grace Wyngaard, for specimen collection and identification, and also for providing access to her laboratory, enabling us to pursue this study. Discussions concerning chromatin diminution in copepods are also appreciated.

I would also like to thank my colleagues, in particular Benoît Pagé who has helped immensely with several tasks throughout this study and unrelated schoolwork. He has also provided with a pleasant work environment in an otherwise lonely laboratory.

Last but not least, I would like to thank my parents for supporting me for more than 25 years as to date, and providing comfort and resources throughout all my studies.

Table of Contents

List of figures

List of tables

List of abbreviations

CBR: Chromosome Breakage Region

CD: Chromatin Diminution

CE: Chromosome Elimination

FISH: Fluorescent *in situ* hybridization

IES: Internal Eliminated Sequence

LTR: Long Terminal Repeat

RASL: Randomly Amplified Shotgun Libraries

Tas: Transposon-like element of *Ascaris*

TE: Transposable Element

1. Introduction:

In animals, it is generally assumed that the DNA contents of the gametes are haploid copies of the somatic cells. However, this has been found to be false in some cases, where the DNA contents of germ cells is noticeably larger than those of somatic cells. The process of chromatin diminution (CD) happens during the differentiation of germinal cells to precursors of somatic cells, where the DNA content of germ cells appears to be reduced during mitosis. The first observation of CD has been made in 1887 (Boveri 1887) in *Pascaris equorum*, a parasitic nematode of horses. Since then, it has also been observed in some other parasitic nematodes, along with occurrences in copepods, hagfish, ciliates, insects, marsupials and the zebra finch (Goday and Rosario Esteban 2001; Kloc and Zagrodzinska 2001; Pigozzi and Solari 2005). CD, in its broadest definition, appears to somewhat vary depending on the organism in which it is observed: in hagfish, where it results in the loss of complete chromosomes, the process has been named chromatin elimination, and in ciliates, where the circular chromosome excises genetic material and amplifies the somatic genome, it is termed DNA elimination. Even though CD has been known of for such a long time, neither its purpose nor its mechanism have been established, but not for lack of hypotheses. Parasitic nematodes remain by far the multicellular organisms in which the mechanism was most studied, while hagfish represent the organisms in which CD was characterized molecularly in the broadest number of species. In copepods however, CD has been well characterized in a cytological manner, while only a few studies have tackled the subject in a molecular fashion.

1.1 In nematodes:

Since nematodes are the species in which the mechanisms of CD have been the most studied, it is important to list the findings which may provide a guideline for the study of CD in other organisms. So far, CD has been reported to occur in ten species of parasitic nematodes (Müller and Tobler 2000). However, studies have primarily been done in three species: *P. equorum*, *Parascaris univalens* and *Ascaris suum*. The location of germline limited heterochromatic fragments on the chromosomes of nematodes have been found to be either at the chromosomal termini or intercalary along the chromosomes. In *P. univalens*, the two pairs of chromosomes contain intercalary heterochromatin fragments. The excision of these fragments leads to fragmentation of the chromosomes during CD, as also observed in *A. suum* (Goday and Pimpinelli 1984; Niedermaier and Moritz 2000). While the heterochromatin is located at the ends of its chromosome pair in *P. univalens*, CD also gives place to fragmentation. It was suggested that this might be the remains of the elimination mechanism of ancient intercalary fragments, similar to *P. equorum* (Goday and Pimpinelli 1984). In both *P. equorum* and *P. univalens*, the germ line chromosomes are polymorphic: the heterochromatin fragments vary in size in different individuals, resulting in chromosomes of different lengths. However, the karyotype remains the same in a given individual (Goday and Pimpinelli 1986). In *P. univalens* and *A. suum*, the chromosomes have characteristic condensation during mitosis of cells undergoing CD. The heterochromatin segments are linked to the future somatic chromosomes by an “euchromatic intercalary region” (EIR), which, unlike the euchromatin, remains decondensed. When anaphase starts, the EIR and

heterochromatin are excised from the chromosomes while the pre-somatic chromosomes migrate to the spindle poles (Niedermaier and Moritz 2000).

As parasitic nematodes are the organisms in which CD was first studied in a cytological manner, they were also the first organisms in which CD was studied at the molecular level. Specifically, the first species was *A. suum*, which reportedly eliminated from 22% to 34% of its genome (Tobler et al. 1972; Moritz and Roth 1976; Pasternak and Barrell 1976; Goldstein and Straus 1978), in which the fraction of repeated DNA would decrease from 23% in spematids to 10% in the larva (Tobler et al. 1972). The repetitive DNA was further characterized using comparative restriction endonuclease analysis. Roth and Moritz (1981) discovered the elimination of two repeated sequences of 125bp and 131bp in size, arranged in tandem arrays. However, these are not completely eliminated as copies are still found in the somatic genome. These very repeats also seem to be preferentially bound by a topoisomerase I enzyme (Jansen and Moritz 1986). Muller et al. (1982) identified another 121bp-long repeat, which can be found as five different variants and possesses a complementary inverted repeat, capable of forming a hairpin structure of 6bp. These tandem arrays were said to account for 70 to 90% of the repeats eliminated from *A. suum*. The authors also speculated that these repeats evolved from a specific undecamer repeat, as follows: GCA(T/A)TT(T/G)TGAT. Streeck et al. (1982) identified a 123bp repeat (similar to the one found by Muller) which can be found in the form of two main variants which they estimated to compose 70% of the eliminated repeats (while the other 30% is suspected to be other related variants).

These repeats also presented a symmetric dyad that would be complementary over a 24bp region.

While tandem repeats occupy a large portion of the eliminated DNA, sequences of other nature have also been found to be eliminated in *A. suum*, in particular retrotransposons and single copy genes. The eliminated retrotransposon, "Tas" for transposon-like element of *Ascaris*, is 7.5kb long and can be found complete with both 5' and 3' long terminal repeats (LTRs) and both primer binding sites. Two variants (Tas-1 and Tas-2) were found, the first being twice as frequent in the genome as the other. While Tas-1 is reduced by approximately one fourth, Tas-2 appears to be completely eliminated after CD (Aeby et al. 1986).

There are three single copy genes, *ALEP-1*, *fert-1* and *aleg-3*, that also seem to be eliminated from the germ-line genome. *ALEP-1* (for *Ascaris lumbricoides* var. *suum* eliminated protein 1), encodes for a protein homologous to the small ribosomal subunit protein S19 (RpS19G, germ-line restricted). Ribosomes containing this specific ribosomal protein homolog seem to be 10 times more abundant in the oocytes than those containing the uneliminated RpS19S (somatic) protein product (Etter et al. 1994). However, it is assumed that CD is not involved in the removal of extra copies of rRNA genes in *A. suum* since no major quantitative difference in rDNA can be observed between the germinal and somatic genome (Tobler et al. 1972; Tobler et al. 1974; Back et al. 1984). *Fert-1*, which is located in the vicinity of *alep-1*, encodes for several transcripts, which are only transcribed after fertilization (hence the name *fert-1*). The transcripts do not contain ORFs for proteins of significant length. Instead, the authors propose that they serve as

structurally functional RNAs, since they are highly stable: one particular transcript can last for up to 41 days, while CD happens during the 3rd or 4th day (Spicher et al. 1994). The third eliminated gene, *aleg-3* (for *A. lumbricoides* var. *suum* eliminated gene 3), is thought to code for a protein, although it did not present homology to other proteins found in databases at the time of study (Huang et al. 1996). All three aforementioned genes are also found to be eliminated from the somatic genome of *P. equorum* (Etter et al. 1991; Spicher et al. 1994; Huang et al. 1996).

Another interesting topic concerning CD is the study of the excision loci, in which parasitic nematodes are also the only multicellular organisms in which it was studied. In *A. lumbricoides*, it was found that specific Chromosome Breakage Regions (CBRs) delimit the locations of the germline restricted DNA, which are subsequently replaced with 2 to 4 kb of telomeric TTAGGC repeats. The specific site of DNA excision in a particular CBR was found to be variable in different individuals: the breakage can take place at several sites in a region of up to 6.5kb long. The excised DNA, which is subsequently degraded during the next cell cycle, has also been observed to have telomeric sequences added to its ends. Telomeric addition, as found following DNA excision, seems to require one to six nucleotides corresponding to the RNA template of the telomerase complex at the addition site (Müller et al. 1991; Jentsch et al. 2002). Telomerase itself is upregulated prior to CD, and remains so until the last pre-diminutive cells undergo CD (Magnenat et al. 1999). Fluorescent in situ hybridizations (FISH) demonstrate that the telomeres at chromosome extremities prior to CD are discarded along with the satellite DNA, and

new telomeres are only detectable from the third cell cleavage post-CD (Niedermaier and Moritz 2000).

CBRs were also found in *P. univalens*, but with some differences. Out of 7 CBRs studied in *A. suum*, 6 have homologous counterparts in *P. univalens*. CBRs 1, 2 and 3 occupy the same loci in *P. univalens*, but are shifted 1 to 2 kb to the side and the span of CBR1 in *P. univalens* is smaller (3.5 kb compared with 6.5 kb in *A. suum*). Three other CBRs present homology (two being partly homologous), and the accompanying excised regions in *P. univalens* are found at different loci in *A. suum*. The authors propose that this demonstrates genomic rearrangement. In general, the CBRs do not seem to be flanked by any common motif, or requirement sequence-wise, that would specifically identify those regions. Moreover, the corresponding CBRs present only 50 to 80% conservation (Bachmann-Waldmann et al. 2004).

The elimination of repeats is not characteristic only to *A. suum*: in *P. equorum*, it has also been found that the DNA lost during CD is mainly (if not entirely) of repetitive nature (Moritz and Roth 1976; Roth 1979). However, the repeats seem do not demonstrate clear bands on agarose gels, but instead produce homogenous smears when digested with a few different restriction enzymes; this was proposed to indicate irregular spacing of the restriction sites within the satellite DNA (Teschke 1985). Repeated sequences were found, and were described as to be composed of two different repeats: a pentamer and a decamer, both present in the form of several variants (accounting for the homogenous smears on agarose gels mentioned previously). The former repeat, pentameric, is defined as having the TTGCA sequence as the leading motif. Although it occurs in small tandem clusters,

the pentameric blocks are composed of a dispersion of lead, major and minor variant fractions, with frequencies within the analyzed DNA satellite fragments of 60%, 16% and 24%, respectively. While these pentameric repeats are largely eliminated from the germline, traces remain in the somatic genome, located at the chromosome ends. The satellite repeat fragments found in the somatic genome are also characterized by the same variant frequencies than in the germline (Teschke et al. 1991). The other component of the repeat satellites is a tandemly repeated decamer with a defined motif of TTTGTGCGTG. It is present in multiple variants which can be categorized under three main motif variants, organized in a randomized manner throughout the repeat fractions. As found by FISH studies, both repeats demonstrate polymorphism in respect to length and ratio of each repeat within the eliminated heterochromatic segments at a given loci (Niedermaier and Moritz 2000).

One particular nematode species represents an exception amongst the other diminutive nematodes. In *Strongyloides papillosus*, CD is involved sex determination of the free-living males. The female karyotype consists of two pairs of chromosomes, one larger than the other, while the karyotype of free-living males reflects that of the female by fragmentation of one the larger chromosomes, resulting in two smaller chromosomes and the elimination of a specific fragment (Triantaphyllou and Moncol 1977; Nemetschke et al. 2010). The CD found in *S. papillosus* not only differs in terms of differentiation mechanism, but also in the contents of the DNA that is lost. Rather, the diminished fragment of *S. papillosus* is homologous to the X chromosome of a closely related species, *Strongyloides ratti*.

Several genes were confirmed to be present in the eliminated fragment, contrarily to only a few genes in other diminutive nematodes (Triantaphyllou and Moncol 1977).

1.2 In hagfish:

In hagfish, the term chromatin diminution is replaced by chromatin elimination (CE), since in most hagfish the elimination of germ line restricted DNA includes both whole heterochromatic chromosomes and heterochromatic fragments of chromosomes. In contrast to nematodes and copepods, the occurrence of CD (or Chromosome Elimination, CE) in hagfish seems to be less of a mosaic: in the order of Myxinida, it was observed that CE is a generalized trait (Nakai et al. 1995). Typically, germ-line restricted heterochromatin in hagfish is located at interspersed loci and/or at chromosome ends, in addition to whole heterochromatic chromosomes (Nakai and Kohno 1987; Nakai et al. 1991). In particular species of hagfish, the eliminated DNA demonstrates length polymorphism, which also applies to the number of heterochromatic chromosomes discarded. This not only occurs between individuals but also between spermatogonia of the same specimen (Kubota et al. 1992; Kubota et al. 1994; Nakai et al. 1995; Kojima et al. 2010). However, some euchromatic DNA also seems to be eliminated during CE in some species, as determined by C-banding experiments (Kubota et al. 1994; Nakai et al. 1995; Kojima et al. 2010).

Molecular studies in hagfish have been applied to a greater number of different species than in nematodes or copepods. Using restriction enzymes in order to identify tandem arrays from genomic DNA, 12 distinct repeat units, identified from 5 different species, were found to be eliminated (at least partially, if not completely). The various repeats can range from 54bp to 180bp, and the number of species found to eliminate such repeats from their germinal genome amounts to 7. Some

repeats are unique to species in which they were found, and other repeats can be found in the germ-line of several species but in different proportions. Moreover, not all repeats are entirely eliminated during CE, and can thus be found in the somatic genome. All repeats found to be eliminated in hagfish species appear to be organized in ladder-like patterns when hybridized against partially digested genomic DNA. This suggests that they are organized in tandem arrays. Also, none seemed to posses any ORF, nor were any homologies found in the DDBJ, EMBL or Genbank databases at the time of the studies (Kubota et al. 1993; Kubota et al. 1997; Goto et al. 1998; Kubota et al. 2001; Kojima et al. 2010). One particular repeat was sequenced along with its homologs found in 5 different species and the sequences showed from 75% to 99% identity with the original sequence. Three direct repeats were observed in the sequences; they can be described as containing three subrepeats, and each seemingly derived from one motif. This putative original motif, which can also be found in other repeats, consists of 28bp in length and contains the conserved region of AGAAGTGAT from the 3rd to 11th position of the motif. The fact that the same eliminated repeats are found in several species but in different proportions suggests that these repeats were present in common the ancestors of the species that share the same repeats, and that further amplification was done after speciation (Nabeyama et al. 2000). To date, all discovered repeated elements in hagfish have been found using the comparative analysis of restriction enzymes on germ and somatic lines.

1.3 In copepods:

As in parasitic nematodes and hagfish, CD in copepods is involved with the differentiation of germ-line cells into somatic cell precursors during the early development of the embryo. So far, reports indicate that CD is found in 17 species and is absent from 13 other species (Dorward 1997; Rasch et al. 2004; Semeshin et al. 2011). The phylogenic occurrence of CD presents a mosaic in copepods, as contrasting occurrence can take place in closely related species. Copepods are the organisms in which we can find the highest amounts of discarded DNA: a reduction of up to 96.4% can be seen between germ-line and somatic cells of *Cyclops kolensis* (Semeshin et al. 2011). The germ-line restricted DNA can be found at different positions along chromosomes: at chromosome ends, at both chromosome ends and at the centrosome, or interstitially dispersed along the chromosomes (Beermann 1977). As in parasitic nematodes and hagfish, copepods also demonstrate polymorphism in term of the length of germ-line restricted heterochromatin, which disappears after the programmed DNA excision of CD. However, this has only been demonstrated inter-individually (Beermann 1977; Chinnappa 1980). Moreover, specificities in the mechanisms of CD can differ in different populations of copepods: *Cyclops strenuus* and *C. kolensis* are two species in which different populations differ in terms of timing of diminution, amount of DNA discarded and the location where the DNA is discarded (at the equatorial plate or at the mitotic poles). These populations consist of the German and Russian populations of each species (Grishanin and Akif'ev 2000). Other particularities include the excision of circular and linear chromatin structure, as observed in two

species of the Cyclops genera, and DNA endoreduplication prior to gametogenesis in *Mesocyclops edax* and *C. kolensis* (Beermann and Meyer 1980; Beermann 1984; Rasch and Wyngaard 2001; Wyngaard et al. 2011).

As for molecular studies in copepods, very little has been found in comparison to nematodes or hagfish. One reason for this is that it is time consuming to obtain even small amounts of DNA to work with (Drouin 2006). As a result, studies performed with the use of several restriction enzymes in order to compare tandem arrays between germ-line and somatic DNA cannot be realistically reproduced with copepods, as done with parasitic nematodes and hagfish. To circumvent this limitation, molecular studies to date have employed methods that consist of random DNA amplification using degenerate primers. One study used partially degenerate primers (DOP-PCR) for the amplification of microdissected granules containing the DNA eliminated during CD of *C. kolensis*. The resulting analysis of these sequences revealed that about half of these could be put together into groups of homologous sequences consisting of 2 to 4 sequences. A few direct repeats, tandem repeats and inverted repeats were also found, along with poly (A) and poly (C) tracks and regions presenting homology to the CELE 46B repeat of *Caenorhabditis elegans*, but none of these are mentioned to be ubiquitous. Something else worth noting is that none of these sequences seemed to code for any known protein sequence at the time (Degtyarev et al. 2002). The same team repeated this approach at a later date. As in the previous study, sequences obtained by DOP-PCR were composed of groups sharing a degree of homology consisting of 2 to 9 sequences. These sequences composed 63% of the overall sequences and

contained an abundance of short repeats, more precisely between 1 and 3 different repeat families per sequence. A total of 24 different repeat motifs were found in these sequences, which the authors describe as being submotifs, part of larger motifs. Again, coding regions could not be found in the obtained sequences at the time of study (Degtyarev et al. 2004).

In the species *M. edax*, approximately 90% of DNA is lost during the process of CD (Rasch and Wyngaard 2001). Unlike molecular studies in *C. kolensis*, *M. edax* was studied by using by RASL, or Randomly Amplified Shotgun Libraries which uses completely randomized 10-mers as primers (Rohwer et al. 2001). Drouin (2006) compared amplified libraries from germ-line and somatic DNA, and observed the reduction of three tandemly repeated sequence arrays consisting of CAAATAGA, CAAATTAAA, and CA repeats. These were found in a proportion of 96% of total pre-diminution sequences, while the proportion of the sequences containing these repeats was reduced to 47% after diminution (Drouin 2006). This finding differs from what was found in *C. kolensis* in terms of the specificity of the sequences that are be eliminated.

Another approach consisted in the amplification of sequences between $(GA)_n$ microsatellites in order to determine their presence before and after chromatin diminution. The amplification was done with $(GA)_9C$ primers on *C. kolensis* embryos before and after CD, and the products were cloned in order to make small libraries. Out of 5 randomly selected clones, only one of these clones seems to not be present in post-diminution DNA (Zagoskin et al. 2008). Otherwise, the presence of several clones of eliminated sequences in *C. kolensis* was tested by PCR on the

somatic genome of different populations of this species, along with a closely related non-diminutive copepod, *C. insignis*. Most sequences were found in all specimens, but some particular sequences were not found in either a population of *C. kolensis* or from *C. insignis* (Zotkevich et al. 2008).

So far, molecular studies of CD in copepods have attempted to circumvent the limitation of small DNA amounts using randomized amplification. However, some aspects remain uncharacterized and would benefit from additional information. An attempt to characterize this programmed genomic rearrangement using restriction enzymes may contribute to further the understanding of CD amongst copepods, and to a further extent, provide a study that can be comparable to those of CD in nematodes or hagfish.

1.4 Other organisms:

Chromatin diminution, or variants of such chromatin rearrangement, can also be found in a variety of other organisms, which include ciliates, insects, marsupials and the zebra finch. Those rearrangement events can resemble those of CD as found in nematodes, hagfish and copepods in several aspects. However, they also differ in fundamental ways.

Ciliated protozoa undergo two simultaneous processes of genomic rearrangements. One includes the fragmentation of the micronucleus (which acts as a germinal nucleus) at determined Chromosome Breakage Sequence (CBS) sites, and produces macronuclear chromosomes. At the same time, the process of DNA elimination results in the elimination of DNA sequences (Internal Eliminated Sequences, or IES), found inside the chromosome fragments, which is followed by the ligation of the flanking sequences. The resulting macronuclear chromosomes are then amplified, and the final product is the macronucleus, which has role comparable to the soma. DNA elimination in ciliates can involve the elimination of up to 95% of the micronuclear genome (Kloc and Zagrodzinska 2001; Mochizuki 2012). It was observed that the IESs are targeted in accordance to their heterochromatic structure, which in turn was demonstrated to require the RNA silencing machinery for its formation (Mochizuki 2012). Interestingly, a transposase was associated with the endonucleolytic activity responsible for the DNA cleavages made during DNA elimination in a species of *Paramecium*. That transposase was named PiggyMac, as it is thought to be domesticated transposase from a *piggy-Bac* transposon (Baudry et al. 2009). One particular proposition dictates that DNA elimination in ciliates may

have a role in silencing transposable elements by their removal from the transcriptionally active macronucleus (Schoeberl and Mochizuki 2011).

The complex behaviour of elimination in sciarid flies has been described in numerous studies. Using *S. coprophila* as a general example, differentiation into somatic cells is a procedure that includes the elimination of three germ line restricted chromosomes (L chromosomes) at the 6^{th} nuclear cycle. This is followed with sex determination at the 7^{th} to 9^{th} nuclei cycle, where females eliminate one X chromosome and males eliminate two (Metz and Schmuck 1931; Dubois 1932; Schmuck and Metz 1932; Metz 1934; De Saint Phalle and Sullivan 1996). Germ cells of the embryo in both sexes eliminate one L chromosome and one X chromosome, both of paternal origin (Rieffel and Crouse 1966). Contrasting from CD in nematodes, copepods and hagfish, meiosis in sciarid flies also exhibit unique chromosome elimination events in sperm. During two meiotic cycles in males, the whole paternal complements of the autosomes (diploid homologues) and X chromosomes are eliminated, while only one copy of the maternal autosomes and L chromosomes are lost. The final product delivers one copy of the maternal autosomes, along with a copy of each parental L chromosomes and the remaining two maternal X chromosomes of (Metz 1926; Schmuck and Metz 1932; Du Bois 1933; Smith-Stocking 1936; Crouse 1966). The female meiosis, however, follows normal meiotic cleavage (without further chromosomal elimination) and yields an egg containing an haploid complement (Schmuck and Metz 1932; Rieffel and Crouse 1966). Most of these events are comparable (with exceptions) in related

species: *S. ocellaris* (Berry 1939; Berry 1941), *S. impatiens* (Crouse et al. 1971), *S. similans* (Metz 1928), and *Trichosia pubescens* (Amabis et al. 1979; Fuge 1994).

The mechanism of chromosome elimination in sciarid flies is well described, although still not thoroughly understood. Some differences sets apart the phenomenon of chromosome elimination in sciarid flies from that observed in nematodes, copepods and hagfish. For one, no fractionation of the eliminated chromatin is observed. Instead, whole chromosomes are eliminated. Another major difference is the presence of two distinct elimination events, one for germ/soma differentiation and another during spermatogenesis. The elimination of the X chromosome, as differentiation mechanism of pre somatic cells, seems dependent upon the presence of a "controlling element", located at the proximity of its centromere (Crouse 1960). Another major difference is the migration of the chromosomes (which are to be eliminated) to a pole using a monopolar spindle, which later extrudes from the nucleus through a budding mechanism (Fuge 1994; Esteban et al. 1997). Two other families of diptera also exhibit chromosome elimination processes similar to that in sciarid flies: in cecidomyiids (Metcalfe 1935; Geyerduszynska 1959; Nicklas 1959; Bantock 1961; Geyerduszynska 1961; Matuszewski 1962; Bantock 1970; Panelius 1971; Stuart and Hatchett 1988; Jazdowskazagrodzinska et al. 1992) and in chironomids (Bauer 1970; Staiber 1988; Staiber 1989; Staiber 1991; Staiber et al. 1997; Staiber 2000; Staiber and Schiffkowski 2000; Staiber 2006), which both demonstrate characteristic differences.

The remaining organisms include marsupials and the zebra finch. Marsupials are also found to exhibit chromosome elimination based on sex differentiation. This

occurs with either the Y chromosome or the X chromosome (male and female, respectively). Moreover, these events tend to be tissue specific instead of dependent upon somatic differentiation (Walton 1971; Close 1979; Murray et al. 1979; Close 1984; McKay et al. 1984; Watson et al. 1998). In the zebra finch (*Taeniopygia guttata*), a similar process to what is observed in dipterians and marsupials takes place: a germ-line restricted chromosome is eliminated during spermatogenesis and oogenesis (in which it is heterochromatic or euchromatic, respectively), and appears to be absent from select somatic tissues (Pigozzi and Solari 1998; Itoh and Arnold 2005; Pigozzi and Solari 2005; Itoh et al. 2009; Goday and Pigozzi 2010; Schoenmakers et al. 2010). It has been found that this chromosome possesses a certain amount of repetitive sequences, although its composition has not been described (Itoh et al. 2009).

1.5 Genome size:

So far, the purpose of CD has remained obscure. Since CD has demonstrated to eliminate large amounts of DNA (up to 96.4% in *C. kolensis*, Semeshin et al. 2010), it is inevitable to suppose that it acts as to regulate genome size. One of the most advocated effects of an increased genome size is the concurrent increase in cell size. Other suggested correlations to genome size include: duration of cell division, development rate and total body size (Dufresne and Jeffery 2011). Cyclopoid copepods in particular have been found to be selectively constrained as per their genome size (Rasch and Wyngaard 2006). In turn, their genome size seems to correlate loosely with body size, but demonstrates a more solid inverse correlation with development rate (Wyngaard et al. 2005). It has also been suggested that, as genome size seems to affect embryonic size in marine copepods, a larger genome may positively affect survival rate of embryos (Wyngaard et al. 1994). Thus, it was speculated that CD (as observed in cyclopoid copepods) is in fact an adaptive trait that confers both the advantages of a larger embryo and of a faster development rate (Wyngaard and Gregory 2001).

1.6 Transposable elements:

Organisms that undergo CD eliminate DNA which is found in a heterochromatin state, as described in respective sections. In heterochromatin, one can find the following: tandem repeats (such as found to be eliminated in parasitic nematodes, hagfish and copepods) and transposable elements (Charlesworth et al. 1994). Considerable portions of genomes can be occupied by sequences derived from said transposable elements: for example the human genome is composed of approximately 45% of such sequences (Lander et al. 2001), the genome of maize is riddled with transposable elements in a proportion of nearly 85% (Schnable et al. 2009), and an estimated 9.4% of the genome of *Daphnia pulex* is derived from such DNA (Colbourne et al. 2011). Moreover, transposable element proliferation is thought to be partly responsible for the variation of genome size in closely related species: for example two of the largest diploid species in the *Oryza* genus in term of genomic size have become so as a result of transposable element proliferation (Zufall et al. 2005). As such elements can be found in heterochromatin, some particular transposable elements even possess the ability to specifically integrate into heterochromatin. This ability is conferred by an extra portion at the C-terminal end of their integrase domain, which recognizes H3 K9 methylation of histones (Gao et al. 2008). Taking into account these facts, it would not be a surprise to find copies of transposable elements amongst the DNA that is excised during CD. It was in fact demonstrated in nematodes (described in the nematode section), but no occurrence has yet been found in copepods or hagfish.

1.7 Hypothesis and Methodology:

In *M. edax*, it was previously demonstrated that CD results in a reduction of specific satellite repeats (Drouin 2006). However, it was not determined if DNA of other nature is also reduced in proportion, or even eliminated. Taking into consideration that CD in copepods results in the elimination of the majority of the heterochromatic DNA, and that heterochromatin usually contains transposable elements (Charlesworth et al. 1994), the hypothesis of this experiment was that CD in *M. edax* results in a reduction of total TEs. The experiment was designed in order to obtain genomic sequences other than the previously described satellite DNA. That is, genomic DNA was digested using *BamHI*, a restriction enzyme that does not cut in the satellite DNA. *M. edax* genomic DNA was obtained from the following specimens: embryos at the 8-cell stage or earlier (for pre-diminution DNA) and antennas (for post-diminution DNA). Using the sequences obtained from these procedures will allow us to compare libraries derived from the pre and post-diminution genomes. Finally, by describing the TE fraction of both libraries, we will be able to determine if the previously stated hypothesis is correct or not.

2. Chapter 1: Chromatin diminution in *Mesocyclops edax* (Crustacea, Copepoda): similarity of the pre- and post-diminution euchromatic genomes.

Christian McKinnon and Guy Drouin

Abstract Chromatin diminution, defined as the developmentally regulated process of partial genome excision, is a process found in numerous taxa such as copepods, nematodes and hagfish. In *Mesocyclops edax*, up to 90% of genomic DNA is lost. Previous studies have shown that the eliminated DNA contained highly repetitive heterochromatic sequences. Here, we digested DNA from germ line and somatic line cells with *BamHI* and produced small libraries of clones to determine whether the eliminated DNA also contained euchromatic sequences. Comparative analyses reveal no dramatic decrease in low copy numbered sequences such as transposable elements, but rather a surprising homogeneity in sequences between the two libraries. In fact, both libraries demonstrate surprising similarity with eDNA from *Cyclops kolensis*, as determined by repeating our analyses on these sequences. Consequently, we suggest that *M. edax* eliminates portions of euchromatic DNA, in addition to the previously characterized satellite sequences.

2.1 Introduction:

The process of chromatin diminution (CD), the elimination of DNA during the differentiation of early embryonic cells into pre-somatic cells, has been known for over a century. It was first observed in a parasitic nematode of horses, *Parascaris equorum* (later renamed *Parascaris univalens*) (Boveri 1887). It has since been identified in several parasitic nematodes, and has been well characterized in *P. univalens* and *Ascaris suum* (Tobler 1986; Muller et al. 1996; Müller and Tobler 2000). Similar processes have also been observed in a variety of organisms, such as copepods, hagfish, ciliates, insects, marsupials and the zebra finch (Beermann 1977; Nakai et al. 1991; Nakai et al. 1995; Dorward 1997; Pigozzi and Solari 1998; Goday and Esteban 2001; Kloc and Zagrodzinska 2001; Mochizuki 2012). In copepods, chromatin diminution has been observed in 17

species, while it is absent from at least 13 other species (Leech and Wyngaard 1996; Dorward 1997; Rasch et al. 2004; Semeshin et al. 2011).

A considerable number of cytological studies have shown that the discarded chromatin is usually composed of heterochromatin, and copepods are no exception (Beermann 1977; Chinnappa 1980; Kloc and Zagrodzinska 2001; Redi et al. 2001; Rasch and Wyngaard 2008). In ciliates, heterochromatin formation requires the RNAi machinery, and DNA elimination is also ultimately dependent upon it (Mochizuki 2012). However, C-banding experiments, indicative of heterochromatin structure (Pimpinelli et al. 1976), have shown instances where not only C-band positive, but also C-band negative chromatin is eliminated during chromatin diminution (Standiford 1989; Kojima et al. 2010). Molecular studies have confirmed that tandemly repeated DNA, known to be a major constituent of heterochromatin (Charlesworth et al. 1994), are indeed eliminated during chromatin diminution in nematodes (Roth and Moritz 1981; Streeck et al. 1982; Teschke et al. 1991; Niedermaier and Moritz 2000) and hagfish (Kubota et al. 1993; Kubota et al. 1997; Goto et al. 1998; Kubota et al. 2001; Kojima et al. 2010). On the other hand, sequences other than tandem repeats have been found to be discarded in nematodes, such as single copy genes, some of which exhibit transcriptional activity (Etter et al. 1991; Etter et al. 1994; Spicher et al. 1994), and Tas retrotransposons (Aeby et al. 1986; Felder et al. 1994). Although the eliminated chromatin is composed in major part of heterochromatic repeated sequences in these organisms, sequences of a different nature seem also to be eliminated, and remain only vaguely described.

Molecular studies on the chromatin diminution of copepods have remained sparse. The main reason for this is that it is difficult and labour-intensive to obtain respectable amounts of pre-diminutive DNA because female egg sacs are very small and contain only a few dozen pre-diminutive embryos. This limitation renders any experimental design such as was used in nematodes and hagfish, i.e. a comparative restriction analysis of the germline genomes using several restriction enzymes, highly impractical. As such, previous molecular studies in copepods have resorted to random (Drouin 2006) or semi-random (Degtyarev et al. 2002; Degtyarev et al. 2004) amplification in order to circumvent this limitation. In *Mesocyclops edax*, a comparative RASL (Randomly Amplified Shotgun Libraries) study of pre and post-diminution DNA (amplified by randomly generated primers) demonstrated a sharp decrease in short tandem repeats (Drouin 2006). In *Cyclops kolensis*, two studies reported the contents of cytoplasmic granules containing DNA eliminated during chromatin diminution, as obtained by amplification using partially degenerated primers (DOP-PCR). They reported an assortment of sequences, of which half corresponded to homologous groupings, and the presence of short direct repeats (Degtyarev et al. 2002) and arrays of motifs that they described as a mosaic structure of sub-motifs within motifs (Degtyarev et al. 2004).

In addition to various repeat fractions, transposable elements (TEs) can also be found in heterochromatin (Charlesworth et al. 1994). The human genome itself is composed of approximately 45% of sequences derived from transposable elements (Lander et al. 2001), whereas this portion is estimated to be 9.4% in the

genome of *Daphnia pulex* (Colbourne et al. 2011). Some retrotransposons possess the ability to integrate specifically into heterochromatin. This capacity is due to the fact that their integrase domain contains an additional portion at the C-terminus, called the chromodomain, which recognizes histone H3 K9 methylation (Gao et al. 2008).

M. edax discards close to 90% of its total DNA contents during chromatin diminution, and the discarded DNA is believed to be mainly heterochromatin (Rasch and Wyngaard 2001). Therefore it is not unreasonable to think that transposable elements might also compose a considerable part of the eliminated DNA of *M. edax*. Here, we compare pre- and post-diminution DNA libraries made from BamHI-digested *M. edax* genomic DNA. The use of BamHI to generate these libraries should not be hindered by DNA methylation, as DNA methylation is thought to be rare in invertebrates (Bird and Taggart 1980). Since the *M. edax* satellite DNA sequences do not contain the restriction site of BamHI, these libraries should therefore represent the euchromatic DNA of this species. Our results show that both the transposable element and non-transposable element content of these two libraries appear similar. We also compare our results with those obtained from *C. kolensis*, and find no discernable difference. This suggests that *M. edax* eliminates portions of euchromatic DNA, in addition to the previously characterized satellite sequences.

2.2 Methods:

2.2.1 Specimens: Specimens were harvested in August 2002 at Lac Shenandoah (Rockinham County, Virginia) and were subsequently identified by Grace Wyngaard, (Department of Biology, James Madison University, Harrisonburg, Virginia). These were kept in 95% ethanol at -20^{o}C until the time of DNA extraction. Eggs sacs with embryos at the 8-cell stage or less were put aside for pre-diminution DNA extraction. For post-diminution DNA, antennas of adult *M. edax* specimens were dissected under a dissection microscope.

2.2.2 DNA extraction, cloning and sequencing: DNA extraction was performed using the Wizard DNA Purification Kit (Promega) using a modified protocol adapted to copepod specimens as described in Sawabe (2005). To maximise the DNA yield from our small samples, glycogen was also used as a carrier during the DNA precipitation step and longer centrifugation times were used. Genomic DNA samples were then digested with the BamHI restriction enzyme.

A mixture of digested DNA and plasmid (pUC19 pre-digested with BamHI, Invitrogen) in a DNA:vector ratio of 3:1 were then ethanol precipitated using ammonium acetate and Polyethylene glycol (PEG-8000), as described in Fregel et al. (2010). Ligations were performed by resuspending directly the pellets of the precipitations in ligation mixes containing 2.5U of ligase in a total volume of 10µL and incubating at room temperature for 2 hours. The resulting ligations were then transformed into *Escherichia coli* "*E. cloni*" chemically competent cells (Lucigen).

The resulting white colonies were then screened for inserts by PCR using M13 forward and reverse primers.

Clones longer than 500 bp were sent to the McGill University/Genome Québec Innovation centre (Montréal, Québec, Canada) to be sequenced using the Applied Biosystems 3730xl DNA Analyzer. The clones were sequenced using the M13 forward primer. Each sequence was manually verified in accordance to its chromatogram as to insure sequence quality.

2.2.2 Sequence analysis: Transposable elements identification was performed using tBlastx (standalone package version 2.2.25+, Camacho et al., 2009), with the Transposable Element library from Repbase Update (release 17.01; Jurka et al., 2005) as the target database. Sequence similarity groupings were identified using Blastn (Camacho et al. 2009) and verified using the ClustalW accessory application included in Bioedit (Thompson et al. 1994; Hall 1999). The sequences were analyzed for tandem repeats using the 4 following programs: BWtrs 1.03.05 (Pokrzywa and Polanski 2010), mreps 2.5 (Kolpakov et al. 2003), Phobos 3.3.12 (Mayer 2006-2010) and Tandem Repeats Finder 4.04 (Benson 1999), using the default settings. Only the repeats found by at least two of these four programs were analyzed further.

The similarity with Gypsy-17_DPu related sequences was determined using sequence alignments made with MUSCLE 3.8.31 (Edgar 2004). Internal gaps were removed using the “gappyout” function of trimAl v1.2rev59 (Capella-Gutierrez et al. 2009) and edges were trimmed manually. Maximum likelihood

trees were generated using PhyML 3.0 (Guindon and Gascuel 2003) based on the substitution models selected using jModelTest 0.1.1 (Posada 2008).

Chromodomain containing sequences were identified using tBlastn (standalone version 2.2.25+, Camacho et al. 2009) and the HMM generated consensus amino acid sequence of retroelement chromodomains (obtained from the GyDB gypsy database, http://gydb.org/index.php/Collection_MRC)

Motif analysis was performed using MEME software on their online web server (http://meme.sdsc.edu/meme/intro.html, version 4.8.1; Bailey 1994). Most of the parameters used were set on default, except the occurrence of motifs, which was set on “any number of repetitions” per sequence, and the maximum width of motifs, which was set to the largest value accommodated by the software (300). Motifs found in the pre-diminution sequences queried against the post-diminution sequences using the FIMO software, included in the meme suite (http://meme.nbcr.net/meme/cgi-bin/fimo.cgi) on default settings.

2.3 Results:

2.3.1 Sequences:

Cloning experiments yielded a total of 70 sequenced clones from pre-diminution eggs and 128 sequenced clones from antennas (which are assumed to be somatic tissue). Prediminution clones are named Nebl-XX (for New Eggs BamHI Ligation) and the postdiminution clones, Abl-XX (Antenna BamHI Ligation). Each library contained 32501 bp and 78173 bp for pre-diminution and post-diminution sequences, respectively.

In our sequences, we were able to find and describe: TE-related sequences, groups of homologous sequences, and sequences similar to protein coding sequences (comparison with online databases using Blastx, http://blast.ncbi.nlm.nih.gov/). Additional analyses were made to study the presence of tandem repeats and chromodomains. All aforementioned analyses were aimed at comparing our pre-diminution library to our post-diminution library. Similar analyses were also performed using sequences previously found to be eliminated from the genome of *C. kolensis* (Degtyarev et al. 2004) in order to compare them with our sequences.

2.3.2 Transposable elements:

To identify transposable element related sequences, both pre and post-diminution libraries were used as query with the standalone version of tBlastx 2.2.25 (Camacho et al. 2009) against the entire Transposable Element database of Repbase Update (release 17.01; Jurka et al., 2005). Table 1 details the respective transposable element portion of each library and lists the different elements recognized in our sequences along with their respective occurrence. As shown, the majority of the TE related post-diminution clones are similar to a particular 8720bp long gypsy LTR retrotransposon found in *D. pulex* (Gypsy-17_DPu-I_Daphnia pulex). In our post-diminution sequences, 28 out of the 35 TE related sequences are similar to this element, while a smaller proportion, 6 out of 10 TE related sequences, are similar to this element in our pre-diminution sequences. The detailed list of these results is shown in Tables S8 and S9. Results found with Repbase Update are in majority concurrent with those found with an online Blastx search (http://blast.ncbi.nlm.nih.gov/) on the non-redundant protein sequences database (nr), listed in Tables S5 and S7. Taking into account both the Repbase and the online Blastx search, we observe no decrease in the overall portion of recognizable TEs; rather, it occupies a significantly larger portion ($p<0.05$) in the post-diminution sequences.

Table 1. Composition and occurrence of transposable elements in pre-diminution and post-diminution sequences.

Transposable element:	Occurrence in pre-diminution clones	Occurrence in post-diminution clones	Total occurrences
LTR retrotransposons (Class)			
BEL (Superfamily)	**4**	**2**	**6**
BEL1-I_SM_BEL_Schmidtea mediterranea	1	0	1
BEL-4-I_NV_BEL_Nematostella vectensis	1	2	3
BEL-26_AA_BEL_Aedes aegypti	2	0	2
Copia (Superfamily)	**0**	**1**	**1**
Copia-130_AA-I_Copia_Aedes aegypti	0	1	1
Gypsy (superfamily)	**6**	**31**	**37**
Gypsy-7_PSo-I_Gypsy_Phytophthora sojae	1	0	1
Gypsy-17_DPu-I_Gypsy_Daphnia pulex	5	28	33
Gypsy-7_PGr-I_Gypsy_Puccinia graminis f. sp. tritici CRL 75-36-700-3	0	1	1
Gypsy-28_Bra-I_Gypsy_Brassica rapa	0	1	1
Gypsy-43-I_NV_Gypsy_Nematostella vectensis	0	1	1
DNA transposons (Class)			
Ginger (superfamily)	**0**	**1**	**1**
Ginger2-1_LS_Ginger2/TDD_Littorina saxatilis	0	1	1
Total:	**10**	**35**	**45**
Percentage of respective library (sequences)	15.7%	25.8%	
Percentage of respective library (total nucleotides of sequences)	20.0%	27.5%	
Percentage of respective library (nucleotides aligned)	9.2%	15.8%	
GC% (library)	42.80%	43.31%	
GC% (TEs)	44.42%	47.04%	
GC% (non TE sequences)	42.39%	41.89%	

To determine whether there are TEs specific to pre-diminution DNA, phylogenetic trees were built based on the sequences similar to the Gypsy-17_DPu-I element from both libraries. Trees based on sequences of the reverse transcriptase domain (Figure 1) or the integrase domain (Figure 2) show that the pre-diminution Gypsy-17_DPu-I sequences are not different from those found post-diminution. In our reverse transcriptase alignment, group 21 and 26 each presented distinct patterns of substitutions. The resulting maximum likelihood tree (Figure 1) illustrates this distinction. As shown, the pre-diminution sequences (Nebl-15, Nebl-23 and Nebl-89) are more closely related to the post-diminution sequences of the same group, rather than pre-diminution sequences of the other group. On the other hand, our integrase alignment (Figure 2) allows us to postulate that the post-diminution sequences of group 25 are more similar to the pre-diminution sequence (Nebl-24) than they are to other post-diminution sequences of group 23. Hence, we found no Gypsy-17_DPu-I variety unique to the germ-line, and the pre-diminution TE related sequences do not appear more divergent than in the post-diminution sequences.

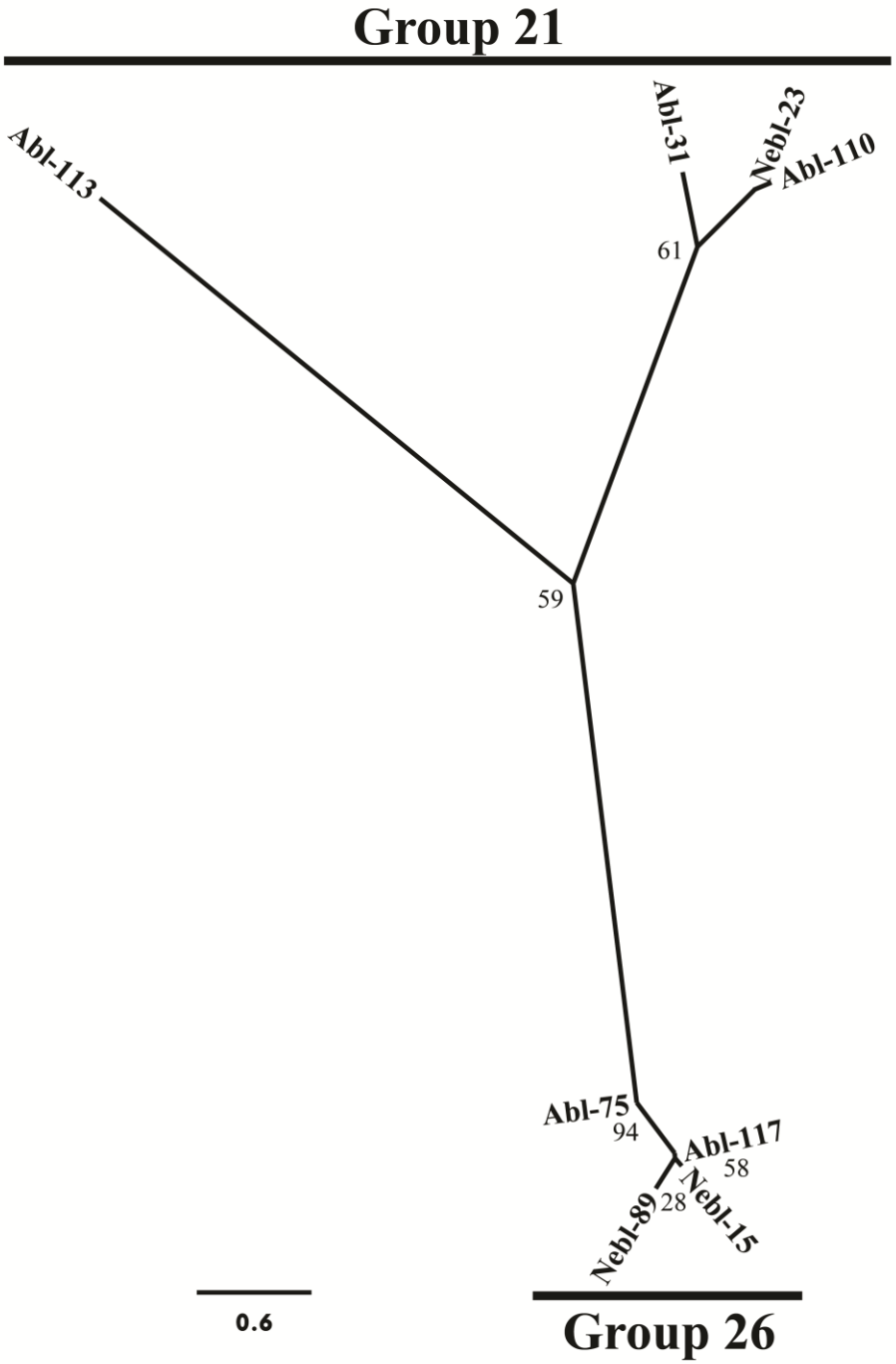

Figure 1. Maximum likelihood tree (1000 bootstrap replicates) of an alignment from select sequences overlapping the reverse transcriptase domain. Sequences belonged to groups 21 and 26, as indicated. The alignment includes 195bp overlapping the Gypsy-17_DPu-I element. Bootstrap percentages are indicated at the nodes.

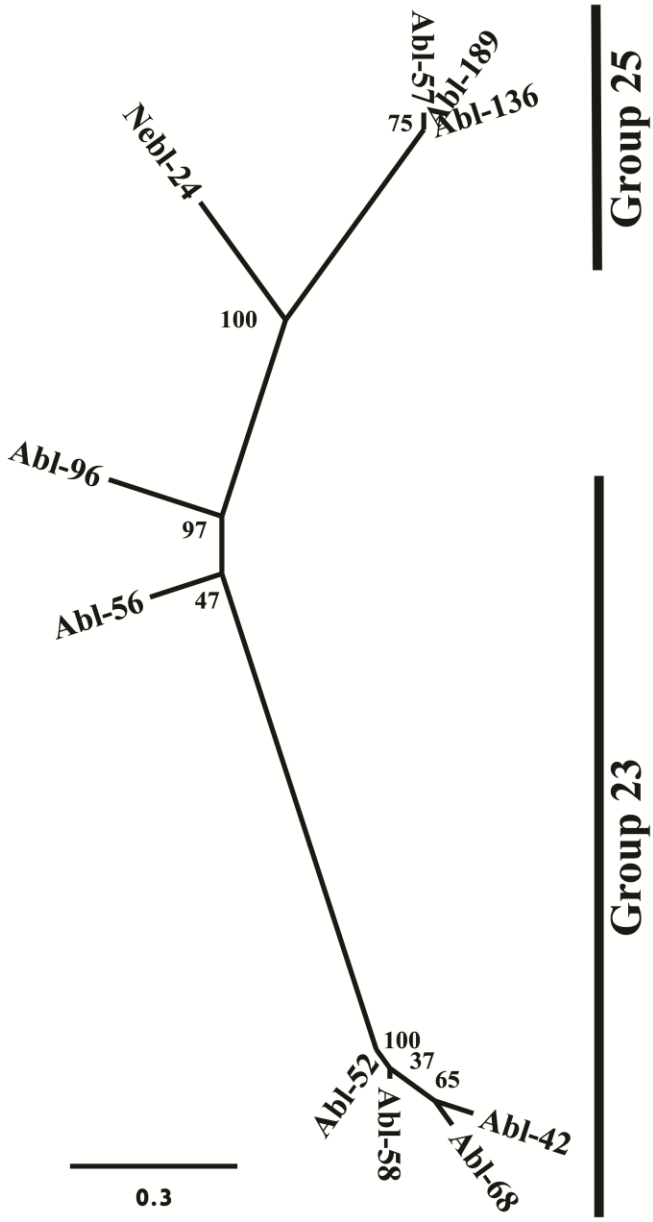

Figure 2. Maximum likelihood tree (1000 bootstrap replicates) of an alignment from select sequences overlapping the integrase domain. Sequences belonged to groups 23 and 25, as indicated. The alignment includes 208bp overlapping the Gypsy-17_DPu-I element. Bootstrap percentages are indicated at the nodes.

2.3.3 Chromodomains:

To determine whether some of our TE related sequences possess a chromodomain, we compared them with a consensus chromodomain sequence. Using tBlastn (Camacho et al. 2009), we could not find any similarity with the consensus chromodomain amongst our sequences. However, we did find similarity in one TE (Gypsy-7_PSo-I from *Phytophthora sojae)* with which one of our sequences was in turn found similar to (Nebl-09). As such, it appears that the possession of a chromodomain is not a feature characteristic of the TEs found in any of our libraries of *M. edax* genomic sequences.

2.3.4 Similar sequence groups:

Out of the total 198 sequences, 49 different sequences (excluding TE-derived sequences) presented a certain degree of similarity with other sequences. The 21 total groups could be classified into 3 different categories: 8 groups contain exclusively pre-diminution sequences, 10 groups containing only post-diminution sequences, and finally the remaining 3 groups contain sequences from both libraries (Supplementary tables S1, S2 and S3).

A comparison between the two libraries was made by analyzing the three categories of groups. Using identity between similar sequences from each group, the average sequence identity (%) was calculated for each category (Table 2). This average was not found significantly different between any of the categories ($p > 0.05$).

Table 2. Groups of similar sequences.

Type of group	Groups	Sequences per group	Total sequences	Sequence identity (average, ±SD, n)
Pre-diminution	8	2 to 3	19	91.3% ±13.6, n=14)
Post-diminution	10	2 to 4	23	89.6% ±13.7, n=17)
Both libraries	3	2 to 3	7	95% ±10.6, n=5

2.3.5 Protein coding sequences:

Another facet of our analysis related to the contents of protein coding sequences between the two libraries. Several of our sequences were found similar to various proteins, as determined using an online Blastx search (http://blast.ncbi.nlm.nih.gov/) on the non-redundant protein sequences (nr) database (Supplementary Tables S4 and S6). In our prediminution sequences, 10 sequences found similar protein sequences, compared with 25 in our post-diminution sequences. These numbers do not represent a significant difference over the total number of sequences in their respective libraries ($p > 0.05$). Furthermore, several of our sequences demonstrated similarities with hypothetical proteins found in the genome of *D. pulex*. Again, the composition of these sequences was not significantly different from one library to the other ($p > 0.05$). Overall, we could not observe any particular difference between the two libraries in terms of protein coding sequences.

2.3.6 Tandem repeats:

Few short tandem repeats were found in both libraries, mostly consisting of various di and tri-nucleotides and other micro and mini satellites. These repeats increase from 1.44% to 1.99% of the pre and post-diminutive libraries, respectively. However, no notable differences could be discerned between the two libraries in term of repeat composition. The tandem repeats previously found by RASL (Drouin 2006) were not found, as expected from our experimental design.

2.3.7 Comparison with C. kolensis:

In order to have a broader view of CD in copepods, we compared our findings with the data available from *C. kolensis* (Sequences AY533039-AY533099 in GenBank; Degtyarev et al. 2004). We also used MEME to compare our sequences with those eliminated from the *C. kolensis* genome.

First, we analysed the sequences eliminated from the *C. kolensis* genome (eDNA) in terms of TEs, similarity groupings, protein coding sequences, tandem repeats and chromodomains, all in the same manner as we analysed our own sequences. Amongst their sequences, 10 were found similar to TEs sequences (Supplementary table S11). This number is not different from the proportion of TE sequences in our pre or post-diminution libraries ($p > 0.05$). None of the *C. kolensis* eDNA sequences had significant similarity with the consensus retroelement chromodomain sequence. However, similarity was found in a TE (Polinton-3-SP element from *S. purpuratus*), with which one of the eDNA

sequence (CkD18) was found similar to, as with our own sequences. Out of the 61 available sequences, 37 could be put into 10 groups of 2 to 10 sequences based on sequence similarity. As for protein coding sequences, similarities were found for 7 of the eDNA sequences from *C. kolensis*, mostly to hypothetical proteins in the *D. pulex* genome (Supplementary table S10). The portion of protein coding sequences is also not different from any of our libraries ($p > 0.05$). Lastly, tandem repeat contents were comparable to what was found in our pre-diminution library. Such repeats composed 1.6% of the *C. kolensis* eDNA sequences. The tandem repeats found were a variety of di and tri-nucleotides, along with mini and microsatellites (results not shown). In general, the contents of the eDNA of *C. kolensis* appear on par with the contents that we found in our pre- and post-diminutive sequences.

We also compared our pre and post-diminution sequences with the eDNA sequences from *C. kolensis* (Degtyarev et al. 2004). Using blastn (Camacho et al. 2009), no similarities were found at the nucleotide level between our sequences and the eliminated sequences of *C. kolensis*. In contrast, tblastx analysis (Camacho et al. 2009) revealed similarities found between the TE-related sequences from both species (Supplementary table S12). Therefore, the only similarities we could find between the eliminated DNA of *C. kolensis* and our *M. edax* genomic sequences are TE-related.

Finally, in order to complete our comparison with the *C. kolensis* eDNA sequences, motif analysis with MEME (Bailey 1994) was performed on our two libraries and subsequently repeated with the *C. kolensis* eDNA sequences. When

performed on the pre-diminution library without the TE sequences, motif analysis revealed a subset of motifs that present a similar, but somewhat different organization in 7 specific sequences (Figure 3). Some of these motifs can also be found more than once per sequence. These motifs appear to be dispersed repeats. In general, the motifs at the same organizational position between the sequences are very similar. However, motifs repeated inside the same sequence are seen to be more variable. A visual representation of this variable organization is illustrated in Figure 3. Otherwise, motif analysis generally revealed motifs arrangements that corresponded to similar sequence groups, as described earlier, and occasional indels were apparent through slight positional shifting of motifs. One post-diminution sequence (Abl-91) also contained some of these motifs, as it was part of a group of similar sequences (group 27, table S3). When repeated with our post-diminution sequences, the MEME analysis did not reveal such variable organization of motifs. Finally, we used FIMO to verify the presence of the pre-diminution motifs in the post-diminution sequences. Occurrences of all pre-diminution motifs were found in the post-diminution sequences.

The MEME analysis was repeated with the *C. kolensis* eDNA sequences in order to be able to make a comparison with our pre- and post-diminutive *M. edax* sequences. The results demonstrated a variety of motif arrangements, some which could be shared amongst several sequences. These motifs arrangements corresponded to groupings of similar sequences. Some motifs could be found more than once per sequence, thus resembling dispersed

repeats. The visual representation provided by MEME demonstrated that some of these arrangements appeared to have inconsistencies, such as the absence of a motif at a given position or as apparent spacers of variable lengths between motifs. However, alignments made with these sequences revealed that any inconsistencies in these arrangements were due to occasional indels. Two examples are provided in figure S1.

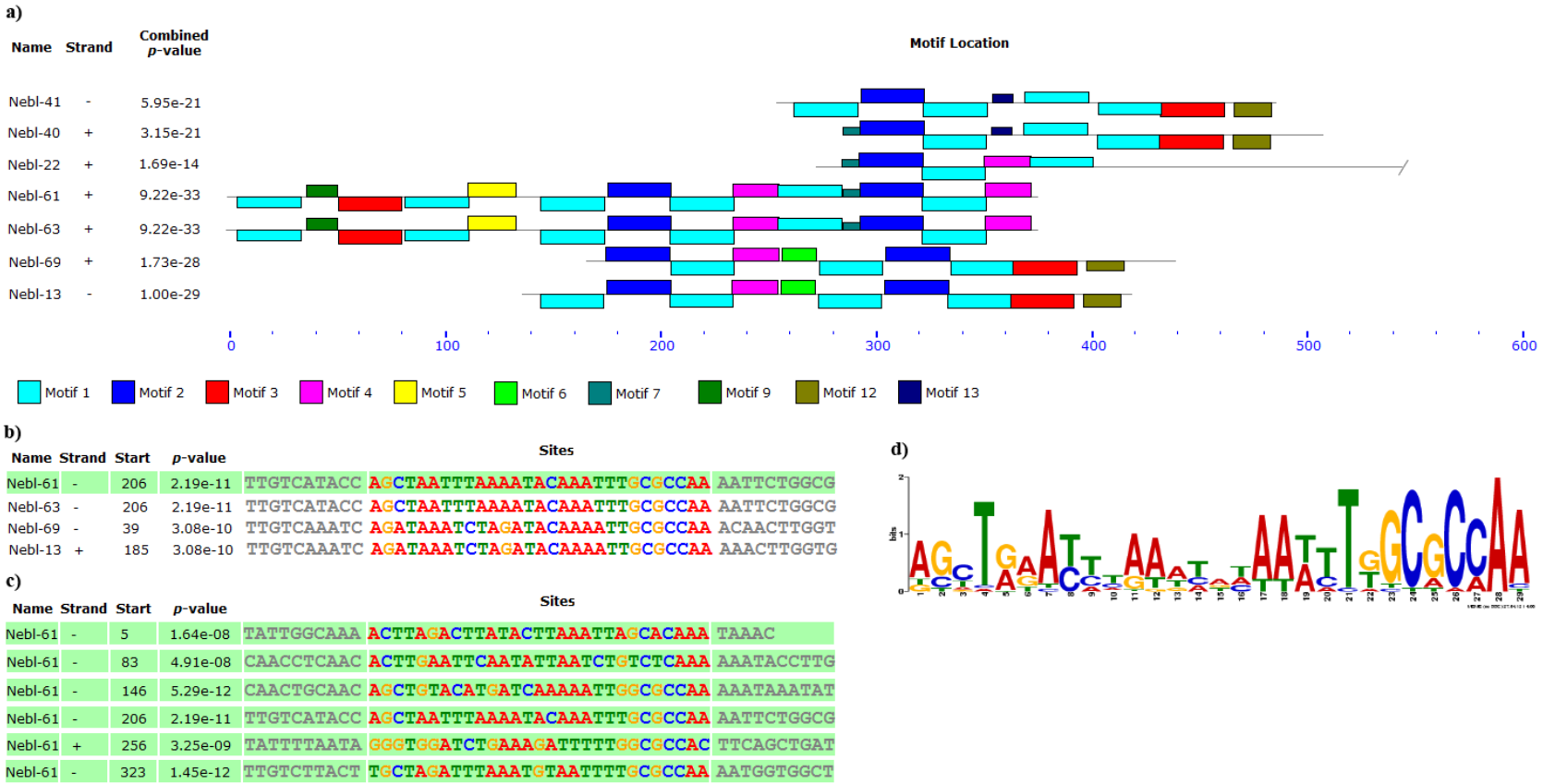

Figure 3. Motifs of variable organization found in the pre-diminution sequences. a) The 7 sequences along with the 10 motifs that are found to vary in terms of organization. The sequences were aligned visually to demonstrate the variation in motif organization. b) Alignment of motif 1 at the same organizational position between 4 different sequences. c) Alignment of motif 1 inside Nebl-61. d) Sequence Logo for motif 1.

2.4 Discussion:

Analyses of the TE related sequences suggest that they do not compose a considerable part of the eliminated DNA, nor do they present a germ-line specific variety. First and foremost, there is no decrease in the overall portion of TE related sequences from the pre- to post-diminution libraries. On the contrary, our sequences seem to demonstrate an increase in the proportion of TEs. Moreover, the sequences similar to Gypsy-17_DPu-I (which constitute the majority of the TE portions of both libraries) do not seem to demonstrate a variant unique to pre-diminution sequences, nor do they seem to be more divergent in the pre-diminution sequences. Finally, we observed only one possible occurrence of chromodomain containing element in our pre-diminution sequences, such as would specifically transpose into heterochromatic structure.

The major limitation of our study, in terms of the composition of TEs, is that our sequences represent only fragments of whole TEs. Thus, it is not possible to determine if our sequences are part of intact copies. Also, detection of TEs is restricted to similarity pairing with known elements from other organisms. However, the resulting observations propose that there is no major elimination of TEs during the process of chromatin diminution in *M. edax*, although we cannot refute that a small amount of TE copies might be eliminated. It is not surprising, since *A. suum* has shown that copies of Tas retrotransposons are found to be eliminated during chromatin diminution, but only in limited amounts (Aeby et al. 1986). Hence, the hypothetical premise that TEs undergo

elimination during chromatin diminution cannot be confirmed, at least not in an observable manner using our sequences.

Amongst the various tandem repeats that were found, we found none of the repeats described in Drouin (2006), as expected from our experimental design. Both libraries contain tandem repeat portions lower than 2%, and a slight increase is observed between the pre and post-diminution libraries. Considering that *M. edax* eliminates up to 90% of DNA, our observations suggest that there are no (or at least, not a considerable amount of) tandem repeats eliminated, other than those previously reported.

As for protein coding sequences, a comparison between our two libraries did not reveal any major difference. Of course, our experimental design does not permit us to observe the elimination of single copy genes, such as observed in *A. suum* (Etter et al. 1991; Spicher et al. 1994; Huang et al. 1996). However, the proportion of protein coding sequences is not different from one library to the other. This is also the case with sequences that were similar to hypothetical protein sequences found in the *D. pulex* genome. Therefore, we cannot infer any difference in the overall composition of protein coding sequences between our two libraries. As such, analyses of our sequences indicate that either no protein coding sequences are eliminated, or conversely, elimination of such sequences would be restricted to an unnoticeable quantity.

2.4.1 Comparable contents of eliminated DNA in *M. edax* and *C. kolensis:*

The comparisons of our sequences with the eDNA of *C. kolensis* (Degtyarev et al. 2004) were done with the aim of acquiring an interspecies view of the matter of CD in copepods. Most aspects studied were revealed to be similar not only between our two libraries, but also when comparing both of them with the *C. kolensis* eDNA. The comparison was first made by repeating the analysis of our sequences with the C. kolensis eDNA, in respect to TEs, protein coding sequences and chromodomains. We also described the presence of similar sequence groups in both our libraries, which were somewhat similar to what was reported in the eDNA of *C. kolensis* (Degtyarev et al. 2002; Degtyarev et al. 2004). Some of our results, however, differ with the original report on the eDNA of *C. kolensis* (Degtyarev et al. 2004). Specifically, we could find similarities to protein coding sequences in the *C. kolensis* eDNA sequences, in which protein coding sequences were reported to be inexistent, and as much can be said about TEs. We assume these differences are due to the dynamic nature and temporal growth of online databases.

Finally, we used the MEME software to discover motifs in our sequences, as was used for the analysis of *C. kolensis* eDNA sequences (Degtyarev et al. 2004). We observed a subset of motifs which presented a certain variable organization in 7 sequences of the pre-diminution library (Figure 3). However, the variable organisational scheme observed in our pre-diminution sequences

appears like products of homology-based crossovers (Smith 1976), such as could also be at the basis for the formation of tandem satellite sequence. Although the MEME analysis did not reveal such motif organization in the post-diminution sequences, the FIMO analysis demonstrated that all pre-diminution motifs also occurred in the post-diminution sequences. We then repeated the MEME analysis with the *C. kolensis* eDNA sequences, in order to have first hand results using the same MEME software version (4.8.1) that we used to analyse our own sequences. The given results demonstrated motifs with arrangements that were in accord with groupings of similar sequences, along with indels demonstrating spacings in motif positions (figure S1). Overall, MEME results appear somewhat similar when comparing our pre-diminution sequences to the *C. kolensis* eDNA sequences.

According to past studies and the present one, these two copepod species demonstrate some similarities but also key differences when it comes to CD at the molecular level (overview in Table 3). So far, evidence suggests that CD in *M. edax* results in a reduction in the total contents of tandem repeats (Drouin 2006), but no significant reduction in terms of TEs, protein coding sequences, or groups of similar sequences (this study). In *C. kolensis*, analysis of the eDNA sequences also demonstrates dispersed repeats (this study). Some occurrences of TE-related sequences and sequences similar to protein coding sequences can also be found in the eDNA of *C. kolensis* (this study). While various tandem repeats can also be found, they do not as a whole appear to occupy an important fraction of the eDNA from *C. kolensis* (this study).

Table 3: Summary of the types, and respective abundances, of sequences found in *M. edax* pre- and post-diminution sequences and in the eliminated sequences from *C. kolensis*.

Category	***M. edax***		***C. kolensis***
	pre-diminution	**post-diminution**	**eliminated DNA**
% Transposable element sequences (% nucleotides)	15.7% (9.22%)	25.8% (15.8%)	16.4% (9.6%)
% protein coding sequences[1] (% nucleotides)	14.3% (8.3%)	19.5% (8.1%)	11.5 % (6.4%)
% sequences with short tandem repeats (% nucleotides)	20.0% (1.44%)	37.5% (1.99%)	27.9% (1.6%)
Dispersed repeats	Dispersed repeats	Dispersed repeats	Dispersed repeats

[1] : other than proteins of transposable element origin

Thus, there appears to be incongruence, but also similarities between the contents of the DNA that is eliminated in *M. edax* and in *C. kolensis*. The incongruence comes from the reduction in satellite DNA, observed in *M. edax* but not *C. kolensis*. Such a disparity is not surprising: comparatively, *A. suum* and *P. univalens* are reported to eliminate repeats of different nature (Roth and Moritz 1981; Müller et al. 1982; Streeck et al. 1982; Teschke et al. 1991; Niedermaier and Moritz 2000). On the other hand, it is possible that the different experimental designs are responsible for the incongruence between these two copepod species. For example, although both species have been studied using whole genome amplification methods, one was studied using completely randomized primers while the other was amplified using partially randomized primers. We believe this matter will be clarified by a repetition of these experiments with the opposite species.

The similarities concern the fact that we can find similar contents in our pre- and post-diminution sequences and the *C. kolensis* eDNA sequences. These contents include TE-related sequences, groups of similar sequences, protein coding sequences, various tandem repeats and dispersed repeats. Considering that the eDNA of *C. kolensis* presents the aforementioned similarities with both our libraries, this may indicate that *M. edax* also eliminates portions of euchromatic DNA (in an unspecific fashion). This is not surprising, as another diminutive copepod, *Acanthocyclops vernalis*, was also proposed to eliminate DNA that is not genetically inert (Standiford 1989).

2.5 Summary:

In summary, our comparative analysis by enzymatic restriction of both pre and post-diminution sequences from the *M. edax* genome does not demonstrate an apparent reduction in the portion of TEs. These findings propose that CD in *M. edax* does not serve to specifically reduce the number of TEs in the somatic genome, as it has been proposed for DNA elimination in ciliates (Schoeberl and Mochizuki 2011). Moreover, the overall composition of protein coding sequences and similar sequence groups seem to be alike in both our pre- and post-diminution libraries. We also did not find any differences with our sequences when we analyzed the eDNA from *C. kolensis* in terms of TEs, protein coding sequences, tandem repeats and dispersed repeats. Thus we suggest that *M. edax* eliminates portions of euchromatic DNA and dispersed repeats, in addition to the previously characterized satellite sequences.

2.6 Acknowledgements:

We thank Dr. Grace Wyngaard (Dept. of Biology, James Madison University, Harrisonburg, Va.) for her help with specimen collection and identification, for providing access to her laboratory, and for discussions on numerous matters concerning the chromatin diminution of copepods. This work was funded by a Discovery grant of the Natural Sciences and Engineering Research Council of Canada.

2.7 Supplementary tables:

Table S1: List and contents of similar sequence groups containing sequences specific to the pre-diminution library.

Group	Pre-diminution clones	Contents (description)
1	Nebl-03 Nebl-05	Similar non-coding sequences
2	Nebl-11 Nebl-12	Similar non-coding sequences that contain a tri-nucleotide repeat and a poly-A tract
3	Nebl-13 Nebl-69	Similar non-coding sequences that contain a tri-nucleotide repeat
4	Nebl-22 Nebl-40 Nebl-41	Similar non-coding sequences; Nebl-22 is only similar over a 78bp long region and contains tri-nucleotide repeats and a 9bp long repeat
5	Nebl-25 Nebl-39 Nebl-64	Similar non-coding sequences
6	Nebl-44 Nebl-70 Nebl-77	Similar non-coding sequences
7	Nebl-57 Nebl-60	Similar non-coding sequences
8	Nebl-73 Nebl-75	Sequences coding for cytochrome c class I (ADR22685.1)
31	Nebl-33 Nebl-53	Sequences similar to hypothetical protein DAPPUDRAFT_277799 (EFX60384.1)

Table S2: List and contents of similar sequence groups containing sequences specific to the post-diminution library.

Group	Post-diminution clones	Contents (description)
9	Abl-5 Abl-10	Similar non-coding sequences
10	Abl-22 Abl-114 Abl-160	Similar non-coding sequences, Abl-114 and Abl-160 contain a tri-nucleotide repeat
11	Abl-23 Abl-87	Non-coding sequences similar over a 61bp long region; Abl-23 contains a 15bp long tandem repeat
12	Abl-29 Abl-161 Abl-177	Similar non-coding sequences that are similar to a TE
13	Abl-32 Abl-147	Similar non-coding sequences that are similar to a TE
14	Abl-62 Abl-81	Similar sequences that contain a poly-C tract and are similar to outer membrane autotransporter barrel domain-containing protein 3 (YP_003979045.1)
15	Abl-64 Abl-83	Similar non-coding sequences
16	Abl-70 Abl-78 Abl-82	Similar sequences that contain a tri-nucleotide repeat and are similar to hypothetical protein DAPPUDRAFT_274695 (EFX61070.1)
17	Abl-76 Abl-88	Similar sequences that contain a 19bp long tandem repeat
18	Abl-80 Abl-85	Similar non-coding sequences; Abl-80 contains a tri-nucleotide repeat
19	Abl-112 Abl-140 Abl-155 Abl-168	Similar non-coding sequences; Abl-112 is only similar over a 150bp long region
20	Abl-154 Abl-163	Similar non-coding sequences that contain a tri-nucleotide repeat
23	Abl-42 Abl-52 Abl-56 Abl-58 Abl-68 Abl-96	Similar non-coding sequences similar to a TE; Abl-56 contains a tri-nucleotide repeat

Table S3: List and contents of sequence groups containing sequences from both libraries.

Group	Pre-diminution clones	Post-diminution clones	Contents (description)
21	Nebl-23	Abl-31 Abl-93 Abl-110 Abl-113	Similar non-coding sequences that are similar to a TE; Abl-31 contains a tri-nucleotide repeat
22	Nebl-32 Nebl-51	Abl-37 Abl-100	Similar non-coding sequences that are similar to a TE
24	Nebl-72	Abl-50 Abl-54 Abl-65	Nebl-72, Abl-54 and Abl-65 contain a tri-nucleotide repeat; Abl-54 is similar to a TE and to peptidase inhibitor 16 GAA55503.1); Abl-65 is similar to hypothetical protein DAPPUDRAFT_120482 (EFX62140.1)
25	Nebl-24	Abl-57 Abl-136 Abl-189	Contain similar non-coding sequences that are similar to a TE
26	Nebl-15 Nebl-89	Abl-75 Abl-84 Abl-97 Abl-117 Abl-135	Similar non-coding sequences similar to a TE; Abl-97 contains a tri-nucleotide repeat
27	Nebl-61 Nebl-63	Abl-91	Similar sequences; Abl-91 contains a 10bp long tandem repeat and is similar to LWamide neuropeptides (EFN85835.1)
28	Nebl-17	Abl-94	Non-coding sequences similar over a 43bp long region; Nebl-17 contains a tri-nucleotide tandem repeat
29	Nebl-27	Abl-121	Similar non-coding sequences that contain tri-nucleotide and a 11bp long tandem repeats
30	Nebl-16	Abl-89	Similar non-coding sequences similar to a TE
32	Nebl-20	Abl-185	Similar non-coding sequences similar to a TE

Table S4: Blastx results[a] (unrelated to TEs) from pre-diminution sequences.

Clone[b]	Protein	Species	Accession	E-value[c]
Nebl-02	glutathione reductase	*Bradyrhizobium japonicum USDA 6*	BAL11379.1	1e-26
Nebl-28	Epa5p	*Candida glabrata*	AAQ82687.1	2e-06
Nebl-29	CBR-ABT-1 protein	*Caenorhabditis briggsae*	XP_002633711.1	1e-05
Nebl-33	hypothetical protein DAPPUDRAFT_277799	*Daphnia pulex*	EFX60384.1	8e-05
Nebl-53	hypothetical protein DAPPUDRAFT_277799	*Daphnia pulex*	EFX60384.1	8e-05
Nebl-73	cytochrome c class I	*Flavobacterium columnare ATCC 49512*	ADR22685.1	4e-35
Nebl-75	cytochrome c class I	*Flavobacterium columnare ATCC 49512*	ADR22685.1	3e-36
Nebl-83	conserved hypothetical protein	*Ixodes scapularis*	EEC15057.1	1e-10
Nebl-86	terminase large subunit	*Mycobacterium tuberculosis H37Ra*	ZP_02548803.1	2e-55
Nebl-90	cyclopropane-fatty-acyl-phospholipid synthase	*Phenylobacterium zucineum HLK1*	YP_002128708.1	3e-14

[a]: only the first significant result from non hypothetical or non putative protein is shown.
[b]: clones determined to be TE-derived (according to Tables S4 and S7) are excluded.
[c]: cut-off of <1[e]-004.

Table S5: Blastx results[a] (related to TEs) from pre-diminution sequences.

Clone	Protein	Species	Accession	E-value[b]
Nebl-09	Polyprotein	*Yarrowia lipolytica*	CAG34127.1	1e-07
Nebl-15	Polyprotein	*Lubomirskia baicalensis*	CAP69654.1	7e-42
Nebl-20	pol protein	*Drosophila melanogaster*	AAM11674.1|AF492764_2	3e-15
Nebl-23	pol protein	*Drosophila melanogaster*	AAM11674.1|AF492764_2	2e-35
Nebl-24	Retrovirus polyprotein	*Dicentrarchus labrax*	CBN80961.1	2e-27
Nebl-32	PREDICTED: similar to polyprotein	*Hydra magnipapillata*	XP_002157550.1	3e-09
Nebl-51	PREDICTED: similar to polyprotein	*Hydra magnipapillata*	XP_002157550.1	5e-11
Nebl-68	zinc knuckle protein	Trichinella spiralis	XP_003370317.1	1e-09
Nebl-88	Pao retrotransposon peptidase superfamily	*Trichinella spiralis*	EFV50150.1	2e-25
Nebl-89	gag-pol polyprotein	*Panicum virgatum*	ABO21746.1	3e-33

[a]: only the first significant result from non hypothetical or non putative protein is shown.
[b]: cut-off of <1^{e}-004.

Table S6: Blastx results[a] (unrelated to TEs) from post-diminution sequences.

Clone[b]	Protein	Species	Accession	E-value[c]
Abl-12	mCG145879	*Mus musculus*	EDL01652.1	6e-07
Abl-15	GA23099	*Drosophila pseudoobscura pseudoobscura*	XP_002133829.1	5e-12
Abl-18	salivary gland secretion 1	*Drosophila melanogaster*	NP_523475.3	7e-07
Abl-35	ATP dependent DEAD-box helicase, putative	*Plasmodium falciparum 3D7*	XP_966219.1	9e-09
Abl-53	hypothetical protein DAPPUDRAFT_274695	*Daphnia pulex*	EFX61070.1	7e-13
Abl-54	peptidase inhibitor 16	*Clonorchis sinensis*	GAA55503.1	3e-06
Abl-59	hypothetical protein DAPPUDRAFT_336040	*Daphnia pulex*	EFX63029.1	6e-08
Abl-62	outer membrane autotransporter barrel domain-containing protein 3	*Achromobacter xylosoxidans A8*	YP_003979045.1	2e-08
Abl-65	hypothetical protein DAPPUDRAFT_120482	*Daphnia pulex*	EFX62140.1	3e-09
Abl-66	AGAP000003-PA	*Anopheles gambiae str. PEST*	XP_001687752.1	1e-36
Abl-70	hypothetical protein DAPPUDRAFT_274695	*Daphnia pulex*	EFX61070.1	3e-16
Abl-72	beta-TRCP/Slimb protein	*Glossina morsitans morsitans*	ADD20087.1	1e-10
Abl-78	hypothetical protein DAPPUDRAFT_274695	*Daphnia pulex*	EFX61070.1	4e-13
Abl-81	outer membrane autotransporter barrel domain-containing protein 3	*Achromobacter xylosoxidans A8*	ADP16330.1	2e-08
Abl-82	hypothetical protein DAPPUDRAFT_274695	*Daphnia pulex*	EFX61070.1	2e-12
Abl-91	LWamide neuropeptides	*Harpegnathos saltator*	EFN85835.1	1e-05
Abl-95	hypothetical protein DAPPUDRAFT_277799	*Daphnia pulex*	EFX60384.1	5e-08
Abl-103	hydroxyproline-rich glycoprotein	*Cryptosporidium hominis*	ACC62005.1	9e-31
Abl-128	hypothetical protein	*Daphnia pulex*	EFX60384.1	7e-23

Abl-148	DAPPUDRAFT_277799 hypothetical protein	*Daphnia pulex*	EFX61855.1	2e-13
Abl-162	DAPPUDRAFT_120854 hypothetical protein	*Daphnia pulex*	EFX60371.1	9e-33
Abl-170	DAPPUDRAFT_344141 hypothetical protein	*Daphnia pulex*	EFX61185.1	7e-11
Abl-172	DAPPUDRAFT_340482 catalase	*Campylobacter jejuni*	Q59296.1	2e-05
Abl-188	hypothetical protein	*Daphnia pulex*	EFX60384.1	2e-08
Abl-190	DAPPUDRAFT_277799 PBR	*Burkholderia cenocepacia*	ACJ54935.1	1e-10

[a]: only the first significant result from non hypothetical or non putative protein is shown.
[b]: clones determined to be TE-derived (according to Tables S6 and S8) are excluded.
[c]: cut-off of $<1^{e}$-004.

Table S7: Blastx results[a] (related to TEs) from pre-diminution sequences.

Clone	Protein	Species	Accession	E-value[b]
Abl-29	ORF3	*Drosophila melanogaster*	AAN87268.1	2e-21
Abl-31	Retrovirus-related Pol polyprotein from transposon 17.6	*Drosophila melanogaster*	P04323.1	2e-45
Abl-37	PREDICTED: similar to polyprotein	*Hydra magnipapillata*	XP_002157550.1	1e-09
Abl-42	pro-Pol polyprotein, partial	*Clonorchis sinensis*	GAA48974.1	4e-30
Abl-52	protease, reverse transcriptase, ribonuclease H, integrase	*Drosophila buzzatii*	CAB39733.1	7e-21
Abl-56	LReO_3	*Oryzias latipes*	BAB83836.1	3e-39
Abl-57	LReO_3	*Oryzias latipes*	BAB83836.1	9e-31
Abl-58	LReO_3	*Oryzias latipes*	BAB83836.1	6e-37
Abl-68	LReO_3	*Oryzias latipes*	BAB83836.1	4e-27
Abl-75	polyprotein	*Lubomirskia baicalensis*	CAP69654.1	3e-41
Abl-84	protease, reverse transcriptase, ribonuclease H, integrase	*Drosophila buzzatii*	CAB39733.1	5e-23
Abl-90	pol polyprotein	*Nosema bombycis*	ABE26654.1	8e-15
Abl-93	pol protein	*Drosophila melanogaster*	AAM11674.1	1e-30
Abl-96	Pol polyprotein TcasGA2_TC002008	*Dicentrarchus labrax*	CBN81191.1	1e-30
Abl-97	polyprotein	*Lubomirskia baicalensis*	CAP69654.1	3e-35
Abl-100	PREDICTED: similar to polyprotein	*Hydra magnipapillata*	XP_002157550.1	1e-11
Abl-110	Retrovirus-related Pol polyprotein from transposon 17.6	*Drosophila melanogaster*	P04323.1	6e-54
Abl-113	ORF B	*Autographa californica nucleopolyhedrovirus*	AAA92249.1	3e-44
Abl-117	gag-pol polyprotein	*Panicum virgatum*	ABO21746.1	4e-39
Abl-120	Retrovirus polyprotein	*Dicentrarchus labrax*	CBN80961.1	4e-23

Abl-135	gag-pol polyprotein	*Panicum virgatum*	EFA00130.1	5e-22
Abl-136	pro-Pol polyprotein, partial	*Clonorchis sinensis*	GAA48974.1	1e-30
Abl-161	TPA: zinc finger protein	*Ciona intestinalis*	FAA00226.1	6e-28
Abl-173	TPA: gag-pol polyprotein	*Schistosoma mansoni*	CAJ00230.1	5e-18
Abl-174	putative reverse transcriptase	*Bacillus cereus W*	ZP_03104392.1	5e-11
Abl-177	TPA: zinc finger protein	*Ciona intestinalis*	FAA00226.1	3e-28
Abl-181	terminase large subunit	*Mycobacterium tuberculosis H37Ra*	ZP_02548803.1	3e-47
Abl-185	pol protein	*Drosophila melanogaster*	AAM11674.1	5e-17
Abl-186	putative reverse transcriptase	*Cicer arietinum*	CAD59768.1	8e-15
Abl-189	TPA: pol polyprotein	*Schistosoma mansoni*	CAJ00226.1	9e-26

[a]: only the first significant result from non hypothetical or non putative protein is shown.
[b]: cut-off of $<1^{e}$-004.

Table S8: tBlastx[a] results from pre-diminution *M. edax* sequences on TEs from Repbase Update (release 17.01).

Clone	Transposable Element	E-Value[b]	Frame[c]	Range[d]
Nebl-09	Gypsy-7_PSo-I_Gypsy_Phytophthora sojae	7e-012	+3/+3	4251..4442
Nebl-15	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	3e-077	-1/+1	2698..3267
Nebl-20	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	4e-027	-3/+2	2333..2497
Nebl-23	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-060	+3/+1	2500..2925
Nebl-24	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	2e-068	+3/+1	4162..4611
Nebl-29	Ginger2-1_LS_Ginger2/TDD_Littorina saxatilis	1e-005	+3/-2	3155..2943
Nebl-32	BEL-4-I_NV_BEL_Nematostella vectensis	1e-009	+3/+2	1601..1696
Nebl-51	BEL-4-I_NV_BEL_Nematostella vectensis	1e-009	+1/+2	1601..1696
Nebl-68	BEL-26_AA-I_BEL_Aedes aegypti	3e-014	-3/+3	1635..1805
Nebl-88	BEL1-I_SM_BEL_Schmidtea mediterranea	2e-033	+2/+1	3091..3402
Nebl-89	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-065	-3/+1	2752..3267

[a]: only the TE with the lowest E-value is shown
[b]: cut-off of <1^{e}-004
[c]: reading frame of query/database sequences
[d]: range positions of the alignment over TE

Table S9: tBlastx[a] results from post-diminution *M. edax* sequences on TEs from Repbase Update (release 17.01).

Clone	Transposable Element	E-Value[b]	Frame[c]	Range[d]
Abl-29	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	5e-046	+2/+1	4072..4368
Abl-31	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	4e-075	+2/+1	2500..3075
Abl-32	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-013	+3/+2	1412..1669
Abl-37	BEL-4-I_NV_BEL_Nematostella vectensis	1e-009	+2/+2	1601..1696
Abl-42	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	9e-060	+2/+1	4063..4461
Abl-50	Copia-130_AA-I_Copia_Aedes aegypti	3e-005	+3/+1	1708..1887
Abl-52	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-047	+1/+1	4063..4368
Abl-56	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	7e-077	+2/+1	4102..4593
Abl-57	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-051	+1/+1	4039..4473
Abl-58	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	3e-077	+1/+1	4063..4602
Abl-61	Gypsy-7_PGr-I_Gypsy_Puccinia graminis f. sp. tritici CRL 75-36-700-3	6e-005	+3/+2	3965..4087
Abl-68	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	3e-054	+3/+1	4102..4431
Abl-75	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-072	+3/+1	2716..3256
Abl-84	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	8e-049	-1/+1	2821..3423
Abl-90	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	3e-037	-2/+1	3337..3645
Abl-93	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	2e-055	+1/+1	2500..2883
Abl-96	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-069	+2/+1	4162..4605
Abl-97	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	2e-060	-1/+1	2899..3267
Abl-100	BEL-4-I_NV_BEL_Nematostella vectensis	9e-010	+3/+2	1601..1696
Abl-110	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	3e-093	+2/+1	2500..3129
Abl-113	Gypsy-17_DPu-I_Gypsy_Daphnia	3e-072	+1/+1	2500..3060

	pulex			
Abl-117	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-074	+2/+1	2677..3267
Abl-120	Gypsy-28_BRa-I_Gypsy_Brassica rapa	4e-026	+2/+1	1966..2253
Abl-135	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	3e-045	-3/+1	2878..3267
Abl-136	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	8e-059	+1/+1	4102..4467
Abl-147	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	9e-014	+1/+2	1412..1669
Abl-161	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	2e-053	+2/+1	4072..4419
Abl-171	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	4e-013	-1/+3	7200..7643
Abl-173	Gypsy-43-I_NV_Gypsy_Nematostella vectensis	9e-034	+3/+3	2952..3374
Abl-177	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	4e-054	+2/+1	4072..4422
Abl-183	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	2e-022	-3/+3	7677..8147
Abl-185	Gypsy-17_DPu-I_Gypsy_Daphnia pulex Gypsy-17_DPu-I_Gypsy_Daphnia pulex	3e-029 3e-029	+1/+2 +1/+1	2327..2497
Abl-189	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	7e-049	+2/+1	4102..4413

[a]: only the TE with the lowest E-value is shown
[b]: cut-off of <1[e]-004
[c]: reading frame of query/database sequences
[d]: range positions of the alignment over TE

Table S10: Blastx results[a] from *C. kolensis* diminution sequences (Degtyarev et al. 2004).

Clone[b]	Protein	Species	Accession	E-value[c]
CkD126	hypothetical protein DAPPUDRAFT_119545	Daphnia pulex	EFX63089.1	6e-08
CkD133	hypothetical protein	Paramecium tetraurelia strain d4-2	XP_001428852.1	3e-07
CkD42	hypothetical protein DAPPUDRAFT_345956	Daphnia pulex	EFX60000.1	4e-10
CkD50	hypothetical protein DAPPUDRAFT_336780	Daphnia pulex	EFX62558.1	5e-11
CkD53	hypothetical protein DAPPUDRAFT_336780	Daphnia pulex	EFX62558.1	4e-06
CkD71	caldesmon	Entamoeba histolytica HM-1:IMSS	XP_001914150.1	9e-06
CkD99	hypothetical protein DAPPUDRAFT_336709	Daphnia pulex	EFX62609.1	2e-05

[a]: only the TE with the lowest E-value is shown.
[b]: clones determined to be TE-derived (according to Table S10) are excluded.
[c]: cut-off of $<1^{e}$-004.

Table S11: tBlastx[a] results from eliminated sequences of *C. kolensis* (Degtyarev et al. 2004) on TEs from Repbase Update (release 17.01).

Clone	Transposable Element	E -Value[b]	Frame[c]	Range[d]
CkD103	Gypsy-22-I_SP_Gypsy_Strongylocentrotus purpuratus	9e-007	-3/+3	4725..4856
CkD120	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	3e-007	+3/+3	120..341
CkD18	Polinton-3_SP_Polinton_Strongylocentrotus purpuratus	5e-007	+2/+1	4516..4758
CkD24	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-006	+3/+1	4246..4329
CkD37	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	6e-014	-3/+1	5419..5604
CkD44	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-007	-1/+3	120..341
CkD5	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	2e-013	-2/+1	5419..5604
CkD51	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	2e-009	+1/+1	5419..5577
CkD55	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	1e-012	-3/+1	5419..5604
CkD92	Gypsy-17_DPu-I_Gypsy_Daphnia pulex	9e-013	-3/+3	7428..7679

[a]: only the TE with the lowest E-value is shown
[b]: cut-off of <1[e]-004
[c]: reading frame of query/database sequences
[d]: range positions of the alignment over the TE

Table S12: tBlastx results from pre- and post-diminution *M. edax* sequences on eliminated sequences of *C. kolensis* (Degtyarev et al. 2004).

Clone	Eliminated sequence from *C. kolensis*	E -Value[a]	Frame[b]	Range[c]
Nebl-24	CkD24	2e-007	+3/+3	3..122
Nebl-32	CkD3	4e-006	+3/+2	56..280
Nebl-51	CkD3	7e-005	+1/+2	68..280
Abl-42	CkD24	2e-008	+2/+3	3..119
Abl-52	CkD24	3e-008	+1/+3	3..119
Abl-56	CkD24	1e-008	+2/+3	3..122
Abl-57	CkD24	5e-008	+1/+3	3..122
Abl-58	CkD24	3e-008	+1/+3	3..119
Abl-68	CkD24	4e-008	+3/+3	6..119
Abl-96	CkD24	2e-008	+2/+3	3..122
Abl-136	CkD24	5e-008	+1/+3	3..122
Abl-171	CkD92	4e-022	-1/-3	261..37
Abl-183	CkD92	5e-014	-3/-3	114..1
Abl-189	CkD24	4e-008	+2/+3	3..122

[a]: cut-off of $<1^{e}$-004

[b]: reading frame of query/database sequences

[c]: range positions of the alignment over the CkD sequence

2.8 Supplementary Figures:

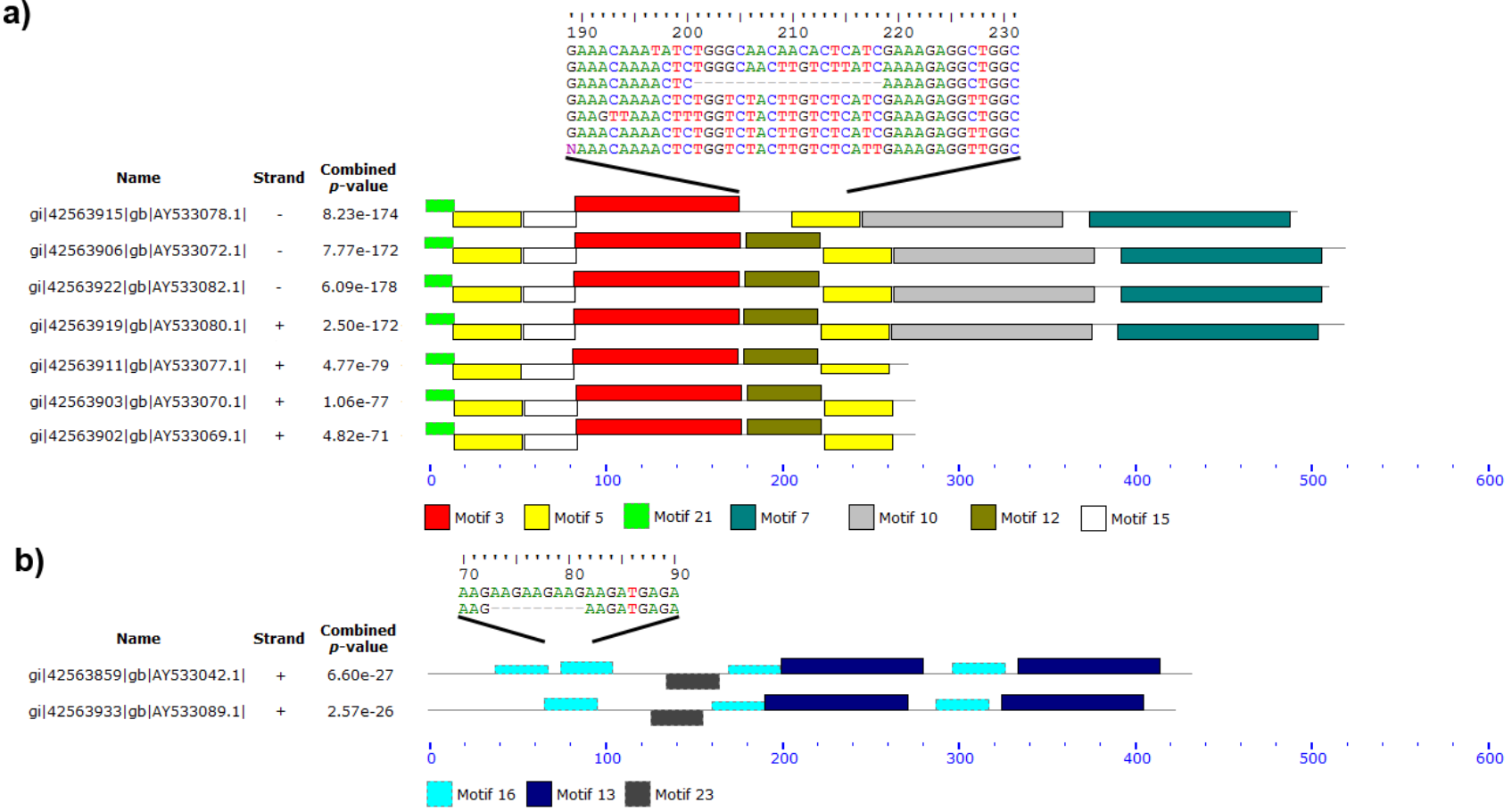

Figure S1: Examples of indels in MEME motif arrangements from *C. kolensis* eDNA sequences. a) Motif arrangement and alignment of 7 sequences, one of which has suffered a 18 bp long deletion causing an apparent absence of motif 12. b) Motif arrangement and alignment of 2 sequences, demonstrating an indel of 9 bp.

3. Conclusion:

CD is an intricate process of genomic rearrangement which remains to be entirely clarified, even more than a century after its discovery. Its purpose has also remained obscure, although several hypotheses have been proposed. In ciliates, it has been proposed that DNA elimination serves to eliminate transposons from the macronucleus (Schoeberl and Mochizuki 2011). Since CD in copepods eliminates large amounts of heterochromatic DNA, and that TEs can usually be found in heterochromatin, our hypothesis was that CD in *M. edax* would result in the reduction of total TEs (Beermann 1977; Chinnappa 1980; Charlesworth et al. 1994; Kloc and Zagrodzinska 2001; Redi et al. 2001; Rasch and Wyngaard 2008).

To test this hypothesis, we produced pre- and post-diminution DNA libraries using BamHI to digest DNA extracted from *M. edax* embryos (pre-diminution) and antennas (post-diminution). With these sequences, we performed several analyses, including a comparison of TE component between the two libraries. Our results were compared with the eDNA of *C. kolensis* by repeating our analyses on those sequences.

Comparison of our pre- and post-diminution libraries did not demonstrate a reduction in total TEs. This suggests that CD in *M. edax* does not lead to the specific reduction of TEs in the soma. Protein coding sequences, tandem repeats groups of similar sequences and dispersed repeats are also similar between the two libraries. Comparison of our libraries with the eDNA sequences of *C. kolensis* also did not reveal any difference in most of the analyses that were undertaken in this study (table 3). This suggests that, in addition to satellite sequences (Drouin, 2006), *M. edax* also eliminates euchromatic DNA. Given that Rasch and Wyngaard (2001)

have shown that endoreduplication occurs during the development of *M. edax*, our results therefore support that some of the endoreduplicated euchromatic DNA is also eliminated during CD.

3.1 Future work:

A next step would be to further characterize the portions of eliminated sequences in *M. edax* using fluorescent in situ hybridizations (FISH). For one, the reduction of previously described satellite repeats could be further described in a quantitative manner. Moreover, our proposition, that euchromatic sequences are eliminated, could be confirmed and also quantitatively described with FISH experiments using a variety of probes selected from the sequences obtained in this study. Finally, taking into consideration that approximately 90% DNA can be eliminated in *M. edax*, FISH studies might determine if other unaccounted sequences are eliminated during CD in *M. edax*. By accounting for the entire DNA that is eliminated during CD in *M. edax*, the purpose of CD in copepods will be one step closer to elucidation. Another possible venue would be to dissect granules containing the eliminated DNA from *M. edax* using a micromanipulator and subsequently amplify those sequences. This would insure the identification of eliminated sequences, and would procure data that can be more comparable with studies done in *C. kolensis*.

4. References:

Aeby P, Spicher A, de Chastonay Y, Müller F, Tobler H. 1986. Structure and genomic organization of proretrovirus-like elements partially eliminated from the somatic genome of Ascaris lumbricoides. *EMBO Journal* **5**(12): 3353-3360.

Amabis JM, Reinach FC, Andrews N. 1979. Spermatogenesis in Trichosia pubescens (Diptera: Sciaridae). *Journal of Cell Science* **VOL 36**: 199-213.

Bachmann-Waldmann C, Jentsch S, Tobler H, Müller F. 2004. Chromatin diminution leads to rapid evolutionary changes in the organization of the germ line genomes of the parasitic nematodes A. suum and P. univalens. *Molecular and Biochemical Parasitology* **134**(1): 53-64.

Back E, Fritz M, Tobler H. 1984. Chromosomal arrangement of the two main rDNA size classes of Ascaris lumbricoides. *Nucleic Acids Research* **12**(3): 1333-1348.

Bailey TE, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent systems for Molecular Biology*: 28-36.

Bantock C. 1961. Chromosome elimination in *Cecidomyiidae*. *Nature* **190**(477): 466-&.

Bantock CR. 1970. Experiments on chromosome elimination in gall midge, *Mayetiola destructor*. *Journal of Embryology and Experimental Morphology* **24**: 257-&.

Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Betermier M. 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. *Genes Dev* **23**(21): 2478-2483.

Bauer H. 1970. Rearrangements between germ-line limited and somatic chromosomes in *Smittia parthenogenetica* (*Chironomidae, Diptera*). *Chromosoma* **32**(1): 1-&.

Beermann S, Meyer GF. 1980. Chromatin rings as products of chromatin diminution in Cyclops. *Chromosoma* **77**(3): 277-283.

Beermann S. 1977. The diminution of heterochromatic chromosomal segments in Cyclops (Crustacea, Copepoda). *Chromosoma* **60**(4): 297-344.

-. 1984. Circular and linear structures in chromatin diminution of cyclops. *Chromosoma* **89**(5): 321-328.

Benson G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research* **27**(2): 573-580.

Berry RO. 1939. Observations on chromosome elimination in the germ cells of Sciara ocellaris. *Proceedings of the National Academy of Sciences of the United States of America* **25**: 125-127.

-. 1941. Chromosome behavior in the germ cells and development of the gonads in Sciara ocellaris. *Journal of Morphology* **68**(3): 547-583.

Bird AP, Taggart MH. 1980. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acids Research* **8** (7): 1485-1497.

Boveri T. 1887. Über Differenzierung der Zellkerne während der Furchung des Eis von Ascaris megalocephala. *Anat Anz, 2, pp 688-693*.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST plus : architecture and applications. *BMC Bioinformatics* **10**.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**(15): 1972-1973.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**(6494): 215-220.

Chinnappa CC. 1980. Bivalent forming race of Mesocyclops edax (Copepoda, Crustacea). *Canadian Journal of Genetics and Cytology* **22**(3): 427-432.

Close R. 1979. Sex Chromosome Mosaicism in Liver, Thymus, Spleen and Regenerating Liver of Perameles Nasuta and Isoodon Macrourus. *Australian Journal of Biological Sciences* **32**(6): 615-624.

Close RL. 1984. Studies on metatherian sex-chromosmes. 11. Rates of sex-chromosome loss during development in different tissues of bandicoots *Perameles nasuta* and *Isoodon macrourus* (*Marsupialia, Peramelidae*). *Australian Journal of Biological Sciences* **37**(1-2): 53-61.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK et al. 2011. The Ecoresponsive Genome of Daphnia pulex. *Science* **331**(6017): 555-561.

Crouse HV, Brown A, Mumford BC. 1971. L-chromosome inheritance and problem of chromosome imprinting in *Sciara* (*Sciaridae, Diptera*). *Chromosoma* **34**(3): 324-&.

Crouse HV. 1960. The controlling element in sex chromosome behavior in *Sciara*. *Genetics* **45**(10): 1429-1443.

-. 1966. An inducible change in state on chromosomes of *Sciara* - Its effects on genetic components of X. *Chromosoma* **18**(2): 230-&.

De Saint Phalle B, Sullivan W. 1996. Incomplete sister chromatid separation is the mechanism of programmed chromosome elimination during early Sciara coprophila embryogenesis. *Development* **122**(12): 3775-3784.

Degtyarev S, Boykova T, Grishanin A, Belyakin S, Rubtsov N, Karamysheva T, Makarevich G, Akifyev A, Zhimulev I. 2004. The molecular structure of the DNA fragments eliminated during chromatin diminution in Cyclops kolensis. *Genome Research* **14**(11): 2287-2294.

Degtyarev SV, Grishanin AK, Belyakin SN, Rubtsov NB, Zhimulev IF, Akifyev AP. 2002. DNA sequences eliminated during chromatin diminution from somatic cell chromosomes of Cyclops kolensis. *Doklady Biochemistry and biophysics* **384**: 148-151.

Dorward HM, Wyngaard, G.A. 1997. Variability and pattern of chromatin diminution in the freshwater Cyclopidae (Crustacea: Copepoda). *Archiv für Hydrobiologie* **Supplement**(107): 447-465.

Drouin G. 2006. Chromatin diminution in the copepod Mesocyclops edax: Diminution of tandemly repeated DNA families from somatic cells. *Genome* **49**(6): 657-665.

Du Bois AM. 1933. Chromosome behavior during cleavage in the eggs of Sciara coprophila (Diptera) in the relation to the problem of sex determination. *Zeitschrift für Zellforschung und Mikroskopische Anatomie* **19**(3): 595-614.

Dubois A. 1932. Elimination of Chromosomes during Cleavage in the Eggs of Sciara (Diptera). *Proc Natl Acad Sci USA* **18**(5): 352-356.

Dufresne F, Jeffery N. 2011. A guided tour of large genome size in animals: what we know and where we are heading. *Chromosome Research* **19**(7): 925-938.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**(5): 1792-1797.

Esteban MR, Campos MCC, Perondini ALP, Goday C. 1997. Role of microtubules and microtubule organizing centers on meiotic chromosome elimination in Sciara ocellaris. *Journal of Cell Science* **110**(6): 721-730.

Etter A, Aboutanos M, Tobler H, Muller F. 1991. Eliminated chromatin of Ascaris contains a gene that encodes a putative ribosomal protein. *Proceedings of the National Academy of Sciences of the United States of America* **88**(5): 1593-1596.

Etter A, Bernard V, Kenzelmann M, Tobler H, Müller F. 1994. Ribosomal heterogeneity from chromatin diminution in Ascaris lumbricoides. *Science* **265**(5174): 954-956.

Felder H, Herzceg A, De Chastonay Y, Aeby P, Tobler H, Müller F. 1994. Tas, a retrotransposon from the parasitic nematode Ascaris lumbricoides. *Gene* **149**(2): 219-225.

Fuge H. 1994. Unorthodox male meiosis in Trichosia pubescens (Sciaridae): Chromosome elimination involves polar organelle degeneration and monocentric spindles in first and second division. *Journal of Cell Science* **107**(1): 299-312.

Gao X, Hou Y, Ebina H, Levin HL, Voytas DF. 2008. Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Research* **18**(3): 359-369.

Geyerduszynska I. 1959. Experimental research on chromosome elimination in *Cecidomyidae* (*Diptera*). *Journal of Experimental Zoology* **141**(3): 391-447.

-. 1961. Spindle disappearance and chromosome behavior after partial-embryo irradiation in *Cecidomyiidae* (*Diptera*). *Chromosoma* **12**(2): 233-&.

Goday C, Esteban MR. 2001. Chromosome elimination in sciarid flies. *BioEssays* **23**(3): 242-250.

Goday C, Pigozzi MI. 2010. Heterochromatin and histone modifications in the germline-restricted chromosome of the zebra finch undergoing elimination during spermatogenesis. *Chromosoma* **119**(3): 325-336.

Goday C, Pimpinelli S. 1984. Chromosome organization and heterochromatin elimination in Parascaris. *Science* **224**(4647): 411-413.

-. 1986. Cytological analysis of chromosomes in the two species Parascaris univalens and P. equorum. *Chromosoma* **94**(1): 1-10.

Goday C, Rosario Esteban M. 2001. Chromosome elimination in sciarid flies. *BioEssays* **23**(3): 242-250.

Goldstein P, Straus NA. 1978. Molecular characterization of *Ascaris suum* and of chromatin diminution. *Experimental Cell Research* **116**(2): 462-466.

Goto Y, Kubota S, Kohno S-I. 1998. Highly repetitive DNA sequences that are restricted to the germ line in the hagfish Eptatretus cirrhatus: A mosaic of eliminated elements. *Chromosoma (Berlin)* **107**(1): 17-32.

Grishanin AK, Akif'ev AP. 2000. Interpopulation differentiation within C-kolensis and C-strenuus strenuus (Crustacea : Copepoda): evidence from cytogenetic methods. *Hydrobiologia* **417**: 37-42.

Guindon S, Gascuel O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* **52**(5): 696-704.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis frogram for Windows 95/98/NT. *Nucl Acids Symp* **Ser.**(41): 95-98.

Huang YJ, Stoffel R, Tobler H, Mueller F. 1996. A newly formed telomere in Ascaris suum does not exert a telomere position effect on a nearby gene. *Molecular and Cellular Biology* **16**(1): 130-134.

Itoh Y, Arnold AP. 2005. Chromosomal polymorphism and comparative painting analysis in the zebra finch. *Chromosome Research* **13**(1): 47-56.

Itoh Y, Kampf K, Pigozzi MI, Arnold AP. 2009. Molecular cloning and characterization of the germline-restricted chromosome sequence in the zebra finch. *Chromosoma* **118**(4): 527-536.

Jansen P, Moritz KB. 1986. Ascaris DNA topoisomerase I binds preferentially to the germ-line-limited DNA. *Naturwissenschaften* **73**(12): 739-741.

Jazdowskazagrodzinska B, Dallai R, Redi CA. 1992. Changes in DNA content and chromosome number during spermatogenesis in the gall midge *Monarthropalpus buxi* (*Cecidomyiidae, Diptera*). *Genome* **35**(2): 244-250.

Jentsch S, Tobler H, Müller F. 2002. New telomere formation during the process of chromatin diminution in Ascaris suum. *International Journal of Developmental Biology* **46**(1): 143-148.

Kloc M, Zagrodzinska B. 2001. Chromatin elimination - an oddity or a common mechanism in differentiation and development? *Differentiation* **68**(2-3): 84-91.

Kojima NF, Kojima KK, Kobayakawa S, Higashide N, Hamanaka C, Nitta A, Koeda I, Yamaguchi T, Shichiri M, Kohno SI et al. 2010. Whole chromosome elimination and chromosome terminus elimination both contribute to somatic differentiation in Taiwanese hagfish Paramyxine sheni. *Chromosome Research* **18**(3): 383-400.

Kolpakov R, Bana G, Kucherov G. 2003. mreps: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Research* **31**(13): 3672-3678.

Kubota S, Ishibashi T, Kohno S. 1997. A germline restricted, highly repetitive DNA sequence in Paramyxine atami: An interspecifically conserved, but somatically eliminated, element. *Molecular and General Genetics* **256**(3): 252-256.

Kubota S, Kuro-o M, Mizuno S, Kohno SI. 1993. Germ line-restricted, highly repeated DNA sequences and their chromosomal localization in a Japanese hagfish (Eptatretus okinoseanus). *Chromosoma* **102**(3): 163-173.

Kubota S, Nakai Y, Kuro-o M, Kohno S. 1992. Germ line-restricted supernumerary (B) chromosomes in Eptatretus okinoseanus. *Cytogenetics and Cell Genetics* **60**(3-4): 224-228.

Kubota S, Nakai Y, Sato N, Kuro-o M, Kohno S. 1994. Chromosome elimination in northeast Pacific hagfish, Eptatretus stoutii (cyclostomata, agnatha). *Journal of Heredity* **85**(5): 413-415.

Kubota S, Takano JI, Tsuneishi R, Kobayakawa S, Fujikawa N, Nabeyama M, Kohno SI. 2001. Highly repetitive DNA families restricted to germ cells in a Japanese hagfish (Eptatretus burgeri): A hierarchical and mosaic structure in eliminated chromosomes. *Genetica* **111**(1-3): 319-328.

Lander ES Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Leech DM, Wyngaard GA. 1996. Timing of chromatin diminution in the free-living, fresh- water Cyclopidae (Copepoda). *Journal of Crustacean Biology* **16**(3): 496-500.

Magnenat L, Tobler H, Müller F. 1999. Developmentally regulated telomerase activity is correlated with chromosomal healing during chromatin diminution in Ascaris suum. *Molecular and Cellular Biology* **19**(5): 3457-3465.

Matuszewski B. 1962. Oogenesis in Mikiola fagi Hart (*Cecidomyiidae, Diptera*). *Chromosoma* **12**(7): 741-&.

Mayer C. 2006-2010. PHOBOS – a tandem repeat search tool for complete genomes. . http://www.rub.de/spezzoo/cm/cm_phobos.htm.

McKay G, McQuade L, Murray J, von Sturmer S. 1984. Sex-chromosome Mosaicism in the Lemur-like Possum Hemibelideus lemuroides (Marsupialia: Petauridae). *Australian Journal of Biological Sciences* **37**(3): 131-136.

Metcalfe ME. 1935. The germ-cell cycle in Phytophaga destructor Say. *Q J Microsc Sci* **77**(308): 585-U510.

Metz CW, Schmuck ML. 1931. Differences between chromosome groups of soma and germ line in Sciara. *Proceedings of the National Academy of Sciences of the United States of America* **17**: 272-275.

Metz CW. 1926. Genetic evidence of a selective segregation of chromosomes in Sciara (Diptera). *Proceedings of the National Academy of Sciences of the United States of America* **12**: 690-692.

-. 1928. Genetic evidence of a selective segregation of chromosomes in a second species of sciara (diptera). *Proceedings of the National Academy of Sciences of the United States of America* **14**: 140-141.

-. 1934. Evidence indicating that in sciara the sperm regularly transmits two sister sex chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **20**: 31-36.

Mochizuki K. 2012. Developmentally programmed, RNA-directed genome rearrangement in Tetrahymena. *Dev Growth Diff* **54**(1): 108-119.

Moritz KB, Roth GE. 1976. Complexity of germline and somatic DNA in Ascaris. *Nature* **259**(5538): 55-57.

Muller F, Bernard V, Tobler H. 1996. Chromatin diminution in nematodes. *BioEssays* **18**(2): 133-138.

Müller F, Tobler H. 2000. Chromatin diminution in the parasitic nematodes Ascaris suum and Parascaris univalens. *International Journal for Parasitology* **30**(4): 391-399.

Müller F, Walker P, Aeby P, Neuhaus H, Felder H, Back E, Tobler H. 1982. Nucleotide sequence of satellite DNA contained in the eliminated genome of Ascaris lumbricoides. *Nucleic Acids Research* **10**(23): 7493-7510.

Müller F, Wicky C, Spicher A, Tobler H. 1991. New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in Ascaris lumbricoides. *Cell* **67**(4): 815-822.

Murray JD, McKay GM, Sharman GB. 1979. Studies on Metatherian sex-chromosomes .9. Sexchromosomes of the greater glider (*Marsupialia petauridae*). *Australian Journal of Biological Sciences* **32**(3): 375-386.

Nabeyama M, Kubota S, Kohno SI. 2000. Concerted evolution of a highly repetitive DNA family in eptatretidae (cyclostomata, agnatha) implies specifically differential homogenization and amplification events in their germ cells. *Journal of Molecular Evolution* **50**(2): 154-169.

Nakai Y, Kohno S. 1987. Elimination of the largest chromosome pair during differentiation into somatic cells in the japanese hagfish *Myxine garmani* (*Cyclostomata, Agnatha*). *Cytogenetics and Cell Genetics* **45**(2): 80-83.

Nakai Y, Kubota S, Goto Y, Ishibashi T, Davison W, Kohno SI. 1995. Chromosome elimination in three Baltic, south Pacific and north-east Pacific hagfish species. *Chromosome Research* **3**(5): 321-330.

Nakai Y, Kubota S, Kohno S. 1991. Chromatin diminution and chromosome elimination in four Japanese hagfish species. *Cytogenetics and Cell Genetics* **56**(3-4): 196-198.

Nemetschke L, Eberhardt AG, Hertzberg H, Streit A. 2010. Genetics, chromatin diminution, and sex chromosome evolution in the parasitic nematode genus strongyloides. *Current Biology* **20**(19): 1687-1696.

Nicklas RB. 1959. An experimental and descriptive study of chromosome elimination in *Miastor* spec. (*Cedicomyiidae, Diptera*) *Chromosoma* **10**(3): 301-336.

Niedermaier J, Moritz KB. 2000. Organization and dynamics of satellite and telomere DNAs in Ascaris: Implications for formation and programmed breakdown of compound chromosomes. *Chromosoma* **109**(7): 439-452.

Panelius S. 1971. Male germ line, spermatogenesis and karyotypes of *Heteropeza pygmaea* winnertz (*Cecidomyiidae, Diptera*). *Chromosoma* **32**(3): 295-&.

Pasternak J, Barrell R. 1976. Quantitation of nuclear-DNA in *Ascaris lumbricoides* - DNA constancy and chromatin diminution. *Genet Res* **27**(2): 339-&.

Pigozzi MI, Solari AJ. 1998. Germ cell restriction and regular transmission of an accessory chromosome that mimics a sex body in the zebra finch, Taeniopygia guttata. *Chromosome Research* **6**(2): 105-113.

-. 2005. The germ-line-restricted chromosome in the zebra finch: recombination in females and elimination in males. *Chromosoma* **114**(6): 403-409.

Pimpinelli S, Santini G, Gatti M. 1976. Characterization of Drosophila heterochromatin. II. C and N banding. *Chromosoma* **57**(4): 377-386.

Pokrzywa R, Polanski A. 2010. BWtrs: A tool for searching for tandem repeats in DNA sequences based on the Burrows-Wheeler transform. *Genomics* **96**(5): 316-321.

Posada D. 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**(7): 1253-1256.

Rasch EM, Lee CE, Wyngaard GA. 2004. DNA-feulgen cytophotometric determination of genome size for the freshwater-invading copepod Eurytemora affinis. *Genome* **47**(3): 559-564.

Rasch EM, Wyngaard GA. 2001. Evidence for endoreduplication: Germ cell DNA levels prior to chromatin diminution in Mesocyclops edax. *Journal of Histochemistry and Cytochemistry* **49**(6): 795-796.

-. 2006. Genome sizes of cyclopoid copepods (Crustacea): Evidence of evolutionary constraint. *Biological Journal of the Linnean Society* **87**(4): 625-635.

-. 2008. Gonomery and chromatin diminution in Mesocyclops longisetus (Copepoda). *Journal of Crustacean Biology* **28**(1): 180-184.

Redi CA, Garagna S, Zacharias H, Zuccotti M, Capanna E. 2001. The other chromatin. *Chromosoma* **110**(3): 136-147.

Rieffel SM, Crouse HV. 1966. The elimination and differentiation of chromosomes in the germ line of sciara. *Chromosoma* **19**(3): 231-276.

Rohwer F, Seguritan V, Choi DH, Segall AM, Azam F. 2001. Production of shotgun libraries using random amplification. *BioTechniques* **31**(1): 108-118.

Roth GE, Moritz KB. 1981. Restriction enzyme analysis of the germ linelimited DNA of *Ascaris-suum*. *Chromosoma* **83**(2): 169-190.

Roth GE. 1979. Satellite DNA properties of the germ line limited DNA and the organization of the somatic genomes in the nematodes *Ascaris suum* and *Parascaris equorum*. *Chromosoma* **74**(3): 355-371.

Sawabe TaI, T. 2005. Wizard® Genomic DNA Purification Kit Provides High-Quality Genomic DNA Template for Molecular Phylogenetic Studies on Copepod crustaceans. *[Internet]*.

Schmuck ML, Metz CW. 1932. The maturation divisions and fertilization in eggs of Sciara coprophila lint. *Proceedings of the National Academy of Sciences of the United States of America* **18**: 349-352.

Schnable PS Ware D Fulton RS Stein JC Wei FS Pasternak S Liang CZ Zhang JW Fulton L Graves TA et al. 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* **326**(5956): 1112-1115.

Schoeberl UE, Mochizuki K. 2011. Keeping the Soma Free of Transposons: Programmed DNA Elimination in Ciliates. *J Biol Chem* **286**(43): 37045-37052.

Schoenmakers S, Wassenaar E, Laven JSE, Grootegoed JA, Baarends WM. 2010. Meiotic silencing and fragmentation of the male germline restricted chromosome in zebra finch. *Chromosoma* **119**(3): 311-324.

Semeshin VF, Omelyanchuk LV, Alekseeva AL, Ivankina EA, Shevelyova NG, Zhimulev IF. 2011. DNA content in nuclei of Cyclops kolensis and Cyclops insignis (Crustacea, Copepoda). *Cell and Tissue Biology* **5**(3): 300-304.

Smith GP. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* **191**(4227): 528-535.

Smith-Stocking H. 1936. Genetic studies on selective segregation of chromosomes in Sciara coprophila Lintner. *Genetics* **21**(4): 421-443.

Spicher A, Etter A, Bernard V, Tobler H, MÃ¼ller F. 1994. Extremely stable transcripts may compensate for the elimination of the gene fert-1 from all Ascaris lumbricoides somatic cells. *Developmental Biology* **164**(1): 72-86.

Staiber W, Schiffkowski C. 2000. Structural evolution of the germ line-limited chromosomes Acricotopus. *Chromosoma* **109**(5): 343-349.

Staiber W, Wech I, Preiss A. 1997. Isolation and chromosomal localization of a germ line-specific highly repetitive DNA family in Acricotopus lucidus (Diptera, Chironomidae). *Chromosoma* **106**(5): 267-275.

Staiber W. 1988. G-Banding of germ line limited chromosomes in *Acricotopus lucidus* (*Diptera, Chironomidae*). *Chromosoma* **97**(3): 231-234.

-. 1989. Multivalent formation and pairing behavior of germ line limited chromosomes in male meiosis of *Acricotopus lucidus* (*Diptera, Chironomidae*). *Genome* **32**(6): 941-945.

-. 1991. Characterization of heterochromatin of germ line limited and soma chromosomes in *Acricotopus lucidus* (*Diptera, Chironomidae*) by differential banding methods. *Hereditas* **114**(1): 91-96.

-. 2000. Immunocytological and FISH analysis of pole cell formation and soma elimination of germ line-limited chromosomes in the chironomid Acricotopos lucidus. *Cell Tissue Res* **302**(2): 189-197.

-. 2006. Chromosome elimination in germ line-soma differentiation of Acricotopus lucidus (Diptera, Chironomidae). *Genome* **49**(3): 269-274.

Standiford DM. 1989. The effects of chromatin diminution on the pattern of C-banding in the chromosomes of Acanthocyclops vernalis Fischer (Copepoda: Crustacea). *Genetica* **79**(3): 207-214.

Streeck RE, Moritz KB, Beer K. 1982. Chromatin diminution in Ascaris suum: Nucleotide sequence of the eliminated satellite DNA. *Nucleic Acids Research* **10**(11): 3495-3501.

Stuart JJ, Hatchett JH. 1988. Cytogenetics of the hessian fly .2. Inheritance and behavior of somatic and germ line limited chromosomes. *Journal of Heredity* **79**(3): 190-199.

Teschke C, Solledar G, Moritz KB. 1991. The highly variable pentameric repeats of the AT-rich germline limited DNA in Parascaris univalens are the telomeric repeats of somatic chromosomes. *Nucleic Acids Research* **19**(10): 2677-2684.

Teschke C. 1985. Unusual cleavage pattern of the germ line limited satellite DNA of Parascaris equorum univalens. *Naturwissenschaften* **72**(4): 207-209.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL-W - Improving the sensitivity of progressive multiple sequences alignmentthrough sequences weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**(22): 4673-4680.

Tobler H, Smith KD, Ursprung H. 1972. Molecular aspects of chromatin elimination in Ascaris lumbricoides. *Developmental Biology* **27**(2): 190-203.

Tobler H, Zulauf E, Kuhn O. 1974. Ribosomal RNA genes in germ line and somatic cells of Ascaris lumbricoides. *Developmental Biology* **41**(2): 218-223.

Tobler H. 1986. The differentiation of germ and somatic cell lines in nematodes. *Results and problems in cell differentiation* **13**: 1-69.

Triantaphyllou AC, Moncol DJ. 1977. Cytology, reproduction, and sex determination of Strongyloides ransomi and S. papillosus. *Journal of Parasitology* **63**(6): 961-973.

Walton SM. 1971. Sex-chromosome mosaicism in pouch young of Marsupials *Perameles* and *Isoodon*. *Cytogenetics* **10**(2): 115-&.

Watson CM, Margan SH, Johnston PG. 1998. Sex-chromosome elimination in the bandicoot Isoodon macrourus using Y-linked markers. *Cytogenetics and Cell Genetics* **81**(1): 54-59.

Wyngaard GA, Goulden CE, Nourbakhsh A. 1994. Life-history traits of the tropical fresh-water copepod *Mesocyclops longisetus* (*Crustacea, Copepoda*). *Hydrobiologia* **293**: 423-427.

Wyngaard GA, Gregory TR. 2001. Temporal control of DNA replication and the adaptive value of chromatin diminution in copepods. *Journal of Experimental Zoology* **291**(4): 310-316.

Wyngaard GA, Rasch EM, Connelly BA. 2011. Unusual augmentation of germline genome size in Cyclops kolensis (Crustacea, Copepoda): further evidence in support of a revised model of chromatin diminution. *Chromosome Research*: 1-13.

Wyngaard GA, Rasch EM, Manning NM, Gasser K, Domangue R. 2005. The relationship between genome size, development rate, and body size in copepods. *Hydrobiologia* **532**(1): 123-137.

Zagoskin MV, Grishanin AK, Korolev AL, Palenko MV, Mukha DV. 2008. Characterization of Cyclops kolensis inter-simple sequence repeats in germline and postdiminution somatic cells. *Doklady Biochemistry and Biophysics* **423**(1): 337-341.

Zotkevich EA, Boykova TV, Akifyev AP, Tereza EP, Pomazkova GI, Melnik NG, Timoshkin OA, Zhymulev IF. 2008. Highly conserved repeated redundant DNA in two Cyclops species. *Doklady Biochemistry and Biophysics* **420**(1): 139-141.

Zufall RA, Robinson T, Katz LA. 2005. Evolution of developmentally regulated genome rearrangements in eukaryotes. *J Exp Zool Part B* **304B**(5): 448-455.