Title
Evolutionary Comparison of X-Chromosome Dosage Compensation Across Caenorhabditis Species

Author
Schartner, Caitlin Marie

Publication Date
2016

Peer reviewed|Thesis/dissertation

Evolutionary Comparison of X-Chromosome Dosage Compensation Across *Caenorhabditis* Species

by

Caitlin Marie Schartner

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Barbara J. Meyer, Chair
Professor Michael B. Eisen
Professor Gian Garriga
Professor Doris Bachtrog

Summer 2016

Evolutionary Comparison of X-Chromosome Dosage Compensation Across *Caenorhabditis* Species

Abstract

Evolutionary Comparison of X-Chromosome Dosage Compensation Across *Caenorhabditis* Species

by

Caitlin Marie Schartner

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Barbara J. Meyer, Chair

Many species determine sex by assessing sex-chromosome dose. In flies, mammals, and nematodes, males have one X chromosome and females or hermaphrodites have two. However, both sexes require the same level of expression for most genes; thus they evolved a process called dosage compensation to equalize X-linked gene expression between the sexes. Failure to carry out this process causes lethality in one sex. Flies, mammals, and nematodes have independently evolved very different dosage compensation mechanisms, showing that multiple strategies can be employed to achieve dosage compensation. We chose to investigate four nematode species with shared ancestry to ask whether and how essential dosage compensation mechanisms are able to evolve over time. Examples of dosage compensation mechanisms in closely related species provide evidence to better understand how genes with different expression constraints can be regulated simultaneously across chromosome-wide territories.

Here, I compare chromosome-wide dosage compensation mechanisms across four *Caenorhabditis* species (less than 30 million years diverged): *C. elegans*, *C. briggsae*, *C. nigoni*, and *C. tropicalis*. We took advantage of improved sequencing technology to create chromosome-level genome assemblies for *C. nigoni* and *C. tropicalis*. We also devised genome editing strategies for the non-model species *C. briggsae*, *C. nigoni*, and *C. tropicalis*. Using genome editing, we created strains for phenotypic and biochemical assays to assess the function of orthologous genes in dosage compensation. Key subunits of the dosage compensation machinery and the genetic hierarchy that regulates the sex-specific function of the machinery share conserved action across species. However, remarkably, the binding sites on the X chromosome and the DNA sequence motifs within binding sites that drive X-specific binding have diverged at least twice within 30 million years of evolution. Since the dosage compensation machinery acts to modify the structure of the *C. elegans* X chromosome by bringing binding sites together, divergence in binding site location could result in a different X-chromosome structure in these species. Future studies of the rapid divergence of binding

sites that characterize X-chromosome structure and X-linked gene expression could bring us closer to understanding how nematode dosage compensation operates.

Contents

List of Figures

List of Tables

Acknowledgments

I would first like to thank my advisor, Barbara Meyer, for her dedication and guidance throughout this project. I couldn't have done it without her. I would also like to thank Te-Wen Lo, Ed Ralston, and Denise Lapidus, for all their work on this project and for all the advice, pep talks, and their friendship. Thanks to Bobby Farboud and Bayly Wheeler for being so good and nice and for allowing me to join their postdoc lunch club. Thanks to Christi Preston, Claire Fassio, Erika Anderson, Nick Fuda, Cathy Pickle, and others. I would also like to thank the staff who poured the plates, washed the dishes, and kept the lab running.

I'm grateful to the worm evolution community, especially Eric Haag, Ron Ellis, and Asher Cutter for their generosity, encouragement, and advice. I would also like to thank Wormbase, Wormbook, and the *Caenorhabditis* Genetics Center.

I must thank Akemi Kunibe, Char Bashore, and Adrianne Pigula for introducing me to the cryomill, and Georjana Barnes for sharing. Grinding worms by hand was one of my least favorite things.

Thanks to those I've learned from outside of lab, including Anne Lachlan, Jose Estrada, and the volunteers, students, organizers of the Prison University Project.

Thanks to Aileen Kelly and Mary Lin, who have been the best roommates and friends. I will always have such happy memories of the time spent in our House o' Dreams. To Alejandra Figueroa-Clarevega and Julie Choe, love yooouu <3!!! Thanks to Adrianne Pigula for being such a generous host in my first days in Berkeley and for all the love and support ever since. I am eternally grateful that she answered my spam email. Thanks to my writing buddy Akemi Kunibe for getting me started. To Adrienne Maxwell, thanks for teaching me zebrafish development, jazz walking, and introducing me to the Cat Club. To Aisha Ellahi, Lauren Wagner, and Adam Merberg, who helped me get through low points, when I didn't even know what I needed. Thanks to my classmates and friends, Char Bashore, Falina Williams, Livy Wilz, Nicole Beier, Rob Thistle, Nina Glazier, Joy-El Talbot, and others.

I am grateful for the experiences I had in Madison. I thank Phil Anderson, Jasmine Parvaz, Leah Frater Rubsam, and Lisa Harms, for their patience and encouragement. Thanks to Colleen Schultz for being such a dear friend, and to Beth Wyman, Hilary Truchan, and Alex Means, for being the greatest team.

I owe so much to my family. My parents, Christine and Dale Schartner, made my education a priority. They supported my interests in math, science, art, and music, and let me find my own way. Thanks to my brothers, Dom, Ted, and Kit, and my sisters, Diane, Sarah and Katie Mak, for their love and support. Thanks to Maximilian Herkender for everything.

Chapter 1

Introduction

Mechanisms of dosage compensation evolved independently in multiple lineages to equalize gene expression between the sexes in response to sex chromosome evolution. In XY species with genetic sex determination, females have two X chromosomes and males have one. Although the X and Y chromosomes originate from a single autosome, the male Y chromosome is subject to degeneration and gene loss and is sometimes lost completely [1]. The loss of genes on the Y chromosome reduces the copy number of X-linked genes in males from two to one, which can be lethal without dosage compensation. Dosage compensation mechanisms differ across species, although the best-studied examples all function through a complex that binds across the X chromosome(s) in one sex. The best-studied examples of dosage compensation, from flies, mammals, and nematodes (figures 1.1 1.2), exemplify the diversity of solutions to the X-chromosome gene dose problem. In flies, the single X chromosome in XY males is upregulated to equal expression in XX females. Conversely, in mammals, one of the two female X chromosomes is inactivated, resulting in expression equal to that in XY males. Nematodes provide a third example, where the two hermaphrodite X chromosomes are both downregulated by half to equal expression from the single male X. In each of these lineages, a unique collection of genes was co-opted for dosage compensation, leading to diverse molecular mechanisms, revealing that very different tactics can be employed to equalize gene expression between the sexes. To understand how this essential process changes over time, characterization of the process among more closely-related species is necessary. The aim of this dissertation is to improve our understanding of transcriptional regulation by investigating the evolution of essential chromosome-wide dosage compensation mechanisms in *Caenorhabditis*.

In this chapter, I begin by discussing the history of nematode research and the recent advances that have made *Caenorhabditis* nematodes robust models for evolutionary studies. I next review the current understanding of dosage compensation mechanisms in diverse organisms, discussing the regulatory mechanisms that activate dosage compensation in one sex, the machinery used to accomplish dosage compensation, and the DNA sequences that drive X-specificity. I also discuss what is known about the evolution of these dosage compensation mechanisms.

In chapter 2, I compare dosage compensation mechanisms in four nematode species, *C. elegans*, *C. briggsae*, *C. nigoni*, and *C. tropicalis*. I describe the contributions we have made to produce high quality, chromosome-level genome assemblies in *C. nigoni* and *C. tropicalis*. I also show that key components of the dosage compensation machinery and the genetic hierarchy regulating sex-specificity are conserved in the *C. briggsae* clade, however the X-chromosome sequences that recruit the dosage compensation machinery have diverged at least twice within 30 million years of evolution.

1.1 *Caenorhabditis* species as model organisms

It's an exciting time to study worms. *C. elegans* is a convenient model organism for many reasons, including small size, short life cycle, and transparent body. However, these features would mean little without the wealth of information and tools scientists have collected and built since nematode research began. In recent years, the discovery of new *Caenorhabditis* species, improvements in sequencing technology, and the development of new tools for genome editing has made *Caenorhabditis* a powerful evolutionary model system. Without these advances, this project would not be possible. Here I describe events leading to this moment in nematode biology.

In the late 1880s, some of the first nematode researchers studied ascarids, the large, parasitic nematodes that infect horses. Eduoard van Beneden and others used these worms to study fertilization, meiosis, and development [2]. The species *Ascaris megalocephala* (*Parascaris equorum*, figure 1.3) has only four chromosomes, which simplified studies of its cytology [2]. In this species, Theodor Boveri proposed that although chromosomes differ in appearance through the cell cycle, they are always present ("chromosome continuity"). This work also contributed to the Sutton-Boveri chromosome theory of heredity (proposed at the same time by Walter Sutton) [2, 3]. Although these horse parasites were useful for cytology, they were not suitable for culture in the laboratory.

Around the same time, species in the genus *Rhabditis* emerged as more appropriate models. Early work in *Rhabditis* began to explore the development of the embryo, gastrulation, post-embryonic development, molting, and the dauer stage [4]. In 1944, Margaret Briggs Gochnauer identified a self-fertilizing hermaphrodite species (*C. briggsae*, initially named *Rhabditis briggsae*) [5] and characterized its growth in culture with several different bacterial species [4]. Identifying a monoxenic culture with a single bacteria species that supports growth was a significant contribution to the development of a nematode model for genetics [4].

Developing a model system was a major endeavor, and recruiting more scientists to work on nematodes made nematode research more valuable. In 1948, Ellsworth Dougherty and Hermione Calhoun wrote a letter to Nature, describing the potential they saw in free-living nematodes as genetic model organisms and calling for other researchers to join them in studying these species [6]. They cited the worms' short life-cycle, small number of cells, and ability to grow on agar plates seeded with bacteria, as well as their "good cytological features and convenient sex patterns". The first *C. briggsae* mutant was characterized by Victor Nigon and Dougherty in 1950 [7] (Nigon is the namesake of the close *C. briggsae* relative, *C. nigoni* [8]). Sydney Brenner became interested in *Caenorhabditis* because he was looking for an animal model with a simple nervous system [9]. He chose *C. elegans* over *C. briggsae* for his groundbreaking 1974 paper, "The Genetics of *Caenorhabditis elegans*", in which he isolated and mapped mutations in about 100 *C. elegans* genes [9]. A

likely reason for this choice was the difference in behavior between *C. briggsae* strains and the N2 Bristol *C. elegans* strain [4]. A spontaneous mutation in the *npr-1* gene that was acquired in the Dougherty lab made the N2 strain far less likely to congregate in clumps or burrow in agar [10]. For better[1] or worse[2], *C. elegans* became the model nematode species.

Scientists developed genetic tools to make *C. elegans* the model it is today. The number of mutant strains has increased from hundreds in Brenner's mutant collection to over a million *C. elegans* strains [12]. *C. elegans* was the first animal to be sequenced in 1998 [13], and now it has the most complete animal genome available. The *C. elegans* genome assembly is nearly complete because it is easier than most genomes to assemble – it is only about 100 Mb long and repetitive regions are relatively short – and because researchers have diligently contributed their annotations and corrections over the past 20 years. *C. elegans* research has led to important discoveries that resulted in three Nobel prizes: the first to Brenner, John E. Sulston, and H. Robert Horvitz in 2002 for their work on development and apoptosis, the second to Andrew Z. Fire and Craig C. Mello in 2006 for the discovery of RNA interference, and the third to Osamu Shimomura, Martin Chalfie, and Roger Y. Tsien in 2008 for the discovery and development of green fluorescent protein (GFP) (Martin Chalfie demonstrated applications of GFP in *C. elegans*).

Improved sequencing technology and new tools for reverse genetics have made it possible to do genetic research in non-model species. The *C. briggsae* genome was published in 2003 [14], then improved by mapping 30,000 polymorphisms between two *C. briggsae* strains in 2010 [15]. Now, nearly complete, *de novo*, chromosome-level assemblies are possible without genetic mapping. As described in chapter 2, we have contributed to the *Caenorhabditis* sequencing effort by sequencing the *C. nigoni* and *C. tropicalis* genomes. Precise, heritable genome editing is now possible with Zinc finger nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), and CRISPR/Cas9. As a lab, we have used all three tools and have developed protocols to use these tools in non-model species [16, 17, 18]. As described in chapter 2, we used TALENs and Cas9 to introduce mutations and epitope tags to interrogate gene function in four *Caenorhabditis* species.

The newly-expanded, rich *Caenorhabditis* phylogeny allows investigation of evolution at various timescales. Dozens of wild *C. elegans* and *C. briggsae* isolates and many new species have been identified across 6 continents recently [19, 20], including 40 of the 50 known

[1] *C. briggsae* lacks environmental RNA interference, which was a major *C. elegans* discovery and a useful tool for research. The *C. elegans* SID-2 transmembrane protein, required to ingest and transport double stranded RNA, is not conserved in *C. briggsae* [11]. Also, gravid *C. elegans* hermaphrodites hold more embryos than *C. briggsae* and their carcasses dissolve in bleach simultaneously, making it much easier to synchronize *C. elegans* populations.

[2] Several *C. briggsae* clade species have been identified at various evolutionary distances, but the first *C. elegans* sister species (more closely related than *C. briggsae*) was only discovered recently (Karin Kiontke, unpublished).

Caenorhabditis species, which were identified in the past 10 years [21, 22]. Features that make *C. elegans* an excellent model are shared with its close relatives, including the ability to survive freezing. Thus, collections may be maintained and shared easily, and they are far less likely to inadvertently accumulate mutations in the laboratory than organisms that require continuous propagation.

Although *Caenorhabditis* species look similar to each other, their genomes are incredibly diverse (figure 1.3). The most divergent *Caenorhabditis* species pairs are less similar than mouse and zebrafish, with approximately 0.7 substitutions per site between *C. briggsae* and *C.* sp. 1 [21]. Also, a range in diversity levels can be found within *Caenorhabditis* species. For example, *C. brenneri*, is hyperdiverse, with diversity at synonymous sites that is 150- and 100-times greater than found in humans and *C. elegans*, respectively [23]. *C. elegans* and *C. briggsae* diverged approximately 15-30 million years ago (figure 1.2) [24], and their sequence divergence is about 0.3 substitutions per site, slightly greater than human and mouse [14, 21]. Androdioecy (having hermaphrodite and male sexes) arose independently at least 3 times in the *Caenorhabditis* lineage: in *C. elegans*, *C. briggsae*, and *C. tropicalis*, also in the *C. briggsae* clade. These three species and the close *C. briggsae* relative *C. nigoni* were selected for this study because we observed divergence between *C. elegans* and *C. briggsae* and wanted to investigate shorter timescales.

1.2 Sex chromosome evolution and the need for dosage compensation

Sex determination mechanisms are incredibly diverse and can evolve quickly [25]. Jonathan Hodgkin showed that they are also highly mutable, by creating *C. elegans* strains with diverse sex determination mechanisms. In these mutant strains, he converted each autosome to a sex chromosome and created temperature-based, maternal, and tetraploid sex determination mechanisms [26]. He showed that very small changes can convert downstream genes in the sex determination pathway into the primary regulator [26].

Species only require dosage compensation when genetic sex determination occurs by a sex chromosome pair in which each chromosome differs in gene content. In general, an autosome first acquires a sex-determining locus and becomes a sex chromosome [27]. Recombination can occur in the homogametic sex (XX female or ZZ male), but additional sex-determining loci and chromosomal inversions suppress recombination in the heterogametic sex (XY male or ZW female), which leads to gene loss (genetic erosion) on the Y or W chromosome [1]. In some cases, including in nematodes, the Y chromosome was completely lost, and sex determination occurs by assessing the X to autosome ratio [1, 28]. Genetic erosion or complete loss of the Y or W chromosome leads to gene copy number imbalance between the sexes. Although most genes are not predicted to be haploinsufficient (insufficient in one copy), si-

multaneous hemizygosity for many genes can be lethal [29].

In Susumu Ohno's seminal book, *Sex chromosomes and sex-linked genes*, he hypothesized that dosage compensation could occur in one of two ways: 1) by upregulation of sex-linked genes specifically in the heterogametic sex, or 2) by upregulation of sex-linked genes in both sexes followed by a decrease in expression in the homogametic sex [30]. The latter mechanism balances X or Z and autosomal expression, like the former, but sex-specific downregulation of X or Z without upregulation in both sexes merely balances sex-chromosome expression between the sexes, leaving expression of X or Z to autosomes unbalanced [31]. The best-studied examples of dosage compensation occur in flies, which upregulate the male X chromosome, and mammals and nematodes, species that turn down X chromosome expression in the XX-female or hermaphrodite.

Fly, mammal, and nematode dosage compensation mechanisms share common themes. In each case, existing cellular molecules were coopted for dosage compensation. Proteins and/or RNAs direct a dosage compensation complex specifically to the X chromosome(s) in one sex. The dosage compensation machinery binds to discrete sites on X and spreads. Dosage compensation mechanisms act on the X chromosome at the level of transcription, in ways that include modifying histones and restructuring chromosomes to affect RNA polymerase II activity. Although these mechanisms arose independently, learning about the various ways species accomplish dosage compensation can teach us general principles of transcriptional regulation. Also, evolutionary studies within these groups may provide some context for the divergence we see among *Caenorhabditis* species.

1.3 Dosage compensation in *Drosophila melanogaster*

Discovery

Dosage compensation was first discovered in *Drosophila melanogaster* (reviewed in [32]). In the 1930s, H.J. Muller showed that X-linked genes with dose-dependent phenotypes, like the *white* gene, which produces red eye pigmentation, had a stronger effect per copy in males than in females. Two copies of the *white* gene were required for red eyes in females, and a single copy produced only pink eyes. In males, a single copy of *white* produced red eyes and two copies produced deeper red eyes. Given the disparity in effect per gene copy, he proposed that sex-specific modifiers exist to compensate for the difference in X-linked gene dose between males and females [33]. In 1965, Mukherjee and Beerman confirmed that a dosage compensation mechanism equalizes transcription between the two female X chromosomes and the single male X in a study that measured transcription radioactively from the large, multi-copy, polytene chromosomes of the fly salivary gland [34]. By 1980, Belote and Lucchesi identified three genes with male-lethal mutant phenotypes, *msl-1*, (male-specific lethal), *msl-2*, and *mle* (*maleless*), then showed that these genes function to increase X-linked transcription in

male dosage compensation [35, 36]. Around the same time, Thomas Cline discovered the female-specific primary regulator of sex determination and dosage compensation, *sex lethal* (*Sxl*), and *daughterless* (*da*), a gene required in mothers for *Sxl* expression in their daughters [37, 38]. In this section, I briefly discuss the sex-specific regulation, machinery, and X-specificity of *Drosophila melanogaster* dosage compensation.

Sex-specific regulation of *D. melanogaster* dosage compensation

In *D. melanogaster*, sex determination and dosage compensation are linked through *sex lethal* (*Sxl*). *Sxl* is a binary switch gene; a gain of function in *Sxl* is male-lethal and a loss of function is female-lethal [37, 39]. The Sxl protein binds to a 3' splice site in the *transformer* (*tra*) transcript, causing an exon with a premature termination codon to be skipped to enable female-specific Tra translation [40, 41, 42]. Sxl protein also promotes female sex determination through other targets [43] and inhibits dosage compensation in the female by binding to the mRNA of the subunit essential for formation of the MSL complex, *msl-2*, to inhibit proper splicing, transport to the cytoplasm, and translation [44, 45, 46]. In XY cells, *Sxl* is inactive, leading to male development and loading of the MSL complex on the single male X chromosome [39, 47].

Sxl is subject to complex regulation to ensure that the protein is only expressed in the female. There are several alternate splice forms of Sxl mRNA that are sex-specific, developmental stage-specific, and tissue-specific [48, 49]. The maternally provided signal transducer *daughterless* (*da*) and the zygotically expressed X signal elements (XSEs), including *sisterless* genes (*sisA*, *sisB* , and *sisC*), promote *Sxl* expression from its early promoter ($\mathrm{SxlP_e}$) [37, 38, 50, 51, 52, 53, 54]. Sxl protein binds its own mRNAs expressed from the constitutive maintenance promoter ($\mathrm{SxlP_m}$) to produce female-specific splice variants without premature stop codons in a positive autoregulatory feedback mechanism [47, 55, 56]. The downstream Tra also feeds back to promote female-specific *Sxl* expression [57].

Early models predicted that inhibitory autosomal signal elements (ASEs) would be critical in sensing the ratio of X chromosomes to autosomes (X:A) [58]. Consistent with X:A sensing, when the X:A ratio is one in either diploid (XXAA) or haploid animals (XA), the animals develop as females [59]. The X:A ratio is 0.5 in males (XYAA), and triploid animals (XXAAA) with an intermediate X:A ratio of 0.67 develop as intersex animals [59]. However, *deadpan* (*dpn*), the strongest ASE identified, only affects sex determination weakly [60]. Instead of assessing the X:A ratio, sex is determined by an X-chromosome counting mechanism that depends on developmental timing. In wild-type embryos, 14 cycles of nuclear divisions occur before zygotic expression begins and the embryo cellularizes, but haploids and triploids cellularize after 15 and 13 cycles, respectively [61]. The time window to assess the X-chromosome dose occurs between the onset of XSE transcription and the cellularization of the embryo. The longer it takes to achieve cellularization, the longer XSE products can accumulate. Haploids have an extra cell cycle to reach that threshold, so they develop

as female, even though they only have one X chromosome [62]. Triploids have one fewer cell cycle, so some cells develop as male and others as female, even though they have two X chromosomes [62].

D. melanogaster dosage compensation machinery

Dosage compensation is achieved by the MSL complex, which increases transcription from the single male X chromosome to approximately equal expression from the two female X chromosomes [35, 36, 63]. The MSL complex, which contains five proteins (MSL1, MSL2, MSL3, MLE, and MOF) and two redundant noncoding RNAs (roX1, and roX2), binds to discrete sites, then spreads to gene bodies across the male X chromosome [64, 65, 66]. MOF (males on first) acetylates H4K16 at genes throughout the genome as a member of the non-specific lethal (NSL) complex and was co-opted to acetylate histones for dosage compensation [67, 68, 69]. The human ortholog hMOF also acetylates H4K16 and non-histone proteins, including p53 [70]. H4K16ac is associated with increased transcription, but the direct mechanism of transcriptional upregulation is unclear [71]. Two models were proposed, a transcriptional initiation model, in which RNA polymerase II recruitment is increased, and an elongation model, in which RNA polymerase II is better able to progress across hyperacetylated gene bodies [72, 73].

MSL-2 contributes to regulation, assembly, and stabilization of the MSL complex. As stated above, the sex-specificity of dosage compensation is directed by *msl-2*, which is only translated in males, in the absence of Sxl. In addition to its role in driving sex-specificity, MSL2 is an E3 ubiquitin ligase that targets itself and the MSL subunits MSL1, MSL3, and MOF for ubiquitination, in part to ensure proper stoichiometry of MSL complex proteins [74, 75, 76]. MSL2 also stabilizes the MSL1 dimers that act as a scaffold for the complex [74, 45, 77]. MSL1-MSL2 dimers also play a key role in binding to specific X-chromosome sites, described below [74, 45, 77].

The RNA components of the MSL complex, roX1 and roX2 (RNA on X), and MLE, an RNA-binding protein homologous to human RNA helicase A, also contribute to MSL complex assembly [78]. The redundant roX1 and roX2 ncRNAs differ in length (3.7kb and 0.6 kb, respectively) and are dissimilar in sequence (except microhomology in roXbox motifs), but they share structural features (roXbox stem-loops) [78, 79]. Mutations in either *roX1* or *roX2* alone do not cause mutant phenotypes, but the complex does not assemble on the X chromosome in the male-lethal *roX* double mutant [78]. MLE, an RNA/DNA helicase, actively remodels roX stem loops to initiate MSL complex assembly [80, 81, 82, 83, 84]. MLE also associates with proteins involved in RNA processing, RNAi, chromatin remodeling, DNA repair, and translation [85, 84].

X-specificity of *D. melanogaster* dosage compensation

The *D. melanogaster* MSL complex binds to discrete sites called chromatin entry sites (CES) or high affinity sites (HAS) on the X chromosome, then spreads to gene bodies [64, 65, 66]. Binding at HAS is positively correlated with higher GC content, enrichment of H3K9ac, H3K36me3 (a mark bound by MSL3 [86]), and a GA-repeat motif named MRE (MSL recognition element) [87]. The MRE motif is functionally important; mutating this sequence disrupts binding [88, 89]. MRE motifs are also enriched at the few MSL-bound autosomal sites [89]. Although the MRE contributes to binding, the MRE motif cannot be driving X-specificity because it is found throughout the genome and is only slightly enriched on the X chromosome (approximately two-fold) [88].

Each component of the MSL complex influences complex assembly on the X chromosome. MSL-1 and MSL-2 are able to bind to the X chromosome at 35-70 discrete HAS in the absence of other protein components [74, 45, 77]. MSL1 and MSL2 require MSL3, MLE, and MOF to spread to active genes across the X chromosome [90, 74, 91]. MOF and MSL3 interactions with MSL1 in particular are critical for the spreading of the complex to the bodies of active genes and to some high affinity sites (HAS) [77]. MSL3 may contribute to the spreading of the MSL complex to expressed genes in *cis* by binding to histone 3 lysine 36 trimethyl marks (H3K36me3) associated with transcription [86]. The *roX1* and *roX2* loci are thought to be the among the first HAS sites bound and are essential nucleation sites for appropriate MSL X-targeting [64, 92, 93]. Autosomal *roX* insertions show that targeting of the MSL complex occurs in *cis*, at *roX* loci [65], and can also occur in *trans*; when MSL components are overexpressed, the autosomally-expressed *roX* transcripts can associate with the MSL on X [92]. Overexpression of MSL proteins only partially rescues lethality in *roX* mutants [93].

Other proteins may also contribute to MSL complex assembly on the X chromosomes. Sequence-specific binding at the MRE motif is facilitated by the CLAMP protein (chromatin-linked adaptor for MSL proteins), which has seven zinc fingers [69, 94]. CLAMP can bind MRE sequences *in vitro* and *in vivo* [95]. The NSL complex may also play a role in MSL complex-binding to HAS on the X chromosome [69]. However, both CLAMP and the NSL complex bind to sequences throughout the genome, indicating that other factors must provide X-specificity [69, 94].

The MSL complex is recruited to HAS and spreads to active gene bodies, however about 25% of active genes on X do not appear to be bound by the MSL [66]. H4K16 acetylation is associated with nearly all active genes on X, which suggests that the complex could associate with other sites on X transiently [96]. Another possibility is that many genes on X are subject to MSL-independent buffering, similar to gene-by-gene mechanisms that regulate expression in autosomal aneuploidies [97]. Perhaps related to MSL-independent buffering, a non-canonical form of dosage compensation occurs at earlier developmental stages, before

MSL-based dosage compensation is established [98].

1.4 Dosage compensation in other flies

Dosage compensation is essential in *Drosophila*, and researchers have asked whether dosage compensation mechanisms are conserved in other flies. Comparisons across related species have provided insight into how other species regulate sex determination and dosage compensation, when this form of dosage compensation arose, and whether changes have occurred in X-specific targeting of the MSL complex between *Drosophila* species. Although some evidence suggests that positive selection has acted on the MSL complex and its binding sites, homologs of MSL complex components and the MRE motif found at binding sites appear to be conserved across *Drosophila*. Lastly, evolution of X-enriched satellite repeats provide an interesting hypothesis to explain potential changes in targeting the MSL complex to the X chromosome.

Divergence in sex-specificity across insects

The *Sxl* gene is conserved across many insect species, including houseflies, mosquitoes, and beetles, but *Sxl* only functions as a master regulator of sex determination and dosage compensation in *Drosophila* [99]. *Sxl* orthologs do not have sex-specific expression, and the medfly (*Ceratitis capitata*) or housefly (*Musca domestica*) *Sxl* orthologs do not cause transformation when ectopically expressed in *D. melanogaster* XY animals (figure 1.4) [100, 101]. Instead, *tra*, which is downstream of *Sxl* in *Drosophila* species, appears to act as the major switch gene for sex determination in many insects [102]. Thus, the *D. melanogaster* mechanism for counting X chromosomes appears to be a recent modification to an ancient pathway.

Conservation of dosage compensation machinery and X-specificity

The dosage compensation mechanism reliant on the MSL complex is at least 55 million years old. It evolved before the split between *Drosophila* and *Chymomyza* (55 MYA) (figure 1.4). MSL-1, MSL-2, and MSL-3 homologs function in the MSL complex in 12 fly species, including 9 *Drosophila* species, *Z. tuberculatus*, *H. pictiventris*, and *C. procnemis*, by male X-specific staining with *D. melanogaster* α-MSL-1, α–MSL-2, and α–MSL-3 antibodies [103]. MSL complex-binding sites in multiple *Drosophila* species were enriched for H4K16ac marks [104].

The RNA component of the MSL is also conserved across $\sim$40 million years. A search among 35 fly species identified 47 new roX orthologs with conserved synteny, roXbox motifs, and roXbox-containing hairpin structures, a subset of which were confirmed to be functional by observing male-specific expression and targeting of lncRNAs to the X chromosome [79]. Conserved roX hairpins also were shown to have function in MSL targeting and H4K16 acetylation in *roX* homologs across 9 *Drosophila* species [105, 106].

In *D. miranda*, the conserved MRE motif has spread to introduce MSL complex-binding sites on newly acquired sex chromosomes. In this species, two fusion events created sex chromosome regions of different ages, referred to as strata: the ancient X chromosome, XL (formed over 60 million years ago (MYA)), the right arm of the X chromosome, XR (15 MYA), and the neo-X chromosome (1-2 MYA) [103, 107, 108]. Each of the three stratum consists of about 20% of the genome [108]. The *D. melanogaster* MRE motif can be identified among XL and XR MSL-binding sites, and a similar motif with weaker consensus to the MRE motif can be identified in MSL-bound sites of the neo-X [107]. Recruitment sequences from the *D. miranda* XL, XR, and neo-X (but not the homologous neo-Y sequence) were able to recruit the MSL complex in *D. melanogaster*[107]. Today, the markers of dosage compensation are fully established on the XR, with MSL-binding distribution, H4K16ac enrichment, and H3K36me3 enrichment similar to the ancient XL, however, dosage compensation appears incomplete on the neo-X, with fewer active genes bound by the MSL [107]. In terms of gene expression, the onset of dosage compensation is developmentally delayed for all *D. miranda* sex chromosome strata and incomplete for the neo-X (expression of some functional neo-Y genes contributes to approximately equalize total neo-X and neo-Y expression between males and females) [108].

MSL-binding sites were acquired on the *D. miranda* neo-X by domestication of a transposable element (TE), called ISX [109]. The ISX element evolved from a TE, called ISY, when a small deletion in ISY created a functional MRE motif. This element was amplified, then retained nearly exclusively on the neo-X [109]. The ISX was further refined by gene conversion events that transferred beneficial mutations among ISX loci [110]. In a similar manner on the XR, a related TE, called ISXR, was apparently domesticated to bind the MSL and amplified in a burst [109]. The ISXR element was more difficult to identify because more time has passed since the XR fusion, allowing more fine-tuning of the binding site sequence and erosion of non-binding ISXR regions [109]. It may be that evidence of more ancient TE rewiring events (in this lineage and others) has been lost [109]. Evidence of TE rewiring was not evident on the neo-X strata in *D. willistoni*, suggesting the evidence was either lost or the MRE motifs were gained by another mechanism [79]. Minor changes in the C/T-rich splicing signal at the 3' end of introns could have created MRE motifs in genes, and consistent with this hypothesis, a strand bias was observed for MRE motifs in introns [79].

Divergent X-enriched satellite repeats may contribute to X-specificity

X-specificity of MSL targeting cannot be explained by the *roX* loci and the MRE motif alone, even in *D. melanogaster*. Additional DNA sequences could contribute to X-specificity and its potential divergence across species. It was long known that the *Drosophila* X chromosome is highly enriched for a complex satellite repeat (SR) with 359 bp monomers related to the 1.688 g/cm^3 satellite [111]. This SR may distinguish X from autosomes in sequence and chromatin structure [112], however there is no direct evidence for its involvement in dosage compensation [113]. Recently, comparative studies have shown that highly enriched

X-chromosome SRs have undergone repeated turnover and divergence in *Drosophila* species, providing an interesting hypothesis to explain MSL X-specificity and potential MSL-targeting differences across species [114, 115].

1.5 Mammalian X-chromosome inactivation (XCI)

Discovery

The mammalian mechanism of dosage compensation was first described in the 1960s, when Mary Lyon proposed random X-chromosome inactivation (XCI) as a mechanism to compensate for increased X-chromosome dose in female mammals [116]. This hypothesis was based on two observations. First, Barr and Bertram identified a deeply staining body in female nuclei (the Barr body) that Ohno and Hauschka later proposed to be one of the two female X chromosomes [117, 118]. They observed that tetraploid females have two of these Barr bodies in each cell and tetraploid males have one [118]. The second important observation was variable penetrance seen in female mice heterozygous for X-linked genes or variegated appearance of X-linked coat color genes [116]. Lyon reasoned that human XCI could account for observed frequencies of X-linked diseases in heterozygous females and for the survival of individuals with Turner syndrome (XO females) and Klinefelter syndrome (XXY males), when most chromosome aneuploidies are lethal [119]. XCI continues to be an important field of study for human health, since XCI can affect many phenotypes directly and indirectly, in complex ways, and likely contributes to differential life expectancy and disease outcomes in females and males [120]. In this section, I review the mammalian dosage compensation, with a focus on the mechanism for targeting one of the two female X chromosomes for inactivation.

Sex-specific regulation and XCI machinery

XCI equalizes gene expression between males and females by "counting" X chromosomes and epigenetically silencing one of the two female X chromosomes in a heritable manner. The X chromosomes are counted and distinguished from autosomes by the presence of the X-inactivation center (Xic). Rastan and Robertson determined that XCI only occurs in cells that contain more than one Xic, which they identified and mapped by staining for Barr bodies in a series of mouse embryonic stem cell lines containing X-chromosome deletions [121, 122]. Further mapping narrowed down the Xic region to a 1 Mb region in human and 450 kb region in mouse that contains several protein-coding and noncoding genes [123, 124]. The key regulator within the Xic, is a a 17 kb long non-coding RNA (lncRNA) expressed only from the inactive X chromosome (Xi) called Xist (X-inactive-specific-transcript) [125, 126, 127]. The *Xist* gene is under complex regulation to ensure proper dosage compensation only in XX animals and only on one of the two X chromosomes.

Many protein and RNA factors appear to contribute to chromosome counting and XCI regulation in mouse by acting on *Xist* expression in *cis* and *trans*. *Tsix*, which encodes a lncRNA, overlaps with *Xist* and is transcribed antisense to *Xist*. *Tsix* is ultimately expressed only on the active X chromosome (Xa) where its transcription represses *Xist* expression in *cis* [124, 128, 129]. XCI initiation is also regulated by pluripotency factors. When present in two copies, Rnf12, the E3 ubiquitin ligase encoded within the Xic, activates XCI in *trans* by targeting Rex1, a pluripotency marker, for degradation [130, 131, 132]. XCI is also tied to differentiation by the pluripotency factors Oct4, Sox2, and Nanog, which regulate *Xist* and/or *Tsix* expression to prevent XCI in pluripotent cells (reviewed in [133]).

The *Jpx* gene encodes a lncRNA that activates *Xist* expression by evicting the zinc finger protein CTCF from the *Xist* promoter [134]. Since CTCF eviction appeared to depend on Jpx dose, *Jpx* was proposed to promote *Xist* expression in *trans* with the Jpx:CTCF ratio acting as the X:A sensing mechanism [134]. It was thought that the two X chromosomes must pair during the X-counting step of XCI, however, inactivation can occur without pairing. In XX-XY heterokaryons, the X in the XY nucleus was inactivated as often as either of the Xs in the XX nucleus, which demonstrates that *trans*-acting (cytoplasmic) factors play a major role in chromosome counting and Xi choice [132]. XCI proceeded even when known pairing regions were deleted on one X chromosome [132]. Also, the Xic-encoded lncRNAs, *Jpx* and *Ftx*, and the X-pairing region, *Xpr*, were shown to contribute to *Xist* activation in heterokaryons, but mostly in *cis* [132].

Xist expression is also regulated on the level of chromatin conformation. Mammalian genomes are organized in megabase-long topologically-associated domains (TADs) that can affect gene expression by bringing promoters and enhancers together or by keeping them apart [135, 136]. Genes and regulatory elements within TADs are more likely to interact with each other even at long range and less likely to interact with loci outside the TAD [137]. Cohesins, named for their ability to hold sister chromatids together during cell division and DNA repair, and CCCTC-binding factor (CTCF), the vertebrate insulator protein, are found at many boundaries between TADs [138]. The *Xist*/*Tsix* locus is the site of a TAD boundary that may reinforce silencing of *Xist* on Xa and activation of *Xist* on Xi (the *Xist* promoter region is insulated from interactions with the *Tsix* promoter region) [139].

CTCF was suspected to play a role in X-chromosome pairing and choice of Xi [140], although as stated above, pairing is not critical for XCI initiation. Instead, the role of CTCF in Xi choice appears to be related to chromatin structure. TADs are not stable loops, but rather exist in multiple conformations across a population of cells, and some interactions within TADs, called master loci, are more important than others for maintaining TAD structure [141]. Cohesin and CTCF bind at sites predicted to be master loci within the Xic, and fluctuations in chromatin structure and interactions at master loci may contribute to the stochastic nature of XCI [141]. Jpx was reported to evict CTCF from the *Xist* locus to activate *Xist* expression on Xi [134].

Once XCI is initiated by monoallelic *Xist* expression, Xist spreading triggers a cascade of events that lead to heritable silencing of most of the genes in *cis* (reviewed in [142, 143, 144, 145]). RNA Polymerase II and transcription factors are first excluded from the Xist RNA compartment as Xist spreads [146]. A number of histone modifiers deplete marks of active chromatin and deposit repressive marks, leaving the Xi enriched for marks including H3K27me3, H4K20me1, H3K9me2, H3K20me3, and H4K20me1. [145]. Replication timing of Xi is delayed, the histone variant macroH2A is incorporated and ubiquitinated, and CG dinucleotides at promoters are methylated [142, 147, 145]. Allele-specific chromatin interaction data shows that the conformation of the dosage compensated Xi is very different from Xa. When XCI is fully established, rather than megabase-long TADs, almost all TADs are lost. Two mega-domains remain, separated by a boundary at a DXZ4 microsatellite, and TAD-like local structure around Xi-expressed "escapee" genes [148, 149, 150].

X-specificity of mammalian XCI

Xist has separable functions in binding and silencing Xi. Xist RNA is capped, polyadenylated, and spliced and contains 6 regions with tandem repeats named A-F. Of these, repeat A is essential for silencing, but not localization [151]. Repeat C appears to be important for binding to Xi [152], and repeats B and F are required to recruit the chromatin-modifying Polycomb Repressive Complex 2 (PRC2) through an interaction with its cofactor Jarid2 [153]. PRC2 methylates histone 3 lysine 27 (H3K27me3). The Yin-Yang 1 (YY1) transcription factor competes with Rex1 to activate *Xist* expression [154] and also binds Xist RNA through repeat C, where it is thought to act as a nucleation center for Xist RNA on Xi [155]. Heterogeneous nuclear ribonuclearprotein U (hnRNPU, also known as SP120 or SAF-A) is another DNA- and RNA-binding protein required for Xist localization to Xi and XCI [156].

Dosage compensation machinery spreads from the Xic to silence genes across the entire Xi. Sequence alone cannot drive binding to *trans* factors, because both X chromosomes have the same sequences. Also, *Xist* transcripts can bind and spread when *Xist* is translocated to an autosome, which shows that spreading is not dependent on X-specific DNA sequences [143]. Recently, new techniques have allowed purification of the largely insoluble Xist RNP complex (Capture hybridization analysis of RNA targets (CHART), chromatin isolation by RNA purification (ChIRP), and RNA-antisense purification (RAP)). These methods employ biotinylated RNA or DNA probes to pull down the complex for mass spectrometry and/or high throughput sequencing [157]. Rather than binding to discrete sites on the X chromosomes by affinity to a specific sequence motif, Xist transcripts and other XCI factors initially spread from the Xic to several-megabase-long domains that are gene-rich and nearby in three dimensions [158, 159]. In later stages, the complex spreads to gene-poor regions, however the complex doesn't spread across the entire Xi; discrete boundaries appear to exist near genes that escape XCI [158]. Although Xist RNP spreading appears largely sequence-

independent, binding correlates with gene-rich regions containing short interspersed nuclear elements (SINEs) and anti-correlates with long interspersed nuclear elements (LINEs) and lamin-interacting sites [160].

Xist binding to the Xi appears dynamic, with localization that differs throughout the cell cycle [161]. Recent stochastic optical reconstruction microscopy (STORM) micrographs suggest that Xist RNA copy number is much lower than predicted (50-100 rather than 300-2,000 copies) [162]. Xist may not literally coat the Xi, but rather interact with Xi through a "hit and run" mechanism in which silencing is established during transient interactions of Xist, PRC2, and Xi chromatin [162].

1.6 XCI evolution in mammals

XCI is conserved across mammals, however several features have diverged. First, sex chromosomes have diverged, resulting in a different subset of genes subject to XCI. Second, the number of genes that escape XCI differs across species, indicating that XCI can be more or less robust. Third, developmental timing of XCI onset, choice of Xi, and the requirement for XCI in extraembryonic tissues differs across species. Fourth, *Xist* is not universal, but rather other lncRNAs can accomplish XCI.

Sex chromosomes have diverged within therian mammals (figure 1.5) (reviewed in [163]). Prototheria (monotremes) like the platypus have multiple pairs of X and Y chromosomes with homology to the chicken Z and W chromosomes. Genes homologous to the platypus X-linked genes are found on autosomes in metatherians (marsupials) and eutherians (placental mammals). The eutherian sex chromosomes evolved from a different autosome pair after the split from prototherians (165 MYA), but before the split from metatherians (150 MYA). The X-chromosomal region shared between eutheria and metatheria is called the X conserved region (XCR), and the X added region (XAR) was created by a sex chromosome to autosome fusion specific to Eutheria.

Within Eutheria, human and mouse differ in the number of genes that escape dosage compensation, with an estimated 15% and 3% of genes expressed from both female X chromosomes, respectively [164, 165]. RNA-FISH experiments in human, mouse, and elephant show that expression of genes in the XCR is generally monoallelic and genes in the XAR are sometimes expressed from both alleles [166]. Mouse XCI may be more robust because intrachromosomal rearrangements have broken up the XAR and incorporated the region into the XCR. Similar experiments show that marsupial XCI is less complete, with tissue-specific variation and more genes that escape from XCI [167]. XCI is also incomplete in the platypus [167].

Mary Lyon correctly predicted that female mammals are mosaic, with one of the two X chromosomes randomly inactivated early in development, but XCI is not random in all mam-

mals. Later, it was discovered that XCI is not random at earlier stages in mouse. The mouse paternal X chromosome (Xp) is inactivated by the 4-cell stage, at which point the maternal X chromosome (Xm) is protected from inactivation by epigenetic imprinting present in the oocyte [168]. At the blastocyst stage, the paternal X is reactivated in the pluripotent cells of the inner cell mass before the second wave of X-chromosome inactivation randomly inactivates one X in each cell (reviewed in [160]). The silencing of the paternal X chromosome could be the ancestral XCI mechanism, since marsupials retain this Xp silencing throughout their life [169]. Humans and rabbits initiate random XCI later than mice, lack imprinted XCI, and do not appear to require XCI in the extraembryonic tissues [170]. XCI is random in horse and mule placentas [171].

The lncRNAs that trigger XCI are rapidly evolving. *Xist* evolved after the split between Metatheria and Eutheria, and might have led to an improvement in XCI [172]. The marsupial *Xist* homolog is a functional protein-coding gene, *Lnx3*, that is expressed in both sexes [173]. Upon closer inspection, at least four of the lncRNA genes in this region encode proteins in marsupials [173]. Another lncRNA, *Rsx* (RNA-on-the-silent-X) appears to play a central role in marsupial XCI [174]. *Rsx* can trigger silencing in *cis* when inserted on an autosome in mouse embryonic stem cells [174]. *Xist* is shared among eutherian mammals, including the "most primitive" placental mammals, elephants and armadillos, and has evolved rapidly through mobile element insertion and changes in splice junctions, which may be a general feature of lncRNA pseudogenization and evolution [173, 172]. In addition, lncRNAs that regulate *Xist* differ between species. Humans lack *Tsix* and another Xic lncRNA, *Linx* [175], and have a primate-specific lncRNA that associates with Xa named *Xact* [176]. Timing of *Xist* expression and initiation of XCI also differ within eutherian mammals. (reviewed in [143]). In humans and rabbits, initial *XIST* expression is neither sex- nor Xi-specific, which suggests that the choice of Xi occurs downstream of *XIST* in these species [170].

1.7 *C. elegans* dosage compensation

Discovery

The development of *C. elegans* as a model system happened at an opportune time. Although fly dosage compensation was first observed in the 1930s and mouse dosage compensation in the 1960s, it wasn't until the 1980s and 1990s that scientists had the tools to dissect the molecular mechanisms of dosage compensation. Around the same time, cellular molecules responsible for *C. elegans* sex determination and dosage compensation were discovered. In *C. elegans*, as in *Drosophila* (but not in mammals), dosage compensation and sex determination are linked ([177], reviewed in [178]).

Experiments with polyploid animals revealed that *C. elegans* sex (and dosage compen-

sation status) is communicated by the X:A ratio rather than X chromosome number. X:A ratios of one result in hermaphrodite development in diploid (2X:2A) and tetraploid (4X:4A) animals, and X:A ratios of 0.5 result in male development (1X:2A and 2X:4A) [179]. Remarkably, rather than developing intersex phenotypes like the fly, 3X:4A animals (X:A of 0.75) develop as hermaphrodites, and 2X:3A triploids (X:A of 0.67) develop as males [179]. Intersex animals were created, but only between the X:A ratios of 0.67 and 0.75, with X-chromosomal duplications in 2X:3A triploids [179]. While intermediate X:A ratios sometimes create intersex animals, ratios above one had detrimental effects on fertility, morphology, and survival; 4X:2A animals died and 3X:2A animals were dumpy with decreased fertility [180]. This demonstrates that increased X-linked expression is not tolerated above a threshold. When the first dosage compensation mutants (*dpy-21*, *dpy-26*, *dpy-27*, and *dpy-28*) were discovered it was noted that their mutant phenotypes are XX-specific and similar to mutants with increased X-chromosome dose [181, 182, 28, 183, 184]. In the first molecular demonstration of *C. elegans* dosage compensation, mutations in three of these genes were shown to cause 2 to 3-fold upregulation of transcription for three X-linked genes in XX-hermaphrodites [28].

Sex-specific regulation in *C. elegans*

In *C. elegans*, the X:A ratio is composed of discrete X- and autosomal signal elements (XSEs and ASEs) that regulate the X-linked, master switch gene, *xol-1* (XO lethal) [185, 186, 187]. ASEs promote and XSEs inhibit *xol-1* transcription [188, 189, 190]. When X:A is low (as in XO males), *xol-1* is expressed at a high level, which inhibits dosage compensation and hermaphrodite fate. In XX hermaphrodites, the X:A ratio is 1 and *xol-1* expression is low. In a second level of regulation, the XSE FOX-1 binds *xol-1* mRNA to create an inactive splice variant in XX-animals [191, 192].

Downstream of *xol-1* in the sex determination pathway are three genes that link sex determination and dosage compensation, *sdc-1*, *sdc-2*, and *sdc-3*, which encode members of the dosage compensation complex (DCC) [193, 194, 195, 196, 197]. Mutations in these genes lead to masculinized XX animals in addition to the XX-specific lethality and dumpiness characteristic of a dosage compensation defect [177, 194, 197]. Expression of the X-linked switch gene, *sdc-2*, is inhibited by *xol-1* in XO males [187]. In XX animals, SDC-2 triggers the assembly of the DCC on the X chromosome and promotes hermaphrodite fate with DCC subunits, SDC-3 and SDC-1, by repressing transcription at the (autosomal) *her-1* promoter [198]. *her-1* is at the top of a cascade that directs male development [198]. Mutations in *her-1* or the downstream *fem-3* rescue *sdc-2* masculinization, but not dosage compensation phenotypes [194]. Mutations in *sdc-1* are generally weaker than *sdc-2* or *sdc-3* mutations in that they affect X-linked gene expression, but they do not cause significant XX-lethality [196]. Temperature-shift experiments demonstrate that *sdc-1* appears to act in establishing dosage compensation, but not in maintaining it at later developmental stages [196].

C. elegans dosage compensation machinery

The *C. elegans* DCC has at least 10 subunits, including the SDC proteins (SDC-1, SDC-2, and SDC-3), a condensin-like core complex called Condensin I^{DC} (MIX-1, DPY-27, DPY-26, DPY-28, and CAPG-1), and others (DPY-30, and DPY-21). Condensins I and II are five-subunit complexes that contain a MIX-1 (SMC-2) and SMC-4 heterodimer and three CAP proteins (Chromosome-Associated Polypeptides) [199]. These complexes condense and resolve chromosomes in preparation for segregation during mitosis and meiosis, and also play roles in DNA repair and transcriptional regulation (reviewed in [200, 201]). In *C. elegans*, Condensin I^{DC} is identical to Condensin I except SMC-4 is replaced with its paralog, DPY-27 [202, 203, 204, 199]. The duplication of SMC-4 may be the key event that allowed condensin co-option for dosage compensation, since mutations in *dpy-27* would not affect Condensin I function or localization. As described below, the Condensin I^{DC} specifically restructures the X chromosome for dosage compensation [205].

Several DCC components were identified with sex-specific Dpy or lethal phenotypes, but only subtle roles in sex determination. *dpy-21* mutations caused dumpy phenotypes in XX hermaphrodites [206, 182], but affected X-linked expression in both sexes [28]. *dpy-26*, *dpy-27*, and *dpy-28* mutations were observed to cause incompletely penetrant lethality, dumpiness, and increased X-linked gene expression in XX hermaphrodites [207, 193]. Mutations in these *dpy* genes affect sex determination only in sensitized backgrounds, including strains with large X duplications or sex determining gene mutations [182, 184]. *dpy-30* mutations affect both sexes because the gene is pleiotropic. DPY-30 functions in both the DCC and a complex homologous to MLL/COMPASS, which trimethylates histone 3 lysine 4 (H3K4me3) and is important for transcriptional activation [208]. *dpy-30* mutations are temperature-sensitive, with almost completely penetrant XX-lethality, rare Dpy escapers, and slow growth and mating defects in males [209].

Some of these components are necessary for X-specificity, with a defined order of dependency for binding the X chromosomes, however the complex lacks obvious DNA-binding domains. SDC-2 is a 344 kD, nematode-specific protein that lacks recognizable protein domains, with the exception of a coiled-coil domain. SDC-2 is able to bind to the X chromosome in the absence of other components [198]. SDC-3, a 250 kD protein, and DPY-30, a 13 kD protein, require only SDC-2 for DCC targeting to X. SDC-3 contains two zinc fingers that are required for dosage compensation, but not sex determination, and a region that shares similarity with the ATP-binding domain of myosin, where mutations disrupt sex determination, but not dosage compensation [197]. The other DCC components follow. At the *her-1* locus, the order is reversed; SDC-3 is able to bind in the absence of SDC-2 [210]. SDC-1 is a 139 kD protein with 7 zinc fingers, hypothesized to be a transcription factor [211], however *sdc-1* is unlikely to be important for DCC targeting to X, since other components do not require SDC-1 for binding.

X-specificity of *C. elegans* dosage compensation

The DCC binds to discrete sites on both hermaphrodite X chromosomes to turn down gene expression by half [192]. The X-specificity of DCC-binding is driven by interactions with recruitment elements on X (*rex* sites), the X-linked sequences that are able to bind the DCC when present on an autosome or an extrachromosomal array [212, 213, 214]. The DCC binds to *rex* sites, then spreads to a second class of sites called *dox* sites (dependent on X) [213]. *dox* sites bind at their endogenous loci, but cannot bind the DCC when detached from the X chromosome [215, 216]. Binding at *C. elegans dox* sites appears to be dependent on transcription levels – the DCC spreads from *rex* sites to *dox* sites that are often found in promoters of highly expressed genes [216]. DCC spreading to *dox* sites appears independent of sequence because a similar pattern of spreading to promoters was observed on an autosome in an X:A fusion [217].

The DNA sequences that drive X-specific binding of the DCC were identified by searching for motifs among *rex* site sequences. Genome-wide chromatin immunoprecipitation (ChIP) experiments identified hundreds of binding sites on the X chromosome, many of which were tested for recruitment ability *in vivo*, bringing the current number of *rex* sites to 47 [213, 214, 218, 216, 192, 208]. Searches among highly-occupied DCC-binding site sequences yielded a 12 base pair X-enriched motif named MEX (motif enriched on X) [218, 216]. A second X-enriched DCC-binding motif, MEX-II, was identified by searching strong *rex* sites that lack the MEX motif (W. Kruesi, unpublished data).

The *C. elegans* condensin-driven DCC restructures the hermaphrodite X chromosomes by bringing strong *rex* sites together [205]. The dosage compensated *C. elegans* X chromosome is organized in approximately 1 Mb TADs (topologically-associated domains), in which chromatin looping insulates each frequently interacting TAD sequence from the next [205]. The autosomes have fewer TAD boundaries and a less defined structure [205]. Many strong TAD boundaries on the X chromosome coincide with strong *rex* sites, and many of these TAD boundaries are weakened or lost in an *sdc-2* mutant where there is no DCC-binding [205]. In contrast, as described above, the mammalian dosage compensated (inactive) X chromosome has less structure than the active X. The mechanistic connection between X-chromosome structure and its effect on gene expression is currently unclear, however evolutionary comparisons may bring us closer to understanding this link in the future.

In chapter 2, I compare dosage compensation mechanisms in four nematode species, *C. elegans*, *C. briggsae*, *C. nigoni*, and *C. tropicalis* (figure 1.3). Although we found that key components of the DCC and the genetic hierarchy driving dosage compensation are conserved across these species, the DCC-binding sites and the sequence-specificity of DCC-binding have diverged. Future experiments will determine whether the divergent *rex* sites in these species have similar or different effects on chromosome structure and gene expression.

1.8 Do ZW species lack chromosome-wide dosage compensation?

Complete, chromosome-wide dosage compensation is not a universal requirement for all species with heteromorphic sex chromosomes. In fact, many ZW species lack complete, chromosome-wide dosage compensation. Dosage compensation was reported to be incomplete in birds [219], the silkworm *Bombyx mori* [220], the parasite *Schistosoma mansoni* [221], the Indian meal moth, *Plodia interpunctella* [222], and the pygmy rattlesnake *S. miliarius barbouri* [223]. Some hypothesized that ZW females are more tolerant of sex-biased expression, thus do not require complete dosage compensation [224]. However, incomplete dosage compensation is not universal for ZW species. It was later discovered that reported sex-bias in *B. mori* Z-linked expression was an artifact caused by errors in microarray data normalization [225]. Also, some ZW moths and butterflies were shown to have complete dosage compensation mechanisms [226, 227]. Hypotheses about ZW dosage compensation can be tested in frog species, which have diverse sex determination mechanisms and sex chromosomes of various ages [228]. Species that lack these complete mechanisms may improve our understanding of the forces that drive the evolution of these complex systems.

There are several possible reasons why some species may require chromosome-wide dosage compensation mechanisms while others do not. First, the evolutionary trajectory that leads to differentiation between X and Y or Z and W chromosomes may differ in some lineages, in ways that affect the requirement for dosage compensation. Second, there may be factors specific to species with female-heterogametic (ZW) systems that favor incomplete dosage compensation or mitigate the consequences of sex chromosome imbalance. Third, some chromosomes may be "better" at becoming sex chromosomes that do not require complete compensation. The particular genes on the sex chromosomes may be more or less dose-sensitive or may be required at different levels in males and females.

The evolutionary trajectory that leads to dosage compensation evolution may differ across species, which could affect the type of dosage compensation mechanism that evolves. A pair of sex chromosomes evolves from an autosome when a sex-determining locus is acquired, but dosage compensation is not required until the X and Y or Z and W differentiate from each other. This happens when recombination between the two is restricted by inversions, for example [229]. Regions within X or Z inversions cannot recombine with the Y or W chromosome without fitness consequences, so Y or W cannot be repaired, and genes are easily lost [230]. Chromosome segments outside of the inversions, in the pseudoautosomal regions, are present in two copies in both sexes, thus do not require dosage compensation [231]. If the decay of the Y or W occurred gradually, then sex-specific regulation for dose-sensitive genes could happen as needed, on a gene-by-gene basis rather than chromosome-wide. Alternatively, rearrangements could move sex-biased or dose-sensitive genes to autosomes [232].

A population genetic model was proposed to explain the difference in extent of dosage compensation between many ZW and XY species [233]. Because ZZ males have two Z chromosomes and ZW females have one, the Z chromosome spends two-thirds of its time in males, which means that Z chromosome is under selection in males two-thirds of the time. In contrast, in XY species, the X chromosome spends only one-third of its time in males. Because of stronger sexual selection, male genes are usually under greater selection, including selection for tightly-regulated expression levels [233]. The combination of stronger selection in males and double the time spent in males means that Z-linked expression is more quickly optimized to male levels, with weaker selection and slower evolution toward female-optimal expression [233].

Sex chromosomes that originate from different autosomes may have different dosage compensation requirements. In support of this hypothesis, we know that some aneuploidies are lethal, while others are not. Individual sex-linked genes may vary in dose-sensitivity, or males and females may have different optimal expression levels. In studies of yeast aneuploidies, it was found that some defects associated with aneuploidy depend on the particular chromosome that is duplicated (karyotype-specific defects) and others are more general [234]. It was also observed that a fraction of yeast genes are subject to buffering, meaning that feedback loops control mRNA expression levels, even when copy number is increased [235].

1.9 Summary

Although the independent fly, mammal, and worm dosage compensation mechanisms differ in many ways, each evolved by co-opting existing cellular machinery. Some components of the dosage compensation machinery are highly conserved and used in other cellular processes today, like condensin in worms, polycomb in mammals, and MOF in flies. Others are more rapidly evolving and have no other known functions, like the worm *sdc* genes, the mammalian *Xist*, and fly *msl-1* and *msl-2*. Nematode dosage compensation is unique in its co-option of condensin, however chromosome structure also appears to play a role in mammalian *Xist* regulation and Xist RNA spreading. In contrast, long noncoding RNAs play a major role in flies and mammals, but none have been found to be involved in dosage compensation in worms. Also, histone modifications are thought to drive silencing in mammals and upregulation in flies, but only one histone modification has been associated with nematode dosage compensation thus far [236].

Sex-specific and X-specific regulation also differs between these species. Flies and worms have sex-specific dosage compensation regulation that is linked directly to sex determination. *Sxl* is the master switch driving sex determination and sex-specific dosage compensation in *Drosophila* species, but not in other flies. Similarly, we see conservation of the switch genes *xol-1* and *sdc-2* between *C. elegans* and *C. briggsae*, but do not find a *sdc-2* homolog in the distant relative *Pristionchus pacificus* (described in chapter 2). In mouse, *Xist* plays a key

role in sex-specific regulation and in coating the inactive X chromosome for dosage compensation, however sex- and Xi-specificity is not determined by *Xist* in humans or rabbits. A different long noncoding RNA appears to play an analogous role to *Xist* in marsupials. These examples show that sex determination and dosage compensation regulation are subject to change.

1.10 Figures

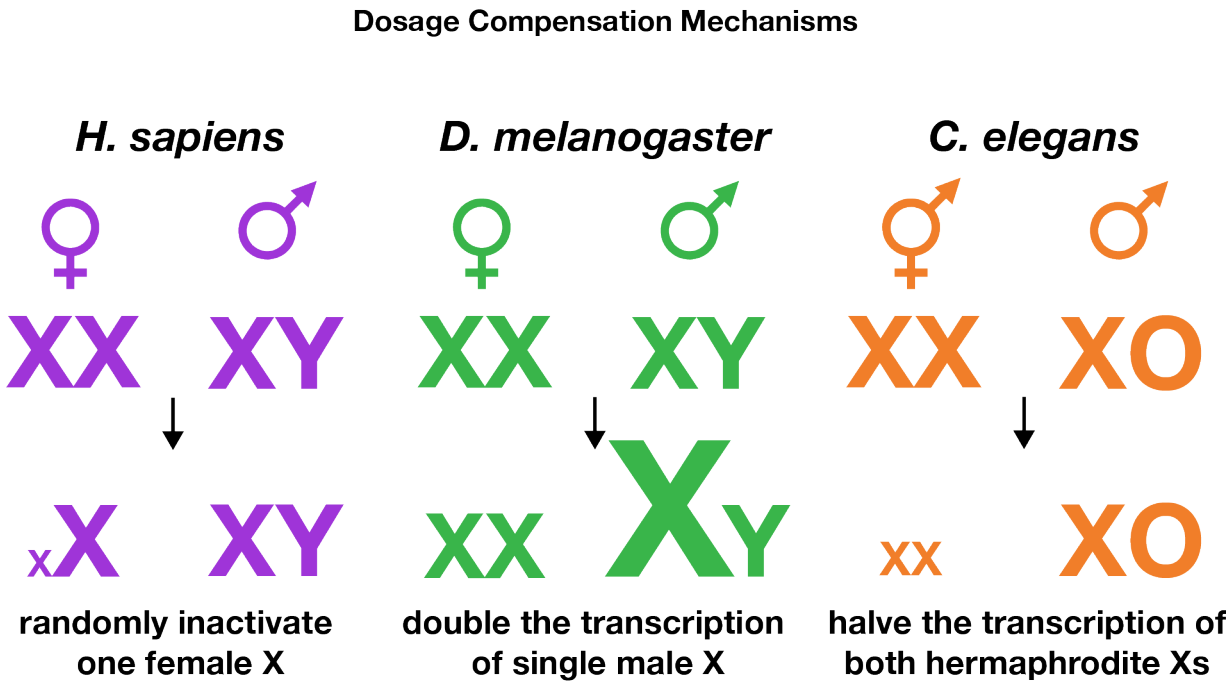


Figure 1.1: Dosage compensation in mammals, flies, and worms.

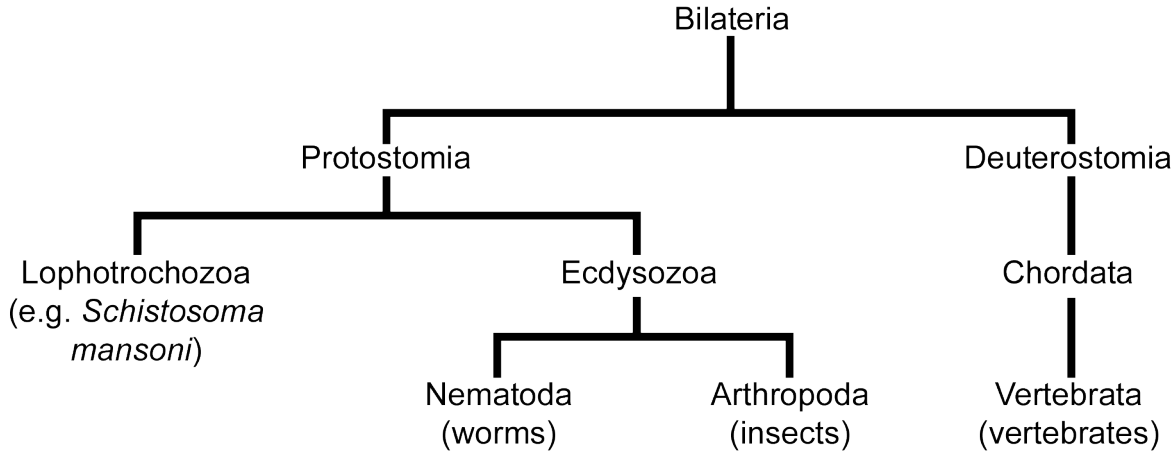


Figure 1.2: Bilaterian phylogeny, adapted from [237]

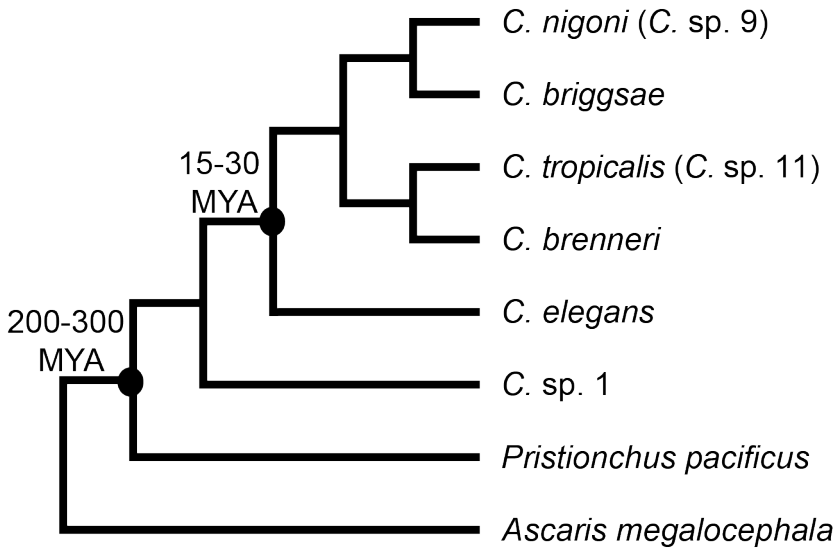


Figure 1.3: Nematode phylogeny, adapted from [238, 24, 22]

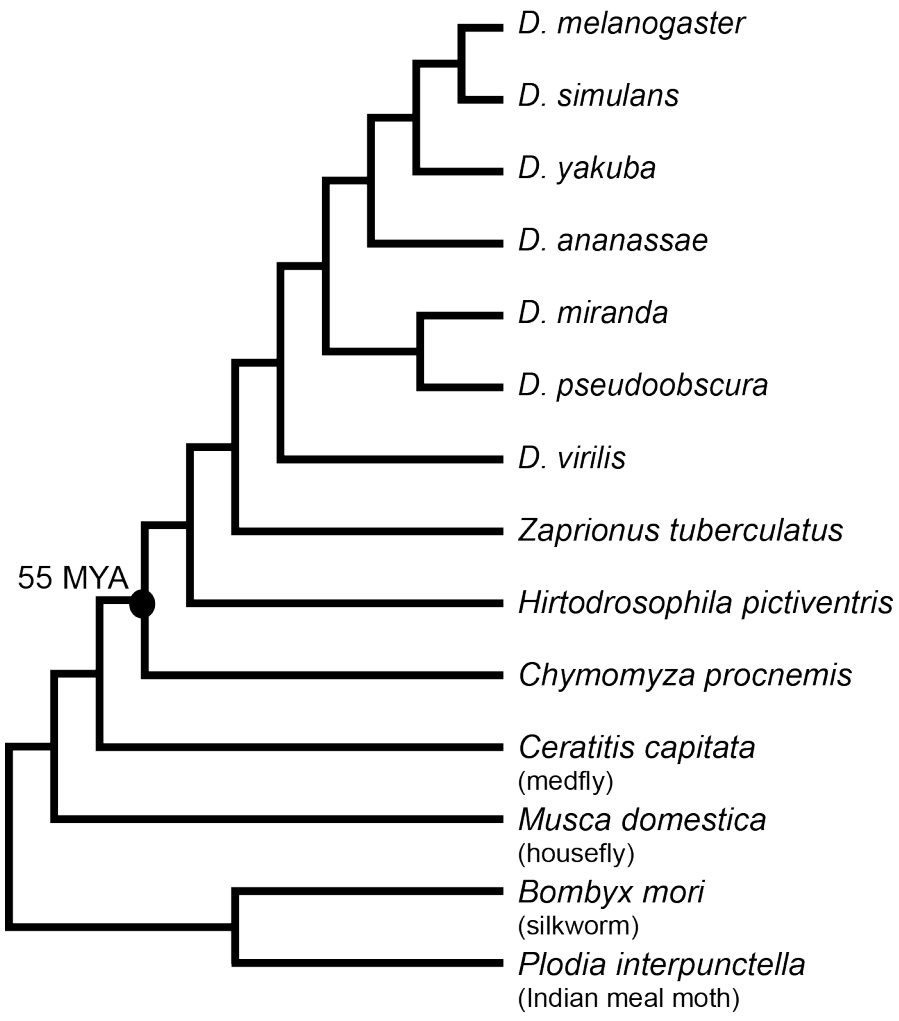


Figure 1.4: Insect phylogeny, adapted from [103, 239, 240, 79]

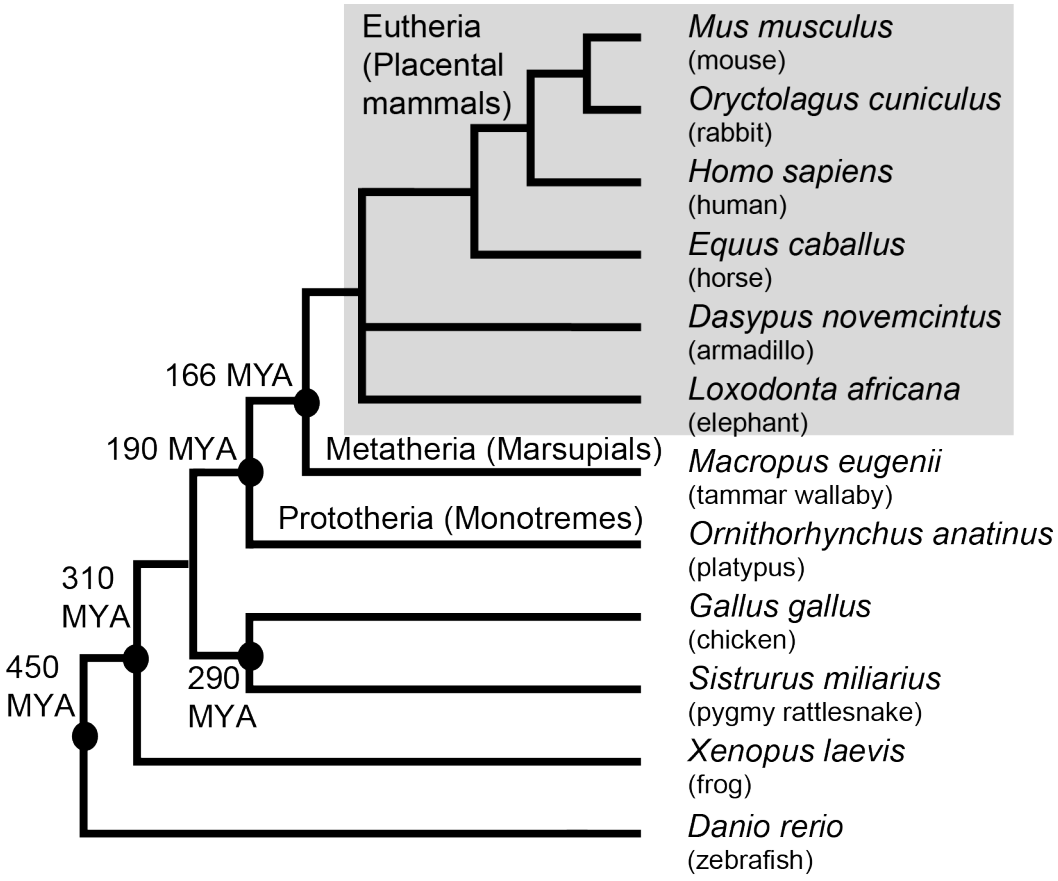


Figure 1.5: Vertebrate phylogeny, adapted from [172, 241]

References

[1] B Charlesworth. "The evolution of chromosomal sex determination and dosage compensation." In: *Current biology : CB* 6.2 (Feb. 1996), pp. 149–162.

[2] Florian Maderspacher. "Theodor Boveri and the natural experiment." In: *Current biology : CB* 18.7 (Apr. 2008), R279–86.

[3] Helga Satzinger. *Theodor and Marcella Boveri: chromosomes and cytoplasm in heredity and development.* Vol. 9. Wellcome Trust Centre for the History of Medicine, University College London, 183 Euston Road, London, NW1 2BE, UK. h.satzinger@ucl.ac.uk, Mar. 2008.

[4] Howard Ferris and W F Hieb. "Ellsworth C. Dougherty: A Pioneer in the Selection of Caenorhabditis elegans as a Model Organism." In: *Genetics* 200.4 (Aug. 2015), pp. 991–1002.

[5] V Nigon and E C Dougherty. "Reproductive patterns and attempts at reciprocal crossing of Rhabditis elegans Maupas, 1900, and Rhabditis briggsae Dougherty and Nigon, 1949 (Nematoda: Rhabditidae)." In: *The Journal of experimental zoology* 112.3 (Dec. 1949), pp. 485–503.

[6] E C Dougherty and H G Calhoun. "Possible significance of free-living nematodes in genetic research." In: *Nature* 161.4079 (Jan. 1948), p. 29.

[7] V Nigon and E C Dougherty. "A dwarf mutation in a nematode; a morphological mutant of Rhabditis briggsae, a free-living soil nematode." In: *The Journal of heredity* 41.4 (Apr. 1950), pp. 103–109.

[8] Marie-Anne Félix, Christian Braendle, and Asher D Cutter. "A streamlined system for species diagnosis in Caenorhabditis (Nematoda: Rhabditidae) with name designations for 15 distinct biological species." In: *PloS one* 9.4 (2014), e94723.

[9] S Brenner. "The genetics of Caenorhabditis elegans." In: *Genetics* 77.1 (May 1974), pp. 71–94.

[10] Patrick T McGrath et al. "Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors." In: *Neuron* 61.5 (Mar. 2009), pp. 692–699.

[11] William M Winston et al. "Caenorhabditis elegans SID-2 is required for environmental RNA interference." In: *Proceedings of the National Academy of Sciences of the United States of America* 104.25 (June 2007), pp. 10565–10570.

[12] Owen Thompson et al. "The million mutation project: a new approach to genetics in Caenorhabditis elegans." In: *Genome research* 23.10 (Oct. 2013), pp. 1749–1762.

[13] C. elegans Sequencing Consortium. "Genome sequence of the nematode C. elegans: a platform for investigating biology." In: *Science (New York, N.Y.)* 282.5396 (Dec. 1998), pp. 2012–2018.

[14] Lincoln D Stein et al. "The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics." In: *PLoS biology* 1.2 (Nov. 2003), E45.

[15] Daniel C Koboldt et al. "A toolkit for rapid gene mapping in the nematode Caenorhabditis briggsae." In: *BMC genomics* 11 (2010), p. 236.

[16] Andrew J Wood et al. "Targeted genome editing across species using ZFNs and TALENs." In: *Science (New York, N.Y.)* 333.6040 (July 2011), pp. 307–307.

[17] Te-Wen Lo et al. "Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions." In: *Genetics* 195.2 (Oct. 2013), pp. 331–348.

[18] Behnom Farboud and Barbara J Meyer. "Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design." In: *Genetics* 199.4 (Apr. 2015), pp. 959–971.

[19] Antoine Barrière and Marie-Anne Félix. "Isolation of C. elegans and related nematodes." In: *WormBook : the online review of C. elegans biology* (2014), pp. 1–19.

[20] Asher D Cutter. "Caenorhabditis evolution in the wild." In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 37.9 (Sept. 2015), pp. 983–995.

[21] Karin Kiontke et al. "Caenorhabditis phylogeny predicts convergence of hermaphroditism and extensive intron loss." In: *Proceedings of the National Academy of Sciences of the United States of America* 101.24 (June 2004), pp. 9003–9008.

[22] Karin C Kiontke et al. "A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits." In: *BMC evolutionary biology* 11.1 (2011), p. 339.

[23] Alivia Dey et al. "Molecular hyperdiversity defines populations of the nematode Caenorhabditis brenneri." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.27 (July 2013), pp. 11056–11060.

[24] Asher D Cutter. "Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate." In: *Molecular biology and evolution* 25.4 (Apr. 2008), pp. 778–786.

[25] Doris Bachtrog et al. "Sex determination: why so many ways of doing it?" In: *PLoS biology* 12.7 (July 2014), e1001899.

[26] Jonathan Hodgkin. "Exploring the envelope. Systematic alteration in the sex-determination system of the nematode caenorhabditis elegans." In: *Genetics* 162.2 (Oct. 2002), pp. 767–780.

[27] Doris Bachtrog et al. "Are all sex chromosomes created equal?" In: *Trends in genetics : TIG* 27.9 (Sept. 2011), pp. 350–357.

[28] B J Meyer and L P Casson. "Caenorhabditis elegans compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels." In: *Cell* 47.6 (Dec. 1986), pp. 871–881.

[29] Megan E Bonney, Hisao Moriya, and Angelika Amon. "Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes." In: *Genes & development* 29.9 (May 2015), pp. 898–903.

[30] Susumu Ohno. *Sex chromosomes and sex-linked genes.* Germany: Springer Berlin Heidelberg, 1967.

[31] Judith E Mank, David J Hosken, and Nina Wedell. "Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict." In: *Evolution; international journal of organic evolution* 65.8 (Aug. 2011), pp. 2133–2144.

[32] John C Lucchesi and Mitzi I Kuroda. "Dosage compensation in Drosophila." In: *Cold Spring Harbor Perspectives in Biology* 7.5 (May 2015), a019398.

[33] H J Muller. "Further studies on the nature and causes of gene mutations." In: *International Congress of Genetics* 6.1 (1932), pp. 213–255.

[34] A S Mukherjee and W Beermann. "Synthesis of ribonucleic acid by the X-chromosomes of Drosophila melanogaster and the problem of dosage compensation." In: *Nature* 207.998 (Aug. 1965), pp. 785–786.

[35] J M Belote and J C Lucchesi. "Male-specific lethal mutations of Drosophila melanogaster." In: *Genetics* 96.1 (Sept. 1980), pp. 165–186.

[36] J M Belote and J C Lucchesi. "Control of X chromosome transcription by the maleless gene in Drosophila." In: *Nature* 285.5766 (June 1980), pp. 573–575.

[37] T W Cline. "Two closely linked mutations in Drosophila melanogaster that are lethal to opposite sexes and interact with daughterless." In: *Genetics* 90.4 (Dec. 1978), pp. 683–698.

[38] T W Cline. "Maternal and zygotic sex-specific gene interactions in Drosophila melanogaster." In: *Genetics* 96.4 (Dec. 1980), pp. 903–926.

[39] E M Maine et al. "The Sex-lethal gene of Drosophila: DNA alterations associated with sex-specific lethal mutations." In: *Cell* 43.2 Pt 1 (Dec. 1985), pp. 521–529.

[40] L R Bell et al. "Sex-lethal, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins." In: *Cell* 55.6 (Dec. 1988), pp. 1037–1046.

[41] J M Belote et al. "Molecular genetics of transformer, a genetic switch controlling sexual differentiation in Drosophila." In: *Developmental genetics* 10.3 (1989), pp. 143–154.

[42] B A Sosnowski, J M Belote, and M McKeown. "Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage." In: *Cell* 58.3 (Aug. 1989), pp. 449–459.

[43] Daniel S Evans and Thomas W Cline. "Drosophila switch gene Sex-lethal can bypass its switch-gene target transformer to regulate aspects of female behavior." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.47 (Nov. 2013), E4474–81.

[44] R L Kelley et al. "Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila." In: *Cell* 81.6 (June 1995), pp. 867–877.

[45] K Copps et al. "Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly." In: *The EMBO journal* 17.18 (Sept. 1998), pp. 5409–5417.

[46] Antoine Graindorge, Clément Carré, and Fátima Gebauer. "Sex-lethal promotes nuclear retention of msl2 mRNA via interactions with the STAR protein HOW." In: *Genes & development* 27.12 (June 2013), pp. 1421–1433.

[47] L R Bell et al. "Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in Drosophila." In: *Cell* 65.2 (Apr. 1991), pp. 229–239.

[48] H K Salz et al. "The Drosophila female-specific sex-determination gene, Sex-lethal, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation." In: *Genes & development* 3.5 (May 1989), pp. 708–719.

[49] Helen K Salz and James W Erickson. "Sex determination in Drosophila: The view from the top." In: *Fly* 4.1 (Jan. 2010), pp. 60–70.

[50] C Cronmiller and T W Cline. "The relationship of relative gene dose to the complex phenotype of the daughterless locus in Drosophila." In: *Developmental genetics* 7.4 (1986), pp. 205–221.

[51] T W Cline. "A female-specific lethal lesion in an X-linked positive regulator of the Drosophila sex determination gene, Sex-lethal." In: *Genetics* 113.3 (July 1986), pp. 641–663.

[52] T W Cline. "Evidence that sisterless-a and sisterless-b are two of several discrete "numerator elements" of the X/A sex determination signal in Drosophila that switch Sxl between two alternative stable expression states." In: *Genetics* 119.4 (Aug. 1988), pp. 829–862.

[53] D Bopp et al. "Developmental distribution of female-specific Sex-lethal proteins in Drosophila melanogaster." In: *Genes & development* 5.3 (Mar. 1991), pp. 403–415.

[54] L Sefton et al. "An extracellular activator of the Drosophila JAK/STAT pathway is a sex-determination signal element." In: *Nature* 405.6789 (June 2000), pp. 970–973.

[55] M Bernstein and T W Cline. "Differential effects of Sex-lethal mutations on dosage compensation early in Drosophila development." In: *Genetics* 136.3 (Mar. 1994), pp. 1051–1061.

[56] M Bernstein et al. "Transposon insertions causing constitutive Sex-lethal activity in Drosophila melanogaster affect Sxl sex-specific transcript splicing." In: *Genetics* 139.2 (Feb. 1995), pp. 631–648.

[57] Scott G Siera and Thomas W Cline. "Sexual back talk with evolutionary implications: stimulation of the Drosophila sex-determination gene sex-lethal by its target transformer." In: *Genetics* 180.4 (Dec. 2008), pp. 1963–1981.

[58] T W Cline. "The Drosophila sex determination signal: how do flies count to two?" In: *Trends in genetics : TIG* 9.11 (Nov. 1993), pp. 385–390.

[59] T W Cline. "The interaction between daughterless and sex-lethal in triploids: a lethal sex-transforming maternal effect linking sex determination and dosage compensation in Drosophila melanogaster." In: *Developmental biology* 95.2 (Feb. 1983), pp. 260–274.

[60] D A Barbash and T W Cline. "Genetic and molecular analysis of the autosomal component of the primary sex determination signal of Drosophila melanogaster." In: *Genetics* 141.4 (Dec. 1995), pp. 1451–1471.

[61] John R ten Bosch, Joseph A Benavides, and Thomas W Cline. "The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription." In: *Development (Cambridge, England)* 133.10 (May 2006), pp. 1967–1977.

[62] James W Erickson and Jerome J Quintero. "Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila." In: *PLoS biology* 5.12 (Dec. 2007), e332.

[63] I Arkhipova, J Li, and M Meselson. "On the mode of gene-dosage compensation in Drosophila." In: *Genetics* 145.3 (Mar. 1997), pp. 729–736.

[64] R L Kelley et al. "Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin." In: *Cell* 98.4 (Aug. 1999), pp. 513–522.

[65] Y Kageyama et al. "Association and spreading of the Drosophila dosage compensation complex from a discrete roX1 chromatin entry site." In: *The EMBO journal* 20.9 (May 2001), pp. 2236–2245.

[66] Gregor D Gilfillan et al. "Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex." In: *Genes & development* 20.7 (Apr. 2006), pp. 858–870.

[67] W Gu, P Szauter, and J C Lucchesi. "Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of Drosophila melanogaster." In: *Developmental genetics* 22.1 (1998), pp. 56–64.

[68] E R Smith et al. "The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation." In: *Molecular and cellular biology* 20.1 (Jan. 2000), pp. 312–318.

[69] Erica Larschan et al. "Identification of chromatin-associated regulators of MSL complex targeting in Drosophila dosage compensation." In: *PLOS Genetics* 8.7 (2012), e1002830.

[70] S Rea, G Xouri, and A Akhtar. "Males absent on the first (MOF): from flies to humans." In: *Oncogene* 26.37 (Aug. 2007), pp. 5385–5394.

[71] Steven P Vensko and Eric A Stone. "Recent progress and open questions in Drosophila dosage compensation." In: *Fly* 9.1 (Jan. 2015), pp. 29–35.

[72] Erica Larschan et al. "X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila." In: *Nature* 471.7336 (Mar. 2011), pp. 115–118.

[73] Thomas Conrad et al. "Drosophila dosage compensation involves enhanced Pol II recruitment to male X-linked promoters." In: *Science (New York, N.Y.)* 337.6095 (Aug. 2012), pp. 742–746.

[74] L M Lyman et al. "Drosophila male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association." In: *Genetics* 147.4 (Dec. 1997), pp. 1743–1753.

[75] Lipeng Wu et al. "The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation." In: *Molecular cell* 43.1 (July 2011), pp. 132–144.

[76] Raffaella Villa et al. "MSL2 combines sensor and effector functions in homeostatic control of the Drosophila dosage compensation machinery." In: *Molecular cell* 48.4 (Nov. 2012), pp. 647–654.

[77] Jan Kadlec et al. "Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1." In: *Nature structural & molecular biology* 18.2 (Feb. 2011), pp. 142–149.

[78] A Franke and B S Baker. "The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila." In: *Molecular cell* 4.1 (July 1999), pp. 117–122.

[79] Jeffrey J Quinn et al. "Rapid evolutionary turnover underlies conserved lncRNA-genome interactions." In: *Genes & development* 30.2 (Jan. 2016), pp. 191–207.

[80] C G Lee et al. "The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation." In: *The EMBO journal* 16.10 (May 1997), pp. 2671–2681.

[81] R Sanjuán and I Marín. "Tracing the origin of the compensasome: evolutionary history of DEAH helicase and MYST acetyltransferase gene families." In: *Molecular Biology and Evolution* 18.3 (Mar. 2001), pp. 330–343.

[82] Sylvain Maenner et al. "ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins." In: *Molecular cell* 51.2 (July 2013), pp. 174–184.

[83] Ibrahim Avsar Ilik et al. "Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila." In: *Molecular cell* 51.2 (July 2013), pp. 156–173.

[84] Simona Cugusi et al. "The Drosophila Helicase MLE Targets Hairpin Structures in Genomic Transcripts." In: *PLoS genetics* 12.1 (Jan. 2016), e1005761.

[85] Simona Cugusi et al. "The Drosophila Helicase Maleless (MLE) is Implicated in Functions Distinct From its Role in Dosage Compensation." In: *Molecular & cellular proteomics : MCP* 14.6 (June 2015), pp. 1478–1488.

[86] Erica Larschan et al. "MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism." In: *Molecular cell* 28.1 (Oct. 2007), pp. 121–133.

[87] Artyom A Alekseyenko et al. "Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context." In: *PLOS Genetics* 8.4 (2012), e1002646.

[88] Artyom A Alekseyenko et al. "A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome." In: *Cell* 134.4 (Aug. 2008), pp. 599–609.

[89] Tobias Straub et al. "The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex." In: *PLOS Genetics* 4.12 (Dec. 2008), e1000302.

[90] M J Palmer et al. "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes a novel protein that associates with the X chromosome in males." In: *Genetics* 134.2 (June 1993), pp. 545–557.

[91] W Gu et al. "Targeting the chromatin-remodeling MSL complex of Drosophila to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities." In: *The EMBO journal* 19.19 (Oct. 2000), pp. 5202–5211.

[92] Yongkyu Park et al. "Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins." In: *Science (New York, N.Y.)* 298.5598 (Nov. 2002), pp. 1620–1623.

[93] Hyangyee Oh, Yongkyu Park, and Mitzi I Kuroda. "Local spreading of MSL complexes from roX genes on the Drosophila X chromosome." In: *Genes & development* 17.11 (June 2003), pp. 1334–1339.

[94] Marcela M L Soruco et al. "The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation." In: *Genes & development* 27.14 (July 2013), pp. 1551–1556.

[95] Guray Kuzu et al. "Expansion of GA Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to Promote Drosophila Dosage Compensation." In: *PLoS genetics* 12.7 (July 2016), e1006120.

[96] Marnie E Gelbart et al. "Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation." In: *Nature structural & molecular biology* 16.8 (Aug. 2009), pp. 825–832.

[97] Philge Philip and Per Stenberg. "Male X-linked genes in Drosophila melanogaster are compensated independently of the Male-Specific Lethal complex." In: *Epigenetics & chromatin* 6.1 (2013), p. 35.

[98] Susan E Lott et al. "Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq." In: *PLoS biology* 9.2 (2011), e1000590.

[99] Zhenguo Zhang, Jan Klein, and Masatoshi Nei. "Evolution of the sex-lethal gene in insects and origin of the sex-determination system in Drosophila." In: *Journal of molecular evolution* 78.1 (Jan. 2014), pp. 50–65.

[100] M Meise et al. "Sex-lethal, the master sex-determining gene in Drosophila, is not sex-specifically regulated in Musca domestica." In: *Development (Cambridge, England)* 125.8 (Apr. 1998), pp. 1487–1494.

[101] G Saccone et al. "The Ceratitis capitata homologue of the Drosophila sex-determining gene sex-lethal is structurally conserved, but not sex-specifically regulated." In: *Development (Cambridge, England)* 125.8 (Apr. 1998), pp. 1495–1500.

[102] Thomas W Cline et al. "Evolution of the Drosophila feminizing switch gene Sex-lethal." In: *Genetics* 186.4 (Dec. 2010), pp. 1321–1336.

[103] I Marín et al. "The dosage compensation system of Drosophila is co-opted by newly evolved X chromosomes." In: *Nature* 383.6596 (Sept. 1996), pp. 160–163.

[104] J R Bone and M I Kuroda. "Dosage compensation regulatory proteins and the evolution of sex chromosomes in Drosophila." In: *Genetics* 144.2 (Oct. 1996), pp. 705–713.

[105] Yongkyu Park et al. "Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex." In: *Molecular cell* 11.4 (Apr. 2003), pp. 977–986.

[106] Seung-Won Park et al. "An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the Drosophila X chromosome." In: *Genetics* 177.3 (Nov. 2007), pp. 1429–1437.

[107] Artyom A Alekseyenko et al. "Conservation and de novo acquisition of dosage compensation on newly evolved sex chromosomes in Drosophila." In: *Genes & development* 27.8 (Apr. 2013), pp. 853–858.

[108] Susan E Lott et al. "Sex-Specific Embryonic Gene Expression in Species with Newly Evolved Sex Chromosomes". In: *PLOS Genetics* 10.2 (Feb. 2014), e1004159–14.

[109] Christopher E Ellison and Doris Bachtrog. "Dosage compensation via transposable element mediated rewiring of a regulatory network." In: *Science (New York, N.Y.)* 342.6160 (Nov. 2013), pp. 846–850.

[110] Christopher E Ellison and Doris Bachtrog. "Non-allelic gene conversion enables rapid evolutionary change at multiple regulatory sites encoded by transposable elements." In: *eLife* 4 (2015), e05899.

[111] T Hsieh and D Brutlag. "Sequence and sequence variation within the 1.688 g/cm3 satellite DNA of Drosophila melanogaster." In: *Journal of molecular biology* 135.2 (Dec. 1979), pp. 465–481.

[112] P V Benos et al. "From sequence to chromosome: the tip of the X chromosome of D. melanogaster." In: *Science (New York, N.Y.)* 287.5461 (Mar. 2000), pp. 2220–2222.

[113] B S Baker, M Gorman, and I Marín. "Dosage compensation in Drosophila." In: *Annual review of genetics* 28.1 (1994), pp. 491–521.

[114] Miguel Gallach. "Recurrent turnover of chromosome-specific satellites in Drosophila." In: *Genome Biology and Evolution* 6.6 (June 2014), pp. 1279–1286.

[115] Miguel Gallach. "1.688 g/cm(3) satellite-related repeats: a missing link to dosage compensation and speciation." In: *Molecular ecology* 24.17 (Sept. 2015), pp. 4340–4347.

[116] M F LYON. "Gene action in the X-chromosome of the mouse (Mus musculus L.)." In: *Nature* 190 (Apr. 1961), pp. 372–373.

[117] M L BARR and E G BERTRAM. "A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis." In: *Nature* 163.4148 (Apr. 1949), p. 676.

[118] S OHNO and T S HAUSCHKA. "Allocycly of the X-chromosome in tumors and normal tissues." In: *Cancer research* 20 (May 1960), pp. 541–545.

[119] M F LYON. "Sex chromatin and gene action in the mammalian X-chromosome." In: *American journal of human genetics* 14 (June 1962), pp. 135–148.

[120] Barbara R Migeon. "Why females are mosaics, X-chromosome inactivation, and sex differences in disease." In: *Gender medicine* 4.2 (June 2007), pp. 97–105.

[121] S Rastan. "Non-random X-chromosome inactivation in mouse X-autosome translocation embryos–location of the inactivation centre." In: *Journal of embryology and experimental morphology* 78 (Dec. 1983), pp. 1–22.

[122] S Rastan and E J Robertson. "X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation." In: *Journal of embryology and experimental morphology* 90 (Dec. 1985), pp. 379–388.

[123] E Heard, P Avner, and R Rothstein. "Creation of a deletion series of mouse YACs covering a 500 kb region around Xist." In: *Nucleic Acids Research* 22.10 (May 1994), pp. 1830–1837.

[124] J T Lee and N Lu. "Targeted mutagenesis of Tsix leads to nonrandom X inactivation." In: *Cell* 99.1 (Oct. 1999), pp. 47–57.

[125] G Borsani et al. "Characterization of a murine gene expressed from the inactive X chromosome." In: *Nature* 351.6324 (May 1991), pp. 325–329.

[126] N Brockdorff et al. "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome." In: *Nature* 351.6324 (May 1991), pp. 329–331.

[127] C J Brown et al. "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." In: *Nature* 349.6304 (Jan. 1991), pp. 38–44.

[128] N Stavropoulos, N Lu, and J T Lee. "A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice." In: *Proceedings of the National Academy of Sciences of the United States of America* 98.18 (Aug. 2001), pp. 10232–10237.

[129] Rafael Galupa and Edith Heard. "ScienceDirectX-chromosome inactivation: new insights into cis and trans regulation". In: *Current opinion in genetics & development* 31 (Apr. 2015), pp. 57–66.

[130] Iris Jonkers et al. "RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation." In: *Cell* 139.5 (Nov. 2009), pp. 999–1011.

[131] Cristina Gontan et al. "RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation." In: *Nature* 485.7398 (May 2012), pp. 386–390.

[132] Tahsin Stefan Barakat et al. "The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing." In: *Molecular cell* 53.6 (Mar. 2014), pp. 965–978.

[133] Alissa Minkovsky, Sanjeet Patel, and Kathrin Plath. "Concise review: Pluripotency and the transcriptional inactivation of the female Mammalian X chromosome." In: *Stem cells (Dayton, Ohio)* 30.1 (Jan. 2012), pp. 48–54.

[134] Sha Sun et al. "Jpx RNA activates Xist by evicting CTCF." In: *Cell* 153.7 (June 2013), pp. 1537–1551.

[135] A C Bell, A G West, and G Felsenfeld. "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators." In: *Cell* 98.3 (Aug. 1999), pp. 387–396.

[136] Jesse R Dixon et al. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." In: *Nature* 485.7398 (May 2012), pp. 376–380.

[137] Johan H Gibcus and Job Dekker. "The hierarchy of the 3D genome." In: *Molecular cell* 49.5 (Mar. 2013), pp. 773–782.

[138] Suzana Hadjur et al. "Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus." In: *Nature* 460.7253 (July 2009), pp. 410–413.

[139] Elphège P Nora et al. "Spatial partitioning of the regulatory landscape of the X-inactivation centre." In: *Nature* 485.7398 (May 2012), pp. 381–385.

[140] Chia-Lun Tsai et al. "Higher order chromatin structure at the X-inactivation center via looping DNA." In: *Developmental biology* 319.2 (July 2008), pp. 416–425.

[141] Luca Giorgetti et al. "Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription." In: *Cell* 157.4 (May 2014), pp. 950–963.

[142] E Heard et al. "Mammalian X-chromosome inactivation: an epigenetics paradigm." In: *Cold Spring Harbor symposia on quantitative biology* 69 (2004), pp. 89–102.

[143] Anne-Valerie Gendrel and Edith Heard. "Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation." In: *Annual review of cell and developmental biology* 30.1 (2014), pp. 561–580.

[144] Cheryl Maduro, Bas de Hoon, and Joost Gribnau. "Fitting the Puzzle Pieces: the Bigger Picture of XCI." In: *Trends in biochemical sciences* 41.2 (Feb. 2016), pp. 138–147.

[145] Stefan F Pinter. "A Tale of Two Cities: How Xist and its partners localize to and silence the bicompartmental X." In: *Seminars in cell & developmental biology* (Apr. 2016).

[146] Julie Chaumeil et al. "A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced." In: *Genes & development* 20.16 (Aug. 2006), pp. 2223–2237.

[147] E S S de Araújo et al. "Stability of XIST repression in relation to genomic imprinting following global genome demethylation in a human cell line." In: *Brazilian journal of medical and biological research = Revista brasileira de pesquisas médicas e biológicas / Sociedade Brasileira de Biofísica ... [et al.]* 47.12 (Dec. 2014), pp. 1029–1035.

[148] Suhas S P Rao et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." In: *Cell* 159.7 (Dec. 2014), pp. 1665–1680.

[149] Xinxian Deng et al. "Bipartite structure of the inactive mouse X chromosome." In: *Genome biology* 16.1 (2015), p. 152.

[150] Luca Giorgetti et al. "Structural organization of the inactive X chromosome in the mouse." In: *Nature* 535.7613 (July 2016), pp. 575–579.

[151] Anton Wutz, Theodore P Rasmussen, and Rudolf Jaenisch. "Chromosomal silencing and localization are mediated by different domains of Xist RNA." In: *Nature Genetics* 30.2 (Feb. 2002), pp. 167–174.

[152] Kavitha Sarma et al. "Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome." In: *Proceedings of the National Academy of Sciences of the United States of America* 107.51 (Dec. 2010), pp. 22196–22201.

[153] Simão Teixeira da Rocha et al. "Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome." In: *Molecular cell* 53.2 (Jan. 2014), pp. 301–316.

[154] Mélanie Makhlouf et al. "A prominent and conserved role for YY1 in Xist transcriptional activation." In: *Nature communications* 5 (2014), p. 4878.

[155] Yesu Jeon and Jeannie T Lee. "YY1 tethers Xist RNA to the inactive X nucleation center." In: *Cell* 146.1 (July 2011), pp. 119–133.

[156] Yuko Hasegawa et al. "The matrix protein hnRNP U is required for chromosomal localization of Xist RNA." In: *Developmental cell* 19.3 (Sept. 2010), pp. 469–476.

[157] Amy Pandya-Jones and Kathrin Plath. "The "lnc" between 3D Chromatin Structure and X Chromosome Inactivation." In: *Seminars in cell & developmental biology* (Apr. 2016).

[158] Matthew D Simon et al. "High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation." In: *Nature* 504.7480 (Dec. 2013), pp. 465–469.

[159] Jesse M Engreitz et al. "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome." In: *Science (New York, N.Y.)* 341.6147 (Aug. 2013), p. 1237973.

[160] Andrea Cerase et al. "Xist localization and function: new insights from multiple levels." In: *Genome biology* 16.1 (2015), p. 166.

[161] C M Clemson et al. "XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure." In: *The Journal of cell biology* 132.3 (Feb. 1996), pp. 259–275.

[162] Hongjae Sunwoo, John Y Wu, and Jeannie T Lee. "The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells." In: *Proceedings of the National Academy of Sciences of the United States of America* 112.31 (Aug. 2015), E4216–25.

[163] Frederic Veyrunes et al. "Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes." In: *Genome research* 18.6 (June 2008), pp. 965–973.

[164] Laura Carrel and Huntington F Willard. "X-inactivation profile reveals extensive variability in X-linked gene expression in females." In: *Nature* 434.7031 (Mar. 2005), pp. 400–404.

[165] Fan Yang et al. "Global survey of escape from X inactivation by RNA-sequencing in mouse." In: *Genome research* 20.5 (May 2010), pp. 614–622.

[166] Shafagh Al Nadaf et al. "A cross-species comparison of escape from X inactivation in Eutheria: implications for evolution of X chromosome inactivation." In: *Chromosoma* 121.1 (Feb. 2012), pp. 71–78.

[167] Janine E Deakin et al. "Unravelling the evolutionary origins of X chromosome inactivation in mammals: insights from marsupials and monotremes." In: *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 17.5 (2009), pp. 671–685.

[168] K D Huynh and J T Lee. "Imprinted X inactivation in eutherians: a model of gametic execution and zygotic relaxation." In: *Current opinion in cell biology* 13.6 (Dec. 2001), pp. 690–697.

[169] Edda Koina et al. "Specific patterns of histone marks accompany X chromosome inactivation in a marsupial." In: *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 17.1 (2009), pp. 115–126.

[170] Ikuhiro Okamoto et al. "Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development." In: *Nature* 472.7343 (Apr. 2011), pp. 370–374.

[171] Xu Wang et al. "Random X inactivation in the mule and horse placenta." In: *Genome research* 22.10 (Oct. 2012), pp. 1855–1863.

[172] Julie Chaumeil et al. "Evolution from XIST-independent to XIST-controlled X-chromosome inactivation: epigenetic modifications in distantly related mammals." In: *PloS one* 6.4 (2011), e19040.

[173] Laurent Duret et al. "The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene." In: *Science (New York, N.Y.)* 312.5780 (June 2006), pp. 1653–1655.

[174] Jennifer Grant et al. "Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation." In: *Nature* 487.7406 (July 2012), pp. 254–258.

[175] Samuel C Chang and Carolyn J Brown. "Identification of regulatory elements flanking human XIST reveals species differences." In: *BMC molecular biology* 11.1 (2010), p. 20.

[176] Céline Vallot et al. "XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells." In: *Nature Genetics* 45.3 (Mar. 2013), pp. 239–241.

[177] A M Villeneuve and B J Meyer. "sdc-1: a link between sex determination and dosage compensation in C. elegans." In: *Cell* 48.1 (Jan. 1987), pp. 25–37.

[178] T W Cline and B J Meyer. "Vive la différence: males vs females in flies vs worms." In: *Annual review of genetics* 30 (1996), pp. 637–702.

[179] J E Madl and R K Herman. "Polyploids and sex determination in Caenorhabditis elegans." In: *Genetics* 93.2 (Oct. 1979), pp. 393–402.

[180] J Hodgkin, H R Horvitz, and S Brenner. "Nondisjunction Mutants of the Nematode CAENORHABDITIS ELEGANS." In: *Genetics* 91.1 (Jan. 1979), pp. 67–94.

[181] J Hodgkin. "X chromosome dosage and gene expression in C. elegans: two unusual dumpy genes". In: *Molecular and general genetics: MGG* 192 (1983), pp. 452–458.

[182] P M Meneely and W B Wood. "An autosomal gene that affects X chromosome expression and sex determination in Caenorhabditis elegans." In: *Genetics* 106.1 (Jan. 1984), pp. 29–44.

[183] P M Meneely and W B Wood. "Genetic analysis of X-chromosome dosage compensation in Caenorhabditis elegans." In: *Genetics* 117.1 (Sept. 1987), pp. 25–41.

[184] J Hodgkin. "Sex determination and dosage compensation in Caenorhabditis elegans." In: *Annual review of genetics* 21 (1987), pp. 133–154.

[185] L M Miller et al. "xol-1: a gene that controls the male modes of both sex determination and X chromosome dosage compensation in C. elegans." In: *Cell* 55.1 (Oct. 1988), pp. 167–183.

[186] C C Akerib and B J Meyer. "Identification of X chromosome regions in Caenorhabditis elegans that contain sex-determination signal elements." In: *Genetics* 138.4 (Dec. 1994), pp. 1105–1125.

[187] N R Rhind et al. "xol-1 acts as an early switch in the C. elegans male/hermaphrodite decision." In: *Cell* 80.1 (Jan. 1995), pp. 71–82.

[188] Jennifer R Powell, Margaret M Jow, and Barbara J Meyer. "The T-box transcription factor SEA-1 is an autosomal element of the X:A signal that determines C. elegans sex." In: *Developmental cell* 9.3 (Sept. 2005), pp. 339–349.

[189] John M Gladden and Barbara J Meyer. "A ONECUT homeodomain protein communicates X chromosome dose to specify Caenorhabditis elegans sexual fate by repressing a sex switch gene." In: *Genetics* 177.3 (Nov. 2007), pp. 1621–1637.

[190] Behnom Farboud et al. "Molecular antagonism between X-chromosome and autosome signals determines nematode sex." In: *Genes & development* 27.10 (May 2013), pp. 1159–1178.

[191] M Nicoll, C C Akerib, and B J Meyer. "X-chromosome-counting mechanisms that determine nematode sex." In: *Nature* 388.6638 (July 1997), pp. 200–204.

[192] Barbara J Meyer. "Targeting X chromosomes for repression." In: *Current opinion in genetics & development* 20.2 (Apr. 2010), pp. 179–189.

[193] J D Plenefisch, L DeLong, and B J Meyer. "Genes that implement the hermaphrodite mode of dosage compensation in Caenorhabditis elegans." In: *Genetics* 121.1 (Jan. 1989), pp. 57–76.

[194] C Nusbaum and B J Meyer. "The Caenorhabditis elegans gene sdc-2 controls sex determination and dosage compensation in XX animals." In: *Genetics* 122.3 (July 1989), pp. 579–593.

[195] A M Villeneuve and B J Meyer. "The regulatory hierarchy controlling sex determination and dosage compensation in Caenorhabditis elegans." In: *Advances in genetics* 27 (1990), pp. 117–188.

[196] A M Villeneuve and B J Meyer. "The role of sdc-1 in the sex determination and dosage compensation decisions in Caenorhabditis elegans." In: *Genetics* 124.1 (Jan. 1990), pp. 91–114.

[197] R D Klein and B J Meyer. "Independent domains of the Sdc-3 protein control sex determination and dosage compensation in C. elegans." In: *Cell* 72.3 (Feb. 1993), pp. 349–364.

[198] H E Dawes et al. "Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate." In: *Science (New York, N.Y.)* 284.5421 (June 1999), pp. 1800–1804.

[199] Gyorgyi Csankovszki et al. "Three distinct condensin complexes control C. elegans chromosome dynamics." In: *Current biology : CB* 19.1 (Jan. 2009), pp. 9–19.

[200] Tatsuya Hirano. "Condensins: organizing and segregating the genome." In: *Current biology : CB* 15.7 (Apr. 2005), R265–75.

[201] Andrew S Belmont. "Mitotic chromosome structure and condensation." In: *Current opinion in cell biology* 18.6 (Dec. 2006), pp. 632–638.

[202] P T Chuang, D G Albertson, and B J Meyer. "DPY-27:a chromosome condensation protein homolog that regulates C. elegans dosage compensation through association with the X chromosome." In: *Cell* 79.3 (Nov. 1994), pp. 459–474.

[203] J D Lieb et al. "MIX-1: an essential component of the C. elegans mitotic machinery executes X chromosome dosage compensation." In: *Cell* 92.2 (Jan. 1998), pp. 265–277.

[204] A V Strunnikov and R Jessberger. "Structural maintenance of chromosomes (SMC) proteins: conserved molecular properties for multiple biological functions." In: *European journal of biochemistry / FEBS* 263.1 (July 1999), pp. 6–13.

[205] Emily Crane et al. "Condensin-driven remodelling of X chromosome topology during dosage compensation." In: *Nature* 523.7559 (July 2015), pp. 240–244.

[206] J Hodgkin. "Male Phenotypes and Mating Efficiency in CAENORHABDITIS ELEGANS." In: *Genetics* 103.1 (Jan. 1983), pp. 43–64.

[207] L DeLong, L P Casson, and B J Meyer. "Assessment of X chromosome dosage compensation in Caenorhabditis elegans by phenotypic analysis of lin-14." In: *Genetics* 117.4 (Dec. 1987), pp. 657–670.

[208] Rebecca R Pferdehirt, William S Kruesi, and Barbara J Meyer. "An MLL/COMPASS subunit functions in the C. elegans dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression." In: *Genes & development* 25.5 (Mar. 2011), pp. 499–515.

[209] D R Hsu and B J Meyer. "The dpy-30 gene encodes an essential component of the Caenorhabditis elegans dosage compensation machinery." In: *Genetics* 137.4 (Aug. 1994), pp. 999–1018.

[210] T L Davis and B J Meyer. "SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome." In: *Development (Cambridge, England)* 124.5 (Mar. 1997), pp. 1019–1031.

[211] M L Nonet and B J Meyer. "Early aspects of Caenorhabditis elegans sex determination and dosage compensation are regulated by a zinc-finger protein." In: *Nature* 351.6321 (May 1991), pp. 65–68.

[212] J D Lieb et al. "The Caenorhabditis elegans dosage compensation machinery is recruited to X chromosome DNA attached to an autosome." In: *Genetics* 156.4 (Dec. 2000), pp. 1603–1621.

[213] Gyorgyi Csankovszki, Patrick McDonel, and Barbara J Meyer. "Recruitment and spreading of the C. elegans dosage compensation complex along X chromosomes." In: *Science (New York, N.Y.)* 303.5661 (Feb. 2004), pp. 1182–1185.

[214] Patrick McDonel et al. "Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex." In: *Nature* 444.7119 (Nov. 2006), pp. 614–618.

[215] Timothy A Blauwkamp and Gyorgyi Csankovszki. "Two classes of dosage compensation complex binding elements along Caenorhabditis elegans X chromosomes." In: *Molecular and cellular biology* 29.8 (Apr. 2009), pp. 2023–2031.

[216] Judith Jans et al. "A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome." In: *Genes & development* 23.5 (Mar. 2009), pp. 602–618.

[217] Sevinc Ercan, Lindsay L Dick, and Jason D Lieb. "The C. elegans dosage compensation complex propagates dynamically and independently of X chromosome sequence." In: *Current biology : CB* 19.21 (Nov. 2009), pp. 1777–1787.

[218] Sevinc Ercan et al. "X chromosome repression by localization of the C. elegans dosage compensation machinery to sites of transcription initiation." In: *Nature Genetics* 39.3 (Mar. 2007), pp. 403–408.

[219] Yuichiro Itoh et al. "Dosage compensation is less effective in birds than in mammals." In: *Journal of biology* 6.1 (2007), p. 2.

[220] Xingfu Zha et al. "Dosage analysis of Z chromosome genes using microarray in silkworm, Bombyx mori." In: *Insect biochemistry and molecular biology* 39.5-6 (May 2009), pp. 315–321.

[221] Beatriz Vicoso and Doris Bachtrog. "Lack of global dosage compensation in Schistosoma mansoni, a female-heterogametic parasite." In: *Genome Biology and Evolution* 3 (2011), pp. 230–235.

[222] Peter W Harrison, Judith E Mank, and Nina Wedell. "Incomplete sex chromosome dosage compensation in the Indian meal moth, Plodia interpunctella, based on de novo transcriptome assembly." In: *Genome Biology and Evolution* 4.11 (2012), pp. 1118–1126.

[223] Beatriz Vicoso et al. "Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation." In: *PLoS biology* 11.8 (2013), e1001643.

[224] Judith E Mank. "The W, X, Y and Z of sex-chromosome dosage compensation." In: *Trends in genetics : TIG* 25.5 (May 2009), pp. 226–233.

[225] James R Walters and Thomas J Hardcastle. "Getting a full dose? Reconsidering sex chromosome dosage compensation in the silkworm, Bombyx mori." In: *Genome Biology and Evolution* 3 (2011), pp. 491–504.

[226] Gilbert Smith et al. "Complete dosage compensation and sex-biased gene expression in the moth Manduca sexta." In: *Genome Biology and Evolution* 6.3 (Mar. 2014), pp. 526–537.

[227] James R Walters, Thomas J Hardcastle, and Chris D Jiggins. "Sex chromosome dosage compensation in Heliconius butterflies: global yet still incomplete?" In: *Genome Biology and Evolution* (Sept. 2015), evv156.

[228] Jacob W Malcom, Randal S Kudra, and John H Malone. "The sex chromosomes of frogs: variability and tolerance offer clues to genome evolution and function." In: *Journal of genomics* 2 (2014), pp. 68–76.

[229] Roberta Bergero and Deborah Charlesworth. "The evolution of restricted recombination in sex chromosomes." In: *Trends in ecology & evolution* 24.2 (Feb. 2009), pp. 94–102.

[230] B Charlesworth. "Model for evolution of Y chromosomes and dosage compensation." In: *Proceedings of the National Academy of Sciences of the United States of America* 75.11 (Nov. 1978), pp. 5618–5622.

[231] B Charlesworth. "Sex chromosomes: evolving dosage compensation." In: *Current biology : CB* 8.25 (Dec. 1998), R931–3.

[232] Tatiana A Gurbich and Doris Bachtrog. "Gene content evolution on the X chromosome." In: *Current opinion in genetics & development* 18.6 (Dec. 2008), pp. 493–498.

[233] Charles Mullon et al. "Evolution of dosage compensation under sexual selection differs between X and Z chromosomes." In: *Nature communications* 6 (2015), p. 7720.

[234] Stacie E Dodgson et al. "Chromosome-Specific and Global Effects of Aneuploidy in Saccharomyces cerevisiae." In: *Genetics* (Feb. 2016), genetics.115.185660.

[235] Audrey P Gasch et al. "Further support for aneuploidy tolerance in wild yeast and effects of dosage compensation on gene copy-number evolution." In: *eLife* 5 (2016), p. 367.

[236] Maxwell Kramer et al. "Developmental Dynamics of X-Chromosome Dosage Compensation by the DCC and H4K20me1 in C. elegans." In: *PLOS Genetics* 11.12 (Dec. 2015), e1005698.

[237] Noriyuki Satoh, Daniel Rokhsar, and Teruaki Nishikawa. "Chordate evolution and the three-phylum system." In: *Proceedings. Biological sciences / The Royal Society* 281.1794 (Nov. 2014), p. 20141729.

[238] André Pires-daSilva and Ralf J Sommer. "Conservation of the global sex determination gene tra-1 in distantly related nematodes." In: *Genes & development* 18.10 (May 2004), pp. 1198–1208.

[239] Richard P Meisel, John H Malone, and Andrew G Clark. "Faster-X evolution of gene expression in Drosophila." In: *PLOS Genetics* 8.10 (2012), e1003013.

[240] José M Eirín-López and Lucas Sánchez. "The comparative study of five sex-determining proteins across insects unveils high rates of evolution at basal components of the sex determination cascade." In: *Development genes and evolution* 225.1 (Jan. 2015), pp. 23–30.

[241] Ivica Letunic and Peer Bork. "Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees." In: *Nucleic acids research* 44.W1 (July 2016), W242–5.

Chapter 2

Evolution of dosage compensation

This work was done in collaboration with Te-Wen Lo, Edward Ralston, Denise Lapidus, Catherine Pickle and Satoru Uzawa. This chapter is to be modified for a manuscript in which Te-Wen Lo and I will share first authorship. My contributions to the data in this chapter include genome sequencing, genome editing in *C. nigoni* and *C. tropicalis* (strains TY5586, TY5754, TY5780, TY5752, TY5771), analysis of mutants (strains TY5586, TY5771, TY5773, and TY5780), ChIP experiments, and data analysis. Te-Wen created and analyzed *C. briggsae* mutants, except TY5773, TY5774, and TY5775, which were created by Edward and Denise, who also created *C. elegans* and *C. tropicalis* mutants for this work. Initial *C. briggsae* ChIP experiments were performed by Te-Wen and analyzed by Edward. William Kruesi identified MEX-II. IF and recruitment assays were performed in collaboration with Te-Wen, Catherine, Satoru, Ed, Denise, and Qian Bian.

The *C. tropicalis* and *C. nigoni* genome sequences are to be published in a manuscript with Erich Schwarz, Da Yin, Erich Haag, and Edward Ralston. Edward assisted with DNA purification. I made sequencing libraries, including separate XX and XO libraries. Genomes were assembled by Erich Schwarz. Da Yin and Erich Haag sequenced *C. nigoni* for the current *de novo* assembly.

2.1 Introduction

Dosage compensation is an essential process that arises in species with chromosome-based sex determination. As XY chromosome pairs evolve, the Y chromosome may erode, leading to an imbalance in gene copy number between XY or X males and XX females or hermaphrodites [1]. Independent, chromosome-wide mechanisms of dosage compensation arose to equalize X-chromosome gene expression between the sexes in mammals, flies, and nematodes. In mammals, one of the two female X chromosomes is randomly, heritably inactivated in each cell early in development. In flies, genes across the single male X chromosome are upregulated approximately two-fold. In *C. elegans*, genes across both hermaphrodite X chromosomes are downregulated by half. In each of these species, different genes were co-opted for X-chromosome dosage compensation. In this study, we asked how dosage compensation mechanisms change over time by comparing mechanisms of dosage compensation across four *Caenorhabditis* species (figure 2.1). We first asked whether homologous machinery accomplishes dosage compensation in these close relatives. We then investigated changes in the molecular mechanism controlling X-specific binding of the dosage compensation machinery.

Dosage compensation and sex determination are linked in *C. elegans*. Sex determination and dosage compensation (*sdc*) genes are important components of the dosage compensation complex (DCC) and the sex determination pathway that trigger dosage compensation and hermaphrodite development in XX animals. Upstream of the *sdc* genes, the X to autosome ratio is communicated by X and autosomal signal elements (XSEs and ASEs). The XSEs and ASEs antagonistically regulate the developmental switch gene, XO lethal (*xol-1*), which is on in males and off in hermaphrodites (figure 2.2) [2, 3, 4]. *sdc-2* is a switch gene that encodes a large protein with a coiled-coil domain and no other recognizable domains. *sdc-2* is repressed by *xol-1* in males, a GHMP kinase family member [5], and is expressed in hermaphrodites, where it acts to trigger DCC binding and signals hermaphrodite fates through DCC binding at the (autosomal) *her-1* locus (figure 2.2) [6]. *xol-1* mutant males die due to inappropriate loading of the DCC and downregulation of the single male X chromosome, and mutations that disrupt *sdc-2* in XX animals lead to masculinization and failure to dosage compensate, the latter of which causes dumpiness or lethality [3].

The *C. elegans* Dosage Compensation Complex (DCC) has 10 defined subunits, including a condensin-like core complex and subunits that recruit the DCC to the X chromosomes [7] (figure 2.3). Condensins are highly conserved, five-member complexes that restructure chromosomes for segregation in meiosis and mitosis throughout eukaryotes. Condensin I and condensin II share a heterodimer of Structural Maintenance of Chromosomes (SMC) family proteins, MIX-1 (SMC-2) and SMC-4, and differ in their three Chromosome-Associated Polypeptides (CAP) proteins (DPY-26, DPY-28, and CAPG-1 in Condensin I and KLE-2, CAPG-2, and HCP-6 in Condensin II) [8]. SMC proteins fold and form dimers at their hinge regions. Their flanking coiled-coil domains associate with each other, and their N- and C-termini form a head region with DNA-binding and ATPase activity [9, 10]. The key step

in the co-option of condensin for dosage compensation may be the *smc-4* duplication that yielded the paralog *dpy-27*, which encodes a DCC subunit. Condensin I^{DC} is the same as Condensin I, except SMC-4 is replaced with DPY-27 [11, 12, 9, 8]. Changes in DPY-27 could allow association with other DCC components (SDC-1, SDC-2, SDC-3, DPY-21, and DPY-30) to restructure the X chromosomes [13] without affecting Condensin I localization or function.

DCC subunits have an order of dependence for binding on the X chromosomes. SDC-2 is thought to bind to the X chromosomes first, since it is found at X chromosome binding sites in the absence of other components and the other components require SDC-2 for binding [14]. SDC-3, a zinc finger protein, and DPY-30, a conserved member of the MLL/COMPASS transcriptional activation complex, depend only on SDC-2 for binding [14, 15]. The rest of the complex depends on these three subunits for X-chromosome binding [14, 15]. SDC-1 and DPY-21 are not necessary for DCC recruitment to the X chromosome, since null mutations do not significantly disrupt DCC binding or reduce viability, but they do have elevated X-chromosome expression and Dpy phenotypes in hermaphrodites [16, 17, 18].

The *C. elegans* DCC is recruited to the X chromosome at specific binding sites called recruitment elements on X (*rex* sites) [19]. These sites are defined by their ability to recruit the DCC in a functional recruitment assay, in which the DNA sequence is introduced at ectopic sites on an autosome or on an extrachromosomal array and stained for colocalization with the DCC. In contrast, sites that are dependent on X (*dox* sites) are only bound by the DCC when present on the X chromosome. The functional recruitment assay was used to narrow the search from the entire X chromosome to the specific X-chromosome sequences that act as *rex* sites. Chromatin immunoprecipitation experiments (ChIP) identified hundreds of DCC-binding sites on the X chromosome. Additional *rex* sites were predicted based on two criteria: first, the sites were highly occupied by the DCC in a wild-type, but not an *sdc-2* mutant strain, and second, they lacked the H3K4me3 enrichment characteristic of transcriptionally active promoters that are usually associated with *dox* sites [15]. The predicted sites that were subsequently tested *in vivo* were all confirmed to be *rex* sites. In *C. elegans*, two highly X-enriched motifs are critical for recruitment at many *rex* sites, the motifs enriched on X, Cel-MEX and Cel-MEX-II ([20], W. Kruesi, unpublished data).

In this study, we compared dosage compensation machinery, the genetic hierarchy, and the DNA binding sites on the X chromosome across four nematode species. Tools were constructed and *C. elegans* protocols were adapted for non-model species. First, chromosome-level genome assemblies were required to investigate DCC recruitment to specific X-chromosome DNA sequences. At the onset, these were only available for *C. elegans* and *C. briggsae*, a species that diverged from *C. elegans* 15-30 million years ago (MYA) (figure 2.1). Second, many protocols for growth and mutagenesis of *C. nigoni*, *C. tropicalis*, and *C. briggsae* had to be adjusted, since, compared to the *C. elegans* N2 strain, their reproductive properties are not as convenient for experiments. We also encountered difficulties in selecting species. *C.*

briggsae and *C. tropicalis* (formerly *C.* species 11) were good choices, since, like *C. elegans*, these androdioecious species consist of self-fertilizing hermaphrodites and males, and as such are subject to far less inbreeding depression than obligate outcrossers [21, 22]. *C. japonica* was initially chosen as an outgroup, but turned out to be a particularly unfortunate choice, since inbreeding depression led to decreased fertility. A genome-edited strain and a population of the wild-type strain died out over a few months. *C. japonica* is not included in this work. Enough divergence was seen within the *C. briggsae* clade that *C. elegans* functions as our outgroup.

Initial sequence analysis and genome editing in *C. briggsae* identified conserved dosage compensation machinery, but divergent DNA binding sites. Since divergence was observed in these close relatives, we decided to investigate species within the *C. briggsae* clade separated by even shorter timescales. We selected *C. nigoni* and *C. tropicalis*, neither of which had a chromosome-level genome assembly. Therefore, we sequenced the *C. nigoni* and *C. tropicalis* genomes to contribute to high quality, chromosome-level genome assemblies. We also had to adapt new genome-editing protocols for these *Caenorhabditis* species (the *C. nigoni* protocol was published in [23], see also [24]) that allowed us to disrupt function and introduce epitope tags in *dpy-27* and *sdc-2* orthologs, to investigate the conservation of function of these genes and the divergence in their DNA-binding patterns. Using these tools, we confirmed that DNA sequence motifs that recruit the DCC to the X chromosome have diverged between *C. elegans* and *C. briggsae*. Although DCC sequence-specificity appears to be shared between the very-close relatives, *C. briggsae* and *C. nigoni*, divergence within the *C. briggsae* clade, between this pair and *C. tropicalis*, indicates rapid coevolution of the DCC and its X-chromosome binding sequences.

2.2 Materials and methods

Strains and maintenance

All strains were maintained at room temperature on NGM plates seeded with *E. coli* strain OP50, unless otherwise specified. Strains are listed (table 2.1). Strains used for mutagenesis or wild-type reference were *C. elegans* N2, *C. tropicalis* JU1373, *C. briggsae* AF16, and *C. nigoni* JU1325 and JU1422. JU1325 was used for *Cni-dpy-27* site-directed mutagenesis. JU1422 is an inbred strain that was used for Illumina sequencing and *Cni-sdc-2* mutagenesis. Strains created by site-directed mutagenesis were backcrossed twice. Homozygous-lethal strains were maintained as heterozygotes.

Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

DNA purification for long-read sequencing

DNA was extracted as follows, with advice from Erich Haag, Da Yin, Erich Schwarz, and Ed Ralston. *C. tropicalis* strain JU1373 was grown on *E. coli* strain OP50 from a single hermaphrodite, then transferred to MYOB plates seeded with *E. coli* strain HB101. Worms were bleached until carcasses dissolved, leaving the embryos relatively bacteria-free. These embryos were plated on fresh MYOB plates with HB101. When bacteria was depleted, worms were separated from agar and debris by sucrose floatation, performed by mixing equal parts 60% sucrose and worms in M9, then centrifuging at 500 x g. Worms were removed from the upper layer of the sucrose gradient, then washed three times in M9. Worms were then washed in disruption buffer (200 mM NaCl, 50 mM EDTA, 100 mM Tris (pH 8.5)) and resuspended in 5 volumes disruption buffer with 0.5% (w/v) SDS (about 3 g pellet in 15 ml total). The sample was frozen at -80°C, then thawed at room temperature. The sample was then incubated at 37°C for 30 minutes with 40 μg/ml RNaseA, then incubated for about 5 hours at 68°C with 2 μg/ml proteinase K . One half volume of phenol was added to the sample, and the tube was gently rotated for 30 minutes at room temperature. One half volume of 24:1 chloroform:isoamyl alcohol was added and mixed gently. The sample was spun at 5 kRPM for 10 minutes. The aqueous layer was transferred to a new tube, then phenol-chloroform extraction was repeated. One volume of 24:1 chloroform:isoamyl alcohol was added to the sample and repeatedly inverted gently to mix. The sample was spun, and the aqueous layer was transferred to a new tube. A tenth volume 3M NaOAc, pH 5.2 was added and mixed. One volume of 100% isopropanol was added and mixed gently. DNA was gently spooled onto a rod, then washed in 70% ethanol. DNA was resuspended in 10 mM Tris-HCl, pH 8. DNA was quantified using Qubit dsDNA high sensitivity assay kit (Invitrogen #Q32851), and run on a 0.8% agarose gel overnight at 20V to verify that high molecular weight DNA was present.

High molecular weight DNA was sequenced to 100X coverage with Pacific Biosciences single-molecule, real-time (SMRT) technology.

Sequence comparisons

The *C. nigoni* and *C. tropicalis* genomes were assembled *de novo* from long-read sequencing data [25, 26], then corrected with Illumina sequencing data by Erich Schwarz (unpublished data).

C. nigoni and *C. tropicalis* predicted genes were identified with Augustus [27] (Erich Schwarz, unpublished data). Homologs of each *C. elegans* DCC component were identified in *C. briggsae*, *C. nigoni*, and *C. tropicalis*. *C. nigoni* and *C. tropicalis dpy-27* and *sdc-2* orthologs were confirmed by partial Sanger sequencing of genomic and complementary DNA (cDNA) libraries (data not shown). A second SDC-2 homolog was identified by BLAST in both *C. nigoni* and *C. tropicalis*; however, the second-best BLAST hit encoded a much smaller predicted protein. We confirmed this homolog was present with Sanger sequencing,

but did not pursue this potential *sdc-2* duplication beyond sequencing. *xol-1* also had a duplication in *C. nigoni* and *C. tropicalis*. In both species, the *xol-1* homologs were found near each other on the same X-chromosome contig.

Pairwise amino acid sequence comparisons were performed with EMBOSS Needle, and multiple sequence alignment was performed with Clustal Omega [28].

Illumina sequencing to identify X-chromosome contigs

DNA libraries were created as described [29], except 100 worms were picked for each sample and worms were sheared in 130 μl lysis buffer by Covaris (10% duty cycle, intensity of 4, 200 cycles per burst, 2 cycles of 60 seconds each). DNA was sequenced with the Illumina HiSeq2000 or the HiSeq4000. Bioanalyzer and sequencing work were performed by the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

Genome editing

TALENs

TALENs were designed using TAL Effector-Nucleotide Targeter (TALE-NT) 2.0 [30], using the Cermak, *et al.*, 2011 [31] architecture, NN for G substitutions, Streubel *et al.* 2012 [32] guidelines, and T only selected for the upstream base. The plasmid kit used for generation of TALENs was a gift from Daniel Voytas and Adam Bogdanove (Addgene kit # 1000000024). TALENs were transcribed *in vitro*, using the mMessage mMachine SP6 Transcription Kit (Catalog # Ambion AM1340). Young adult *C. tropicalis* hermaphrodites and *C. nigoni* females were injected with 1 to 1.5 μg/μl TALEN mRNA and 0.05 μg/μl single-stranded 200 bp oligo repair template (tables 2.2, 2.3). Progeny were screened as described in [23], except where large insertions were expected, PCR products were run on 2% agarose gel without Cel-1 digestion.

Cas9

Cas9 mutagenesis was performed by DNA injection [33] or ribonucleoprotein (RNP) injection [34]. Concentrations used for DNA injection follow: 50 ng/μl *Peft-3::Cas9::tbb-2* expression plasmid, 120 ng/μl *rol-6* guide RNA expression plasmid, 10 ng/μl target guide single guide RNA (sgRNA) expression plasmid, and 5 ng/μl *Cbr-myo-2::gfp* marker plasmid. For ribonucleoprotein (RNP) injection, the following concentrations were used: 15.3 μM Cas9 protein, 12 μM Dharmacon CRISPR RNA (crRNA) *rol-6*, 30 μM Dharmacon crRNA target, 42 μM Dharmacon tracerRNA, 0.42 μM *rol-6* repair oligo, and 1 μM target repair oligo (tables 2.2, 2.3). Purified Cas9 protein was purchased from QB3 Macrolab, Berkeley, California. RNAs were purchased from GE Healthcare Dharmacon Inc.

Mutagenesis strategy

We chose to add an in-frame 3X-FLAG-encoding sequence to *Cbr-sdc-2* and *Ctr-sdc-2* near the start codon because *C. elegans* strain TY4573, with an extrachromosomal array encoding FLAG-tagged SDC-2 at the corresponding locus, produced robust IF and ChIP results with α-FLAG antibody [15]. The resulting FLAG::SDC-2 strains are *Cbr-sdc-2(y716)* (TY5775) and *Ctr-sdc-2(y675)* (TY5743) (table 2.1). The FLAG strains complement the *sdc-2* null.

Structural maintenance of chromosomes (SMC) proteins have well-characterized, conserved structures, so the DPY-27 insertion site was chosen based on homology to an accessible SMC-3 site near the hinge region of *S. cerevisiae* SMC3 ([35], Vincent Guacci, personal communication). We aligned SMC-3 and SMC-4 amino acid sequences from diverse species with *Caenorhabditis* DPY-27 sequences to properly identify the homologous proline-rich region between regions of high conservation near the hinge. 3X-FLAG::DPY-27 strains follow: *dpy-27(y679)* (TY5753), *Cbr-dpy-27(y706)* (TY5774), *Cni-dpy-27(y683)* (TY5754), and *Ctr-dpy-27(y677)* (TY5752). Templated *dpy-27* mutations were introduced at the same sites to introduce a premature stop codon and disrupt function in *Cni-dpy-27(y709)*/+ (TY5780) and *Ctr-dpy-27(y703)*/+ (TY5771), and a 52 bp deletion *Cbr-dpy-27(y705)*/+ (TY5773) (table 2.1).

Western blot

FLAG-tagged DCC components were visualized by Western blot. ChIP extracts (5-15 μl) or 30 to 60 young adult worms were frozen in sample buffer (31.25 mM Tris-HCl, pH 8.5, 5% glycerol, 1% SDS, 2.5% β-mercaptoethanol, 0.125 mg/ml bromophenol blue), boiled for 10 minutes, centrifuged at 13,000 x g for 10 minutes, then run on a 3-8% tris acetate gel. Proteins were transferred to nitrocellulose at 30V overnight. Blots were probed with M2 monoclonal mouse anti-FLAG antibody (Sigma-Aldrich Corporation Catalog #F3165) and donkey anti-mouse antibody conjugated to horseradish peroxidase (Jackson ImmunoResearch Laboratories Catalog #115-035-008). Chemiluminescence was detected with the Western-Bright Sirius Blotting Detection kit (Advansta Corporation Catalog #K-12043-D20).

Immunofluorescence

Young adult worms were cut in egg buffer (25 mM HEPES, pH 7.4, 118 mM NaCl, 48 mM KCl, 0.2 mM CaCl, 0.2 mM MgCl), then fixed in 2% formaldehyde for 5 minutes under a coverslip on a Superfrost Plus slide. Slides were frozen in liquid nitrogen, then coverslips were quickly removed with a razor blade. Slides were washed for 10 minutes at room temperature in PBS-T (1X PBS, 1mM EDTA, 0.5% Triton-X) before and after antibody incubation steps. Antibodies were applied at 1:200 in PBS-T and incubated for 6 to 16 hours at room temperature. Slides were mounted in ProLong antifade (Thermo Fisher Catalog #P36934) with 1 μg/ml DAPI. Nearly all IF images were taken with a Leica TCS

SP2 confocal microscope, except some *Cbr-rex-02* recruitment assays, which were taken with a Leica TCS SP8.

ChIP-seq

Extract preparation

Worms were grown for ChIP at room temperature on approximately 60 MYOB plates seeded with *E. coli* strain HB101. When most of the bacteria on the plates was depleted, mixed stage worms and embryos were removed from the plate in M9 buffer and allowed to settle in a separatory funnel. Worms were separated from agar and debris by sucrose floatation, performed by mixing equal parts 60% sucrose and worms in M9, then centrifuging at 500 x g. Worms were removed from the top layer of the sucrose gradient, then washed three times in M9. Protease inhibitor cocktail III (2 μl, Millipore #EMD 539134-1SET) was added per ml of M9 to wash the pellet, then after centrifugation, volume was reduced to a 1:1 worms:M9 mixture. This mixture was added dropwise into liquid nitrogen. Worm carcasses were disrupted by grinding in liquid nitrogen with a SPEX 6870 Freezer/Mill (2 minutes on, 1 minute off, 4 cycles, 15 cycles per second).

ChIP extracts were made with individual worm strains and also with pooled strains. Pooled samples were combined after sucrose flotation and freezing, but before grinding. Extracts were fixed in 48 ml of 1% formaldehyde for 10 minutes, quenched with 2.5 ml of 2.5 M glycine, then spun down at 4800 x g. Extracts were washed three times in PBS plus protease inhibitor, once in FA buffer (50mM HEPES/KOH pH7.5, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 150mM NaCl) plus protease inhibitor, then resuspended in 1 ml FA buffer plus protease inhibitor per gram of pellet. Extracts were sheared by Covaris S2 (20% duty factor, power level 8, 200 cycles per burst) for a total of 35 minutes of processing time (60 seconds on, 45 seconds off, 35 cycles), then centrifuged in a tabletop centrifuge at 4°C for 15 minutes at max speed. Supernatant was transferred to a new tube.

ChIP

Extract protein concentrations were approximately 10 mg/ml as quantified by BCA assay (Fisher Scientific Catalog #PI 23227). Extracts (750 μl samples) were rotated with 20 μl 1 mg/ml M2 mouse monoclonal anti-FLAG antibody or 1 mg/ml mouse IgG antibody (Sigma-Aldrich Corporation Catalog #I5381) at 4°C overnight. Protein G Sepharose Magnetic Beads (50 μl per sample, GE Life Sciences Catalog #28-9440-08) were washed twice in FA buffer, then extract and antibody were added to the washed beads. Extracts and beads were rotated at 4°C for 4 to 6 hours. Beads were washed by rotating at room temperature in 1 ml of each of the following buffers: 5 minutes in FA buffer twice, 10 minutes in FA with 1M NaCl, 10 minutes in FA with 0.5M NaCl, 10 minutes in TEL (0.25 M LiCl, 1% NP40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0), and 5 minutes in TE twice. The ChIP sample was shaken at 65°C in 150 μl elution buffer (1% SDS in TE with 250 mM NaCl) and eluted, then wash and elution was repeated. Samples were then

treated with 2μl 10 mg/ml RNaseA at room temperature for 1 to 2 hours. Crosslinking was reversed and proteins were digested with 2 μl of 10 mg/ml proteinase K overnight at 65°C. DNA was purified by precipitation with two volumes ethanol after adding the carrier GlycoBlue (Ambion Catalog #AM9515).

Library preparation and sequencing

ChIP-seq libraries were prepared as follows, with MinElute PCR purification steps following each enzymatic reaction. DNA ends were prepared for adapter ligation using the End-It DNA End-Repair kit (Epicentre Biotechnologies Catalog #ER81050). Addition of the 3' A was performed with Klenow Fragment (3→5 exo-) in NEBuffer 2 (New England Biolabs Catalog #M0212L and #B7002, respectively) with 0.2 mM dATP. Solid-phase reversible immobilization (SPRI) beads (Beckman Coulter Genomics Catalog #A63881) were used for size selection. A ratio of 0.5:1 bead solution to DNA sample was used to remove fragments larger than approximately 700 bp from supernatant, then the ratio was increased to 1:1 in order to retain fragments larger than approximately 150 bp. NEXTflex DNA barcodes (Bioo Scientific Catalog #514102) were added with Quick ligase (New England BioLabs Catalog #M2200), and libraries were PCR amplified for 16 cycles. Libraries were run on a 2% agarose gel. DNA fragments between 200 and 400 bp were selected by gel extraction with the Qiagen gel extraction kit (Catalog #28704) or by Pippin prep (Sage Science). DNA fragmentation and concentration were measured by 2100 Bioanalyzer (Agilent Technologies) using the high sensitivity DNA analysis kit (Agilent Technologies Catalog #5067-4626). DNA was sequenced with the Illumina HiSeq2000 or the HiSeq4000. Pippin prep, bioanalyzer, and sequencing work were performed by the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

Analysis

Reads were filtered with CASAVA 1.8, then aligned to reference genomes with Bowtie [36], allowing 2 mismatches and no duplicate sites. Read depth was assessed with SAMtools mPileup [37]. Peaks were called with Model-Based Analysis of ChIP-Seq (MACS2) [38], with callpeak using *C. elegans* effective genome size and including a maximum of 10 duplicate reads. When peak calling failed, the additional parameters "–nomodel" and "–shiftsize 100" were used: α-Cbr-SDC-2 (BMCS205B vs. C), α-Cni-DPY-27 (BMCS206A vs. C), α-DPY-27 vs. IgG (WSK3), α-Cbr-DPY-27 vs. IgG (TL21), and α-Cbr-MIX-I vs. IgG (TL21) (table E.1) . Motifs were identified with MEME [39, 40].

In vivo recruitment assays

Young adult worms were injected with 5 ng/μl *C. elegans Pmyo-2::gfp* co-injection marker plasmid (pPD118.33, expressed in the pharynx), 10 ng/μl DNA to be tested for DCC recruitment, and 65 ng/μl pGEM 7Z+ as filler plasmid. The protocol was refined by the addition of a *C. briggsae*-specific *Cbr-Pmyo-2::gfp* (Ed Ralston, unpublished data). Stain-

ing was performed as above, with the addition of fluorescence *in situ* hybridization (FISH) after fixation and before antibody staining. An AlexaFluor 555 probe was created with *C. elegans Pmyo-2::gfp* co-injection marker plasmid with the FISH Tag DNA Orange kit (ThermoFisher Scientific Catalog #F32948) as described. After fixation, slides were washed three times in PBS-T for 10 minutes, then water was removed in 95% ethanol for 10 minutes. 15 μl hybridization solution (30% formamide, 10% dextran sulfate, 1-10 ng probe DNA in 3X SSC) was placed on each slide and covered with a glass coverslip. FISH probes were hybridized in a temperature-controlled slide chamber (Bio-Rad ALD0211 Alpha Unit Block Assembly) with the following program: 80°C for 10 minutes, 0.5°C per second to 50°C, °C for 1 hour, 0.5°C per second to 45°C, 45°C for 1 hour, 0.5°C per second to 40°C, 40°C for 1 hour 0.5°C per second to 38°C, 38°C for 1 hour, 0.5°C per second to 37°C, then 37°C overnight. The next day, slides were washed at 39°C three times in each of the following buffers: 50% formamide in 2X SSC (300 mM NaCl, 30 mM sodium citrate) for 15 minutes, 25% formamide in 2X SSC for 10 minutes, 2X SSC for 10 minutes, 1X SSC for 1 minute. Slides were washed three times in PBS-T before antibody incubation (as described above).

Images were taken as described above, deconvolved with Huygens Professional (Scientific Volume Imaging, The Netherlands, http://svi.nl), and analyzed in Priism [41]. Overlap of FISH staining with DCC staining was scored as recruitment.

Phenotypic analysis

L4 worms were transferred to individual plates for mating or self-fertilization to analyze brood size, lethality, and morphological phenotypes (dumpy, small, or sick). Parents were moved and embryos were counted approximately every 12 hours until laying stopped. Hatched worms were counted and phenotypes were assessed when progeny reached the L4 or young adult stage. Single worms were genotyped by PCR with the following primers: *Cni-dpy-27* with CS553 (CATCGTTCTTCGCTCTGGAGTACGG) and CS556 (GCTTCTTCCATGTCTTTAGCCAACAG), *Ctr-dpy-27* with CS362 (GACGAAGGAGGATGTGAAGAAGGCTATC) and CS443 (CTCCAAAAACTCACAAAGACTCTG), and *Cbr-dpy-27* with ER418 (GACACATGAGGACTGCATAGCAG) and ER419 (GCTCCGGATCGTTCGATGAGTC).

Most strains with insertions or deletions were compared to wild-type strains by gel electrophoresis with 2% agarose, however the 14 bp deletion in strain *Cni-sdc-2(y516)/+* (TY5586) was separated from a wild-type band on a 4% 3:1 agarose gel (GenePure Catalog # A00158).

2.3 Results

The DCC is conserved across *Caenorhabditis*

Condensin subunits that restructure chromosomes for proper segregation during cell division were co-opted in *C. elegans* to restructure the X chromosomes for hermaphrodite-specific gene repression during dosage compensation [7, 13]. In what may be the critical event for condensin co-option, condensin subunit SMC-4 was duplicated to create the paralog, DPY-27, which has a unique role in the DCC. We asked whether a condensin-driven DCC accomplishes dosage compensation in the *C. briggsae* clade by examining the roles of DPY-27 orthologs in three species: *C. briggsae*, *C. nigoni*, and *C. tropicalis*. We identified DPY-27 orthologs in each species, then used genome editing to FLAG-tag each ortholog for biochemical analysis and to create null alleles for functional analysis.

As background necessary for comparing *C. elegans* and *C. briggsae*, *C. elegans dpy-27* accomplishes its XX-specific role in dosage compensation through its X-chromosome localization. Mutations in *dpy-27* cause maternal-effect lethality in XX animals, with escapers that have a Dpy phenotype. XO homozygotes are not affected. The maternal effect causes homozygous XX hermaphrodites from heterozygous mothers to be slightly Dpy, but their XX progeny that lack a maternal *dpy-27* contribution are dead or severely Dpy [42]. Matings between homozygous *dpy-27* mutant mothers and wild-type fathers produce healthy XX heterozygotes, indicating zygotic rescue [42].

We asked first whether *Cbr-dpy-27* mutations cause maternal-effect or recessive phenotypes in *C. briggsae*. With genome editing, we created a 52 bp deletion near the hinge region in *Cbr-dpy-27* that caused a frameshift and in-frame stop codon (*Cbr-dpy-27(y705)*). If the mutation is recessive, we would expect 25% of the progeny from a selfed heterozygous mother to be homozygous, and therefore have a mutant phenotype. A maternal effect would rescue homozygotes with heterozygous mothers. Consistent with recessive inheritance and no maternal effect, *Cbr-dpy-27(y705)*/+ heterozygotes produced an average of 70% non-Dpy progeny, 13% Dpy or small progeny, and 17% embryonic lethality (30% Dpy or dead progeny is close to the expected 25%) (table 2.4).

We confirmed that these phenotypes were specific to *Cbr-dpy-27(y705)* homozygotes by PCR genotyping. Among a representative sample of 319 living progeny from three selfed *Cbr-dpy-27(y705)* heterozygotes, 14% were Dpy homozygotes, 63% were non-Dpy heterozygotes, 21% were non-Dpy and wild-type at the *Cbr-dpy-27* locus, and less than 2% did not fit into these categories (figure 2.4). Similar results were obtained in a strain with a *Cbr-dpy-27* deletion that removed a region including the entire first exon (allele *Cbr-dpy-27(y436)*) (Te-Wen Lo, unpublished data).

We then tested for zygotic rescue by mating homozygous *Cbr-dpy-27(y705)* hermaphrodites

to wild-type males. Although few cross-progeny were produced, all of these were non-Dpy (data not shown), indicating zygotic rescue in *C. briggsae*, as seen in *C. elegans*. We also confirmed that the sterility we observed was sex-specific, as follows. None of the five homozygous hermaphrodites we carefully tracked produced embryos (table 2.4). In contrast, males homozygous for the *Cbr-dpy-27* mutation (that were also homozygous for a *Cbr-him-8* mutation) were mated to *Cbr-she-1* females (spermless hermaphrodites [43]) to produce non-Dpy cross-progeny.

We also tested Cbr-DPY-27 for conserved localization by IF, indicating a conserved function in the DCC. An antibody raised to a Cbr-DPY-27 peptide colocalized with X-chromosome FISH in hermaphrodites, confirming X-localization is conserved in *C. briggsae* (figure 2.5, Te-Wen Lo, unpublished data). The α-FLAG antibody also labeled the DCC in a genome edited strain encoding 3X-FLAG-tag near the hinge region of Cbr-DPY-27, confirming that the tag is accessible in *C. briggsae* (figures 2.6, 2.7). For further confirmation that the FLAG tag is specific, α-Cbr-DPY-27 and α-FLAG colocalize on the hermaphrodite X chromosomes in this strain (data not shown). Also, we confirmed that the FLAG insertion does not disrupt DCC function in a mating in which the FLAG::Cbr-DPY-27 complemented the null, indicating no dosage compensation phenotype (data not shown).

We next asked whether the DPY-27 ortholog in the very close *C. briggsae* relative, *C. nigoni*, has conserved function. Cbr-DPY-27 and Cni-DPY-27 are nearly identical; they share 92% identity and 94% similarity at the amino acid level (figure 2.8a). As expected for conserved DCC function, FLAG::Cni-DPY-27 localization was consistent with that of other DPY-27 orthologs (figures 2.9, 2.6). To disrupt function, we inserted a 3X-FLAG tag and an in-frame stop codon at the same site in the Cni-DPY-27 hinge region, allele *Cni-dpy-27(y709)*.

If the *Cni-dpy-27(y709)* mutation were recessive, we would expect homozygous XX-animals to be Dpy or dead, however general lethality and other non-specific phenotypes in this strain complicated our analysis (described below). Out of 356 total embryos produced in one mating between one heterozygous female and one homozygous male, we would expect 25% to be Dpy or dead (none of the XO-males and half of the XX-females), to our surprise, 37% died before hatching, and the excessive lethality was not sex-specific.

We genotyped the *Cni-dpy-27* locus in 219 of the viable progeny to determine whether the *Cni-dpy-27(y709)* allele caused sex-specific, recessive dumpiness in the survivors. Approximately half (51%) were male and half (49%) were female or too small to accurately sex, which could indicate that the *Cni-dpy-27(y709)* allele or background mutation(s) caused general lethality. As described below, it appears to be the latter. Of the adult males, 46% were heterozygous, 54% were homozygous, consistent with the 50:50 ratio expected if XX-specific function is conserved. Also as expected, none of the males were Dpy or small. Of the surviving progeny that were female or too small to identify, we would expect more than half

to be non-Dpy heterozygotes and the remainder to be Dpy homozygotes if the mutation is XX-specific recessive. Instead, 16% were Dpy or small heterozygotes, 57% were non-Dpy heterozygotes, 25% were Dpy or small homozygotes, and 2% were non-Dpy homozygotes (figure 2.4). Either the mutation is semi-dominant or background mutation(s) affect phenotype. Similar to *Cbr-dpy-27*, homozygous mutant *Cni-dpy-27(y709)* females from heterozygous mothers either died as embryos or had severe morphological defects (with two exceptions that could be potentially attributed to experimental error), which is consistent with a lack of maternal effect in this *Cni-dpy-27* mutant.

Phenotypic analysis of the *Cni-dpy-27(y709)*/+ strain was complicated, primarily by inbreeding depression in this gonochoristic species. We know that lethality and mutant phenotypes were non-specific because fertility and lethality did not correlate with parental genotype at the *Cni-dpy-27* locus and were highly variable (table 2.4). For example, in one mating between parents that were both wild-type at the *Cni-dpy-27* locus, we would expect about half of the progeny to be non-Dpy female and half to be non-Dpy male, however, of 135 embryos, 13% were non-Dpy female, 8% were non-Dpy male, 18% were Dpy or small, and 61% died before hatching. Consistent with inbreeding depression, outcrossing *Cni-dpy-27(y709)* heterozygous and homozygous males to the parent strain JU1325 improved fertility and viability in the F1 progeny to wild-type levels. However, a baseline of 9% lethality and 7% mutant phenotypes were observed in progeny of the wild-type strain (egg-to-adult viability of 84%). Consistent with this finding, the egg-to-adult viability for *C. nigoni* matings has been previously reported at 75% ±7% [44] and 82% [45].

We also asked whether mutations in the *C. tropicalis dpy-27* ortholog cause maternal-effect or recessive phenotypes. Ctr-DPY-27 shares 53% similarity with Cbr-DPY-27 and 54% similarity with Cel-DPY-27. Using an allele *Ctr-dpy-27(y703)* in which a 3X-FLAG sequence and a premature stop codon were inserted near the hinge region, we observed recessive function, rather than a maternal effect. As expected for recessive function, a representative sample of self-progeny from four *Ctr-dpy-27(y703)*/+ heterozygotes produced 75% non-dumpy hermaphrodites and 25% dumpy or dead progeny (9% and 16% of an average 189 embryos, respectively) (table 2.4). PCR genotyping confirmed that the *Ctr-dpy-27(y703)* mutation is recessive; all Dpy or small progeny were homozygous, none of the homozygotes were non-Dpy, and all animals heterozygous or wild-type at the *Ctr-dpy-27* locus were non-Dpy (figure 2.4).

In addition to conserved mutant phenotypes within the *C. briggsae* clade, staining of FLAG::Ctr-DPY-27 (allele *Ctr-dpy-27(y677)*) was consistent with conserved X-localization of this DCC subunit (figure 2.6). ChIP-seq data described below confirms that the DPY-27 ortholog in each species binds to the X chromosome.

The genetic hierarchy that controls X- and sex-specificity for the dosage compensation process is conserved in the *C. briggsae* clade

Since the DCC-specific condensin I^{DC} subunit DPY-27 is conserved in the *C. briggsae* clade, we next asked whether the genetic hierarchy controlling X- and sex-specificity for the dosage compensation process is also conserved. In *C. elegans*, the XX-specific DCC component SDC-2 is the prime DCC recruiter. XX *sdc-2* mutants die because the DCC fails to load on the X chromosomes, and rare survivors are masculinized [3]. The master switch *xol-1* is required to repress expression of *sdc-2* in XO animals [2]. XO *xol-1* mutants die because the DCC is inappropriately loaded on their single X chromosome [3]. *xol-1 sdc-2* XO double mutants are viable, and *xol-1 sdc-2* XX double mutants are dead [14]. We created mutations in *Cbr-sdc-2*, *Cni-sdc-2*, *Ctr-sdc-2*, and *Cbr-xol-1*, and also inserted FLAG tags following the start codon in *Cbr-sdc-2* and *Ctr-sdc-2* for functional and biochemical analysis.

We first discuss the *sdc-2* gene orthologs. Cbr-SDC-2 and Cel-SDC-2 share only 26% identity and 43% similarity along 3248 and 2962 amino acids, respectively (figure 2.8b). Despite amino acid sequence divergence, *Cbr-sdc-2* has conserved function in sex determination and dosage compensation. *Cbr-sdc-2* mutations caused 98% XX-specific lethality and rare XX survivors were Dpy and masculinized (figure 2.10C) (Te-Wen Lo, unpublished data). All XO hemizygous *Cbr-sdc-2* mutant males were viable and phenotypically wild-type (figure 2.10B). As in *C. elegans*, XX *Cbr-sdc-2* mutant homozygotes failed to load the DCC as assessed by IF with α-Cbr-DPY-27 antibody (figure 2.10C) (Te-Wen Lo, unpublished data). Also, FLAG::Cbr-SDC-2 (allele *Cbr-sdc-2(y716)*) localization is consistent with conserved DCC targeting to X as assessed by IF with α-FLAG antibody (data not shown). ChIP-seq data described below confirms that Cbr-DPY-27 and Cbr-SDC-2 colocalize on the *C. briggsae* X chromosomes.

We asked whether the *C. nigoni sdc-2* ortholog is functionally conserved compared to the very similar *C. briggsae sdc-2* ortholog. Cni-SDC-2 and Cbr-SDC-2 share 89% identity and 94% similarity (figure 2.8b). We analyzed the phenotype of a mutant strain with a 14 bp deletion that caused a frameshift 38 amino acids into the *Cni-sdc-2* gene (allele *Cni-sdc-2(y516)*). In matings between heterozygous females and hemizygous males, nearly equal numbers of heterozygous and homozygous mutant *Cni-sdc-2(y516)* females survived to adulthood (figure 2.11). Although homozygous females survived to adulthood, these were sterile with more subtle morphological defects including protruding gonad, absence of oocytes, and abnormal germline morphology (figure 2.12). These morphological defects were also observed in some heterozygous females, but not in the wild-type strain (figure 2.12). As in *C. elegans* and *C. briggsae*, hemizygous mutant males were viable, wild type in appearance, and produced progeny upon mating (figure 2.11), consistent with a lack of function for *Cni-sdc-2* in XO males. Although we saw XX-specific sterility and not the expected highly penetrant XX-lethality, we do not know if this difference is caused by divergence in *Cni-sdc-2* function because the 14 bp deletion and frameshift is so close to the start that it

may not be a true null allele.

We next asked whether *C. tropicalis sdc-2* mutations cause XX-specific lethality or sex determination defects. Ctr-SDC-2 shares 43% similarity with *C. elegans* and *C. briggsae* SDC-2, the same level of sequence conservation observed between the functionally conserved Cel-SDC-2 and Cbr-SDC-2 (figure 2.8B). An 11 kb deletion in *Ctr-sdc-2* was created in the *Ctr-dpy-27(y677)* (FLAG::Ctr-DPY-27) background that removed most of the gene (from amino acid 80 to 3316, strain *Ctr-dpy-27(y677)*; *Ctr-sdc-2(y719)*/+). As expected for a gene with conserved DCC function, *Ctr-sdc-2(y719)* caused hermaphrodite-specific embryonic lethality (Denise Lapidus, unpublished data). Outcrosses were performed with hemizygous *Ctr-sdc-2(y719)* males, which were viable and fertile, consistent a lack of function for *Ctr-sdc-2* in XO males (D. Lapidus, unpublished data).

Consistent with XX-lethality caused by inability to dosage compensate, we determined that Ctr-DPY-27 requires *Ctr-sdc-2* for proper localization. As in *C. elegans* and *C. briggsae*, a subset of embryos from selfed *Ctr-sdc-2(y719)* heterozygotes failed to load the FLAG::Ctr-DPY-27 on the X chromosomes as assessed by IF with α-FLAG antibody (figure 2.13). IF staining of the FLAG-tagged Ctr-SDC-2 with α-FLAG antibody also supports conserved X-localization of the DCC in hermaphrodites (figure 2.6). As described below, ChIP-seq data confirms that Ctr-DPY-27 and Ctr-SDC-2 colocalize on the *C. tropicalis* X chromosomes.

We have shown above that *sdc-2* orthologs are required for dosage compensation and hermaphrodite development in the *C. briggsae* clade, so we also asked whether *Cbr-xol-1* represses *Cbr-sdc-2* in a conserved genetic hierarchy. Although Cbr-XOL-1 shares only 17% amino acid identity and 31% similarity with *C. elegans* XOL-1 (figure 2.8c), a *Cbr-xol-1* mutation caused fully-penetrant XO-lethality, similar to *C. elegans xol-1* mutations (figure 2.10D). By analyzing the double mutant, we determined that *Cbr-sdc-2* mutations suppress *Cbr-xol-1* mutations. As in *C. elegans*, XO *Cbr-sdc-2 Cbr-xol-1* double mutants were viable and XX *Cbr-sdc-2 Cbr-xol-1* double mutants were dead or masculinized, indicating conserved epistatic interactions driving sex determination and dosage compensation (figure 2.10E) (Te-Wen Lo, unpublished data).

Although the genetic hierarchy driving X- and sex-specific dosage compensation is conserved between *C. elegans* and *C. briggsae*, other species in the *C. briggsae* clade have two predicted *xol-1* homologs that might participate in sex determination (appendix C). In both *C. nigoni* and *C. tropicalis*, the two *xol-1* homologs were predicted within a single X-chromosomal contig. The two *C. tropicalis xol-1* homologs were also identified in a genome assembly created independently in another laboratory, thus are unlikely to be alleles of the same gene. Although sequence divergence is high (figure 2.8c), each homolog has a predicted structure similar to the *C. elegans* XOL-1 GHMP kinase structure [46]. The evolutionary history and relative importance of each *xol-1* homolog for sex determination is currently unclear.

Creating chromosome-level genome assemblies to understand X-chromosome DCC sequence-specificity across species

Before we could investigate the evolution of DCC targeting to the X chromosome, we needed chromosome-level genome assemblies. When this project began, the *C. briggsae* genome was nearly complete, with nearly all sequences assigned in order to chromosomes, but the *C. nigoni* and *C. tropicalis* genomes were not. Neither of these were chromosome-level, and much worse, the *C. nigoni* sequences were contaminated with *C.* species 7 sequences. To create an improved assembly, we sequenced the *C. nigoni* genome with Illumina sequencing (assemblies by Erich Schwarz). We aligned the ($>$ 10,000) *C. nigoni* contigs to the *C. briggsae* genome to create a draft chromosome-level assembly based on synteny. This chromosome-level assembly was useful for our initial analysis. More recently, a new, *de novo* assembly was created with long PacBio reads and short Illumina reads by Da Yin and Eric Haag. The long reads resulted in a much better assembly of 211 contigs. Again, we assigned contigs to chromosomes and put them in order based on synteny with *C. briggsae*.

In an independent analysis, we assigned *C. nigoni* contigs to the X chromosome or to autosomes, as follows. Adult XX females and XO males were sequenced separately. The number of reads per contig was normalized to the total read depth for each library. The normalized number of reads in the XX libraries was divided by the normalized number of reads from the XO libraries to get an XX/XO read depth ratio. Because there are two X chromosomes for every two sets of autosomes in females and only one X chromosome for every two sets of autosomes in males, the XX/XO ratio should be higher for X-chromosome contigs than for autosomal contigs (figure 2.14). Contigs assigned to the X chromosome by synteny with *C. briggsae* were confirmed to be X-chromosome contigs by XX/XO read depth ratios. In addition, four contigs that share homology with *C. briggsae* autosome sequences were found to be X contigs by their XX/XO read depth ratio. We also found an error in the *C. nigoni* assembly by assessing XX/XO read depth ratios. One contig with homology to both *C. briggsae* chromosome V and the X chromosome had a greater XX/XO ratio on the end that aligned to the X chromosome, consistent with an assembly error in this contig.

We sequenced the *C. tropicalis* genome with short Illumina reads and long PacBio reads to create a high quality *de novo* assembly. For this work, we are using an intermediate assembly with 33X PacBio read coverage. Although we expect the assembly to improve when the full 100X coverage is used, this 33x genome assembly is only 141 contigs, which is very good for a *de novo* assembly. We sequenced XX hermaphrodites and XO males separately to assign contigs to the X chromosome or to autosomes, as described above for the *C. nigoni* genome (figure 2.15). *C. tropicalis* contigs did not align well to the *C. briggsae* genome, so we did not order the contigs by synteny.

ChIP-seq experiments identify discrete DCC binding sites in four species

With chromosome-level genome assemblies, we were able to ask whether DCC-binding profiles are similar across species, reflecting a similar mode of dosage compensation. In *C. elegans*, the DCC binds to discrete sites and acts at a distance to control gene expression chromosome-wide [20], however alternate models of chromosome-wide dosage compensation exist. For example, the *Drosophila melanogaster* DCC binds and spreads to broad regions across gene bodies to locally upregulate gene expression [47]. We used ChIP-seq in our genome-edited strains to determine whether the discrete DCC-binding peaks observed in *C. elegans* are conserved or some other pattern of DCC-binding occurs in species across the *C. briggsae* clade.

To make comparisons across species, we were able to use the same α-FLAG antibody across species to bind the same epitope in genome edited strains, as close as possible to the same part of the complex. We expect the ChIP experiments to be as close to directly comparable as possible. We made extracts for ChIP-seq with each FLAG strain individually and also with three and four species' FLAG::DPY-27 strains pooled together in a single tube to control for ChIP conditions (table E.1). We compared our results to two different types of negative controls: 1) chromatin from each extract was immunoprecipitated with pre-immune IgG antibody, which allowed us to subtract non-specific signal at highly ChIP-able regions, and 2) a control for the α-FLAG antibody was performed with wild-type strains that lack a FLAG-tag. We validated our approach by comparing *C. elegans* ChIP-seq results across α-DPY-27 and α-FLAG antibodies and between ChIP libraries with *C. elegans* alone or pooled with other species (figure 2.16). We called ChIP-seq peaks in each library and compared peak heights from one library to another in pairwise correlation plots (appendix H). We see conserved strong peaks with differences in rank order across libraries (figures H.1, H.2). Both IgG and α-FLAG negative controls were similar to each other (data not shown). We used the IgG control for the analysis here.

Consistent with a conserved dosage compensation style, ChIP-seq experiments identified discrete, narrow X-chromosome peaks in *C. elegans*, *C. briggsae*, *C. nigoni*, and *C. tropicalis* (figures 2.16, 2.17, 2.18, 2.19). X-chromosome DCC peaks also have similar GC content across *C. elegans*, *C. briggsae*, *C. nigoni*, and *C. tropicalis*. We found that GC content was much higher at DCC-binding sites than the genomic average in all four species. The 250 bp regions centered on the top 500 peak summits had GC content between 45% and 47%. X-chromosome, autosome, and total genome GC content was consistently between 35% and 38%.

A review of *C. elegans* DCC binding to X chromosomes

We analyzed our results in other species by comparison to *C. elegans*. Here, we review the properties of *C. elegans* X-chromosome DCC-recruitment sites (*Cel-rex* sites) (figure F.2). Motif searches among *Cel-rex* site sequences identified two important DNA sequence motifs, Cel-MEX (12 bp) and Cel-MEX-II (26 bp) ([20] and W. Kruesi, unpublished data). Cel-MEX-II was initially identified among strong *Cel-rex* site sequences that lack Cel-MEX motifs. Some *Cel-rex* sites have a single Cel-MEX or Cel-MEX-II, and some have clusters of Cel-MEX, Cel-MEX-II, or both motifs (figure F.1). The motif score for a sequence refers to the natural log of the probability of finding such a match to the consensus matrix in the genome, given the GC content of the genome. These motifs are highly enriched on the *C. elegans* X chromosome compared to autosomes at ln(P) scores less than (better than) -15 or -16 (figures 2.20, 2.21). They were shown to be important for binding to *Cel-rex* sequences in three assays. One, arrays containing *Cel-rex* sequences recruited the DCC *in vivo*, but the same sequence minus the motif(s) did not. Two, new binding sites on the X chromosome were created by inserting these motifs into a site that was previously not bound by the DCC, then assessed for DCC occupancy by ChIP-qPCR. Third, small biotinylated DNA fragments with motifs were better able than similar fragments lacking these motifs to pull down DCC subunits from crude extract in an *in vitro* assay. Some *Cel-rex* sites do not have either Cel-MEX or Cel-MEX-II motifs, thus DCC sequence-specificity is not fully understood in *C. elegans*.

DCC targeting to X chromosomes has diverged in *C. briggsae*

Given that orthologs of the condensin I^{DC} subunit DPY-27 and the DCC-loader SDC-2 bind to the X chromosomes in species of the *C. briggsae* clade (figures 2.17, 2.18, 2.19), we next asked whether the DCC sequence-specificity is also conserved in these species. Here we show 5 lines of evidence that support a divergence in X-chromosome DCC targeting sequences between *C. elegans* and *C. briggsae*. First, X-enriched motifs that drive recruitment of the DCC to the X chromosomes in *C. elegans* were not enriched on the X chromosome in *C. briggsae* (figures 2.20a, 2.21a). Second, strong matches to *C. elegans* motifs on the *C. briggsae* X chromosome were not bound by the *C. briggsae* DCC (figures 2.20b-c, 2.21b-c). Third, *C. elegans rex* site sequences introduced in an extrachromosomal array in *C. briggsae* generally do not recruit the *C. briggsae* DCC (table 2.5). Fourth, new motifs that differ from Cel-MEX and Cel-MEX-II were identified in *C. briggsae rex* sequences that contribute to DCC recruitment in *C. briggsae* (figures 2.22, 2.23, 2.24). Fifth, these *C. briggsae* X-enriched motifs were not enriched on the *C. elegans* X chromosome (figures 2.25, 2.26).

Our first indiction that DCC targeting has diverged in *Caenorhabditis* was the lack of X-enrichment of Cel-MEX and Cel-MEX-II in *C. briggsae*. Although high-scoring Cel-MEX and Cel-MEX-II motifs were highly enriched on the *C. elegans* X chromosome, they were not enriched at any score in *C. briggsae*, nor were they enriched in the strongest ChIP-seq peaks

(figures 2.20, 2.21 and table F.4). Rare Cel-MEX and Cel-MEX-II motifs on the *C. briggsae* X chromosome were not bound by the *C. briggsae* DCC. Recruitment assays in *C. briggsae* demonstrated that extrachromosomal arrays carrying two *C. briggsae* X regions, each with a strong Cel-MEX motif (scores -17.56 and -16.93), failed to recruit the *C. briggsae* DCC (table 2.5).

We next asked whether *C. elegans rex* sequences could drive recruitment in *C. briggsae* and whether *C. briggsae* and *C. elegans rex* sites are found at homologous loci. We tested *C. elegans rex* sites that were in genes or promoters so the homologous regions could be selected in *C. briggsae* with greater confidence. *Cel-rex-3* and *Cel-rex-4* each contained a Cel-MEX motif (ln(P) of -14.72 and -15.8, respectively), *Cel-rex-33* contained three Cel-MEX motifs (-15.46, -15.45, and -13.23), and *Cel-rex-39* contained two strong Cel-MEX-II motifs (ln(P) of -20.85 and -21.3). These consistently recruited the *C. elegans* DCC to high copy number arrays, but did not recruit in *C. briggsae* (or the level of recruitment was not above background) (table 2.5). We also tested *C. briggsae* sequences in genes homologous to the genes that contain *Cel-rex-4*, *Cel-rex-33*, and *Cel-rex-39* with a Cel-MEX (ln(P) of -15.8), no *C. elegans* motifs, and two weak Cel-MEX-II motifs (ln(P) of -14.3 and -12.6), respectively. These *C. briggsae* sequences did not recruit the complex to arrays in either *C. briggsae* or *C. elegans*, showing that at least some syntenic sites were not conserved (table 2.5).

To discover motifs that could be important for DCC recruitment in *C. briggsae*, we tested highly occupied sites identified with ChIP-seq to determine whether these were *C. briggsae rex* sites. The α-Cbr-DPY-27 peptide antibody bound weakly, so we initially identified a dozen strong peaks. All 12 recruited the DCC and were named *C. briggsae rex* sites (*Cbr-rex-01* to *Cbr-rex-12*) (table 2.5, Te-Wen Lo, unpublished data). ChIP-seq with our FLAG-tagged *Cbr-dpy-27* and *Cbr-sdc-2* strains identified hundreds of strong peaks, including each *Cbr-rex* site (figure 2.17). Only one Cel-MEX motif and one Cel-MEX-II motif with a score below ln(P) of -15 were found among the top 100 *C. briggsae* DCC-binding sites, and neither was in a defined *Cbr-rex* site (tables 2.17, F.3). Instead, two *C. briggsae* motifs were identified among *Cbr-rex* sequences (figures 2.25, 2.26, and table F.3). Cbr-MEX (13 bp) resembles Cel-MEX (they share a strong AGGG consensus) and Cbr-MEX-II (30 bp) doesn't resemble either *C. elegans* motif. Although these motifs are not found in many peaks, they are X-enriched in *C. briggsae* and highly clustered at six of the 12 *Cbr-rex* sites (table F.3). These motifs were not X-enriched in *C. elegans* (figures 2.25a, 2.26a).

The Cbr-MEX and Cbr-MEX-II motifs were tested for their contributions to DCC binding in *C. briggsae* through *in vivo* recruitment assays (table 2.5). The *Cbr-rex-01* sequence contains four Cbr-MEX motifs (ln(P) of -15.57, -15.57, -14.63, -14.47) and one Cbr-MEX-II (ln(P) of -27.58). A DNA fragment containing the intact *Cbr-rex-01* sequence recruited strongly in *C. briggsae* (84%), but a fragment with scrambled Cbr-MEX sequences or a scrambled Cbr-MEX-II had moderate to poor recruitment (38% and 24% weakly recruiting, respectively) (figure 2.23). Scrambling all five motifs reduced DCC recruitment to

background levels (6%). Similar with *Cbr-rex-02* sequences, the intact sequence with a Cbr-MEX (ln(P) of -14.39) and a Cbr-MEX-II (ln(P) of -22.76) recruited in 100% of nuclei (figure 2.24). Scrambling Cbr-MEX alone, Cbr-MEX-II alone, or both severely reduced binding (13%, 27%, and 15% weakly recruiting, respectively). These *Cbr-rex-01* and *Cbr-rex-02* recruitment series support the functional importance of both Cbr-MEX and Cbr-MEX-II in *C. briggsae* DCC recruitment.

As in *C. elegans*, X-enriched motifs are not found at all *C. briggsae rex* sites. Five of the twelve *C. briggsae rex* sites lack strong *C. elegans* or *C. briggsae* motifs, including *Cbr-rex-08*. The *Cbr-rex-08* sequence recruits the *C. briggsae* DCC to extrachromosomal arrays, but does not recruit the *C. elegans* DCC *in vivo*. Thus some of the divergence we see in DCC-binding specificity may be caused by other motifs yet to be identified.

C. nigoni and *C. briggsae* use similar DCC binding motifs

The divergence in DCC recruitment motifs between *C. elegans* and *C. briggsae* (diverged 15-30 MYA) raised the question of when and how often changes in DCC recruitment occurred. We identified DCC binding sites in the very close *C. briggsae* relative, *C. nigoni* to determine whether binding patterns have diverged over such a short timescale. The strongest *C. briggsae* and *C. nigoni* peaks were conserved at homologous loci. Also supporting the conservation of DNA sequence specificity of the *C. nigoni* DCC, motif searches among strong *C. nigoni* peak sequences identified motifs very similar to Cbr-MEX and Cbr-MEX-II, called Cni-MEX (15 bp) and Cni-MEX-II (28 bp). Both of these were X-enriched in *C. nigoni* and enriched in peaks in *C. nigoni* and *C. briggsae* (figures 2.27, 2.28). The Cni-MEX motif is also peak-enriched in *C. elegans*, but may not contribute to DCC recruitment in *C. elegans* since the peak sequences that contain Cni-MEX also contain strong, non-overlapping *C. elegans* motifs that could recruit the *C. elegans* DCC alone. Similar to *C. briggsae*, motif clustering was observed at 7 of the top 12 *C. nigoni* binding sites (table F.5). Also, Cel-MEX and Cel-MEX-II were not X-enriched nor generally DCC-bound in *C. nigoni* (figures 2.20, 2.21). Unfortunately, we were unable to generate array-carrying transgenic lines in *C. nigoni*, so recruitment ability was not tested.

Although *C. briggsae* and *C. nigoni* motifs differ from *C. elegans* motifs, they might be similar enough to contribute to *C. elegans* DCC recruitment when many copies of the rare[1] sequences with dense motif clusters are present, and *vice versa*. *Cbr-rex-04*, which contains four distinct motifs (Cbr-MEX -13.8, Cbr-MEX-II -19.09, Cni-MEX -15.45, and Cni-MEX-II -16.3), and *C. nigoni* peak 11 (homologous to *Cbr-rex-07*), which contains six distinct motifs (including Cni-MEX -18.25 and -18.86 and Cni-MEX-II -22.87), each recruit the *C. elegans* DCC to high copy number arrays (tables 2.5, F.3, F.5). *C. nigoni* peak 11

[1]There are only six *C. briggsae* DCC-binding sites with four or more non-overlapping *C. briggsae* or *C. nigoni* motifs.

also has a weak match to the *C. elegans* MEX-II motif (-14.19), however the weak MEX-II score alone would not be predictive of binding in *C. elegans*. In the reverse direction, *C. elegans rex-32* sequences recruit the *C. briggsae* DCC to an array (table 2.5). *Cel-rex-32* contains six distinct motifs, 4 of which are strong (Cel-MEX of scores -17.65, -18.97, -18.97 and Cel-MEX-II of scores -21.89, -12.4, -12.15) (table F.1).

DCC binding motifs in *C. tropicalis* differ from motifs in the other species, indicating rapid divergence of X-chromosome DCC recruitment

Since we see divergence in DCC-binding DNA sequence motifs between *C. elegans* and *C. briggsae*, but conservation between the very close relatives, *C. briggsae* and *C. nigoni*, we next sought to identify DCC-binding DNA sequence motifs in *C. tropicalis* to learn more about when and how this change occurred (figure 2.29). If DCC recruitment in *C. tropicalis* appeared to rely on similar DNA sequence motifs as *C. elegans*, it would indicate that the ancestral DCC recruitment mechanism was more like *C. elegans*, and a change occurred in the *C. briggsae* lineage after the split with *C. tropicalis*. If DCC recruitment mechanisms were similar between *C. tropicalis* and *C. briggsae*, it would indicate conserved DCC recruitment mechanisms within the *C. briggsae* clade, and a change occurred before these species split or in the *C. elegans* lineage. Lastly, if DNA sequence motifs that differed from *C. elegans* and *C. briggsae* motifs were identified at strong DCC-binding sites in *C. tropicalis*, it would indicate that DCC recruitment mechanisms diverged more than once.

We show below that the mechanism of targeting the DCC to the X chromosome has diverged in *C. tropicalis* from both *C. elegans* and *C. briggsae*. We found a new motif at highly-occupied DCC binding sites in *C. tropicalis*, called Ctr-MEX (21 bp) (figure 2.30). This motif was highly enriched on the *C. tropicalis* X chromosome and among strong peaks in *C. tropicalis*, but was not X- or peak-enriched in the other species. Nine of the top 15[2] occurrences of this motif on the *C. tropicalis* X chromosome were found in sites highly occupied by the *C. tropicalis* DCC. Further supporting divergence in DCC recruitment mechanism, *C. elegans*, *C. briggsae*, and *C. nigoni* motifs do not predict DCC binding in *C. tropicalis*. Cel-MEX and Cel-MEX-II were not enriched on *C. tropicalis* X-chromosome contigs (figures 2.20, 2.21). High scoring Cel-MEX and Cel-MEX-II motifs (ln(P) $\leq$ -15) on the *C. tropicalis* X-chromosome contigs were not bound or were very weakly bound by the *C. tropicalis* DCC, with only one exception in a site that also contains a Ctr-MEX (table F.6). Although *C. briggsae* motifs were X-enriched in *C. tropicalis*, they were not enriched at strong DCC binding sites (figures 2.25, 2.26). Cni-MEX and Cni-MEX-II were neither X-enriched nor peak-enriched in *C. tropicalis* (figures 2.27, 2.28, and table F.6).

[2]The top 15 Ctr-MEX scores range from ln(P) of -25.68 to -17.91. Motifs of scores -22.66, -20.58, -20.02, -19.03, and -17.98 were unbound. A motif with score -20.76 was weakly bound.

In addition to observed differences in DNA sequence motifs at binding sites, we also found that highly occupied DCC-binding sites were not often conserved at homologous sites across *C. elegans*, *C. briggsae*, and *C. tropicalis*. The *Cbr-rex-01* site is an interesting exception. This site is homologous to *C. elegans rex-34*, *C. nigoni* peak 1, and *C. tropicalis* peak 20. The peak spans an exon in *C. elegans* C41A3.1, a gene encoding a fatty acid synthase. We have already shown that the five *C. briggsae* motifs found at *Cbr-rex-01* contribute to DCC recruitment in *C. briggsae* (figure 2.23). We also found that, like *Cbr-rex-08* and *C. nigoni* peak 11, the motif-dense *C. nigoni* peak 1 sequence is able to recruit the *C. elegans* DCC (table 2.5). The *C. elegans rex-34* sequence contains four *C. elegans* motifs and the *C. tropicalis* sequence only has one weak Cni-MEX-II motif (ln(P) of -13.54). This peak spans an exon, so we were able to align the sequences precisely (appendix D). We can see at this locus that these motifs are related to each other, and we can see the precise changes that occured to convert one motif to another or to create or destroy a motif. This site is under increased constraint as a functional gene, yet species-specific motifs were able to arise.

A common motif identified among *C. elegans*, *C. briggsae*, and *C. tropicalis* peaks, but not *rex* sites, may be predictive of *dox* sites

While searching for potential *rex* motifs, we found a strikingly similar motif in the top 600 DCC peaks in *C. elegans*, *C. briggsae*, and *C. tropicalis*. Although the motif was not X-enriched, the similarity across species warranted a closer look. We averaged these nearly-identical motifs and called the resulting motif "top 600". The "top 600" motif is enriched in the top 200 FLAG::DPY-27 peak sequences in *C. elegans* and *C. tropicalis* (figure 2.31a). It is also found at many lowly occupied binding sites in *C. elegans*, *C. briggsae*, *C. nigoni*, and *C. tropicalis* (figure 2.31b-c). This motif is rarely found in *rex* sites and was common among known *C. elegans dox* sites (tables F.1, F.3).

The "top 600" motif is not predictive of binding because many occurrences of this motif are not bound, but it may help predict which peak sequences are *dox* sites. In *C. elegans*, *dox* sites are often found in the promoters of highly expressed genes, while *rex* sites are more often found in intergenic regions. However, the distributions of *C. nigoni* and *C. tropicalis* peaks relative to promoters, coding, 3' UTRs, or intergenic regions are very different from peak distributions in *C. elegans* and *C. briggsae* (figure 2.32). *C. nigoni* is less gene-dense, and has far more peaks in intergenic regions than the other species. *C. tropicalis* is more gene-rich, and more peaks are found in coding regions in this species. Both species have fewer promoter peaks than *C. elegans* or *C. briggsae*, so the "top 600" motif may be a better predictor for *dox* sites than proximity to genes in these species, where the *C. elegans* pattern doesn't hold. Functional recruitment assays described below will determine whether DCC peaks with the "top 600" motif are *rex* or *dox* sites.

2.4 Discussion

Condensin-driven dosage compensation is conserved in the *C. briggsae* clade

Dosage compensation mechanisms can arise by co-opting existing cellular molecules. In *C. elegans*, condensin I was co-opted to restructure the X chromosomes specifically in hermaphrodites to balance X-chromosome expression between the sexes. Here we showed that species in the *C. briggsae* clade also use the condensin I^{DC} component DPY-27 for dosage compensation, indicating that co-option occurred before the split between *C. elegans* and *C. briggsae*.

Duplication of the condensin I component *smc-4* may be the key event that allowed condensin-driven dosage compensation to evolve. When a duplication occurs, the new paralog may evolve new functions without disrupting the original gene's function in a process called neofunctionalization. The protein-protein interactions that formed between SMC-4 paralog, DPY-27, and sex-specific condensin loaders could have allowed the highly conserved condensin I subunits to be recruited to the X chromosome as part of the DCC for dosage compensation.

We showed that DPY-27 is functionally conserved in the *C. briggsae* clade as is SDC-2. *C. elegans* SDC-2 is the key component that triggers dosage compensation and hermaphrodite sexual differentiation in XX animals and plays a major role in DCC recruitment to *rex* sites. We showed that DPY-27 and SDC-2 orthologs colocalize at discrete sites on the X chromosomes of XX animals in the *C. briggsae* clade as they do in *C. elegans*, and mutations that disrupt their function cause phenotypes consistent with dosage compensation defects. Together, these data indicate that this condensin-driven form of dosage compensation is at least 15-30 million years old.

The distant relative *Pristionchus pacificus* may have an unrelated dosage compensation mechanism

If duplication of *smc-4* was a critical event that made condensin co-option possible, then this mechanism of dosage compensation might not be present in species that diverged before *dpy-27* arose. From the currently available *Pristionchus pacificus* genome sequence, it appears that *dpy-27* arose after these species split (200-300 MYA [48]). Two paralogs of *smc-4* were identified in *Pristionchus pacificus*, but these arose from separate *smc-4* duplication events that occurred in the *Pristionchus* lineage and not the common ancestor of *Pristionchus* and *Caenorhabditis* (Christian Roedelsperger, personal communication). Also, *Pristionchus pacificus* lacks a clear SDC-2 homolog (Christian Roedelsperger, personal communication). The current best match to SDC-2 in *P. pacificus* only aligns along 9 percent of the length (with BLAST). Thus, any X-chromosome-wide mechanism of dosage compen-

sation that may exist in *Pristionchus* could be unrelated to the current *C. elegans* and *C. briggsae* mechanism, and the *Caenorhabditis* mechanism likely evolved since the split from *Pristionchus*.

Although *P. pacificus* lacks SDC-2 and has SMC-4 paralogs that are distinct from DPY-27, it appears that some form of dosage compensation occurs in *P. pacificus* today. The *P. pacificus* X chromosome was identified and found to be largely homologous to the *Caenorhabditis* X chromosome, with few translocations, consistent with an ancient origin [29]. The number of X-chromosome genes with sex-biased expression[3] in *P. pacificus* appears to be lower than the number of X-chromosome sex-biased genes in *Caenorhabditis* species [29], arguing against less-extensive dosage compensation in *Pristionchus*. More X-linked genes would be expected to have higher expression in XX- compared to XO- animals if dosage compensation were absent.

Given that the nematode X chromosome is ancient and that the *Pristionchus* mechanism appears to have different origins, the current *Caenorhabditis* mechanism probably didn't arise at the same time as the sex chromosomes evolved. It is possible that an ancestral dosage compensation mechanism was replaced with the condensin I^{DC}-containing DCC in *Caenorhabditis*. It is also possible that the ancestral species survived with only a partial dosage compensation mechanism or gene-by-gene dosage compensation rather than a complete, chromosome-wide mechanism. An important caveat is that we do not know when the Y chromosome was lost completely. If the last common ancestor of *Pristionchus* and *Caenorhabditis* had a Y chromosome, then X-chromosome genes that were also on the Y chromosome would not require dosage compensation in the ancestor. Investigation of the *P. pacificus* dosage compensation mechanism could shed light on the origin of nematode dosage compensation or not, depending on how dosage compensation has changed in the *Pristionchus* lineage over the past 200-300 million years.

DNA sequence motifs at *rex* sites

Key DCC subunits are conserved across *C. elegans*, *C. briggsae*, *C. nigoni*, and *C. tropicalis*, however the DNA sequence-specificity that drives recruitment of the DCC to the X chromosome has diverged. The *C. elegans* motifs, Cel-MEX and Cel-MEX-II, were shown to be critical for DCC recruitment to many *C. elegans rex* sites ([19, 20, 7], Will Kruesi, unpublished data), but these are clearly not important for DCC binding in species in the *C. briggsae* clade. These motifs are not X-enriched in *C. briggsae*, *C. nigoni*, or *C. tropicalis*, so they could not be used to distinguish the X chromosome from autosomes in these species. Also, Cel-MEX and Cel-MEX-II motifs that are found on X chromosomes are generally not

[3]Different expression levels in XX- compared to XO- animals could be caused by dosage compensation or sex determination. The different sexes may differ in optimal expression levels.

bound by the DCC in these species.

Functional recruitment data confirmed that Cel-MEX and Cel-MEX-II are not able to drive recruitment in *C. briggsae* as they do in *C. elegans*. Sequences on the *C. briggsae* X chromosome with strong Cel-MEX motifs that were not DCC-bound in ChIP experiments were also unable to recruit the *C. briggsae* DCC to high copy number extrachromosomal arrays. We also asked whether *C. elegans rex* site sequences were able to recruit the *C. briggsae* DCC. The four classes of *C. elegans rex* sites contain 1) a Cel-MEX motif, 2) a Cel-MEX-II motif, 3) a cluster of Cel-MEX, Cel-MEX-II, or both motifs, or 4) no known X-enriched motifs. Of the five *C. elegans rex* sites tested, four did not recruit the *C. briggsae* DCC (one with a single Cel-MEX, two with clusters of motifs, and one with no defined motifs). Only *Cel-rex-32*, which contains a dense cluster high-scoring motifs (three Cel-MEX and one Cel-MEX-II within 500 bp), recruited the *C. briggsae* DCC to high copy number extrachromosomal arrays. More functional recruitment data would strengthen our understanding of *C. briggsae* DCC sequence-specificity. In particular, we plan to test two *C. elegans rex* sequences that have only a single strong Cel-MEX motif, *Cel-rex-36* and *Cel-rex-37*, for ability to recruit the *C. briggsae* DCC.

We identified new motifs that contribute to DCC binding in *C. briggsae*. Cbr-MEX and Cbr-MEX-II are enriched on the X-chromosomes and in DCC-binding sites in *C. briggsae* and *C. nigoni*, but not in *C. elegans*, and were shown to be important for recruitment to *Cbr-rex-01* and *Cbr-rex-02*. Similar to *C. elegans*, Cbr-MEX, Cbr-MEX-II, or both motifs are found in dense clusters at some strong *Cbr-rex* sites. However, Cbr-MEX and Cbr-MEX-II are more rare and fewer DCC-binding sites have these motifs in *C. briggsae* and *C. nigoni* compared to DCC-binding sites with Cel-MEX and Cel-MEX-II in *C. elegans*. We did not find differences in DCC recruitment mechanisms between the very close relatives *C. briggsae* and *C. nigoni*. The Cni-MEX and Cni-MEX-II motifs are very similar to Cbr-MEX and Cbr-MEX-II and were found at homologous DCC-bound loci in *C. nigoni*. We may be missing important DCC-binding motifs that drive recruitment to other *rex* sites, but it is also possible that *C. briggsae* and *C. nigoni* have fewer *rex* sites than *C. elegans*.

We found that *C. elegans rex* site sequences are not sufficient for DCC recruitment in *C. briggsae*, with one motif-dense exception, so we also asked whether the reverse is true. *C. briggsae* and *C. nigoni* DCC-binding sites were tested for their ability to recruit the *C. elegans* DCC. *Cbr-rex-08* lacks *C. elegans* and *C. briggsae* X-enriched motifs and was not able to recruit the *C. elegans* DCC, consistent with a divergent recruitment mechanism. However, rare *C. briggsae* and *C. nigoni* motif-dense sequences (from *C. briggsae* and *C. nigoni*) do recruit the *C. elegans* DCC. We plan to test the *Cbr-rex-02*, *Cbr-rex-03*, and *Cbr-rex-05* sequences that are less motif-dense to determine whether strong Cbr-MEX-II motifs can recruit the *C. elegans* DCC alone. These sites lack strong Cbr-MEX and Cni-MEX motifs that resemble Cel-MEX. Preliminary evidence shows that cross-species recruitment to *Cbr-rex-02* and *Cbr-rex-05* is patchy and weak.

C. tropicalis DCC-bound sequences differ from both *C. elegans* and *C. briggsae*, which shows that DCC recruitment evolves rapidly. The Ctr-MEX motif is only X-enriched in *C. tropicalis* and does not resemble other *rex* motifs. Also, *C. elegans*, *C. briggsae*, or *C. nigoni* X-enriched, DCC-binding motifs are generally not DCC-bound in *C. tropicalis*. Unfortunately, we were not able to generate any heritable transgenic arrays to test sequences for DCC recruitment in *C. tropicalis* or *C. nigoni*. It may be possible to test for *C. tropicalis* or *C. nigoni* DCC recruitment by IF in first-generation array-carrying progeny. If so, we may identify *Ctr-rex* and *Cni-rex* sites and investigate motifs for their contribution to DCC binding this way. If not, DNA sequences could be inserted with genome editing, then tested for DCC binding with ChIP-qPCR or ChIP-seq. Since, the genome editing approach is labor-intensive and low throughput, we also consider other methods for future studies. A higher throughput method is to use an *in vitro* assay, in which *C. elegans* FLAG::SDC-2 from a crude worm extract binds to biotinylated DNA probe sequences with sequence-specificity (developed by K. Brejc). We now have FLAG::Ctr-SDC-2 and FLAG::Cbr-SDC-2 strains which would allow us to investigate DCC sequence-specificity this way in *C. tropicalis* and *C. briggsae*.

A new DNA sequence motif was found at DCC peaks in all four species, but does not correlate with *rex* sites

Broadening our search for *rex* site motifs unexpectedly led us to identify a potential *dox* site motif. The "top 600" motif was identified among the top 600 *C. elegans*, *C. briggsae*, and *C. tropicalis* DCC peaks. It is also present in many *C. nigoni* peaks. This motif is not X-enriched, so it cannot be important for distinguishing X from autosomes in any of these species. Also, it was present in only three *C. elegans rex* or predicted *rex* sites and none of the *C. briggsae rex* sites (tables F.1, F.3). In contrast, this motif was found in 16 *C. elegans dox* sites half of which are in promoters. This is the first motif we have identified that is associated with *dox* sites.

C. elegans dox sites are often found at highly-expressed gene promoters, which led to the hypothesis that open chromatin, polymerase, or some other factor at promoters may facilitate DCC spreading from *rex* to *dox* sites in *cis* [49, 20]. Improved annotation of transcription start sites (TSS) and better ChIP resolution (ChIP-seq rather than ChIP followed by microarray) determined that binding occurs just upstream of the TSS [50]. However, promoters are not predictive of *dox* sites, and not all DCC-bound genes are dosage compensated [20, 50], so it appears to be more complicated than DCC spreading to act on the promoters with the highest gene expression. Also fewer DCC-bound sites are found in promoters among the top 500 peaks in *C. nigoni* (88 promoter peaks) and *C. tropicalis* (181) compared to *C. elegans* (297) or *C. briggsae* (334) (figure 2.32), so *dox* sites are unlikely to correlate as well with promoters in these species. In other words, it is highly unlikely that most of the non-

promoter sites are *rex* sites in these species, especially considering that potential X-enriched *rex* motifs were only found in tens rather than hundreds of sites. While DCC spreading to *dox* sites could be facilitated by open chromatin or transcription-associated factors, it seems that other mechanisms must be in effect.

Presence of a "top 600" motif may be more predictive of a *dox* site peak than the peak's proximity to a TSS. This motif is more often DCC-bound in *C. elegans* when it is found in promoters, where *dox* sites are often found. However, gene density and DCC peak distribution differ among species in the *C. briggsae* clade. The "top 600" motif is often found in intergenic DCC peaks in *C. nigoni* and in peaks in coding regions in *C. tropicalis*. In *C. nigoni* and *C. tropicalis*, the overall distribution of DCC-binding sites in promoters, coding regions, 3' untranslated regions, and intergenic regions appears to be random compared to the overall proportion of the genome that lies in these regions (figure 2.32). It may be more accurate to predict DCC peaks that are *dox* sites by presence of the "top 600" motif rather than location in a promoter, especially in these species. To test this prediction, we plan to determine whether highly occupied *C. briggsae* DCC-binding sites that contain the "top 600" motif are able to recruit the DCC in *C. briggsae*.

A note regarding genome shrinkage

Reproductive mode played a large role in shaping these species' genomes. Hermaphroditism evolved three separate times in *Caenorhabditis*, in *C. elegans*, *C. briggsae*, and *C. tropicalis* [22]. These androdioecious species have 20 to 40% smaller genomes than their closest obligate outcrossing relatives, due to a phenomenon called genome shrinkage [51, 52]. These species each gained the ability to self-fertilize, so inbreeding became far more common and mating became less frequent. The effective population size decreased by half, resulting in a decrease in selective power [52]. Hermaphroditism also causes genome shrinkage through segregation bias. In male meiosis, autosomes with deletions are more likely to segregate with the X chromosome [53], which likely contributed to gene loss in androdioecious species [51, 52]. Also, androdioecious species have reduced sexual selection that has had large effects on mating behaviors and phenotypes [52].

Divergence of DCC binding sites and implications for X chromosome structure

The *C. elegans* DCC actively restructures the X chromosome by bringing *rex* sites together [13]. Several of the *C. elegans rex* sites that interact most frequently coincide with boundaries between topologically-associated domains (TADs). TADs are DNA domains that have frequent interactions within the domain, but are insulated from sequences outside of the domain. In DCC mutants, the *rex* sites interact far less frequently and many TAD boundaries at these sites are diminished. The dosage compensated X chromosome has distinct TADs that are about 1 Mb long, and the non-compensated X and the autosomes have fewer TADs

and weaker TAD boundaries [13]. In contrast to the highly structured, dosage compensated X chromosomes in nematodes, mammalian X chromosome inactivation (XCI) results in two mega-domains on the inactive X chromosome, and very few TAD-like structures, compared to a highly structured active X chromosome [54, 55, 56].

Throughout the genome, mammalian TAD structure facilitates interactions between promoters and enhancers, with TAD boundaries often occuring at CTCF binding sites [54]. TAD structure appears to reinforce expression of the long noncoding RNA, Xist, on the inactive X chromosome and its repressor, Tsix, on the active X chromosome [57, 58]. TAD boundaries insulate sequences within domains from those without, and genes within TADs are thought to be co-regulated [59], however, the *C. elegans* DCC appears to act at a distance, with no clear co-regulated chromosomal domains ([20], [50], and Bayly Wheeler, unpublished data). Also, long-range enhancers are not common in *C. elegans* [60, 61], which suggests that insulation of specific enhancer-promoter pairs may not be a driving mechanism in nematode dosage compensation. We don't yet know what role the DCC function of creating or strengthening TAD boundaries plays in regulating gene expression

While one could hypothesize that there are so many *C. elegans rex* sites and Cel-MEX and Cel-MEX-II motifs on X because they are important, it seems that *rex* sites are redundant. *C. elegans* apparently can spare six of the *rex* sites that influence TAD boundaries and still be viable. Recently Qian Bian and Erika Anderson deleted six *C. elegans rex* sites at DCC-dependent TAD boundaries, which resulted in no apparent loss of viability (unpublished data). They will next determine whether the structure of the X chromosome lacking these six *rex* sites is more similar to the X chromosome of the DCC mutant or whether other *rex* sites can compensate by forming new TAD boundaries to restructure the X chromosome in this mutant. More extensive studies will determine whether these mutants have subtle dosage compensation defects. The strain with six *rex* sites removed supports the hypothesis that the precise locations of *rex* sites can change and that dosage compensation is a robust process in *C. elegans*. It may be that all that is required for chromosome-wide dosage compensation are a few *rex* sites on X and a good mechanism to spread the DCC to *dox* sites.

We found that the precise location of strong *rex* sites has diverged in the *C. briggsae* clade. With one exception, known *C. briggsae rex* sites are not syntenic with *C. elegans rex* sites or strong *C. tropicalis* DCC-binding sites, which may affect the X-chromosome structure in these species. Also, the total *rex* site number in species across the *C. briggsae* clade is yet unknown. In *C. briggsae* and *C. nigoni* especially, *C. briggsae rex* motifs are only found at a few binding sites. It remains to be seen whether we are missing important recruitment motifs at yet-unknown *rex* sites. An alternative hypothesis is that the few, strong *C. briggsae* and *C. nigoni rex* sites may play a stronger role in these species.

On selection

To put nematode dosage compensation evolution in context, we compare the conservation of key DCC components and rapid divergence of their DNA binding sites with the remarkable conservation of the DNA-binding motifs of transcription factors. Transcription factors have ancient origins and conserved function; key players in many gene regulatory networks are conserved across species that are more than half a billion years apart [62, 63, 64, 65]. Most transcription factors have conserved DNA-binding specificity, but binding motifs are easily gained and lost within *cis*-regulatory modules, so the binding sites might be in slightly different locations [66, 67]. Rare changes in DNA sequence-specificity generally occur in transcription factor families, where duplications allow one paralog to diverge without affecting the function of the other [67]. The DNA sequence changes observed among transcription factor families are often subtle, since the protein folds that directly bind DNA have constrained binding capabilities [68, 67]. In contrast, nematode dosage compensation arose more recently, and we see far greater divergence in DNA-binding specificity in species less than 30 million years apart.

We were surprised to see so much divergence in the DNA sequence-specificity of the DCC in comparison to transcription factors, however the DCC is very different from a transcription factor. First, transcription factors are often involved in multiple developmental processes, and epistatic interactions can constrain their evolution [69]. SDC-2 and SDC-3 are DCC components important for sequence-specific recruitment to *rex* sites for dosage compensation and recruitment to different sequences at the autosomal *her-1* gene to inhibit male development, respectively. Thus these genes are not pleiotropic. Second, transcription factors often have a well-characterized protein fold that interacts with DNA, but SDC-2 does not. SDC-3 only has two zinc finger domains, which are not sufficient on their own for sequence-specificity. Third, transcription factors need to bind near the genes they regulate, but the DCC acts at a distance to regulate genes across the entire chromosome. Also, *rex* site redundancy may allow gain and loss of *rex* sites without affecting function. If DCC-binding sites are not constrained to specific loci and the protein(s) that bind DNA are not constrained by other developmental or tissue-specific functions, it may be relatively easy to change DCC recruitment mechanisms, which would explain the rapid divergence we see.

C. elegans, *C. briggsae*, and *C. nigoni* each have two different *rex* site motifs, which could help explain how the DCC and the X chromosome coevolved. The DCC is recruited to the X chromosome in at least two different ways, so one motif may be free to diverge without disrupting dosage compensation, which is essential in XX animals. As the DNA-sequence specificity for one motif changed, instances of the new motif could be preferentially maintained on the X chromosome and lost on the autosomes until the new motif was also enriched on the X chromosome. This could also be true in *C. tropicalis* if it has another *rex* site motif not yet identified, or if the interactions with a secondary motif are currently "broken" in this species. "Intermediate" stages, where one motif or the other has low binding affinity and/or

X-enrichment could either be less than optimal for dosage compensation or one motif could compensate for the loss of binding interactions with the other. Some transcription factors bind at primary and secondary DNA motifs, but binding-specificity is generally conserved at both [67]. Transcription factors can also bind to *cis*-regulatory sites indirectly, through interactions with other transcription factors [69]. Cel-MEX, Cbr-MEX, and Cni-MEX differ from each other in subtle ways, so it is likely that the same component binds to this motif in each species. Cbr-MEX-II and Cni-MEX-II are nearly identical, but differ from Cel-MEX-II. It is unclear whether a conserved protein domain interacts with both Cel-MEX-II and Cbr-MEX-II.

X-incompatibility was observed in interspecies hybrids of the closely related species *C. briggsae* and *C. nigoni*, but this species barrier appears to be unrelated to dosage compensation [70]. The X-incompatibility primarily caused hybrid male lethality that was suppressed by a mutation in *Cbr-him-8* [70]. Also, we didn't find differences in dosage compensation between these species (the same DCC-bound DNA sequence motifs were found at homologous DCC-binding sites). Although these species share dosage compensation mechanisms, changes in DCC-binding mechanisms could contribute to other speciation events.

At this point we can only speculate on the forces that caused a change in DCC-binding site sequence-specificity in *Caenorhabditis*. Unfortunately, testing for positive or negative selection would be difficult and perhaps uninformative in nematodes at this time. First, the specific DCC protein(s) and protein domains that interface with X-chromosome binding sites have not been determined. The prime candidate for DNA sequence recognition (because it binds first and independently in *C. elegans*), is SDC-2, a large protein with only a coiled-coil domain and no consistent structural predictions. To determine whether SDC-2 is under positive or negative selection, we would need to align *sdc-2* DNA sequences across many strains and multiple species. DNA sequence alignments would be difficult because the SDC-2 ortholog sequences do not align well across the entire length at the amino acid level. Also, we would need to know which regions are important for binding to make sense of the data, but it is impossible to know this without structural information or candidate DNA-binding domains. Second, it is possible that there are DCC components that are important for sequence-directed recruitment at *rex* sites that we haven't identified yet. These could be protein or RNA factors. Third, most *rex* sites are not found in coding regions, except in *C. tropicalis*, so aligning sequences may be difficult within or between species. We have only found one *rex* site that can be aligned across all four species. Together, we are unable to address the specific changes in the DCC that led to a change in *rex* site motifs at this time.

Conclusions

This study provides a striking example of rapid divergence of protein-DNA interactions. We sequenced the *C. nigoni* and *C. tropicalis* genomes and developed reverse genetic tools in these non-model species to investigate the molecular mechanism of dosage compensation.

We found that DCC recruitment mechanisms differed from *C. elegans* in *C. briggsae* and *C. nigoni*. Highly occupied DCC binding sites were found at different X-chromosome loci, and X-enriched DNA sequence motifs that are important for binding have diverged. Further divergence in *C. tropicalis* demonstrates that the sequence-specificity of DCC recruitment has diverged more than once. This work raises many questions that can now be addressed, including whether species in the *C. briggsae* clade have fewer *rex* sites than *C. elegans* and whether fewer are needed for dosage compensation. Future studies will determine how these evolving *rex* sites affect X-chromosome structure and gene expression, bringing us closer to understanding how nematode dosage compensation works.

2.5 Tables and figures

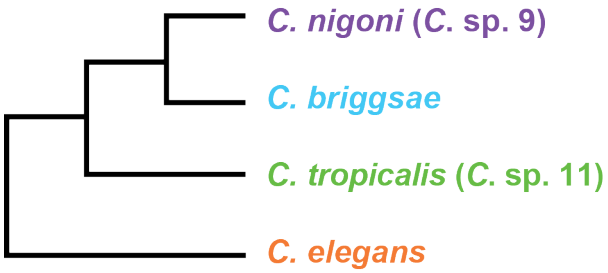


Figure 2.1: **The *Caenorhabditis* phylogeny** (adapted from [22]). *C. elegans* diverged from *C. nigoni*, *C. briggsae*, and *C. tropicalis* 15 to 30 million years ago. *C. elegans*, *C. briggsae*, and *C. tropicalis* are androdioecious species (consists of male and hermaphrodite sexes). Hermaphroditism evolved independently in each of these lineages. *C. nigoni* is gonochoristic (consists of male and female sexes), like the common ancestor and the other *Caenorhabditis* species. *C. briggsae* and *C. nigoni* are close enough relatives to produce (gonochoristic) fertile interspecies hybrids.

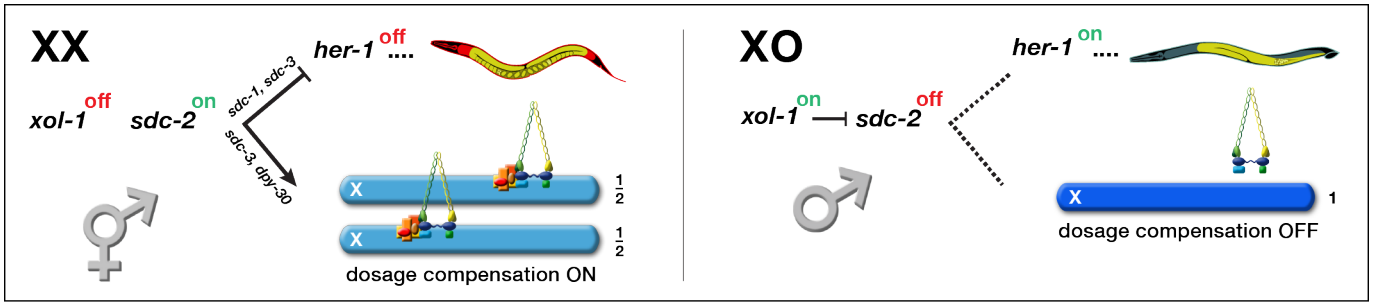


Figure 2.2: **Genetic hierarchy controlling *C. elegans* sex determination and dosage compensation.** *xol-1* triggers the male fate in XO embryos by repressing the XX-specific gene *sdc-2*, which triggers hermaphrodite fate. SDC-2 acts with SDC-1 and SDC-3, both zinc finger proteins, to induce hermaphrodite development by repressing *her-1*, a male sex determining gene. SDC-2 acts with SDC-3 and DPY-30, a subunit of the MLL/COMPASS chromatin modifying complex, to load the DCC onto X, activate dosage compensation, and reduce expression by half. *sdc-2* is the sole gene required by all DCC subunits to bind to X. Without *sdc-2*, *her-1* is on, and the DCC fails to bind X, causing masculinization and death of XX embryos. Without *xol-1*, XO embryos repress her-1 and activate dosage compensation, causing feminization and death.

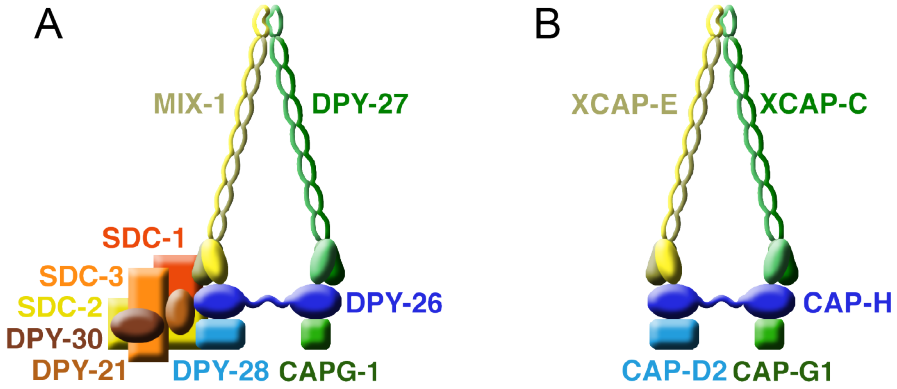


Figure 2.3: **The *C. elegans* dosage compensation complex (DCC) shares subunits with condensins.** A) The *C. elegans* DCC has 10 known protein subunits, four of which are shared with condensin I (MIX-1, CAPG-1, DPY-26, and DPY-28). DPY-27 is an SMC-4 homolog that is unique to the DCC. SDC-2 plays a critical role in loading the DCC on the X chromosomes and triggering hermaphrodite fate in *C. elegans*. SDC-2 is able to bind to *rex* sites in the absence of other components. SDC-3 and DPY-30 are also key players in DCC recruitment to the X chromosomes. These require SDC-2 for binding and are required to stabilize the complex and/or bring the complex to the X chromosome. SDC-3 also plays a key role in triggering hermaphrodite fate in *C. elegans*. DCC components DPY-21 and SDC-1 are not required for DCC recruitment at *rex* sites. A one-to-one homolog exists for each DCC component in *C. briggsae*, *C. nigoni*, and *C. tropicalis*. B) Condensins are conserved throughout eukaryotes. The frog condensin is represented here.

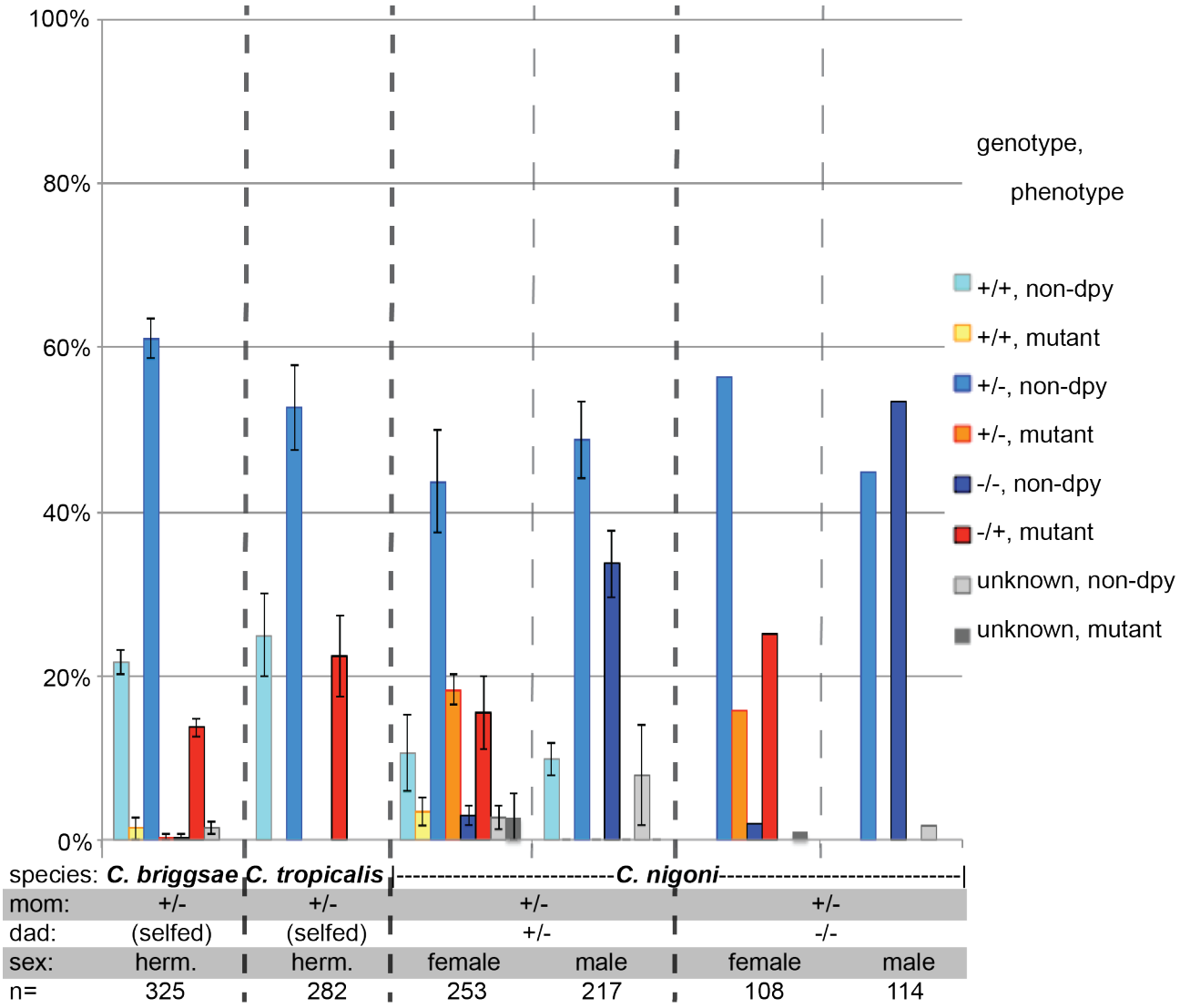


Figure 2.4: ***dpy-27* ortholog mutations cause defects consistent with conserved function in the *C. briggsae* clade** (Continued on the following page.)

Figure 2.4: ***dpy-27* ortholog mutations cause defects consistent with conserved function in the *C. briggsae* clade.**

Genotypes of *Cbr-dpy-27(y705)*/+, *Ctr-dpy-27(y677)*/+, and *Cni-dpy-27(y709)*/+ progeny were assessed by single worm PCR. The bars are color-coded as follows: non-dumpy adults were labeled in shades of blue (wild-type at the *dpy-27* locus in light blue, heterozygous in medium blue, and homozygous mutant in dark blue) and dumpy or small worms were labeled in yellow (wild-type at the *dpy-27* locus), orange (heterozygous), and red (homozygous mutant). Grey bars refer to the PCR reactions that failed (light grey were non-dumpy and dark grey were dumpy or small). Error bars represent the standard error of the mean.

Heterozygous *Cbr-dpy-27(y705)* and *Ctr-dpy-27(y677)* hermaphrodites were allowed to self-fertilize. Surviving homozygous progeny were dumpy. Heterozygotes and wild-type hermaphrodites were unaffected. Only 2% of the *C. briggsae* self-progeny were exceptional (two dumpy wild-type, one dumpy heterozygote, and 1 non-dumpy heterozygous male). As in *C. elegans*, these *dpy-27* ortholog mutations are recessive. Unlike *C. elegans dpy-27* mutants, maternal contributions do not rescue homozygous progeny from heterozygous mothers.

Cni-dpy-27(y709) heterozygotes were crossed to heterozygous and homozygous mutant males. Some heterozygous and nearly all homozygous females were dumpy or small. Ten of the 253 female progeny of heterozygous parents were dumpy or small, but scored wild-type at the *Cni-dpy-27(y709)* locus. None of the males were phenotypically mutant, consistent with a XX-specific dosage compensation defect. Many progeny died, most likely due to inbreeding depression (table 2.4). Consistent with inbreeding depression, the deaths did not correlate well with mutant genotype at the *Cni-dpy-27* locus.

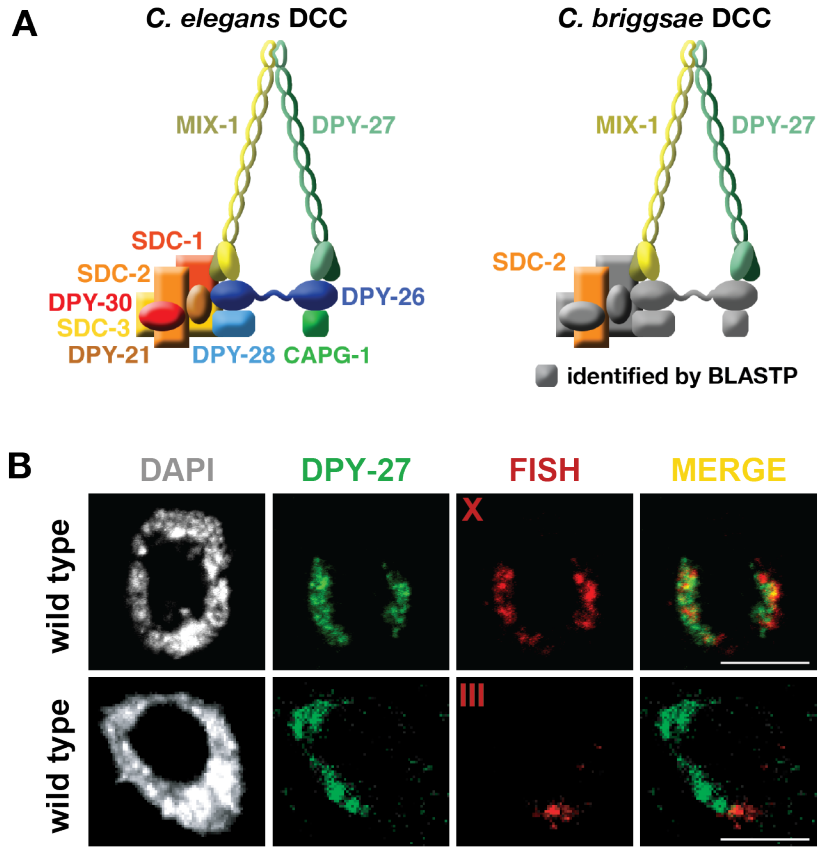


Figure 2.5: **The *C. briggsae* DCC binds to the X chromosomes.** A) One-to-one homologs exist in *C. briggsae* for each DCC gene. Functional conservation was demonstrated for *C. briggsae* DPY-27, SDC-2, and MIX-1 orthologs. B) The Cbr-DPY-27 protein colocalizes with the X-chromosomes by FISH-IF. In the top row, the X chromosomes are labeled with FISH probes, and chromosome III is labeled in the bottom row.

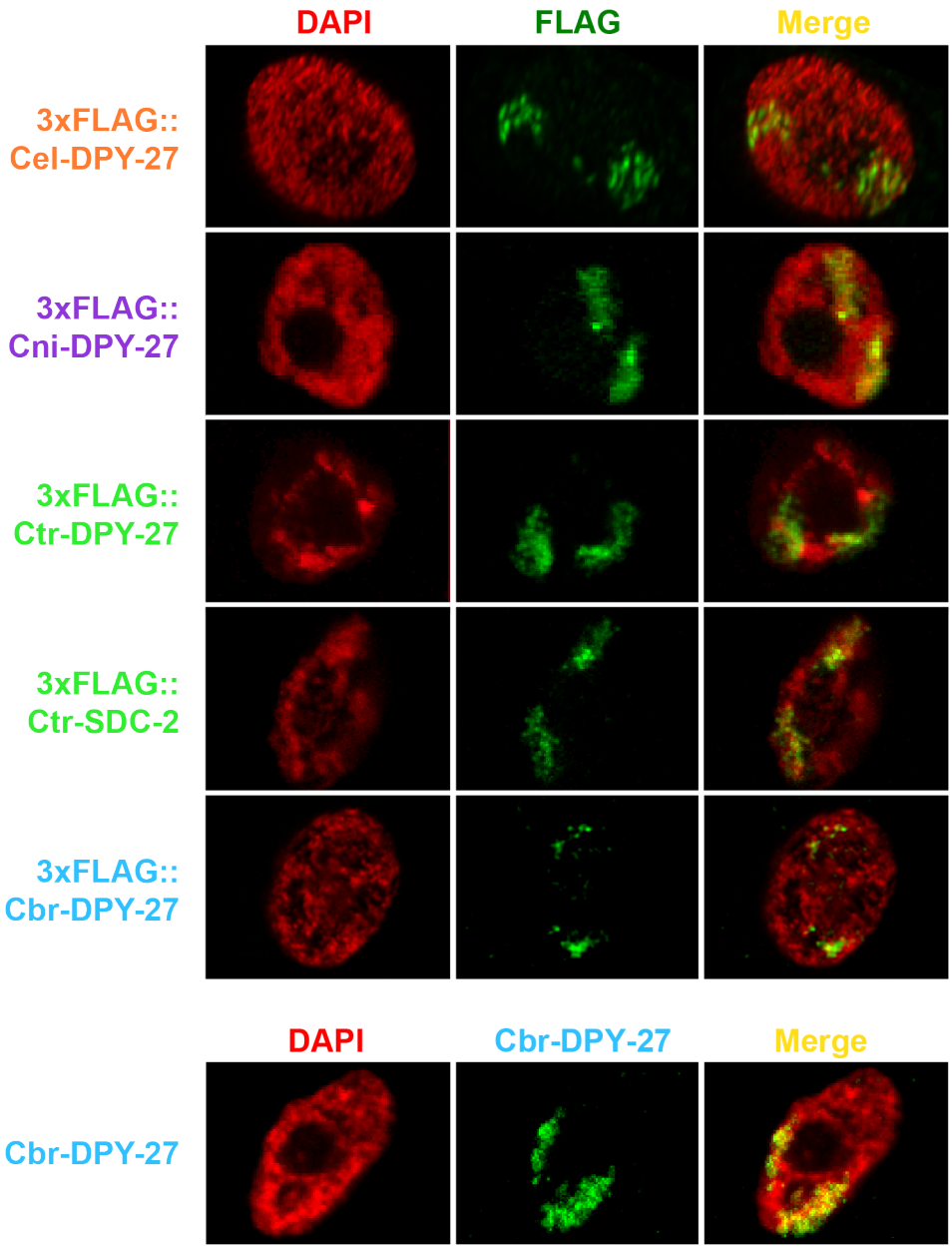


Figure 2.6: **Immunofluorescence (IF) images of FLAG-tagged strains created with genome editing are consistent with a conserved DCC.** In these confocal images, DNA was stained with DAPI and tagged proteins were stained with α-FLAG antibody. The *C. briggsae* nucleus in the bottom row was stained with α-Cbr-DPY-27 antibody. Image slices with two visible chromosomal domains were selected. ChIP-seq experiments confirmed that DPY-27 and SDC-2 orthologs bind to X chromosomes (figures 2.16, 2.17, 2.18, 2.19).

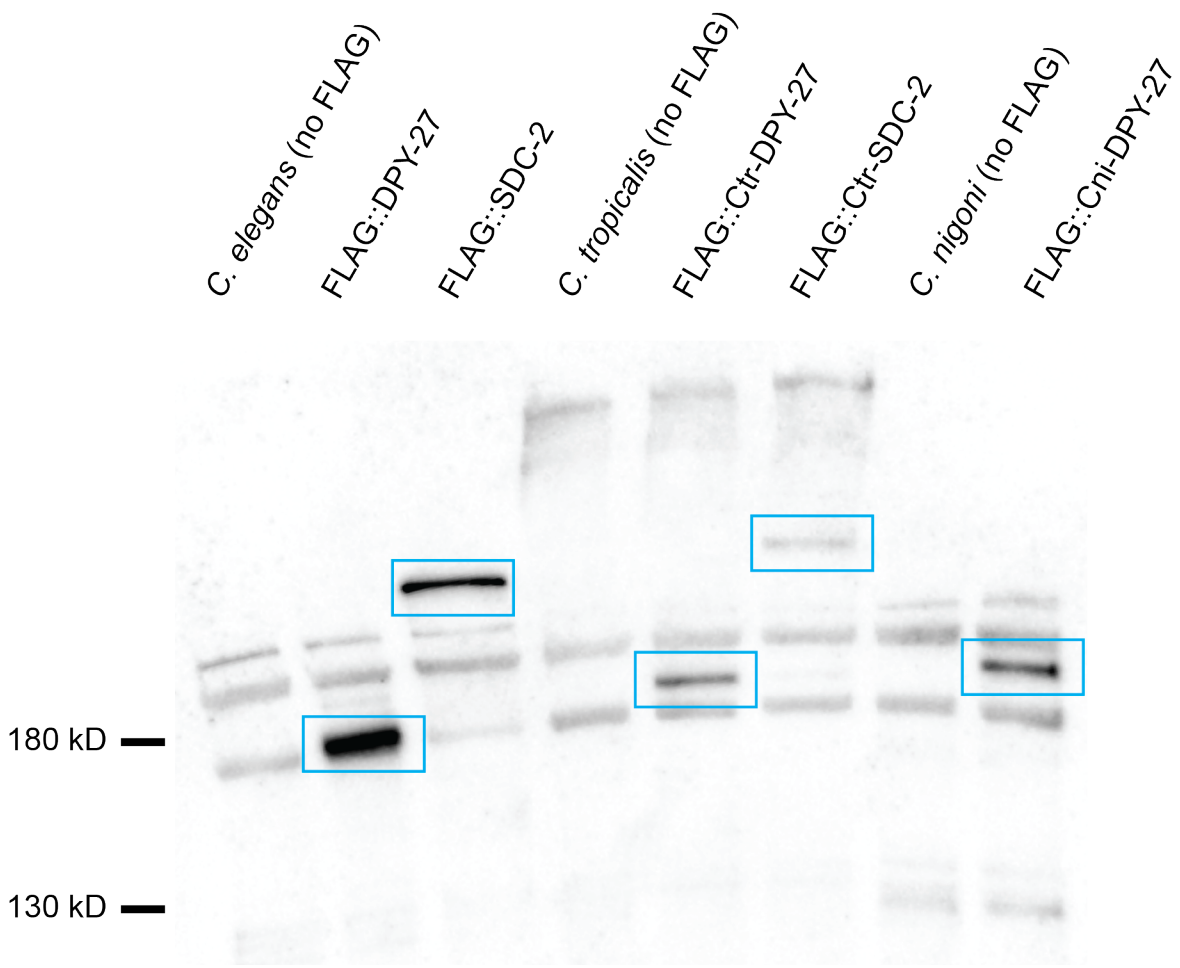


Figure 2.7: **Western blot of FLAG-tagged DCC strains created with genome editing.** *C. elegans*, *C. tropicalis*, and *C. nigoni* FLAG-tagged proteins were imaged by Western blot with α-FLAG antibody (FLAG::Cbr-DPY-27 and FLAG::Cbr-SDC-2 not shown). FLAG-specific bands are labeled with blue boxes. Sizes are consistent with full-length proteins.

A DPY-27

	C. briggsae	*C. tropicalis*	*C. elegans*
C. nigoni	92	34	38
	94	51	57
	2	17	10
C. briggsae		35	38
		53	56
		12	12
C. tropicalis			38
			54
			14

% identical
% similar
% gaps

B SDC-2

	C. briggsae	*C. tropicalis*	*C. elegans*
C. nigoni	89	27	27
	94	42	43
	1	24	24
C. briggsae		27	26
		42	43
		23	24
C. tropicalis			26
			42
			27

C XOL-1

	Cbr-XOL-1	Cni-g27926	Cni-g27847	Ctr-g5907	Ctr-g5908
C. elegans XOL-1	17	16	20	21	21
	31	30	36	34	38
	36	39	26	37	33
C. briggsae XOL-1		67	76	22	21
		70	85	42	36
		24	1	17	27
C. nigoni g27926			29	17	21
			65	31	36
			26	36	27
C. nigoni g27847				59	17
				65	31
				26	36
C. tropicalis g5907					20
					39
					25

Figure 2.8: **DPY-27, SDC-2, and XOL-1 are conserved in the *C. briggsae* clade.** Pairwise comparisons show percent amino acid identity, similarity, and gaps in A) DPY-27, B) SDC-2, and C) XOL-1 homologous sequences. For comparison, 12,155 *C. elegans*/*C. briggsae* ortholog pairs were found to be 75% identical on average with a standard deviation of 18% amino acid conservation [71]. *C. nigoni* and *C. tropicalis* have two *xol-1* homologs. In these species, both homologs are found on the X chromosome.

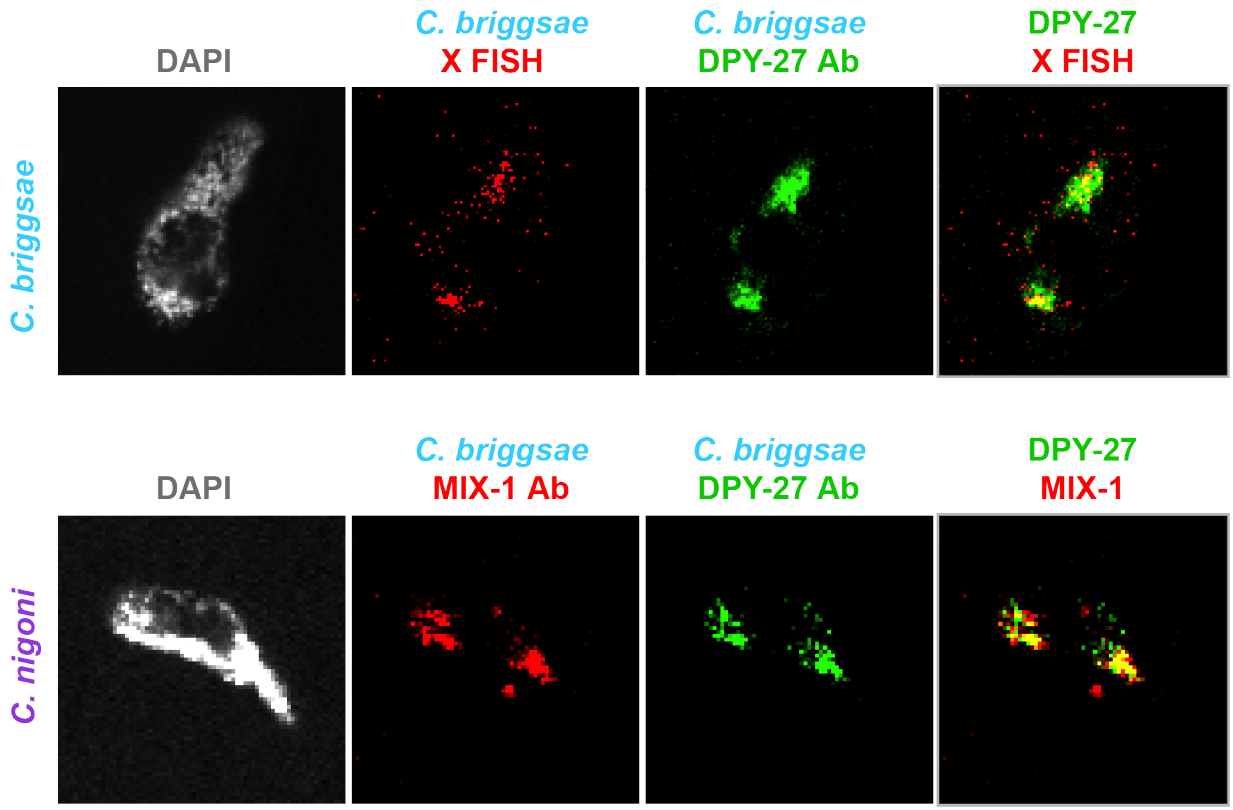


Figure 2.9: **Cbr-DPY-27 and Cbr-MIX-1 antibodies label the *C. nigoni* DCC.** In *C. briggsae*, DNA was stained with DAPI, the X chromosome was stained with X FISH probes, and the DCC was stained with α-Cbr-DPY-27 peptide antibody. The *C. briggsae* DCC colocalized with the X chromosome. In *C. nigoni*, the DNA was stained with DAPI and the DCC was stained with α-Cbr-MIX-1 and α-Cbr-DPY-27 antibodies. Cni-DPY-27 and Cni-MIX-1 colocalize, presumably on the X chromosome. (X-chromosome localization was confirmed with ChIP-seq.)

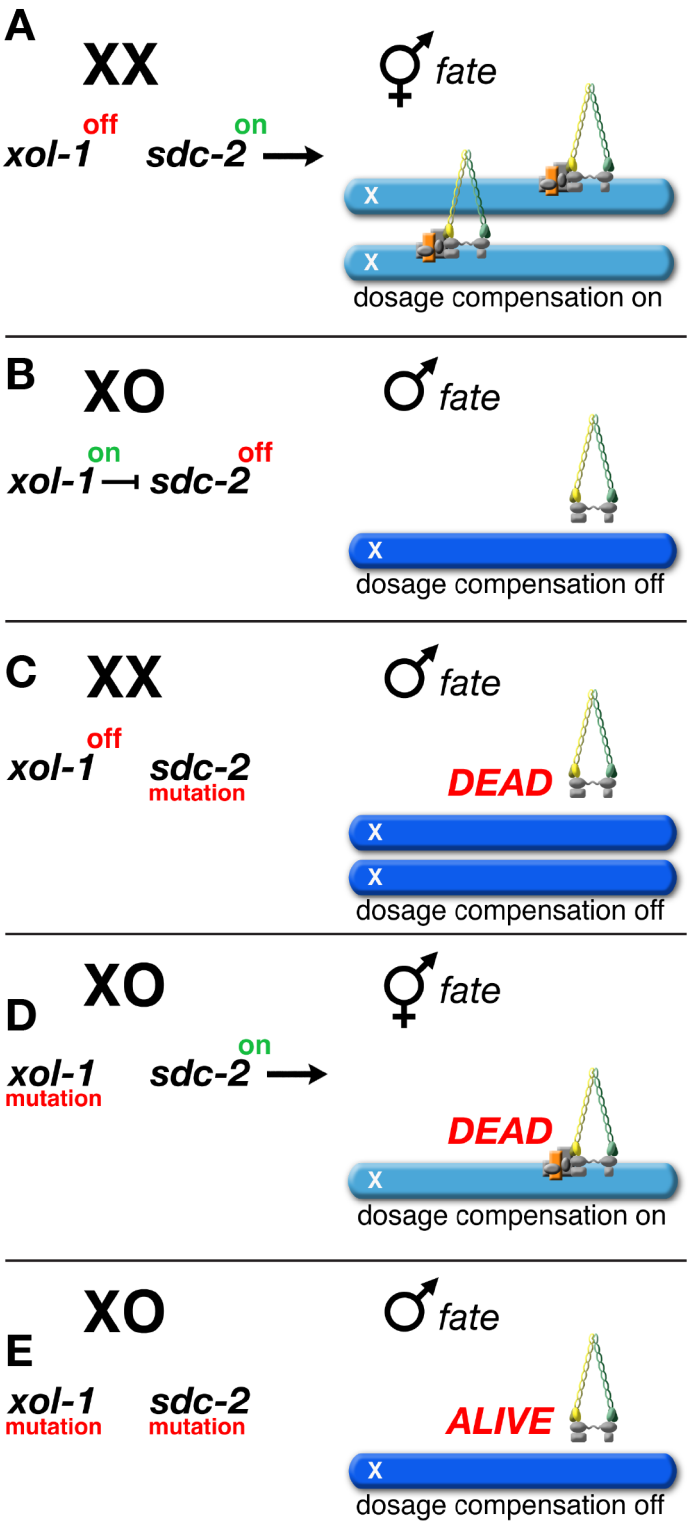


Figure 2.10: **The genetic hierarchy.** (Continued on the following page.)

Figure 2.10: **The genetic hierarchy directing sex determination and dosage compensation.**
A) In *C. elegans*, the genetic switch *xol-1* is off in XX hermaphrodites and *sdc-2* is expressed. *sdc-2* triggers loading of the DCC on the X chromosomes and hermaphrodite fates.
B) In XO males, *xol-1* is expressed and represses *sdc-2*. The DCC fails to load on the single male X chromosome and the animal develops as a male.
C) In an XX *sdc-2* mutant, the DCC fails to load, leading to embryonic lethality. Rare survivors are masculinized. XO *sdc-2* mutants are unaffected.
D) In an XO *xol-1* mutant, *sdc-2* is expressed, leading to inappropriate loading of the DCC on the single male X and XO lethality. XX animals are unaffected.
E) XO *xol-1 sdc-2* double mutants are rescued. The DCC fails to load on the X chromosome and the animals develop as males. The double mutant confirms that *sdc-2* acts downstream of *xol-1*.

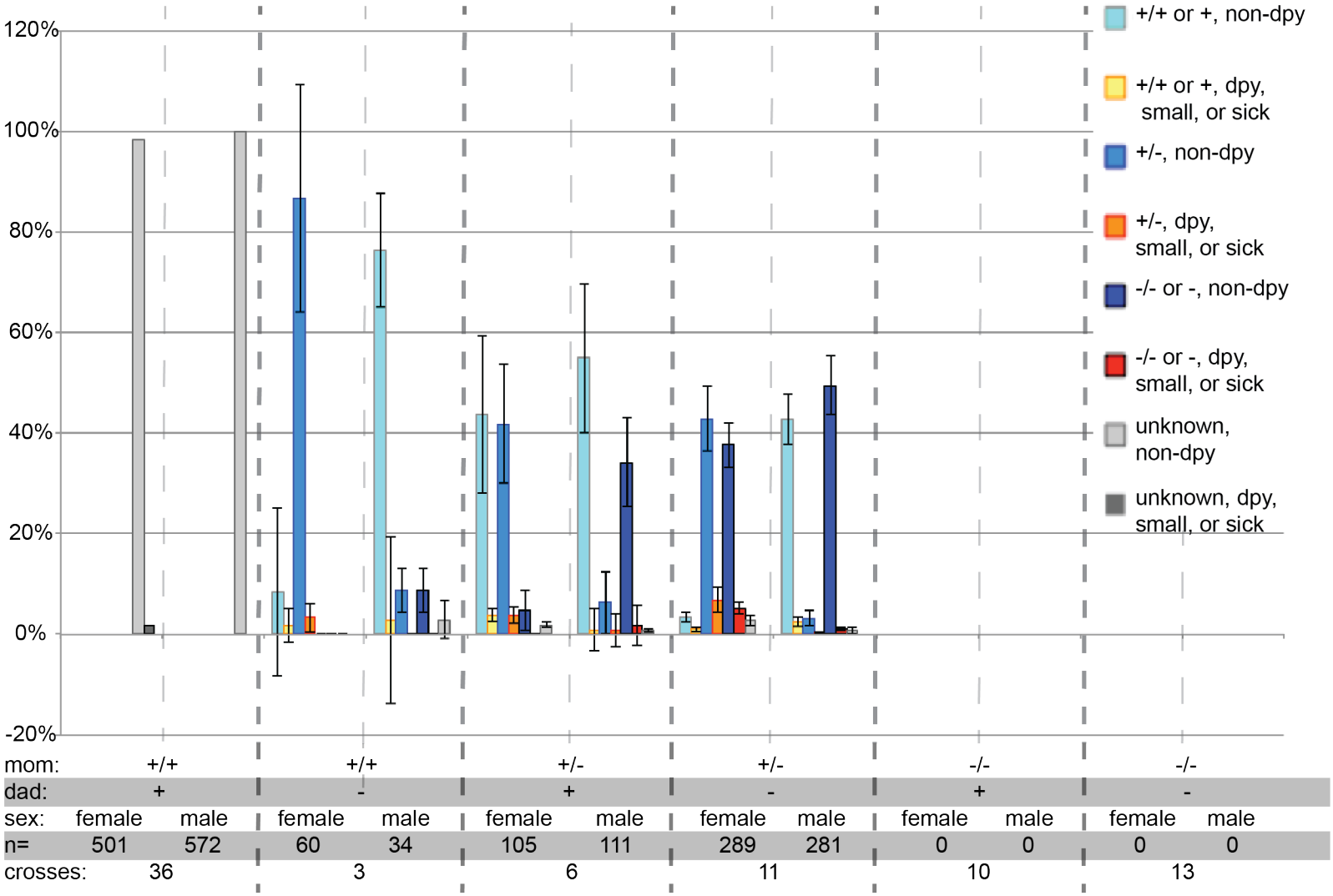


Figure 2.11: **A *Cni-sdc-2* mutation causes recessive female sterility** (Continued on the following page.)

Figure 2.11: **A *Cni-sdc-2* mutation causes recessive female sterility.**

Genotypes of *Cni-sdc-2(y516)* progeny were assessed by single worm PCR. The bars are color-coded as follows: non-dumpy adults were labeled in shades of blue (wild-type at the *dpy-27* locus in light blue, heterozygous in medium blue, and homozygous mutant in dark blue) and dumpy or small worms were labeled in yellow (wild-type at the *dpy-27* locus), orange (heterozygous), and red (homozygous mutant). Grey bars refer to the PCR reactions that failed (light grey were non-dumpy and dark grey were dumpy or small). Error bars represent the standard error of the mean.

Homozygous mutant females were sterile, but not dumpy, small or sick. Hemizygous mutant males were unaffected.

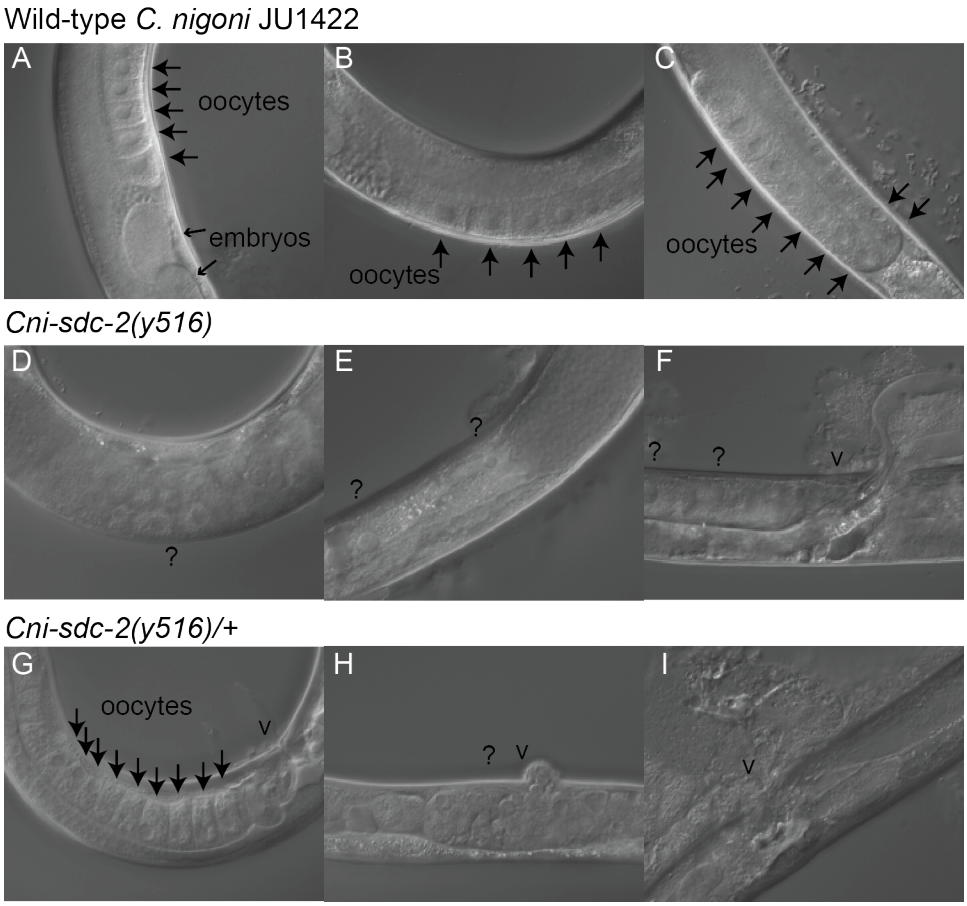


Figure 2.12: ***Cni-sdc-2(y516)* mutants have germline and gonad morphology defects.** A-C) Oocytes and sometimes embryos are visible in wild type female gonads. B-F) Homozygous *Cni-sdc-2(y516)* lack normal oocytes and have abnormal gonad morphology, including protruding gonad, consistent with observed sterility. G-I) Some heterozygous females appear wild type while others have mutant phenotypes.

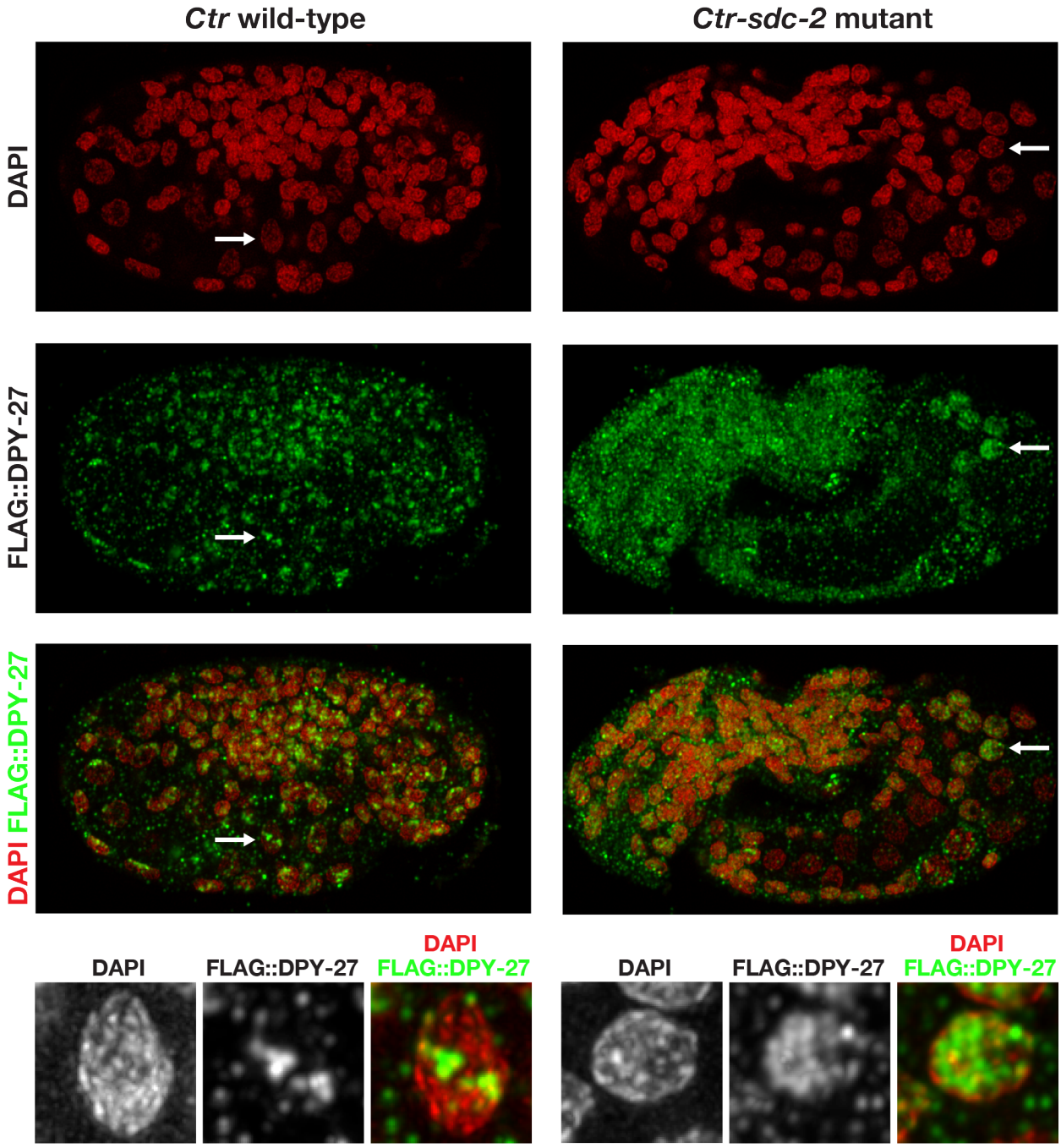


Figure 2.13: ***C. tropicalis sdc-2* mutants fail to load the DCC on the X chromosomes.** FLAG::Ctr-DPY-27 localization is consistent with X-chromosome DCC localization in the *Ctr-dpy-27(y677)* strain. In the *Ctr-dpy-27(y677)*; *Ctr-sdc-2(y719)* strain, *Ctr-sdc-2* is mutant, so the FLAG::Ctr-DPY-27 fails to load, which leads to embryonic lethality.

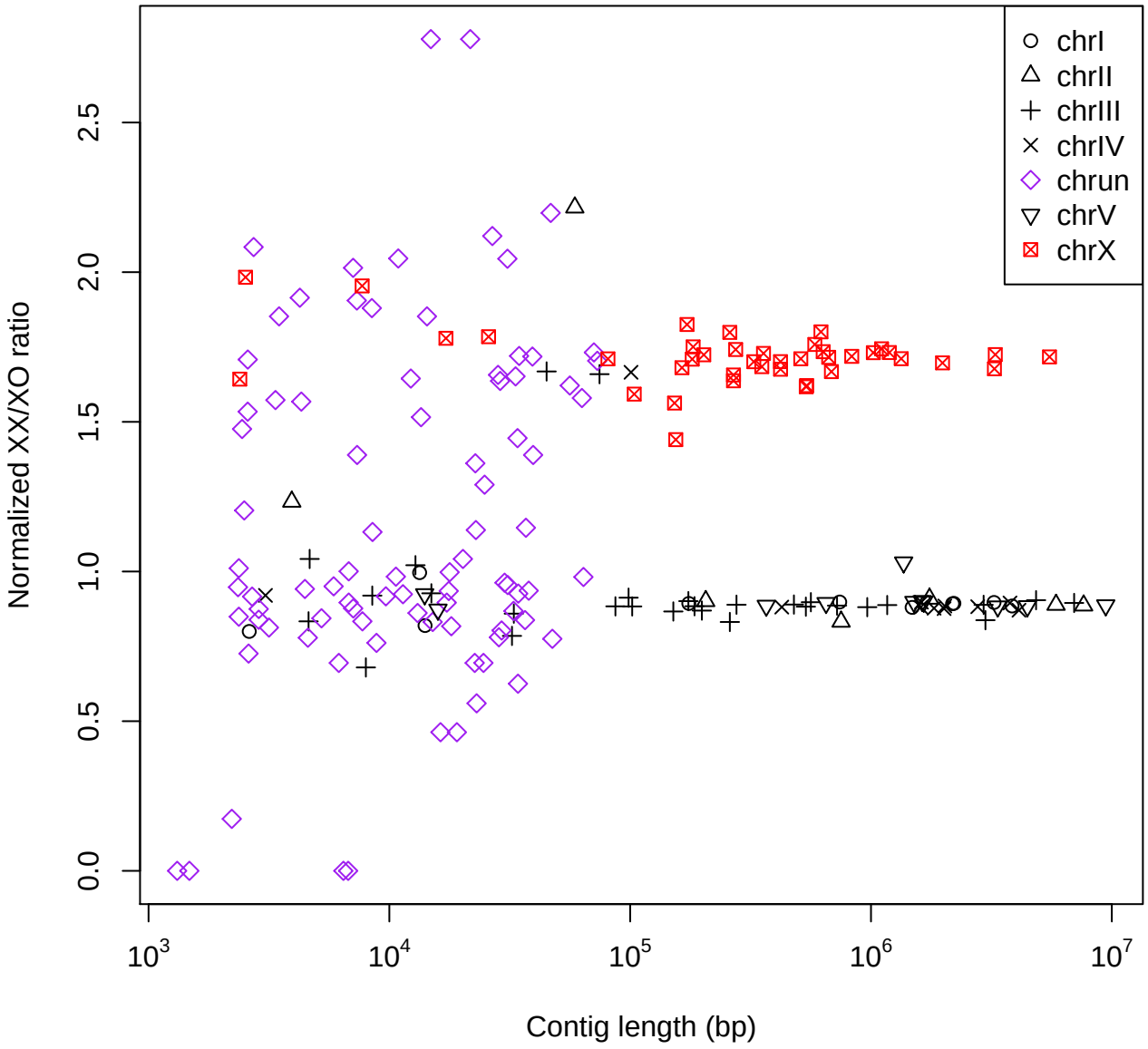


Figure 2.14: **The *C. nigoni* genome assembly** (Continued on the following page.)

Figure 2.14: **The *C. nigoni* genome assembly.**
The *C. nigoni* genome was sequenced to 100X coverage with Pacific Biosciences SMRT sequencing, resulting in 129 Mb in 211 contigs with a contig N50 of 3.3 Mb. Normalized read depth ratios of XX and XO libraries were plotted against contig length. These contigs were reordered and assigned to chromosomes using progressiveMauve with *C. briggsae* as a reference genome. Unassigned contigs were labeled "chrun". Contigs assigned to X by progressiveMauve were confirmed with XX/XO ratios. The final genome assembly was created as follows: progressiveMauve chromosome assemblies and contig orders were retained, with the following exceptions, 4 autosomal contigs and 5 unassigned contigs with ratios greater than 1.5 and length greater than 40 kb were assigned to "X_random" and unassigned contigs were ordered from highest to lowest XX/XO ratio rather than by contig length.

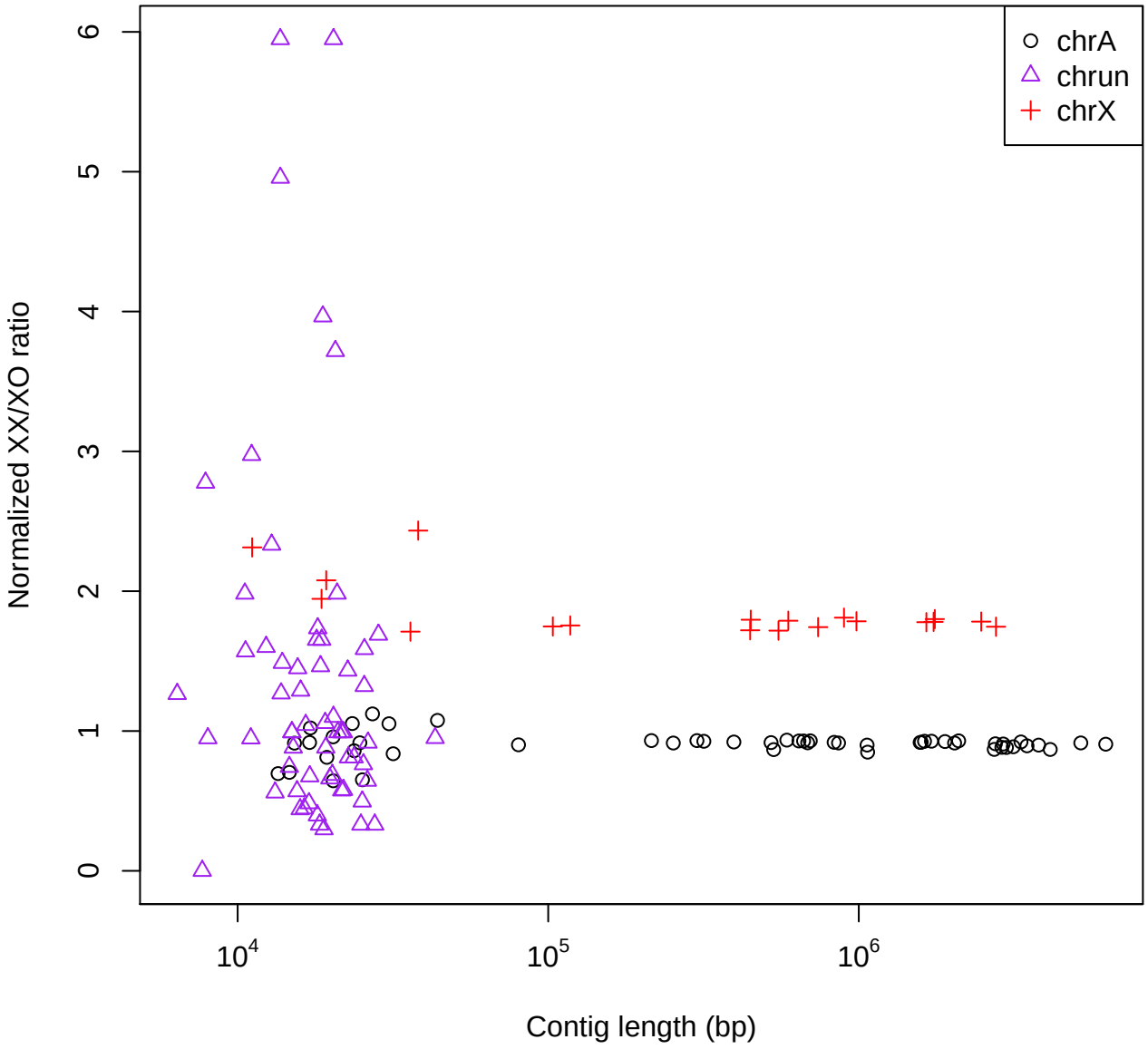


Figure 2.15: **The *C. tropicalis* genome assembly** (Continued on the following page.)

Figure 2.15: **The *C. tropicalis* genome assembly.**
The *C. tropicalis* genome was sequenced to 33X coverage with Pacific Biosciences SMRT sequencing and error corrected with Illumina reads, resulting in 82.8 Mb in 141 contigs with a contig N50 of 2.7 Mb. *C. tropicalis* contigs were assigned to the X chromosome or the autosomes by comparing read depth of hermaphrodite (XX) and male (XO) sequencing libraries. Contigs were assigned to X if the normalized XX/XO ratio was greater than 1.5 and at least 75 reads mapped to the contig across the six libraries. Contigs were assigned to "chrA" if ratios were less than 1.5 and at least 100 total reads mapped to the contig. A draft genome assembly was created by joining sequences in order of contig size.

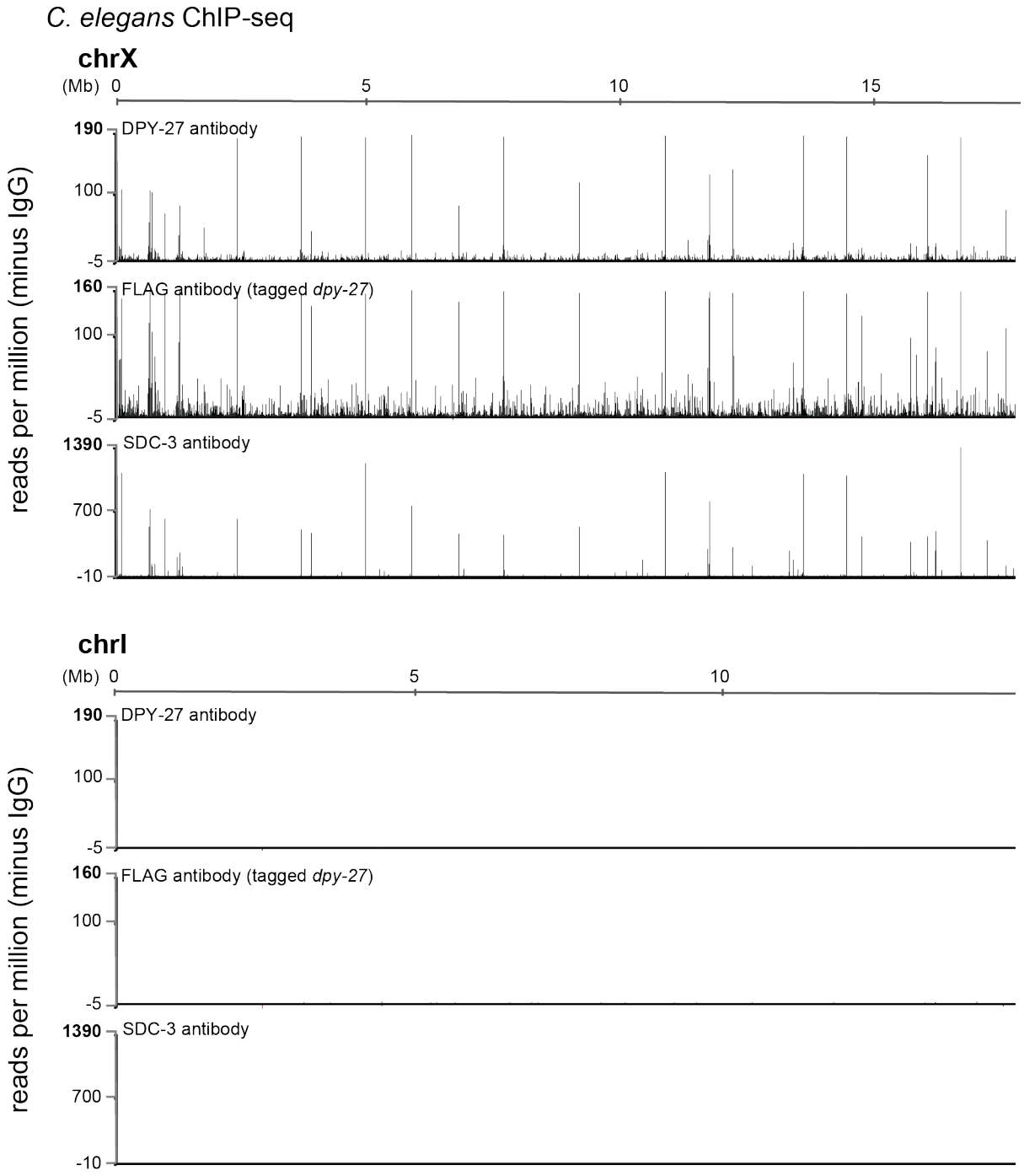


Figure 2.16: **ChIP-seq identifies discrete peaks on the X chromosome in *C. elegans*.** ChIP was performed with α-FLAG antibody in FLAG::DPY-27 and with α-DPY-27 and α-SDC-3 antibody in the wild-type reference strain, N2. ChIP with all three antibodies identified the same highly-occupied DCC-bound sites.

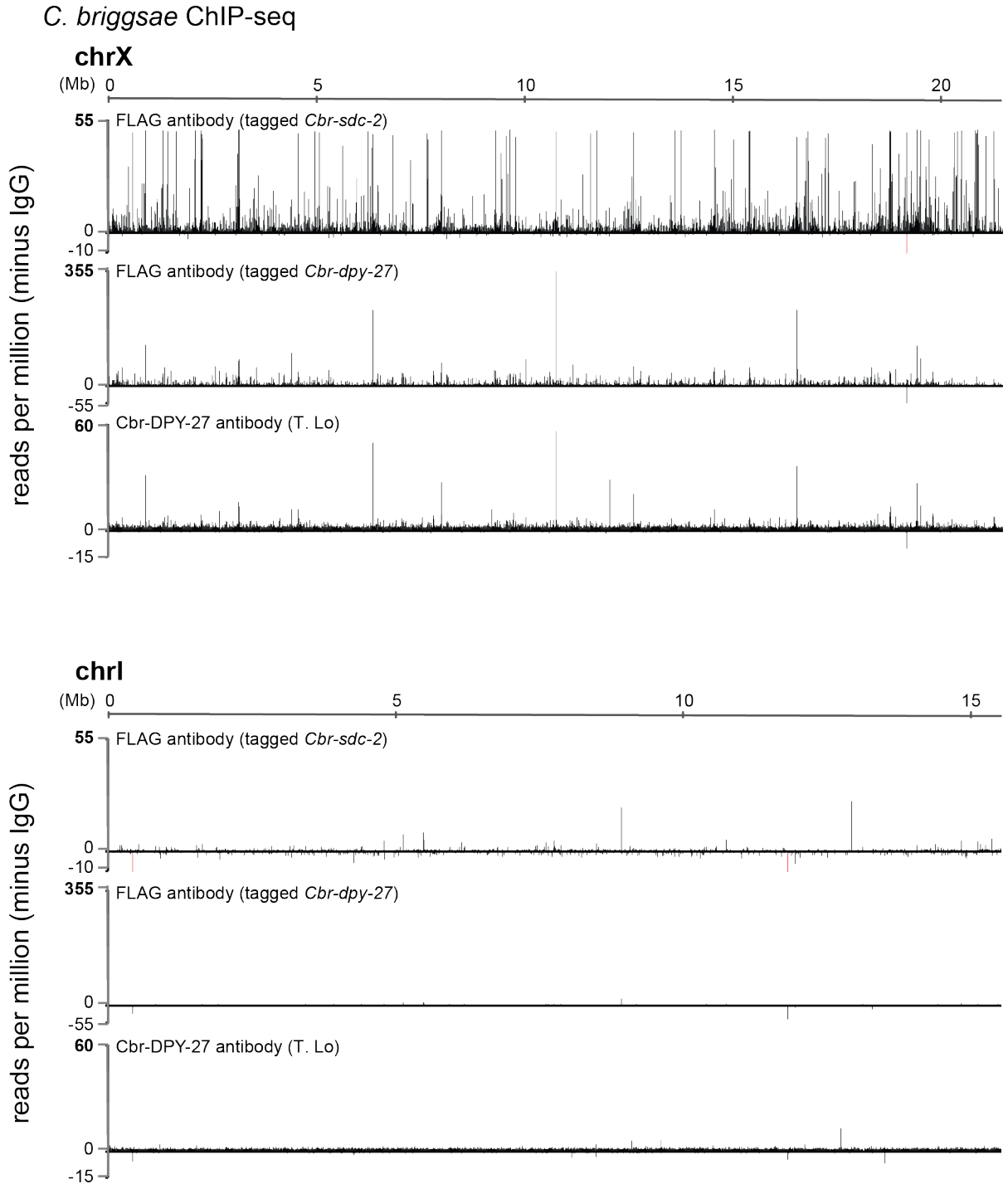


Figure 2.17: **ChIP-seq identifies discrete peaks on the X chromosome in *C. briggsae*.** (Continued on the following page.)

Figure 2.17: **ChIP-seq identifies discrete peaks on the X chromosome in *C. briggsae*.** ChIP was performed with α-FLAG antibody in FLAG::Cbr-SDC-2 and FLAG::Cbr-DPY-27 strains and with α-Cbr-DPY-27 peptide antibody in the wild-type reference strain, AF16. ChIP in all three strains identified the same highly-occupied DCC-bound sites.

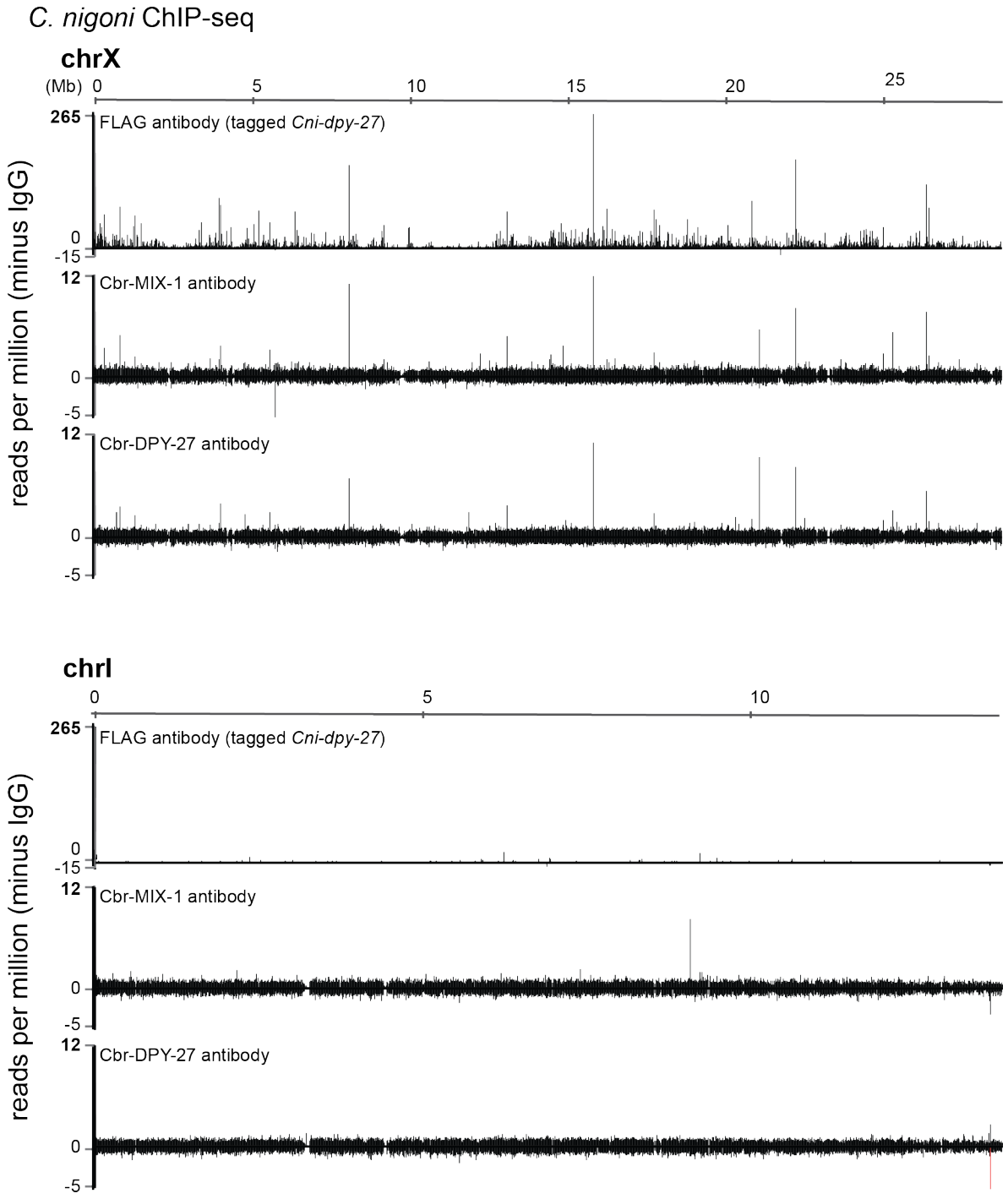


Figure 2.18: **ChIP-seq identified discrete peaks on the X chromosome in *C. nigoni*.** (Continued on the following page.)

Figure 2.18: **ChIP-seq identified discrete peaks on the X chromosome in *C. nigoni*.** ChIP-seq was performed with α-FLAG antibody in a FLAG::Cni-DPY-27 strain and with α-Cbr-DPY-27 and α-Cbr-MIX-1 peptide antibodies in the wild-type reference strain, JU1325. The strongest DCC-binding sites were identified with all three antibodies.

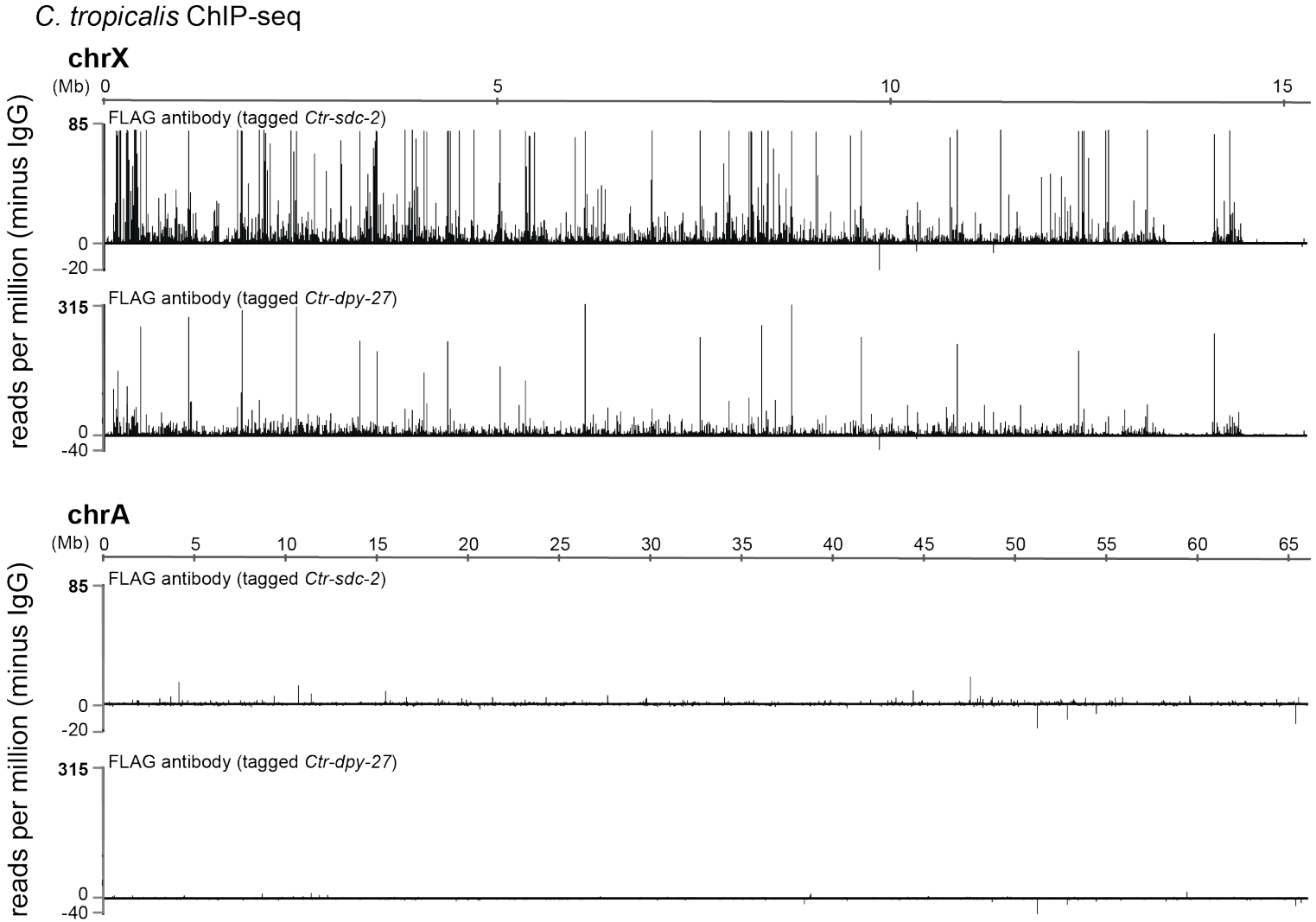
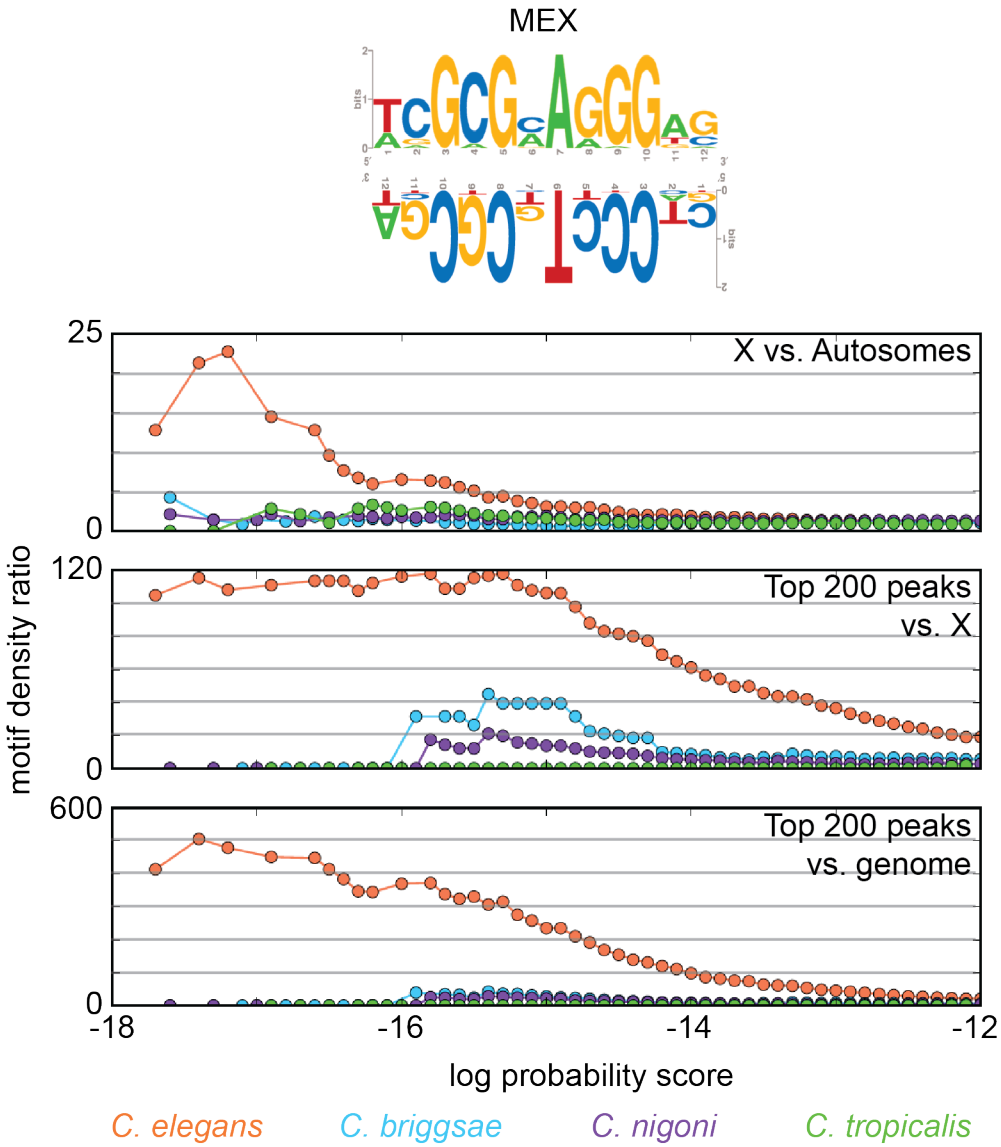
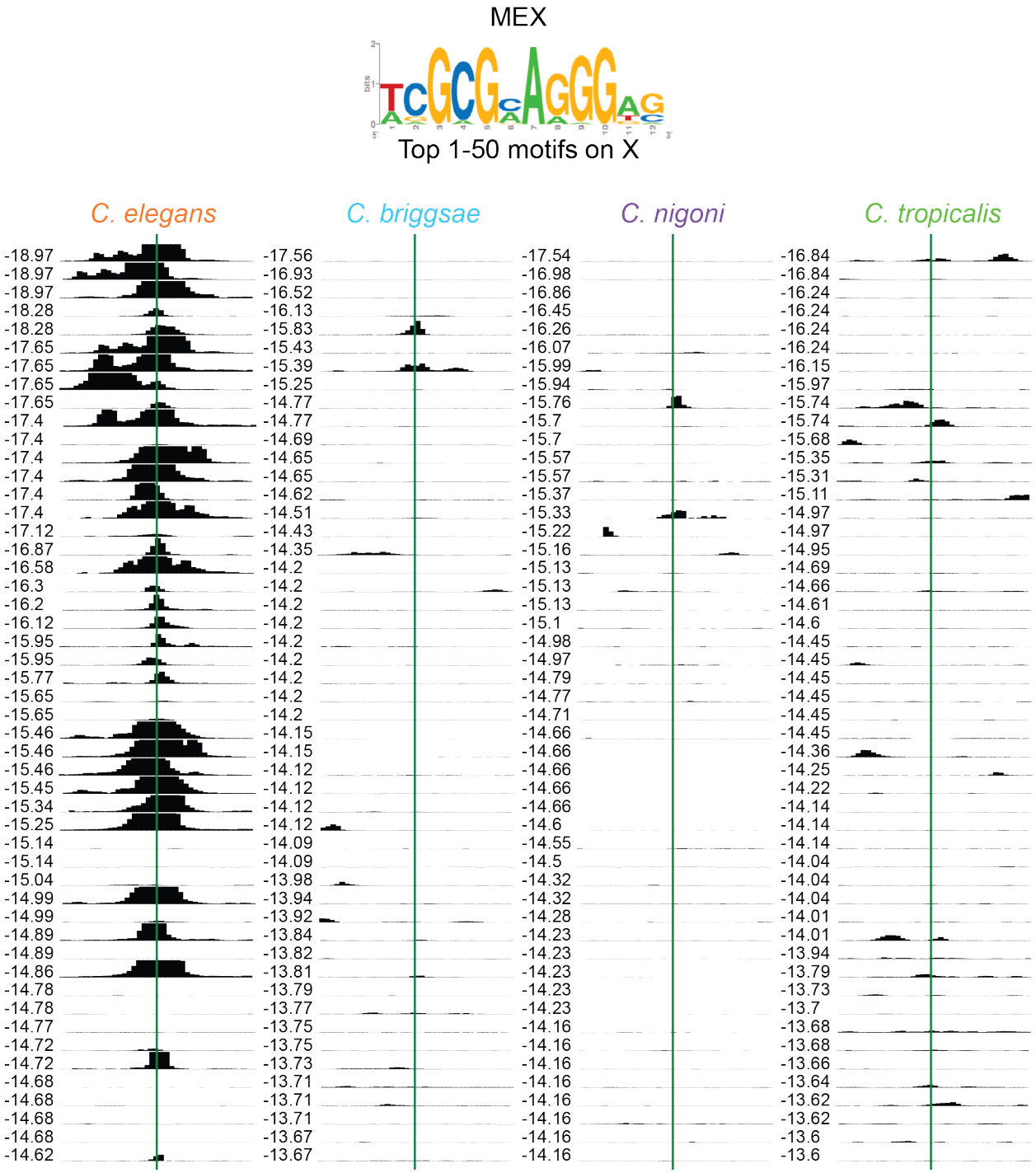


Figure 2.19: **ChIP-seq identifies discrete peaks on the X chromosome in *C. tropicalis*.** ChIP was performed with α-FLAG antibody in FLAG::Ctr-SDC-2 and FLAG::Ctr-DPY-27 strains. ChIP in both strains identified the same highly-occupied DCC-bound sites.

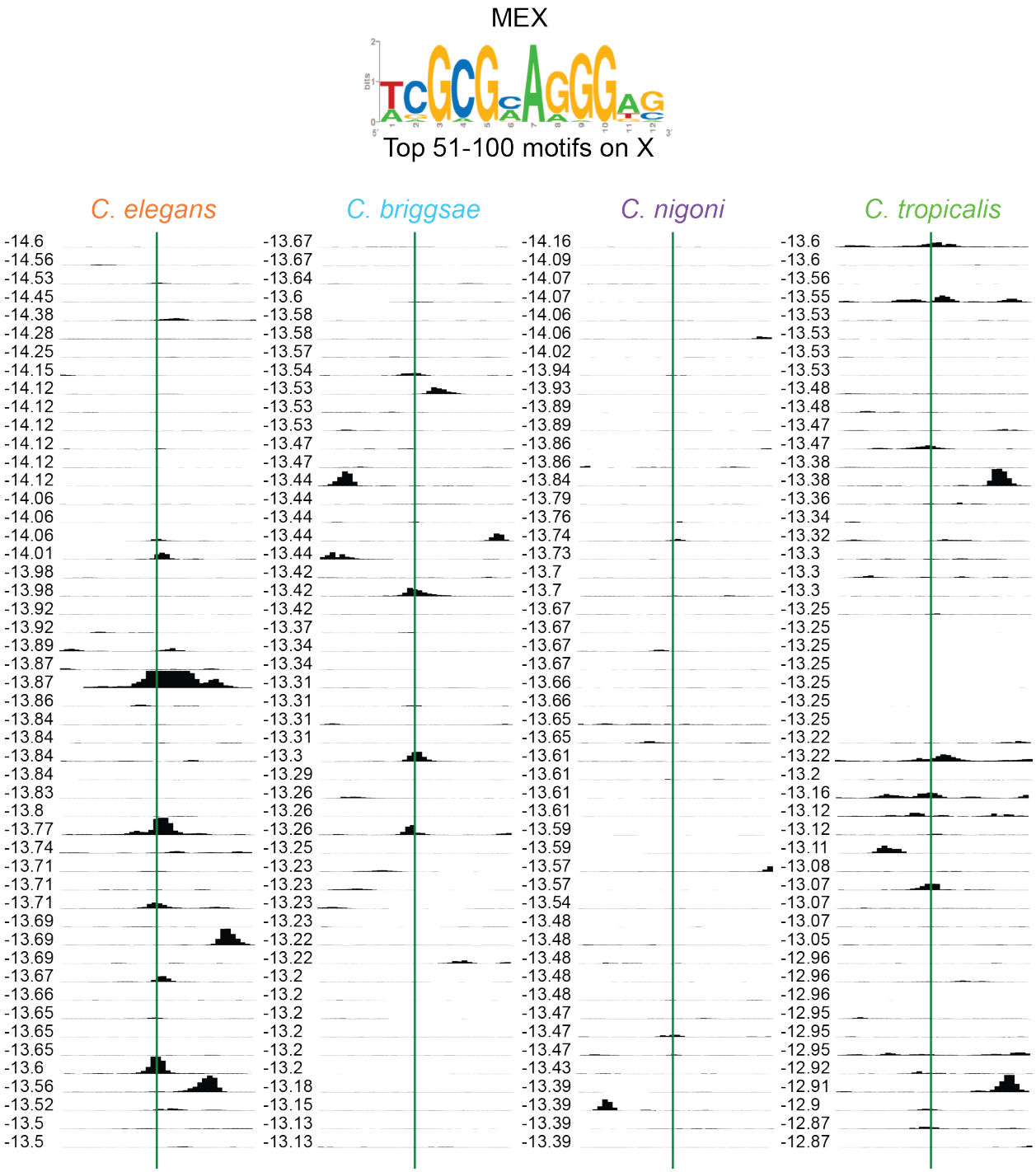
Figure 2.20: Cel-MEX a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cel-MEX motif is X-enriched and peak-enriched only in *C. elegans*.** The Cel-MEX motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

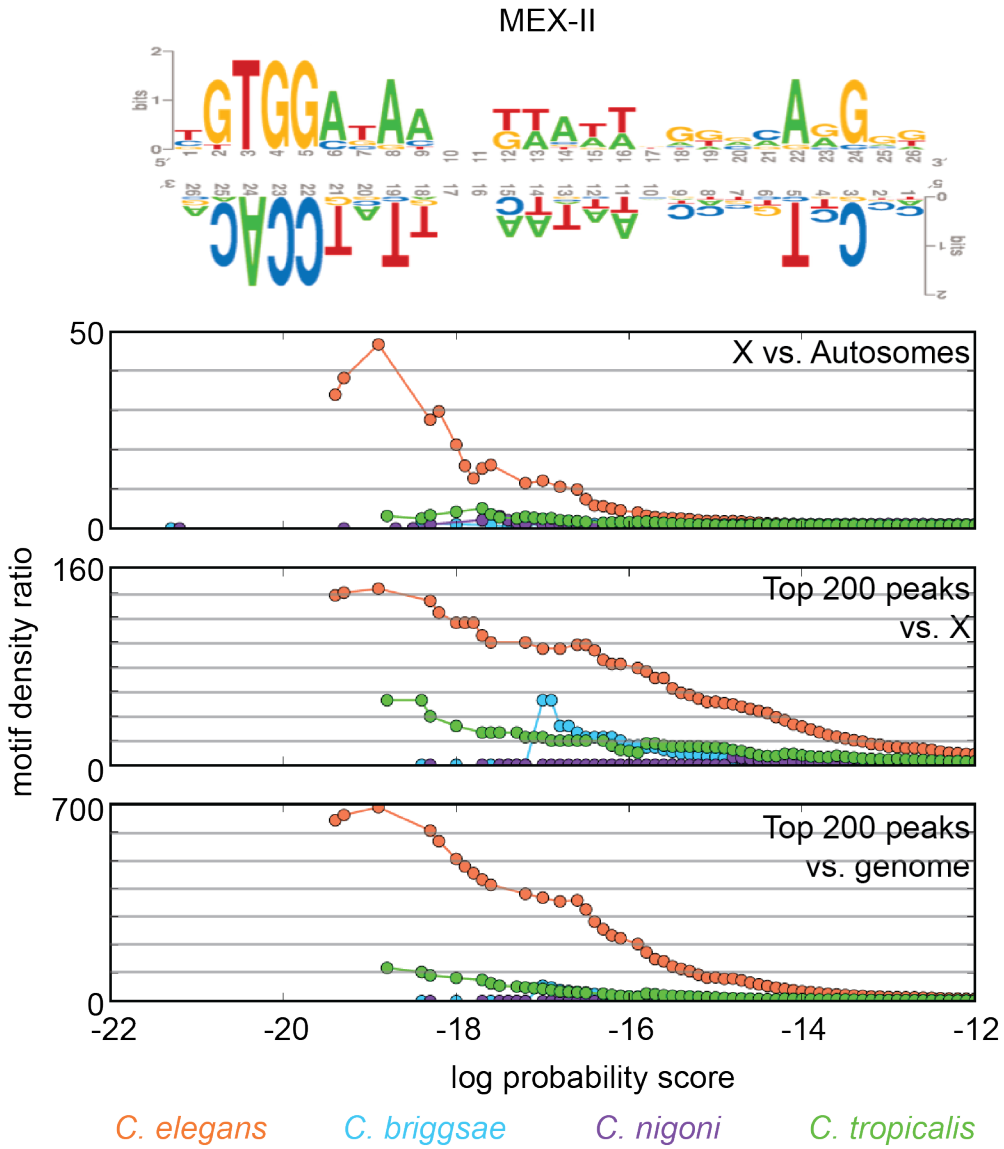


(b) **The *C. elegans* DCC is bound at most of the top 50 Cel-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Cel-MEX motifs on the X chromosome in four species.

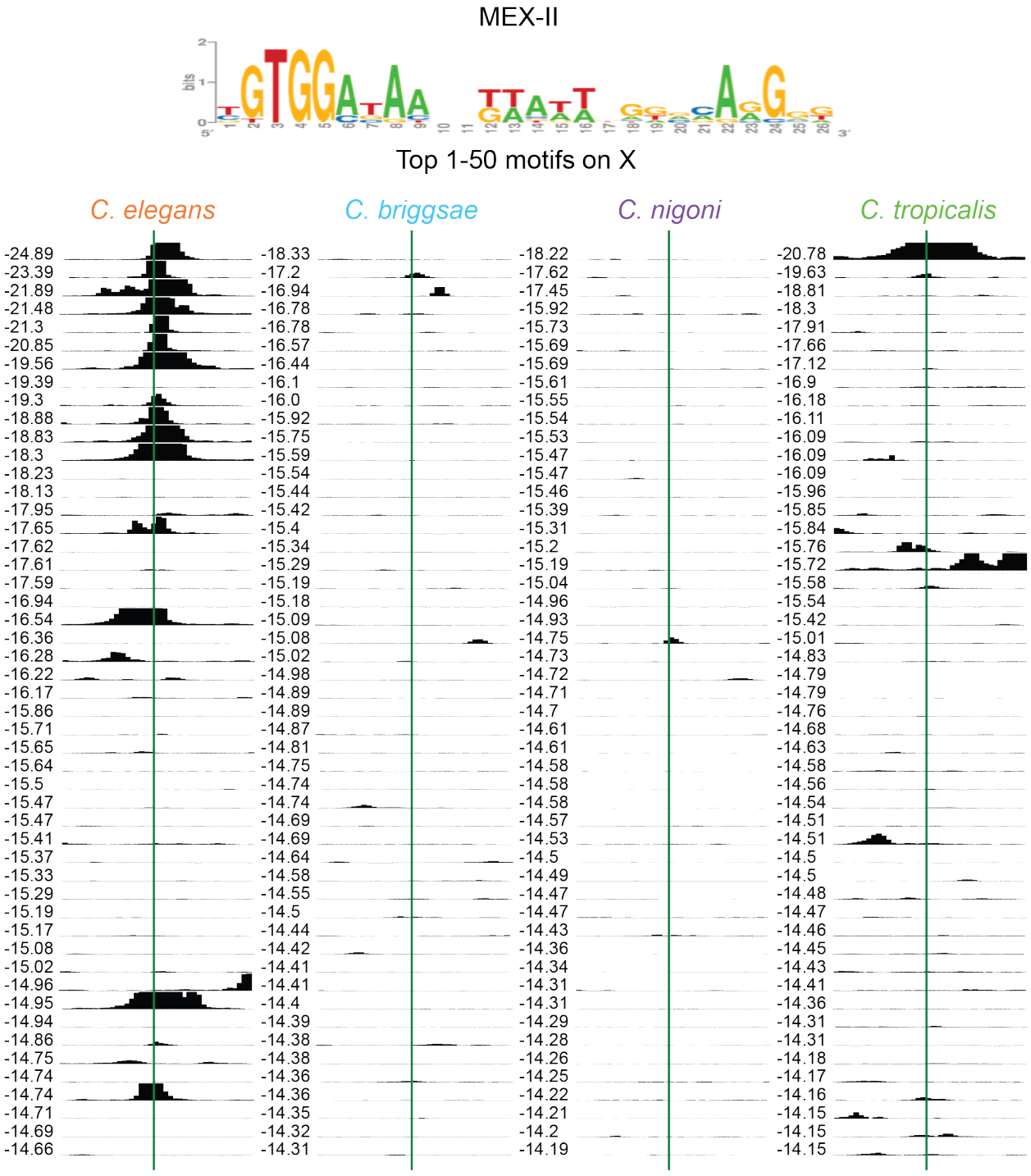


(c) **The *C. elegans* DCC is also bound at some of the top 51-100 Cel-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cel-MEX motifs on the X chromosome in four species.

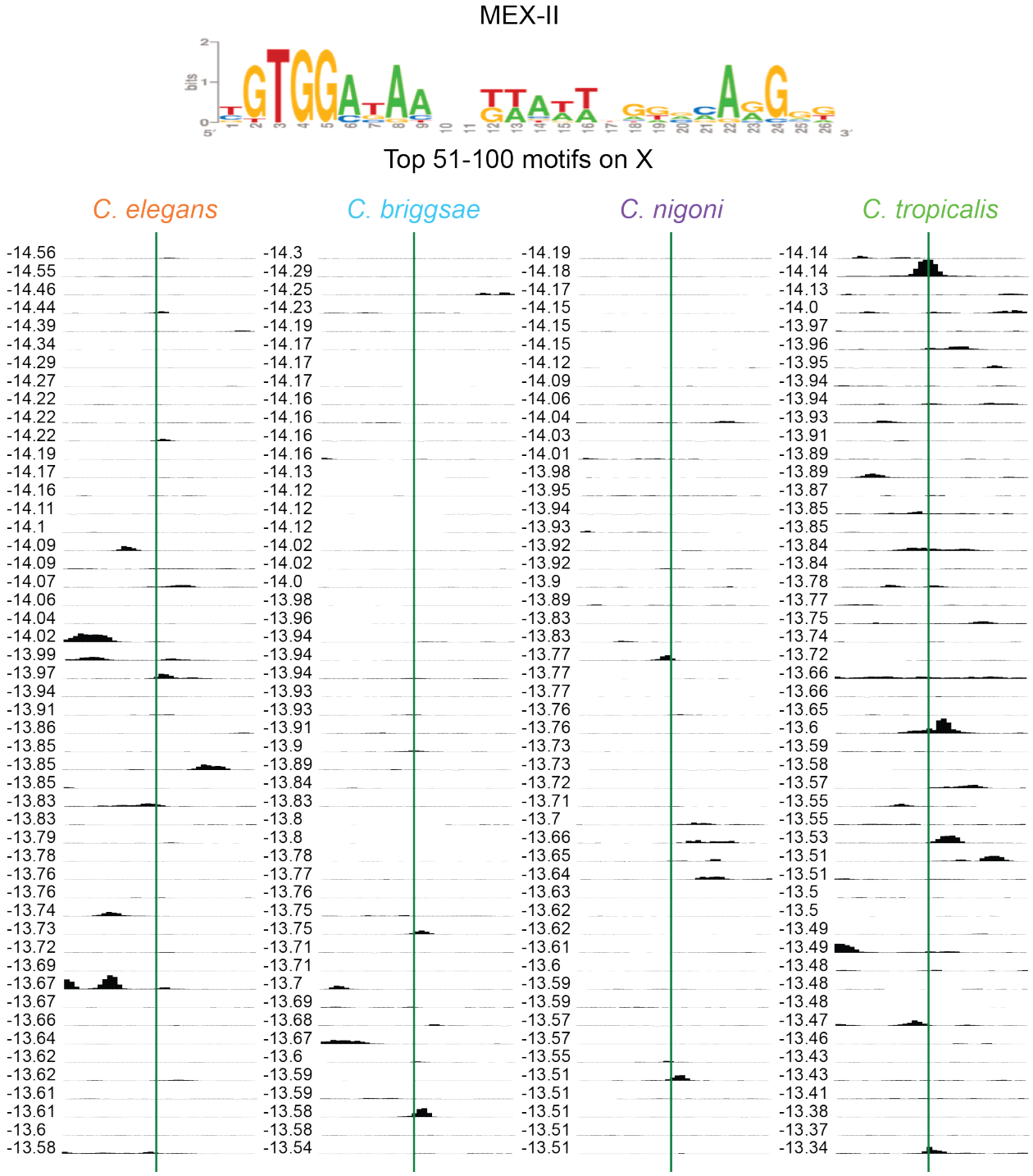
Figure 2.21: Cel-MEX-II a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cel-MEX-II motif is X-enriched only in *C. elegans* and peak enriched in *C. elegans* and *C. tropicalis*.** The Cel-MEX-II motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.



(b) **The *C. elegans* DCC is bound at many of the top 50 Cel-MEX-II motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Cel-MEX-II motifs on the X chromosome in four species.



(c) **The DCC is bound at few of the top 51-100 Cel-MEX-II motifs on the X chromosome across species.** The ChIP-seq signal is plotted at the top 51-100 Cel-MEX-II motifs on the X chromosome in four species.

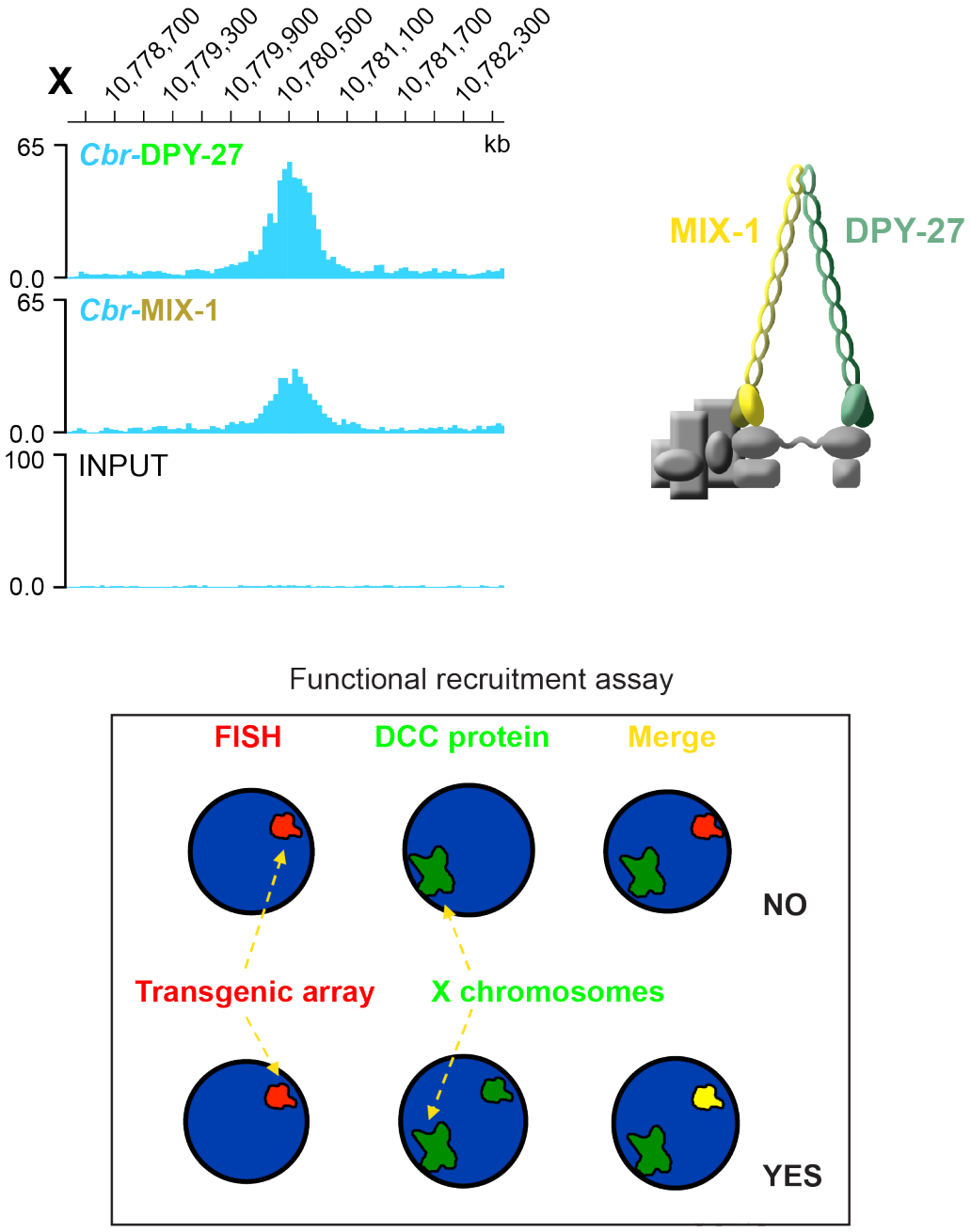


Figure 2.22: **Defining *rex* and *dox* sites in *C. briggsae*.** (Continued on the following page.)

Figure 2.22: **Defining *rex* sites in *C. briggsae*.** DCC-bound loci were identified by chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq). In this example, Cbr-DPY-27 and Cbr-MIX-1 peptide antibodies identified a highly occupied DCC-binding site. These DNA sequences were then injected to create extrachromosomal arrays. Arrays were stained with fluorescence *in situ* hybridization (FISH) probes. If DCC antibodies colocalized with the array, the site was defined as a recruitment element on X (*rex* site). Very strong recruitment titrated the DCC from the X chromosomes, so all DCC staining colocalized with the array. If the site was DCC-bound by ChIP-seq, but did not recruit in the functional assay, it would be defined as dependent on X (a *dox* site). "Flat" regions did not bind the DCC by either ChIP-seq or recruitment to an array.

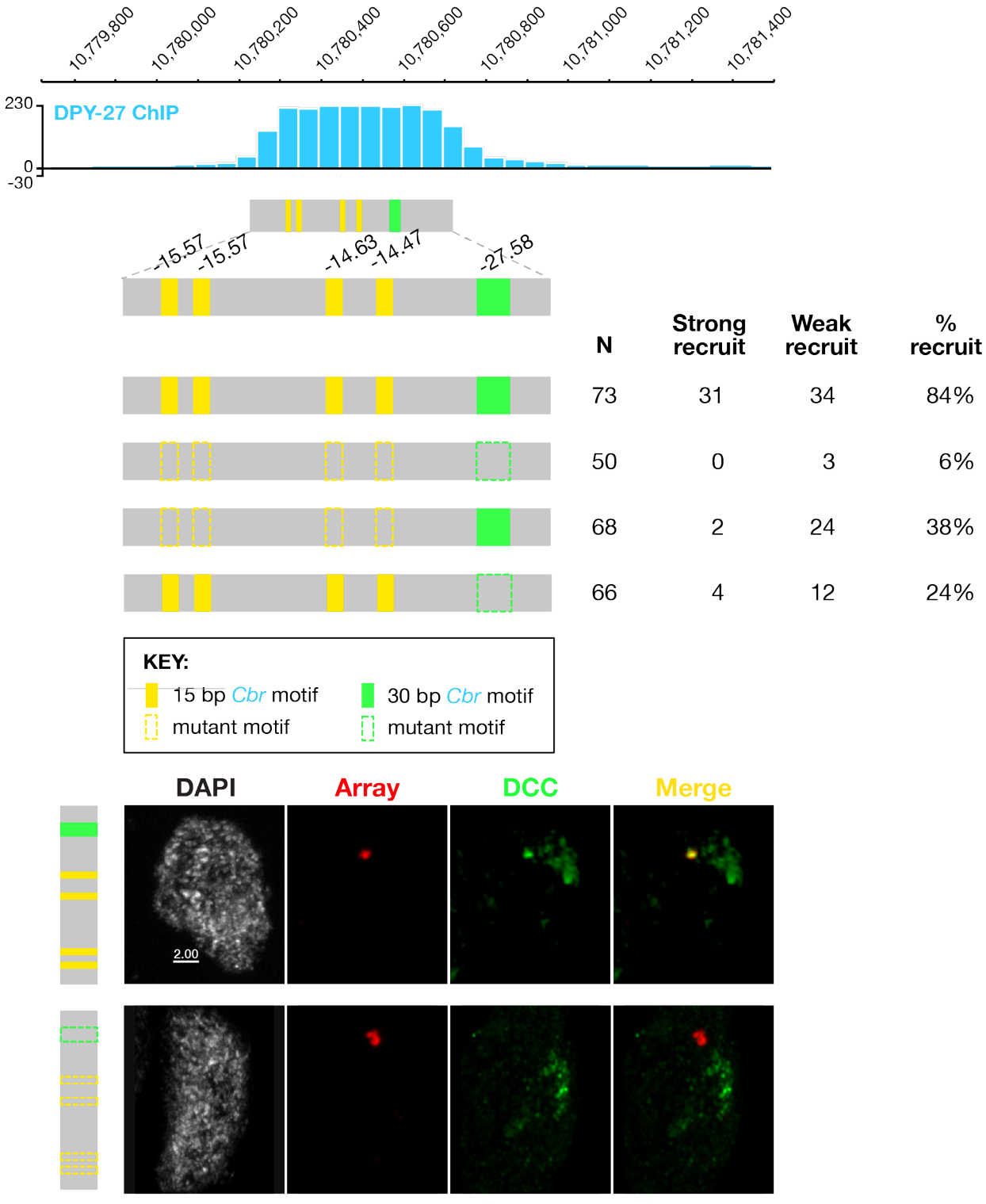


Figure 2.23: ***Cbr-rex-01* recruitment** (Continued on the following page.)

Figure 2.23: ***Cbr-rex-01* recruitment depends the Cbr-MEX and Cbr-MEX-II motifs.** The ChIP-seq signal at the *Cbr-rex-01* locus and the location and scores of Cbr-MEX and Cbr-MEX-II motifs are shown. An extrachromosomal array containing a 500 bp oligonucleotide with the *Cbr-rex-01* sequence recruited the *C. briggsae* DCC (84% recruitment) in first generation array-carrying progeny. The same sequence, with Cbr-MEX-II and four Cbr-MEX motifs scrambled failed to recruit (6% recruitment). Scrambling either Cbr-MEX or Cbr-MEX-II motifs reduced recruitment (38% and 24%, respectively), confirming that this *Cbr-rex* site depends on both Cbr-MEX and Cbr-MEX-II for full recruitment. In these confocal images, DNA was stained with DAPI, the extrachromosomal array was stained with FISH, and the DCC was stained with α-Cbr-DPY-27 antibody.

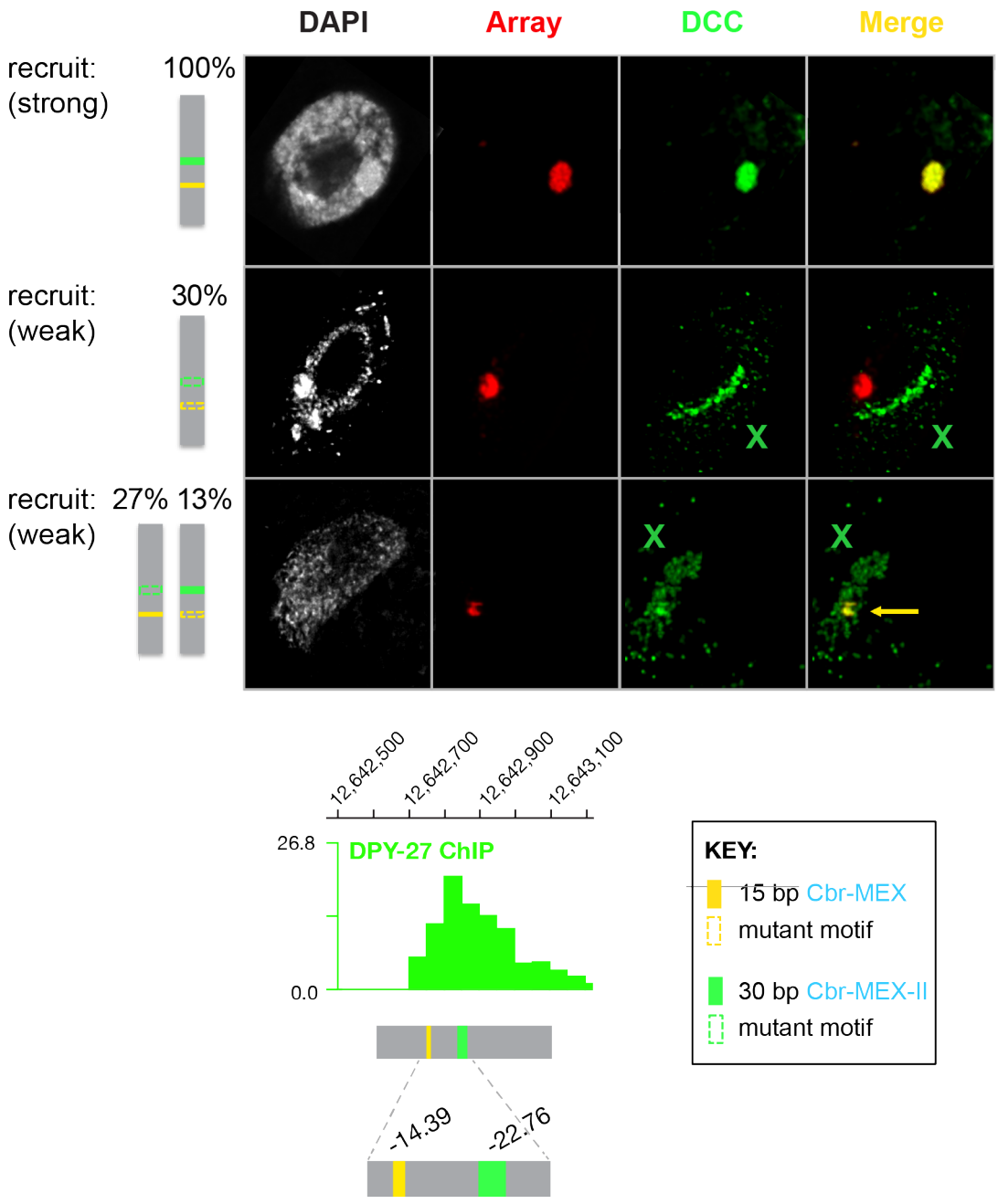
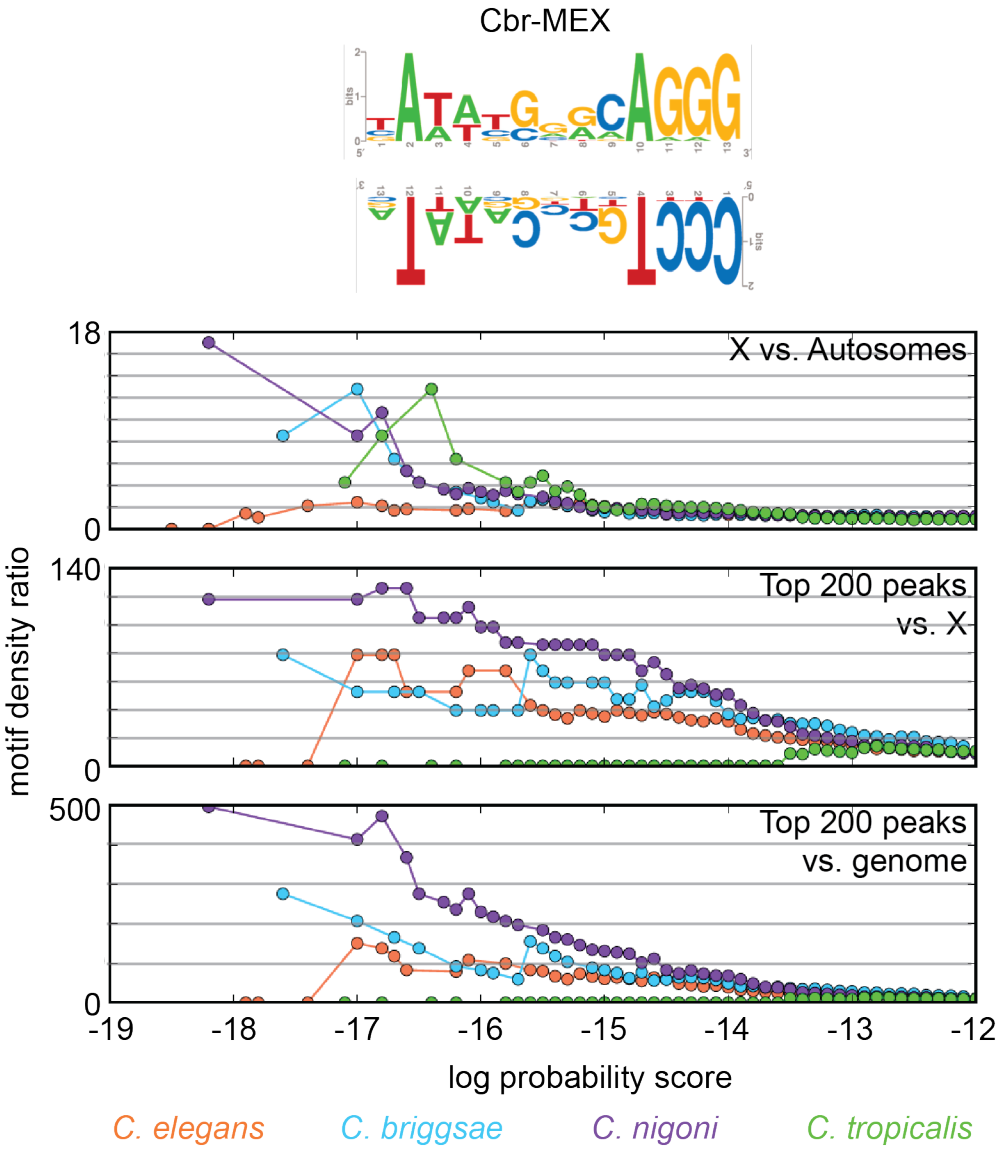


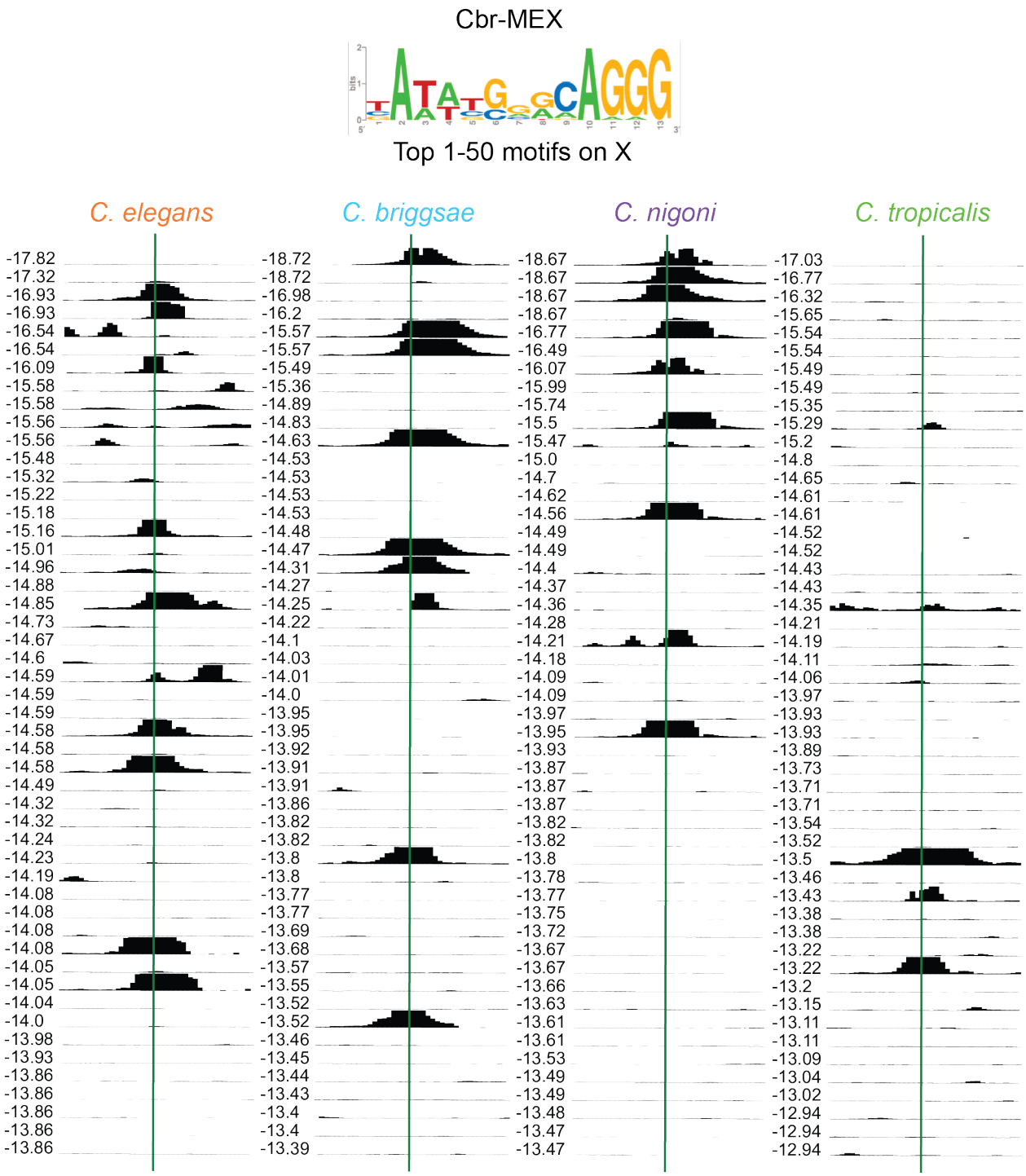
Figure 2.24: ***Cbr-rex-02*** **recruitment depends on Cbr-MEX and Cbr-MEX-II motifs.** (Continued on the following page.)

Figure 2.24: ***Cbr-rex-02* recruitment depends on Cbr-MEX and Cbr-MEX-II motifs.** In these confocal images, DNA was stained with DAPI, the extrachromosomal array was stained with FISH, and the DCC was stained with α-Cbr-DPY-27 antibody. The *Cbr-rex-02* sequence recruited the *C. briggsae* DCC 100% of the time. Sequences lacking Cbr-MEX and Cbr-MEX-II motifs had diminished DCC recruitment, confirming that both motifs contribute to DCC recruitment at this locus. The ChIP-seq signal at the *Cbr-rex-02* locus and the location and scores of Cbr-MEX and Cbr-MEX-II motifs are shown.

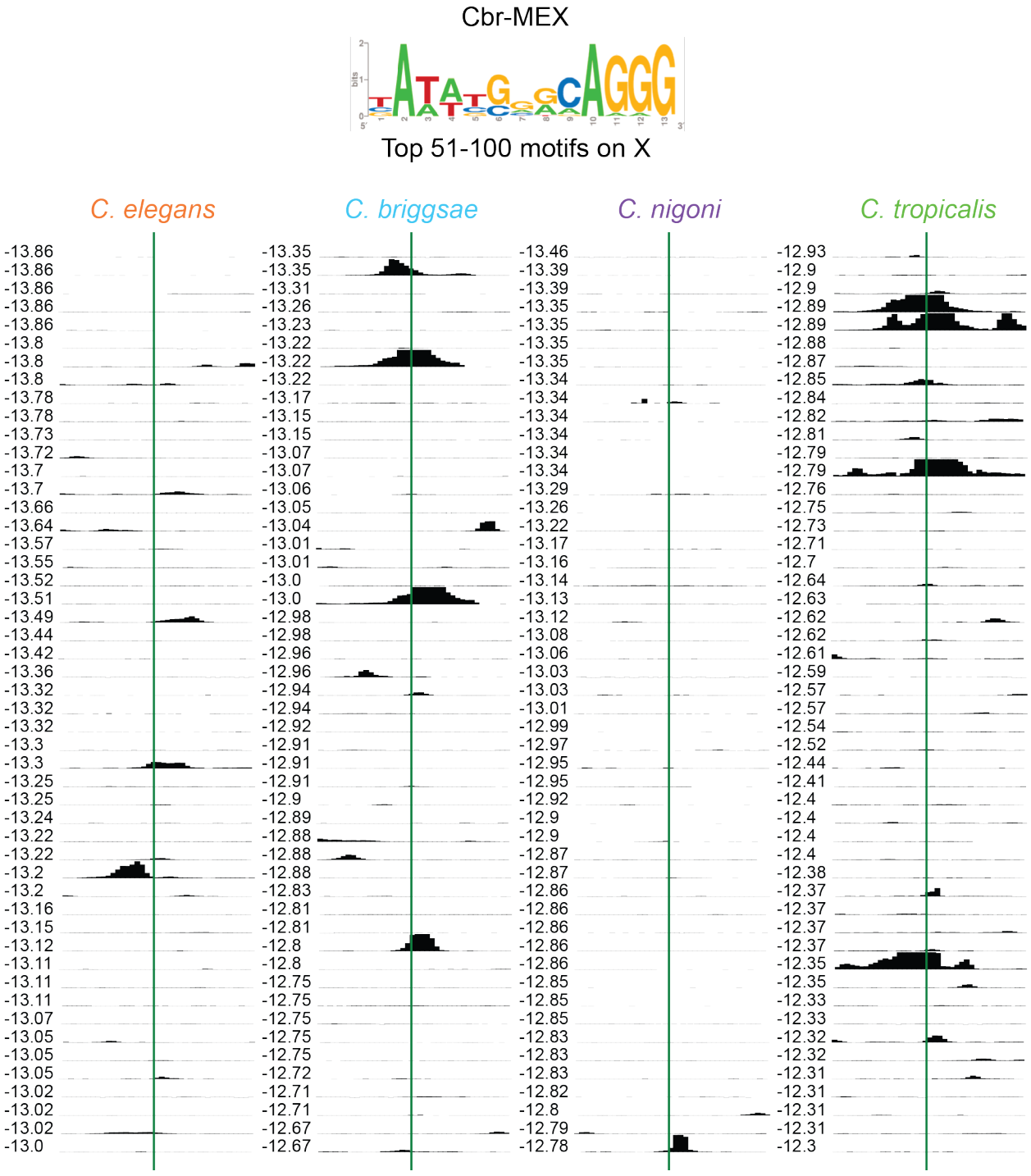
Figure 2.25: Cbr-MEX a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cbr-MEX motif is X-enriched in the *C. briggsae* clade and peak enriched in *C. briggsae* and *C. nigoni*.** The Cbr-MEX motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

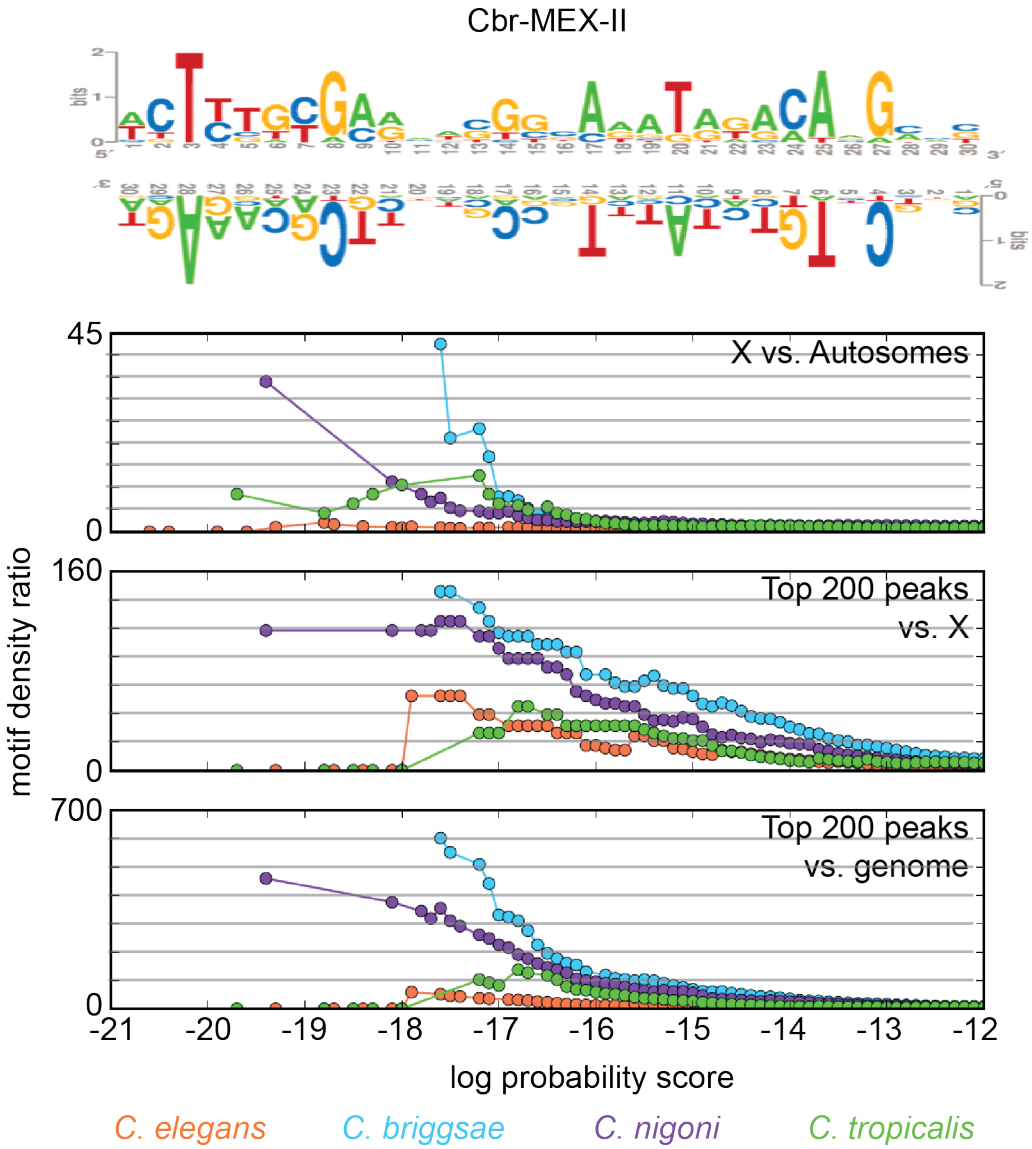


(b) **The DCC is bound at many of the top 50 Cbr-MEX motifs on the X chromosomes. Other species-specific motifs are also often found at these DCC-bound *C. elegans* and *C. tropicalis* sites.** The ChIP-seq signal is plotted at the top 50 Cel-MEX motifs on the X chromosome in four species.

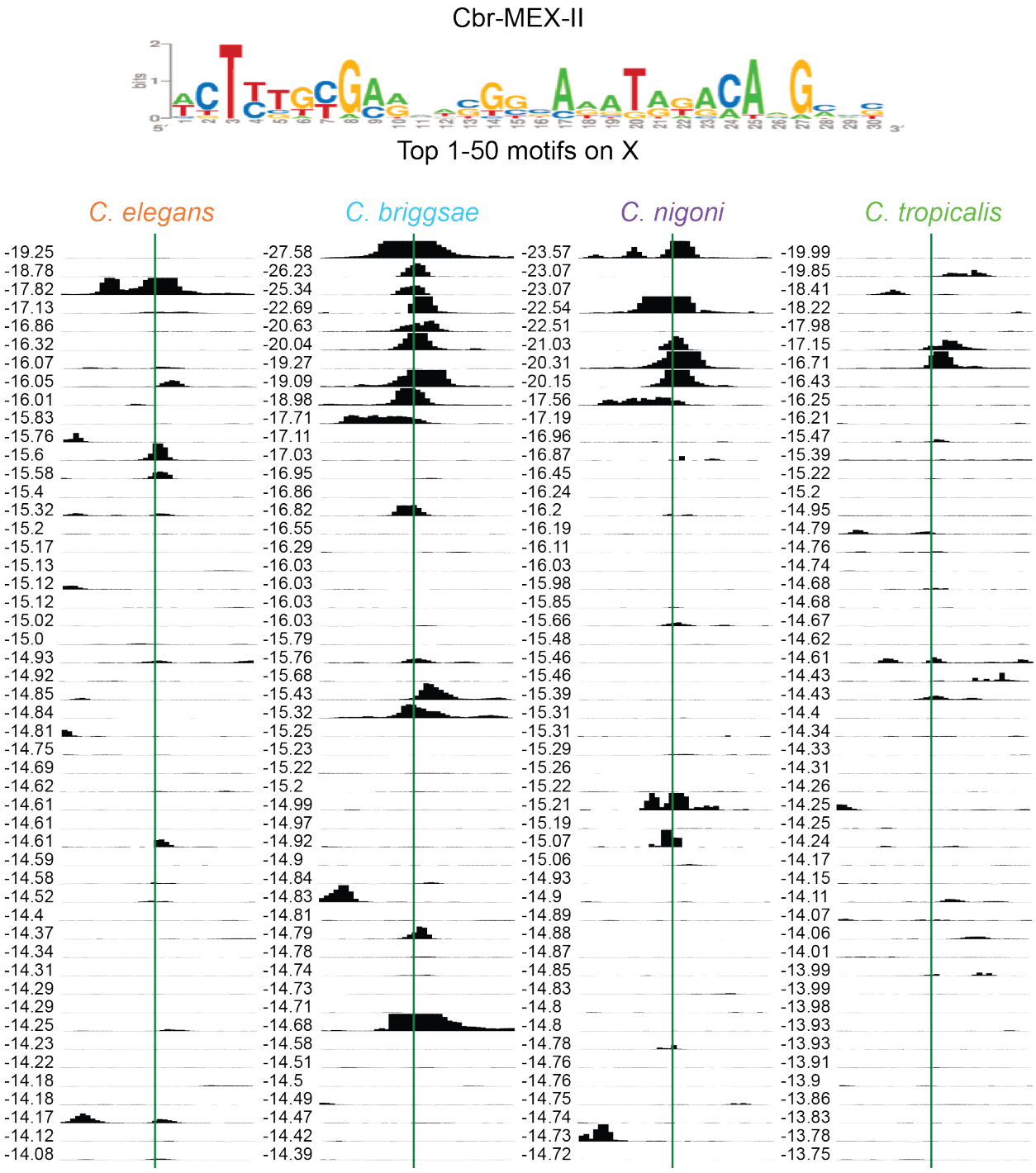


(c) **The DCC is bound at some of the top 50 Cbr-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cbr-MEX motifs on the X chromosome in four species.

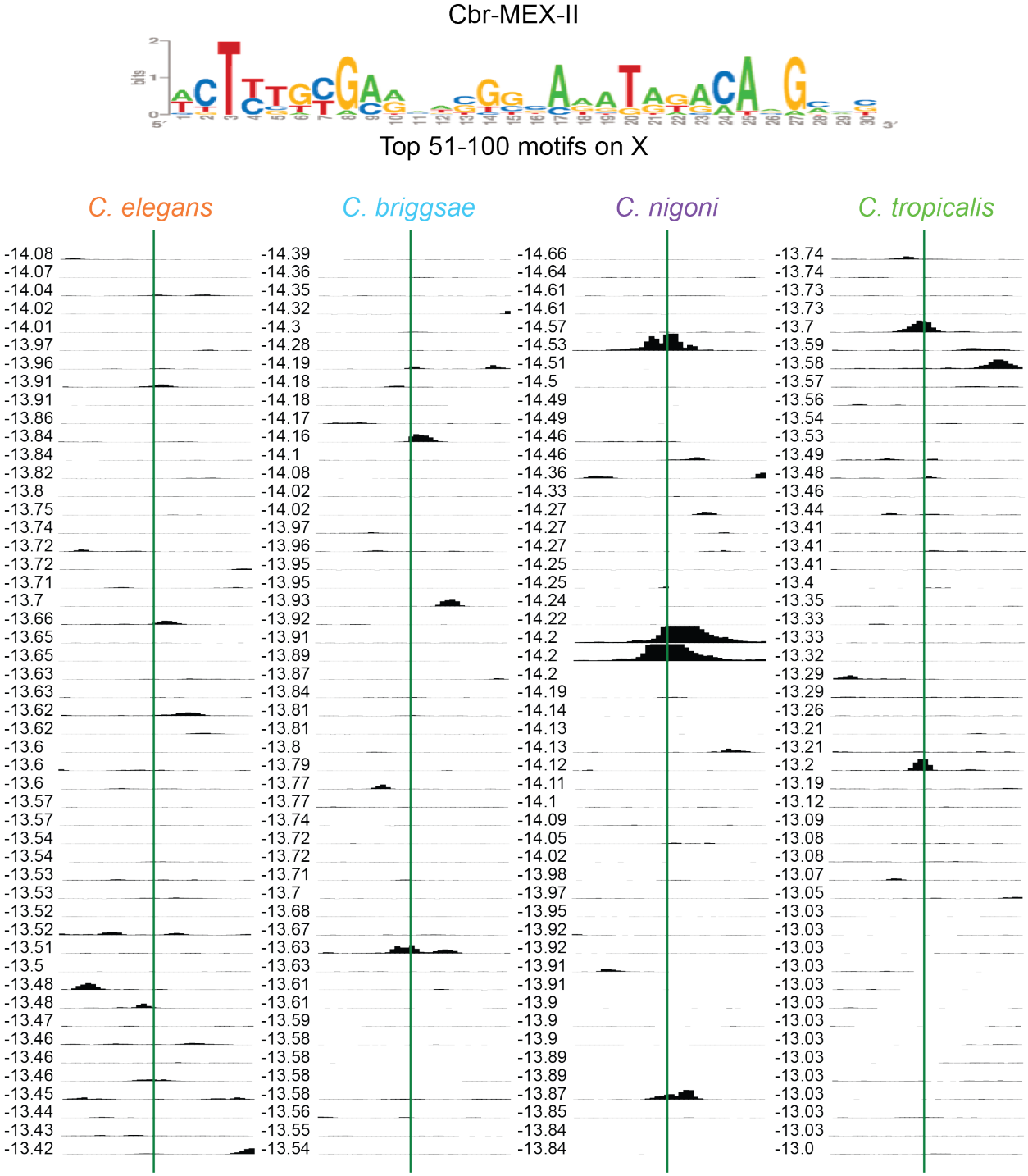
Figure 2.26: Cbr-MEX-II a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cbr-MEX-II motif is X-enriched in the *C. briggsae* clade and peak enriched in *C. briggsae* and *C. nigoni*.** The Cbr-MEX-II motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

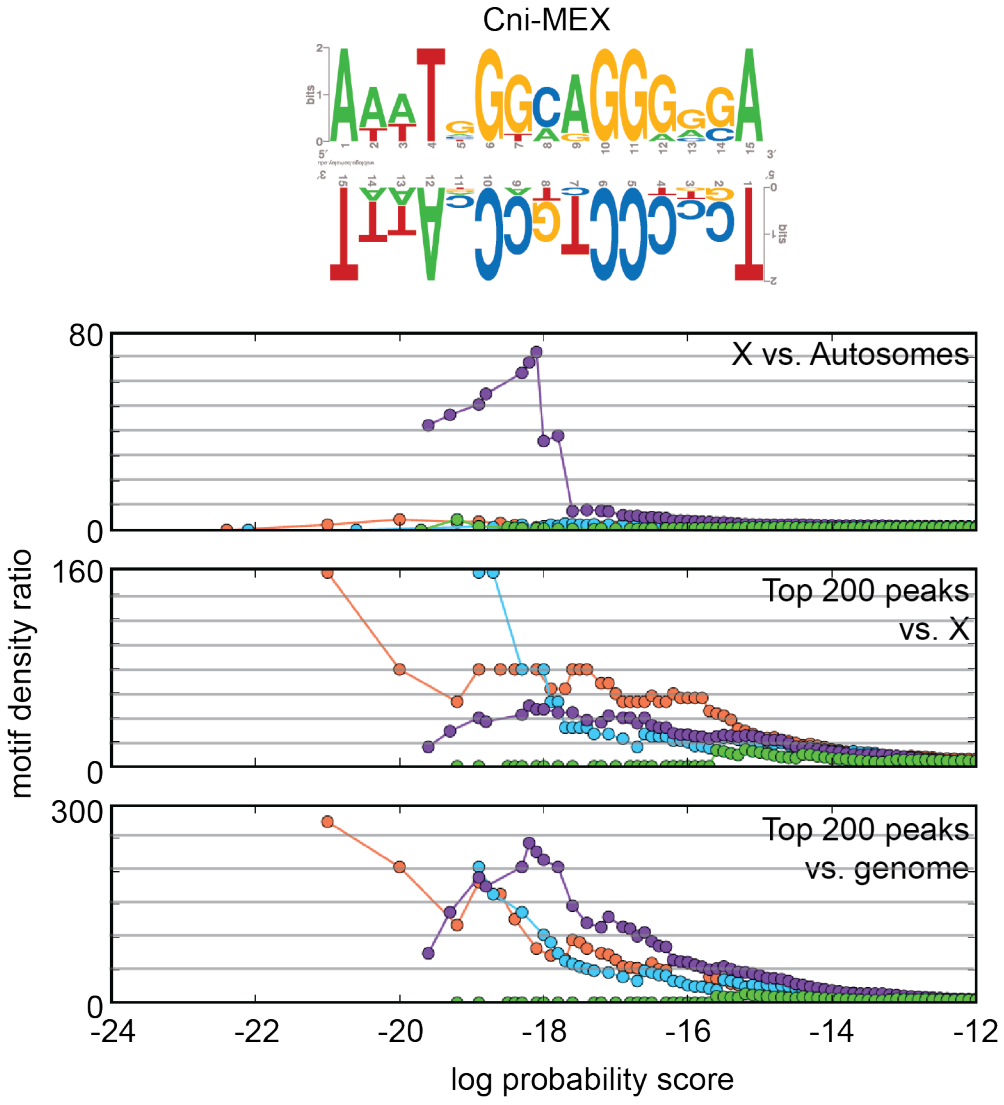


(b) **The DCC is bound at the strongest of the top 50 Cbr-MEX-II motifs on the X chromosome in *C. briggsae* and *C. nigoni*.** The ChIP-seq signal is plotted at the top 50 Cbr-MEX-II motifs on the X chromosome in four species.

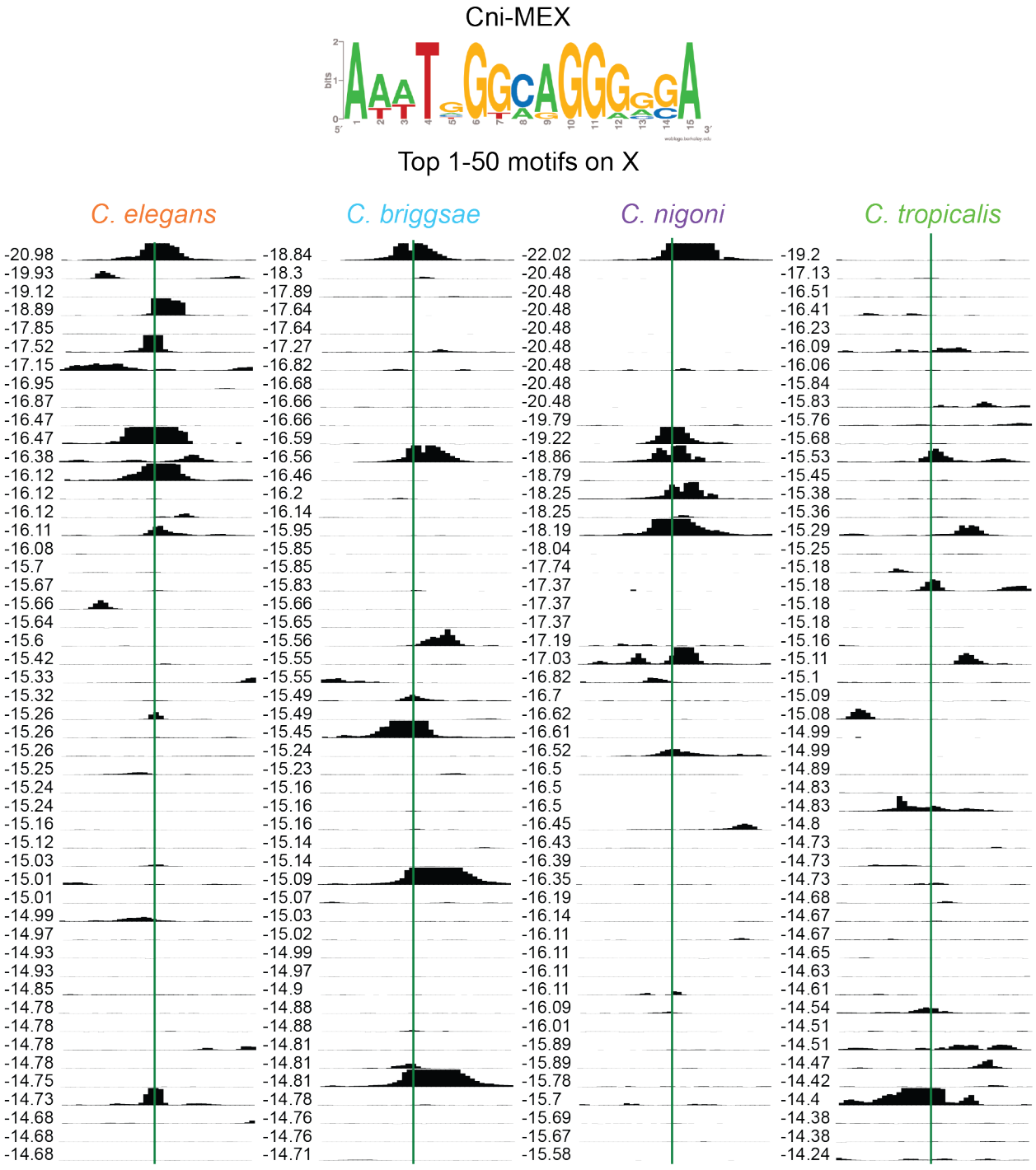


(c) **The DCC is bound at few of the top 51-100 Cbr-MEX-II motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Cbr-MEX-II motifs on the X chromosome in four species.

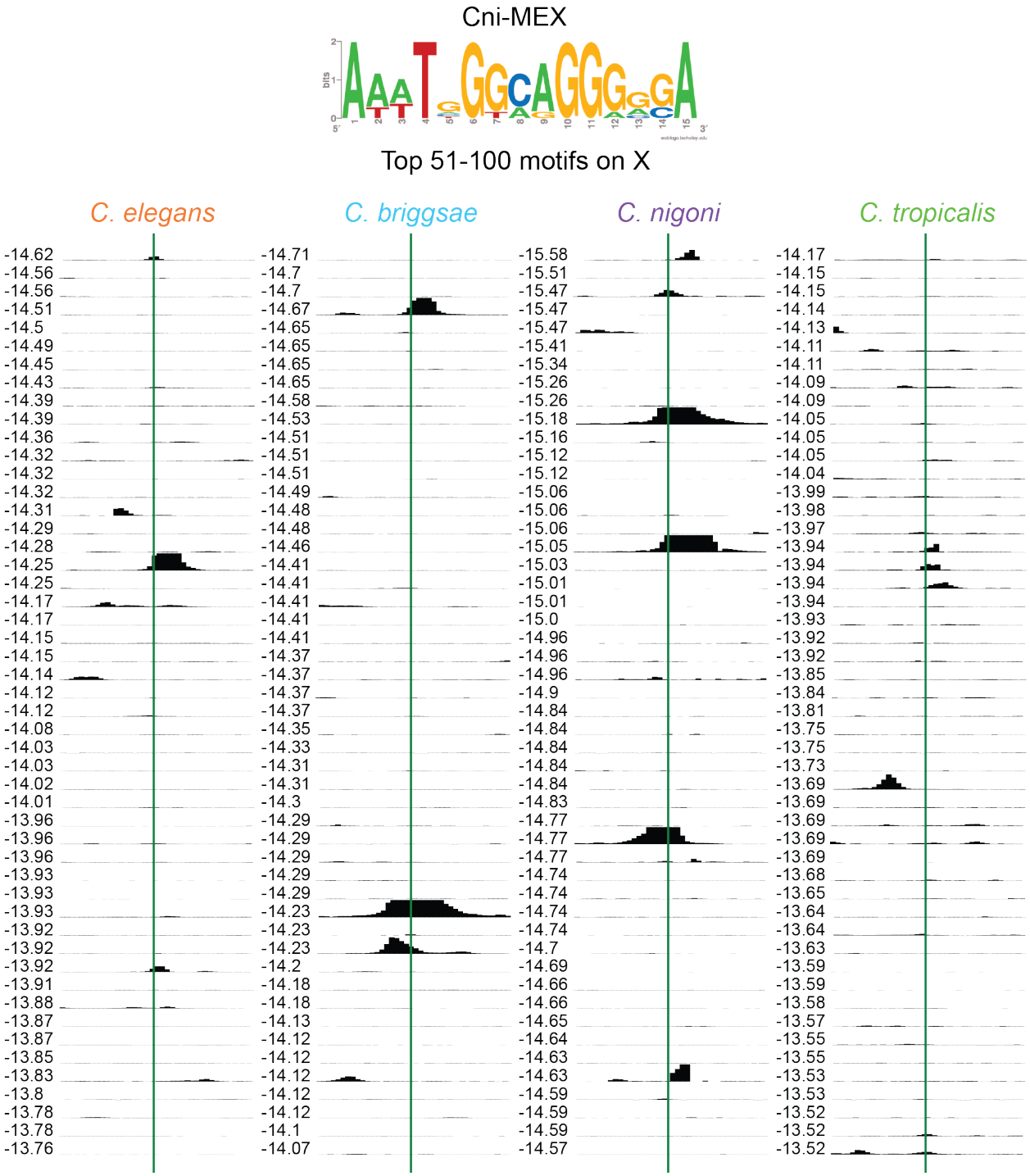
Figure 2.27: Cni-MEX a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni-MEX motif is only X-enriched in *C. nigoni* and is peak enriched in *C. nigoni*, *C. briggsae*, and *C. elegans*.** The Cni-MEX motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

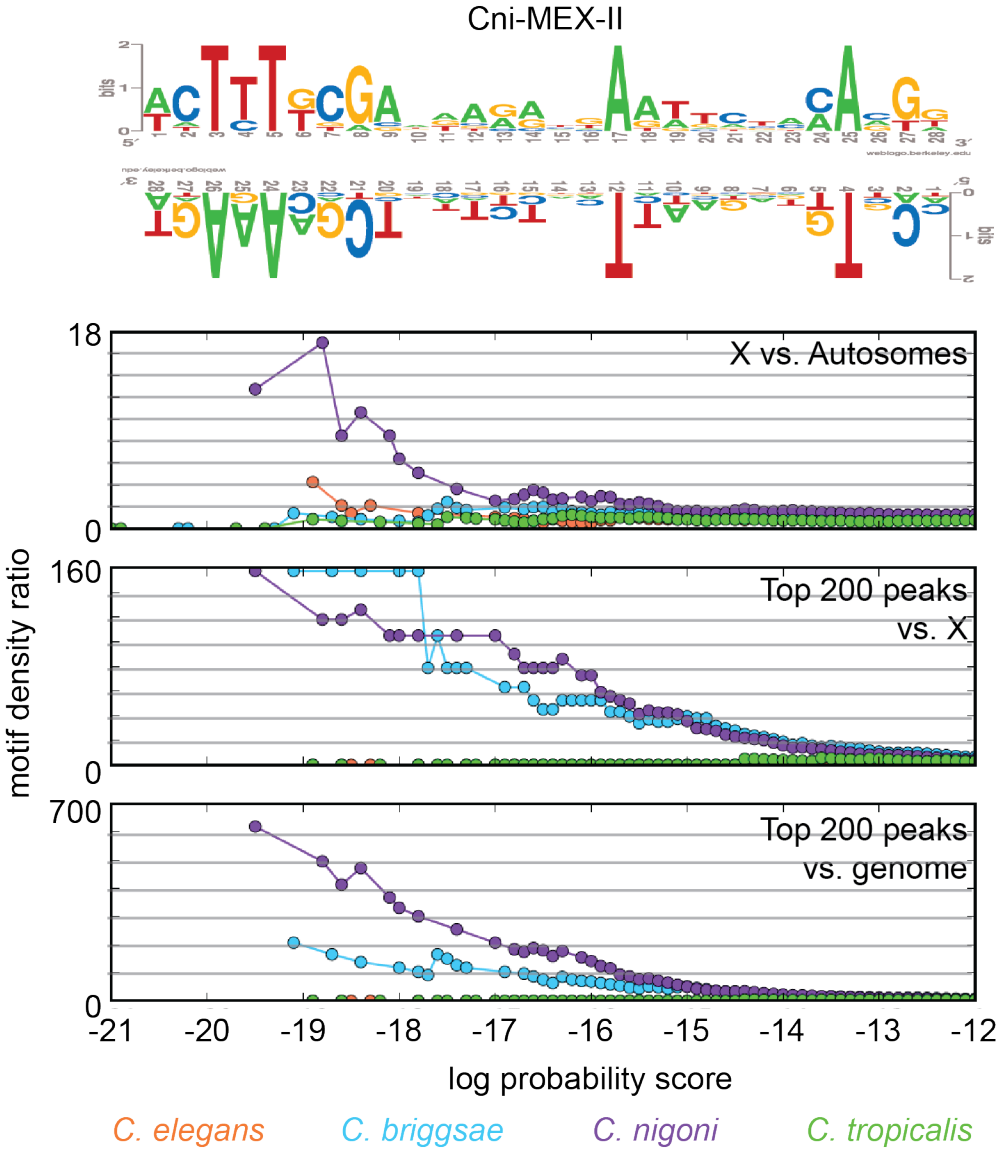


(b) **The DCC is bound at many of the top 50 Cni-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Cni-MEX motifs on the X chromosome in four species.

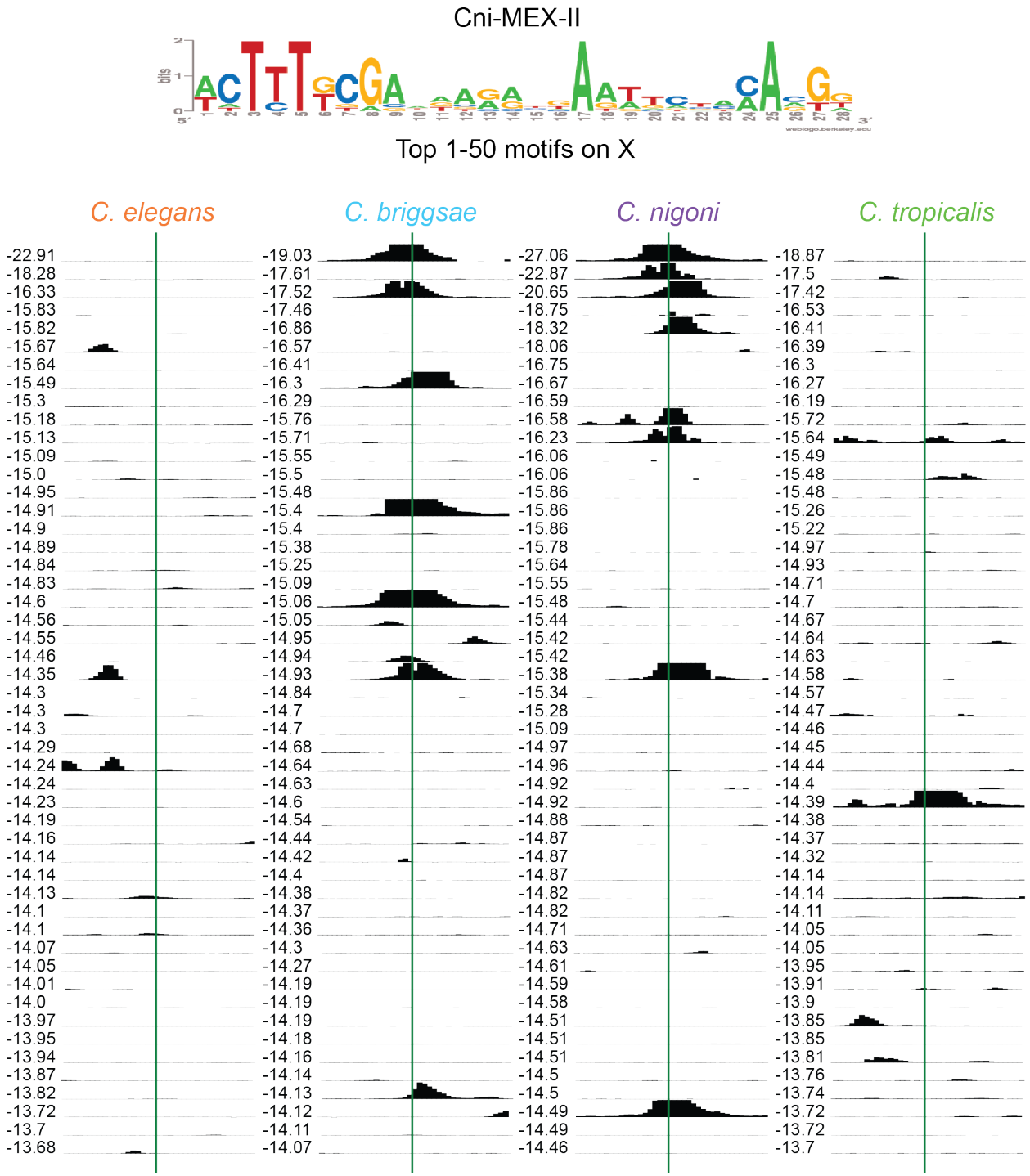


(c) **The DCC is bound at some of the top 51-100 Cni-MEX-II motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni-MEX-II motifs on the X chromosome in four species.

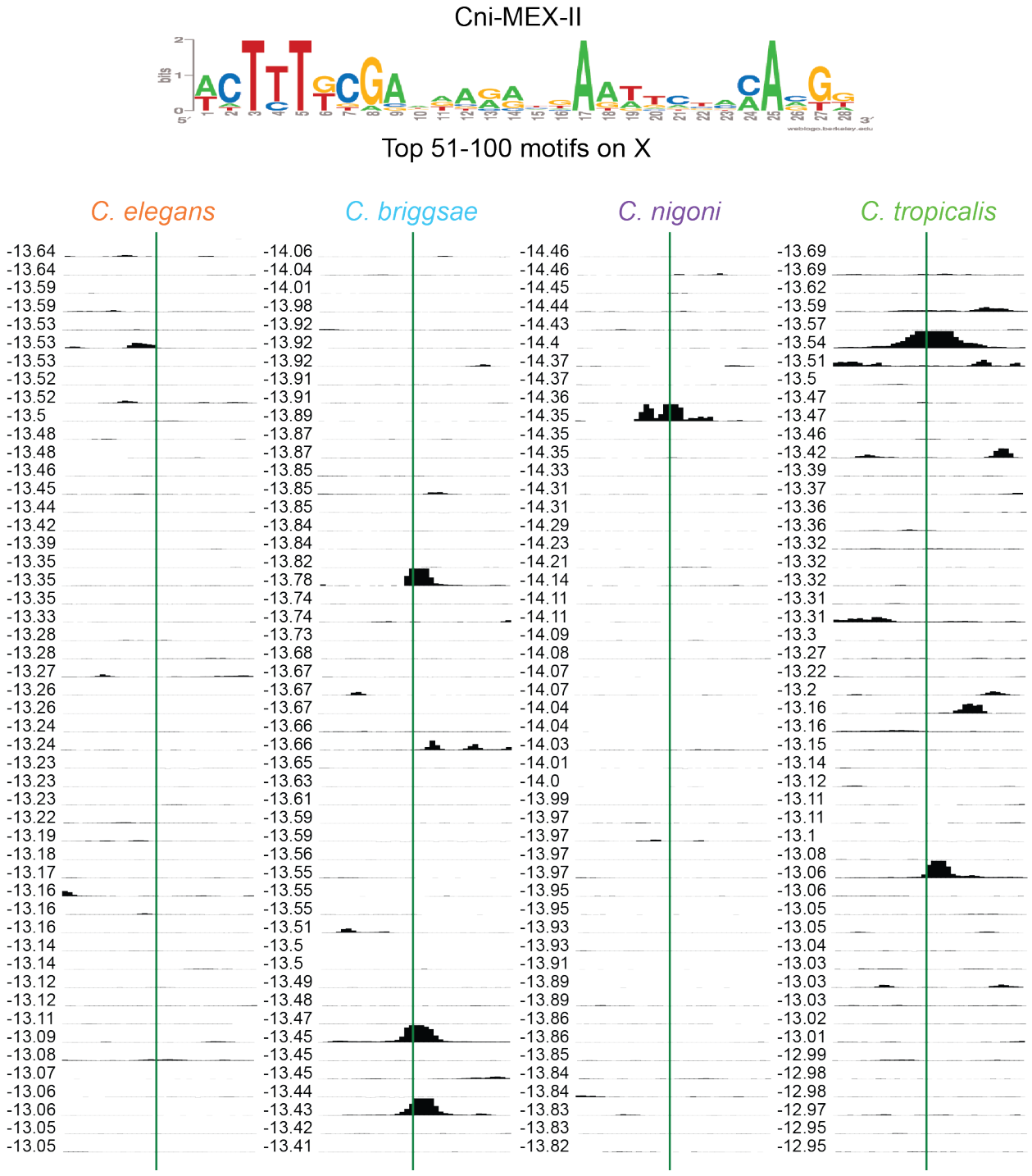
Figure 2.28: Cni-MEX-II a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni-MEX-II motif only X-enriched in *C. nigoni* and is peak enriched in *C. nigoni* and *C. briggsae*.** The Cni-MEX-II motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.



(b) **The DCC is bound at many of the top 50 Cni-MEX-II motifs on the X chromosome in *C. nigoni* and *C. briggsae*.** The ChIP-seq signal is plotted at the top 50 Cni-MEX-II motifs on the X chromosome in four species.



(c) **The DCC is bound at a few of the top 51-100 Cni-MEX-II motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni-MEX-II motifs on the X chromosome in four species.

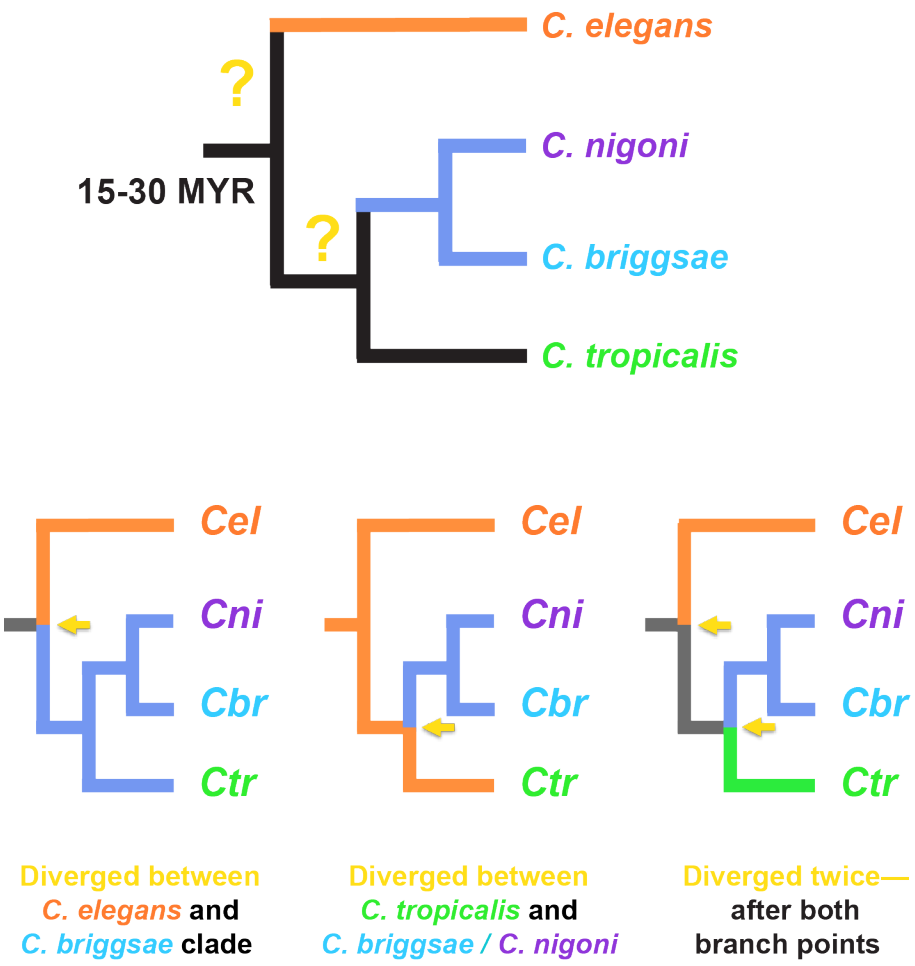
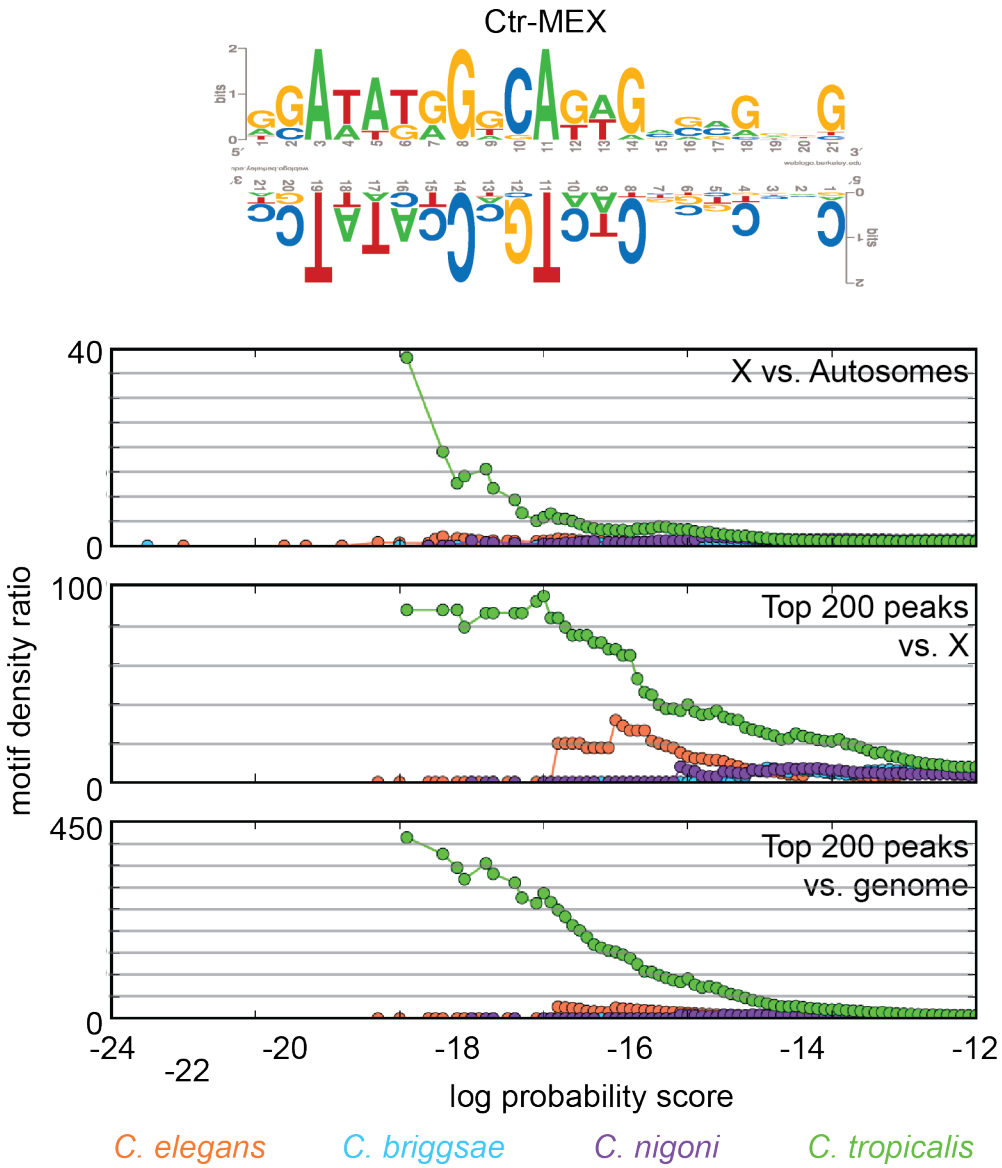
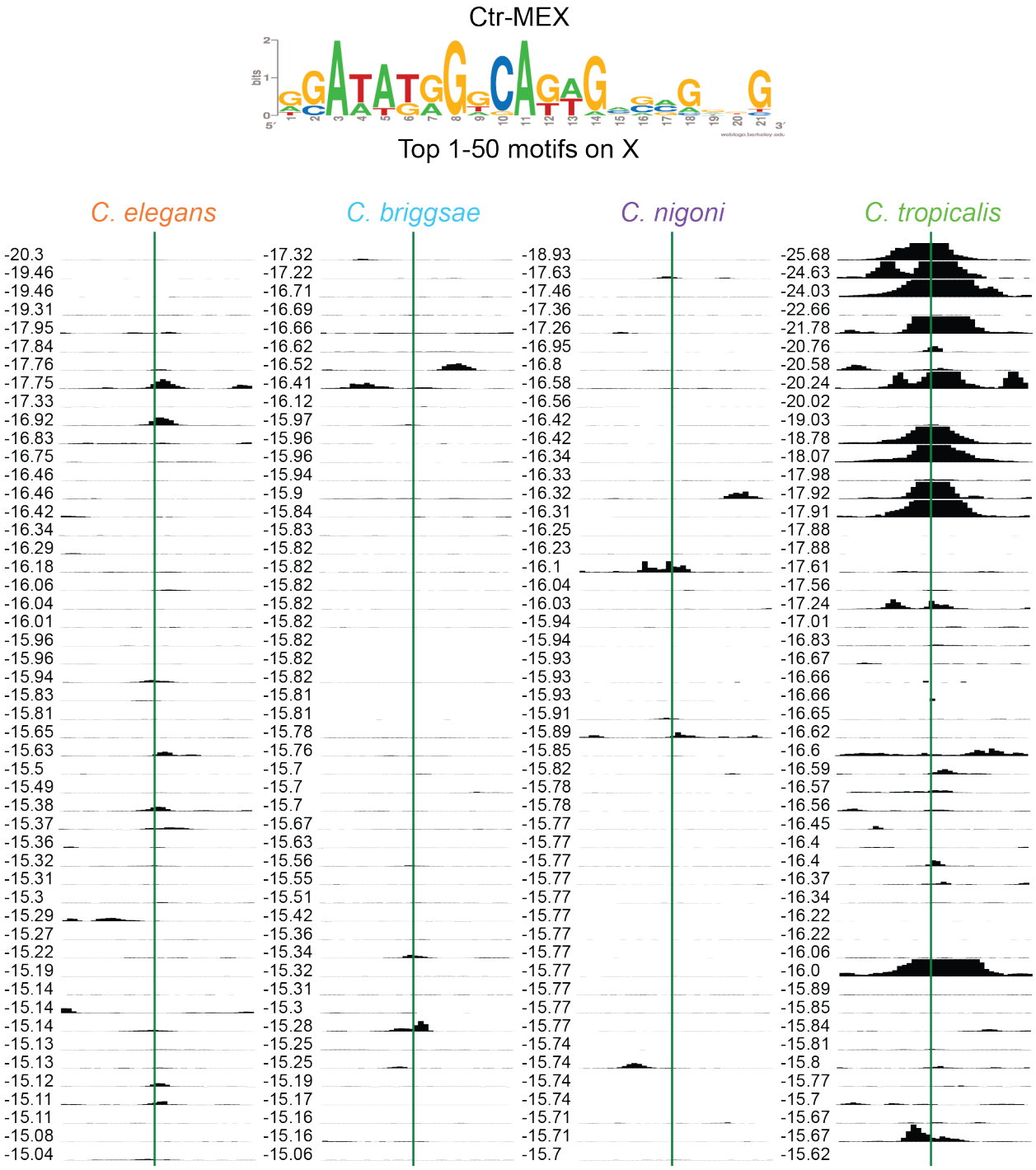


Figure 2.29: **DCC recruitment mechanisms diverged between *C. elegans* and *C. briggsae*. Analysis in *C. tropicalis* gives insight into the timeline of X-chromosome-targeting mechanism divergence.** If *C. tropicalis* X-targeting mechanisms were shared between *C. elegans* and *C. tropicalis*, it would indicate divergence in the *C. briggsae*/*C. nigoni* lineage. If *C. tropicalis* shared mechanisms with *C. briggsae*, it would indicate divergence in the *C. elegans* lineage or before *C. tropicalis* and *C. briggsae* split. Lastly, if *C. tropicalis* DCC-targeting to the X chromosome differed from both *C. elegans* and *C. briggsae*, it would indicate that multiple events changed DCC sequence-specificity and recruitment mechanisms evolve rapidly.

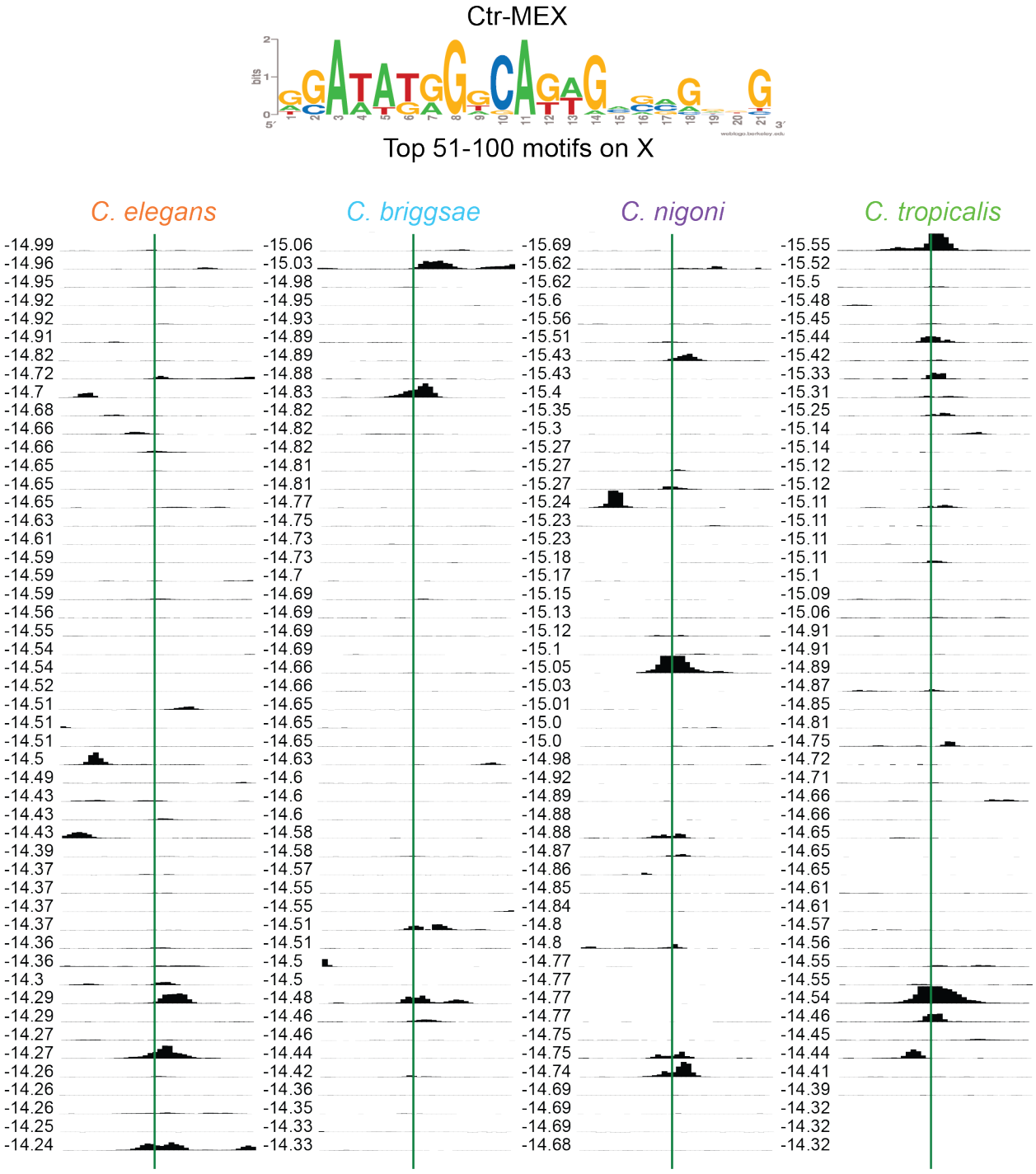
Figure 2.30: Ctr-MEX a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Ctr-MEX motif only X-enriched and peak-enriched in *C. tropicalis*.** The Ctr-MEX motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

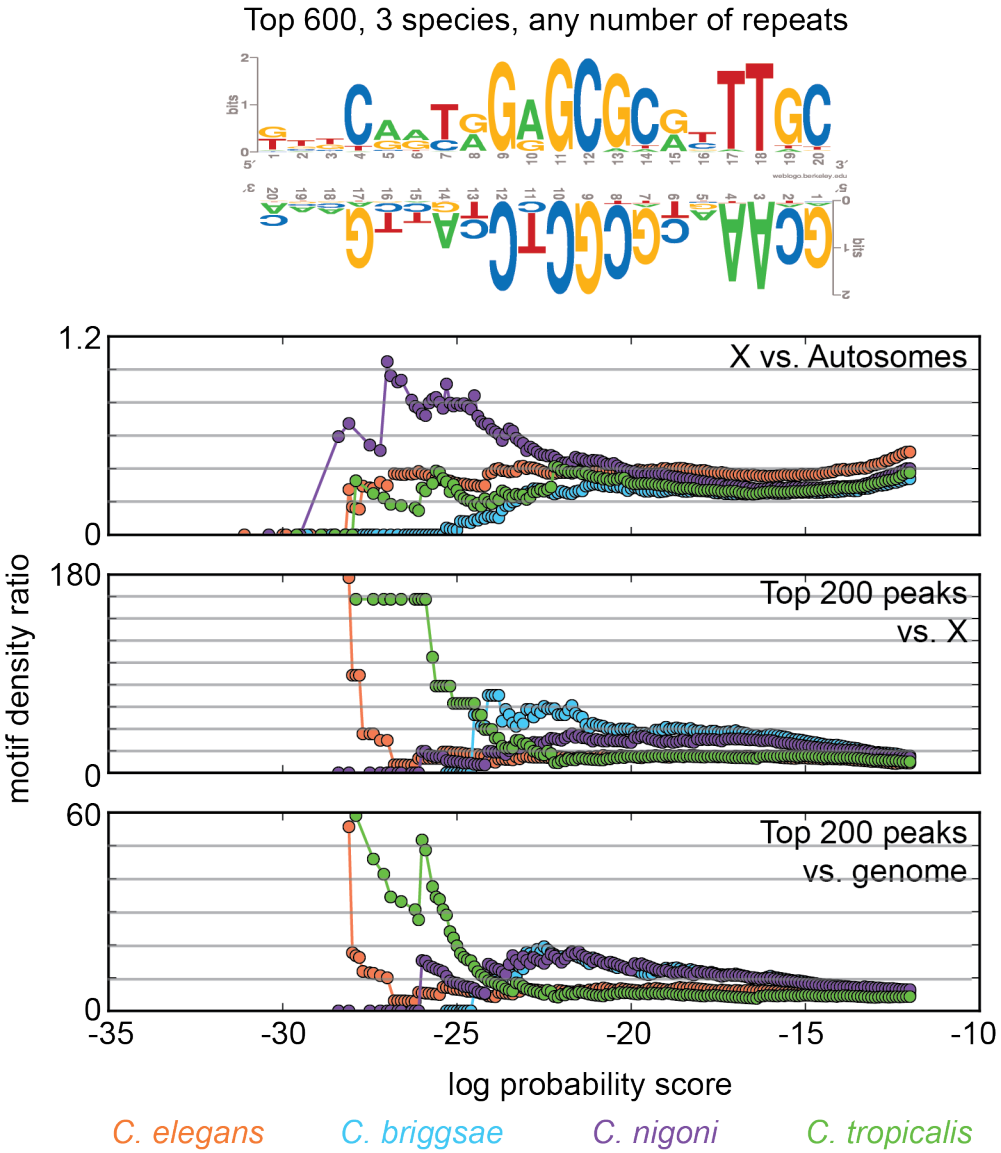


(b) **The DCC is bound at the strongest of the top 50 Ctr-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Ctr-MEX motifs on the X chromosome in four species.

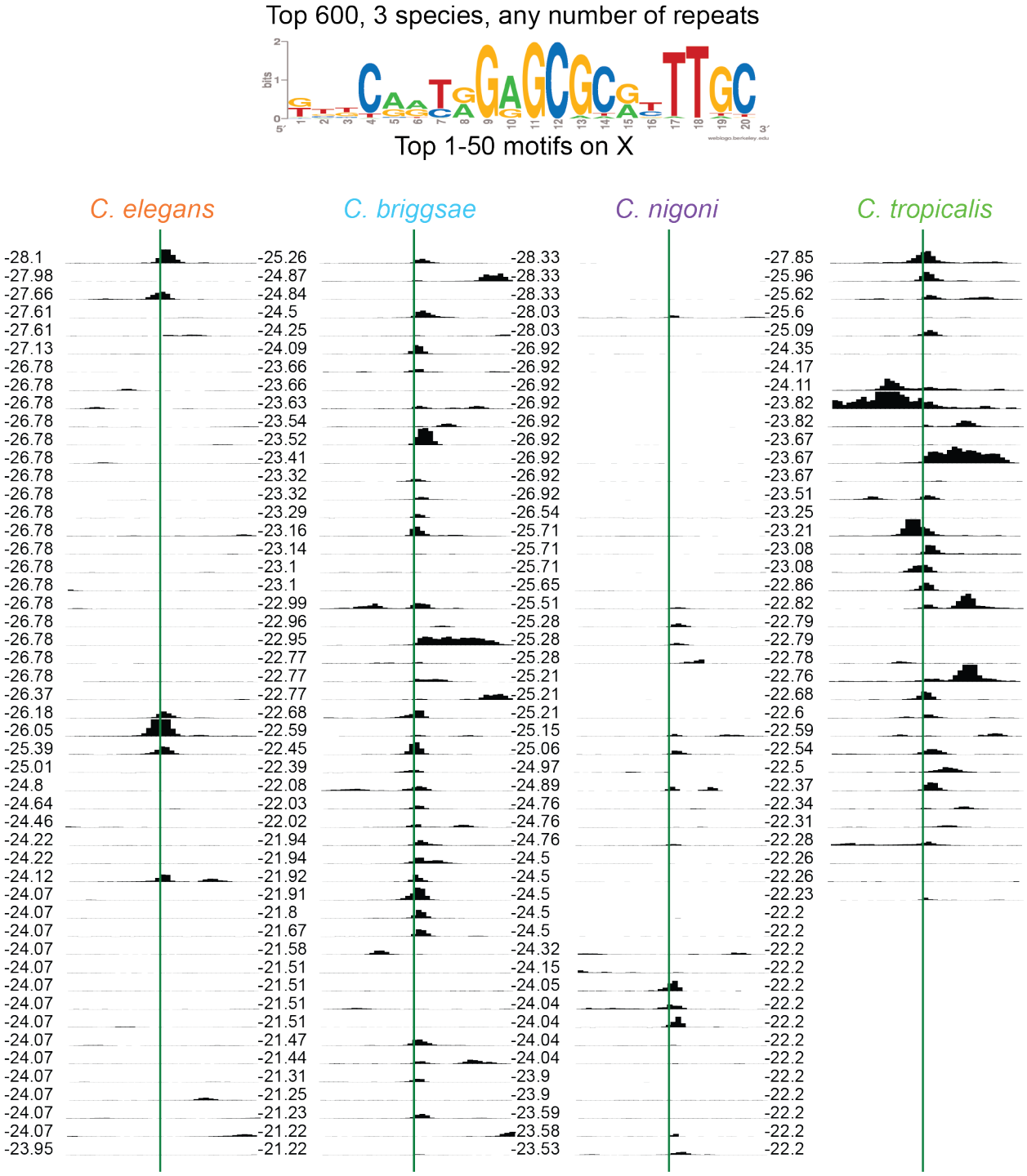


(c) **The DCC is bound at some of the top 50 Ctr-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Ctr-MEX motifs on the X chromosome in four species.

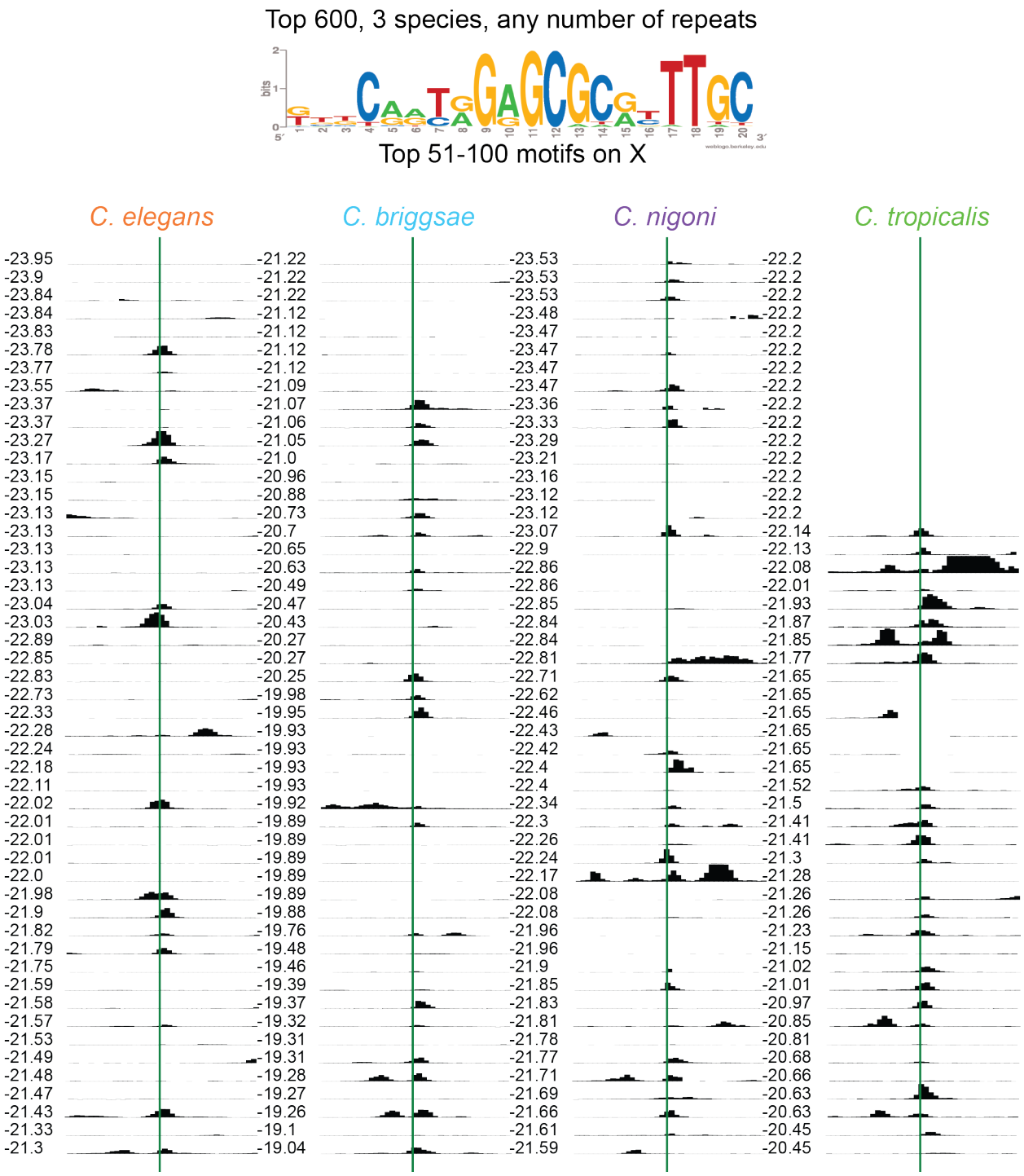
Figure 2.31: "Top 600" motif a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The "top600" motif is not X-enriched, but is peak-enriched in *C. elegans* and *C. tropicalis*.** The "top 600" motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.



(b) **The DCC is bound at many of the top 50 "top 600" motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 "top 600" motifs on the X chromosome in four species.



(c) **The DCC is bound at many of the top 51-100 "top 600" motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 "top 600" motifs on the X chromosome in four species.

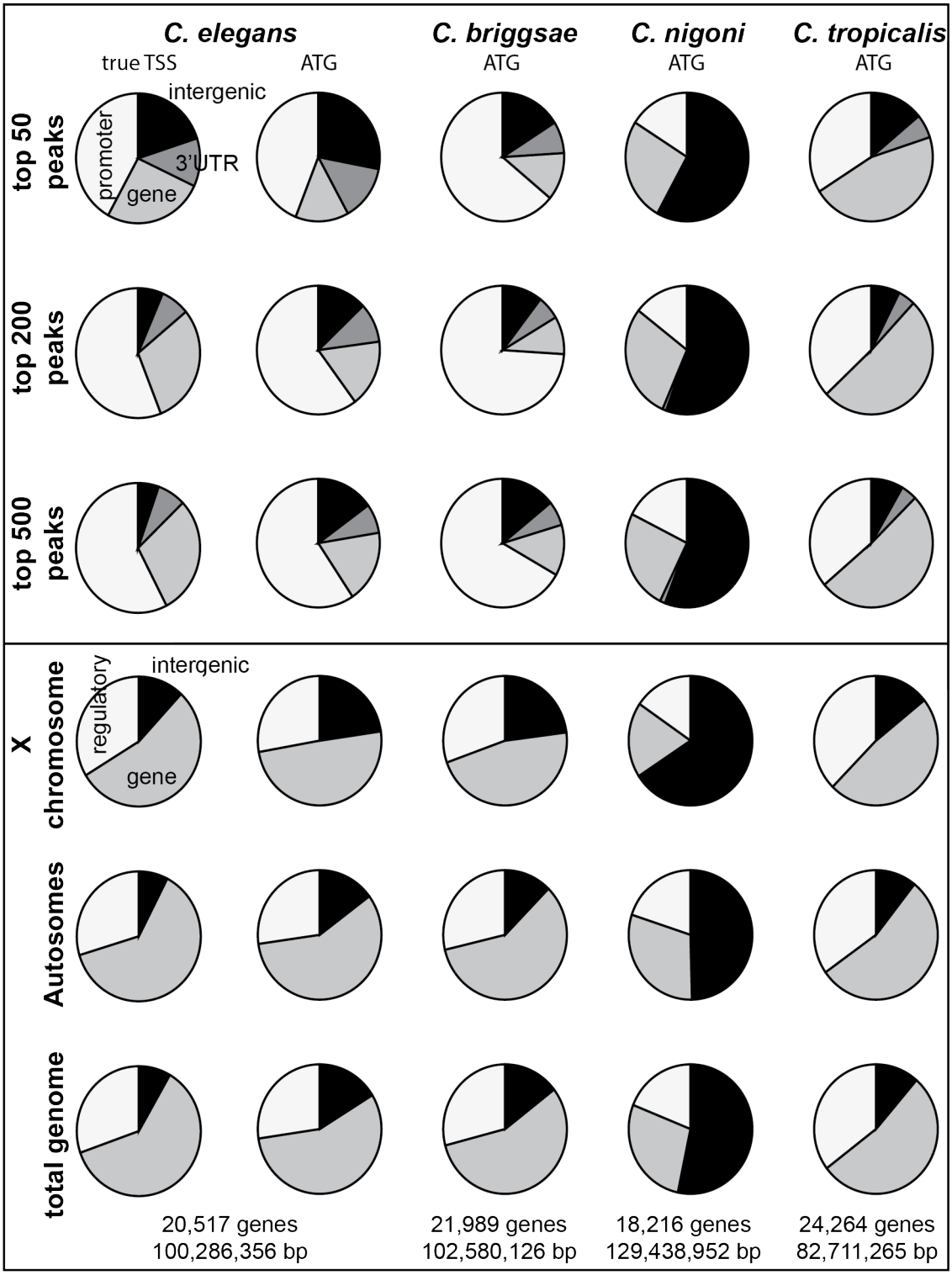


Figure 2.32: **Peak and genome composition** (Continued on the following page.)

Figure 2.32: **The distribution of DCC-bound peaks relative to gene loci and the genome composition and gene density have diverged in the *C. briggsae* clade.** The top three rows of pie charts show the distribution of the top 50, 200, and 500 DCC-binding site loci relative to genes. Promoters are defined as 2 kb upstream of the transcription start site (true TSS) or the start codon (ATG). The TSS calls were only available for *C. elegans*. The gene regions include both exons and introns (and also the 5' UTR for the true TSS calls). 3' untranslated regions (3' UTRs) are defined as 500 bp downstream of the stop codon. If the site was found in a region where the promoter and a 3'UTR overlapped, the distance from the peak summit to the gene was taken into account.

The next three rows of pie charts reflect the total proportion of the X chromosome, the autosomes, and the total genome found in these regions. *C. elegans* binding site preferences are not conserved across the *C. briggsae* clade.

C. elegans rex sites are often highly-occupied by the DCC and found in intergenic regions. We see enrichment of the top 50 peaks in intergenic regions compared to the X chromosome or the genome, but not the top 200 or 500 peaks, in *C. elegans*. This pattern was not observed in the other species. *C. elegans dox* sites are often found in promoters. Enrichment of DCC-binding peaks in promoters was also pronounced in *C. briggsae*, but not in *C. nigoni* or *C. tropicalis*. *C. nigoni* and *C. tropicalis* peak distributions resemble the overall genome proportions.

	C. elegans	
N2	reference strain	
TY5753	*dpy-27(y679)*	3X-FLAG insertion at amino acid 689
TY4573	*sdc-2(y74)* X; *yEx992[FLAG-sdc-2 + myo2::gfp]*	Array carrying *3X-FLAG::sdc-2* in an *sdc-2* mutant background
	C. briggsae	
AF16	reference strain	
TY5005	*Cbr-dpy-27(y436)*	Deletion (NHEJ) including entire first exon. Maintained as a heterozygote.
TY5773	*Cbr-dpy-27(y705)*	52 bp deletion (NHEJ, frameshift) starting at amino acid 689. Maintained as a heterozygote.
TY5774	*Cbr-dpy-27(y706)*	3X-FLAG insertion at amino acid 702
TY5775	*Cbr-sdc-2(y716)*	3X-FLAG insertion at amino acid 17
	C. nigoni	
JU1325	reference strain	
JU1422	inbred reference strain	
TY5754	*Cni-dpy-27(y683)*	3X-FLAG insertion at amino acid 702 in JU1325
TY5780	*Cni-dpy-27(y709)*	Insertion of 3X-FLAG followed by a stop codon at amino acid 702 in JU1325. Maintained as a heterozygote.
TY5586	*Cni-sdc-2(y516)*	14 bp deletion (NHEJ, frameshift) starting at amino acid 38 in JU1422. Maintained as a heterozygote.
	C. tropicalis	
JU1373	reference strain	
TY5743	*Ctr-sdc-2(y675)*	3X-FLAG insertion following start codon
TY5781	*Ctr-dpy-27(y677)*; *Ctr-sdc-2(y719)*	11025 bp deletion from amino acid 80 to 3316. Maintained as a heterozygote.

TY5752	*Ctr-dpy-27(y677)*	3X-FLAG insertion at amino acid 690
TY5771	*Ctr-dpy-27(y703)*	3X-FLAG followed by a stop codon insertion at amino acid 690. Maintained as a heterozygote.

Table 2.1: **Strain table**

Locus	
TALEN target sequences	
Cni-dpy-27 hinge	
T ACGGTAAAACAGTGAACGA	CCCCAAGAAACTTCCTGGC A
Cni-sdc-2 5' end	
T CACAAGAATCAGTGTCAGT	CACCAGCGCAATCCAAGGAG A
Ctr-dpy-27 hinge	
T TTGATAGGAATGTCAAGGAG	TGCGCCAGAGTCTTTGTG A
Ctr-sdc-2 exon 3	
T CTTCTGACATACCATTGGCT	CTGAAGGCCGGCTATTTCCC A
Locus	**Cas9 guide RNA sequence**
Cel-dpy-27 hinge	GCGCTCTGGAGTACGGTAAAA
rol-6	GTGAGACGTCAACAATATGG
Cbr-dpy-27 hinge	GTGATACTGTAGTAGCAACGG
Cbr-sdc-2 (near start)	ATCATCATCACAAGAATCAG
Cbr-rol-6	GTGAGACGTCAACAATACGG
Ctr-sdc-2 deletion (near start)	GTGATGGGCTCCTTGATTGG
Ctr-sdc-2 deletion (near end)	GATCATCGCCGAGATCAAAG
Ctr-rol-6	GTGAGACGTCAACAGTATGG

Table 2.2: **TALEN and Cas9 target sequences.**

Primer	Sequence
Cel-dpy-27(y679) 3X-FLAG insertion template	gaagctcccgcgcgcatcgttcttcgctctggagtacggtaaaacagta aacgatcaagctgcagctGACTACAAAGACCATGACG GTGATTATAAAGATCATGACATCGACTACA AGGATGACGATGACAAGgcagctgcacatatggataag ccaactaaacttcctggcaaacgcttgttcgacaaagtgcaatgca
rol-6 repair template	TGTGGGTTGATATGGTTAAACTTGGAGCAG GAACCGCTTCCAACCGTGTGcgctgcCAACAAT ATGGAGGATATGGAGCCACTGGTGTTCAGC CACCAGCACCAAC
Cbr-dpy-27(y706) 3X-FLAG insertion template	gaagctcccgcgcgcatcgttcttcgctctggagtacggtaaaacAgt aaacgatcaagctgcagctGACTACAAAGACCATGAC GGTGATTATAAAGATCATGACATCGACTAC AAGGATGACGATGACAAGgcagctgcacatatggata agccaactaaacttcctggcaaacgcttgttcgacaaagtgcaatgca
Cbr-sdc-2(y716) 3X-FLAG insertion template	CAGACAGCTCAGTGAATCATCATCACAAGA AgctgcagctGACTACAAAGACCATGACGGTGAT TATAAAGATCATGACATCGACTACAAGGAT GACGATGACAAGgcagctgcaTCAGTGGCAGTA ATTACTGTAGTTCCATCAC
Cbr-rol-6 repair template	GTGGGTTGATATGGTCAAGCTTGGAGCTGG AACCGCTTCAAACAGAGTGcgctgcCAACAATA CGGAGGATACGGAGCCAGTGGAGTTCAGCC ACCAGCACCAAC
Cni-dpy-27(y683) 3X-FLAG insertion template	GCTTCCTCGCGCATCGTTCTTCGCTCTGGAG TACGGTAAAACAGTGAACGATCAGCATgctgca gctGACTACAAAGACCATGACGGTGATTATAA AGATCATGACATCGACTACAAGGATGACGA TGACAAGgcagctgcaATGAATAACCCCAAGAAA CTTCCTGGCATGCGATTGTTCGACAAAGTG CACTGCAAGG
Cni-dpy-27(y709) premature stop plus 3X-FLAG insertion template	GAAGCTcCCgCGtGCATCGTTCTTCGCTtTGG AGTACGGTAAAACAGTGAACGATCAGCATgg taccGACTACAAAGACCATGACGGTGATTATA AAGATCATGACATCGACTACAAGGATGACG ATGACAAGttagtATGAATAACCCCAAGAAAC TTCCTGGCATGCGATTGTTCGACAAAGTGC ACTGCAAGGATCC

Primer	Sequence
Ctr-dpy-27(y677) 3X-FLAG insertion template	GATGCAAAGATCCGCTTTCTGTTTTGTGGA GTTTGATAGGAATGTCAAGGAGTACAGAgctg cagctGACTACAAAGACCATGACGGTGATTAT AAAGATCATGACATCGACTACAAGGATGAC GATGACAAGgcagctgcaATGAATGTTGCGCCA GAGTCTTTGTGAGTTTTTGGAGGCCTTTTT TGGGAGGGGATG
Ctr-dpy-27(y703) premature stop plus 3X-FLAG insertion template	GATGCAAAGATCCGCTTTCTGTTTTGTGGA GTTTGATAGGAATGTCAAGGAGTACAGAGG TACCGACTACAAAGACCATGACGGTGATTA TAAAGATCATGACATCGACTACAAGGATGA CGATGACAAGTTAGTATGAATGTTGCGCCA GAGTCTTTGTGAGTTTTTGGAGGCCTTTTT TGGGAGGGGATG
Ctr-sdc-2(y675) 3X-FLAG insertion template	ATTAATTTCATCTAACATGTTTCAGATTGTA AACCCTATGTGACTTGCACAACAATGGACTA CAAAGACCATGACGGTGATTATAAAGATCA TGACATCGACTACAAGGATGACGATGACAA GGCTAAAAAGAAGTTTCGTCAGACCGGCGG GTGTGCGAAAAACGAAGTGATCACACTGTC GCCAGAGGAACTATTTAT
Ctr-sdc-2(y719) 11kb deletion template	CCAAACAAGTCAAAGAGACAATCTTAAACA CAGGTACCGATCTCGGCGATGATCGAAATG CCACGCTC
Ctr-rol-6 repair template	gttaaacttggagccggaaccgcttcaaacagagtGCgCTgCcaG caAtaCggaggatatggagccagtggagttcagccacc

Table 2.3: **Genome editing repair templates.**

***Cbr-dpy-27(y705)* (strain TY5773)**							
Parent (♀)		Embryos per brood (SEM)	% ♀(SEM)	% ♂(SEM)	% Dead (SEM)	% Affected (SEM)	Number of broods
Cbr-dpy-27 +/+		239 (30.8)	92 (4.4)	< 1 (0.2)	7 (4.4)	< 1 (0.1)	6
Cbr-dpy-27 +/-		220 (32.3)	69 (2.7)	< 1 (0.2)	17 (2.6)	13 (1.7)	8
Cbr-dpy-27 -/-		0	0	0	0	0	5
***Ctr-dpy-27(y703)* (strain TY5752)**							
Parent (♀)			% ♀	% ♂			
Ctr-dpy-27 +/+		164 (14)	101 (1.8)	-1 (1.8)	-1 (1.8)	*leq* 1 (0.1)	7
Ctr-dpy-27 +/-		189 (11.5)	75 (1)	< 1 (0.1)	9 (1.2)	16 (1.4)	11
Ctr-dpy-27 -/-		0	0	0	0	0	5
***Cni-dpy-27(y709)* (strain TY5780) and *C. nigoni* (reference strain JU1325)**							
Parents: ♀	x ♂		% ♀	% ♂			
Cni-dpy-27 +/+	+/+	135	13	8	61	18	1
Cni-dpy-27 +/+	-/-	13 (13)	0	0	96	4	2
Cni-dpy-27 +/-	+/+	117 (18)	19 (10.4)	23 (11.5)	47 (20.9)	12 (1)	2
Cni-dpy-27 +/-	+/-	172 (38.6)	17 (5.3)	19 (5.9)	42 (13.5)	22 (11.5)	8
Cni-dpy-27 +/-	-/-	128 (65.8)	17 (2.2)	27 (2.9)	49 (6.4)	6 (2)	5
JU1325	*Cni-dpy-27* +/- or -/-	355 (89.5)	45 (1.9)	42 (2.1)	5 (1.8)	7 (2.3)	11
JU1325		310 (66.9)	39 (3)	45 (11.2)	9 (10.3)	7 (1.6)	9
JU1325 (fresh thaw)		479 (34.3)	43 (4.7)	42 (4)	9 (5.8)	7 (3)	7

Table 2.4: ***dpy-27* ortholog mutations cause defects consistent with conserved function in the *C. briggsae* clade.** (Continued on the following page.)

Table 2.4: ***dpy-27* ortholog mutations cause defects consistent with conserved function in the *C. briggsae* clade.**

Phenotypes of *Cbr-dpy-27(y705)*, *Ctr-dpy-27(y703)*, and *Cni-dpy-27(y709)* progeny were assessed as follows. Total number of embryos per brood were averaged, as were the percent non-dumpy (separated by sex), percent that failed to hatch, and percent affected, meaning dumpy or small. "n" refers to the number of broods. These strains were maintained as heterozygotes (no balancers are available for these species).

Similar to *C. elegans dpy-27*, *Cbr-dpy-27(y705)*/+ and *Ctr-dpy-27(y703)*/+ brood counts were consistent with recessive lethality and dumpiness (30.2% and 25.1% dumpy or dead, respectively). However, unlike *C. elegans dpy-27* mutants, maternal contributions do not rescue homozygotes with heterozygous mothers in these species. Homozygous *dpy-27* mutant hermaphrodites were sterile (no embryos were counted among homozygous hermaphrodites analyzed). Few males were produced in selfed hermaphrodites.

C. nigoni dpy-27 phenotype analysis was complicated by non-specific lethality. A cross between siblings wild-type at the *Cni-dpy-27* locus produced 135 embryos, of which only 21 percent were phenotypically normal. Nearly half of the progeny from heterozygous *Cni-dpy-27(y709)* mothers died. High levels of lethality and progeny with mutant phenotypes did not correlate with parental *Cni-dpy-27* genotype. Crossing heterozygous and/or mutant dads to strain JU1325 reduced the percent lethal and percent affected to wild-type levels. Also, a freshly-thawed wild-type strain produced more embryos on average than a strain that was propagated in the laboratory for months, but egg-to-adult viability was the same (both of these produced about 9% dead embryos and 7% with mutant phenotypes).

Table 2.5: Recruitment assays

C. briggsae rex DNA									
start	end	name	motifs (ln(P) score)	size	type	Cbr recruitment		Cel recruitment	
10778971	10781620	*Cbr-rex-01*	Cbr-MEX (-15.57, -15.57, -14.63, -14.47) Cbr-MEX-II (-27.58)	500	PCR	yes	54/59		
12639963	12642242	*Cbr-rex-02*	Cbr-MEX (-14.39) Cbr-MEX-II (-22.76)	2279	PCR	yes	91/101		
19468419	19469368	*Cbr-rex-03*	Cbr-MEX-II (-12.36, -20.04) Cni-MEX (-13.65) Cni-MEX-II (-13.2, -13.43) Ctr-MEX (-13.65)	461	PCR	yes	65/74		
6357847	6359296	*Cbr-rex-04*	Cbr-MEX (-13.8) Cbr-MEX-II (-19.09) Cni-MEX (-15.45) Cni-MEX-II (-16.3)	1449	PCR	yes	58/68	yes	19/19
3152854	3153354	*Cbr-rex-05*	Cbr-MEX-II (-18.98) Cni-MEX-II (-13.45, -12.51)	500	PCR	yes	44/45		
18811174	18811674	*Cbr-rex-06*	Cbr-MEX (-13.35) Cbr-MEX-II (-15.43) Cni-MEX-II (-14.13)	500	PCR	yes	50/68		
8026271	8026771	*Cbr-rex-07*	Cbr-MEX (-12.26, -18.72) Cni-MEX (-18.84) Cni-MEX-II (-14.93, -17.52)	500	PCR	yes	63/65		
16590530	16590979	*Cbr-rex-08*	Ctr-MEX (-13.52)	500	PCR	yes	19/52	no	0/27
3135115	3135615	*Cbr-rex-09*	Cbr-MEX (-12.8)	500	PCR	yes	53/62		
895711	895896	*Cbr-rex-10*	–	185	PCR	yes	44/55		
4562615	4563115	*Cbr-rex-11*	–		PCR	yes	48/54		
19564735	19565235	*Cbr-rex-12*	Cni-MEX (-14.67)	500	PCR	yes	61/77		
C. briggsae test DNA (flat regions)									
8040565	8042583	flat 1	Cel-MEX (-17.56)	2018	PCR	no	0/66,0/25		
5887364	5889456	flat 2	Cel-MEX (-16.93)	2092	PCR	no	0/55		
12488134	12490210	flat 3	Cbr-MEX-II (-17.53)	2076	PCR	no	0/83,0/98		
20917253	20919096	flat 4	Cbr-MEX-II (-19.4)	1843	PCR	no	0/69,0/75		
11761774	11764047	*Cbr-mom-1* (syntenic to *rex-33*)	–	2274	TOPO	no	3/48	no	3/78

Table 2.5: Recruitment assays

11761774	11764047	*Cbr-mom-1* (syntenic to *rex-33*)	–	2274	TOPO	no	0/55	no	3/55
4958820	4961920	*Cbr-sdc-2* promoter (syntenic to *rex-4*)	Cel-MEX (-15.8)	3101	TOPO	no*	3/75	no	21/75
18463282	18464899	CBG07595 (syntenic to *rex-39*)	Cel-MEX-II (-14.3, -12.6)	1618	TOPO	no	0/46	no	20/97
			C. briggsae test DNA (modified peak sequences)						
10779110	10779299	*Cbr-rex-01*	Cbr-MEX-II (-27.58)	190	oligo	no*	7/48,8/50,0/66		
10779110	10779299	*Cbr-rex-01* minus Cbr-MEX-II	–	160	oligo	no	7/165,0/58		
10780211	10780710	*Cbr-rex-01*	Cbr-MEX (-15.57, -15.57, -14.63, -14.47) Cbr-MEX-II (-27.58)	500	oligo	yes	65/73		
10780211	10780710	*Cbr-rex-01* scrambled 5 motifs	–	500	oligo	no	3/50		
10780211	10780710	*Cbr-rex-01* scrambled 4 Cbr-MEX	Cbr-MEX-II (-27.58)	500	oligo	yes*	26/68		
10780211	10780710	*Cbr-rex-01* scrambled Cbr-MEX-II	Cbr-MEX (-15.57, -15.57, -14.63, -14.47)	500	oligo	yes*	16/66		
12641154	12641341	*Cbr-rex-02* "Small peak"	Cbr-MEX-II (-22.76)	190	oligo	yes	56/81,15/27		
12641154	12641341	*Cbr-rex-02* minus Cbr-MEX-II	–	160	oligo	no	0/30,1/80		
12642606	12643183	*Cbr-rex-02*	Cbr-MEX (-14.39) Cbr-MEX-II (-22.76)	577	PCR	yes	36/85,71/81, 45/50		
12642606	12643183	*Cbr-rex-02* minus Cbr-MEX-II	Cbr-MEX (-14.39)	547	PCR	yes*	104/211		
12642606	12643183	*Cbr-rex-02*	Cbr-MEX (-14.39) Cbr-MEX-II (-22.76)	577	TOPO	yes*	4/69,23/37, 15/92,167/183		
12642606	12643183	*Cbr-rex-02* minus Cbr-MEX-II	Cbr-MEX (-14.39)	547	TOPO	yes*	54/54,32/52, 102/122,26/75		
12642606	12642988	*Cbr-rex-02*	Cbr-MEX (-14.39) Cbr-MEX-II (-22.76)	495	oligo	yes	43/43		
12642606	12642988	*Cbr-rex-02* scrambled 2 motifs	–	493	oligo	no*	11/37		
12642606	12642988	*Cbr-rex-02* scrambled Cbr-MEX	Cbr-MEX-II (-22.76)	493	oligo	no*	6/45		
12642606	12642988	*Cbr-rex-02* scrambled Cbr-MEX-II	Cbr-MEX (-14.39)	495	oligo	no*	6/22		

Table 2.5: Recruitment assays

18819889	18820486	TLpeak16	–	598	PCR	no	0/128,0/37		
18819889	18820486	TLpeak16 minus potential motif	–	568	PCR	no	0/59,0/73		
			C. elegans rex DNA						
start	end	name	motifs (ln(P) score)	size	type	Cbr recruitment		Cel recruitment	
11360041	11362399	*Cel-rex-3* (F42E11.1)	Cel-MEX (-14.72)	2358	TOPO	no	11/77	yes	12/14,
11360041	11362399	*Cel-rex-3* mutant (F42E11.1)	–	2358	TOPO	no	0/48	yes*	23/41
11520647	11522647	*Cel-rex-4* (*sdc-2* promoter)	Cel-MEX (-15.8)	2001	TOPO	no	1/116	yes	16/16
2996004	2998096	*Cel-rex-32*	Cel-MEX (-17.65, -18.97, -18.97) Cel-MEX-II (-21.89, -12.4, -12.15)	2092	TOPO	yes	51/58	yes	45/45
6295287	6297381	*Cel-rex-33*	Cel-MEX (-15.46, -15.45, -13.23)	2094	TOPO	no	0/53	yes	63/63
14812297	14814299	*Cel-rex-39* (T04C10.3)	Cel-MEX-II (-20.85, -21.3)	2003	TOPO	no	11/52	yes	54/56
			C. nigoni peak DNA						
start	end	name	motifs (ln(P) score)	size	type	Cbr recruitment		Cel recruitment	
13067270	13069325	*C. nigoni* peak 11	Cni-MEX (-18.25, -18.86) Cni-MEX-II (-22.87, -16.23) Cbr-MEX (-16.07) Cbr-MEX-II (-14.53)	2055	PCR	yes		yes	43/47
15784116	15786449	*C. nigoni* peak 1	Cni-MEX (-22.02, -15.05, -13.39) Cni-MEX-II (-12.98, -15.38) Cbr-MEX (-13.95) Cbr-MEX-II (-22.54)	2333	PCR			yes	

Table 2.5: Recruitment assays

Table 2.5: **An *in vivo* functional recruitment assay identified *Cbr-rex* sites and motif sequences that contribute to DCC recruitment in *C. briggsae*.** This table shows the X-chromosome location, name of site, motif scores, fragment size, source of DNA injected (PCR product, synthesized double-stranded oligo, or PCR fragment cloned in a TOPO vector), and recruitment results in *C. briggsae* and/or *C. elegans*. Recruitment is reported as a "yes" or "no", with an asterisk if the level of recruitment was intermediate or variable across lines. Raw counts are included as a ratio (number of nuclei that recruit the DCC over total scored) for each transgenic line. Twelve *C. briggsae* DCC-binding sequences were shown to be *Cbr-rex* sites. The *Cbr-rex-04* sequence also recruits the *C. elegans* DCC. *Cbr-rex-01* and *Cbr-rex-02* sequences that were modified to remove or scramble the Cbr-MEX and/or Cbr-MEX-II motifs were less able to recruit the *C. briggsae* DCC. Four *C. briggsae* "flat" regions were tested for recruitment. These X-chromosome loci were called "flat" regions because they had background levels of ChIP-seq signal. Two contained a single strong Cel-MEX-II motif and two had a single strong Cbr-MEX-II motif. These also failed to recruit the *C. briggsae* DCC in the *in vivo* assay. *C. elegans rex* sites and *C. briggsae* regions homologous to *C. elegans rex* sites were also tested. Of these, only *C. elegans rex-32* was able to recruit the *C. briggsae* DCC. Two strong *C. nigoni* DCC-binding sites recruit the *C. elegans* DCC. One of these was also tested in *C. briggsae*; it recruited the *C. briggsae* DCC as well.

References

[1] B Charlesworth. "The evolution of chromosomal sex determination and dosage compensation." In: *Current biology : CB* 6.2 (Feb. 1996), pp. 149–162.

[2] L M Miller et al. "xol-1: a gene that controls the male modes of both sex determination and X chromosome dosage compensation in C. elegans." In: *Cell* 55.1 (Oct. 1988), pp. 167–183.

[3] C Nusbaum and B J Meyer. "The Caenorhabditis elegans gene sdc-2 controls sex determination and dosage compensation in XX animals." In: *Genetics* 122.3 (July 1989), pp. 579–593.

[4] John M Gladden and Barbara J Meyer. "A ONECUT homeodomain protein communicates X chromosome dose to specify Caenorhabditis elegans sexual fate by repressing a sex switch gene." In: *Genetics* 177.3 (Nov. 2007), pp. 1621–1637.

[5] John Gately Luz et al. "XOL-1, primary determinant of sexual fate in C. elegans, is a GHMP kinase family member and a structural prototype for a class of developmental regulators." In: *Genes & development* 17.8 (Apr. 2003), pp. 977–990.

[6] Diana S Chu et al. "A molecular link between gene-specific and chromosome-wide transcriptional repression." In: *Genes & development* 16.7 (Apr. 2002), pp. 796–805.

[7] Barbara J Meyer. "Targeting X chromosomes for repression." In: *Current opinion in genetics & development* 20.2 (Apr. 2010), pp. 179–189.

[8] Gyorgyi Csankovszki et al. "Three distinct condensin complexes control C. elegans chromosome dynamics." In: *Current biology : CB* 19.1 (Jan. 2009), pp. 9–19.

[9] A V Strunnikov and R Jessberger. "Structural maintenance of chromosomes (SMC) proteins: conserved molecular properties for multiple biological functions." In: *European journal of biochemistry / FEBS* 263.1 (July 1999), pp. 6–13.

[10] Kim Nasmyth and Christian H Haering. "The structure and function of SMC and kleisin complexes." In: *Annual review of biochemistry* 74 (2005), pp. 595–648.

[11] P T Chuang, D G Albertson, and B J Meyer. "DPY-27:a chromosome condensation protein homolog that regulates C. elegans dosage compensation through association with the X chromosome." In: *Cell* 79.3 (Nov. 1994), pp. 459–474.

[12] J D Lieb et al. "MIX-1: an essential component of the C. elegans mitotic machinery executes X chromosome dosage compensation." In: *Cell* 92.2 (Jan. 1998), pp. 265–277.

[13] Emily Crane et al. "Condensin-driven remodelling of X chromosome topology during dosage compensation." In: *Nature* 523.7559 (July 2015), pp. 240–244.

[14] H E Dawes et al. "Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate." In: *Science (New York, N.Y.)* 284.5421 (June 1999), pp. 1800–1804.

[15] Rebecca R Pferdehirt, William S Kruesi, and Barbara J Meyer. "An MLL/COMPASS subunit functions in the C. elegans dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression." In: *Genes & development* 25.5 (Mar. 2011), pp. 499–515.

[16] P M Meneely and W B Wood. "An autosomal gene that affects X chromosome expression and sex determination in Caenorhabditis elegans." In: *Genetics* 106.1 (Jan. 1984), pp. 29–44.

[17] A M Villeneuve and B J Meyer. "The role of sdc-1 in the sex determination and dosage compensation decisions in Caenorhabditis elegans." In: *Genetics* 124.1 (Jan. 1990), pp. 91–114.

[18] M L Nonet and B J Meyer. "Early aspects of Caenorhabditis elegans sex determination and dosage compensation are regulated by a zinc-finger protein." In: *Nature* 351.6321 (May 1991), pp. 65–68.

[19] Patrick McDonel et al. "Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex." In: *Nature* 444.7119 (Nov. 2006), pp. 614–618.

[20] Judith Jans et al. "A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome." In: *Genes & development* 23.5 (Mar. 2009), pp. 602–618.

[21] Elie S Dolgin et al. "Inbreeding and outbreeding depression in Caenorhabditis nematodes." In: *Evolution; international journal of organic evolution* 61.6 (June 2007), pp. 1339–1352.

[22] Karin C Kiontke et al. "A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits." In: *BMC evolutionary biology* 11.1 (2011), p. 339.

[23] Te-Wen Lo et al. "Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions." In: *Genetics* 195.2 (Oct. 2013), pp. 331–348.

[24] Qing Wei et al. "Rapid creation of forward-genetics tools for C. briggsae using TALENs: lessons for nonmodel organisms." In: *Molecular Biology and Evolution* 31.2 (Feb. 2014), pp. 468–473.

[25] Sergey Koren et al. “Hybrid error correction and de novo assembly of single-molecule sequencing reads.” In: *Nature biotechnology* 30.7 (July 2012), pp. 693–700.

[26] Konstantin Berlin et al. “Assembling large genomes with single-molecule sequencing and locality-sensitive hashing.” In: *Nature biotechnology* 33.6 (June 2015), pp. 623–630.

[27] Mario Stanke and Stephan Waack. “Gene prediction with a hidden Markov model and a new intron submodel.” In: *Bioinformatics (Oxford, England)* 19 Suppl 2 (Oct. 2003), pp. ii215–25.

[28] Weizhong Li et al. “The EMBL-EBI bioinformatics web and programmatic tools framework.” In: *Nucleic Acids Research* 43.W1 (July 2015), W580–4.

[29] Sarah Elizabeth Albritton et al. “Sex-Biased Gene Expression and Evolution of the X Chromosome in Nematodes”. In: *Genetics* (July 2014), pp. 1–27.

[30] Erin L Doyle et al. “TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction.” In: *Nucleic acids research* 40.Web Server issue (July 2012), W117–22.

[31] Tomas Cermak et al. “Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.” In: *Nucleic acids research* 39.12 (July 2011), e82.

[32] Jana Streubel et al. “TAL effector RVD specificities and efficiencies.” In: *Nature biotechnology* 30.7 (July 2012), pp. 593–595.

[33] Behnom Farboud and Barbara J Meyer. “Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design.” In: *Genetics* 199.4 (Apr. 2015), pp. 959–971.

[34] Alexandre Paix et al. “High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes.” In: *Genetics* 201.1 (Sept. 2015), pp. 47–54.

[35] Thomas Eng, Vincent Guacci, and Douglas Koshland. “Interallelic complementation provides functional evidence for cohesin-cohesin interactions on DNA.” In: *Molecular biology of the cell* 26.23 (Nov. 2015), pp. 4224–4235.

[36] Ben Langmead. “Aligning short sequencing reads with Bowtie.” In: *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* Chapter 11 (Dec. 2010), Unit 11.7.

[37] Heng Li et al. “The Sequence Alignment/Map format and SAMtools.” In: *Bioinformatics (Oxford, England)* 25.16 (Aug. 2009), pp. 2078–2079.

[38] Yong Zhang et al. “Model-based analysis of ChIP-Seq (MACS).” In: *Genome biology* 9.9 (2008), R137.

[39] T L Bailey and C Elkan. "Fitting a mixture model by expectation maximization to discover motifs in biopolymers." In: *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology* 2 (1994), pp. 28–36.

[40] Timothy L Bailey et al. "MEME SUITE: tools for motif discovery and searching." In: *Nucleic acids research* 37.Web Server issue (July 2009), W202–8.

[41] H Chen et al. "IVE (Image Visualization Environment): a software platform for all three-dimensional microscopy applications." In: *Journal of structural biology* 116.1 (Jan. 1996), pp. 56–60.

[42] J D Plenefisch, L DeLong, and B J Meyer. "Genes that implement the hermaphrodite mode of dosage compensation in Caenorhabditis elegans." In: *Genetics* 121.1 (Jan. 1989), pp. 57–76.

[43] Yiqing Guo, Shirley Lang, and Ronald E Ellis. "Independent recruitment of F box genes to regulate hermaphrodite development during nematode evolution." In: *Current biology : CB* 19.21 (Nov. 2009), pp. 1853–1860.

[44] Joanna L Kozlowska et al. "Genetic variation for postzygotic reproductive isolation between Caenorhabditis briggsae and Caenorhabditis sp. 9." In: *Evolution; international journal of organic evolution* 66.4 (Apr. 2012), pp. 1180–1195.

[45] Gavin C Woodruff et al. "Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of Caenorhabditis nematodes." In: *Genetics* 186.3 (Nov. 2010), pp. 997–1012.

[46] Lawrence A Kelley et al. "The Phyre2 web portal for protein modeling, prediction and analysis". In: *Nature Protocols* 10.6 (May 2015), pp. 845–858.

[47] John C Lucchesi and Mitzi I Kuroda. "Dosage compensation in Drosophila." In: *Cold Spring Harbor Perspectives in Biology* 7.5 (May 2015), a019398.

[48] André Pires-daSilva and Ralf J Sommer. "Conservation of the global sex determination gene tra-1 in distantly related nematodes." In: *Genes & development* 18.10 (May 2004), pp. 1198–1208.

[49] Sevinc Ercan et al. "X chromosome repression by localization of the C. elegans dosage compensation machinery to sites of transcription initiation." In: *Nature Genetics* 39.3 (Mar. 2007), pp. 403–408.

[50] William S Kruesi et al. "Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation." In: *eLife* 2 (2013), e00808.

[51] Cristel G Thomas et al. "Simplification and desexualization of gene expression in self-fertile nematodes." In: *Current biology : CB* 22.22 (Nov. 2012), pp. 2167–2172.

[52] Janna L Fierst et al. "Reproductive Mode and the Evolution of Genome Size and Structure in Caenorhabditis Nematodes". In: *PLOS Genetics* 11.6 (June 2015), e-1005323–25.

[53] John Wang et al. "Chromosome size differences may affect meiosis and genome size." In: *Science (New York, N.Y.)* 329.5989 (July 2010), pp. 293–293.

[54] Suhas S P Rao et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." In: *Cell* 159.7 (Dec. 2014), pp. 1665–1680.

[55] Xinxian Deng et al. "Bipartite structure of the inactive mouse X chromosome." In: *Genome biology* 16.1 (2015), p. 152.

[56] Luca Giorgetti et al. "Structural organization of the inactive X chromosome in the mouse." In: *Nature* 535.7613 (July 2016), pp. 575–579.

[57] Elphège P Nora et al. "Spatial partitioning of the regulatory landscape of the X-inactivation centre." In: *Nature* 485.7398 (May 2012), pp. 381–385.

[58] Luca Giorgetti et al. "Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription." In: *Cell* 157.4 (May 2014), pp. 950–963.

[59] François Le Dily et al. "Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation." In: *Genes & development* 28.19 (Oct. 2014), pp. 2151–2162.

[60] Peter G Okkema and Michael Krause. "Transcriptional regulation." In: *WormBook : the online review of C. elegans biology* (2005), pp. 1–40.

[61] Jeb Gaudet and James D McGhee. "Recent advances in understanding the molecular mechanisms regulating C. elegans transcription." In: *Developmental dynamics : an official publication of the American Association of Anatomists* 239.5 (May 2010), pp. 1388–1404.

[62] L V Goodrich et al. "Conservation of the hedgehog/patched signaling pathway from flies to mice: induction of a mouse patched gene by Hedgehog." In: *Genes & development* 10.3 (Feb. 1996), pp. 301–312.

[63] Franck Pichaud and Claude Desplan. "Pax genes and eye organogenesis." In: *Current opinion in genetics & development* 12.4 (Aug. 2002), pp. 430–434.

[64] Veronica F Hinman et al. "Developmental gene regulatory network architecture across 500 million years of echinoderm evolution." In: *Proceedings of the National Academy of Sciences of the United States of America* 100.23 (Nov. 2003), pp. 13356–13361.

[65] Lucia Ciglar and Eileen E M Furlong. "Conservation and divergence in developmental networks: a view from Drosophila myogenesis." In: *Current opinion in cell biology* 21.6 (Dec. 2009), pp. 754–760.

[66] Feng Yue et al. "A comparative encyclopedia of DNA elements in the mouse genome." In: *Nature* 515.7527 (Nov. 2014), pp. 355–364.

[67] K R Nitta et al. "Conservation of transcription factor binding specificities across 600 million years of bilateria evolution". In: *eLife* (2015).

[68] Arttu Jolma et al. "DNA-binding specificities of human transcription factors." In: *Cell* 152.1-2 (Jan. 2013), pp. 327–339.

[69] Trevor R Sorrells et al. "Intersecting transcription networks constrain gene regulatory evolution." In: *Nature* 523.7560 (July 2015), pp. 361–365.

[70] Vaishnavi Ragavapuram, Emily Elaine Hill, and Scott Everet Baird. "Suppression of F1 Male-Specific Lethality in Caenorhabditis Hybrids by cbr-him-8." In: *G3 (Bethesda, Md.)* 6.3 (2015), pp. 623–629.

[71] Lincoln D Stein et al. "The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics." In: *PLoS biology* 1.2 (Nov. 2003), E45.

Appendix A

DPY-27 alignment

```
Cbr-DPY-27  MSAAKRRAVSAEPTDPDGPEGPPVAVDMPDVNLSDDQRAMYKYKLNIADPLNEKNLEAEF   60
Cni-DPY-27  MSAAKRRAVSAEPTEPDGPEGPPVAVDMPDVNLSDDQRAMYKYKLNIADPLNEANLEAEF   60
DPY-27      MQPFKRRALTSDDDRPYAD----TD-SMPEVDLDVDRRRQYMEQLNIFDDVSSGAYMLEL   55
Ctr-DPY-27  MPEPKRRAVADRNDRPVAA----KSKKAIDFEMAPSRRKQMMDAMKIVDQAP-DDNKPEF   55
            *   ****::     * .         .  :.::  .:*      ::* *         *:

Cbr-DPY-27  ELGKKTIEMQLNENDSLLDIVVGPKYKDFEADPDGKRVIIQDIIVHNFKSYKGSHQLGPF  120
Cni-DPY-27  ELGKKTIEMQINENDSLLDIVVGPKYKDFEADPDGKRVIIQDIIVHNFKSYKGSHQLGPF  120
DPY-27      EAAEN--GVKYDEKEDLLNVQIPPKYEDQISDPDGNRMIILNIYVENFKSYAGKHILGPF  113
Ctr-DPY-27  DTDRH--GFEFDDDEDLLDIFIADKPSDLIADPEGRRLIIKDIFVDNFKSYHGRHQLGPL  113
            :  ..   .: ::.:.**:: :  * .*  :**:*.*:** :* *.***** * * ***:

Cbr-DPY-27  HKNLTMVMGPNGSGKSNIIDALLFVFGFKSKRIRAQSLVNLIHDDRIANSKETTTTIKMA  180
Cni-DPY-27  HKNLTMVMGPNGSGKSNIIDALLFVFGFKSKRIRAQSLVNLIHDDRIAG-KDTVSNIKFA  179
DPY-27      HKNLTMILGPNGSGKSNVIDALLFVFGFKAGKIRTKKLSALINSGGNY---------ESC  164
Ctr-DPY-27  HKNLTMILGPNGSGKSNVIDALLFVFGFRAKKIRTTKLTSLIHVGEEE--------AESA  165
            ******::*********:**********:: :**: .*  **.             : .

Cbr-DPY-27  KVEILFQQIEDIDEE---KYVVSPGEAFVIARTITREGSSTYQLNNSNVQFRVIEQQLSK  237
Cni-DPY-27  KVEIHFQEIEDIDEE---KYIVVPRHDFVIARTITREGTSSYSINDSPSTFRAIEQLLSR  236
DPY-27      SVTIMFQMVKDMPVENYDKYEVLTDNCVCITRTINRENNSKYRIDDKDASQKDVQELLLR  224
Ctr-DPY-27  MVEIVFQVIKDVDKE---KYIVDPKECFTISRSIHLDNTSNYFYNNQVTSQKFIQSLLVN  222
             * * ** ::*:  *   ** *   . . *:*:*  : .*.*  ::.    : ::. * .

Cbr-DPY-27  VHIDLTHNRFLILQGEVEAISQMKHTSGNRDEPGMLEYIEELVGTQRFVEPINQLSHLTA  297
Cni-DPY-27  VHIDLTHNRFLILQGEVEAISQMKHTSGNRDEPGMLEYIEELVGTQRFVEPINQLSHLTA  296
DPY-27      AGIDMTHNRFLILQGEVEAIALMKPTSKNPNEEGMLEYIEDIVGTNRFVAPISKLMHRVS  284
Ctr-DPY-27  AGIDMTHNRFLILQGEVEAISQMKPVSTKADEEGMLEYIEDIVGTNRYVEAIAKLTHKVK  282
             . **:***************: ** .* : :* *******::***:*:*  * :* * .

Cbr-DPY-27  LLELKVSQYHASCRQHAGHLEKFRAAMAAGVGYLNNQNAINMCKGLMIRGNIRYGMQMRQ  357
Cni-DPY-27  LLELKVSQYQASCRQHAGHLEKFRAAMAAGVGYLNNQNAINMCKGLMIRGNIRYGMQMRQ  356
DPY-27      LLEHKSSQYGASVRRHEGHLKVFEKAMVIGMAYLNTFNNLNYLRGIRVKHNLCRYAETMR  344
Ctr-DPY-27  TLEFKSSQYVAICRRHTTLLKEFAPSMQGGVKYVNAVNNLNQIKGFIYKHELALAKAAKQ  342
             ** * *** *  *:*   *: *  :*  *: *:*  * :*  :*:  : ::       :

Cbr-DPY-27  AAEEALIRRKDELDDVVYTATEARKALREKEREEREIDAELTELTKKKIDAEEEVAKLHD  417
Cni-DPY-27  AAEEALIRRKDELDDVVYTTTEARKALREKEREEKEIDAELTEMTKKKIDAEEEVAKLHA  416
DPY-27      DAKMSLVTRTGELEENKDIMLEAKDEVRKKETHERSLNSIVTELENKRIDWQSKKNDWHA  404
Ctr-DPY-27  ESDEAREQEMAKLEEAKAEMLQNKNDLRAAERAERAAAEKTNRLTTEKTTVEQQITDWTS  402
             :. :   .  :*::      : :. :*  *  *:      ..: .::   :.:  .
```

```
Cbr-DPY-27  TGNQIRINVKSANSVLVKCEKEADKLKEELEQLREVPVAARVNIQNMQEELEQIRLKANE  477
Cni-DPY-27  TDNQIRHNVKSANLALVKCEQEAEQLKEELEQLREVPVAARVNIQNMQEELEQIRLKANE  476
DPY-27      RDAKRKQGLKSCTQDLGKLMKERDEARREKFEIETAPENARISKQNMQLEWDQLKEQENV  464
Ctr-DPY-27  RDHKLKAQIKAAASELRQMDVEIAKLTDELKQSQEAPEKSKANIENMLVEMQQMLESKNK  462
               : :  :*:.   * :   *  :   *  : . .*  :: . :**  * :*:  . *

Cbr-DPY-27  IDKSLTSNIQKYDNKIGKERGQTHEIEQEHKVATDAYSKAKSEYQLLLSEF---NLKRED  534
Cni-DPY-27  IDKCLTSNIQKYDNKIGKERGQTHEIEQEHKVATDAYSKAKSEYQLLLSEF---NLKRED  533
DPY-27      CQRTATENLIKYDQKSSADRAKHDDLEKKLSDELLQSMRAKAELDVSESELKDMTIMMEQ  524
Ctr-DPY-27  LEKVYTANLQKFDAKSTIERDKVAMLNEKSERQAQEIYNLQSQIQDFEAELRDMKVTGTG  522
              ::  * *: *:* *   :* :   :::: .       . ::: :   :*:   .:

Cbr-DPY-27  EENRQALADCEQKLKT-------EEAKMTGLQKELEATQEPYNEAKTNVTASETTLGTMR  587
Cni-DPY-27  EENRQSLADCEQKLKT-------EEAKMTGLQKELDALQEPYNEAKNNVTASETTLRTMR  586
DPY-27      GQKR--VDELKGTLQTMMAENIRDNTELNAVTTELQDRKLKFDK----------AVEKL-  571
Ctr-DPY-27  DEKR--VVEMKKKLENIMHQNKQEVERLKQHQNAADEWSAKKNE----------QLGRI-  569
              ::*  : : : .*:.       :  .:.   .  :  .   ::          :  :

Cbr-DPY-27  HHLTGVESRLQSTIDELNY-----LSHEDSQRNLRGKTTKVMYQLKESGKFTPFIGRLGD  642
Cni-DPY-27  HHLTGVESRLQSTIDELNY-----LSHEDSQRNLRGKTTKVMYQLKESGKFTPFIGRLGD  641
DPY-27      PHLKSTEQLLRSKKYELDQEVIEASNTQEV--TYRHQATAKLHELKEAGLFPGFKGRLGD  629
Ctr-DPY-27  PGLNGTIKLLRNQKYSLDRKVDELEDRGDGIYDNRHNNTTMLHKWKEDGRLPGFLGRLGD  629
              *... . *:.   .*:       .  :     * : *  ::: ** * :  * *****

Cbr-DPY-27  LAHVDEEYDAVMSTIFAGNLDFLVVKTHEDCIAAIDLLYKLKLPRASFFALEYGKTVNDQ  702
Cni-DPY-27  LAHVDEEYDAVMSTIFAGNLDFLVVKTHEDCIAAIDLLYKLKLPRASFFALEYGKTVNDQ  701
DPY-27      LASIPIKFDTAISTVFFAQLDYHVVQTSDECRIGIGFCHEYKLPRTTFVFLDHLKDTDTS  689
Ctr-DPY-27  LASISKKFDAAISTIFGHHLDYQVTQTKEDVKKAINLLIEHKMQRSAFCFVEFDRNVKEY  689
             ** :  ::*:.:**:*  .**: *.:* ::   .* :  : *: *::*  ::. : ..

Cbr-DPY-27  HMD-KPTKLPGKRLFDKVQCKDPDIRRCYYSIMGDILLAKDMEEAVKLDKKGGGRFRVCT  761
Cni-DPY-27  HMN-NPKKLPGMRLFDKVHCKDPDIRRCYYSIMGDTLLAKDMEEAVILDKKGGGRFRVCT  760
DPY-27      GMD-STMKFPAERLFDKIHCVNPEIRREFYFLIHDILVVDSLEEATRIDKKYPGRHRYCT  748
Ctr-DPY-27  RMNVAPESFPAPRLFDQIRFENDDIREIYYHIVGDTLVVDTLEEATRLDKKYRGKYPLCN  749
             *:    .:*. ****:::  : :**. :* :: * *:.. :***. :***  *:.  *.

Cbr-DPY-27  MKGGLIERSGALTGGGSVNRGRIQTSEIYQRYEAEEFTSTSTDSERNAHREKLVTRKEQF  821
Cni-DPY-27  LDGGLVEKSGALTGGGGVNRGRIQTSEIYQRYEAEEFTSTSSDSERYAHREKLVTRKEQF  820
DPY-27      LNGSILNRSGALTGGGKPTTGRIRNDNNPNMSGVKKVD----LSKLRAAQEKHNH---AL  801
Ctr-DPY-27  YVGDCLERNGSITGGGRPARNRMRTDSLPISHQHDNRKQ--NDSKINAQMQSVAA---QL  804
              *  :::.*::****     *::...       .:       *:  *  :.       :
```

```
Cbr-DPY-27  TRERKTMMEKIALE-ERNLASLKPILDSFGPKIAGLDEMIIETQGRINSHKLTIHSLTSR   880
Cni-DPY-27  TRERKTMVEKIALE-ERNFASLKKILDPLVPKIAGLEEMIIETQGRINSHKLTIHSLTSR   879
DPY-27      -----EAHLKLQLKQEEIRADNGPIIK----QLEIRKRELIMSTKE---QKTRIAELKSS   849
Ctr-DPY-27  A----NTEDQLRVVNEAI-AECEPQIKYHVEQVAKLKKQIVFNEAAVKSLTESIADLELS   859
                      :: :  *   *.    :.    ::   .. :: .       .  * .*

Cbr-DPY-27  LDTA-----GDSTNAEQELRNMQARLNKLSETVQETEAVVARTGAKVTENARKFELIHDK   935
Cni-DPY-27  LDTA-----GDSPNADQELREMQARLNELSETVQETEAVVARTGAKVTENARKFELIHDK   934
DPY-27      IAAHERRMVNYREVTVEDLDEKRAQIADLKRQVEESQKSSAKIKQQIEQYKRKMDRMFME   909
Ctr-DPY-27  SAVR----PHHIECSEEELSTRKGVVAHMKKQLVDEQKIASQIKKDRDAGEVKARKMFDE   915
              .           : ::*   :. : .:.. : : :   ::   .      *   :. :

Cbr-DPY-27  LIRQNRVQLEEHQNRMKELEAEMAKDQALITNSPEQIRACEQKLAALKATIEDKSAAAGV   995
Cni-DPY-27  LIRQNRVQLEEHQNRMRELEAEMAKDQALITNSPEQIRACEQKLAALRATIAEKSAAAGV   994
DPY-27      LVQKNKDSIEQAKDRMGQLEQDIARQTAIIENNPSHLEQAEKKLSELEHMCLEKRSEADA   969
Ctr-DPY-27  LVGKHKEQLRLTTERIEQMEADIARERAMLENNPAHITAVKKQLKDLGESYKVTSGVARQ   975
            *: :.: .:.   :*: ::* ::*:: *:: *.* ::   :::*  *      . . *

Cbr-DPY-27  RGRA-------EKEFNDIQLAEGTTRLDRTLNEWRAMNKEADAIKADRKLKEQEYQRALV  1048
Cni-DPY-27  RGRA-------EKEFNDIQLAEGTTRLDKTLDEWRTMTREADEIKADRKLKEQEYQRALV  1047
DPY-27      LAQLEVGEDVKGIDIINAQLQTSTASIDAQRARYTEA-------VAARREADAAYQTTVD  1022
Ctr-DPY-27  YSEVDSVVHNREEEENQEKLRVVSTDLKVALEDYTRV-------SNERVAADKKYQESLE  1028
              ..          :  : :*   :: :.     :            *   :  ** ::

Cbr-DPY-27  EQKEKQVIYNETLDLVNETVAQVAQLEESLLPIDDNWLEPESLDSTVQYVRIGDPDFDDK  1108
Cni-DPY-27  EQKEKQVIYNETLDLVNETVAQVAQLEESLLPIDDNWLEPESLDSTVQYVRIGDPDFDDK  1107
DPY-27      NYNMVKQTYDELMRIIDDLENKTMADNAELDIIESAWMQPEKLYPPGKFVRYNDPDIAAK  1082
Ctr-DPY-27  VYRGMSANMEEINKMIDKAEGKIDHYENLLEEVANGWLTAESLDPSAKYCRTWEDDFQEK  1088
              .  .   :*   :::.   :    :  *  : . *:  *.*    :: *  : *:  *

Cbr-DPY-27  VSEGALVMPNDVLAMIEPYREQYTLAVSEIHLESEIIAFVDKMTARKQNLEAQAESFRVQ  1168
Cni-DPY-27  VSEGALVMPNDVLAMIEPYREQYTLAVSEIHVEKEILVFVDKMKARMKNLEAQAESFRVQ  1167
DPY-27      MTDGHVVLPYECISMIEPHREAYEEHEARMLEDDV----FEDTANKICKLEKDVDKFRRE  1138
Ctr-DPY-27  VNDGYLIMPEEVDADIIDYRSLYESTPVTVQAPGN----IQQLKGMLHNLEVTAENFRIQ  1144
            :.:* :::* :  : *  :*. *      :         .:.      :**  .:.** :

Cbr-DPY-27  YDEKGISQYVMMVSFQMSEQTAARKYRAKLAAHRKKLNELRQARLSEFSEALAFLGTTTQ  1228
Cni-DPY-27  YDEKGISQYVMMVSFQMSEQTAARKYRAKLAAHRKKLNELRQARLSEFSEALAFLGTTTQ  1227
DPY-27      FDNKGVRDYAMIVSLLMNEVTSAKKFSDKLKAHREKLNELRMARFNEFSEALAFLGTTTQ  1198
Ctr-DPY-27  HDEKGITHYATLVSLQLNELTSASKYVDKLHKHRVKLHDLKMARYEEFSQALSFLGTTTQ  1204
            .*:**: .*. :**: :.* *:* *:  **  ** **.:*: ** .***:**:*******
```

```
Cbr-DPY-27  MLYQLITNGGDASLKFVEEGKSSDPFSGGIKFSVRPAKKSWKVIQNLSGGEKTLASLCFV  1288
Cni-DPY-27  MLYQLITNGGDASLKFVEEGKSSDPFSGGIKFSVRPAKKSWKVIQNLSGGEKTLASLCFV  1287
DPY-27      MLYQLITNGGDASLKFVEEGKSTDPFDGGIKFSVRPAKKSWKLIENLSGGEKTLASLCFV  1258
Ctr-DPY-27  MLYQLITNGGDASLKFVEEGRSMDPFSGGIKFSVRPATKSWKLIENLSGGEKTLASLCFV  1264
            ********************:* ***.**********.****:*:***************

Cbr-DPY-27  FAMHHFRATPLYVMDEIDAALDINNVRLIANYIKNSDRTRNAQFIIISLRNQMFDLGPRL  1348
Cni-DPY-27  FAMHHFRATPLYVMDEIDAALDINNVRLIANYIKNSDRTRNAQFIIISLRNQMFDLGPRL  1347
DPY-27      FAMHHYRPTPLYVMDEIDAALDLNNVSLIANYIKHSERTRNAQFIIISLRNQMFEVGNRL  1318
Ctr-DPY-27  FAMHHFRATPLYVMDEIDAALDLNNVRLIANYIKNSERTRNAQFVIISLRNQMFEVGNRL  1324
            *****:* **************:*** *******.*:*******:*********::* **

Cbr-DPY-27  VGIYKVDGCTGNVVVNPETVETSKRYTQKFLDQKRKEAYLRQKELEGAEDEQPEPSPVPG  1408
Cni-DPY-27  VGIYKVDGCTGNVVVNPETVETSKRYTQKFLDQKRKEAYLRQKELEGAEDEQPEPSPVAP  1407
DPY-27      LGIYKIDGKTYNIMVDPIAVEIKNRPILKIFEEEIKRR-------EKLRRAEIEPEI---  1368
Ctr-DPY-27  IGIYKTDGSTKHVIINPDKIDEINKGARKTLDNELKELMRKKKREERRARGEEDPEDEEE  1384
            :**** ** * .:::::*  ::  ::   * :::: *.       *     : :*.

Cbr-DPY-27  RRKFEGENMKTGKHKRIFSGPRTPKFAAPLNLKDFGIGSSDEDESD--------------  1454
Cni-DPY-27  --------------------------------RHFGIGSSDEDESD--------------  1421
DPY-27      DLSNGLSNVVIAPKRKQRRLE-------MLKLSDFGLDD---DSDLPEFNRFPPATRREL  1418
Ctr-DPY-27  QLAHSMQRVSLANKRLVYSTPSVWSCPFMSPMGPMGLAPVGEDCPTDSFIHYYS------  1438
                                          :*:     *

Cbr-DPY-27  -----EEDQQPIKSRIHAGIIRRIKDIALEEEDRTPSDSEYEESTIGGSYVEEDVQSEAP  1509
Cni-DPY-27  -----EEDQQPIKSRIHAGIIRRIKDIALEEEDRTPSDSEYEESTIAGSYVEEDVQSEAP  1476
DPY-27      SVEDSDEDDEPVRRRPRRQV----E-----------------EEDEEDELIEEATPSPPP  1457
Ctr-DPY-27  -----CCDDNPFQCCFHFETWAIVIFGI------------IGITVIVGSLFIAGKLLMAP  1481
                  *::*.:   :                                . .       *

Cbr-DPY-27  SAGRPVETDREGSYTNFDEEGDEPIRKKKRRKVAKEYEDASDLESTPTPTRDPSPVVQTR  1569
Cni-DPY-27  SAGPPVETDREGSYTNFDEEGDEPIRKKKRRKVAKEYEDASDLESTPTPTRDPSPVVQTR  1536
DPY-27      ------------------------------------------------------IVVQRR  1463
Ctr-DPY-27  ----------------------------------------------------GSKQRG  1487
                                                                     *

Cbr-DPY-27  SRRSRL                                                        1575
Cni-DPY-27  SRRSRF                                                        1542
DPY-27      VRRSRH                                                        1469
Ctr-DPY-27  NGRV--                                                        1491
              *
```

Appendix B

SDC-2 alignment

```
Cbr-SDC-2  ---------------------------MDKTGRTRQLSESSSQESVAVITVVPSPAQPKQ   33
Cni-SDC-2  MI-------------HSITTHQDLTNIMDRTGRTRQRSKSSSQESVSVITVVPSPAQSKE   47
Ctr-SDC-2  MAKKKFRQTGGCAKNEVITLSPE-ELFIKNAALQLTVNLNSKTENVEVIELEDSPVPDED   59
SDC-2      ------------------------------------------------------------    0

Cbr-SDC-2  KLN-------------------------------QQHKK---QKDGER------------   47
Cni-SDC-2  KHH-------------------------------QLHKK---QKDQER------------   61
Ctr-SDC-2  KSNSSITTVTTKQVKETILNTGKPPIKEPITSISAEEKNDNESSEGEPNYQVDSEGEETY  119
SDC-2      ------------------------------------------------------------    0

Cbr-SDC-2  -------------RKREQQKDRDLDAILTVNDRRQRPQASA-SASLQTAY--VSPKKHFV   91
Cni-SDC-2  -------------RRLEQQQDRDLETILTVNERRQEPQASA-SARLQPAY--VSPKKHLV  105
Ctr-SDC-2  EAWKYRYKKYRKRSLLEYQRD-DFSSD-GDSDKESDPKLQTDGSDVNRANVSHGPTKKKK  177
SDC-2      ------------------------------------------------------------    0

Cbr-SDC-2  RNRPPSMAVPSPKKSNPVSHEKSRLSTSHSPIQKP-IQKSIQKPPTSDA-----------  139
Cni-SDC-2  QNRPPTMAVPSPTKSNPASQQQSRLAPSHCPIQKQ-TQKPIQNPPTSNA-----------  153
Ctr-SDC-2  RNKSPC------------S-SRSKLLPETNVVDYQLSDIPVNNRPYEDDLTGYPLFDHRR  224
SDC-2      ------------------------------------------------------------    0

Cbr-SDC-2  ----------------------CASTSFHQNPSAKLIPIYP-----RSPKSGPKHSVTF  171
Cni-SDC-2  ----------------------RASSSFHKNPPAKLIPIYP-----RSPKSGPKNSVTF  185
Ctr-SDC-2  RVTPDYEDMVQENDFFYDPLSLYTGEGDFDDRELATLSPFNYNEEMNGIEEQGPSKDFDL  284
SDC-2      ------------------------------------------------------------    0

Cbr-SDC-2  RHPISSHQSSLSSHEDSPPATTYPTSLAKLSASVDHYQQSI--------------PQSSF  217
Cni-SDC-2  RHPISSHQSPPSSHEDSPPATTSQPSLAKLSASVDHYQQPI--------------RQSSF  231
Ctr-SDC-2  DLPYLSHQLAEAF----------IMTLTKE--SVNQNNDGGNDSDDNGESSEDSDSDSDS  332
SDC-2      ------------------------------------------------------------    0

Cbr-SDC-2  QQLTPREQEITSILQNSSSQS----ETQSKQKT----------VDRYS---------NSP  254
Cni-SDC-2  QQLTPREQETTSTHQNSKSQLEIQRQTQSKQNP----------VDRYS---------NSP  272
Ctr-SDC-2  DDYEGDESETEN-----GCGTDAESAEKSNSEAVDGMEFDAEPIEDFRENQAEKQRKDSV  387
SDC-2      ---MSDESE--L-----GNQSE---------------------MESFN------------   17
                 *.*                                  :: :
```

```
Cbr-SDC-2  EFLQPVLPK----------------QSKKPQQRPQQKIEKSHHPKVVSSVDMLNTKYNRA  298
Cni-SDC-2  EFLQPEPPK----------------QAKKPQQRPQMK---SRKPEIVPRAEMLNTKYNRA  313
Ctr-SDC-2  DFFTEELIEKHIKYDKGASKKAEPETVEISMEEPEQLILSA-SGSEVDPEDMIRDREKQ-  445
SDC-2      -------------------------------------ESD-SPDEADPDVVIIHDIVHL   38
                                                      . .    ::     :

Cbr-SDC-2  VLATDDEE-SQDSGSRSQDSSSPEFIRTVKSPNQKMAQGKHNSGHDSGCATSS-----SQ  352
Cni-SDC-2  VLATDDEESSQGSGSRSQDSSSPEFIRTVRSTNQKLAQDRHTSDHDSGFATSL-----SQ  368
Ctr-SDC-2  RL-AELNFDIREAKR-LKEQKE---TRSLRQ--ANAAPGETPVDTETEHSTSPNKVTASR  498
SDC-2      RASTTGDYSQSEIGK-LPEQN------------TFFLPGR--------------------   65
               :  :           :..                  .

Cbr-SDC-2  NE---QRRKRAQLSKKVLSVVVEEDEGADDEEETPPREIQVS------------------  391
Cni-SDC-2  DE---QRKKRPQLSKKVLSVVAEEDEGADDEEETPPREIQSS------------------  407
Ctr-SDC-2  DGRNVEWKQMQSLLKYHLHPEDADPEI-----EDRPRDIECESTDWAVYQKGLKEFLEKT  553
SDC-2      -------------VKRNISSNDSDVIIDEDEIPDGAIRITSD-THFIGSSRGTSE-----  106
                        *  :     :             *  .

Cbr-SDC-2  -------------GGSEDSDIIEVFPNRSDRDATRPPKSRRSEKKSKKQNRRSRTPSREP  438
Cni-SDC-2  -------------EGSVDSDIIEVFPNRKDREATRPPESQKPKKKSKKQNRRSRTPSREP  454
Ctr-SDC-2  SPVHRYDPRNDVSCGIEEWKIIQSGISQARER---------AEKAEKARKRKQAIAEKER  604
SDC-2      --LGDF------E--MDEQEFLNITIEENGNE---------QELEEHLR---NAYRH-EE  143
                         : .:::   ..  .           :  .: :   .     *

Cbr-SDC-2  SVVIDEE----------EAPPKKRTRRRLKKEKDPMDVGTRRHKMRRFIHIVYGRPRPVK  488
Cni-SDC-2  SVVIEEEE--------EEQPVQKRTRRRLKKEKDPMDVGTRRHKMRRFIHIVYGRPRPVK  506
Ctr-SDC-2  REREEEERE-------------KLKRREIPKYTYDLLDPNKNMKSRKQVEVVKNEVRPVI  651
SDC-2      EECFEEEDDIIELPPLPVKPAVKKPRRKLPKHL-SIESGSTAKTSKLVAEVVHDHPRPVN  202
               :**               *  **.: *    :   .   . :   .:*  . ***

Cbr-SDC-2  YKMKALTIKKYRALHQKRTRVTRQISNHIVPQYHREPEK--GKRNVPEYTVAETVESYLD  546
Cni-SDC-2  YKMKALTIKKYRALHQKRTRVTRQISNHMIPQYHREPEK--GKRNVPEYTVAETVESYLD  564
Ctr-SDC-2  YRMKALSSAKTRLLYAKRTRVTHQVANHKIPQYHLESHENTRFRDVPDKTVAQTLTCYLE  711
SDC-2      YRMKPAVTDDGKVVEQKRTRVTRNIMSHTIPQYHLEGEE-TEFGRVKESTLSKTIEQYLQ  261
           *:**     . : :  ******::: .* :**** * .:      * : *:::*:  **:

Cbr-SDC-2  VSKTMMQKSSKHHDELVAVGVDYDNSVKMLHFGRTMKKHSCKQKRLKFQMSWWPKDTPDE  606
Cni-SDC-2  VSKTMMQKSSKHHDELVAVGVDYDNSVKMLHFGRTMKKHSCKQKRLKFQMSWWPKETPEE  624
Ctr-SDC-2  ASKPFFNQSDRFHDELVATAVEYDRNVKMLHFGTSMKKHSCRQKRVKFQTLWWKRKRTPF  771
SDC-2      AGKLVSPKCDQFREQIVATAVEYDGSVKMLQFENALKKHSGKQKRLKYQTGWWKASKSHY  321
            ..* .  :..:.:::::**..*:** .****:*  ::**** :***:*:*  **  .
```

```
Cbr-SDC-2  KLKRKGGIRTTRITCYTPYRIIDDPYLFKNHWSFCPKSNAPLQVIRKYYLKPMTRRRTTD   666
Cni-SDC-2  KLKRKGGVRTTRITCYTPYRIIDDPYLFKNHWSFCPQSNAPLQAIRKYYLKPMTRRRTTD   684
Ctr-SDC-2  KKARSGKLHGPK---KPVFRVKEDPILYACHMTLRSTRFSYLQRSLDQLRSRLRVRRNTD   828
SDC-2      ERAVNGYVAMPK---TPVLSISDDPVLYKHHSLFPKNQSSELEKINVQLRIRLNSKRQNN   378
           :   .* :   :        : :** *:  *  :     : *:         :  :* .:

Cbr-SDC-2  DIILDTSYFVREFYLGKAFISLRVTRSSDIPYVYVPPIMQCGYYPYSAVTVENKKFYLAA   726
Cni-SDC-2  DIILDTSYFVREFYLGKAFISLRVTRSSDIPYVYVPPIMQCGYFPYSAVTVENKKFYLAA   744
Ctr-SDC-2  DVIMDSTYFVREFFLTKLCVSFRITRSSDIPLAFLPPTLKAGYFPFSVVEKEDQMHYLMC   888
SDC-2      DVIPDSSYFVREFLMQKHSISLRMNRSSDLPELFVPPTLECGYFPQDAVTVQQQEHYLMM   438
           *:* *::****** : *  :*:*:.****:*  ::** ::.**:* ..*  ::: .**

Cbr-SDC-2  RFREAQLEYFNITYRDIKPWQGFKVGTITDSELYYFHCLGKHIHGFWLIWEKIGRCNVDK   786
Cni-SDC-2  RFREAQLEYFNITYRDIEPWQGFKVGTITDSELYYFHCLGKHIHGFWLIWEKIGRCNVDK   804
Ctr-SDC-2  RYREAQKEYFNLSYFDIKPPPEFEVDDIKGEELLNFHNKGRHIHGFFLVWQTTESFYEDE   948
SDC-2      RFEEAQDEYHNITYRSIAPPVEFQVGTISAKELHKFHRIGRHIHGFFVVWENKFPEYDES   498
           *:.*** **.*::* .* *   *:*  *. .**  **  *:*****:::*:.      :.

Cbr-SDC-2  LKRYTNRRYLVDMFNFQFFPLDVDIKKWELRLRIAFDTVTAYNLHLAEVLRINKPVFDSL   846
Cni-SDC-2  FKRYTNRRYLVDMFNFQFFPLDVDIKKWELRLRIAFDTVTAYNLHLAEVLRINKPVFDSL   864
Ctr-SDC-2  DGNLRGKRYLVDMYFKLKFPLDIKYERWETRLKLAFDRLIIYNLHFSEILRANRPLFNQL  1008
SDC-2      GICCPRKRYLVDMFNLICFPLYTEYEQWESRLRVAFDKTIVYNLHLSEILRCNRPVFDFL   558
                 :******:    ***  . ::** **::***    ****::*:** *:*:*: *

Cbr-SDC-2  TRNPSFYKAVTLKEIVHLMLEQGINPKYYMNSCGKREFYNWGLEKTNEDYLSAYFIICGG   906
Cni-SDC-2  TRNPSFYKAITLKEIVHLMLEQGINPKYYMNSCGKREFYNWGLEKSNEDYLSAYFIICGG   924
Ctr-SDC-2  VKNPAIFQSLTLDEMLKLMNEQEIDTYYFIHTVGHDQFYDWGKTLADTNYLSAFMIICGG  1068
SDC-2      SKNKSMLQPITLKEIVYLIEQSNMDAKSFAVKFGLRTFYDHGRATSNKDYLSAFLIITGG   618
            :* :: : :**.*:: *: :. ::   :  . *   **: *   :: :****::** **

Cbr-SDC-2  SKIIKDNRKFDLEHTRAHIDD--QENTAAITRKGEVLHMHIPMTPSEILVHLDNFKYKKN   964
Cni-SDC-2  SKILKDNRKFDLKNTRAHIDD--QENTAAITRKGEVLHMHIPMTPSEILVHLDNFKYNKH   982
Ctr-SDC-2  SKILKRNAKCKMSRPQVLFDERNPMNSVIIDTDGEHFILKDEKNQPFKKFIISSFGGKTT  1128
SDC-2      AKVVTEEIDSERLR---VF-NSDYMESGVLTSSGDVYTFEFDKIPNNYQI---SIGCNAD   671
           :*::. : . .  .    : :    ::  :  .*:   :.         .   .:  :

Cbr-SDC-2  FE----RLIVHGPMTPEEQVITNLIADTPR--CPTVTPQEAPKKTVRLRTTVMTRKELMQ  1018
Cni-SDC-2  FE----RLIVHGPMTPEEQVITNLITDTPR--CPTVTPQEAPKKTVRLRTAVMTRKELMQ  1036
Ctr-SDC-2  TV-----ERPEKVIDEVLINTTRIISDMEKEKKEDPAPQRRPANSNPTECSRLTRKQLIS  1183
SDC-2      GVAEMEQEDVRHELSECSSRITRIIGDSKKPEKIIARPLVK--TNQNDGMKFFTRKDLLN   729
                     .  :       *.:* *  :       *     ..       :***:*:.
```

```
Cbr-SDC-2  VRNKIYQ----IPKPKKVKNKPQKPARKKPGRKPETQVDKGLKKHDFDILYMASDIESDY  1074
Cni-SDC-2  VRNKIYH----IPKPKKVKTKPQKPARKKPGRKPETQVDKGLKKHDFDILYMASDIESDY  1092
Ctr-SDC-2  FRKKTYKYKDDPPEKKKPDKPEKPPRKKPGRKPETLQDKGMKRHDFKTRFLASDAESEY  1243
SDC-2      YRIKLYDPSYVVPRA--KKQIVNEPAKKKPGRKSKTRYDAAMQQNNFEIEGVPSDVDSEF   787
            * * *.     *.        ::* :****** :*  * .::::.:*.   : ** :*::

Cbr-SDC-2  ENYLSGDENYEDNIPKLTRSQSSESIFADIYYTDYQFDKVSGFHRQRINYIVHPIAQRKK  1134
Cni-SDC-2  ENYLSGDENYEENIPKLTRSQSSESIFADIYYTDYQFDKVSGFHRQRINHIVHPIAQRKK  1152
Ctr-SDC-2  EGYLSQSEDVNDNVPTLKRTHSSDTIFMDAAYRDRRFDRVSWMHQEKVDISTESYAKKRK  1303
SDC-2      EGYLSDSENVFQKPSKLMRSTSSDSVFIDYQYREKMFLDVSWFHQQKMIDRSLPPLKKRK   847
           * *** .*:  ::  .* *: **:::* *  * :  *  ** :*::::        :::*

Cbr-SDC-2  RKMNRHVKKHLLRYRMLALEGVAFKEMLECYHGNLPKLGQEITKTKKAIRV-NGYKRFKV  1193
Cni-SDC-2  RKMNRHVKKHLLRYRMLALEGVAFKEMLECYHGNLPKLGQEITKTKKAIRV-NGYKRFKV  1211
Ctr-SDC-2  RKMNRIRQKHTLRFFMLESESLAFREMLECYHNNMPRLTIEVSKTKNCIRNPYRFERLKV  1363
SDC-2      RKMNRIYHKHSVRYTMLQANGCAFTEMYRCYDKILPCGTKEIARTKNAIRFPHRFRTYNI   907
           *****  :** :*: **  :. ** ** .**.  :*    *:::**:.**    :.  ::

Cbr-SDC-2  PKVFEKGDSAKIDSIGDLLKEMVTFTVAAEHSSTRAANGVARIAQRGRIMQRLTTDNLKP  1253
Cni-SDC-2  PKVFEKGDSAKIDSIGDLLKEMVTFTVAAEHSSTRAANGVARIAQRGRIMQRLTTDNLKP  1271
Ctr-SDC-2  PRVYGYGDAPFIDVVADILRNCVTHCVALEHASTRAGNGLSYHIQKARTQRREIVRNLPI  1423
SDC-2      PQVYGPGDKQLITEVFGVVKDVITRATGFESASIRTANDIAQAVYDANIARRELLENLEP   967
           *:*:  **   *  :  :::: :*  .. * :* *:.* ::     ..  :*    **

Cbr-SDC-2  T-NFGLPALSYLAMEILTRVKMTGRTMMEEEKRNLRESVNQSYINYMSLLPHERRLLDAK  1312
Cni-SDC-2  T-NSGLPALSYLAMEILTRTKMTGRRMMEEEKENLRESVRQSYINYMSLLPSERRILDAK  1330
Ctr-SDC-2  KKGYNLPKAPYLAMEFLSLAPVISRAWLEEQKRKLFEQLYDLTHPFMNLPPHMQVFKQSM  1483
SDC-2      SDNGILPSPAYLAIEMLSHQKMSGRLCLESARKDVQNNVDKMYNDYMDLDPLDKELHFEI  1027
           .    **   ***:*:*:   : .*  :*. :..: :.: .    :*.* *  : :

Cbr-SDC-2  IRNQKDRHIA---LQNERYETERMKNLSKRQGFIRYDQRSLQAQHKREEARRLNRLKHSE  1369
Cni-SDC-2  IRNQQEKHIA---LQNERYETERMKNLSKRQGFIRYDQRSLQAQHKREEERRLNRLKHSE  1387
Ctr-SDC-2  L---KIKQADDEFDEFFRDQKERVKQLDRSYRTIVFAPRSLEAQHKKEYERRLARKAHAQ  1540
SDC-2      SQSIRQSKLNESLEEYERN---RERQLAKTLKTVPMDKRSQAALARREEKRRESRRKLAD  1084
               :  :      :  *    * ::* :    :    **  *  ::*  **  *   ::

Cbr-SDC-2  KETAFRKEQRY-----IAELKAAGGDVPTK-------EWI-----RKRIQEEEAEEAAKD  1412
Cni-SDC-2  KETAFRREQRY-----IAELKAAGGDVPTR-------EWI-----RKRIQEEEAEEAAKD  1430
Ctr-SDC-2  KTAARKKMSAEERARAIREEKENEGAVKAAKKQADDRKKLLTSLHMREVDLRERRELTVM  1600
SDC-2      KYAEQRRMMASTRRLEKR------------------------TTQKQVDPE---TIQRL  1116
           * :  ::                                      :.:: .
```

```
Cbr-SDC-2  RKA-EEEQRE-----RKRMQEIEAARLLKEKERRKAAIEKEKLDEAV---ATKLLKEK--  1461
Cni-SDC-2  RRA-EEEQRE-----RKRQQELEAVRLLKEKERRDAAIEKEKLDEAM---ATRLLKEK--  1479
Ctr-SDC-2  ARLDRIEKRERRYHNRRKRAAIKKALLAREKERL---EEKERLEEKRRLEEKRRLEEKQR  1657
SDC-2      RREDEVRKR-KRFEEEDRR-----GMIRRREERV---ALQEKV--------DRMLEEGLR  1159
             :  . .:*      . :        : :.:**      :*::         : *:*

Cbr-SDC-2  --------------DERERKRIEMERIQAILRESSALMKEAAEKERQKQLEE--------  1499
Cni-SDC-2  --------------EELERKRIETERIQAILRESDALMKEAAEKERLKKLEE--------  1517
Ctr-SDC-2  LEEKQRLEEAERER----KKKIEKETARIR----AEMKKEEEQKQKQQSLKKQAIKRKQD  1709
SDC-2      LEKV---REAERIRQQQEEERIEMETILIS----RRVREEEEEKMRLERLRK--------  1204
                          .::** *           : :*  :* : : *.:

Cbr-SDC-2  -------------------------DKRRK-------NAEKSQSESEEELRRLDRQRHEA  1527
Cni-SDC-2  -------------------------EKRRK-------DAEKSQSESEEELRRLDRQRNEA  1545
Ctr-SDC-2  LEIEENMRKLVEKRKIELLKETLEAHRREKEKEEAEAKRLKLEKEKAEEEKRLKLEKEKA  1769
SDC-2      -------------------------AER-----EREQERLK--REEEEERKRLEQ----L  1228
                                  .*         .  *   *. ** :**.

Cbr-SDC-2  RRLKVLEREKKRSEEEKTMEAMWLQRQKE-----------LAEMKRRQ------------  1564
Cni-SDC-2  RRLKILEREKKRSAEEKAMEAMWLQRQKE-----------LAEMKRRQ------------  1582
Ctr-SDC-2  EEAERLRREKEK---AKEVERLKLQKKKEEEAKRLQIEN--EKAERAAKLKLEREKAEEA  1824
SDC-2      REAEKLKAEIEK---ENE---RKLQEERTR--KALELERKIEEIKRVSTLK------DMF  1274
           .. : *. * ::    :      **.::              : :*

Cbr-SDC-2  --------------EEETAKSL-AAVKIPKTVTTSLYRLAQKLDKEMIAIAEEKLYSRTV  1609
Cni-SDC-2  --------------EEQTAKSL-AAVKTPKTVTTSLYRLAQKYNKEMIAIAEEKLYSRTV  1627
Ctr-SDC-2  ERLKLEEERKKKELRKQASKKNVHIVTSPEVDLLTVYKNPLE-----FEIARSSFRVIDC  1879
SDC-2      GPLPIAKE------NEQTEKDFQILLDDHELTLLTISRDPLN---EKYQEARTEFERLDI  1325
                      .::: *.    :   :    :: :   :        *. .:

Cbr-SDC-2  MLVIRESEDKFGAFLRKTRNFNLRVFTAYFSRFF--DKNRFAQKNSDNDLYDNIAKCIHY  1667
Cni-SDC-2  MLVIRESEPKFAAFLRTKRDFNLRIFMSYFSRFF--DKNRFDEE-ENNDFYDNVAKCIHY  1684
Ctr-SDC-2  REVLLEYAELFGLIVMNSLRTPEGELVRYLVNLAEIHPKRPIVTAGCEQLYENIASSFIF  1939
SDC-2      KSMLLRKAEKLIDVLTIHYDVPIEQTCRYFTSSIESNENRMAVNEQLNKLFENMANCFTF  1385
             :: .    :  .:             *:      . :*       :.:::*:*..: :

Cbr-SDC-2  CPTFN--------DYKFVLDMKKIVKHLSSDMKHRIKKYMNSDPKRSGS-------ESPV  1712
Cni-SDC-2  CPTTN--------DYKFVLDMQKIVKNLSSDMRHRVKKYMSSDPQGSGS-------ESPV  1729
Ctr-SDC-2  KKVEN-----NMMKWIFSYDLDVLLRNIPSDVRTLIEKCRELRITDPELSFSLFSAKTPV  1994
SDC-2      NIQDGENGLQSKRKWDFQFKKCAVFDGVSQSTVNFIEEKMRENTKK----KHLATPKTVI  1441
                     .: *  .   :.  : ..    :::                  :: :
```

```
Cbr-SDC-2  SNA-AF----SPEYE--PSPEPESEILADSLSESDDADGPDELL--PEAYETGQTDLTAT  1763
Cni-SDC-2  SNA-SF----SPEYE--PSPEPESEILADSLSESDDPDGPDELL--PEAYELGQTDLTAA  1780
Ctr-SDC-2  ATQDSLMRDVFPRFTNVSQDEPELNVSIPDASMLS-HIGTVA----NASVSF----VHSK  2045
SDC-2      SIDTSLLKQSLLRSH--ARFDPDISLYAQN-HTAN-SIGDVTLKMSNYSLDFATQSIHDK  1497
           :   ::      .       :*: .:   .    .   *         : .     :

Cbr-SDC-2  HWS--------------------------------RSPSVDSQDGDESQNSYKG---RCF  1788
Cni-SDC-2  RWS--------------------------------RSPSVGSEDGGESQNSFKG---RGF  1805
Ctr-SDC-2  QWS--MTPRTSK-KTGRVRKLFTSDE----------PSVGHS---NEENGVMVDETAVW  2088
SDC-2      ELAEKATPKKGPTVRRHIKNLFGSEKVIVRRSLAAGKPASLNSEDSDSEDSREGSPVAEF  1557
           . :                                   *:   .   :.::.        :

Cbr-SDC-2  HPMVALRSTFWRLIEMSENALEAQNEQ-LYRNEFRNYIVRRRSFRKAGVPYAVGVYAASC  1847
Cni-SDC-2  HPMVALRSTFWRLIEMSENSLEPQNEH-LYRNEFRNYVVRRRSFRKAGVPYAVGVYAASC  1864
Ctr-SDC-2  IPPDTDEGLFWQLIGVYEDSGKTIEERTRMNEQFLAYLTSKPDLSNNSIHWLFSILGACY  2148
SDC-2      -LPTNPVCSFWKLVVKIENSTT-DKEKTELCEDLDKLILRKDDLFSKSLKWMFPLLATFY  1615
                    **:*:   *::    :*:    :::   :  : .: . .: : . : .:

Cbr-SDC-2  VLLTGSMYDPCGRREQSPLRMPGEVIEIDNNDPDLKGVIDRVAQLGVVFHQAN-RSPLNL  1906
Cni-SDC-2  VLLTGGMYDPRGHREQSPLRMPGEVIEIDNNDPDLKGVIDRVAQLGVVFHQAN-RSPLNL  1923
Ctr-SDC-2  VIVVGKKKQDEKLQ---DVIYEGHQYDEISTDNQVKKVVKNLTTLAWFFQAAHQESQKNL  2205
SDC-2      VLLSNAVLNENE-----E-IISDKNQTGVTKDEILKSTINDLMIIAAYFEEGS-RERSNL  1668
           *::     :              .     ..*  :* .:. :  :.  *. .  ..  **

Cbr-SDC-2  KQLCKWNGFRQACDLIDELYEFIMGVYCKLQLDQVFKDELDDETKIREAFRFIATKFVPL  1966
Cni-SDC-2  KQLCKWNGFRQACDLIDELYEFIMGVYCKLQLDQVFKDELDDETKIREAFRFIATKFVPL  1983
Ctr-SDC-2  EALIAHNGFQQVFSTLETMIQTIYDLFTALNISTFVRADVSVEDGLLVIFTKIGDECERM  2265
SDC-2      RKMISMNGFSVVFNRVILFAKKTCTLAKELESNSRSLSGYVIEDL----FESLLAEIERT  1724
           . :   ***  . . :  : :    :   *: .         *      *  :  :

Cbr-SDC-2  LSVHCGVKKSQVAKWRYEEVTIGRCCVNMTEYKQPTVNTTNEFILKQNAQQFSRITAIVN  2026
Cni-SDC-2  LSVHCGVKKSQVAKWRYEEVTIGRCCVNMTEYKQPTVNTTNEFILKQNAQQFSRITAIVN  2043
Ctr-SDC-2  IASDYSVPAAVASIMRNDEDDT----VNHILEKRP----YRPSRLHRDLEHEERINTIIK  2317
SDC-2      MRQELGSSVRKTGKLERDFEEI----VKLIQNEKKLA--LSHKSHKNDENRRFRLNTVVK  1778
           :  . .     ..  . :        *:    ::           :.: ::  *:.::::

Cbr-SDC-2  WYQYLVEKGKSKIEDMRSNAMNAIAWKRRQYHIM-SPMPATSDQEEDDEESPIKIIIPDD  2085
Cni-SDC-2  WYQYLMEKGKSKIEDMRSNAMNAIAWKRRQYHLM-SPMPATSDQEEDDEDGPIKIIIPDD  2102
Ctr-SDC-2  WYHGVQTHQAEEIKSIKHFALQKAA---DQYAERIKQLEAIQS----------------Q  2358
SDC-2      WYDAIICHCKEELTQAIVDAFP---------------LNAITK----------------N  1807
           **. :  :  .:: .    *:                : *  .               :
```

```
Cbr-SDC-2  VNLLSPRKVTPRTLTPRTPTLHVTKDFVIDKNNGKDAEASAT--------------VRHV  2131
Cni-SDC-2  VDLLSPRKITPRTLTPQKPTLHVTKDFVIDKDNNKDAEASTS--------------VRPV  2148
Ctr-SDC-2  VNTPKPRN-------KFVKIFDDVDNFEYD--GGDESGGSRGNTRANSPDSDILLIEKEP  2409
SDC-2      K---------------------ETSHVAME--NGDDEAMLSDT-----SDNQMS------  1833
                                 ....  :    .:

Cbr-SDC-2  VSPYQHPFVQNIGQL----GEVCQQQSTVYVSGSYNISSEKEKYEKKIRYLIERREEIEK  2187
Cni-SDC-2  VSPYQHPFVQNIGPG----GEV--QKSTVYVSGSYNVPSEKEKYEKKIRYLMERREEIEK  2202
Ctr-SDC-2  ETSYEAIVQDRVSKTPAHLEFLESSEYYVKVESTVRMLSKKICKAQLDLFKKQRALEIEK  2469
SDC-2      TTDYQM--PKNICRN---SE-IFPEDAFAKAYAVVRIPSKKERAQMLSVYRKKNAQSGCV  1887
             : *:    ..:         :  ..  . . .  .: *:*        :  :.  .

Cbr-SDC-2  ENSLNNNIPPSADMFINNLWRAIERRISVFPGGIKIMTGLHKKIQRPHILDSEFKIYIM-  2246
Cni-SDC-2  ENSLNYNTPPSDDMFINNLWRAIERRVLVFPGGIKIMTGQHKKVQRPHILGSEFKIYIM-  2261
Ctr-SDC-2  ENMFPTKLPTFEQPFLQHTWRTLARSFYYLSNREKDMMRLFNKYNEIHQQRSFSRIHPLI  2529
SDC-2      ENKGLSRMPKFEEPFVDSVWRTIEKRINNMTHSEEKQIKRFIPVSRSHKLNEKVKFYAMV  1947
           **    . *   : *::  **:: : .  :    :     .   .. *   .  ::: :

Cbr-SDC-2  ---SRDARGKRFPEEFPEYKHDWFKYTRISIEPR--KYQAYEDTILNSFPHEIMCKTEFR  2301
Cni-SDC-2  ---SRDARCKRLPEEFTEYKHDWYNYTRMSVEPR--KYQAYEDTVLNSFPHEVMCKKEFR  2316
Ctr-SDC-2  HDISIAAQNDFLAPELREDPDNWAFYKKLEVGQGLDACRESEQKVLDLFNHIPYTRREFG  2589
SDC-2      MIQERDSRDTRLFNSKFQDDNLWHCYSKSSLNHE-----KMESRILQHIEHTVLSKSNFN  2002
              .  ::   :  .  :  . *  *.: .:          *. :*: : *    : :*

Cbr-SDC-2  KMQWTVPRQFGPPKKAIEFFTDLDKYRDLELYKQYLSEGELPFNIKIYRHLWFMGSLFAE  2361
Cni-SDC-2  KMQWTVPRQFGPPKKAIEFFTDLDKYRDLDLYKQYLSEGELPFNIKVYRHLWFMGSIFAE  2376
Ctr-SDC-2  KMKWEVPRKNGKTIHALEFFTDLEKYRAGKLYRKYATSGFLPFKFYVYDHLWFMGALTPT  2649
SDC-2      QMKWSVQCVNGNKKDAIHYFTDLYKYRSESEFRSALSCGKLKFNFKVYTHLWFMGNLLPT  2062
           :*:* *    *   .*:.:**** ***  . ::.  : * * *:: :* ****** :

Cbr-SDC-2  GIAEDWHDDGLPGGFCGACTDGTVIFVKKCTCIFHQDHYDDKFIYTHCNIKKELNGVERL  2421
Cni-SDC-2  GIAEDWHDDGIPGGYCGACTDGTVIFVKKCTCIFHQNHYEEKFIYTHCNIKKELNGVERL  2436
Ctr-SDC-2  SYCLDSHEDL-GNGVCAGCTEGSVFVIPHCTCEEHLDVKRNTFIYTCFKKGVEAGGVNRI  2708
SDC-2      SYNPDSHDDK-LFVPCSGCTSGDVIIIHKCTCAYHNDTFSDKFIYANTSLPVGIDKVTRL  2121
           .   * *:*      *..**.* *:.: :***  * :   :.***:  .       * *:

Cbr-SDC-2  TGRFVCEHGPSSVLVLVDEDKRPKGV------------YEVKNPAYTTHDAKLRIVARKT  2469
Cni-SDC-2  TGRFVCEHGPSSVLVLVDEDKRPKGV------------YQVKEPAYTNHDAKLRIVARKT  2484
Ctr-SDC-2  LGRFVCEHGPSSFLVLENEQNDNDNQRVVPSKRPTDRPFEPDMNRCIVFDSKLRVIKRKT  2768
SDC-2      VGRFVCEHGPSSFLILEHCSANVDAN----------IPFE---SENVEFSAELRIVKRKT  2168
           ***********.*:* . .   .              ::        ..::**:: ***
```

```
Cbr-SDC-2  MHAQIRKCFANVPQTIRERSQESTTNS-------DSSGSSTDS----LQNSVD-------  2511
Cni-SDC-2  MHAQIRKCFANVPRSIPEGSEKSPTNS-------DSSGSSTDS----LQDSVD-------  2526
Ctr-SDC-2  MYRDLLQAVRTHPKPERRIIRNEPNELASQSDSCDD-SSDSDDSEQVFEDSDIEVKEA--  2825
SDC-2      MHSQLVKTFAEEHTHLRDASRHRAIST----VTLDSSGSGRSTRCEIFEDSPSEDENDEN  2224
            *: :: : .           ..   .        *. .*  .     :::*

Cbr-SDC-2  -----EFGNPLIV-SKVQPNIVENAKELYKRFSRLKEGKITLD-----KP------KKMR  2554
Cni-SDC-2  -----EFGNPVIV-SKVQPNIVDHAKELHKRFSRLKEGKITLE-----KP------KTMR  2569
Ctr-SDC-2  SPDPTSFSDFTFHSNAYETLKKKKTE--EQEALRREAERI----RNRLEQ--RRIYKQRH  2877
SDC-2      QLDTTRIG------RKIDPIIV--DS--DKAYLIAEGERMALRIKRLLDPELQKFRSKNF  2274
                 :.         :       .   :     :  ::        .       .

Cbr-SDC-2  TWRSKSVDSYRKAFEVKHRPGLTATQSLIDLTDLENHAKLKMEKAKQTMIEELNIEKDV-  2613
Cni-SDC-2  TWRSKSADSYEKAFEVKHRPGLTATQSLIDLTDLEIHAKWKMENAKKTMIEDFKIAEEV-  2628
Ctr-SDC-2  GIRSSSADAPVRE-SPGTIKYYYCSQSVTDLTQLSKYVEKKMERTRLKLKLEFPEDEDAI  2936
SDC-2      VSRSKSVDAPKTS-KQ--KTVIRRSQSVCDLNDVNEYAQKKVRNTKDSFATLFRDHEYST  2331
             **.*.*:     .         :**: **.::. :.: *:..:: .:   :   :

Cbr-SDC-2  ------RLDSETMD-TRLFEGIHNISEANNFRLLLELFTLGPAAEEPTAKYCKTRYIKIQ  2666
Cni-SDC-2  ------RLDSETMD-TRLFEGIHNLNEANNFRLLLELFTSGLSDEEPTTKTCHRNYIKIQ  2681
Ctr-SDC-2  -----PLGDEEILNSHIVFNGTGNASNPQKFQVIASMLECGMKEEIENPKN--STIHMIR  2989
SDC-2      RRTYEEQLNNELLDVVTTFGGASNVSADKKYNILASILAFEKEVQLVNDKN--GELFKTV  2389
                  :.* ::    * *  * .  :::.:: .::      :  . *

Cbr-SDC-2  EHLKQHSLMRVYGQDKENVPRFDED--KKFQGGQPISALMHEYYAFMQYIKRTMRAAKNH  2724
Cni-SDC-2  EQLKEHSLMRVYGQDKENVPRFDED--KKFQGGQPISALMNEYYAFMQYIKRTMRAAKNH  2739
Ctr-SDC-2  AAFEDRGLLKIVGDSIENKPQYDLM--GNYLSGKLTSITLNAYYQFLPFINDTYRRARLA  3047
SDC-2      SNLVQRNSLQHVKGVILAEDNQTLRSTDNTSEVFPESKAVNEYLKFEIYKRKMMVNAKLM  2449
             : ::  ::          .       :       *  :. *  *  : .     *:

Cbr-SDC-2  VAANRRLRFNE---AQFEYFHMIYQKVFNLNLHLFEHLLHQISKHTFTPYALHHAEHKGD  2781
Cni-SDC-2  VAANRRLRFNE---AQFEYFHMIYQKVFNLNLHLFEHLLHQISKHTFTPYALHHAEHKGD  2796
Ctr-SDC-2  PDSL-LGTLRTPENKPFKNLLACYEKIFRFNYYLVQHFLEVSLARIFNTSAVYCAEINQD  3106
SDC-2      ADTVKDLKLKHAEYRPFAKLIATYDSIFKFNVYLFEHFLNCISKHVFNPYAIYCEETRPT  2509
             :     :.      *  :   *:.:*.:* :*.:*:*.    : *.  *::  * .

Cbr-SDC-2  L---TKIRTVLARMKIDLPTVMNSFFNIEPMKRQIHELRQLSEFCQKSEMDCHIATLGRY  2838
Cni-SDC-2  L---SKIQATLSRIKSGLQTVMNSFFNPEPMKRQIHELRQLSEFCQKAEMDCHIATLGRY  2853
Ctr-SDC-2  I---SLIEKNLAKLRHIVPLFLCQLFNTSPIRRQLRDLNEIKDQITEYDLDCHIASLCRY  3163
SDC-2      GTELSKFQLTLKLIETSMPTVLSMLFNTEPLRRQLSELSEIHKKVRSEDLACTIASLCRY  2569
               : :.  *  :.  :  .:  :** .*::**: :* :: .   . :: * **:* **
```

```
Cbr-SDC-2  AIERIRVPQSAEKVFDDYPWINQS-VHKDTIDLLRFDSGETVPDGFDSR---TFNEQLMK  2894
Cni-SDC-2  AIERIRVPQSVGKVFDDYPWVNQS-VHKDTIDLLRFDSGETVPDGFDSR---TFNRQLMK  2909
Ctr-SDC-2  AITRIRVPQTAEKHLNEFSWVNHMAKHQDTFEIFHLNVDETLPRDFNLD---KFAILLNQ  3220
SDC-2      AIERIRIPQTADKRLCDFSWLNSAEDHRETVSFIRLTLEHTLPDMKTENEQTRFVEFLKE  2629
           ** ***:**:. * : :: *:*    *::*..::::   .*:*          *   * :

Cbr-SDC-2  NFYNP---IDVFEQSTTLRPKGESAELNLYNAFYTQCDGFFAKFERMMPH-GAMDPKMKT  2950
Cni-SDC-2  NFYNP---IDVIEQSTTLLPKGESAELNLYNAFYAQCDAFFAKFERMMPRGGAMDPKMKT  2966
Ctr-SDC-2  AYYEEYEGESMKHSWESFLAPRRTGGVAVLRAYFNQTESFFEEVERSLPN-ELVDPKTKA  3279
SDC-2      AEGFHFSYKFVEAQCKTFVRNHGDSKQAFFTAFYNQNEAFYGSLQKFMSN-GTIDPKMKL  2688
                     :  .  ::      .   .  *:: * :.*: ..:: : .   :*** *

Cbr-SDC-2  YHQHQAFIRLYEIAKG-NRIVDRTDVTRMNDTDVIMLYTAFVSNPDVETDAGADLDCLNQ  3009
Cni-SDC-2  YHQHQAFIRLYKIAKG-NRIVDRTDVTRMNDTDVIMLYTAFVSNPDVETDAGADLDCLNQ  3025
Ctr-SDC-2  FYEIQFFEKIRQILVTKTHFPTSEDIAKMGPLNAAALISAMIEMPRSNHLSFNLSHSVDA  3339
SDC-2      YYQHQAFLRLHNIVKKRSHIITSDDYHRSSDVCKAMLLSEIVSNPKIAQEAYISGSVLDR  2748
           ::: * * :: :*    .::    *  :        * : ::. *     :      ::

Cbr-SDC-2  LYMQLSKQKAVPCPINPSLIGTTFVVFDHHLVVSMVREPFVFLADLHFNFT---PMKSRG  3066
Cni-SDC-2  LYMQLSKQKAVPCPINPSLIGTTFVVFDHHLVVSMVREPFIFLADLHFKFT---PMRSRG  3082
Ctr-SDC-2  FYEEMLNYEAKLLPVCPSTIGTTFVCFEKDLYFSVVKENEVLVNEK-YPSKQLSQ---DD  3395
SDC-2      MYTSLCKIKAKMPLISPSYIGTSLTCFEDELLFSAVREAKVHTDTR-VVFRSKSCMRPNE  2807
           :* .: : :*    : ** ***::. *:..* .* *:*  :

Cbr-SDC-2  RIIEA--VSGSCVINLLMDSNSDKIRIEMR--PKSVQTKGDRLCFELD-HETLTRAGSID  3121
Cni-SDC-2  RIVEA--VSGSCVVNLLMDSNSDKVRIEMR--PKSVTCKGDRLCFELD-HETLANAGSIE  3137
Ctr-SDC-2  RHYYVVYNLPESHFNASLNKKDGKLELSYH--RDNSYQVDDRFSVAFASHFHVFPYPKWD  3453
SDC-2      KAGDA--NFKTCKVTLLVNLETALLSMVFKSRDQSEIDKDDRLDIDILDEEVIKPIIDWN  2865
           :   .      . ..  :: :   : :  :   ..     **: . :  .  :    . :

Cbr-SDC-2  GVLKFVVSQRFNKLQEQFEMQPQV--RPF-----KSRRGILENRNIINELVSSDEQDKSS  3174
Cni-SDC-2  GVMQMVVTKHFNKLQEQFEMQPLV--RPF-----KSRRGILENRNIMNELLASDEQDVPP  3190
Ctr-SDC-2  DISDVKLTGKMNVLRKLL-----DSLVPTQPNRDDFRVGRLENPVIERKS-----NRA--  3501
SDC-2      RIFETFIQPTYNTLFSRMEKRERVSILPEN------PLGRLENYAFTNPN-----QDK--  2912
            : .  :    * * . :         *          * ***  : .       :

Cbr-SDC-2  TSSCRMSERTIDPNYVGFLHTHKELKHLSEVSKNMREYFITNRRPGSRKRSVPP------  3228
Cni-SDC-2  TSSGRTSERTVDPNYVGFLHTPEEMKHLSEVSKNMREYYITDRRRGCRKRPVAP------  3244
Ctr-SDC-2  --------RVIANRSVRF-------------NESVRQYDLQGSDTEGED-----------  3529
SDC-2      ----------------D--------------CQAVLEYIDVASDTDAEESIEDPLDIVEM  2942
                                         : : :*         ..
```

```
Cbr-SDC-2  -ASPHIPPHMNPKRIRFSHKY                    3248
Cni-SDC-2  -AIPHVPPHMNPKRIRFSHKY                    3264
Ctr-SDC-2  --------------------                     3529
SDC-2      TLKRALPRSMSPSSKRRRMR-                    2962
```

Appendix C

XOL-1 alignment

```
XOL-1         ------------------------------------------------------------    0
Csp11_g5907   ------------------------------------------------------------    0
Csp11_g5908   ------------------------------------------------------------MS    2
Cni_g27847    ------------------------------------------------------------    0
Cni_g27926    MKAWESVKTGFLIRDRRDRRGARLYALPTVPFYRTIPPQPSSFRLSTFLFQRSEQFQPIC   60
Cbr-XOL-1     ------------------------------------------------------------    0

XOL-1         ----------------------------------------------------MQVEAN    6
Csp11_g5907   ----------------------------------------------------MSLEQQ    6
Csp11_g5908   DREPVGREPRIQG-----------RGNQEVSSEEDEEE--------------GEGSAIQQ   37
Cni_g27847    ----------------------------------------------------MEN    3
Cni_g27926    HVAVWGRINRLFRCFPHEYEQLETWTNLQITAASHTHSTYHLTTVFFPILHHIHMSARFK  120
Cbr-XOL-1     ------------------------------------------MDKTNFP------------    7

XOL-1         SERRVKILGIDRSENSPVLTYMET-----------EDD-----PNFRNSKLAAAPHTVHM   50
Csp11_g5907   PSPSSNV-------EQTSVCWKEG-----------VDVKKIFNELPEGGLLSMASHTVKL   48
Csp11_g5908   --PEPSDEPD--YEAKKMETYARTPVLDHIQEVIDSIDADGRVVDPSSITVGVAPHVISI   93
Cni_g27847    PHLQPPIKYD--FQVHKSSSF--------------PVEEDVGGDLLDSKKVGIAPHVVDK   47
Cni_g27926    TKRLSWINQT--YNLLSTMTFMCI-----SLPLFQLKSKDVGDNFPGPKIVGIAPHVVDK  173
Cbr-XOL-1     ----PPINHD--FHVH-----KSS-----SL----PIEEDVGDNAPGPKIVGIAPHVVDK   47
                                                                :. * *.:

XOL-1         MDSGFLAINRQCLVKGKAILAREPKSSNEHMIDDLPKHAHDQHTLSILRDFIDQLKLHNV  110
Csp11_g5907   IGCAYVATNKYCFVEGAIIKGDRHAVSF-------ALCPNTEIDNEIVTDFMAMFDLKNK  101
Csp11_g5908   AGFAFLATNKRCSVRSEVIWNRKAVNVSEHRIH-FPELSNTTEPEEFVKAFLGELNLQFV  152
Cni_g27847    LGSLIMAVNKYCKATTRVESKSRPRSAEEHRIE-MEDSFHGDIMRGLIRNLLRDLKLTRV  106
Cni_g27926    MGSLIMAVNKYCKATTRVESPSRPRSAEEHRIE-MADSFHGDIMRGLIRNLLRDLKLTRV  232
Cbr-XOL-1     MGSLIMAVNKYCKATTRVESPSKPRSAEEHIIE-MADSFHGDIRRGLIRNLLRDLELTRV  106
                    :* *: * .         .                .      ::  ::  :.*

XOL-1         YEINFYDPLDSSGKLAVIPMLIALWKCMLASETDICDQEVLKS------IMNSVIAKFEL  164
Csp11_g5907   YSIVVHNSMDSDYRLSYTTLFVAVWKLLQVNEHHIGDENDLET------RMYSIFQEHDI  155
Csp11_g5908   HKITIRGALNYSGQISYPTMAVAIWKSLATE-------FGI-CLMQDL---INKLMKFDL  201
Cni_g27847    YRIQITGEHDYNGRVSQIAVLVAIWRSLKSLEHPVDSTFGEDCIWSNAETFYAMVKDYNF  166
Cni_g27926    YRIQITGEHDYNGRISQTAVLVAIWRSLKSFEHPYDRTFGEDCIWSNAETFYAMVKNYNF  292
Cbr-XOL-1     YRIQITGEHDYNGRISQTAVLVAIWRSLKSFEHPYDRTFGEDCIWSNAETFYAMVKDYNF  166
              : * .    : . :::   : :*:*: :                          . ..::
```

```
XOL-1         QIPCKNAVIDATLSGSREEVHII--AEDG--SLENSNG----------TTEHFNKKHDLV  210
Csp11_g5907   EPEDKEAVFRATVEGSSYELFFKTFPEDGRDRLVRTKA----------KIEKFTDSCDLV  205
Csp11_g5908   KVEEQEAVLATV-LNGSERWIQL-KPERPSDHTGIPRGSVLRENDANRLSVDLIKNLSFL  259
Cni_g27847    ESEAKLSVLGATIFENESKYFY---DEKMDENIRNPSK------FRTLSSITMHANYDFA  217
Cni_g27926    ESEAKLSVLGATIFDNELKYFY---DEKMDEDIQNPSG------FRTLSSVGMLAEYDFV  343
Cbr-XOL-1     ESAAKLSVLGATIFDNELKYFY---DEKMDEDIQNPKG------FRTLSSVGMLAEYDFV  217
              :   : :*: :.      .       *                        :  . .:

XOL-1         FVKTDLHPEDFTPQMFPSQAKAKLLRDAFNNEEDEDTFP--DILVPAYMTAHSKN--RVR  266
Csp11_g5907   YLRTDAHPYPVESDACATFSDCDELKLKFTD----DFFD--ADDVAQNLTVFSKKRLAMK  259
Csp11_g5908   FVRTDLHNSTFVEKLLANEDD---LRE-FKQNAKNAGYMEEEDLIGKTMEQYSRTRIARY  315
Cni_g27847    FATTNLHSPKYTPELFWDKDN---PQGRLERVKENIEDI--HNNLAHQMIHFSDQRVNST  272
Cni_g27926    FATTNLHTPKYTPELFWDTGN---PQARFERVKENVHYP--DDNLAHQMVYFSDQRVASK  398
Cbr-XOL-1     FVSTNLHTPKYTPELFWDKEN---PKARFERVKENVHYP--DDNFAHQMVYFSDRRVAST  272
              :  *: *      .      .    :  :               .   :  .*

XOL-1         ----QEDYTCLEVEFDSQVALEKLM---NEHEQVEGFEVQQGGILVALKKDSFFDDELIE  319
Csp11_g5907   ----NRNKSCSTLDIDLFDALSKYYNENNADRLVKGFEVQPGGVMIAMKKNKIEKSKFPT  315
Csp11_g5908   GSNSKEKYECLPIEVDSYTALDKFRS----TISLAGVQVQQGGVLLVMKKGEYFNGGL-L  370
Cni_g27847    AIPP---SPIVKLTHQGL--HAVNRN----ESELIGFEIQQGGFLVVLKKGVFLADHTWM  323
Cni_g27926    SIPP---SPIVKLAHQGL--HAVNRN----ESDLIGFEVQQGGFLVVLKKGVFLADHTWM  449
Cbr-XOL-1     SIPP---SPIVKLTHQGL--HAVNRN----ESDLIGFEIQQGGFLVVLKKGVFLADHTWM  323
                            :  :                 : *.::* **.::.:**

XOL-1         --KIAYSAQYYLSMTHFSNRISIPLFSSLVFLTVSIVINAMCHKSIFCKRVISRLPFP--  375
Csp11_g5907   --IMKDIA------SSFKSNVSEVWFEVL---------RPGIRASIVDQGMISELKLG--  356
Csp11_g5908   EQIANDIA-----NGSR-TEISTITIDVIKLS-------GGVLLADQK--IALDMRKET-  414
Cni_g27847    IQIARKLA-----DNDESHSLEKIYFHLLE---------PGKQSGELSSSLLKSMDENHK  369
Cni_g27926    IQIASKLV-----ENDESNSLEEINFHLLE---------PGKQSAELSSSLLRSMDENHH  495
Cbr-XOL-1     IQIASKLV-----ENDESNSLEEIYFHLLE---------PGKRSEEWKPNLLKYMDENHH  369
                     .            :.   :  :                     :   :

XOL-1         -HCQILKLSHFST-------GRTFLSYLSI-IAKCTPISHI----NQSNILPAQNKIFAI  422
Csp11_g5907   -KYKFLSVTRPVAVTPEIDLKRKRHEEHGY-GNSCSDSEA-----E-------IDDLYDE  402
Csp11_g5908   -GYTFNVTAVE---ILQTCLKRSADDMLSEEGPSAKKKRGRR------RRNA--------  456
Cni_g27847    ATVTILKKKFE---FAE----PKVFGLTSEEESKAKPSPGKQQPRGSSNRSLSNDDTFDT  422
Cni_g27926    ATVSISKKTVQ---FAE----PKFFGLTSEEEPKTKPPSQKSKPRDPADRSLSNDDTFDT  548
Cbr-XOL-1     ATVSLKKKTVQ---FAE----PIFFGLTSEEEPKNKPPSHNPKPRDPADRSLSNDDTFDT  422
                  :                        .    . .
```

XOL-1	KQFS-	426
Csp11_g5907	TSTKK	407
Csp11_g5908	-----	456
Cni_g27847	ASTS-	426
Cni_g27926	AST--	551
Cbr-XOL-1	ASTS-	426

Appendix D

Conserved DCC-binding site alignment

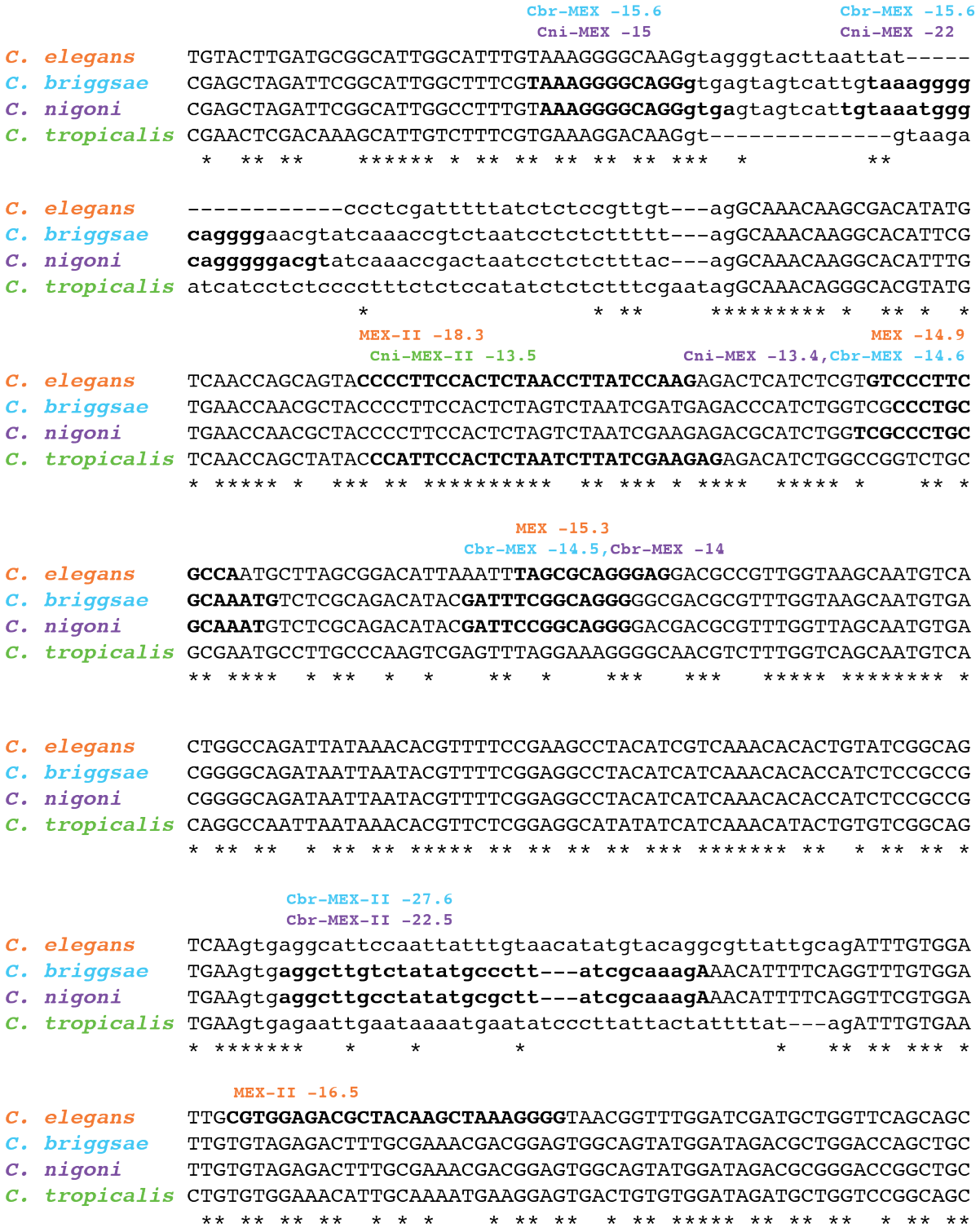


Figure D.1: One highly occupied site is conserved across all four species.

Appendix E

ChIP-seq libraries

Sample	Strain	Antibody	Conditions	Unique reads
CS070A	Cbr-AF16	rabbit α-Cbr-DPY-27	by hand	10258373
CS070B	Cbr-AF16	mouse IgG	by hand	10941228
CS070C	Cni-JU1325	rabbit α-Cbr-DPY-27	by hand	9262848
CS070D	Cni-JU1325	rabbit α-Cbr-MIX-1	by hand	9261745
CS070E	Cni-JU1325	mouse IgG	by hand	8998431
BMCS083A	Cni-JU1325	rabbit α-Cbr-DPY-27	1% fix, by hand	8913240
BMCS083B	Cni-JU1325	rabbit α-Cbr-MIX-1	1% fix, by hand	9075481
BMCS083C	Cni-JU1325	mouse IgG	1% fix, by hand	7460499
BMCS083D	Cni-JU1325	rabbit α-Cbr-DPY-27	750mM NaCl, by hand	8687406
BMCS083E	Cni-JU1325	rabbit α-Cbr-MIX-1	750mM NaCl, by hand	7144241
BMCS083F	Cni-JU1325	mouse IgG	750 mM NaCl, by hand	7149511
BMCS083G	Cni-JU1325	rabbit α-Cbr-DPY-27	500 mM NaCl, by hand	7996246
BMCS083H	Cni-JU1325	rabbit α-Cbr-MIX-1	500 mM NaCl, by hand	16204001
BMCS083I	Cni-JU1325	mouse IgG	500 mM NaCl, by hand	8772033
BMCS083J	Cni-JU1325	rabbit α-Cbr-DPY-27	300 mM NaCl, by hand	3935985
BMCS083K	Cni-JU1325	rabbit α-Cbr-MIX-1	300 mM NaCl, by hand	7755092
BMCS083L	Cni-JU1325	mouse IgG	300 mM NaCl, by hand	10178461
BMCS112A	Ctr-TY5743	mouse α-FLAG (Ctr-SDC-2)	1 min x 2 cycles (cryomill)	9049872
BMCS112B	Ctr-TY5743	mouse α-FLAG (Ctr-SDC-2)	2 min x 4 cycles (cryomill)	8052071
BMCS112C	Ctr-TY5743	mouse α-FLAG (Ctr-SDC-2)	2 min x 6 cycles (cryomill)	9034624
BMCS112D	Ctr-TY5743	mouse α-FLAG (Ctr-SDC-2)	3 min x 6 cycles (cryomill)	8002417
BMCS112E	Ctr-TY5743	rabbit IgG	2 min x 4 cycles (cryomill)	7008693
BMCS112F	Ctr-TY5752	mouse α-FLAG (Ctr-DPY-27)	13 min by hand	9271880
BMCS112G	Ctr-TY5752	mouse α-FLAG (Ctr-DPY-27)	1 min x 2 cycles (cryomill)	10373003
BMCS112H	Ctr-TY5752	mouse α-FLAG (Ctr-DPY-27)	2 min x 4 cycles (cryomill)	15304614
BMCS112I	Ctr-TY5752	mouse α-FLAG (Ctr-DPY-27)	2 min x 6 cycles (cryomill)	14036816
BMCS112J	Ctr-TY5752	rabbit IgG	2 min x 4 cycles (cryomill)	10193773
BMCS112K	Ctr-JU1373	mouse α-FLAG	13 min by hand	11302705
BMCS185A	Cel-N2	rabbit α-DPY-27 (DPY-27)	2 min x 4 cycles (cryomill)	41280400
BMCS185B	Cel-TY5753	mouse α-FLAG (DPY-27)	2 min x 4 cycles (cryomill)	44325254

Sample	Strain	Antibody	Conditions	Unique reads
BMCS185C	Cel-TY5753	mouse IgG	2 min x 4 cycles (cryomill)	37821232
BMCS185D	Ctr-TY5743	mouse α-FLAG (Ctr-SDC-2)	2 min x 4 cycles (cryomill)	43066623
BMCS185E	Ctr-TY5743	mouse IgG	2 min x 4 cycles (cryomill)	35343853
BMCS186A	Ctr-TY5752	mouse α-FLAG (Ctr-DPY-27)	2 min x 4 cycles (cryomill)	43399584
BMCS187B	Ctr-TY5752	mouse IgG	2 min x 4 cycles (cryomill)	57330743
BMCS187D	Cni-TY5754	mouse α-FLAG (Cni-DPY-27)	2 min x 4 cycles (cryomill)	35175581
BMCS187C	Cni-TY5754	mouse IgG	2 min x 4 cycles (cryomill)	42382763
BMCS187A	Ctr-TY5752, Cni-TY5754, Cel-TY5753	mouse α-FLAG (DPY-27)	2 min x 4 cycles (cryomill)	31259755, 14254509, 29824535
BMCS186C	Ctr-TY5752, Cni-TY5754, Cel-TY5753	mouse IgG	2 min x 4 cycles (cryomill)	24480241, 18009720, 28568479
BMCS186B	Ctr-JU1373, Cni-JU1325, Cel-N2	mouse α-FLAG (control)	2 min x 4 cycles (cryomill)	19604913, 19051473, 25297369
BMCS186D	Ctr-JU1373, Cni-JU1325, Cel-N2	mouse IgG	2 min x 4 cycles (cryomill)	12783784, 12845864, 17508899
BMCS203A	Ctr-TY5752, Cni-TY5754, Cel-TY5753, Cbr-TY	mouse α-FLAG (DPY-27)	2 min x 4 cycles (cryomill)	24030746, 28686578, 48474328, 33793810
BMCS203B	Ctr-TY5752, Cni-TY5754, Cel-TY5753, Cbr-TY	mouse IgG	2 min x 4 cycles (cryomill)	17709377, 28761936, 45987549, 31984954

Sample	Strain	Antibody	Conditions	Unique reads
BMCS204A	Ctr-JU1373, Cni-JU1325, Cel-N2, Cbr-AF16	mouse α-FLAG (control)	2 min x 4 cycles (cryomill)	12248167, 9859369, 24630221, 25352773
BMCS204B	Ctr-TY5743	mouse α-FLAG (Ctr-DPY-27)	2 min x 4 cycles (cryomill)	103226184
BMCS204C	Ctr-TY5743	mouse IgG	2 min x 4 cycles (cryomill)	83915035
BMCS205A	Cbr-TY5774	mouse α-FLAG (Cbr-SDC-2)	2 min x 4 cycles (cryomill)	21303658
BMCS205B	Cbr-TY5775	mouse α-FLAG (Cbr-DPY-27)	2 min x 4 cycles (cryomill)	160173398
BMCS205C	Cbr-TY5775	mouse IgG	2 min x 4 cycles (cryomill)	47600003
BMCS206A	Cni-TY5754	mouse α-FLAG (Cni-DPY-27)	2 min x 4 cycles (cryomill)	102579398
BMCS206C	Cni-TY5754	mouse IgG	2 min x 4 cycles (cryomill)	102382362

Table E.1: **ChIP experiment table.** This table includes the name of each ChIP library, the species and strain(s), the antibody and target for immunoprecipitation, the conditions, and the number of unique reads (separated by species).

Appendix F

Motif positions relative to ChIP-seq peaks

C. elegans motifs at *rex/dox* sites

rex/dox	site	peak rank	start	end	MEX (≥15)	MEX-II (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	top600 (≥25)
rex-01	IG	12	4395434	4395674	14.72	23.39	16.09		17.52			
rex-02	IG		1908940	1909086			14.59		12.11			
rex-03	C	406	11361204	11361318	14.62							
rex-04	C		11521744	11522154								
rex-05	C		11472531	11472774			15.56		12.68			
rex-06	3'	22	12362157	12364129	13.87 17.65		14.85				13.06	
rex-07	3'	20	11922233	11924309	12.12							
rex-08	IG	18	11090336	11095474	17.4 18.97	12.13 19.56	12.35 12.88 14.58					
rex-09	IG	2088	11682762	11689913								
rex-11	P	1927	11447281	11454464		13.55			12.46			
rex-13	3'	2169	12162920	12170355							12.26	22.24
rex-14	P	16	8036153	8037002	15.46 17.4	12.04 14.95						
rex-16	3'	21	11937383	11938556			12.62		16.12			
rex-17	IG	97	8047824	8050033	16.2							
rex-18	P	199	1378811	1381002		13.03						
rex-19	C	113	1491363	1493563	16.12							
rex-20	P	622	1681700	1683859			12.73		12.37 12.68			
rex-21	IG	8	1888103	1889829							12.37	
rex-22	C	1010	4009755	4011626	12.62						12.24	
rex-23	P	11	4208061	4210232	12.83 17.4 17.65	12.21		17.82		12.24		
rex-24	IG	15	7180798	7182818		14.74						
rex-25	C	267	8403257	8405357								
rex-26	IG	339	10352973	10354960					15.26			
rex-27	C	747	10621901	10624108								
rex-28	3'	110	10667026	10669154	15.77	19.3						
rex-29	C	188	10755366	10757490	15.95							
rex-30	C	489	11221742	11224062								
rex-31	C	273	12729135	12731379	12.49 16.3 17.65							
rex-32	P	10	2996004	2998096	17.65 18.97 18.97	12.15 12.4 21.89						

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of TSS, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. elegans motifs at *rex/dox* sites

rex/dox	site	peak rank	start	end	MEX (≥15)	MEX-II (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	top600 (≥25)
rex-33	C	14	6295287	6297381	13.23 15.45 15.46							
rex-34	C	13	5428461	5430561	14.86 15.25	16.54 18.3				13.04	12.17	
rex-35	IG	30	16680887	16683118	12.61 15.46							
rex-36	P	19	11898254	11900355	17.4				12.88			
rex-37	P	102	8810059	8812187	18.28							12.19
rex-38	3'	2339	5858592	5860766	17.4	12.45	14.58				12.2	
rex-39	C	39	14812187	14814187		20.85 21.3						
rex-40	IG	3	805232	807232	12.95 13.36 14.89		15.16		12.86		14.72	
rex-41	IG	531	17543789	17545789	13.0 13.0							
rex-42	IG	31	17180524	17182524	12.38		12.14		14.73			
rex-43	IG	25	13699827	13701827		18.83						
rex-44	C	4	1322307	1324307		24.89			12.37 14.25			
rex-45	P	5	1343906	1345906		21.48	14.58					
rex-46	IG	34	15735132	15737132	12.99		16.93		18.89			
rex-47	P	17	9465490	9467180	14.99							
dox-01	P	505	992186	994224				12.4			12.1 12.1 12.77 12.99 14.22	
dox-02	P	635	1913591	1915698					13.21		12.19	23.55
dox-03	C	756	2115780	2117802							13.7	
dox-04	P	353	2287187	2289387								19.72 23.13
dox-05	P	219	4253904	4256104								
dox-06	C	164	4264788	4266930								
dox-07	C	135	4388004	4390204								12.91
dox-08	C	272	5811217	5813303								14.46
dox-09	C	176	6840022	6842222					12.1			20.27
dox-10	P	396	7193086	7195286								
dox-11	P	37	8028040	8030207								
dox-12	P	238	8038417	8041199			12.81	12.02				16.44

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of TSS, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. elegans motifs at *rex/dox* sites

rex/dox	site	peak rank	start	end	MEX (≥15)	MEX-II (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	top600 (≥25)
dox-13	C	238	8041039	8047329							12.15	
dox-14	C	630	8050059	8053247								12.5
dox-15	C	506	8052696	8059136								12.5
dox-16	C	435	8555767	8558014				12.17			13.79	
dox-17	P	295	9271062	9273202								
dox-18	C	70	9336965	9339038								
dox-19	C	194	10187476	10189677					12.94			
dox-20	C	112	10523983	10526091								
dox-21	C	180	10554557	10556646								
dox-22	P	48	10567613	10569810								19.44
dox-23	P	374	10575531	10577669								17.48
dox-24	C	2128	10590425	10592599								
dox-25	P	1920	10594026	10596099					12.5			
dox-26	C	665	10596600	10598678	13.34			14.93			12.79	12.78
dox-27	C	972	10617664	10619901								
dox-28	P	1657	10628303	10630385								
dox-29	P	145	10636607	10638723								
dox-30	P	467	10678264	10680430								
dox-31	P	138	11206026	11208149								
dox-32	3'	1976	11209673	11211824					13.12			
dox-33	C	413	11214498	11216620								
dox-34	P	473	11247156	11249344								
dox-35	C	749	11251509	11253630								
dox-36	P	469	11256155	11258230	12.31 12.68							
dox-37	P	152	11261445	11263583								21.43
dox-38	P	1324	11289251	11291177				12.03				
dox-39	P	615	11296087	11298204								12.82
dox-40	C	71	11298792	11300952								19.52
dox-41	C	1214	11305724	11307886								
dox-42	C	860	11308835	11310973						12.13		
dox-43	P	420	11338120	11340208								
dox-44	P	1100	11364922	11367066								
dox-45	C	542	11367106	11369259							14.66	12.59
dox-46	P	23	12392897	12395090					12.24			
dox-47	P	334	12633999	12636190							12.86	13.57
dox-48	P	252	15724214	15726369								
dox-49	P	174	17183351	17185550								
Prex-1	C	1	409967	410411							12.84	

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of TSS, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. elegans motifs at *rex/dox* sites

rex/dox	site	peak rank	start	end	MEX (≥15)	MEX-II (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	top600 (≥25)
Prex-2	P	1943	1223811	1224252	12.22						16.06	
Prex-3	C	106	1302248	1302694								
Prex-5	P	5	1344723	1345339		21.48	14.58					
Prex-6	P	6	1389289	1389737								
Prex-7	P	7	1626688	1627339	13.13 13.27		16.93		13.53 20.98		12.9	
Prex-8	3'	879	2226807	2227469								
Prex-9	C	1285	2580361	2583697								
Prex-10	C	602	2731602	2732174								
Prex-11	C	53	2809447	2809974								
Prex-12	P		5908089	5910222								
Prex-13	P		5910404	5910979								
Prex-14	P	89	7334211	7334654							14.29	
Prex-15	C	195	8015963	8020064		15.65						
Prex-16	P	466	8737047	8737627								17.91
Prex-18	IG		12188324	12188853								
Prex-19	IG		12191486	12195475		12.29						
Prex-20	C	268	12452082	12452582								
Prex-21	P	428	12889607	12890339								
Prex-22	IG	24	13514210	13514861	13.77							
Prex-23	P	333	13696121	13697825								
Prex-24	IG	25	13700365	13701307		18.83						
Prex-25	P	515	13919448	13920850								
Prex-26	P	812	14026764	14027300		12.71						
Prex-27	C	107	14480082	14485107								
Prex-28	3'	26	14525403	14526539	16.58 17.4							
Prex-29	3'	34	15736448	15736947	12.99		16.93		18.89			
Prex-30	3'	28	16056164	16056821	13.36		14.05 14.08		13.18 16.47			
Prex-31	P	77	16205847	16206335		12.45 17.65	12.27					
Prex-32	IG	1533	16940563	16943412								
Prex-33	IG	4543	16943844	16944697								

Table F.1: Motifs at *C. elegans rex/dox* sites

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of TSS, 3' = 500 bp downstream of stop codon
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. elegans motifs at BMCS203A vs BMCS203B peaks

peak rank	*rex*	site	start	end	MEX (≥15)	MEX-II (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	top600 (≥25)
1	*Prex-1*	C	409940	410440							12.84 (193)	
2		P	768039	768539								26.05 (18)
3	*rex-40*	3'	806428	806928	12.95 (-1), 13.36 (-28), 14.89 (75)		15.16 (72)		12.86 (-68)			
4	*rex-44*	C	1322953	1323453		24.89 (-80)			14.25 (-103)			
5	*Prex-5*	P	1344809	1345309		21.48 (-32)	14.58 (-7)					
7	*Prex-7*	C	1626890	1627390	13.13 (-35), 13.27 (-53)		16.93 (7)		13.53 (-70), 20.98 (9)		12.9 (10)	
8	*rex-21*	IG	1888367	1888867							12.37 (-61)	
9		IG	1909305	1909805							13.0 (67)	
10	*rex-32*	IG	2996897	2997397	17.65 (-93), 18.97 (-42), 18.97 (133)	21.89 (-113)						
11	*rex-23*	P	4208953	4209453	17.4 (31), 17.65 (69)			17.82 (15)		12.24 (-12)		
12	*rex-01*	IG	4395356	4395856	14.72 (-12)	23.39 (22)	16.09 (53)		17.52 (55)			
13	*rex-34*	C	5429271	5429771	14.86 (-4), 15.25 (41)	16.54 (180), 18.3 (-24)				13.04 (-50)	12.17 (-46)	
14	*rex-33*	C	6296269	6296769	13.23 (-40), 15.45 (-3), 15.46 (25)							
15	*rex-24*	IG	7181505	7182005		14.74 (78)						
16	*rex-14*	P	8036178	8036678	15.46 (-38), 17.4 (-84)	12.04 (98), 14.95 (-33)						
17	*rex-47*	IG	9465527	9466027	14.99 (51)							
18	*rex-08*	IG	11093884	11094384	17.4 (18), 18.97 (-83)	12.13 (1), 19.56 (-102)	12.35 (-18), 14.58 (19)					
19	*rex-36*	P	11898945	11899445	17.4 (107)				12.88 (68)			

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of TSS, 3' = 500 bp downstream of stop codon
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. elegans motifs at BMCS203A vs BMCS203B peaks

peak rank	*rex*	site	start	end	MEX (≥15)	MEX-II (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	top600 (≥25)
21	*rex-16*	IG	11937214	11937714	15.34 (-83)		12.62 (-33)		16.12 (-37)			
22	*rex-06*	3'	12362858	12363358	13.87 (-145)		14.85 (-144)					
24	*Prex-22*	IG	13514305	13514805	13.77 (-63)							
25	*Prex-24*	IG	13700589	13701089		18.83 (-57)						
26	*Prex-28*	3'	14525580	14526080	16.58 (29), 17.4 (-8)							
27		P	15840978	15841478								12.3 (-220)
28	*Prex-30*	3'	16056155	16056655	13.36 (165)		14.05 (-12), 14.08 (111)		13.18 (-16), 16.47 (107)			
29		3'	16209605	16210105		18.88 (2)						
30	*rex-35*	IG	16681755	16682255	12.61 (13), 15.46 (86)							
31	*rex-42*	IG	17181094	17181594	12.38 (-61)		12.14 (41)		14.73 (43)			
34	*Prex-29*	3'	15736339	15736839	12.99 (5)		16.93 (-44)		18.89 (-48)			
35		P	4711819	4712319								18.22 (-48)
38		P	14746823	14747323					13.16 (5)			
39	*rex-39*	C	14813234	14813734		20.85 (-3), 21.3 (-20)						
41		P	6378278	6378778								14.05 (-1)
43		P	16147484	16147984							12.5 (211)	12.19 (-135)
46		P	1319760	1320260			13.2 (216)		12.11 (-242)		12.16 (-248)	12.96 (-26)
50		P	9958445	9958945								23.03 (93)
51		P	12017827	12018327								23.27 (8)
52		IG	1437032	1437532	13.6 (12)			15.6 (2)	12.32 (11)			
55		P	14169562	14170062								17.31 (27)
62		C	785299	785799							12.66 (-20)	
67		P	13686893	13687393		12.46 (86)						
71	*dox-40*	P	11299425	11299925								19.52 (-12)
72		P	1128438	1128938								28.1 (-31)
73		P	13527155	13527655				12.12 (-35)				
75		P	5841033	5841533								12.25 (-30)
76		P	6742078	6742578		12.65 (148)						
77	*Prex-31*	IG	16205906	16206406		17.65 (-18)	12.27 (-234)					

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of TSS, 3' = 500 bp downstream of stop codon
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. elegans motifs at BMCS203A vs BMCS203B peaks

peak rank	*rex*	site	start	end	MEX (≥15)	MEX-II (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	top600 (≥25)
79		P	14708024	14708524							14.27 (-64)	18.66 (33)
80		P	876889	877389				13.41 (70)				19.49 (-21)
81		P	16612510	16613010						12.71 (-71)		
87		P	14818544	14819044							12.49 (240)	
88		C	13144929	13145429							13.26 (-48)	
89	*Prex-14*	P	7334283	7334783							14.29 (-237)	
90		IG	5475032	5475532		12.81 (222)						
92		P	6207569	6208069							12.04 (-121)	
94		P	16327132	16327632	16.87 (-4)							
97	*rex-17*	IG	8048533	8049033	16.2 (5)							
102	*rex-37*	P	8810805	8811305	18.28 (-11)							12.19 (127)
108		P	1454269	1454769					12.65 (-77)			
110	*rex-28*	3'	10667768	10668268	15.77 (-37)	19.3 (-3)						
113	*rex-19*	C	1492104	1492604	16.12 (-14)							
115		P	7478352	7478852							12.03 (-61)	
117		P	1370820	1371320	13.29 (-17)							
121		P	949464	949964	12.06 (-100)							
125		P	5275141	5275641								14.35 (25), 22.02 (33)
128		P	15040596	15041096								20.98 (-31)
129		P	7823067	7823567	12.13 (10)	12.67 (83)						
133		P	3605637	3606137				12.36 (-190)				18.8 (23)
134		P	4606170	4606670								12.1 (250)
135	*dox-07*	P	4388873	4389373								12.91 (32)
136		P	16040080	16040580							13.9 (-29)	
137		P	2688493	2688993							12.42 (-3)	
139		P	17392030	17392530								21.2 (2)
143		P	5763800	5764300					12.03 (-42)		17.75 (-41)	16.67 (34)
144		P	13519015	13519515								15.19 (15)
146		P	14557197	14557697					16.11 (15)		13.01 (-180)	
149		P	14101855	14102355								13.82 (49), 15.04 (-57)

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of TSS, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. elegans motifs at BMCS203A vs BMCS203B peaks

peak rank	*rex*	site	start	end	MEX (≥15)	MEX-II (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	top600 (≥25)
151		P	1035013	1035513								23.78 (24)
152	*dox-37*	P	11262194	11262694								21.43 (1)
153		P	1856052	1856552								16.11 (81)
157		3'	9113648	9114148					13.01 (-36)		16.92 (-8)	
163		P	16331785	16332285								16.98 (30)
167		IG	1338857	1339357	15.95 (-26)							
168		3'	13665228	13665728							12.97 (-76)	
171		P	11467177	11467677								12.21 (-129)
172		P	14983158	14983658								20.74 (-10)
176	*dox-09*	P	6840755	6841255					12.1 (-37)			20.27 (-23)
177		P	11656534	11657034								19.28 (13)
181		P	8551622	8552122								25.39 (2)
182		IG	16024515	16025015							12.75 (-83)	
185		P	1450271	1450771								18.55 (-8)
188	*rex-29*	C	10756055	10756555	15.95 (70)							
189		IG	16726676	16727176								14.21 (73)
192		P	17434111	17434611				14.61 (-18)				
194	*dox-19*	C	10188355	10188855					12.94 (245)			
199	*rex-18*	IG	1379470	1379970		13.03 (6)						
201		P	14588963	14589463								19.11 (-22)
203		P	17151007	17151507				15.58 (-15)				
204		P	3694123	3694623				12.69 (-129)				
207		P	13726775	13727275								27.66 (20)
208		P	5685878	5686378					12.22 (3)			
209		3'	6097474	6097974	12.94 (-18)							
210		C	1937864	1938364				12.34 (4)				
211		IG	8205794	8206294								15.27 (-45)
212		P	16813867	16814367								21.07 (2)

Table F.2: Motifs at the top 200 *C. elegans* peaks on the X chromosome

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of TSS, 3' = 500 bp downstream of stop codon
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. briggsae motifs at *Cbr-rex*/flat sites

Cbr-rex/flat	site	peak rank	start	end	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
Cbr-rex-01 (Big peak, peak1, peak13)	C	3	10778971	10781620	14.47 14.63 15.57 15.57	27.58	13.44 14.23 14.81 15.09	12.67 15.06	12.49 13.48			
Cbr-rex-02 (Small peak)	P	20	12642606	12643183	14.25	22.69	13.62	13.78				
Cbr-rex-03 (peak4)	IG	5	19468419	19469368		12.36 20.04	13.65	13.2 13.43	13.65			
Cbr-rex-04 (peak2)	IG	2	6357847	6359296	13.8	19.09	15.45	16.3				
Cbr-rex-05 (peak6, peak9)	P	14	3152854	3153354		18.98		12.51 13.45				
Cbr-rex-06 (peak15)	IG	23	18811174	18811674	13.35	15.43	14.23	12.8 14.13				
Cbr-rex-07 (peak12)	IG	12	8026271	8026771	12.26 12.58 18.72		16.56 18.84	14.93 17.52				
Cbr-rex-08 (peak5)	C	29	16590530	16590979					13.52			
Cbr-rex-09 (peak8)	3'	1	3135115	3135615	12.8							
Cbr-rex-10 (peak3, peak7)	C		895711	895896								
Cbr-rex-11 (peak10)	C	167	4562615	4563115								
Cbr-rex-12 (peak17)	P	7	19564735	19565235			14.67					
Cbr-flat1 (ER333-334)	3'	11014	5887364	5889456						16.93		
Cbr-flat2 (ER331-332)	C		8040565	8042583						17.56		
Cbr-flat3 (ER335-336)	IG	13506	12488134	12490210		17.11						

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. briggsae motifs at *Cbr-rex*/flat sites

Cbr-rex/flat	site	peak rank	start	end	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
Cbr-flat4 (ER337-338)	C		20917253	20919096		19.27						

Table F.3: Motifs at *C. briggsae rex* sites

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. briggsae motifs at BMCS203A vs BMCS203B peaks

peak rank	*Cbr-rex*	site	start	end	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
1	*Cbr-rex-9*	3'	3135328	3135828	12.8 (-92)							
2	*Cbr-rex-04*	IG	6358388	6358888	13.8 (-6)	19.09 (-90)	15.45 (81)	16.3 (-146)				
3	*Cbr-rex-01*	C	10780259	10780759	14.47 (-19), 14.63 (-59), 15.57 (-191), 15.57 (-165)	27.58 (96)	13.44 (-63), 14.23 (-17), 14.81 (-163), 15.09 (-189)	12.67 (-106), 15.06 (69)	12.49 (-16), 13.48 (-188)			
4		IG	16577972	16578472	13.0 (-142), 13.22 (13), 13.52 (49), 14.31 (-56)			19.03 (108)	12.4 (2)	12.19 (-143)		
5	*Cbr-rex-03*	IG	19468465	19468965		12.36 (-163), 20.04 (15)	13.65 (53)	13.2 (-190), 13.43 (-60)	13.65 (75)			
6		3'	895208	895708	12.6 (-62)	14.68 (6)		15.4 (35)				
7	*Cbr-rex-12*	P	19564715	19565215			14.67 (-115)					
8		P	330040	330540								23.52 (-71)
10		P	10051062	10051562						12.67 (-19)		16.94 (-31)
12	*Cbr-rex-7*	IG	8026057	8026557	12.26 (107), 12.58 (155), 18.72 (22)		16.56 (24), 18.84 (157)	14.93 (100), 17.52 (237)				
14	*Cbr-rex-5*	P	3152694	3153194		18.98 (147)		12.51 (176), 13.45 (47)				
17		IG	9320883	9321383		15.32 (90)						
18		IG	14841429	14841929								15.14 (-72)
20	*Cbr-rex-02*	P	12642599	12643099	14.25 (-97)	22.69 (-32)	13.62 (-95)	13.78 (-3)				
22		P	3553718	3554218								18.62 (-135)
23	*Cbr-rex-06*	IG	18811101	18811601	13.35 (226)	15.43 (-101)	14.23 (222)	12.8 (100), 14.13 (-50)				
24		P	13557782	13558282					12.51 (-36)			
26		P	12819192	12819692					12.86 (-16), 14.83 (-79)			
29	*Cbr-rex-08*	C	16590490	16590990					13.52 (-71)			
31		C	18835197	18835697		26.23 (-7)						17.18 (-196)
32		P	4101074	4101574						12.28 (-3)		

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. briggsae motifs at BMCS203A vs BMCS203B peaks

peak rank	*Cbr-rex*	site	start	end	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
35		P	18204375	18204875						15.83 (3)		
40		C	2499811	2500311								15.91 (-61)
42		P	10573339	10573839							13.16 (-135)	
45		P	2670795	2671295		14.79 (-26)						21.91 (-9)
50		P	220173	220673								22.45 (25)
52		C	4102174	4102674		12.06 (-53)						
54		P	18959177	18959677		16.82 (50)						
58		P	10630781	10631281						12.54 (-38)		13.34 (45)
61		P	15452734	15453234		20.63 (-130)						
62		P	10191350	10191850						13.3 (20)		
64		3'	8157690	8158190								21.07 (-39)
66		P	18822418	18822918		25.34 (24)						16.38 (-176)
67		P	12526700	12527200								18.24 (10)
68		P	12299540	12300040								19.95 (-41)
69		P	9819329	9819829					13.26 (41)			
70		P	10633891	10634391					15.28 (-35)		12.61 (169)	
71		P	18524878	18525378		13.2 (-59)						
73		IG	1317264	1317764			12.07 (245)					
74		C	19496825	19497325								17.24 (34)
75		P	9683755	9684255								12.63 (45)
78		P	7068359	7068859								16.02 (41)
79		3'	22389	22889					12.83 (-216)			17.66 (89)
80		C	6486981	6487481					12.85 (-128), 13.49 (171)			22.95 (-85)
81		P	3334401	3334901					12.83 (226)			23.16 (-3)
83		IG	2089624	2090124			12.84 (53)	12.62 (-59)				
86		3'	3060774	3061274								12.09 (6)
90		IG	2833018	2833518					12.06 (183)			
91		P	17595228	17595728		14.16 (-35)						

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. briggsae motifs at BMCS203A vs BMCS203B peaks

peak rank	*Cbr-rex*	site	start	end	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
92		P	18457563	18458063						12.93 (-60)		12.31 (-208)
95		P	7144745	7145245					12.13 (151)			
96		3'	18354874	18355374							16.94 (-249)	
98		P	18260793	18261293								18.29 (-67)
101		P	11289945	11290445								16.97 (20)
102		P	11734324	11734824					12.13 (87)			
103		P	17701203	17701703								20.25 (30)
104		P	1305779	1306279								14.09 (14)
105		P	15426933	15427433					12.33 (-135)			
107		P	17711626	17712126					15.03 (-241)			
108		C	199459	199959								12.86 (60)
115		P	10499947	10500447							13.58 (-41)	
122		P	13983253	13983753								16.11 (-60)
124		P	17699699	17700199								24.09 (-2)
126		P	8178994	8179494					12.51 (62)			
127		C	4587894	4588394								19.26 (-75)
130		P	5353011	5353511								16.03 (-6)
131		P	15053270	15053770								18.77 (13)
132		IG	9610991	9611491		12.15 (-158)						
133		P	16781908	16782408								17.33 (47)
135		IG	10822437	10822937						13.42 (37)		12.49 (204)
140		P	1955536	1956036			12.05 (-111)					
141		P	19236535	19237035								22.68 (-5)
144		P	14502830	14503330					13.11 (-68)			
145		P	18690262	18690762								12.86 (10)
147		P	1314999	1315499				12.56 (-180)				
149		P	17819258	17819758								21.8 (-22)
150		3'	6843186	6843686								12.57 (207)
153		P	9446929	9447429								21.05 (-60)
154		C	11968106	11968606					12.51 (-142), 12.7 (3)			

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. briggsae motifs at BMCS203A vs BMCS203B peaks

peak rank	*Cbr-rex*	site	start	end	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
157		P	6460961	6461461								19.28 (-23)
158		P	15465391	15465891								19.37 (-55)
159		IG	20907155	20907655						13.26 (70)		
165		P	12196692	12197192		12.12 (-118)						
166		P	19696213	19696713						12.35 (161)		
169		P	18562422	18562922	12.18 (-197)							18.91 (5)
170		P	15880960	15881460								21.67 (-28)
173		P	8314659	8315159					13.35 (12)			
177		P	10151240	10151740								14.28 (18)
179		P	19225259	19225759								17.61 (-35)
181		P	4960597	4961097	12.45 (141)	13.63 (161)			14.48 (73)	15.39 (73)		
183		3'	13247957	13248457								22.99 (-28)
185		P	14781711	14782211								18.0 (-8)
188		P	15558505	15559005								15.94 (-88)
192		P	12564704	12565204								19.03 (-55)
193		P	13453770	13454270								24.5 (-48)
194		IG	891181	891681			12.55 (-115), 15.49 (55)					
195		IG	5751938	5752438					12.34 (64)			
197		P	16755276	16755776								18.91 (50)
200		IG	3084882	3085382					13.18 (99)			

Table F.4: Motifs at the top 200 *C. briggsae* peaks on the X chromosome

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. nigoni motifs at BMCS203A vs BMCS203B peaks

peak rank	site	start	end	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
1	C	15784955	15785455	13.39 (-47), 15.05 (-173), 22.02 (-147)	12.98 (85), 15.38 (-90)	13.95 (-3), 14.56 (-43), 15.5 (-175), 16.77 (-149)	22.54 (112)			13.4 (-172)	
2	IG	22181312	22181812	13.25 (3), 15.18 (-59), 18.19 (39)	14.49 (0), 27.06 (102)	18.67 (-61), 18.67 (43)	14.2 (-97), 14.2 (63)			13.06 (-58), 13.9 (-10)	
3	IG	26301384	26301884	19.22 (47)	12.75 (38), 18.32 (-66)		20.15 (9)			15.05 (48)	
4	P	8079237	8079737	14.77 (53)	20.65 (-165)		20.31 (-109)				
5	IG	345612	346112								18.68 (-92)
7	IG	1307681	1308181	13.11 (-33)							
8	IG	842520	843020		14.35 (29)	12.58 (-68)	15.21 (0)				
10	C	17716573	17717073	12.56 (35), 17.03 (-85)	16.58 (-2)	14.21 (-87)	23.57 (-31)				
11	P	13068024	13068524	18.25 (-191), 18.86 (-58)	16.23 (-47), 22.87 (20)	12.55 (-60), 16.07 (-108), 18.67 (-193)	14.53 (-20)			14.19 (-190)	
12	IG	3971909	3972409			12.78 (-82)					
14	IG	18760715	18761215							12.43 (-58)	
17	C	17887340	17887840							14.74 (-98)	
18	IG	4003926	4004426		13.73 (128)		15.07 (99)			13.05 (75)	
20	IG	20047977	20048477								18.01 (3)
23	IG	24950827	24951327				21.03 (-9)		13.1 (-103), 13.5 (37)		15.91 (-214)
26	C	14842482	14842982							13.12 (52)	
28	IG	3338467	3338967								13.25 (-22), 18.59 (-54)
30	C	26387444	26387944	14.63 (-146)							
34	IG	24401219	24401719					15.76 (-13)			
35	IG	238173	238673								22.24 (36)
36	IG	4857268	4857768								22.4 (-67)
39	IG	4199888	4200388								23.07 (13)
43	C	25984682	25985182	15.58 (-198)							24.05 (-9)
44	IG	5064418	5064918	13.42 (197)							
45	P	17258717	17259217								17.63 (3)

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. nigoni motifs at BMCS203A vs BMCS203B peaks

peak rank	site	start	end	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
48	P	12747749	12748249				13.87 (-157)				
56	C	9178970	9179470	12.15 (250)							
57	P	14701604	14702104								12.57 (46)
58	P	17600321	17600821								19.83 (15)
59	IG	22192347	22192847					12.4 (22)			
61	IG	51831	52331								18.04 (71)
62	IG	24343063	24343563							13.44 (242)	16.85 (24)
65	C	15639997	15640497					12.49 (8)			13.26 (91)
67	C	6421651	6422151					12.42 (-20)			
68	IG	23769858	23770358							12.57 (-232)	
69	IG	24534184	24534684							16.1 (69)	
72	P	16760677	16761177	13.48 (-60)							
75	C	15826154	15826654					13.38 (45)	12.08 (-156)		12.45 (212)
76	IG	23874242	23874742								23.33 (-18)
82	IG	3657244	3657744							12.93 (99)	
86	IG	5368872	5369372								17.21 (-40)
88	IG	1269664	1270164					12.61 (-16)			
92	C	17217173	17217673				12.0 (-123)				
95	IG	1260474	1260974								13.97 (25)
96	C	15087568	15088068								16.01 (54)
97	IG	23608515	23609015							12.68 (100)	
98	P	17334121	17334621		12.53 (172)						24.04 (-39)
100	C	8220545	8221045				17.56 (171)				
102	C	7344596	7345096							12.31 (42)	
104	IG	24602185	24602685				12.83 (-30)				
105	C	26329495	26329995								17.14 (2)
106	IG	14469926	14470426	16.52 (27)							
108	C	6470022	6470522			12.39 (19)		15.33 (-44)		14.38 (-44), 14.6 (-197)	
110	IG	219566	220066							12.14 (113)	
111	IG	20136689	20137189								16.2 (-66)
112	P	17634832	17635332								19.82 (-186)
114	C	25813407	25813907				12.13 (146)				
115	P	8186258	8186758								19.89 (-23)

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. nigoni motifs at BMCS203A vs BMCS203B peaks

peak rank	site	start	end	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
117	P	6958572	6959072								17.87 (15)
120	IG	23757803	23758303								19.68 (7)
122	C	16786844	16787344								23.47 (-34)
127	IG	24422296	24422796								18.97 (-31)
131	IG	23759270	23759770								21.66 (13)
132	C	16985130	16985630							12.65 (0)	13.24 (-43)
133	IG	6029463	6029963							12.33 (56)	
135	IG	20390508	20391008								22.71 (-7)
138	IG	22465841	22466341				12.43 (170)				
139	C	27445649	27446149								14.33 (-29)
141	IG	838282	838782	15.47 (44)							
144	IG	22344545	22345045							12.34 (-109)	21.5 (16)
150	C	15713122	15713622								12.05 (169)
151	IG	19181267	19181767								17.02 (-83)
161	P	6360016	6360516								12.89 (-37)
163	C	12806548	12807048						14.75 (10)		
168	IG	8134429	8134929								20.56 (-32)
171	C	8823991	8824491								16.45 (-16)
175	IG	19755441	19755941								15.45 (26)
176	IG	24759250	24759750								13.79 (-8)
177	P	16331104	16331604								18.43 (19)
178	IG	4883976	4884476							12.68 (-218)	
180	IG	18841261	18841761								17.28 (-26)
182	IG	267601	268101								21.85 (0)
184	IG	374260	374760							12.21 (-73), 15.43 (-122)	
185	IG	894717	895217								18.02 (132)
188	C	6362125	6362625								12.15 (-159)
189	IG	18401410	18401910								21.71 (-19)
190	IG	19731078	19731578								20.99 (-9)
193	IG	18400933	18401433							12.72 (-190)	
194	IG	1396624	1397124								13.51 (-243)

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

***C. nigoni* motifs at BMCS203A vs BMCS203B peaks**

peak rank	site	start	end	Cni-MEX (≥16)	Cni-MEX-II (≥17)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Ctr-MEX (≥16)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
198	C	14945575	14946075						12.96 (186)		
201	IG	20790644	20791144							12.26 (-165)	
205	IG	24635391	24635891							12.54 (179)	12.3 (175), 18.63 (3)
Autosomal site on chrIII											
173	3'	16369630	16370130								25.92 (30)
Autosomal site on chrV											
187	3'	11321784	11322284								13.05 (25)

Table F.5: Motifs at the top 200 *C. nigoni* peaks on the X chromosome

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. tropicalis motifs at BMCS203A vs BMCS203B peaks

peak rank	site	start	end	Ctr-MEX (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
1	P	176816	177316	13.95 (-91)		12.62 (60)		12.25 (33)			
2	P	293337	293837	13.83 (28)							
3	C	461844	462344	21.78 (13)							
4	P	1078823	1079323	12.12 (-80)							
5	P	1756951	1757451		12.79 (30)			14.39 (23)			
6	C	2452520	2453020	18.07 (3)							
7	C	3262666	3263166	18.78 (46)							
8	C	3486186	3486686	17.92 (6)	13.22 (3)		12.7 (5)				
9	C	4087861	4088361		12.35 (83)		12.19 (213), 14.4 (85)				
10	P	4391654	4392154	17.91 (28)							
11	P	5068569	5069069	12.96 (-51)							
13	IG	6152943	6153443	16.0 (-10)	13.5 (-3)		12.11 (200), 12.74 (-7)			12.2 (150), 20.78 (15)	
14	IG	7626998	7627498	24.63 (-53)			12.45 (-46)				
15	IG	8410557	8411057	25.68 (89)	12.89 (86)		13.04 (88)				
16	IG	8792617	8793117	12.09 (44), 13.52 (-52)							
17	P	9686530	9687030	24.03 (-10)							
18	IG	10913737	10914237	12.13 (-46), 20.24 (-98)	12.89 (-87)		13.04 (-91)				
20	C	14211893	14212393					13.54 (-86)			
21	P	1748535	1749035	12.23 (-18)							
24	C	13350564	13351064	14.54 (80)							
25	P	4121492	4121992	15.55 (-28)							
26	C	1982129	1982629			16.71 (-43)		13.06 (-70)			
31	C	166541	167041	13.11 (-40)							
32	C	11731509	11732009								17.6 (-36)
33	C	400989	401489	13.82 (1)							23.21 (157)
35	C	6447295	6447795								12.48 (-84)
40	3'	7007269	7007769							14.14 (60)	
41	P	2895403	2895903	12.9 (31)							12.39 (-61)
42	C	723790	724290								18.93 (53)
54	C	156736	157236			12.15 (31)					
58	P	2384229	2384729	14.13 (56)							
61	C	6783020	6783520								21.93 (-77)
68	C	6251889	6252389				13.15 (-180)				
72	C	2061061	2061561							12.83 (22)	
73	P	242389	242889		13.43 (-86)		13.13 (-174)		12.11 (-87)		
76	P	8281689	8282189							13.6 (-118)	

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

C. tropicalis motifs at BMCS203A vs BMCS203B peaks

peak rank	site	start	end	Ctr-MEX (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
79	C	9232365	9232865								19.36 (19)
80	P	7086927	7087427								13.91 (42)
85	C	10275698	10276198	12.53 (186), 15.67 (209)							
89	C	9135882	9136382			12.41 (-27)					15.09 (53)
92	C	2126124	2126624			13.2 (63)					
93	IG	5446571	5447071		12.24 (162)						
96	C	5305827	5306327	12.66 (-45)							
97	P	6890883	6891383								21.77 (-13)
99	C	12849573	12850073				15.18 (2)				
100	P	1184225	1184725								16.04 (-45)
108	P	5621653	5622153								27.85 (8)
114	P	11017501	11018001								20.44 (-28)
115	IG	117122	117622						12.07 (-26)		
117	P	2748695	2749195								20.63 (-1)
118	P	2505433	2505933			12.25 (-22)					14.21 (-38)
124	C	2519551	2520051	13.26 (-6)							
125	P	14247304	14247804	12.14 (170)							
126	C	2167251	2167751			13.7 (51)					16.28 (-58)
127	3'	195547	196047					12.38 (-66)			
128	P	11107550	11108050								21.41 (15)
129	C	5364936	5365436								17.71 (19)
130	P	2482284	2482784				15.53 (14)				
133	C	11190456	11190956								15.48 (8)
134	P	13415802	13416302								12.2 (7)
135	P	6261734	6262234	13.23 (-124)							
136	C	882664	883164								20.05 (11)
142	P	10843751	10844251			17.15 (-164)					
145	P	14187560	14188060								13.82 (-75)
150	C	2580796	2581296								19.68 (-23)
151	P	2352173	2352673						12.19 (44)		
152	C	7385902	7386402								16.25 (-22)
161	3'	271339	271839	12.5 (-174)							
163	P	306793	307293	12.67 (64)							15.61 (-41)
167	C	11475854	11476354				12.14 (-39)				
170	P	11471958	11472458								12.76 (44)
176	C	7520906	7521406	14.44 (223)							
179	P	8471761	8472261			12.55 (-98)					
180	C	1931158	1931658								12.05 (1)

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon

Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

***C. tropicalis* motifs at BMCS203A vs BMCS203B peaks**

peak rank	site	start	end	Ctr-MEX (≥16)	Cbr-MEX (≥15)	Cbr-MEX-II (≥16)	Cni-MEX (≥16)	Cni-MEX-II (≥17)	MEX (≥15)	MEX-II (≥16)	top600 (≥25)
181	P	4611371	4611871	12.08 (-73)							16.27 (138)
182	P	127552	128052	13.25 (-54)							
183	C	3192224	3192724								22.37 (-43)
191	C	11648759	11649259	14.46 (6)							16.17 (45)
192	P	3105319	3105819								25.96 (-11)
193	P	8026087	8026587								14.92 (74)
194	C	2117729	2118229	13.64 (-8)	12.04 (-11)						
195	C	6226988	6227488								20.22 (-22)
197	C	1094673	1095173	13.25 (33)							
198	P	14464973	14465473								12.81 (-43), 18.38 (27)

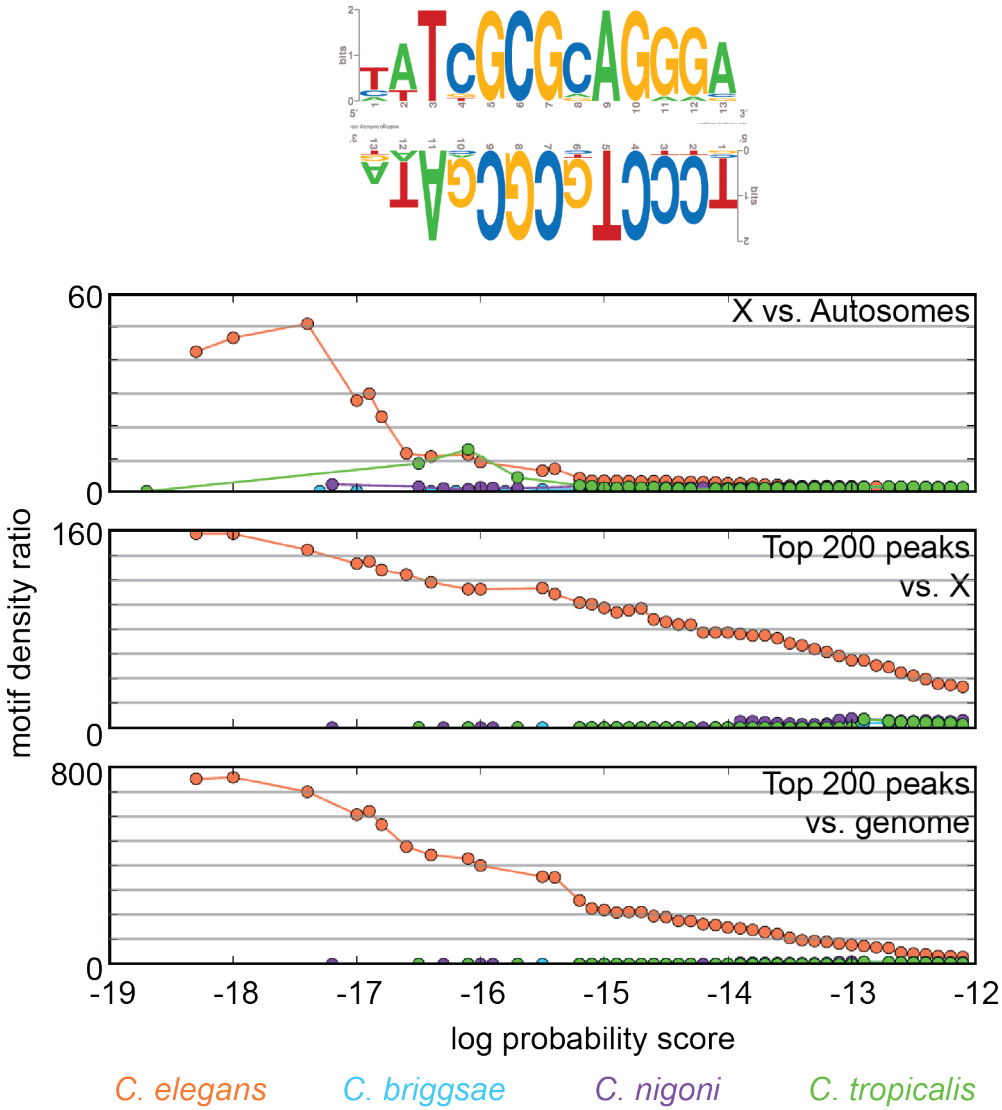
Table F.6: Motifs at the top 200 *C. tropicalis* peaks on the X chromosome

abbreviations: IG = intergenic, C = coding, P = 2 kb upstream of ATG, 3' = 500 bp downstream of stop codon
Motif scores are shown in absolute values of the log probability score. All motifs with scores better than 12 are shown.Highlighted scores are better than or equal to the cutoffs listed below each motif in the header. Peaks were called with 4 species tagged *dpy-27* pooled (FLAG vs. IgG). Distance in base pairs from the peak summit to the motif midpoint are shown in parentheses. Negative numbers are upstream of the summit in the reference genome assembly.

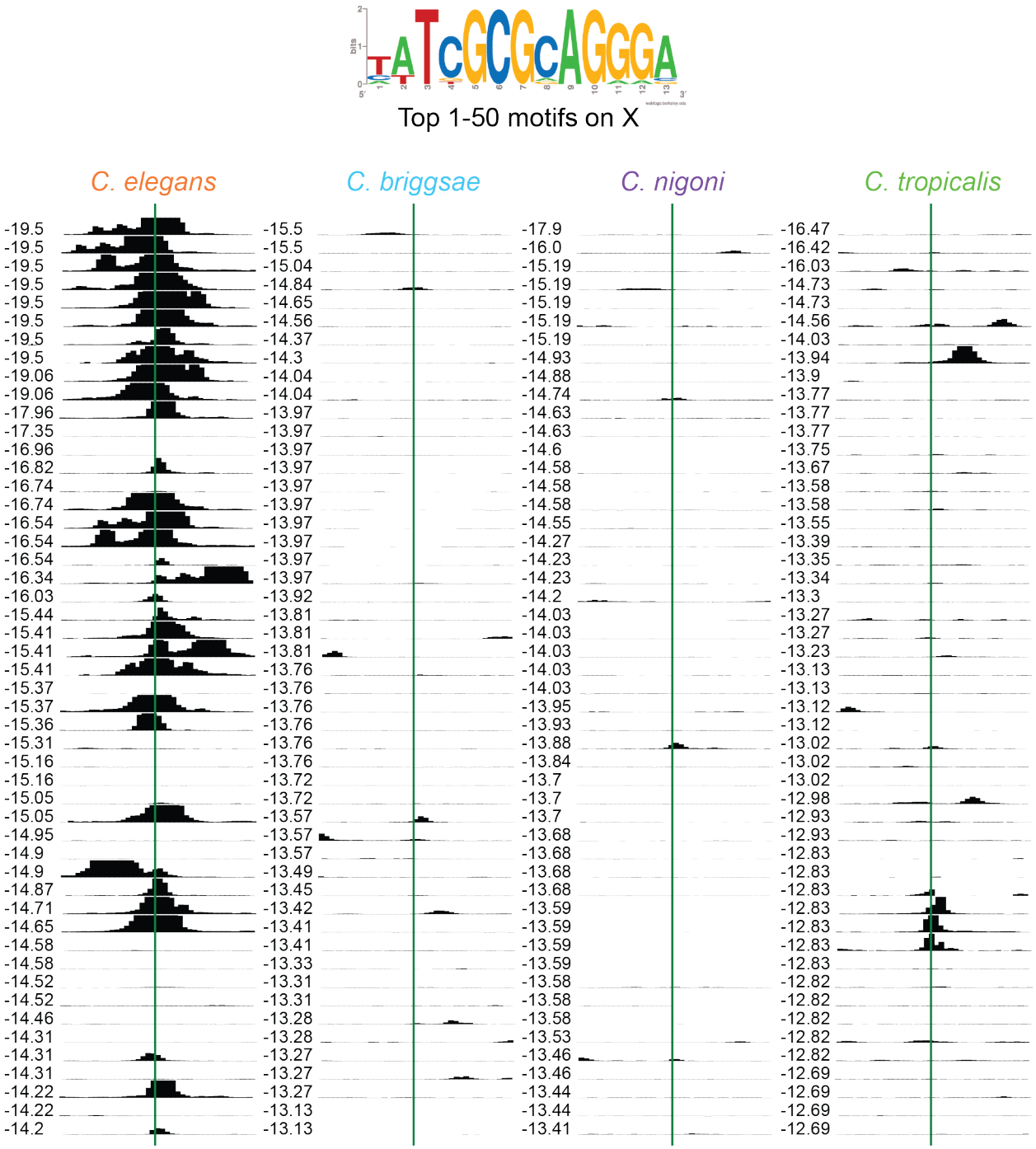
Appendix G

Additional motifs

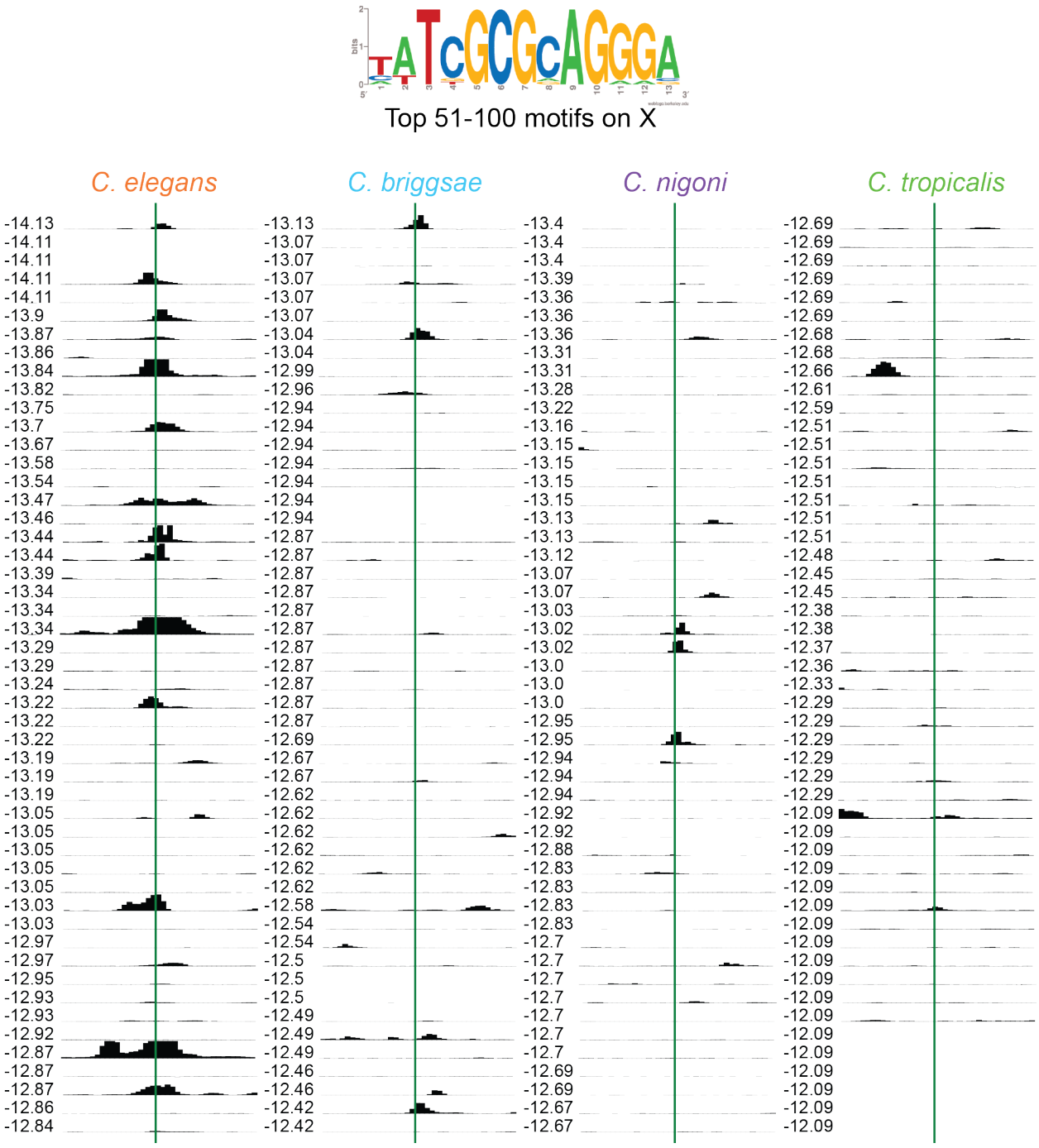
Figure G.1: Cel-bMEX-13bp a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cel-bMEX-13bp motif is X-enriched and peak-enriched only in *C. elegans*.** The Cel-bMEX-13bp motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

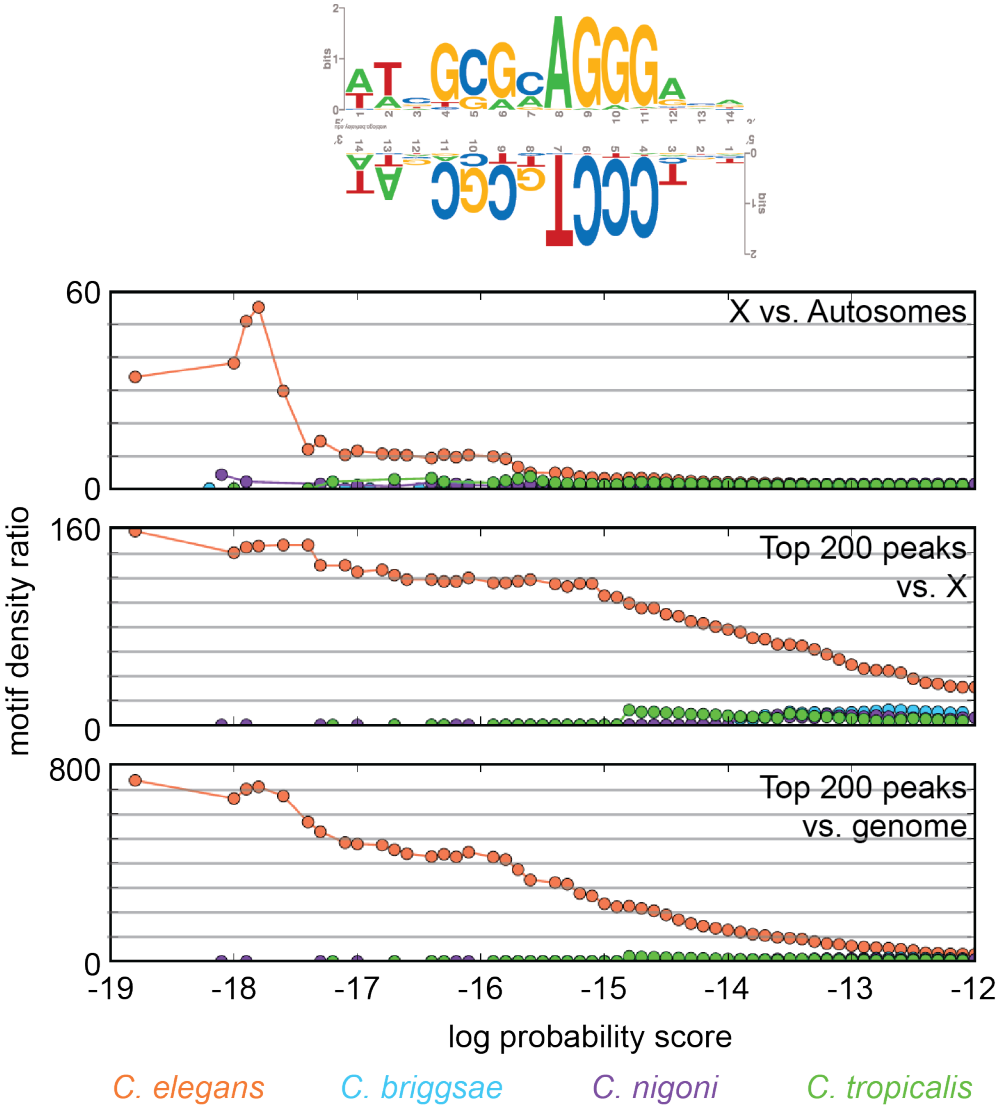


(b) **The *C. elegans* DCC is bound at most of the top 50 Cel-bMEX-13bp motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Cel-bMEX-13bp motifs on the X chromosome in four species.

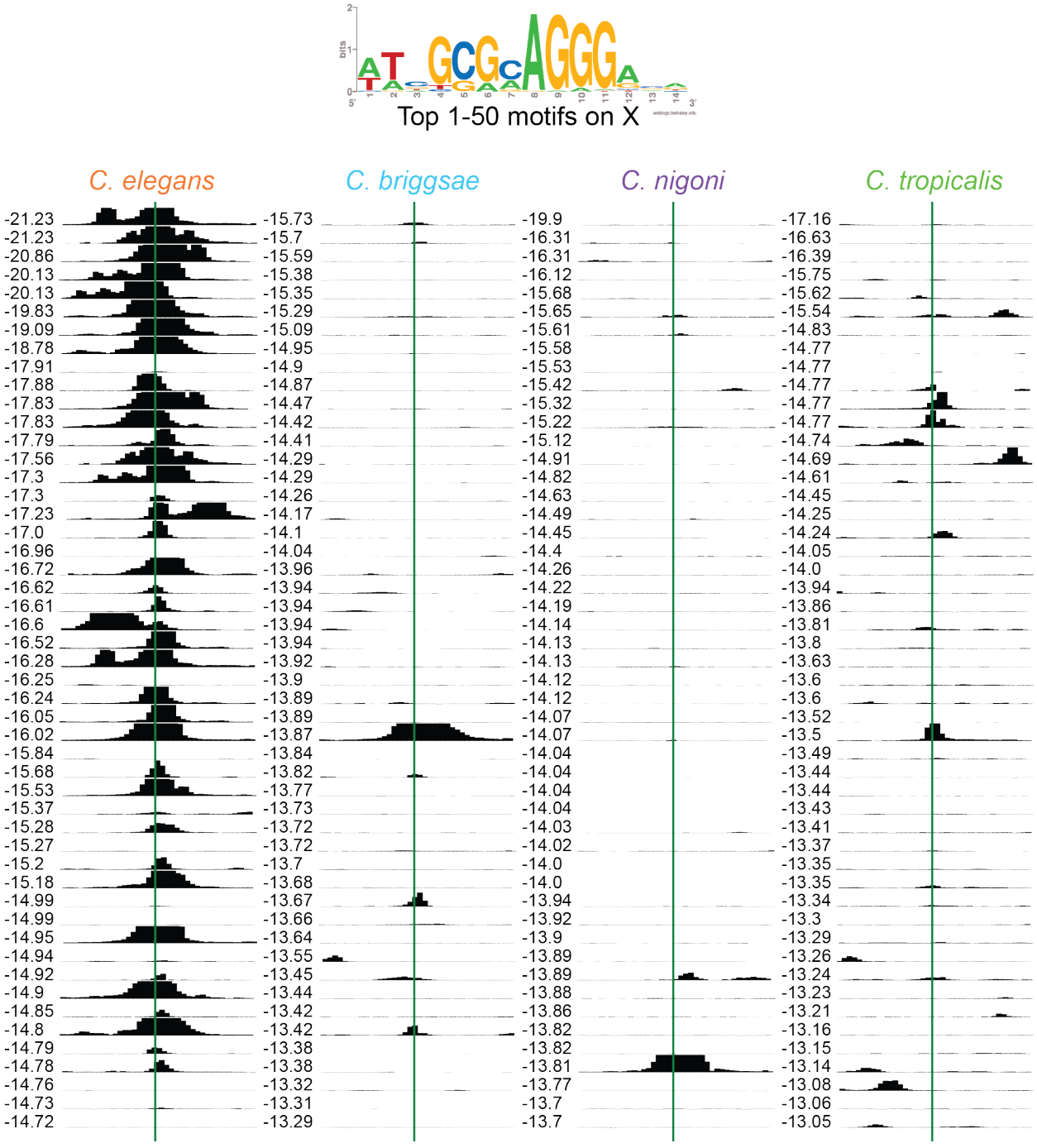


(c) **The *C. elegans* DCC is also bound at many of the top 51-100 Cel-bMEX-13bp motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cel-bMEX-13bp motifs on the X chromosome in four species.

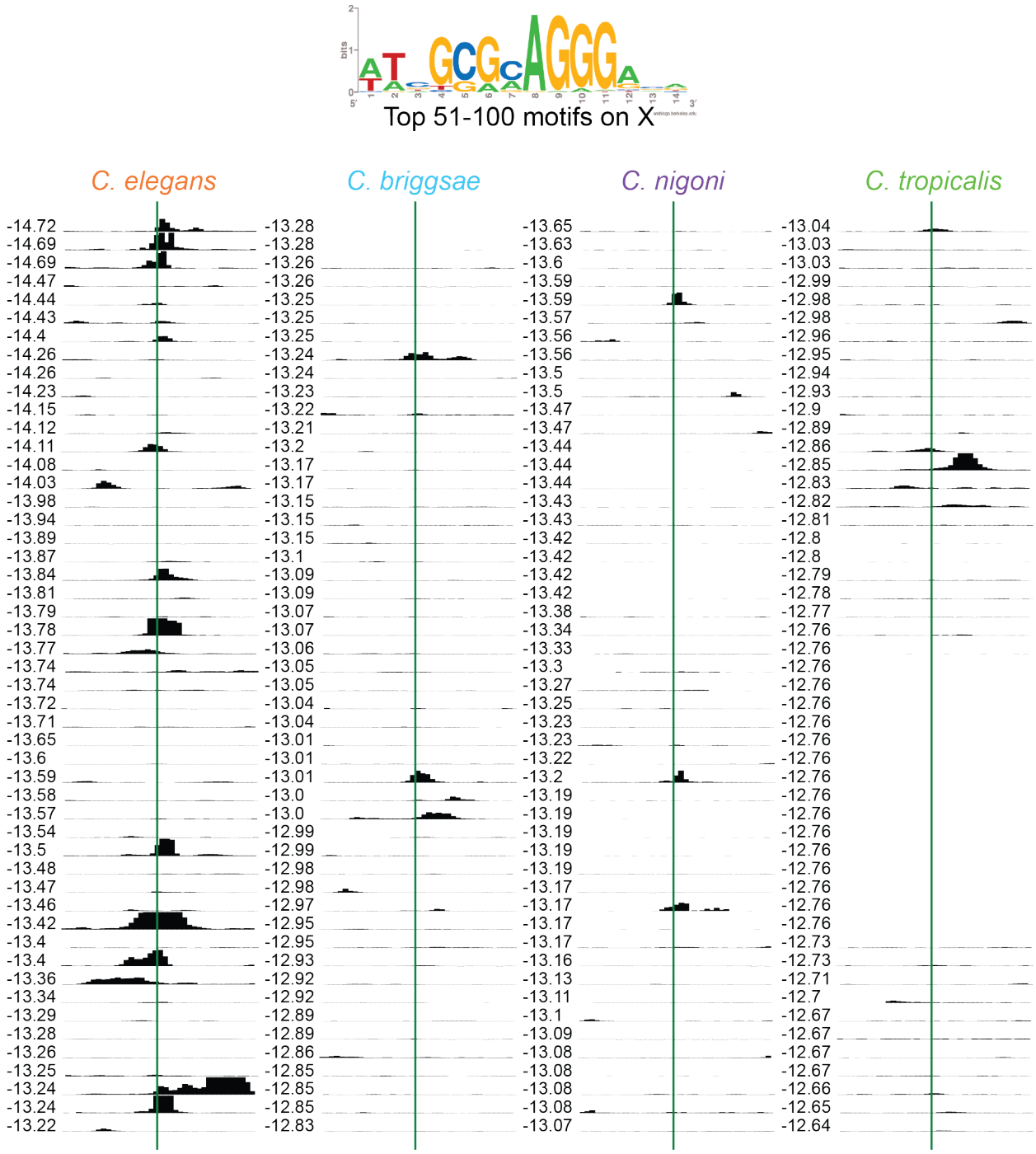
Figure G.2: CS181 Cel-MEX a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The CS181 Cel-MEX motif is X-enriched and peak-enriched only in *C. elegans*.** The CS181 Cel-MEX motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

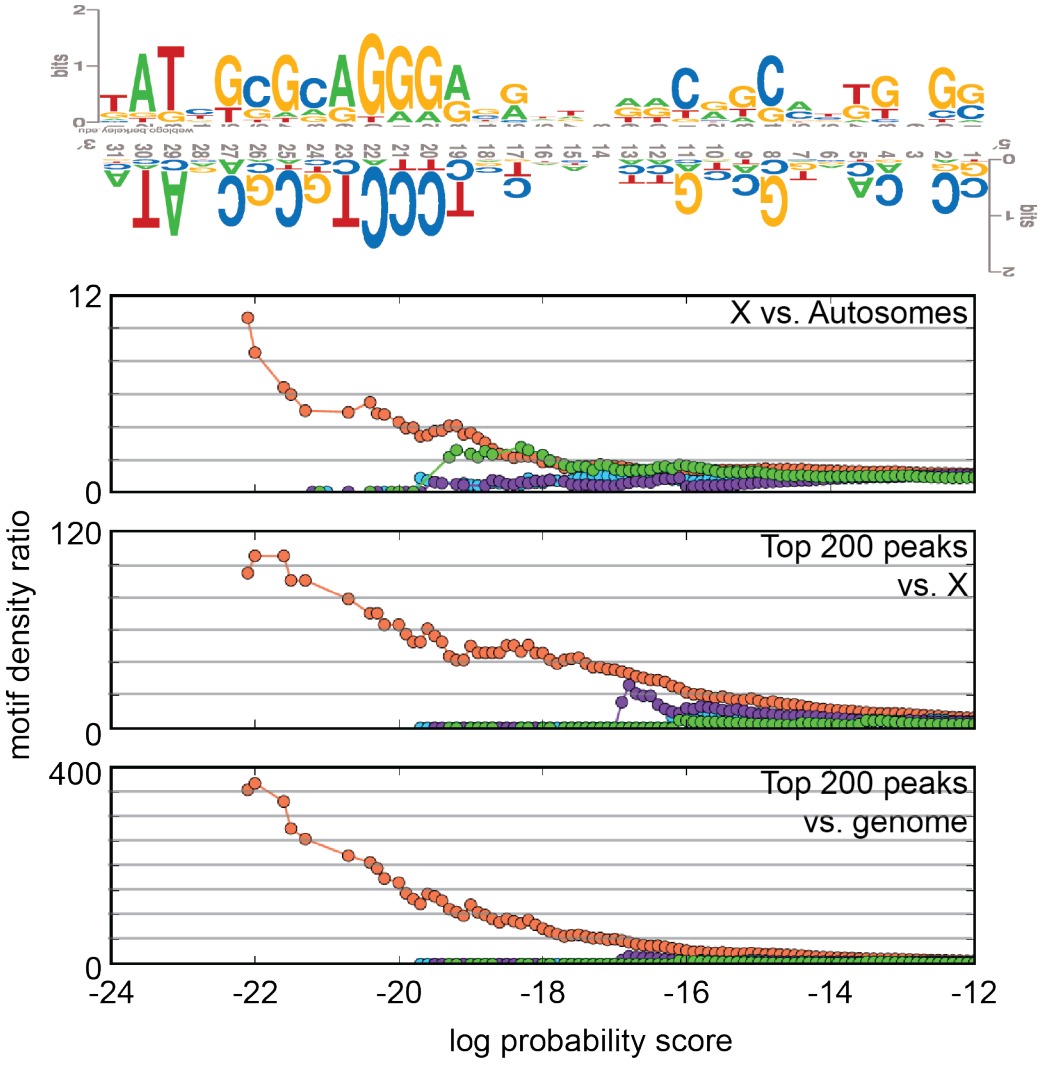


(b) **The *C. elegans* DCC is bound at most of the top 50 CS181 Cel-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 CS181 Cel-MEX motifs on the X chromosome in four species.

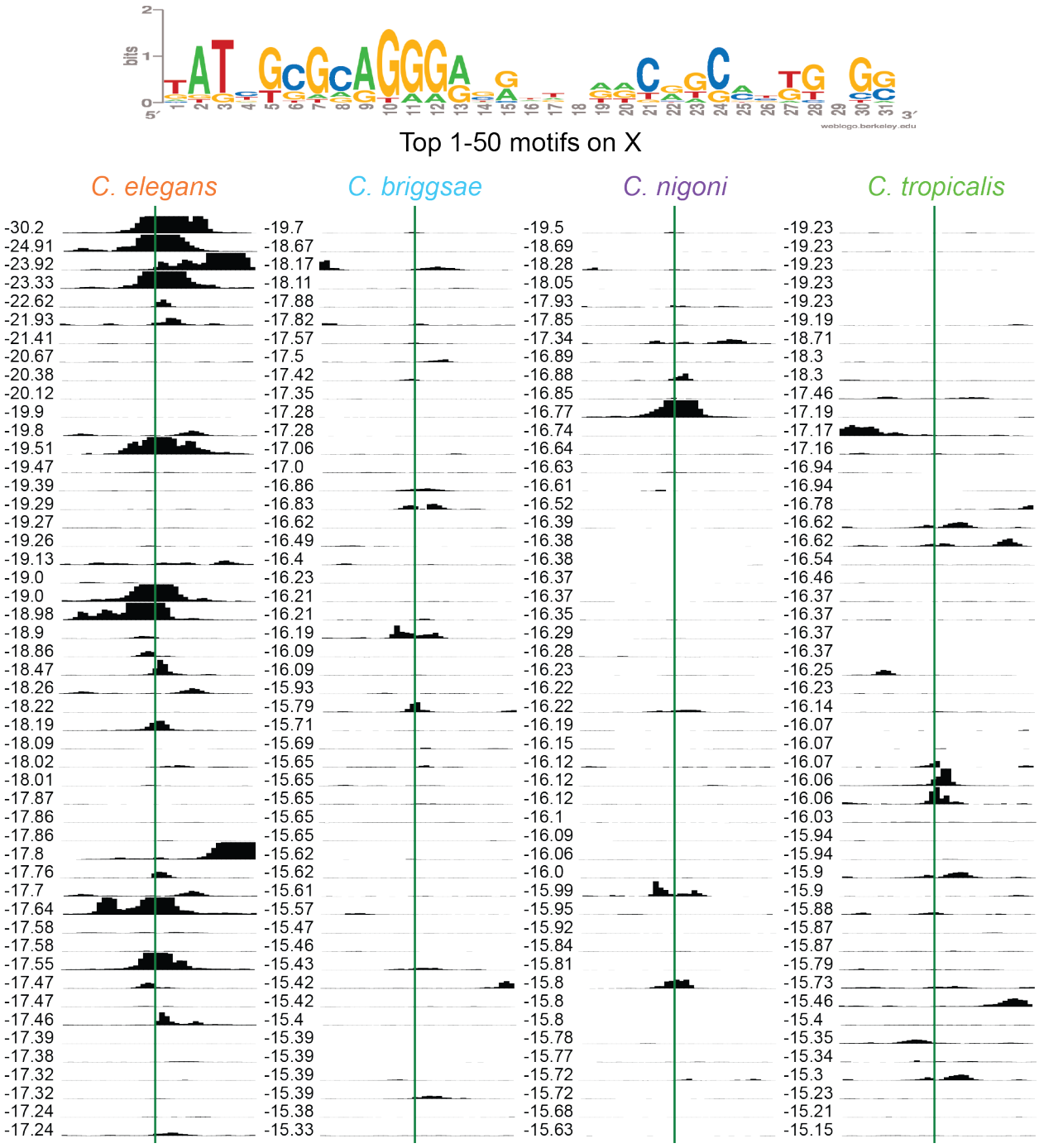


(c) **The *C. elegans* DCC is also bound at many of the top 51-100 CS181 Cel-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 CS181 Cel-MEX motifs on the X chromosome in four species.

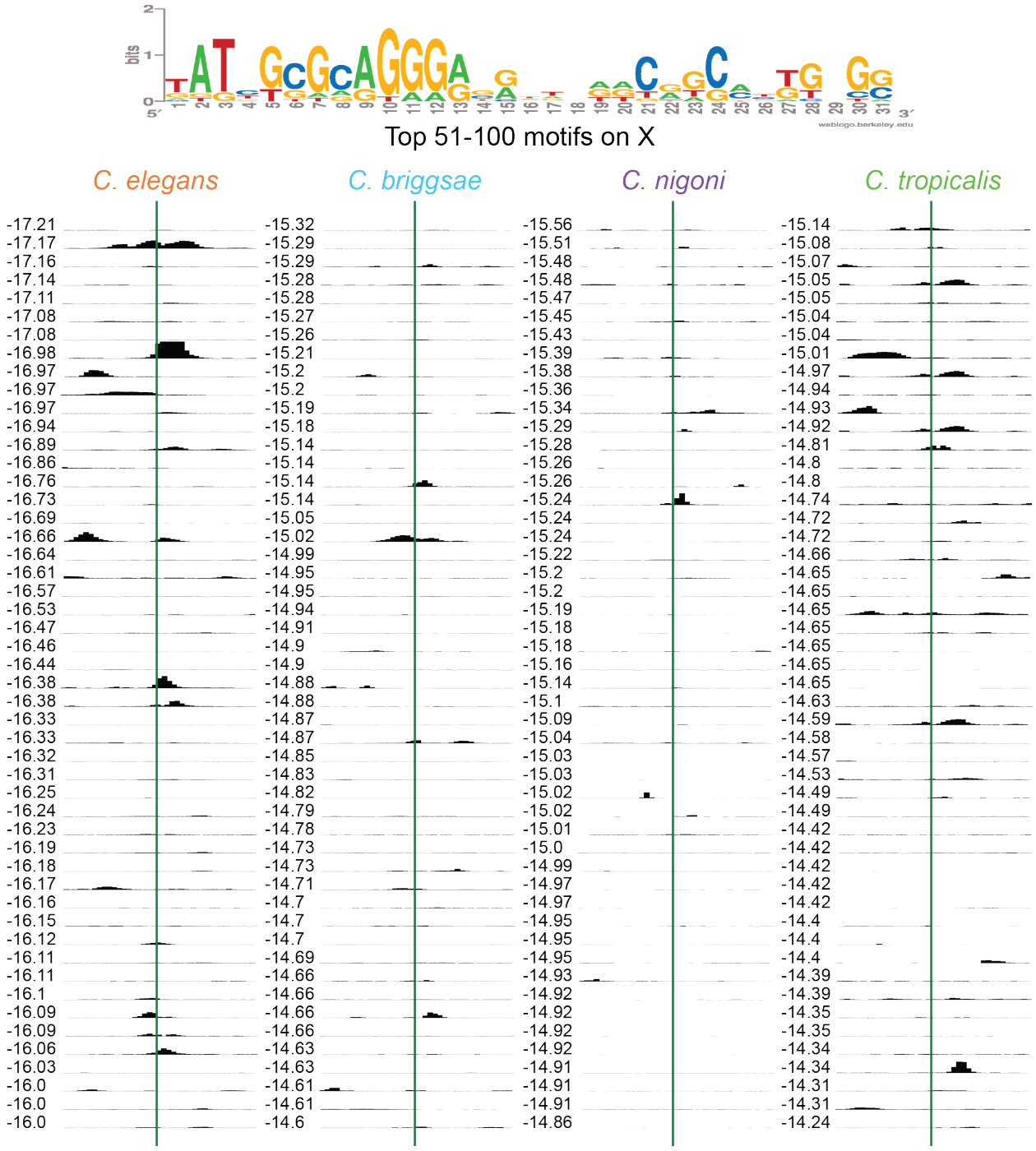
Figure G.3: Cel-bMEX-31bp a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cel-bMEX-31bp motif is X-enriched and peak-enriched only in *C. elegans*.** The Cel-bMEX-31bp motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

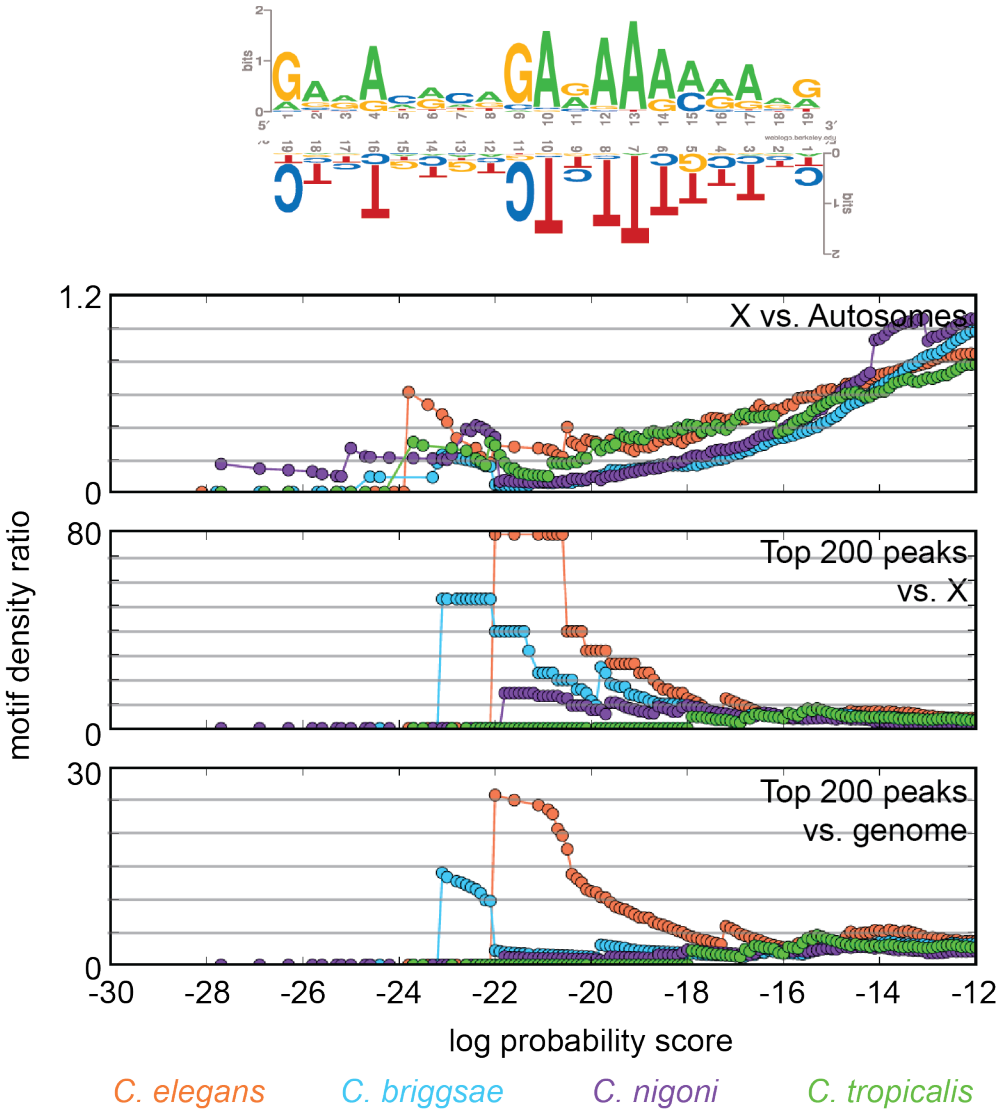


(b) **The *C. elegans* DCC is bound at many of the top 50 Cel-bMEX-31bp motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Cel-bMEX-31bp motifs on the X chromosome in four species.

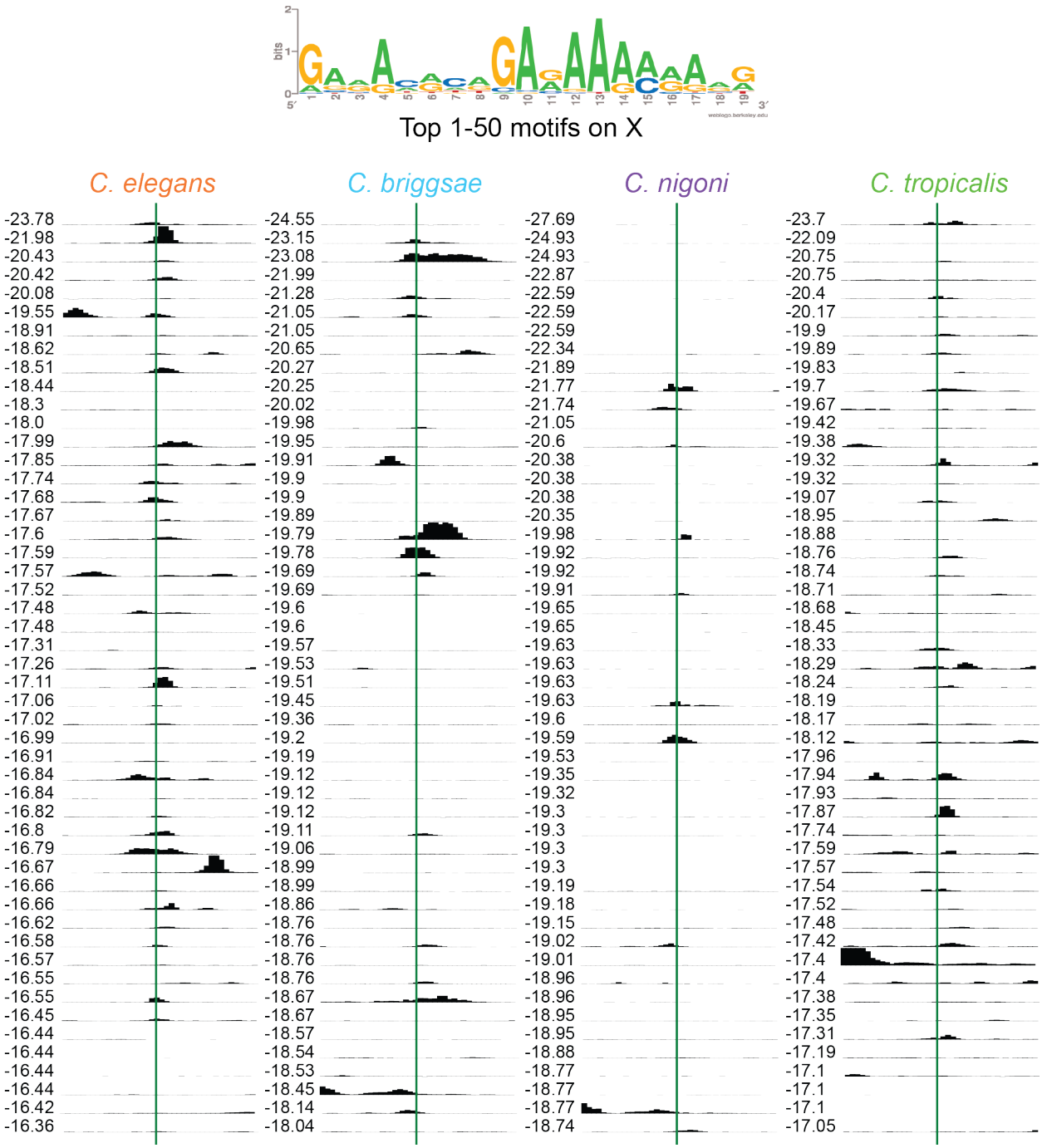


(c) **The *C. elegans* DCC is also bound at some of the top 51-100 Cel-bMEX-31bp motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cel-bMEX-31bp motifs on the X chromosome in four species.

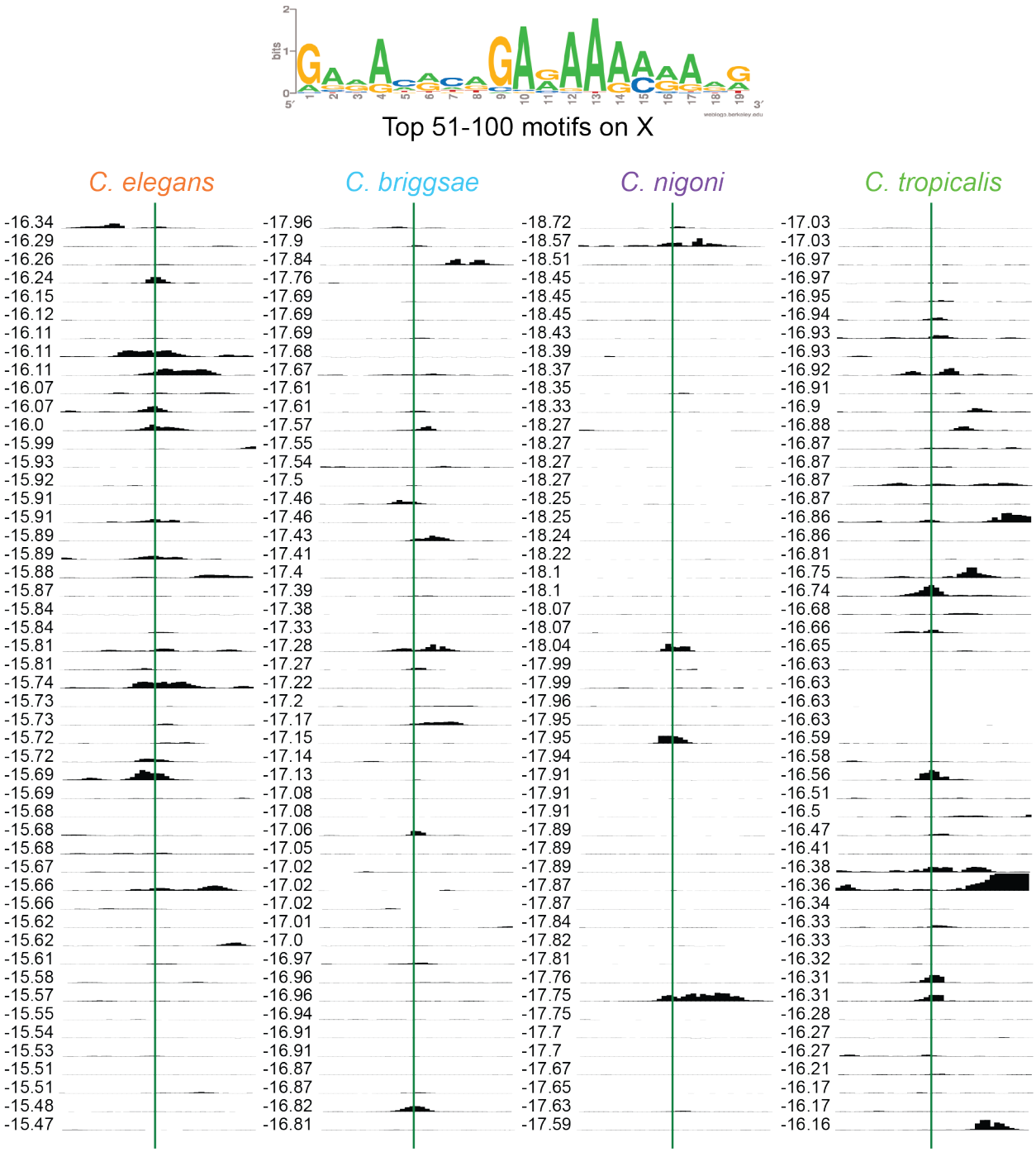
Figure G.4: Cbr20-Cni14-MEME2 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cbr20-Cni14-MEME2 motif is not X-enriched, but is found in peaks in each species.** The Cbr20-Cni14-MEME2 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

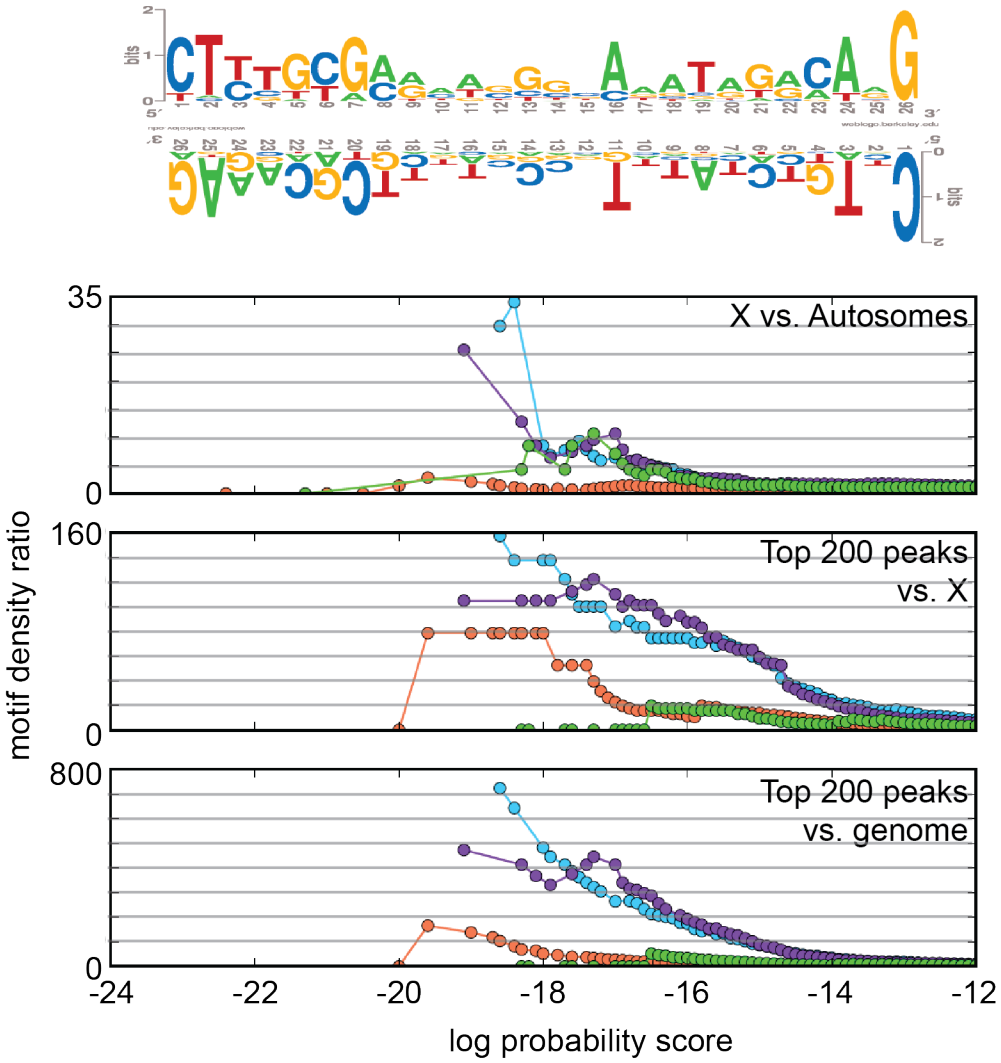


(b) **The DCC is bound at many of the top 50 Cbr20-Cni14-MEME2 motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 50 Cbr20-Cni14-MEME2 motifs on the X chromosome in four species.

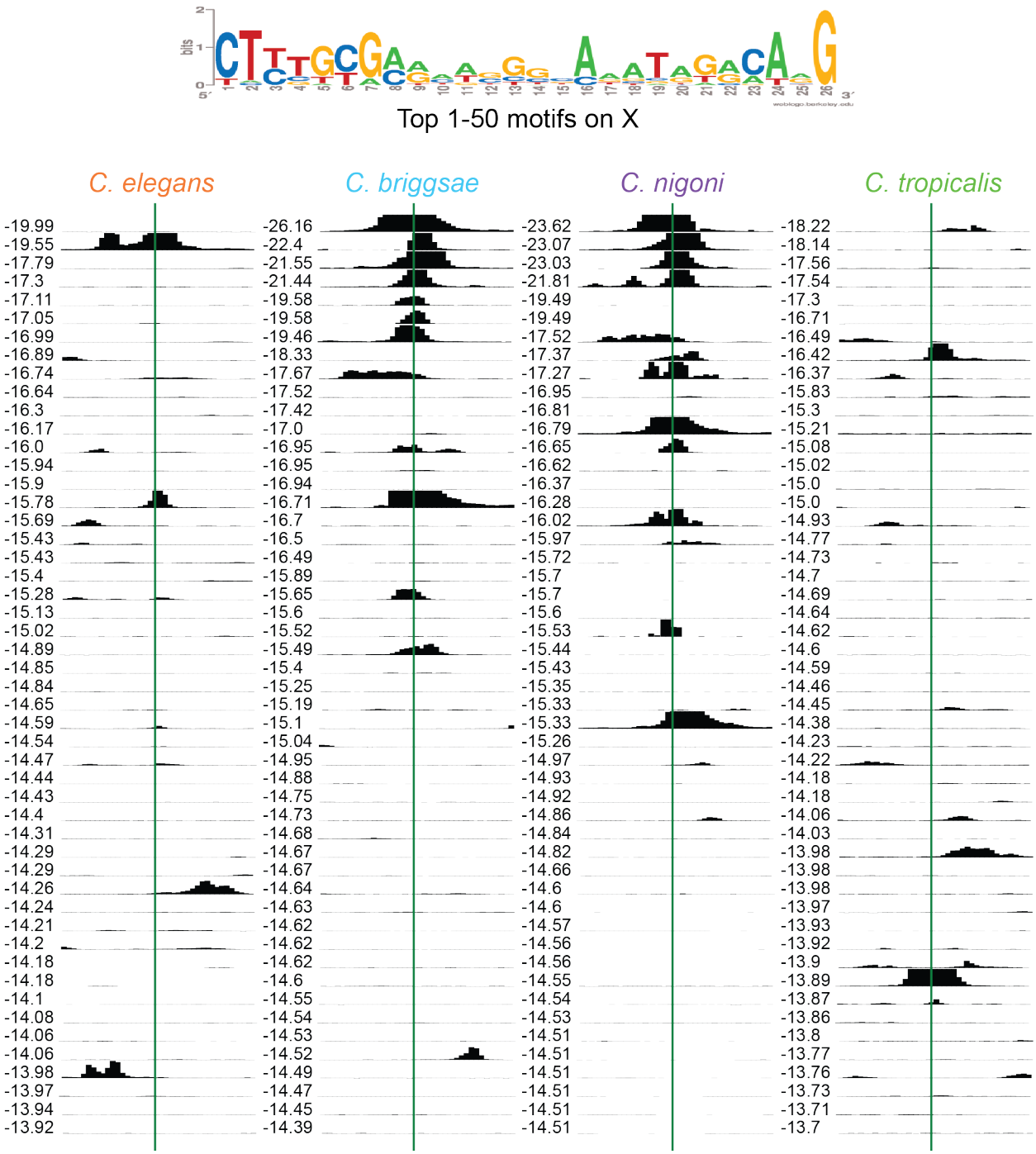


(c) **The DCC is also bound at many of the top 51-100 Cbr20-Cni14-MEME2 motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 51-100 Cbr20-Cni14-MEME2 motifs on the X chromosome in four species.

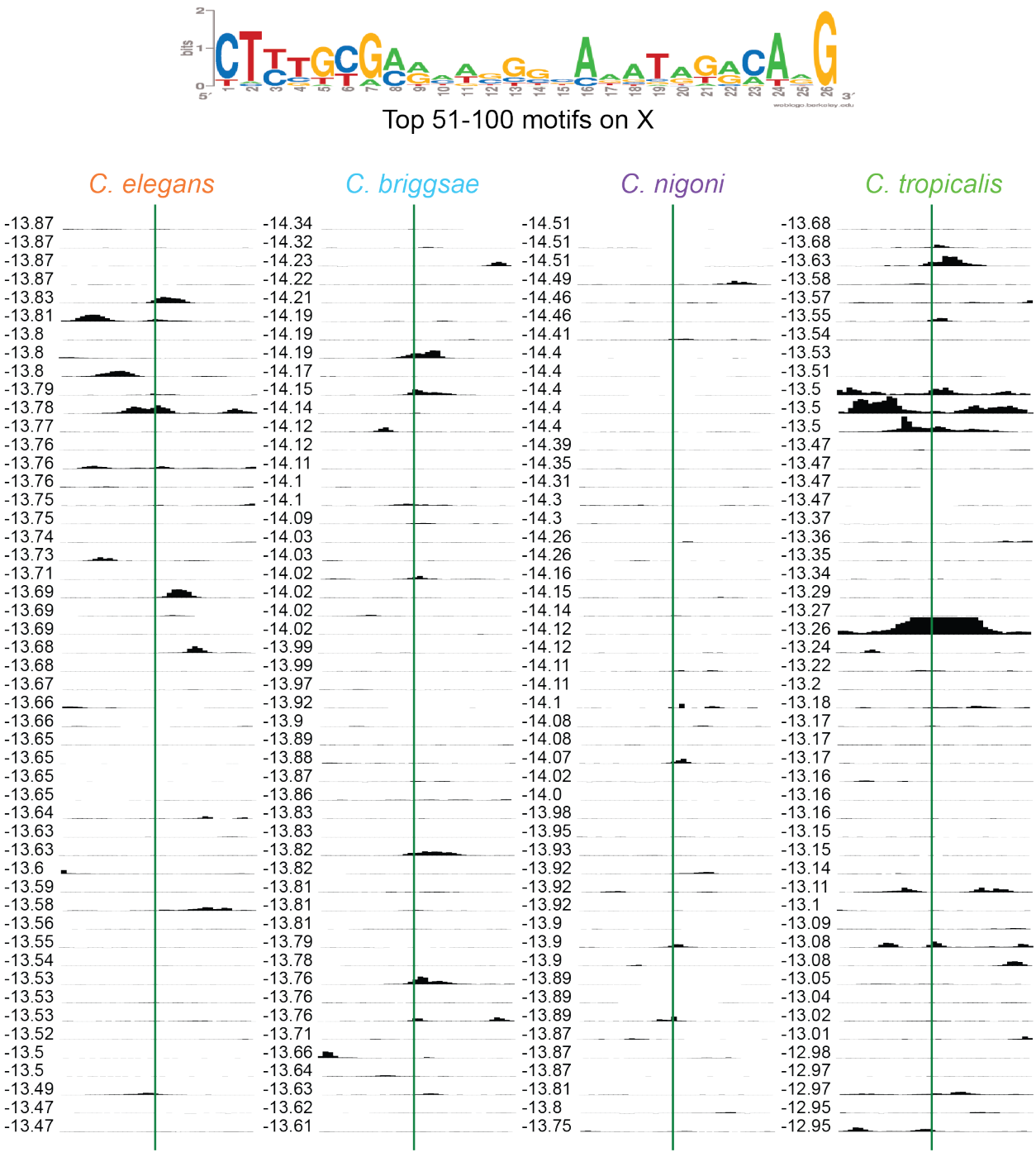
Figure G.5: Cbr20-Cni14-MEME1 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cbr20-Cni14-MEME1 motif is X-enriched and peak-enriched in *C. briggsae* and *C. nigoni*.** The Cbr20-Cni14-MEME1 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

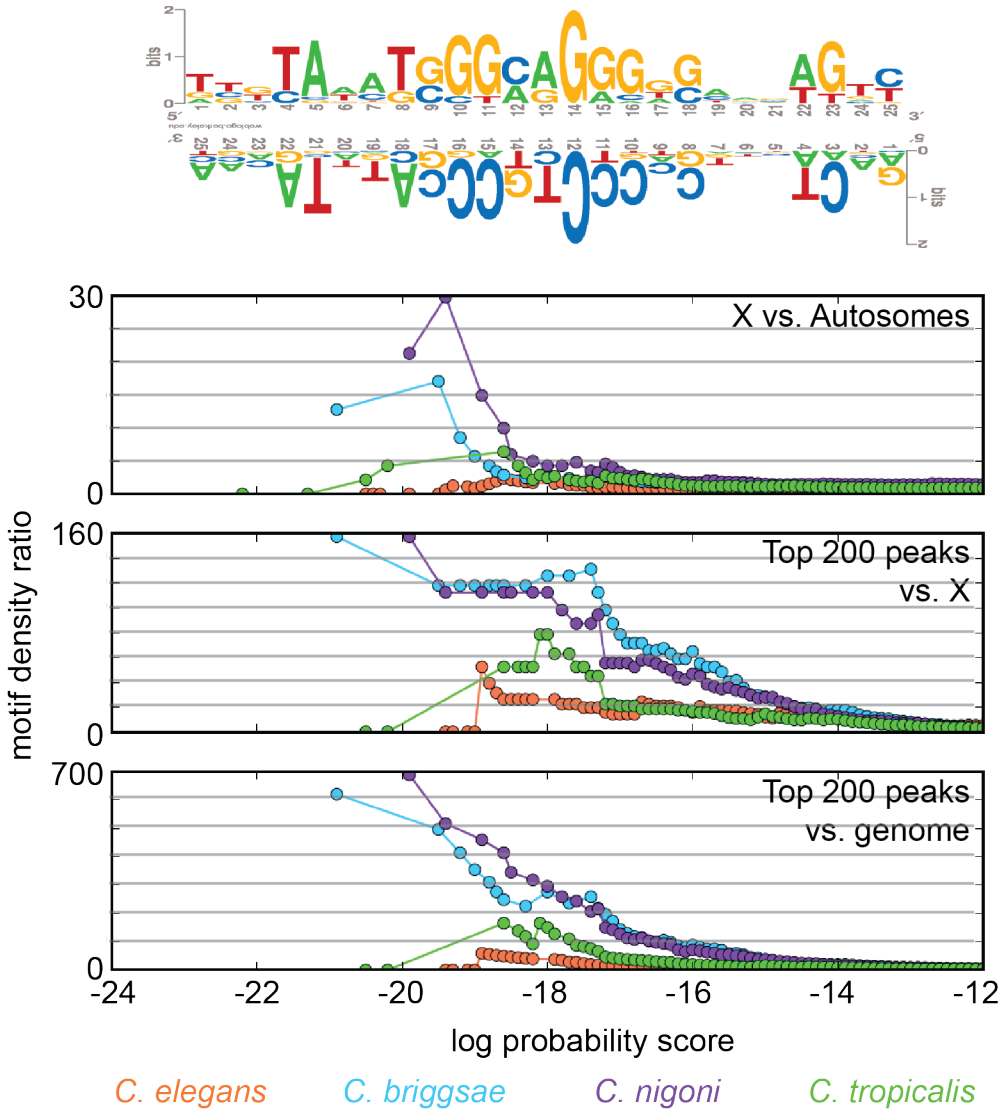


(b) **The DCC is bound at the strongest of the top 50 Cbr20-Cni14-MEME1 motifs on the X chromosome in *C. briggsae* and *C. nigoni*.** The ChIP-seq signal is plotted at the top 50 Cbr20-Cni14-MEME1 motifs on the X chromosome in four species.

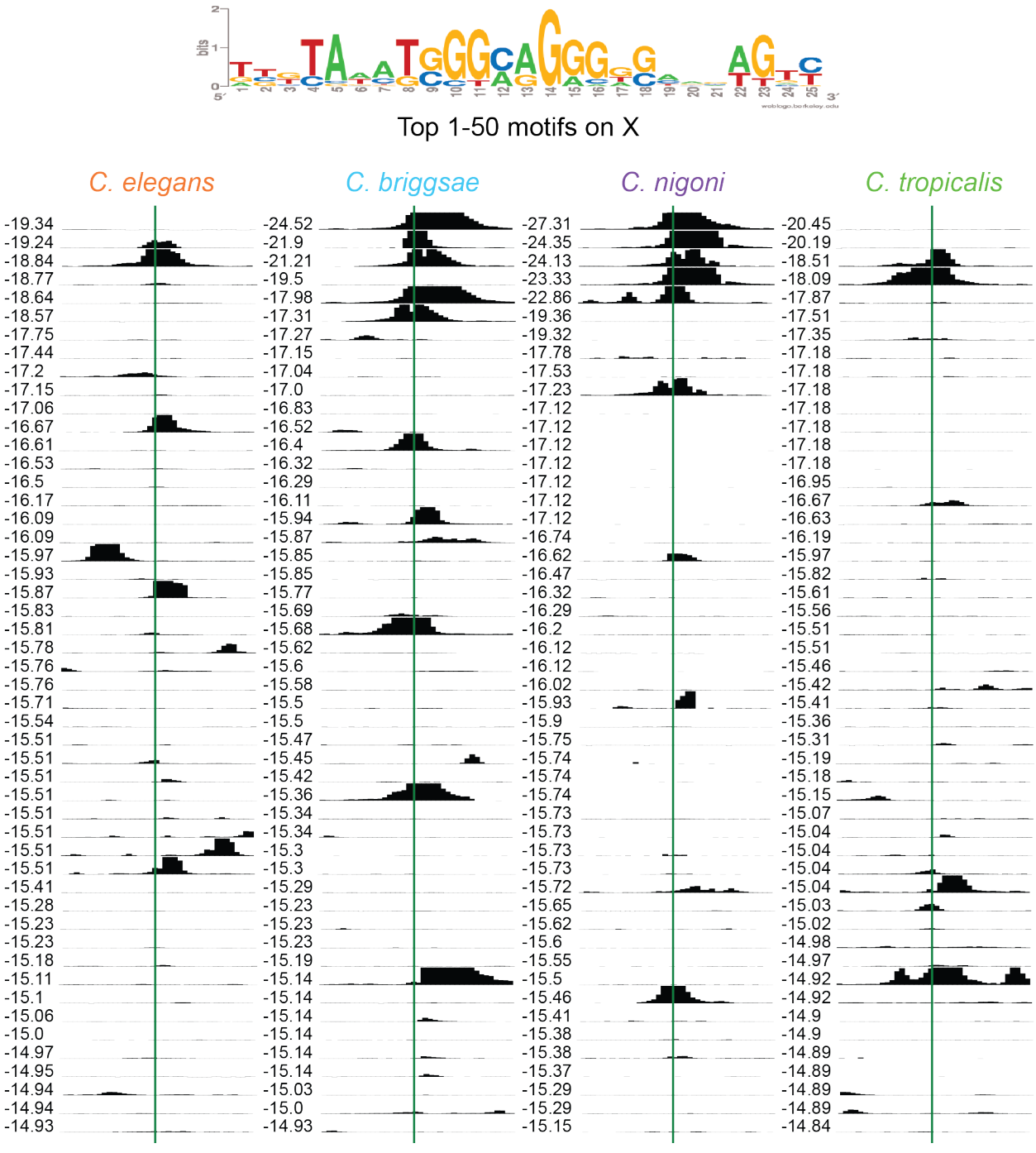


(c) **The DCC is bound at few of the top 51-100 Cbr20-Cni14-MEME1 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cbr20-Cni14-MEME1 motifs on the X chromosome in four species.

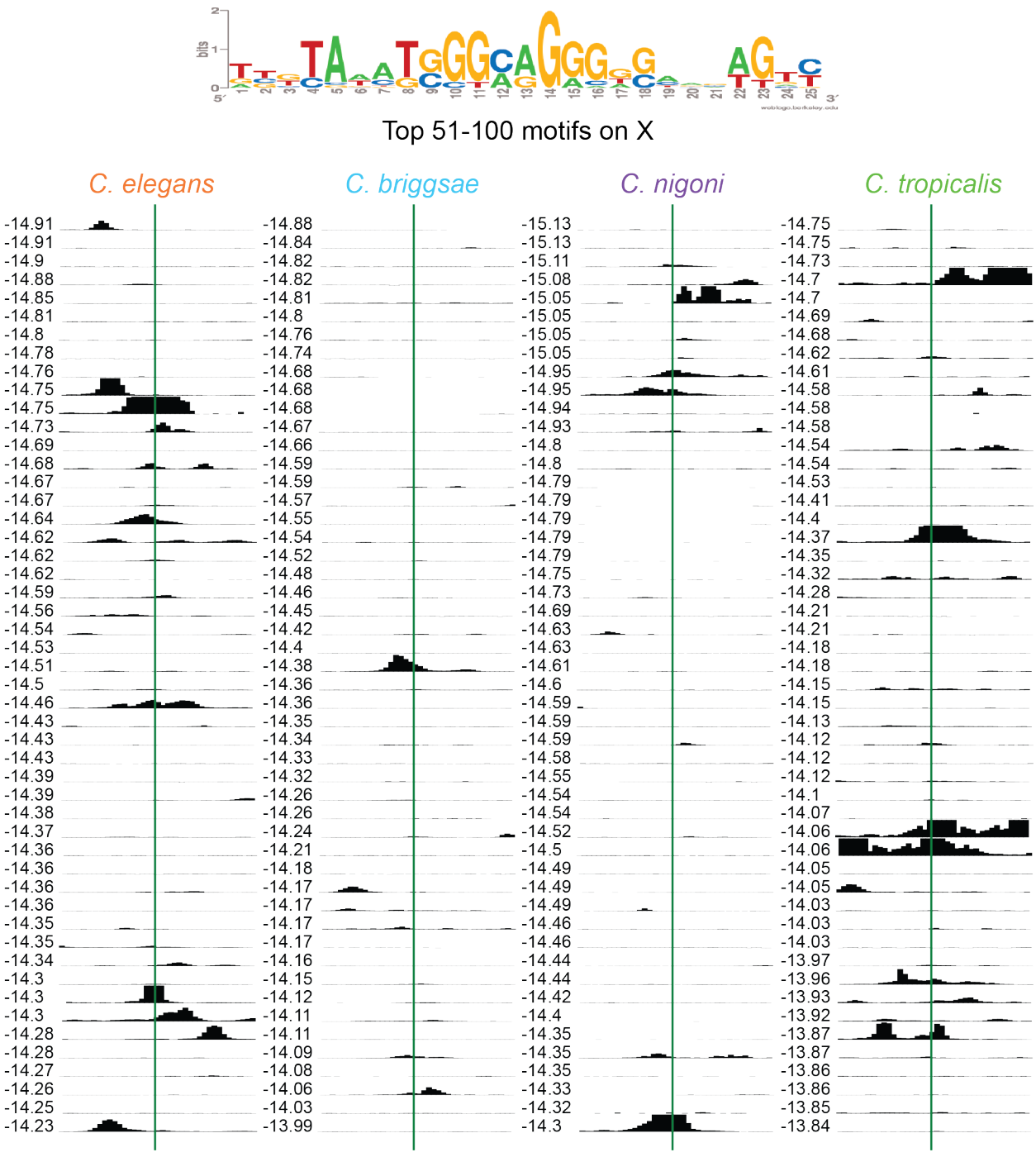
Figure G.6: Cbr20-Cni14-MEME3 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cbr20-Cni14-MEME3 motif is X-enriched and peak-enriched in *C. briggsae* and *C. nigoni*.** The Cbr20-Cni14-MEME3 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

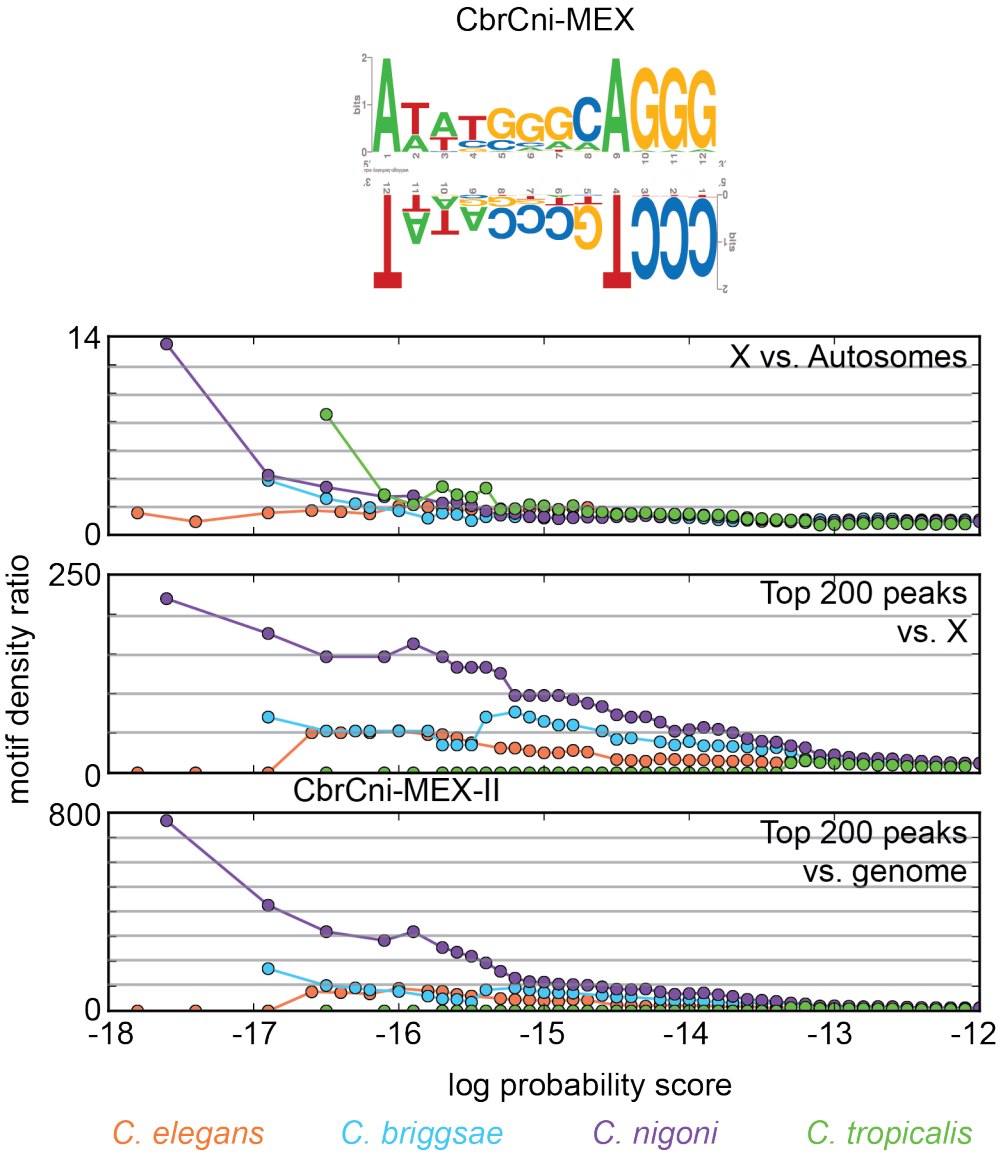


(b) **The DCC is bound at the strongest of the top 50 Cbr20-Cni14-MEME3 motifs on the X chromosome in *C. briggsae* and *C. nigoni*.** The ChIP-seq signal is plotted at the top 50 Cbr20-Cni14-MEME3 motifs on the X chromosome in four species.

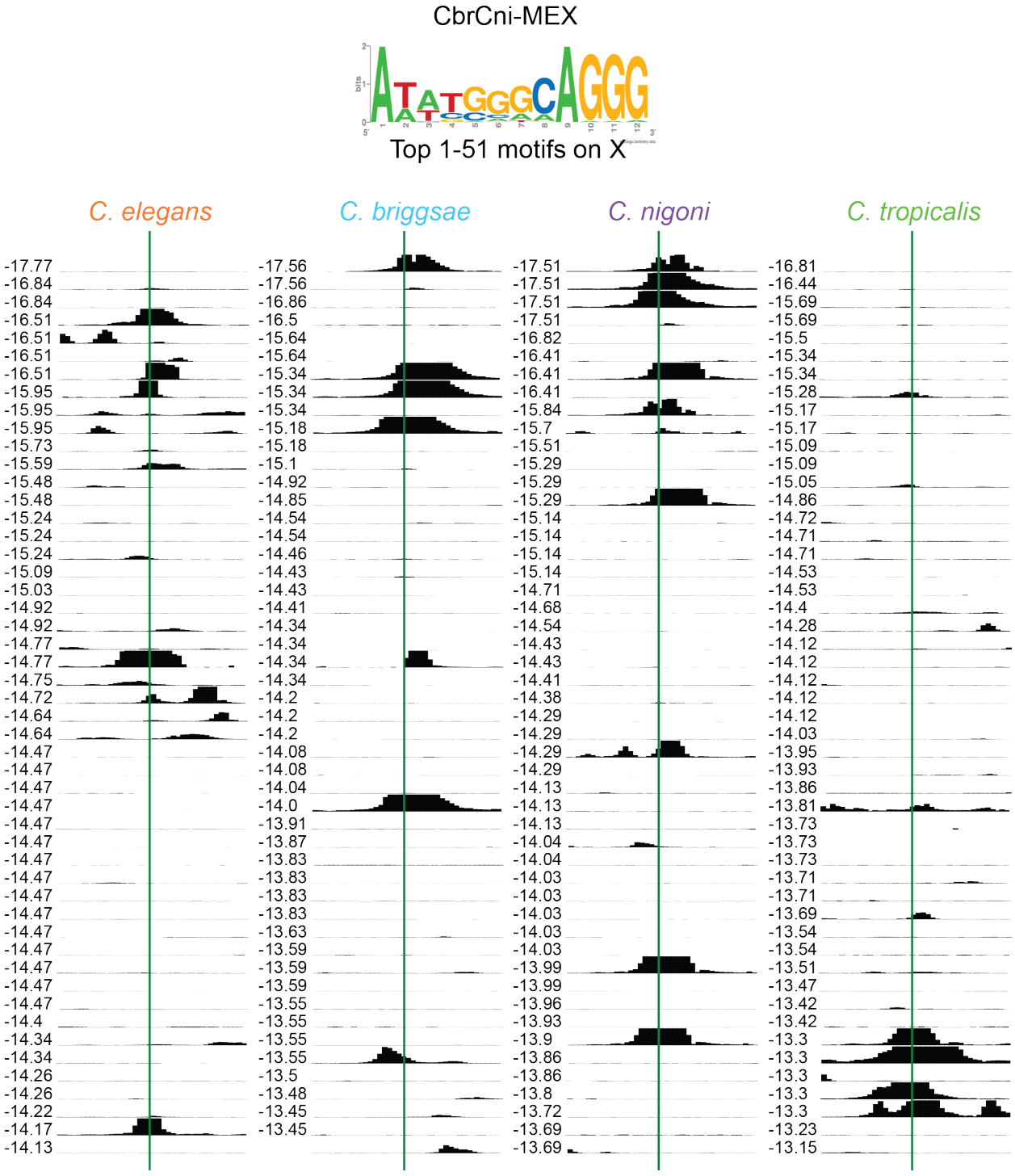


(c) **The DCC is also bound at some of the top 51-100 Cbr20-Cni14-MEME3 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cbr20-Cni14-MEME3 motifs on the X chromosome in four species.

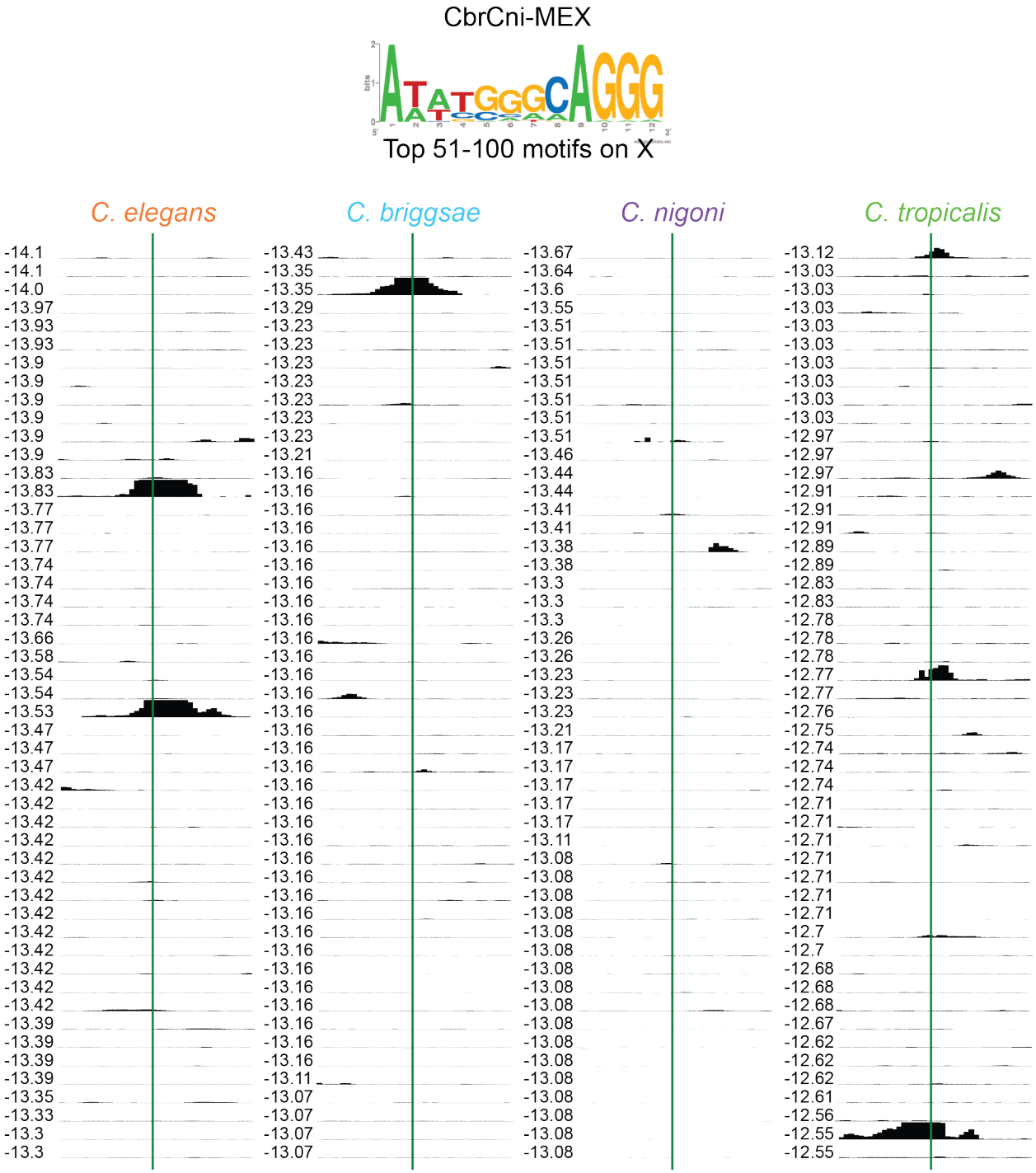
Figure G.7: CbrCni-MEX a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The CbrCni-MEX motif is X-enriched in *C. nigoni* and *C. tropicalis* and peak-enriched in *C. nigoni* and *C. briggsae*.** The CbrCni-MEX motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

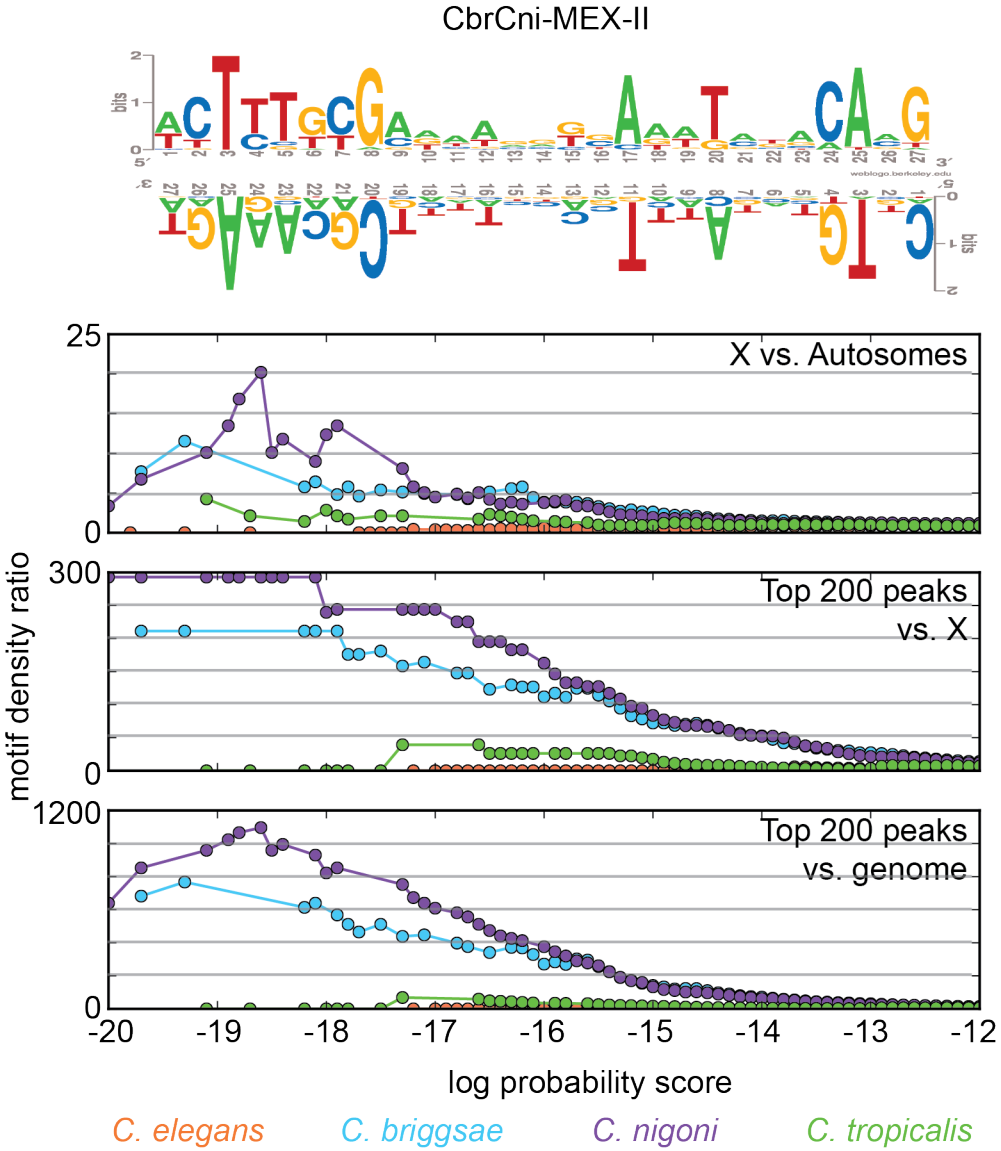


(b) **The DCC is bound at many of the top 50 CbrCni-MEX motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 50 CbrCni-MEX motifs on the X chromosome in four species.

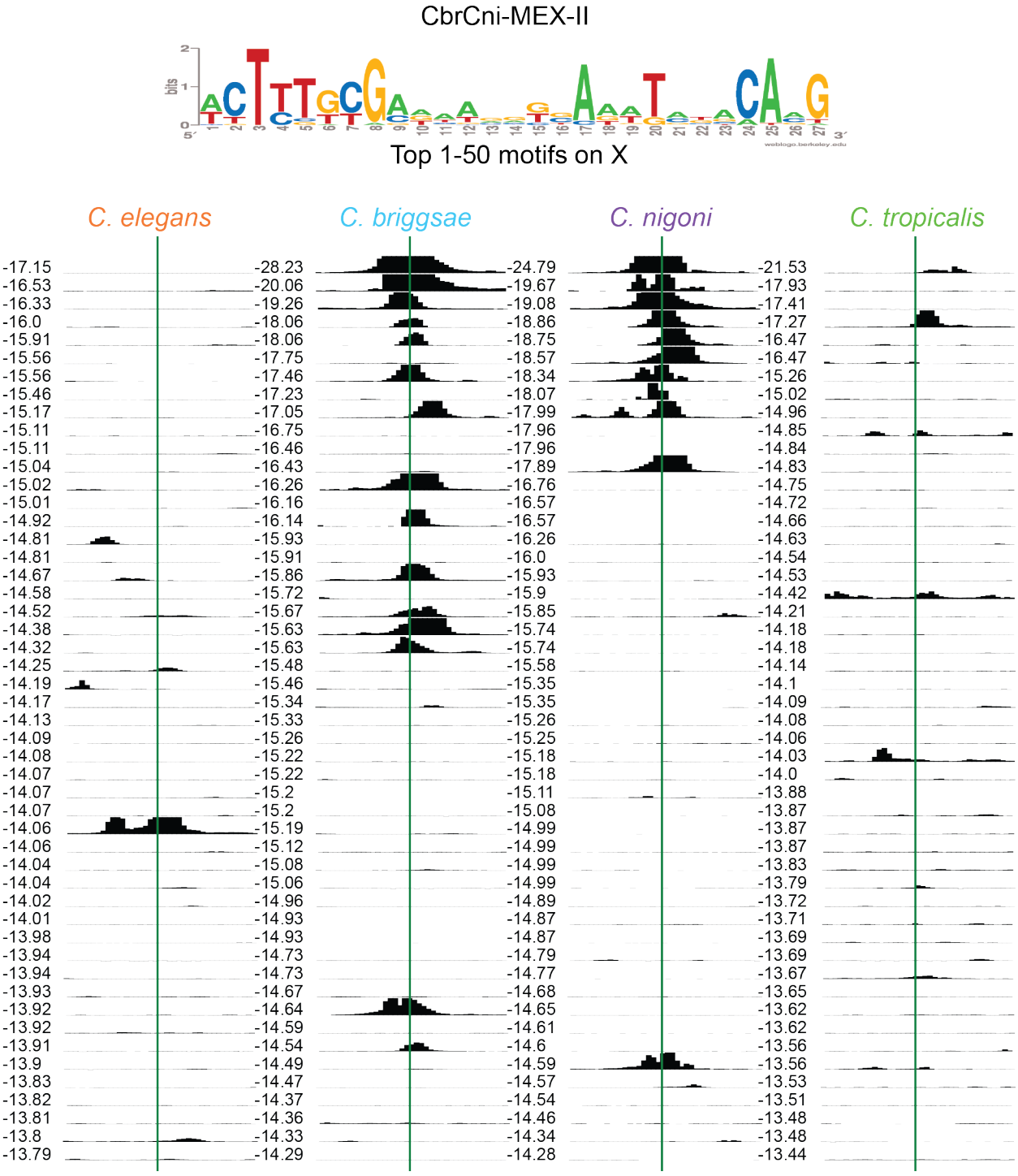


(c) **The DCC is bound at few of the top 51-100 CbrCni-MEX motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 CbrCni-MEX motifs on the X chromosome in four species.

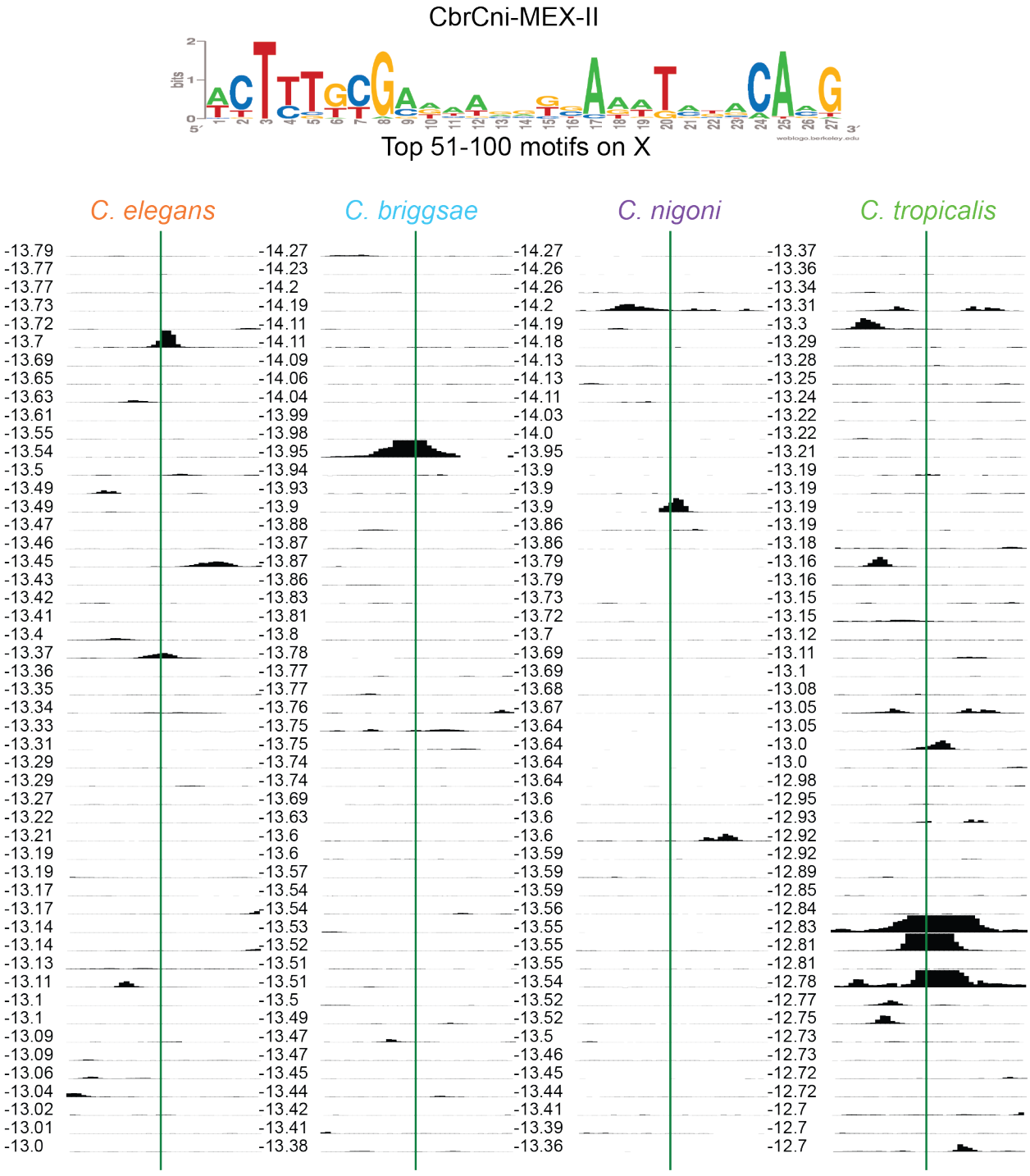
Figure G.8: CbrCni-MEX-II a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The CbrCni-MEX-II motif is X-enriched and peak-enriched in *C. nigoni* and *C. briggsae*.** The CbrCni-MEX-II motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

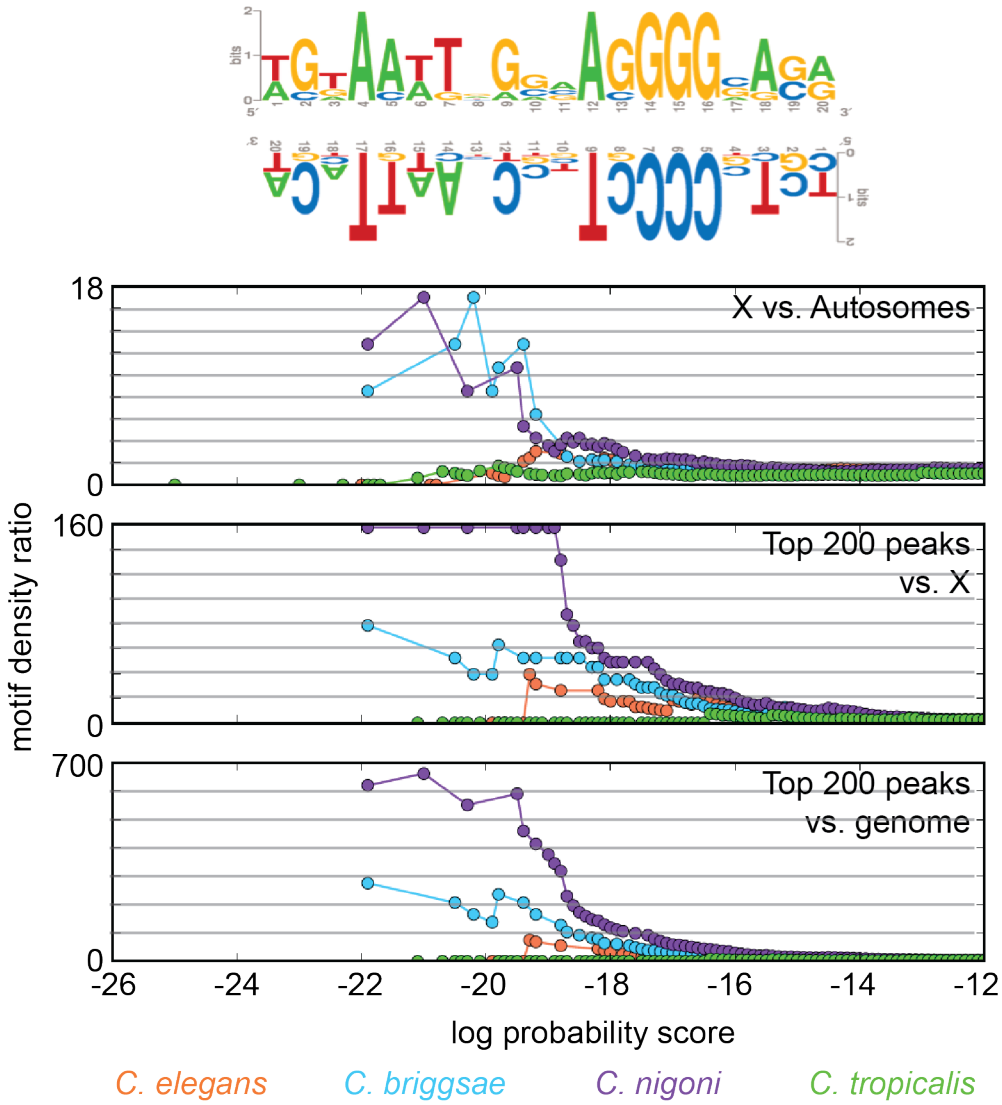


(b) **The DCC is bound at the strongest of the top 50 CbrCni-MEX-II motifs on the X chromosome in *C. briggsae* and *C. nigoni*.** The ChIP-seq signal is plotted at the top 50 CbrCni-MEX-II motifs on the X chromosome in four species.

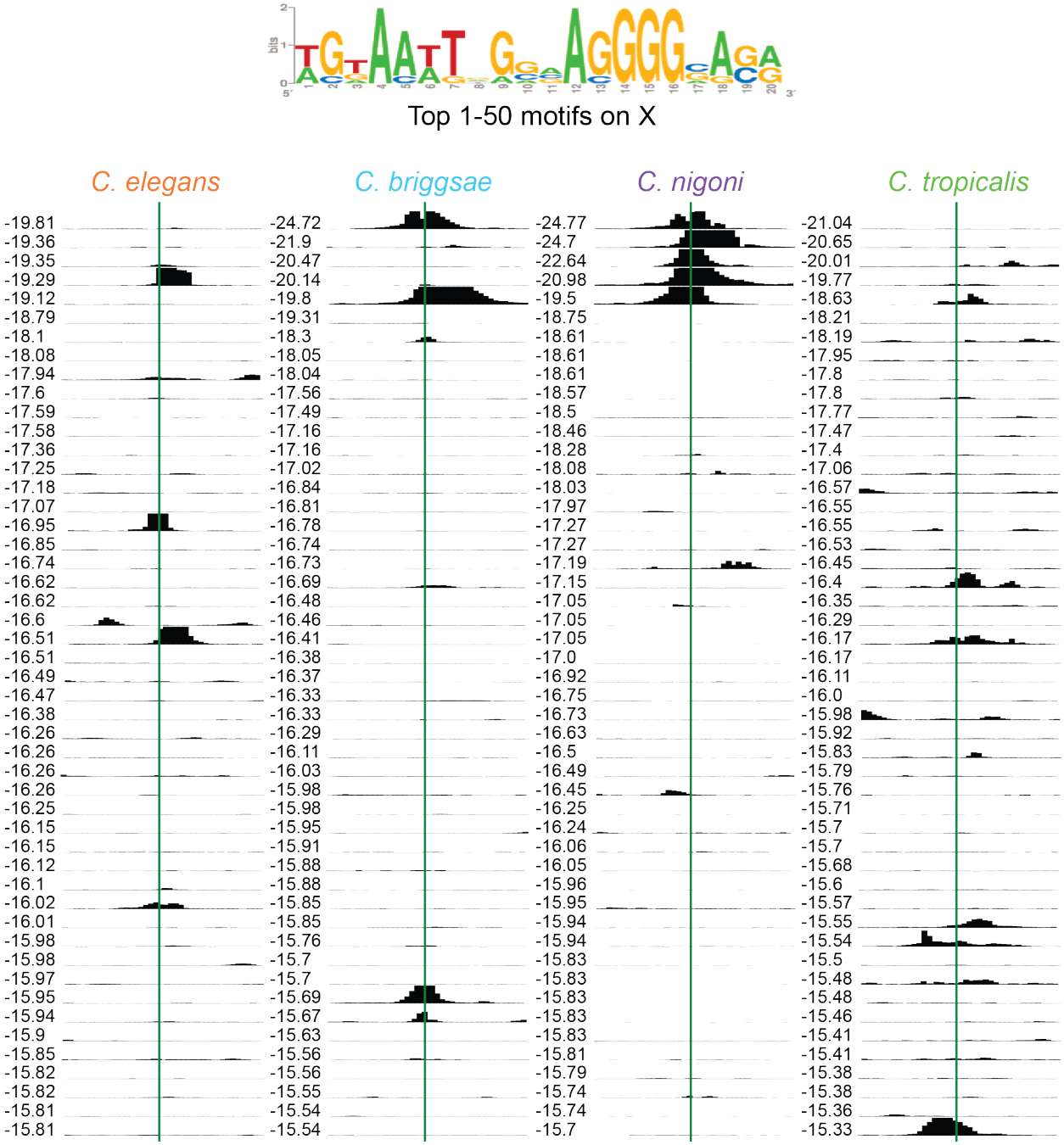


(c) **The DCC is bound at few of the top 51-100 CbrCni-MEX-II motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 CbrCni-MEX-II motifs on the X chromosome in four species.

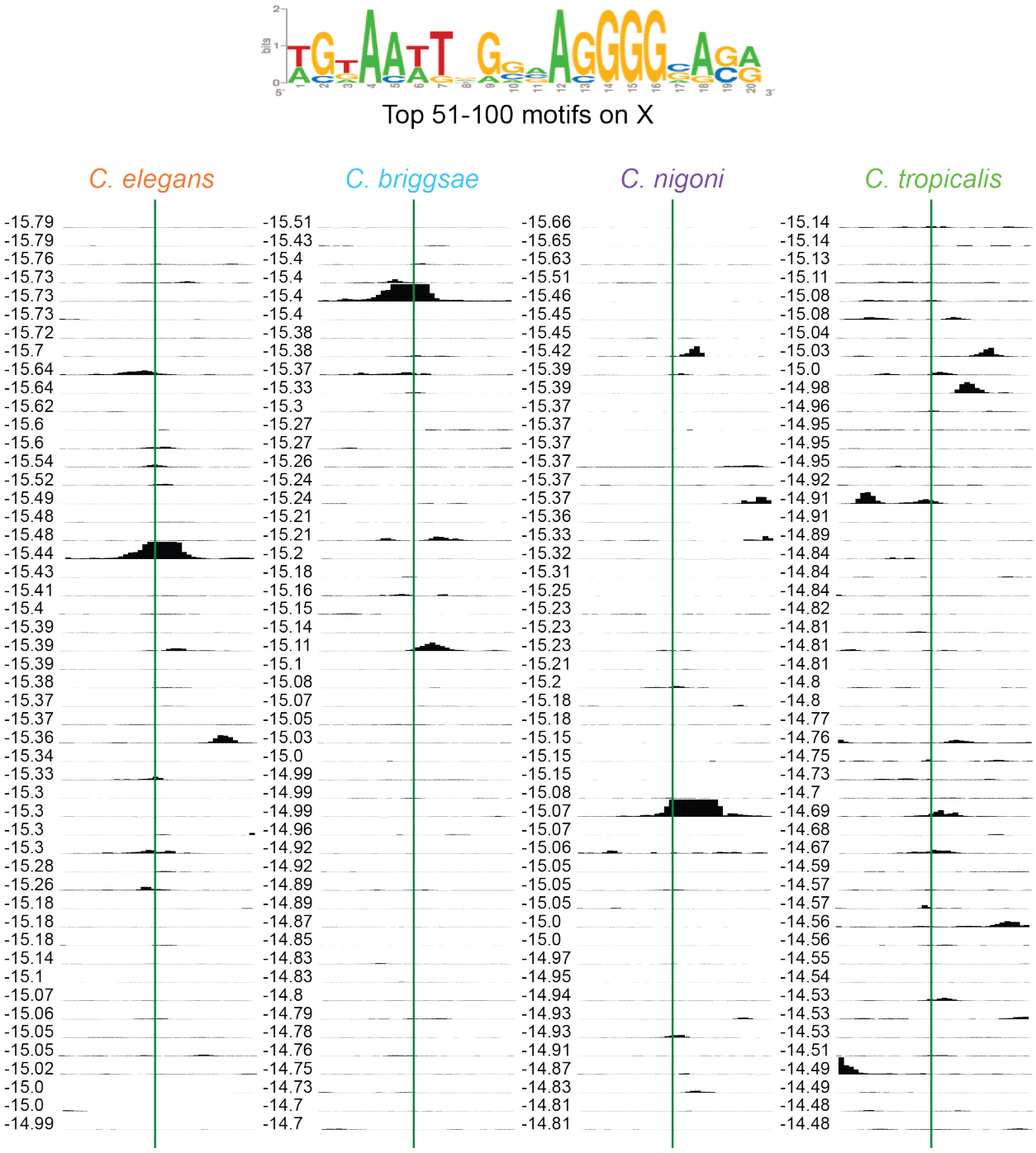
Figure G.9: Cni-MEME1 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni-MEME1 motif is X-enriched and peak-enriched in *C. nigoni* and *C. briggsae*.** The Cni-MEME1 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

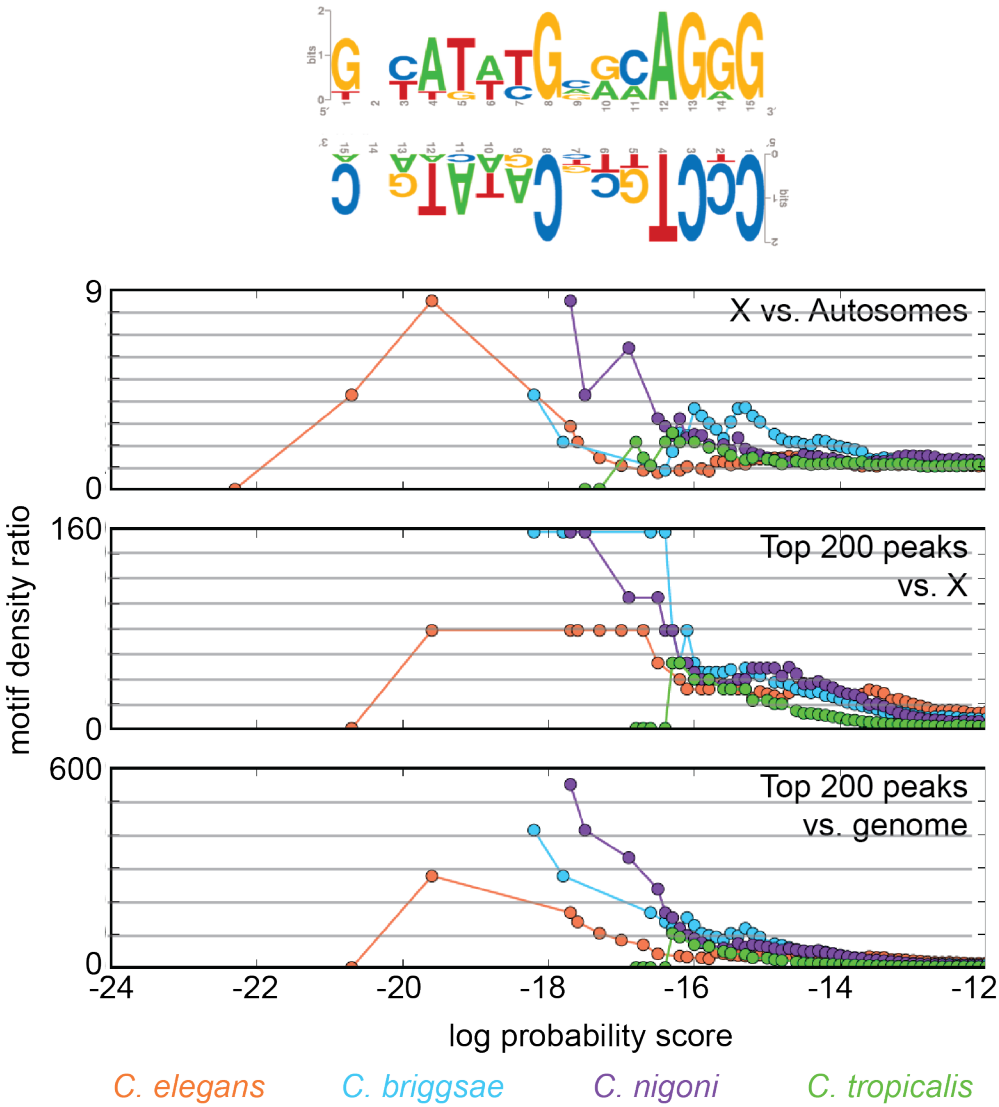


(b) **The *C. nigoni* DCC is bound at the five strongest Cni-MEME1 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Cni-MEME1 motifs on the X chromosome in four species.

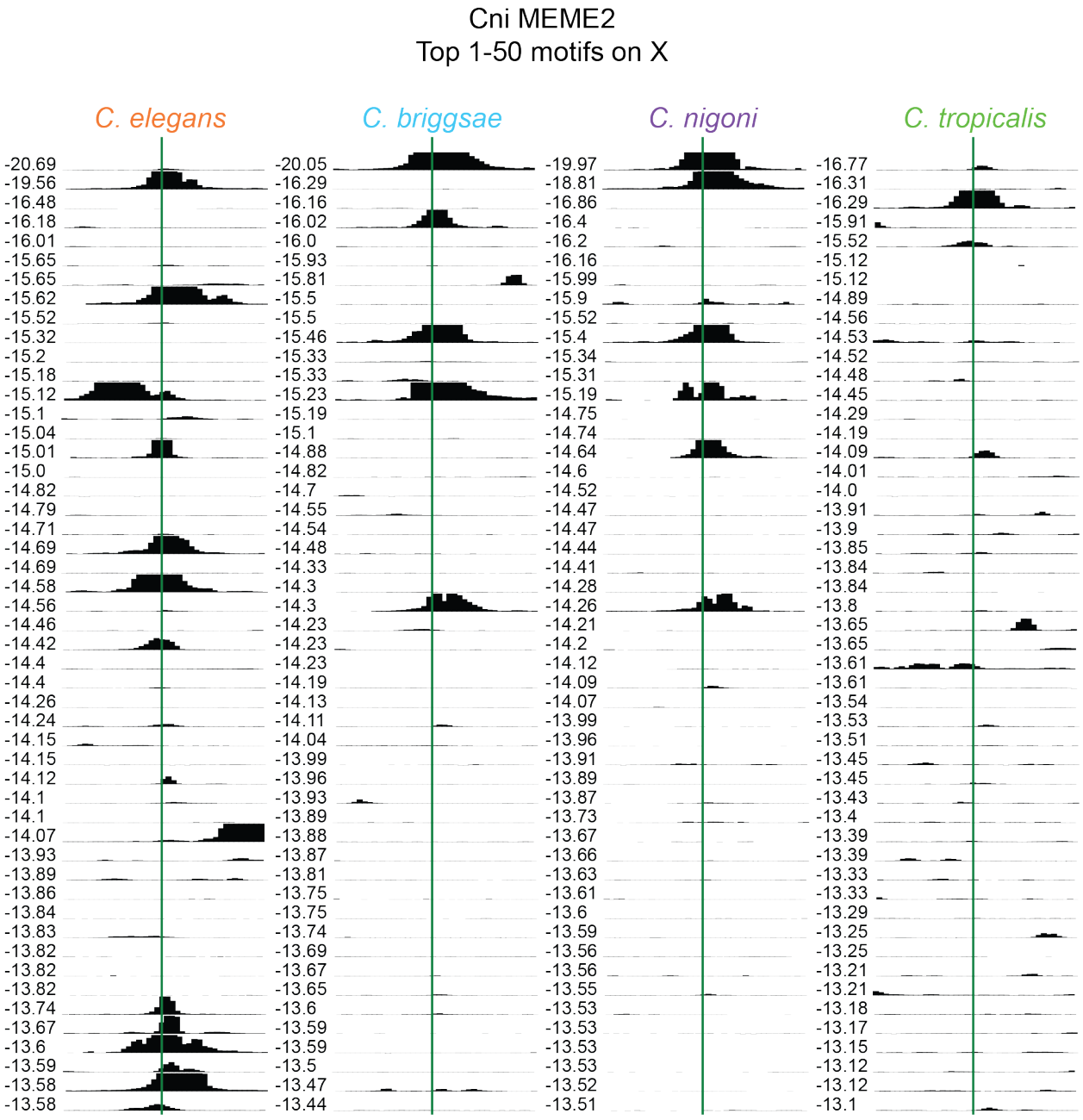


(c) **The DCC is bound at few of the top 51-100 Cni-MEME1 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni-MEME1 motifs on the X chromosome in four species.

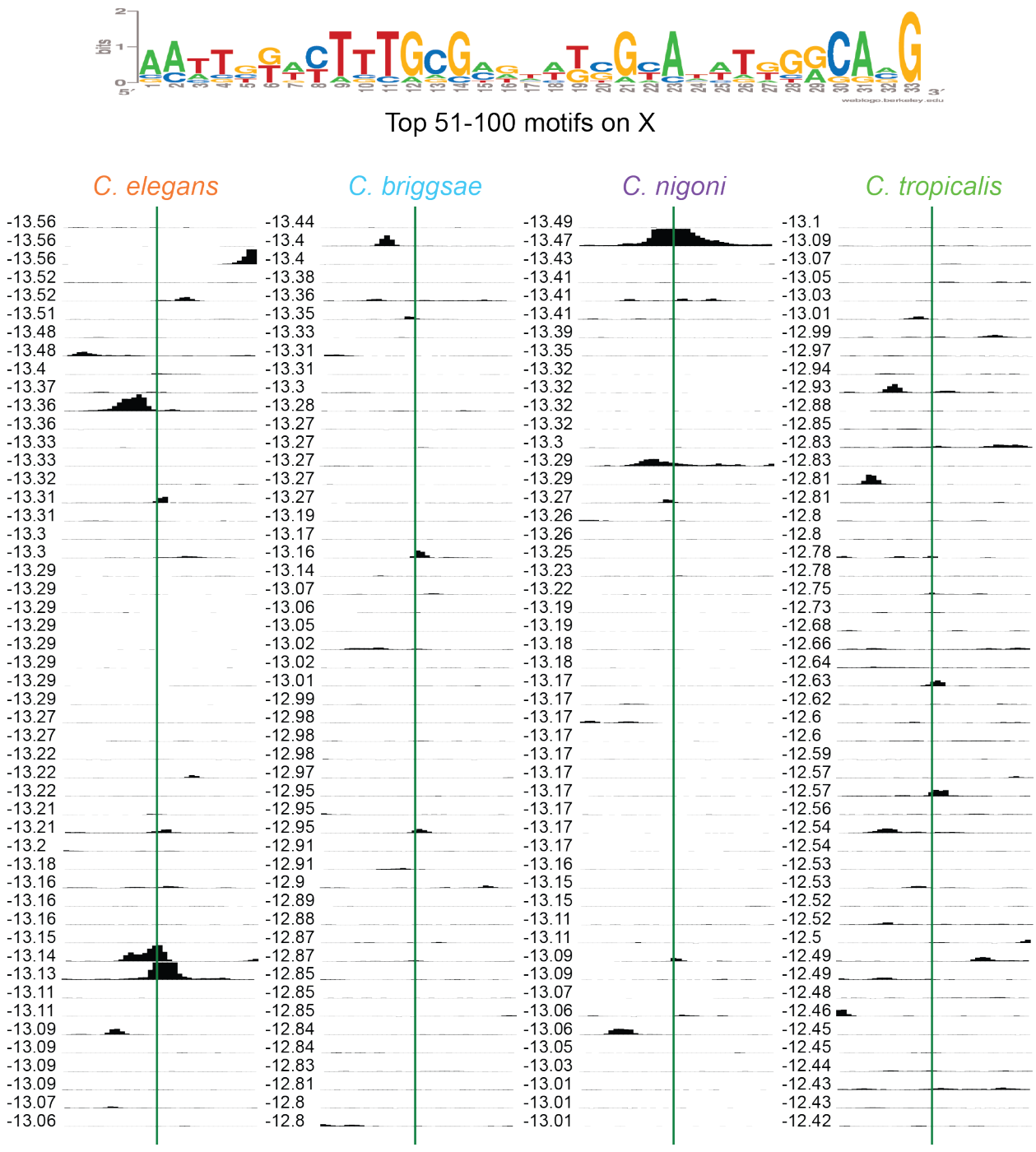
Figure G.10: Cni-MEME2 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni-MEME2 motif is X-enriched and peak-enriched in *C. nigoni* and *C. briggsae*.** The Cni-MEME2 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

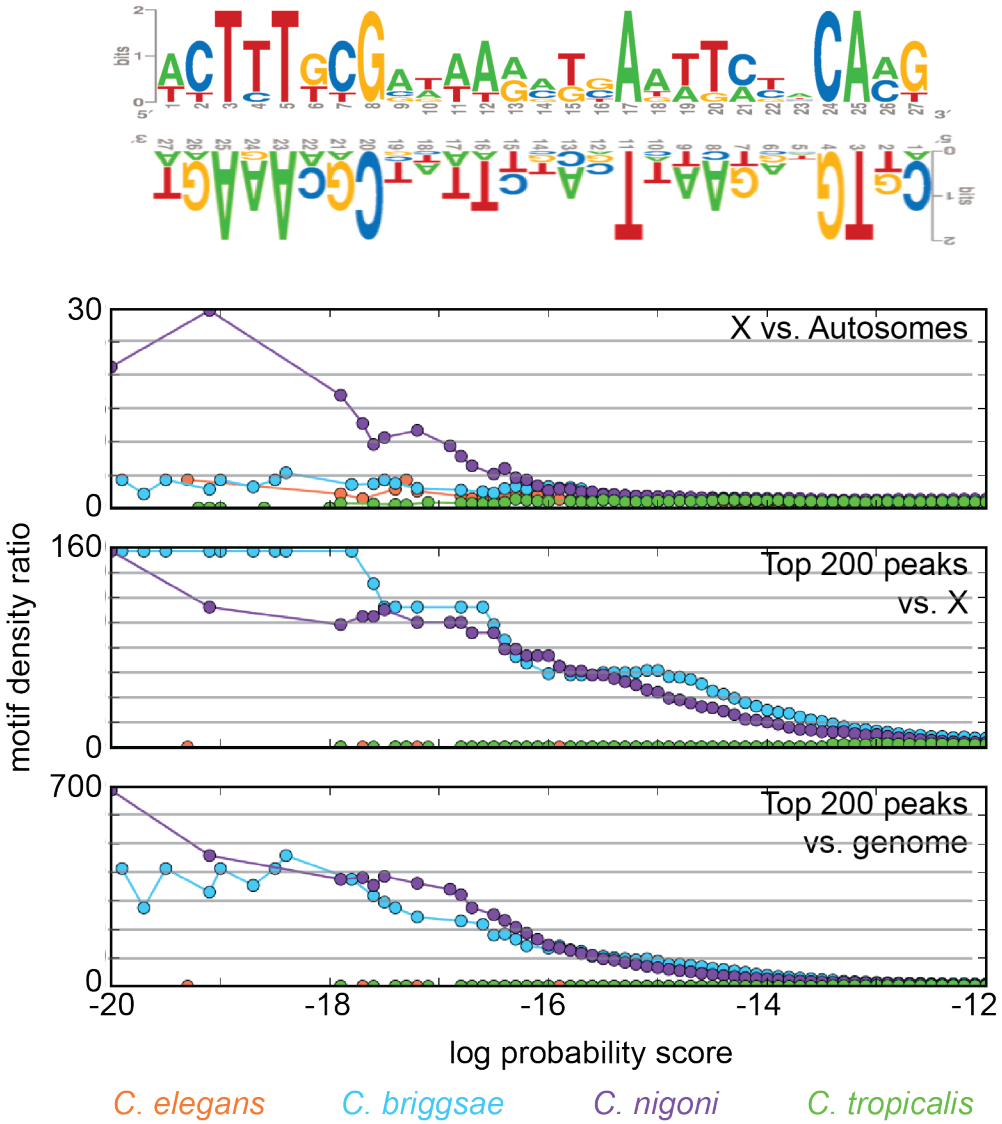


(b) **The DCC is bound at many of the top 50 Cni-MEME2 motifs on the X chromosome in *C. nigoni*, *C. briggsae*, and *C. elegans*.** The ChIP-seq signal is plotted at the top 50 Cni-MEME2 motifs on the X chromosome in four species.

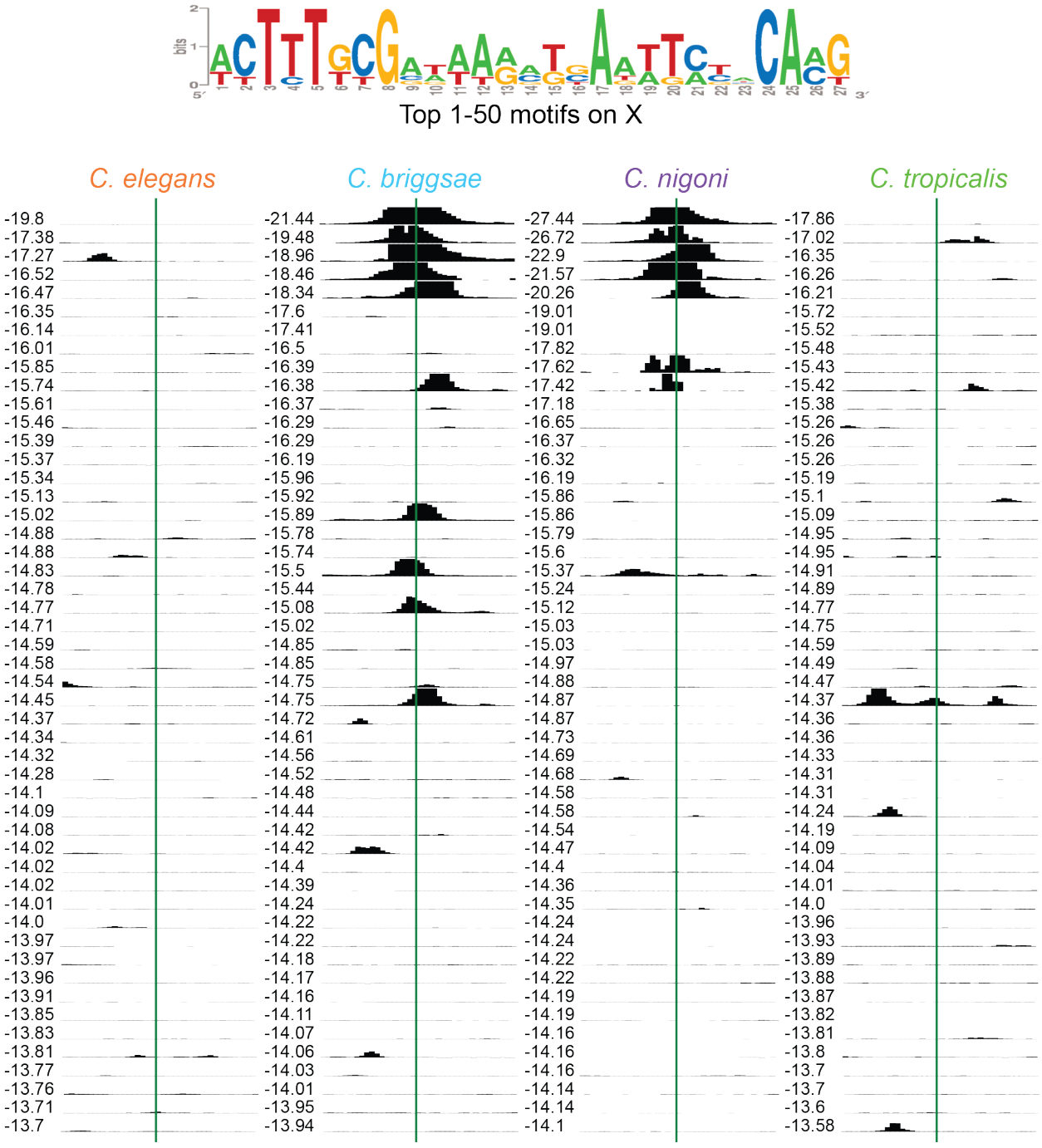


(c) **The DCC is bound at few of the top 51-100 Cni-MEME2 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni-MEME2 motifs on the X chromosome in four species.

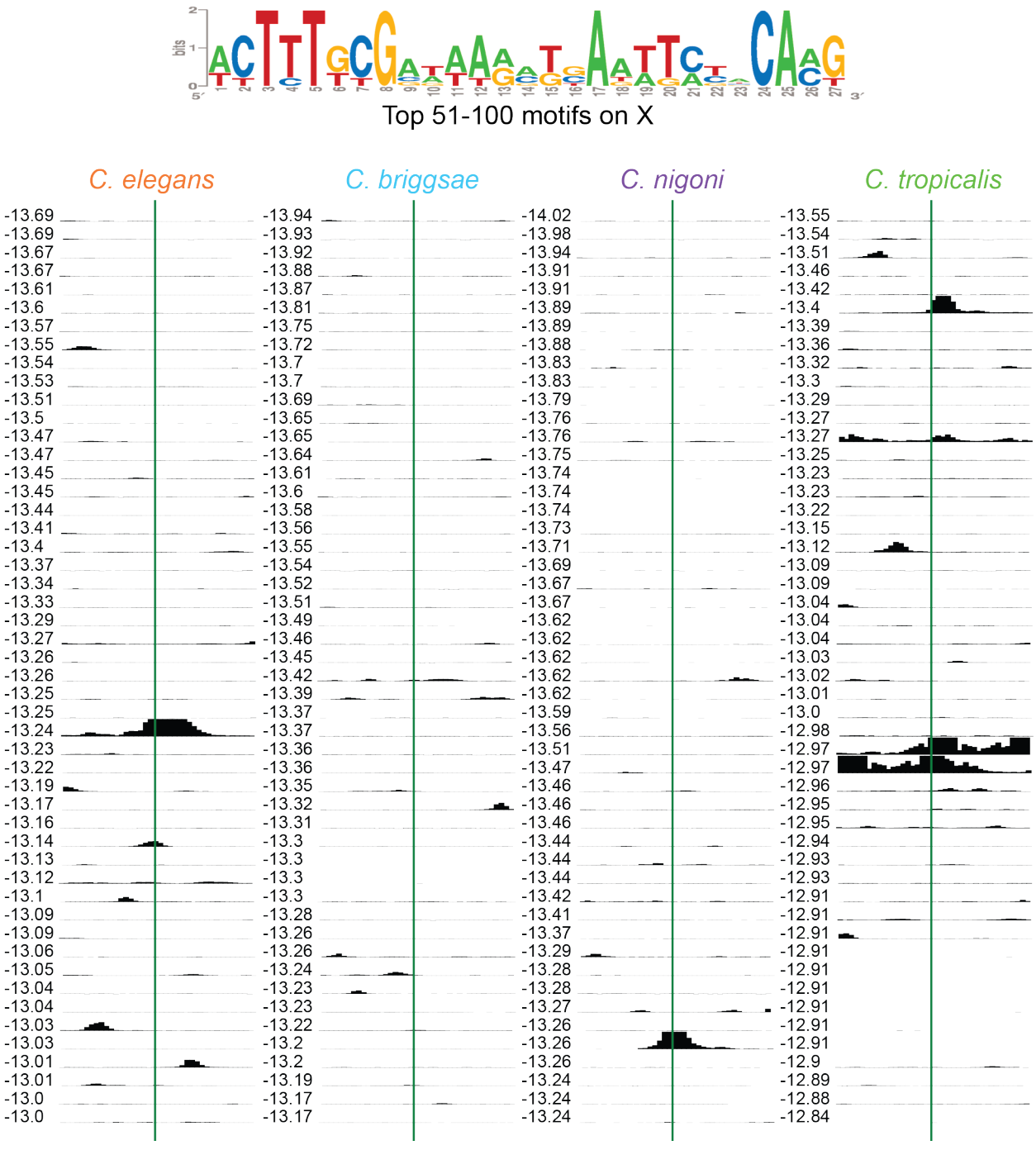
Figure G.11: Cni-MEME3 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni-MEME3 motif is X-enriched in *C. nigoni* and peak-enriched in *C. nigoni* and *C. briggsae*.** The Cni-MEME3 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

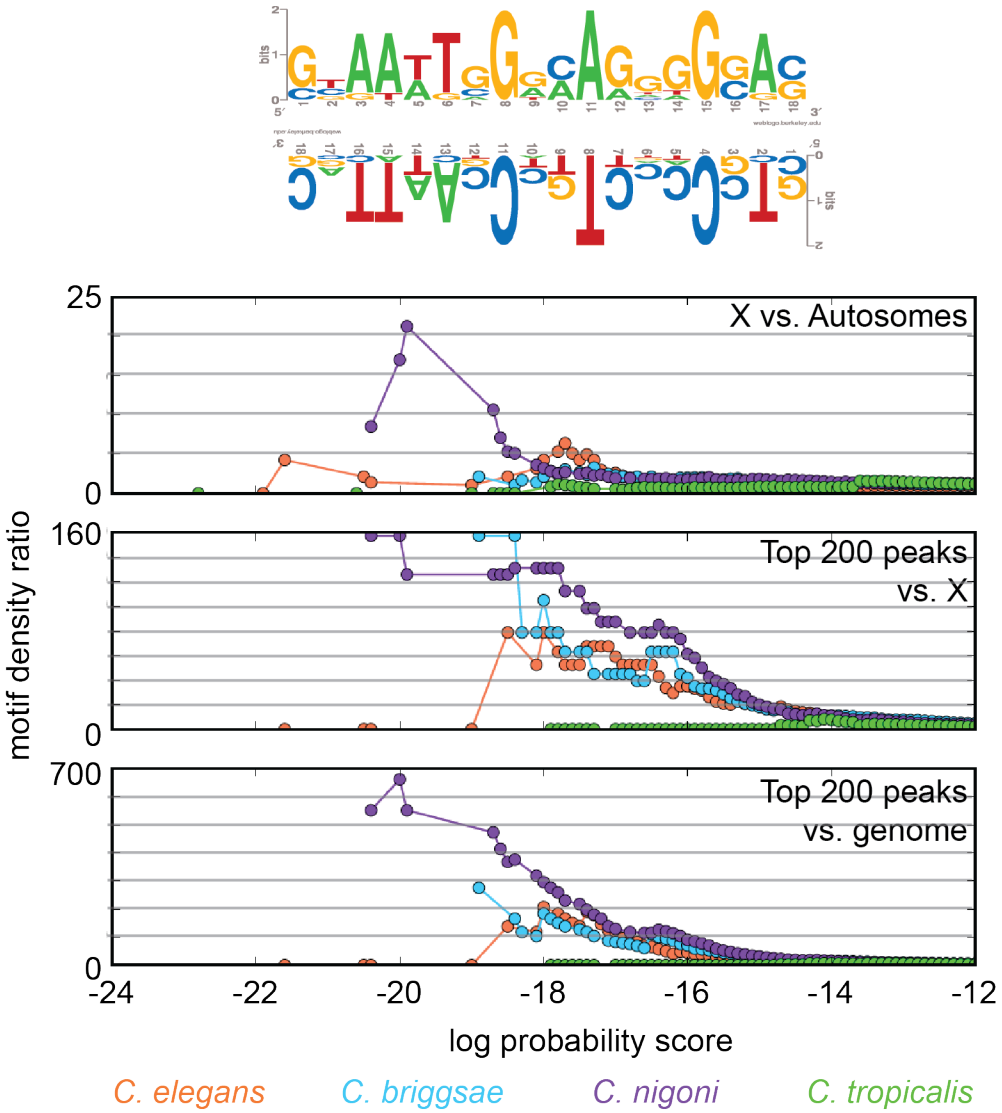


(b) **The DCC is bound at the strongest Cni-MEME3 motifs on the X chromosome in *C. briggsae* and *C. nigoni*.** The ChIP-seq signal is plotted at the top 50 Cni-MEME3 motifs on the X chromosome in four species.

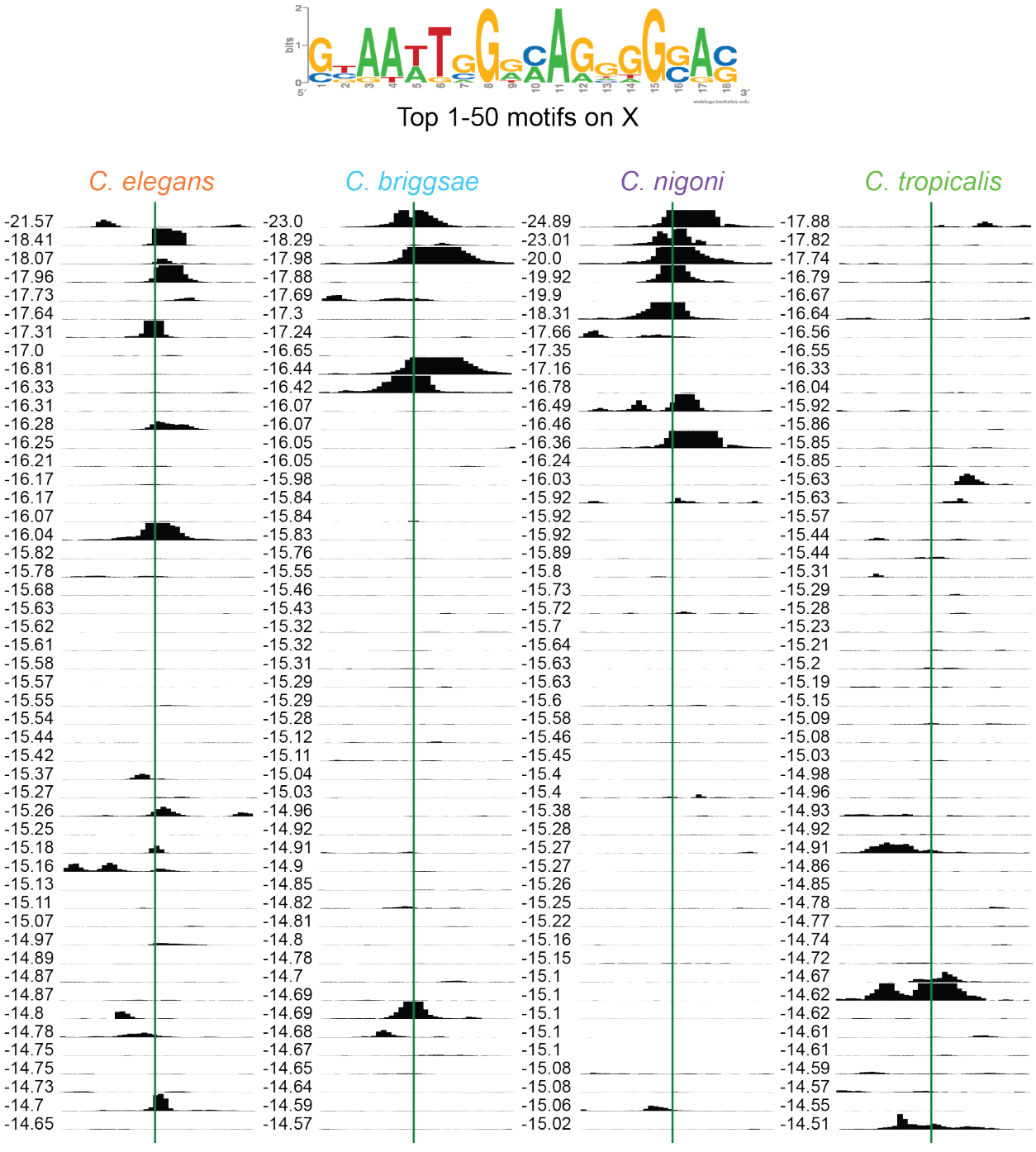


(c) **The *C. elegans* DCC is bound at few of the top 51-100 Cni-MEME3 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni-MEME3 motifs on the X chromosome in four species.

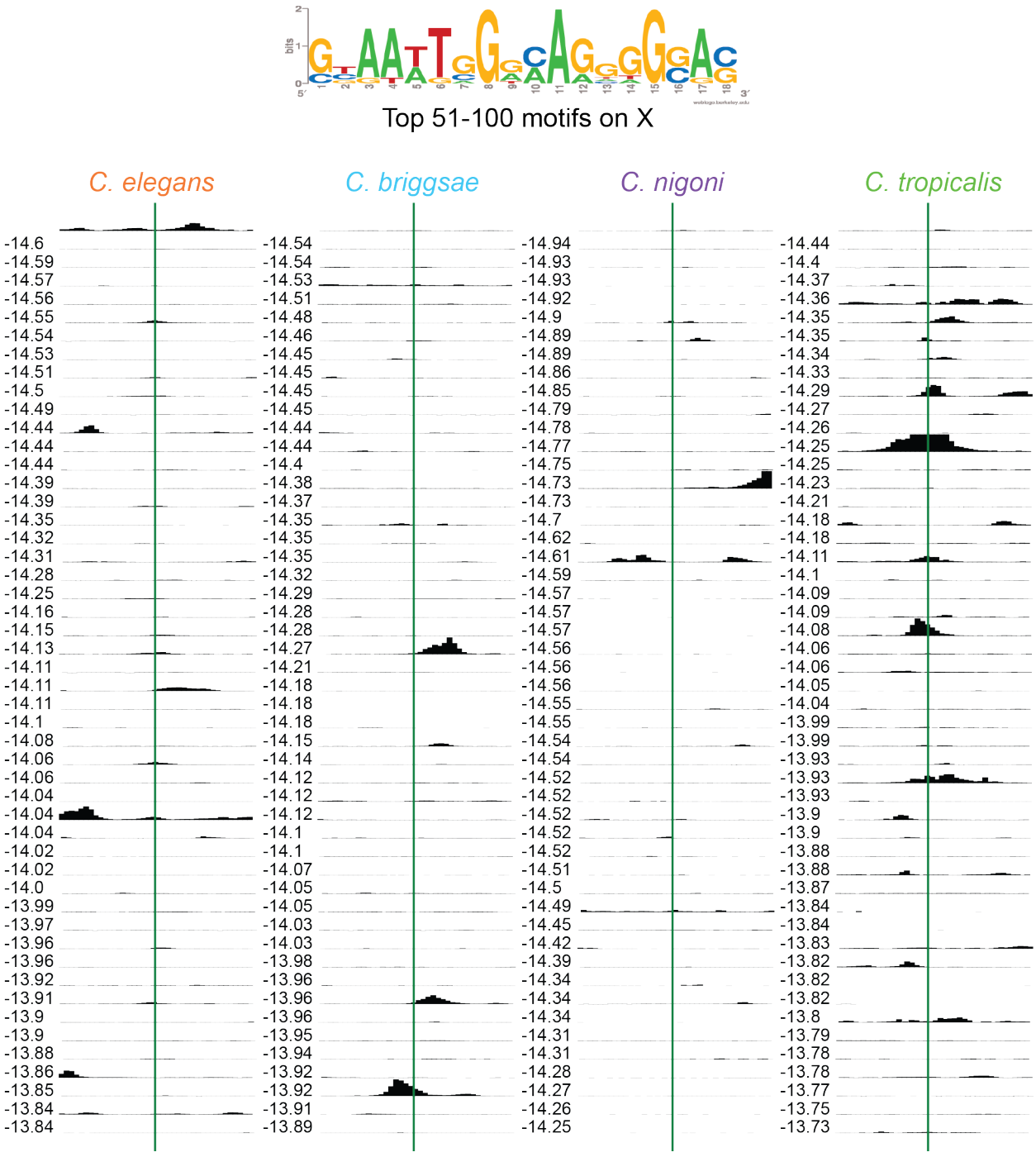
Figure G.12: Cni-top10-MEME1 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni-top10-MEME1 motif is X-enriched in *C. nigoni* and peak-enriched in *C. nigoni* and *C. briggsae*.** The Cni-top10-MEME1 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

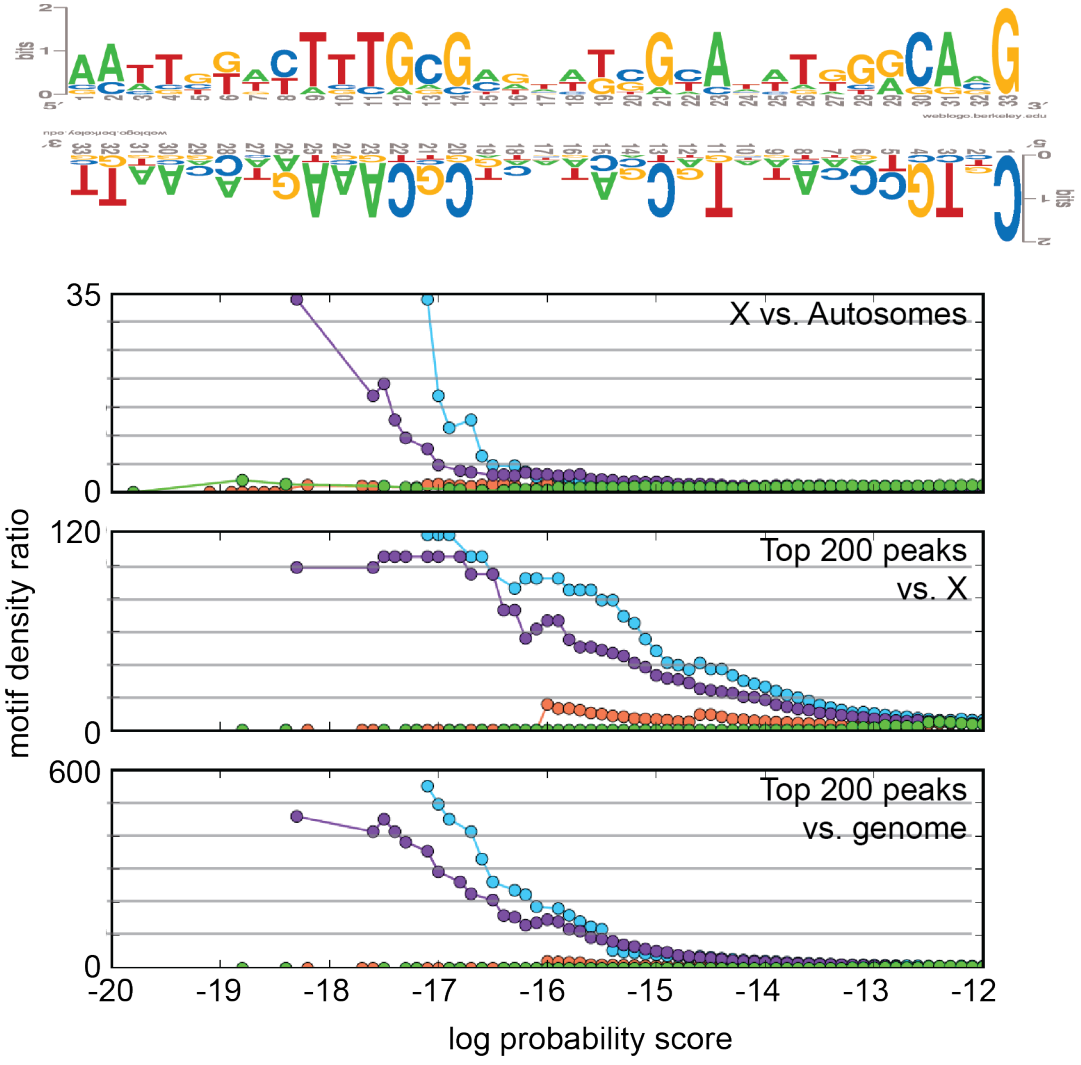


(b) **The DCC is bound at the strongest Cni-top10-MEME1 motifs on the X chromosome in *C. briggsae*, *C. nigoni*, and *C. elegans*.** The ChIP-seq signal is plotted at the top 50 Cni-top10-MEME1 motifs on the X chromosome in four species.

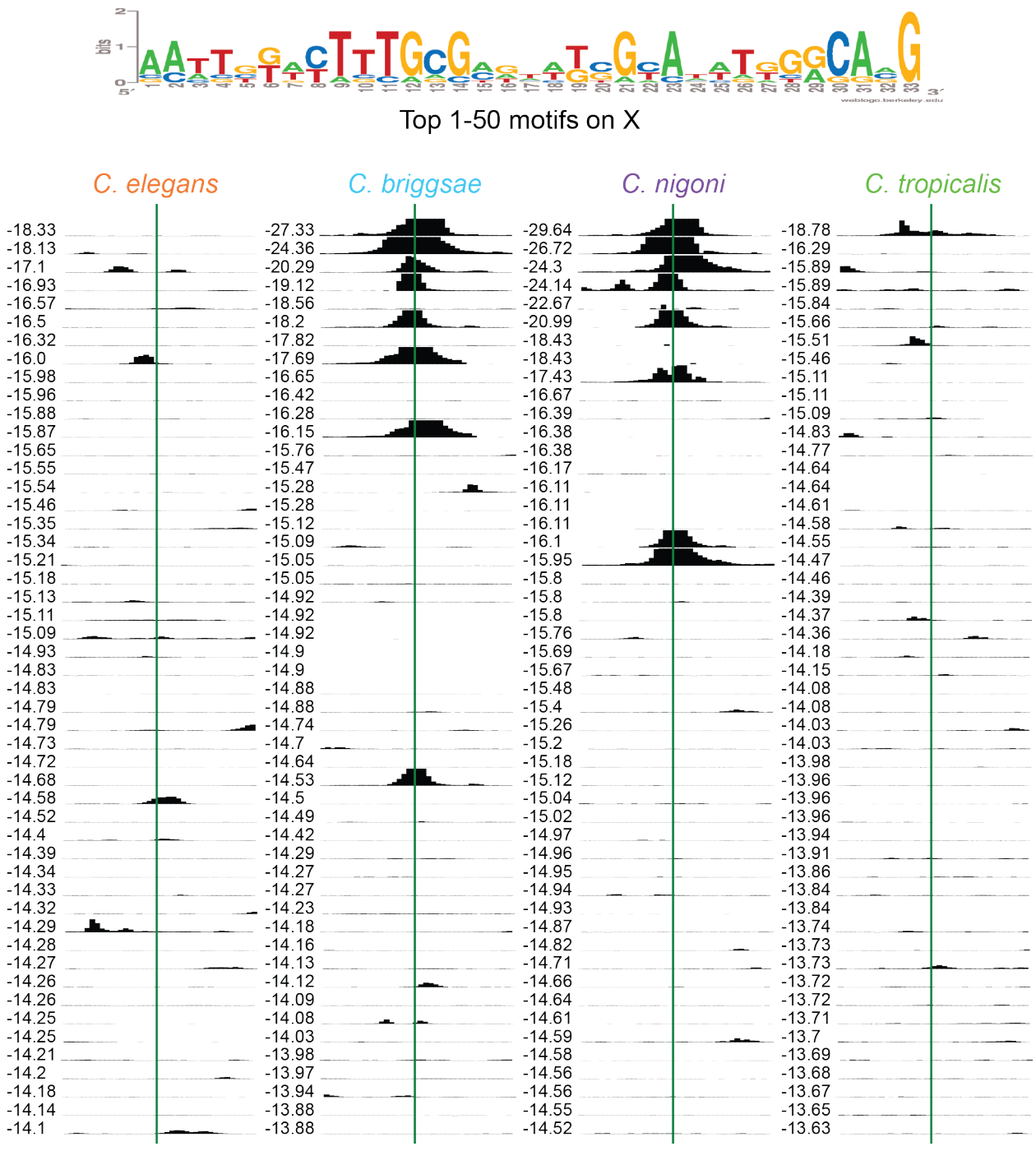


(c) **The *C. elegans* DCC is bound at few of the top 51-100 Cni-top10-MEME1 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni-top10-MEME1 motifs on the X chromosome in four species.

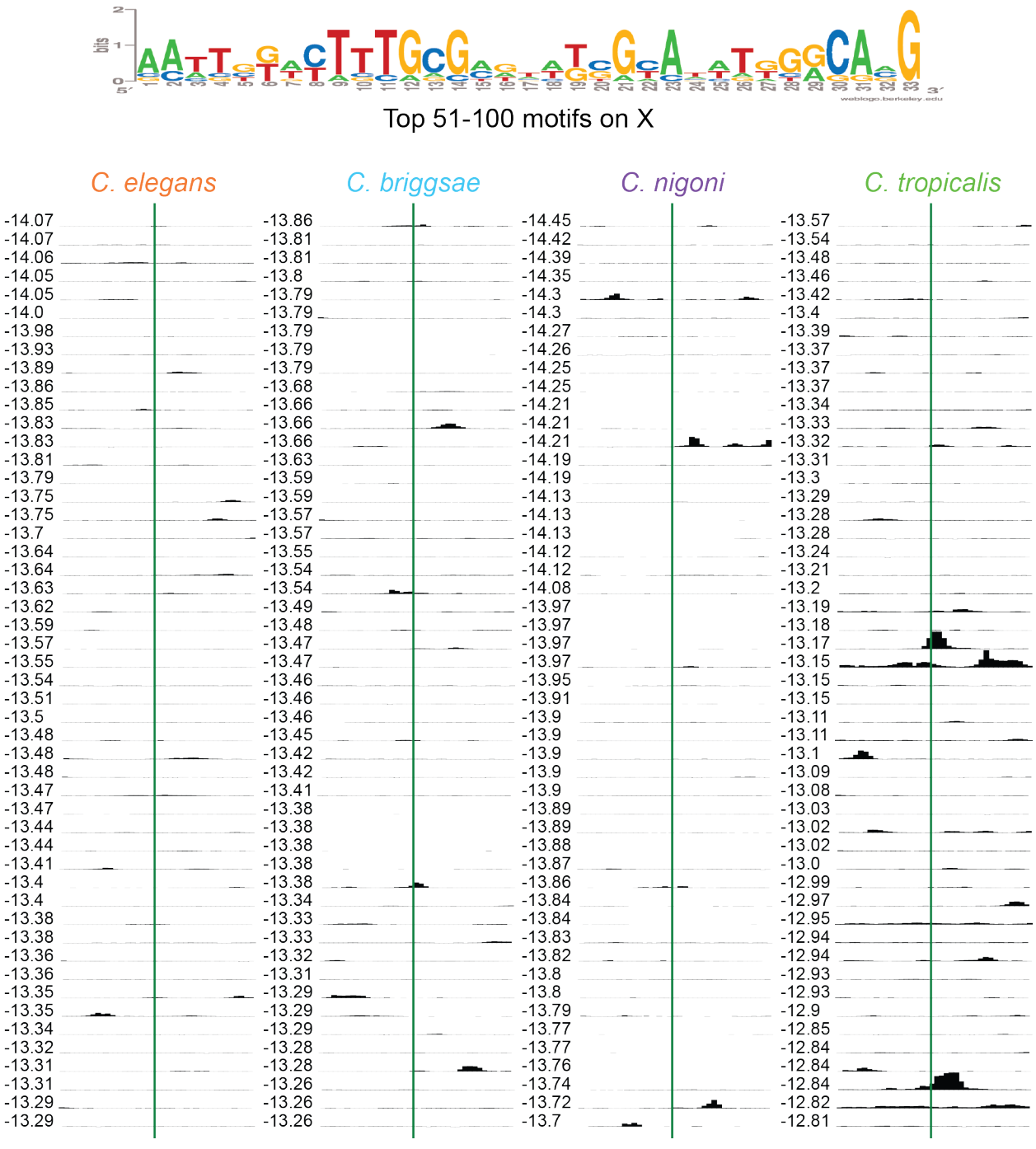
Figure G.13: Cni-top10-MEME2 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni-top10-MEME2 motif is X-enriched and peak-enriched in *C. nigoni* and *C. briggsae*.** The Cni-top10-MEME2 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

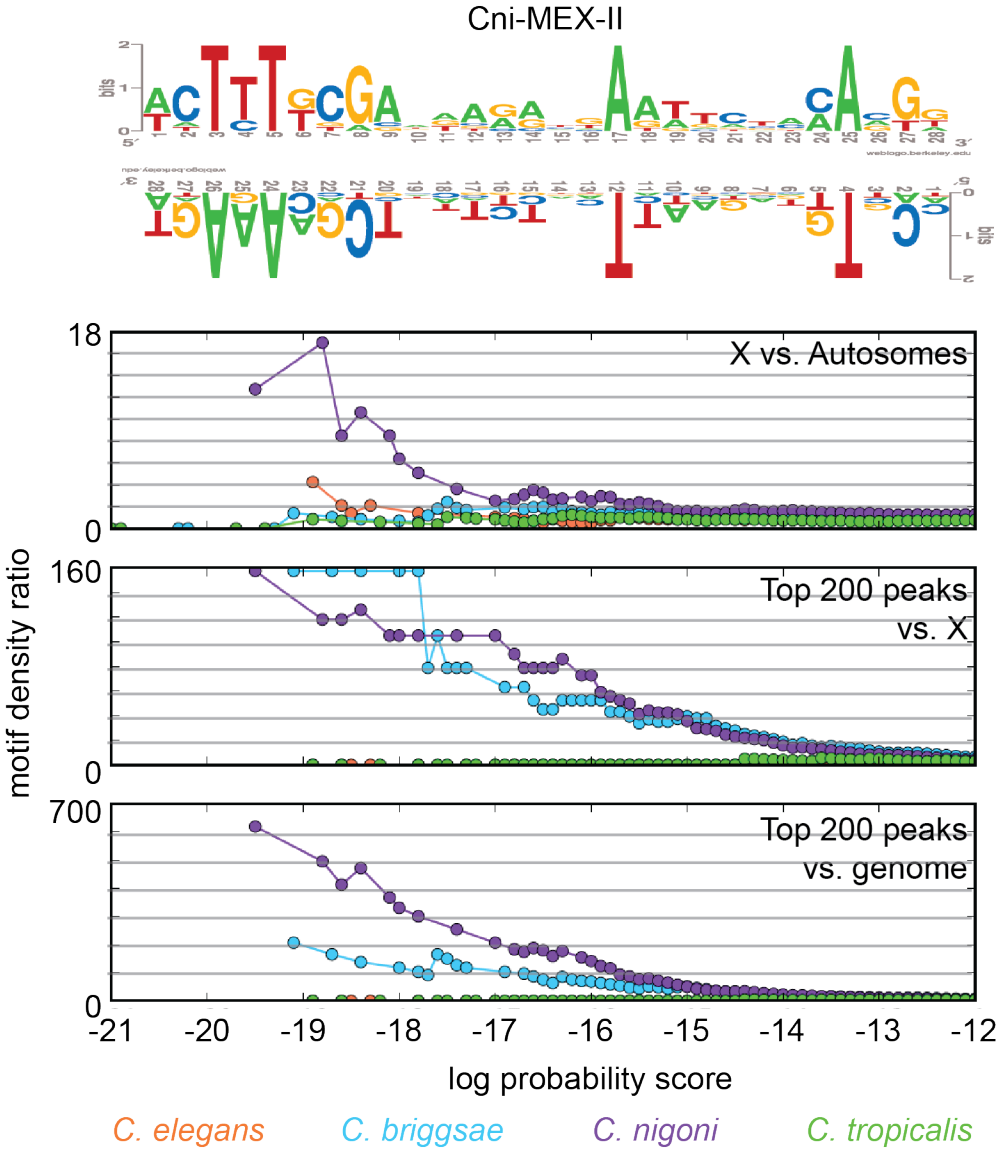


(b) **The DCC is bound at the strongest Cni-top10-MEME2 motifs on the X chromosome in *C. nigoni* and *C. briggsae*.** The ChIP-seq signal is plotted at the top 50 Cni-top10-MEME2 motifs on the X chromosome in four species.

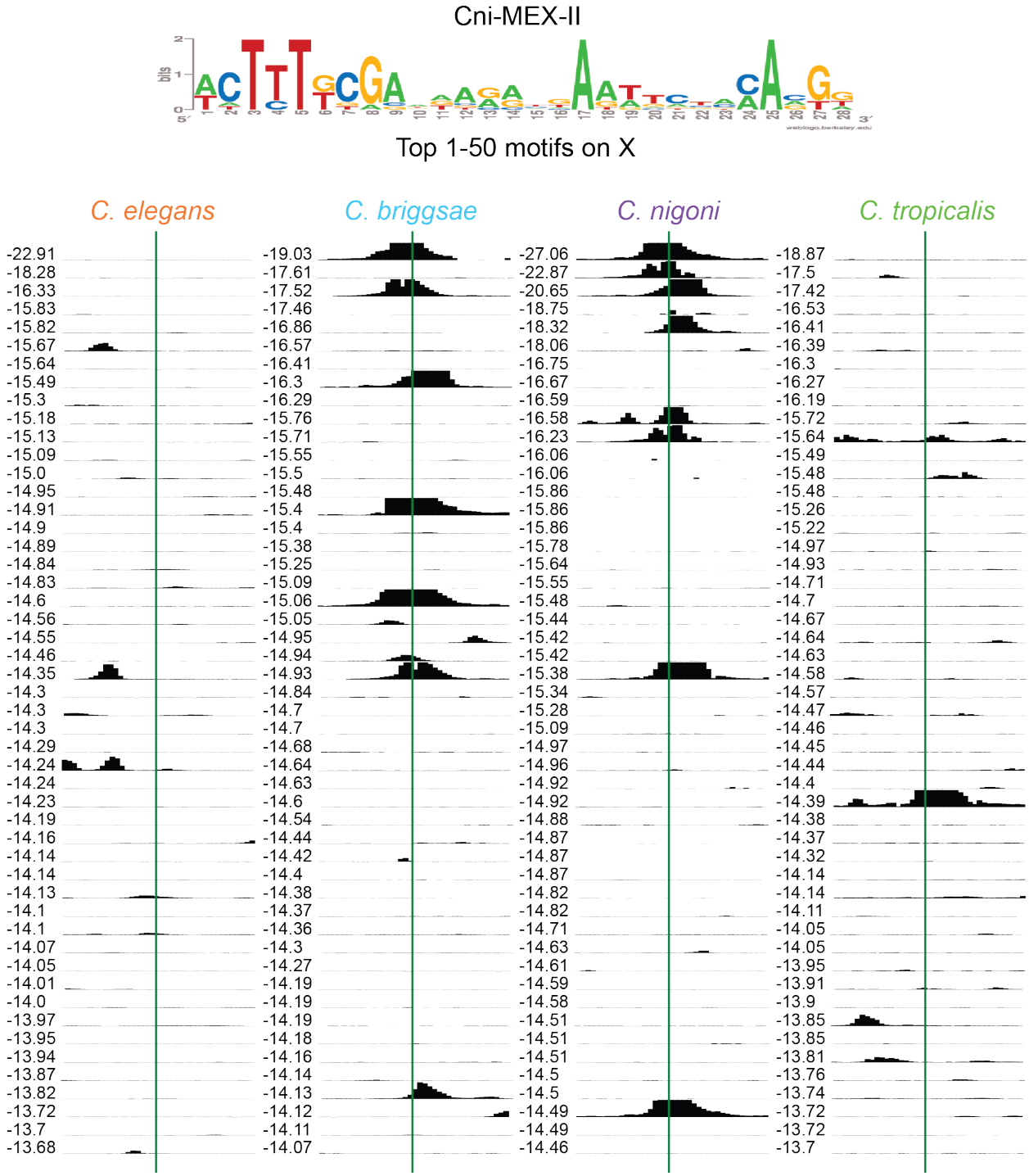


(c) **The DCC is bound at few of the top 51-100 Cni-top10-MEME2 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni-top10-MEME2 motifs on the X chromosome in four species.

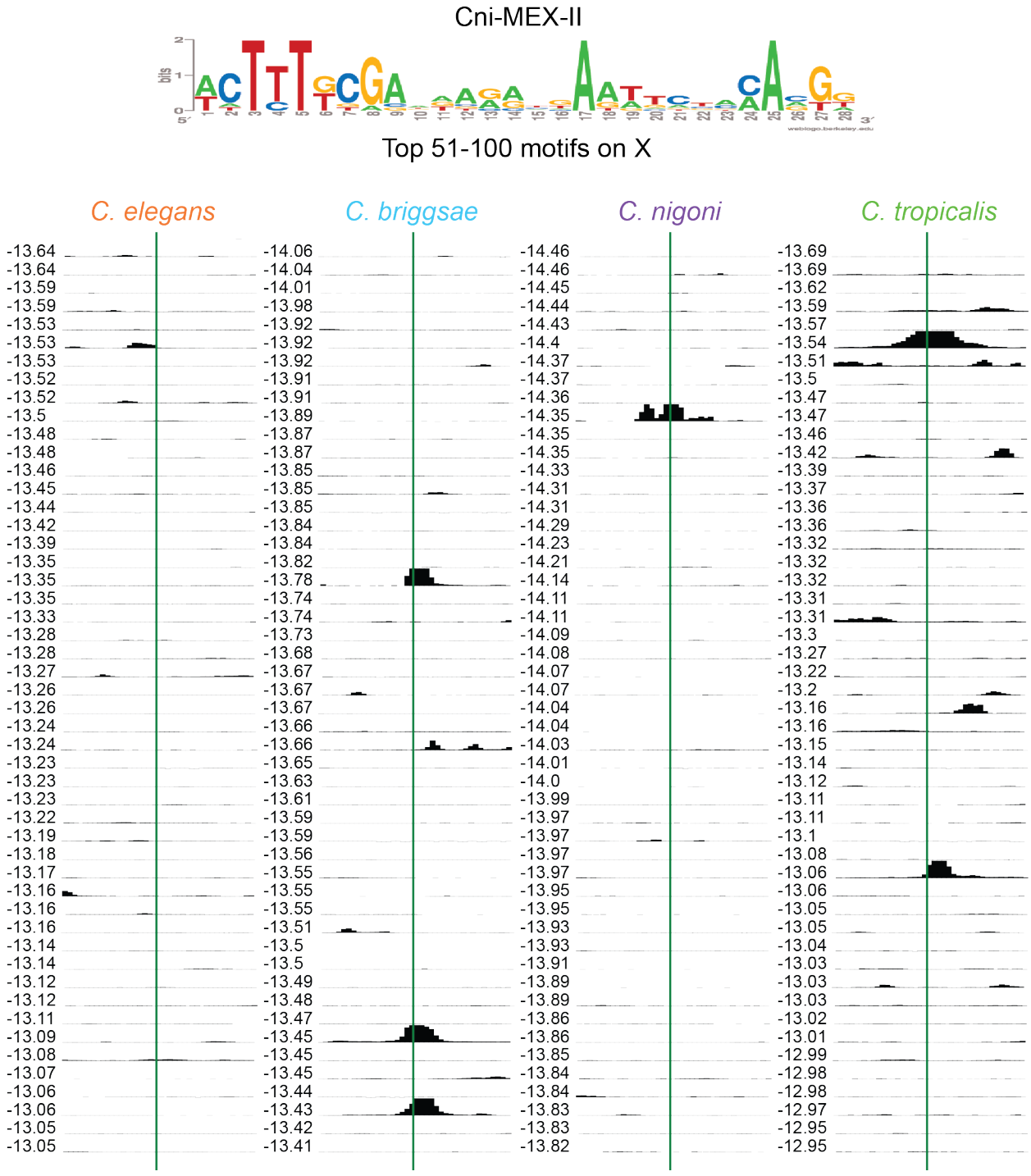
Figure G.14: Cni-top10-MEME3 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni-top10-MEME3 motif is X-enriched in *C. nigoni* and peak-enriched in *C. nigoni* and *C. briggsae*.** The Cni-top10-MEME3 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

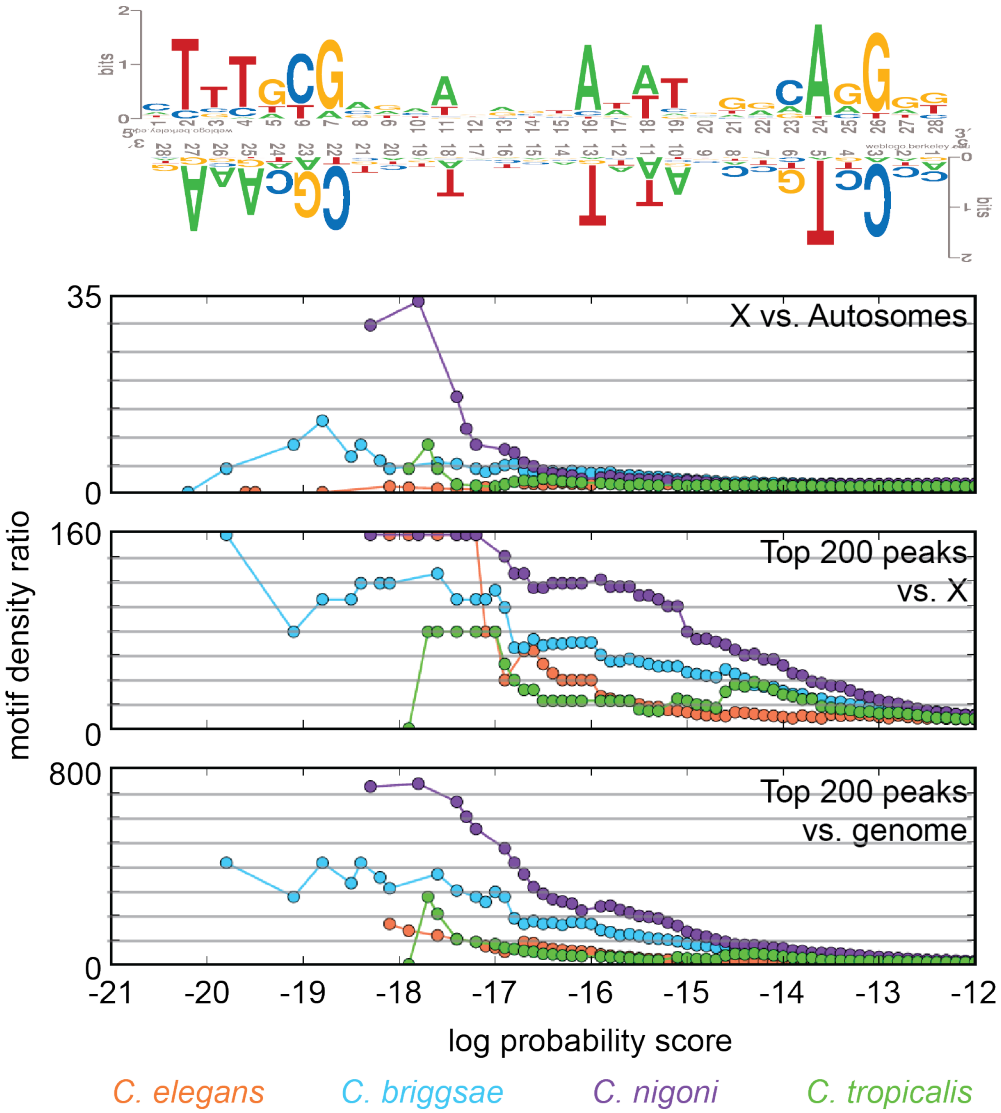


(b) **The DCC is bound at many of the top 50 Cni-top10-MEME3 motifs on the X chromosome in *C. briggsae* and *C. nigoni*.** The ChIP-seq signal is plotted at the top 50 Cni-top10-MEME3 motifs on the X chromosome in four species.

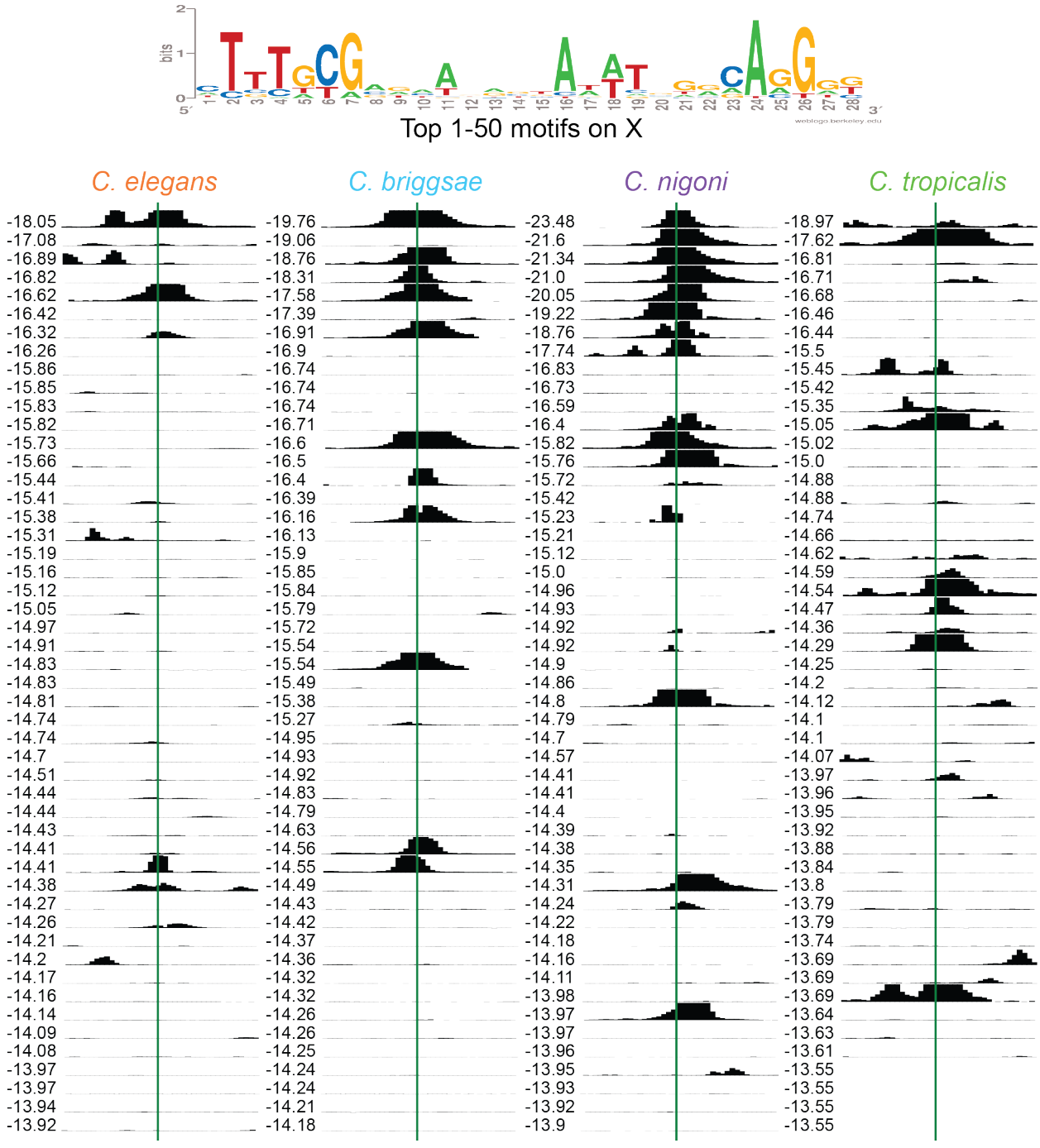


(c) **The DCC is also bound at some of the top 51-100 Cni-top10-MEME3 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni-top10-MEME3 motifs on the X chromosome in four species.

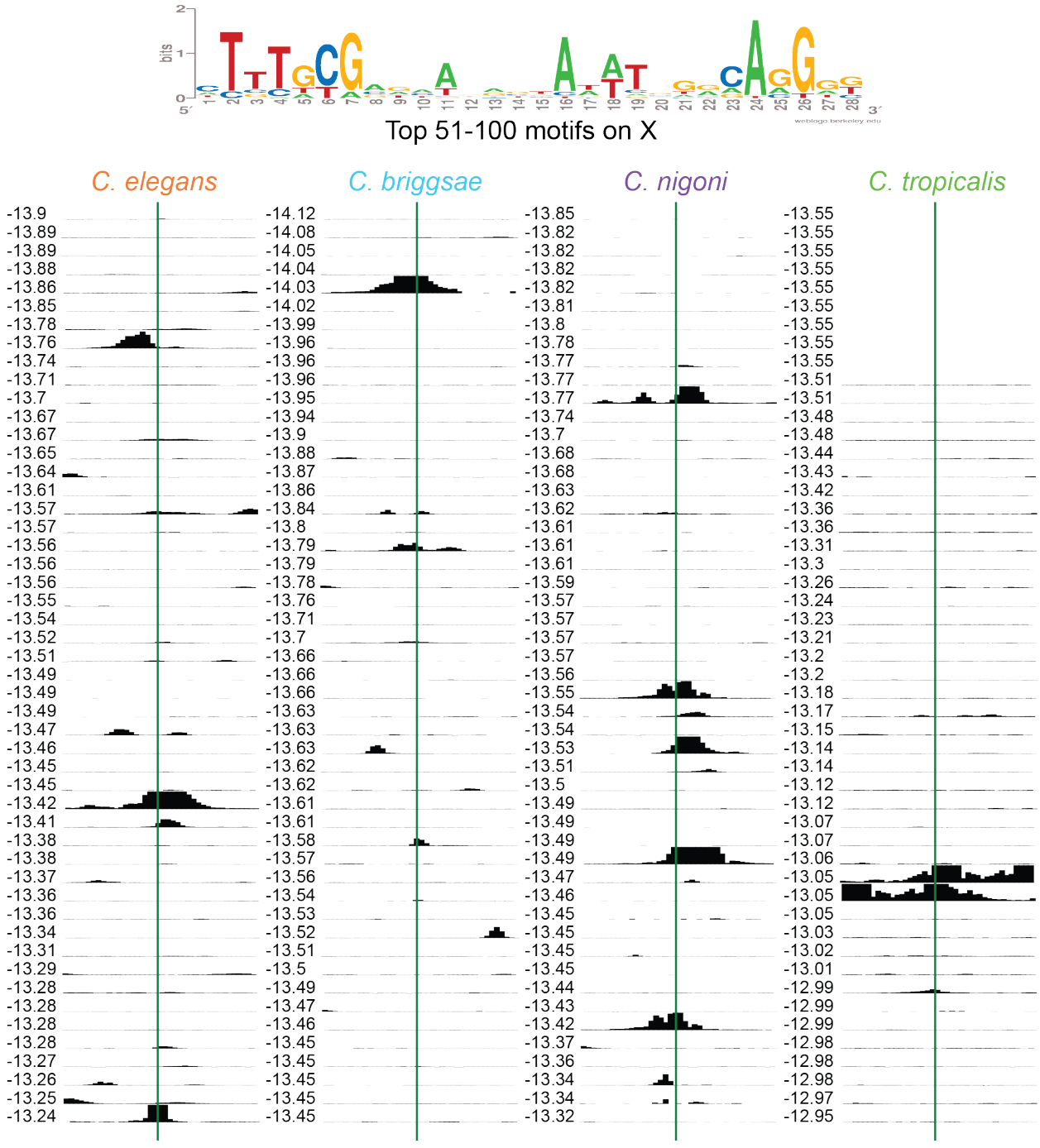
Figure G.15: CS179Cni10-1 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The CS179Cni10-1 motif is X-enriched in *C. nigoni* and peak-enriched in *C. nigoni*, *C. briggsae*, and *C. elegans*.** The CS179Cni10-1 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

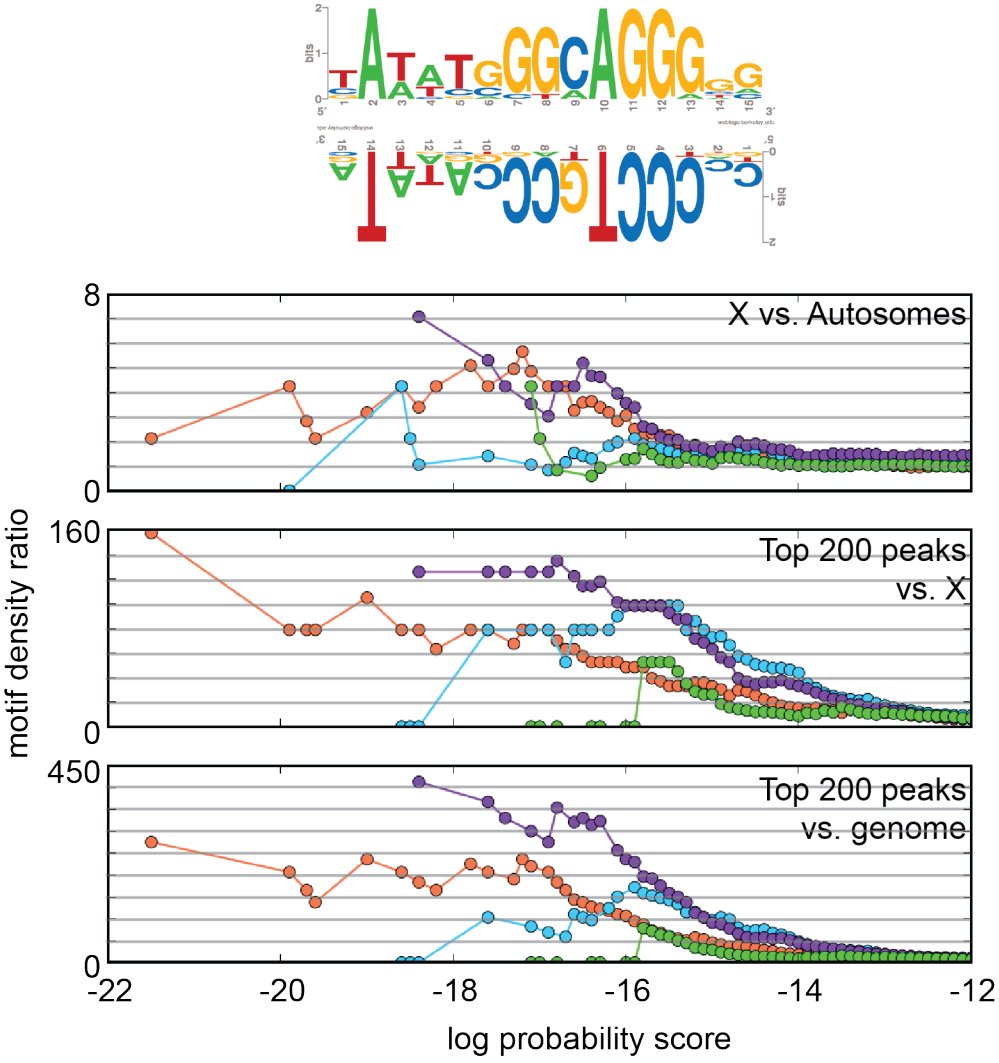


(b) **The DCC is bound at most of the top 50 CS179Cni10-1 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 CS179Cni10-1 motifs on the X chromosome in four species.

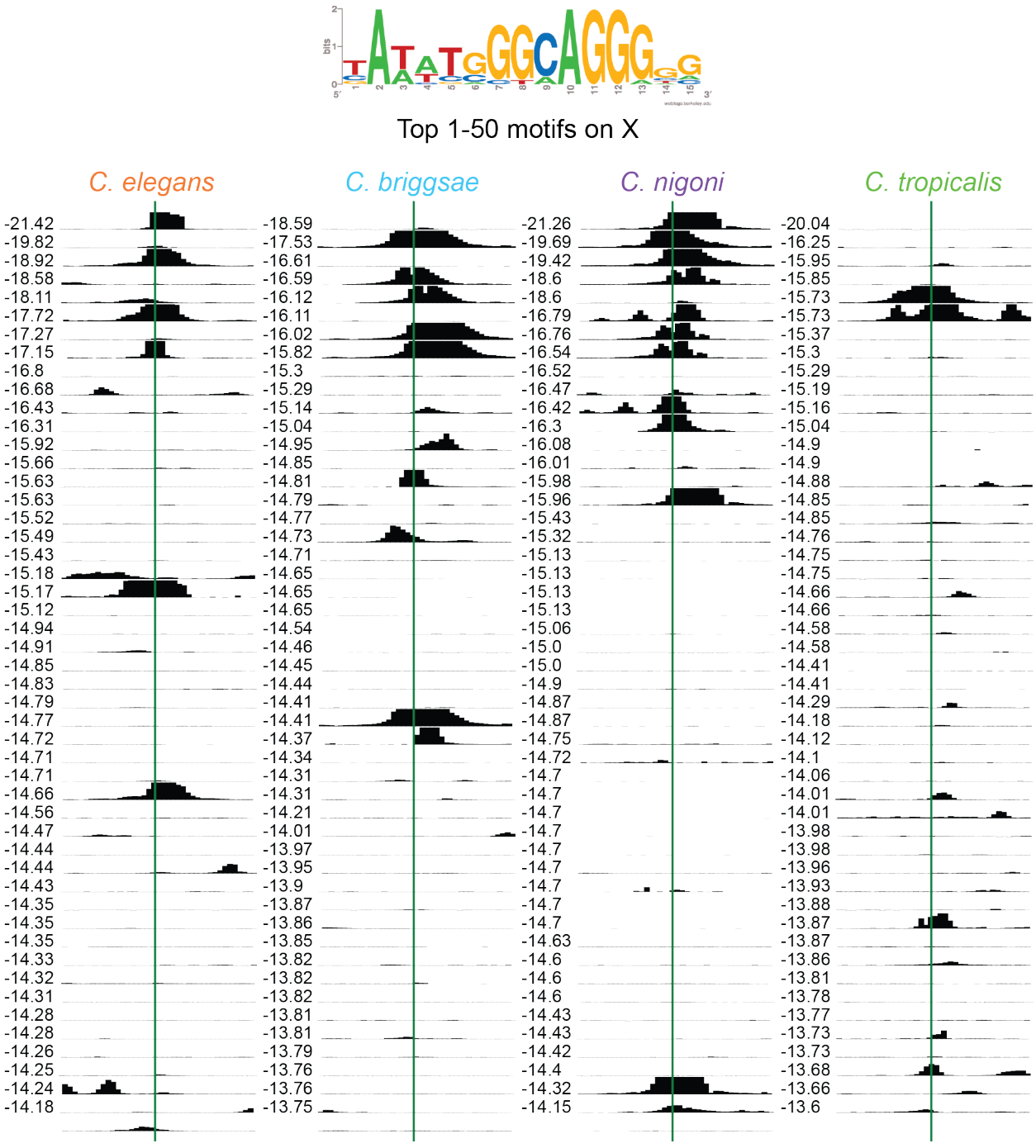


(c) **The DCC is also bound at many of the top 51-100 CS179Cni10-1 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 CS179Cni10-1 motifs on the X chromosome in four species.

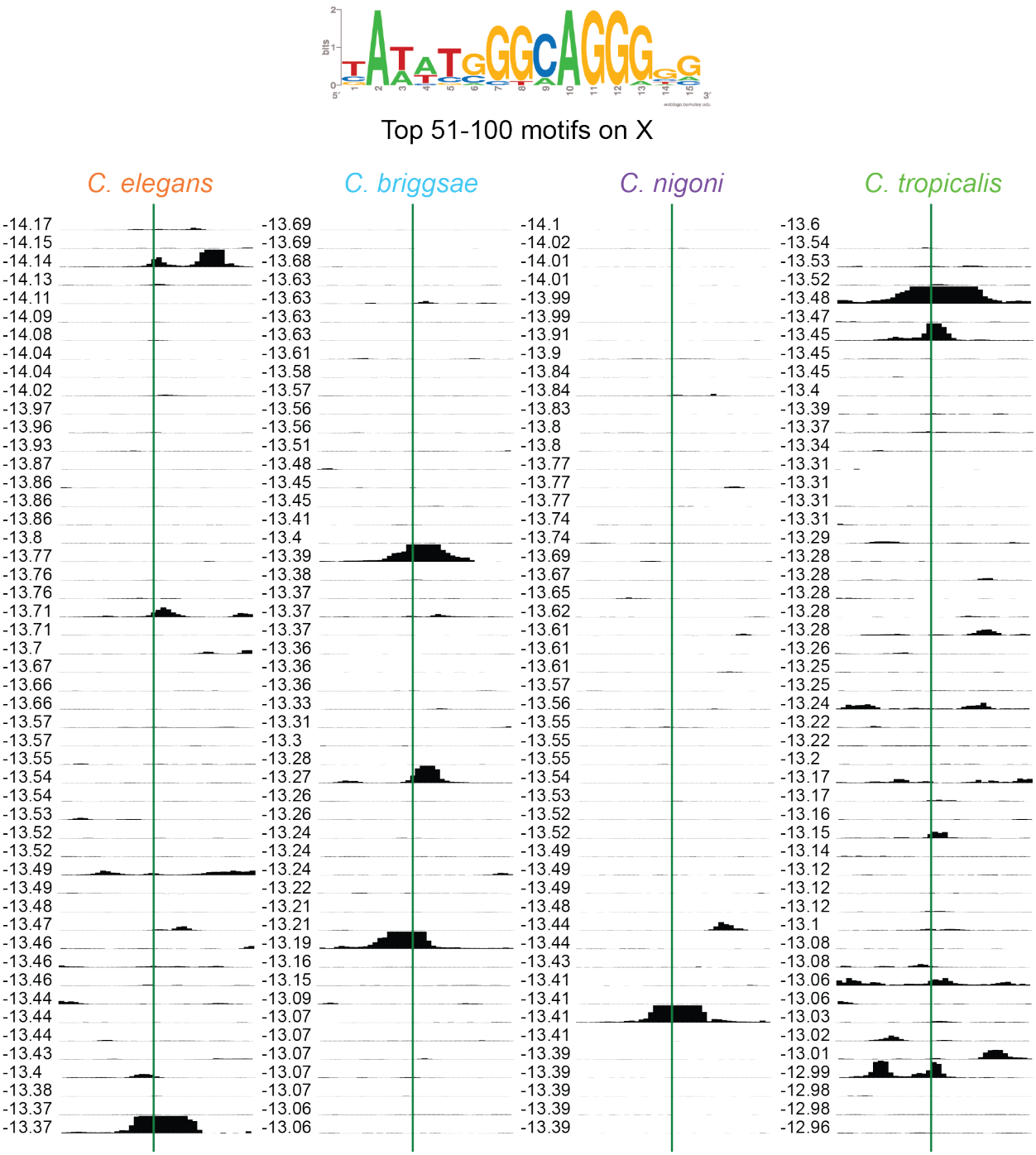
Figure G.16: CS179Cni15-3 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The CS179Cni15-3 motif is X-enriched and peak-enriched in *C. nigoni* and *C. elegans*.** The CS179Cni15-3 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

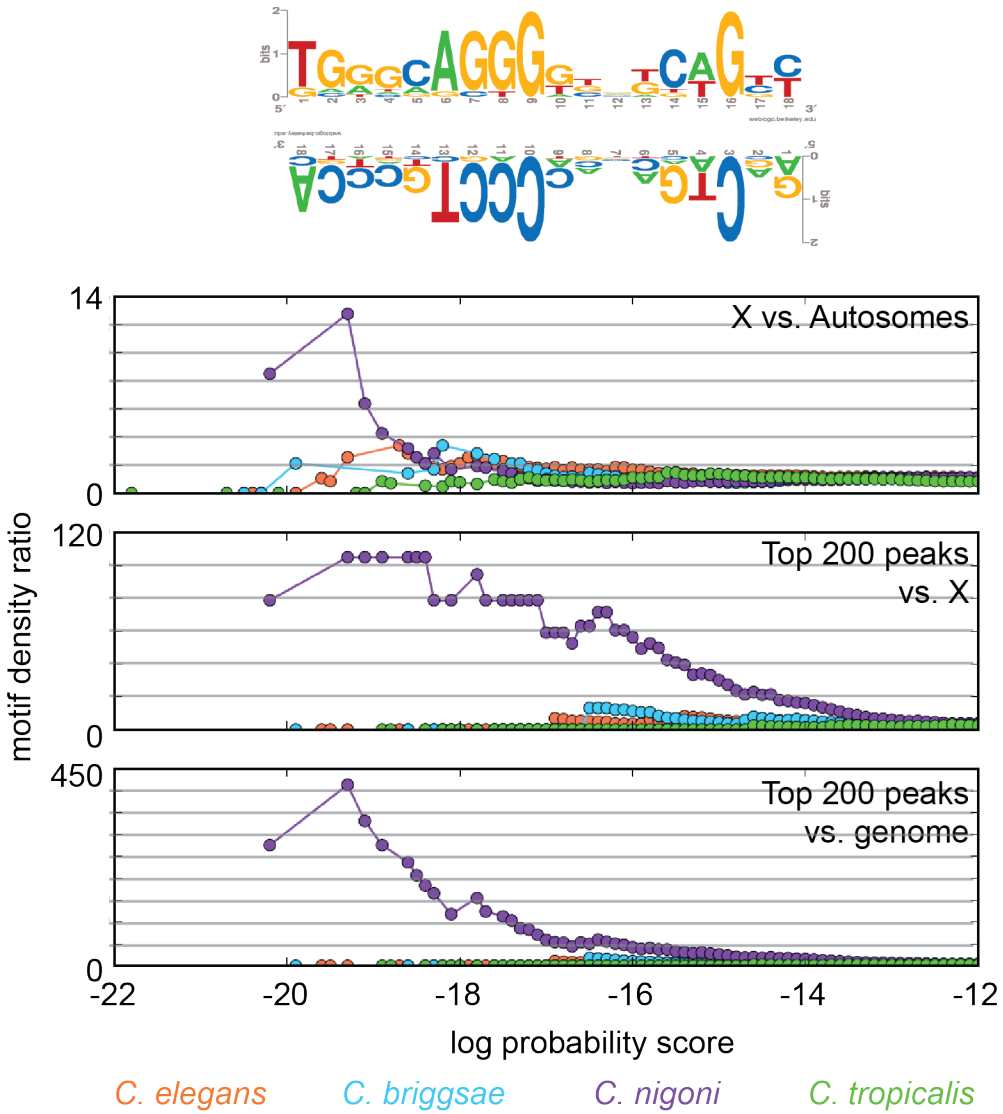


(b) **The DCC is bound at most of the top 50 CS179Cni15-3 motifs on the X chromosome in *C. briggsae*, *C. nigoni*, and *C. elegans*.** The ChIP-seq signal is plotted at the top 50 CS179Cni15-3 motifs on the X chromosome in four species.

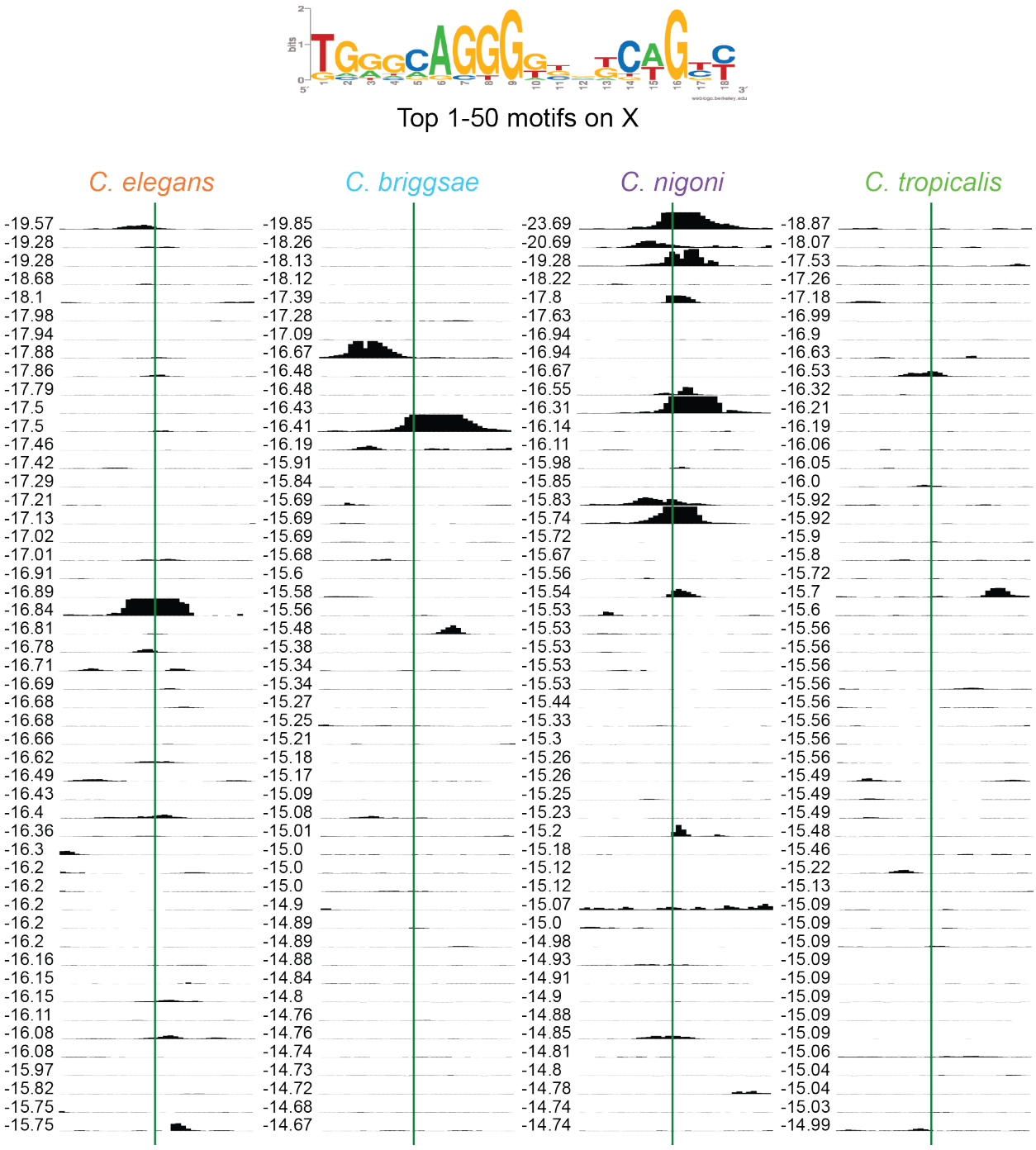


(c) **The DCC is also bound at some of the top 51-100 CS179Cni15-3 motifs on the X chromosome in all four species.** The ChIP-seq signal is plotted at the top 51-100 CS179Cni15-3 motifs on the X chromosome in four species.

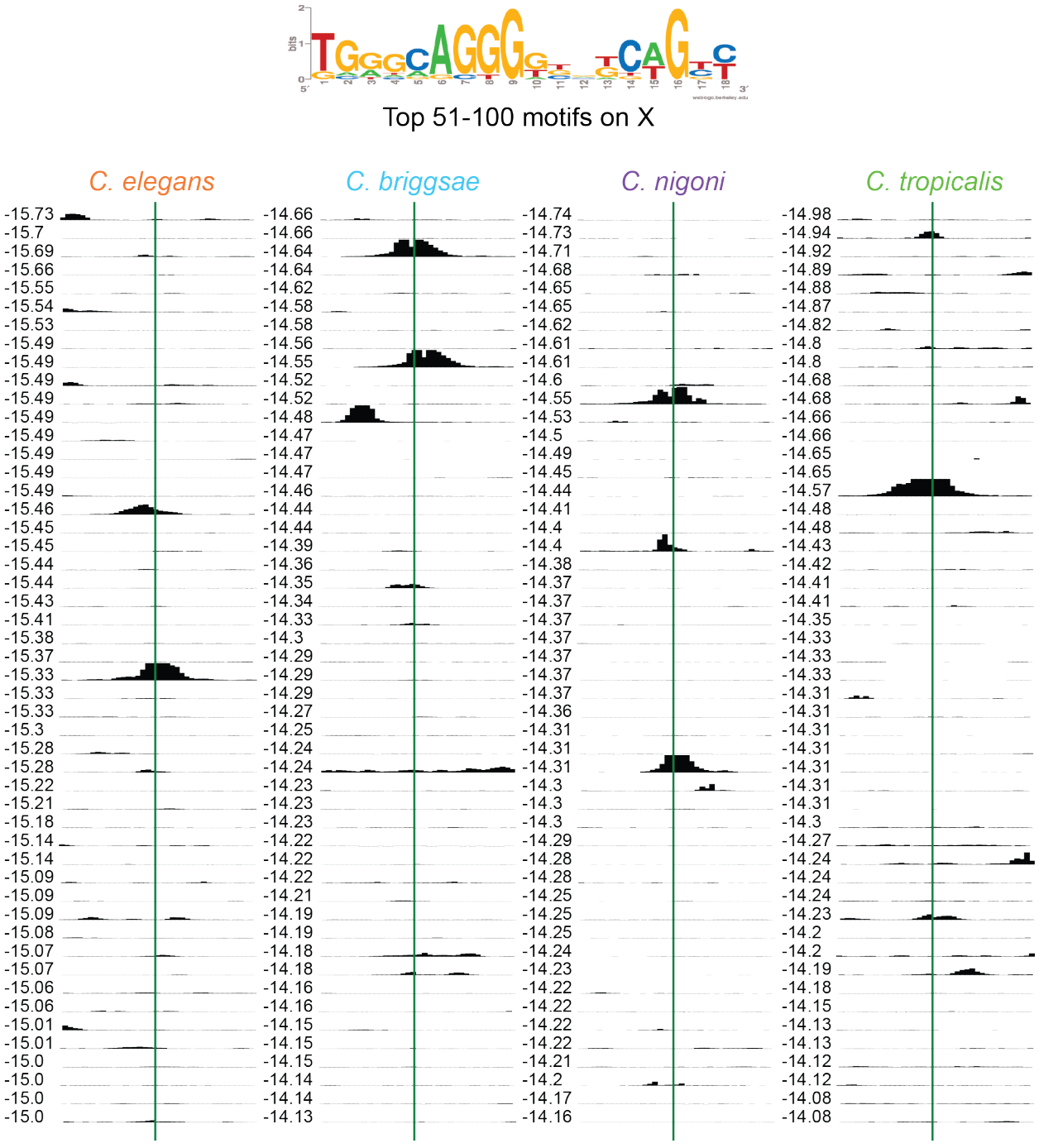
Figure G.17: Cni14-MEME1 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni14-MEME1 motif is X-enriched and peak-enriched in *C. nigoni*.** The Cni14-MEME1 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

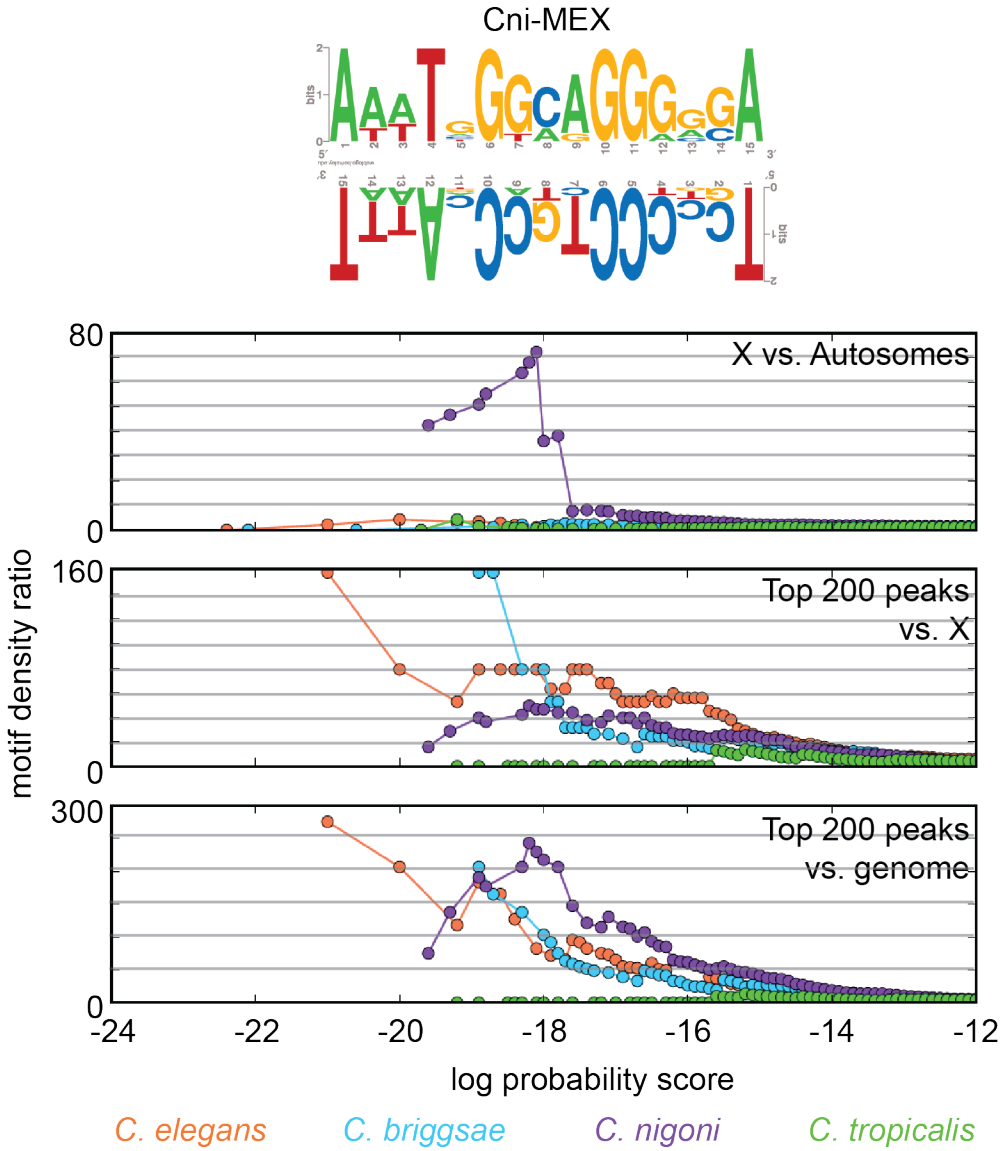


(b) **The DCC is bound at some of the top 50 Cni14-MEME1 motifs on the X chromosome in *C. nigoni*, *C. briggsae*, and *C. elegans*.** The ChIP-seq signal is plotted at the top 50 Cni14-MEME1 motifs on the X chromosome in four species.

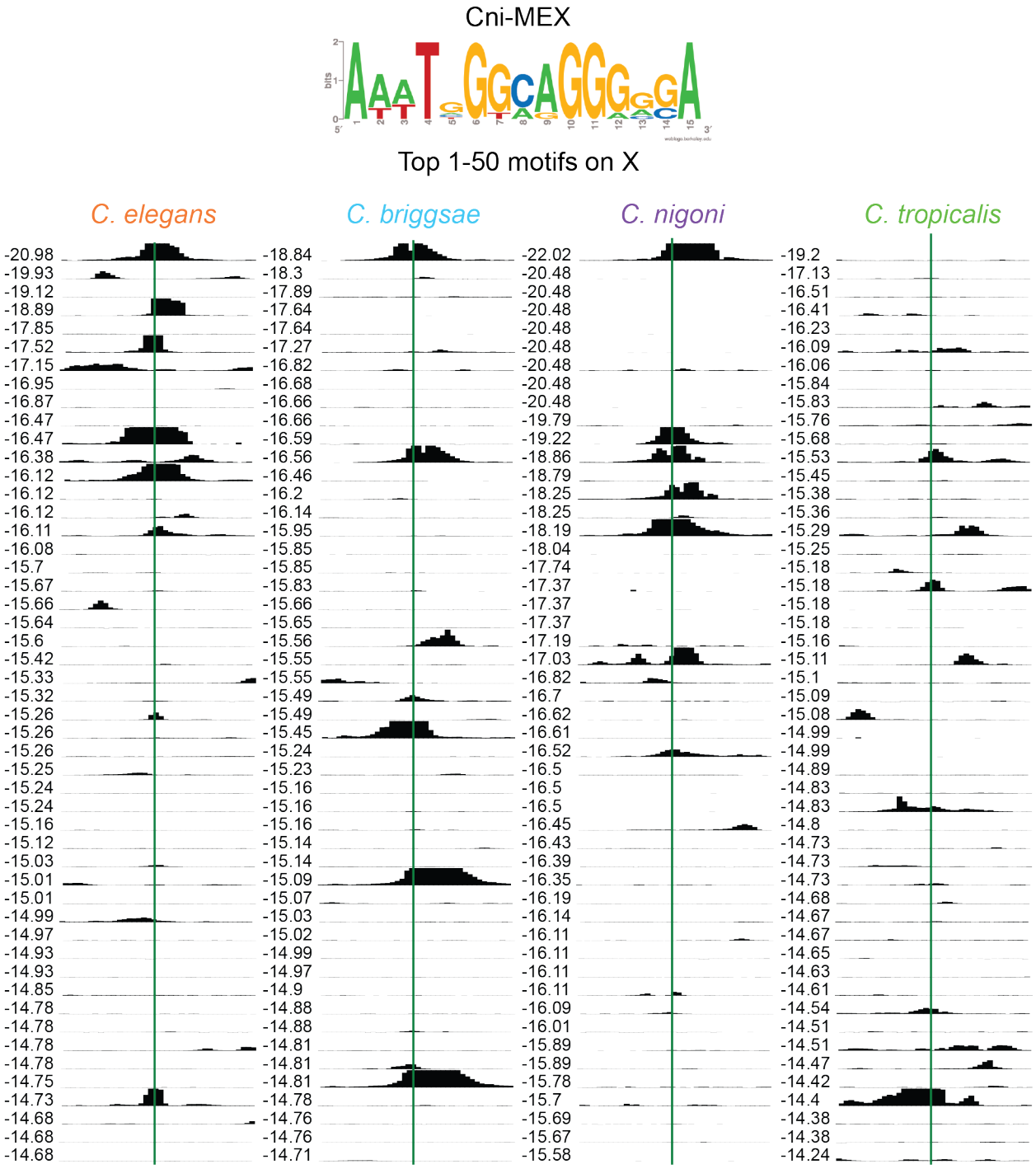


(c) **The DCC is also bound at some of the top 51-100 Cni14-MEME1 motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 51-100 Cni14-MEME1 motifs on the X chromosome in four species.

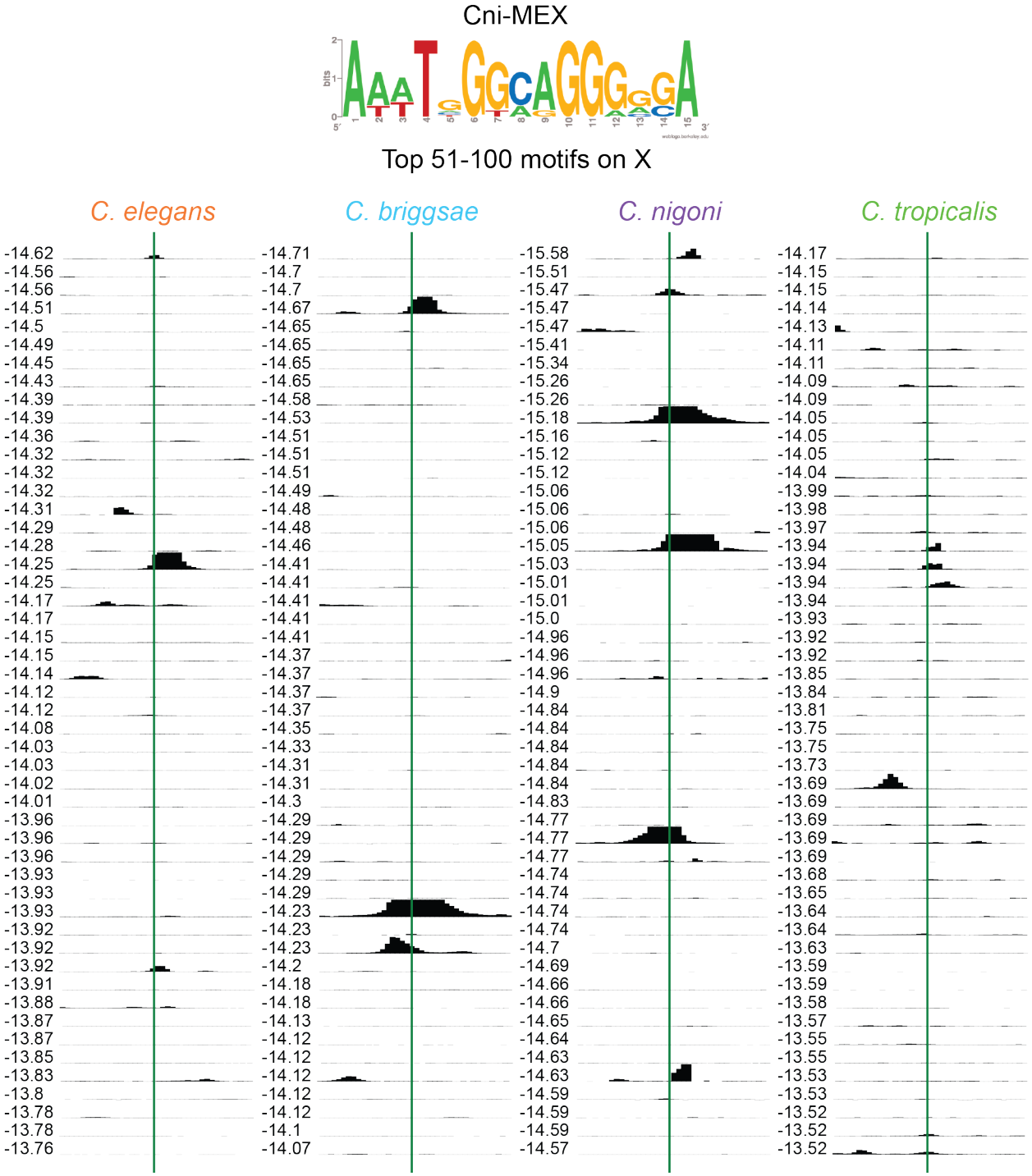
Figure G.18: Cni14-MEME2 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Cni14-MEME2 motif is X-enriched in *C. nigoni* and peak-enriched in *C. nigoni*, *C. briggsae*, and *C. elegans*.** The Cni14-MEME2 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

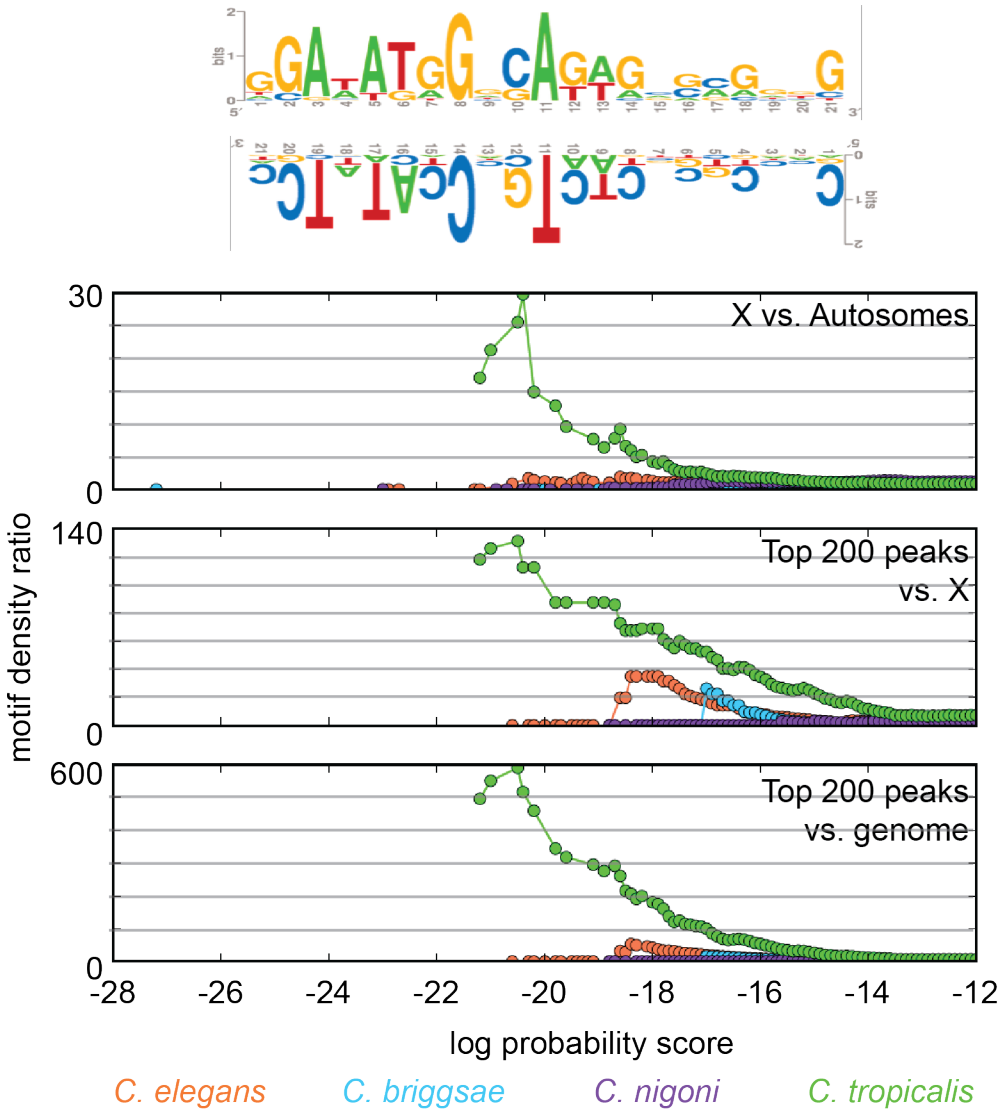


(b) **The DCC is bound at many of the top 50 Cni14-MEME2 motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 50 Cni14-MEME2 motifs on the X chromosome in four species.

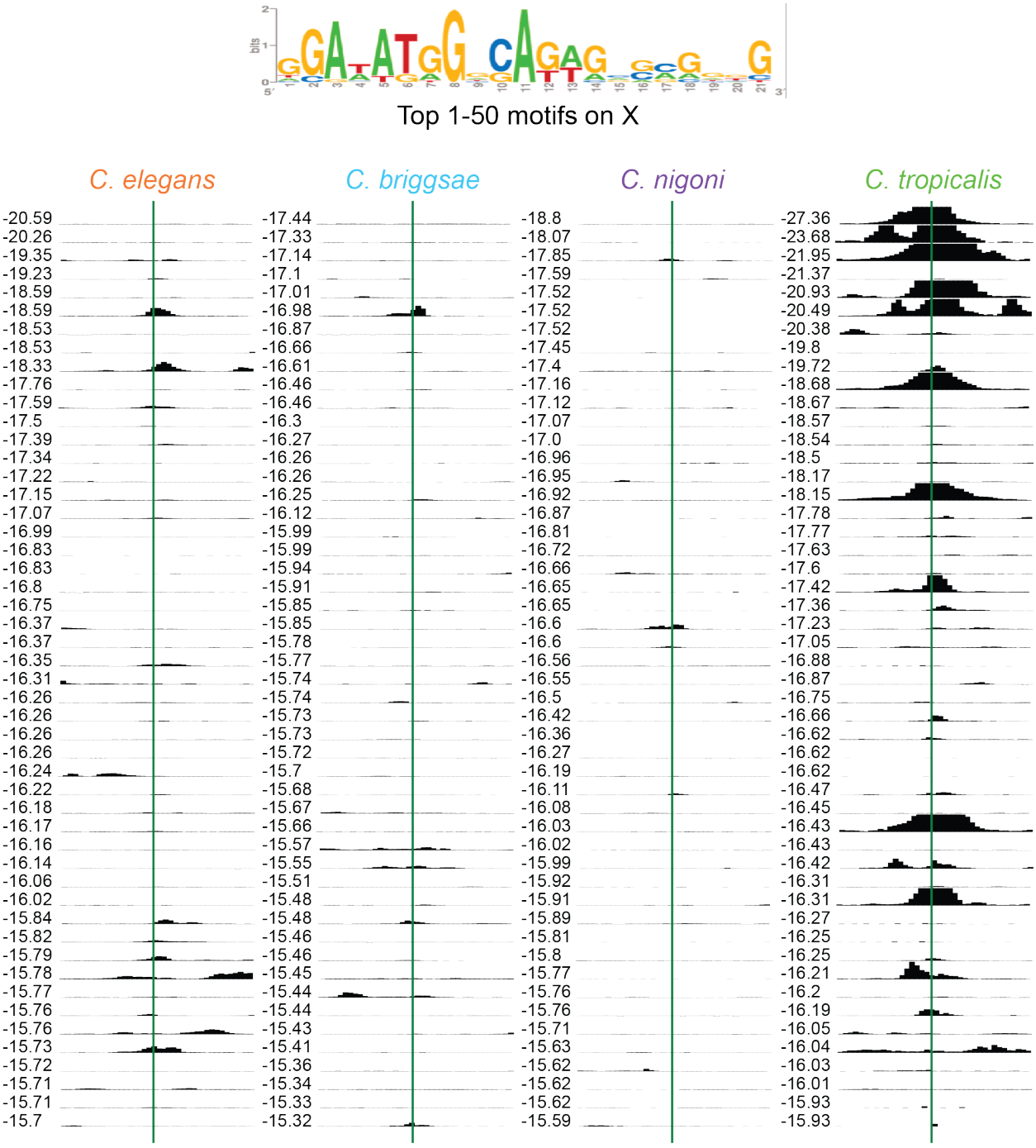


(c) **The DCC is also bound at some of the top 51-100 Cni14-MEME2 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Cni14-MEME2 motifs on the X chromosome in four species.

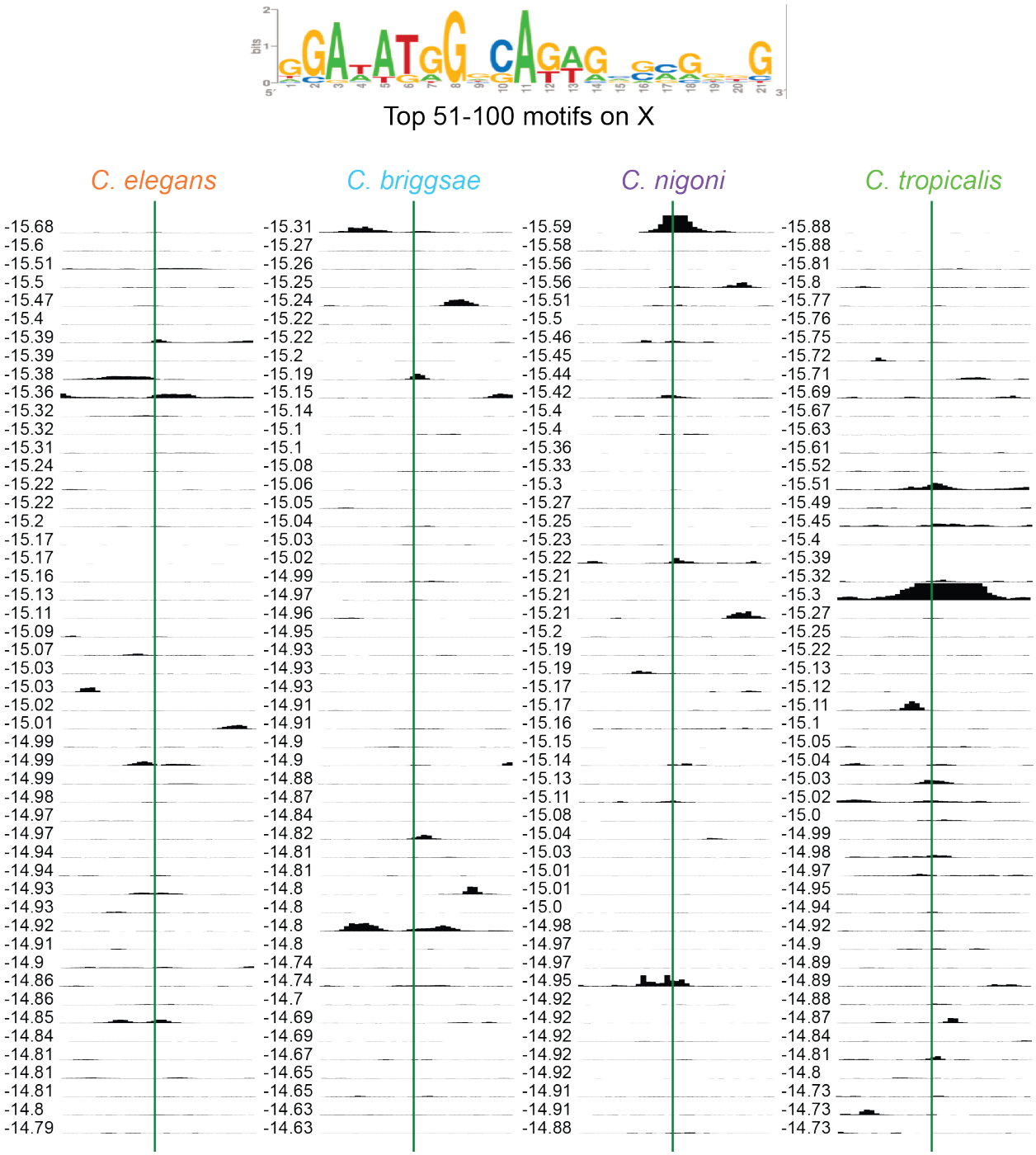
Figure G.19: Ctr-top17 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The Ctr-top17 motif is X-enriched and peak-enriched in *C. tropicalis*.** The Ctr-top17 motif is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

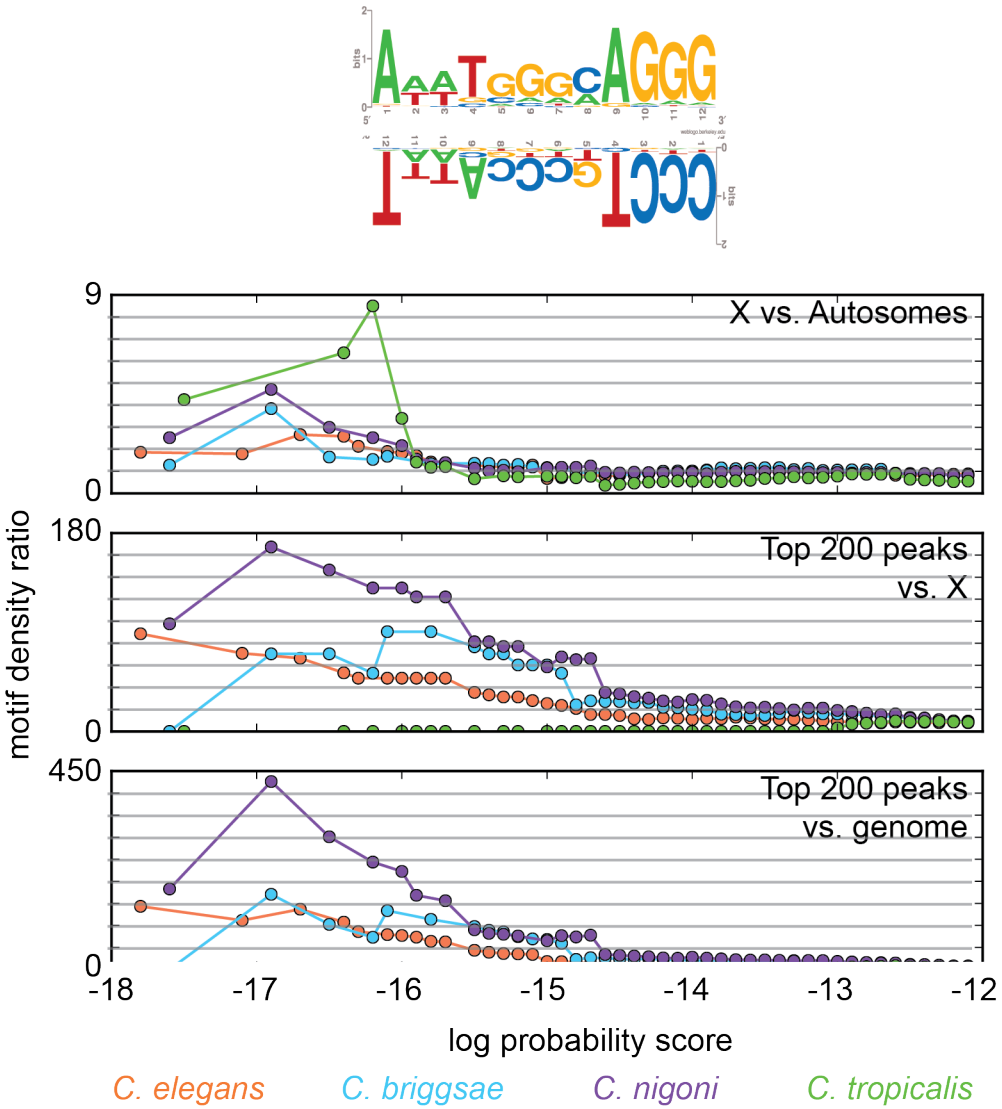


(b) **The *C. tropicalis* DCC is bound at most of the top 50 Ctr-top17 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 50 Ctr-top17 motifs on the X chromosome in four species.

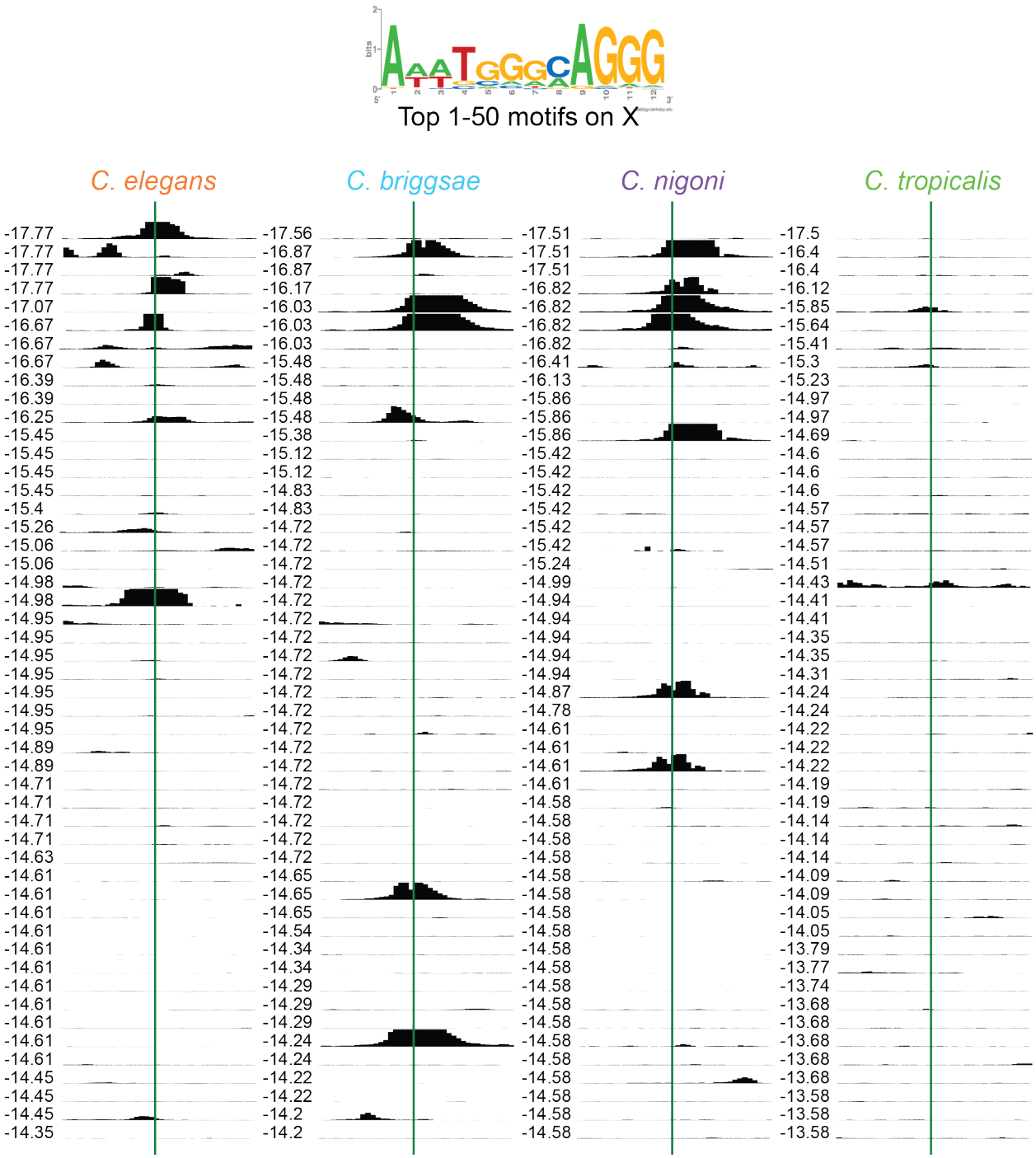


(c) **The *C. tropicalis* DCC is also bound at some of the top 51-100 Ctr-top17 motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 Ctr-top17 motifs on the X chromosome in four species.

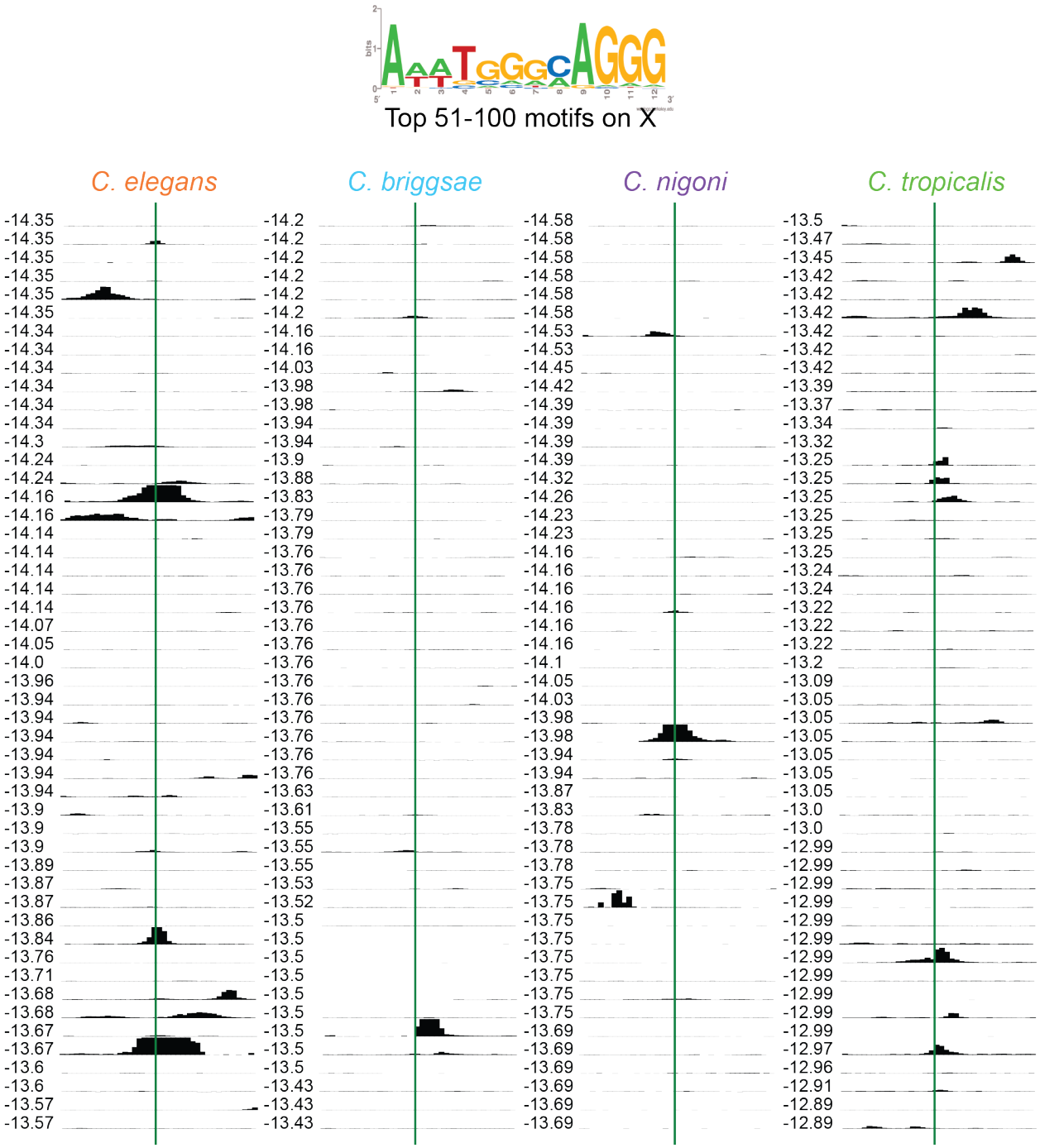
Figure G.20: Averaged 8 *C. briggsae* and *C. nigoni* MEX-like motifs a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The average of 8 MEX-like motifs is X-enriched in *C. tropicalis* and peak-enriched in *C. nigoni* and *C. elegans*.** The average of 8 MEX-like motifs is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

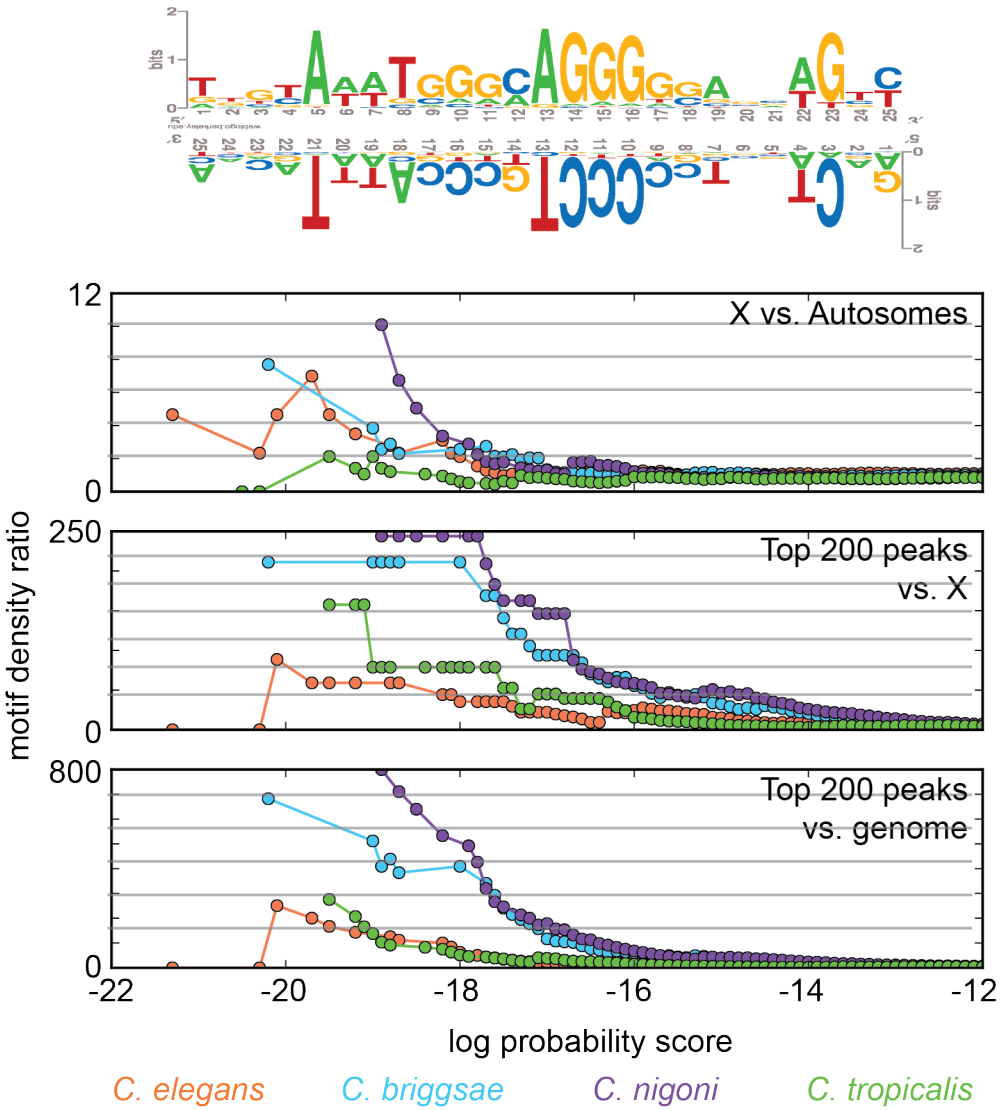


(b) **The DCC is bound at many of the top 50 average of 8 MEX-like motifs on the X chromosome in *C. briggsae*, *C. nigoni*, and *C. tropicalis*.** The ChIP-seq signal is plotted at the top 50 average of 8 MEX-like motifs on the X chromosome in four species.

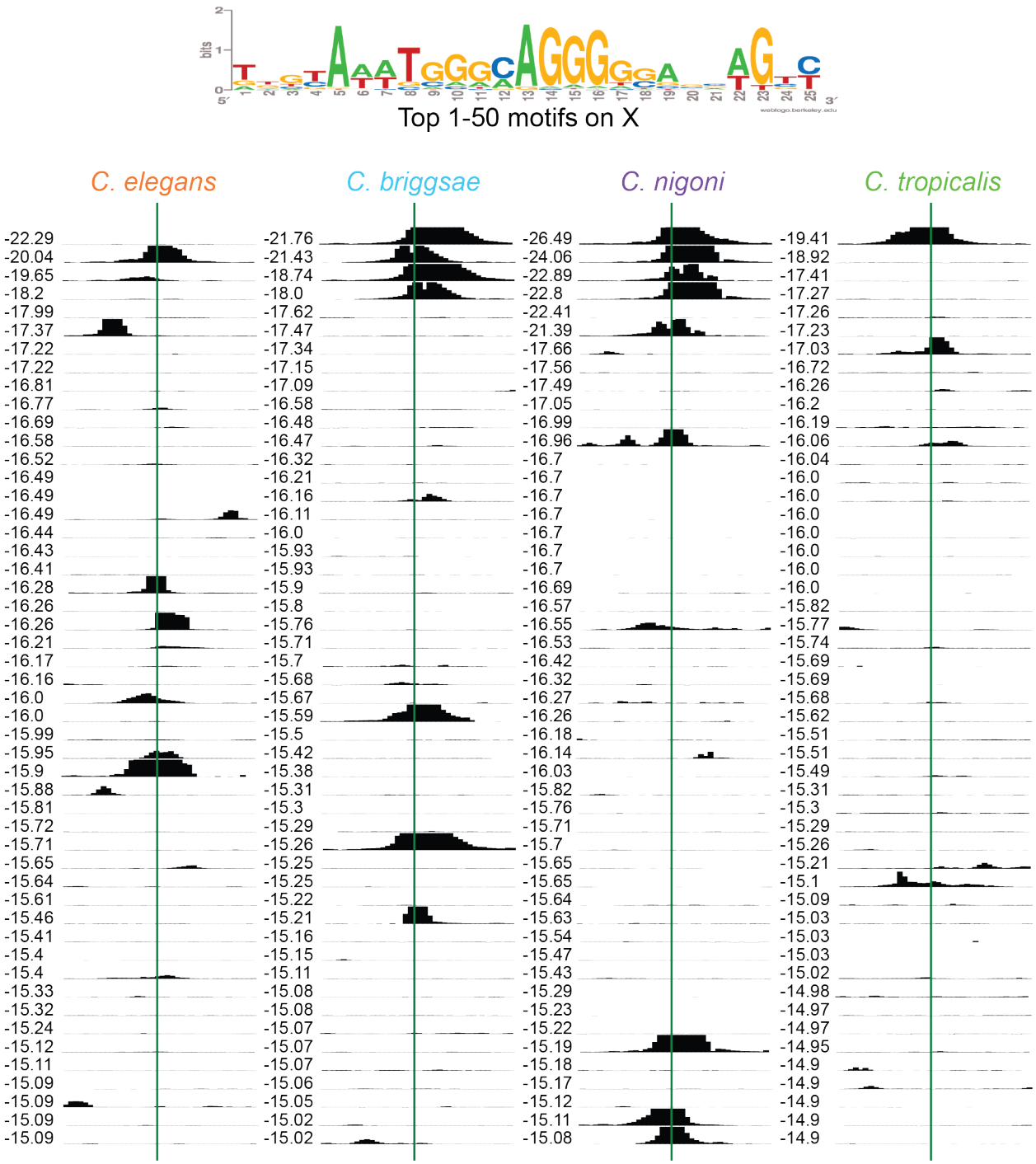


(c) **The DCC is also bound at some of the top 51-100 average of 8 MEX-like motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 average of 8 MEX-like motifs on the X chromosome in four species.

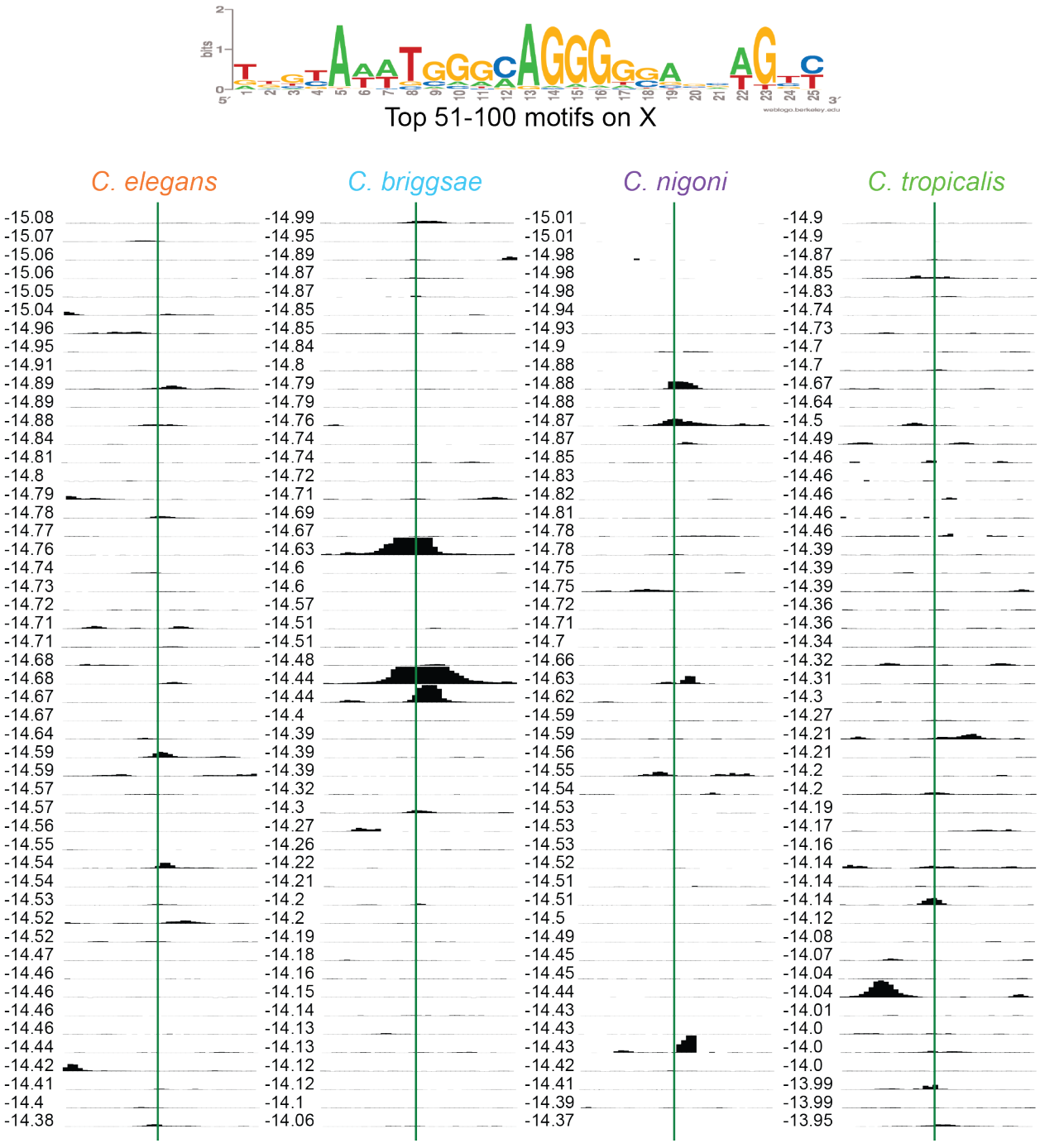
Figure G.21: Averaged 8 *C. briggsae* and *C. nigoni* MEX-like motifs, full length a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The average of 8 MEX-like motifs (extended) is X-enriched in *C. elegans*, *C. briggsae*, and *C. nigoni* and peak-enriched in *C. briggsae*, *C. nigoni*, and *C. tropicalis*.** The average of 8 MEX-like motifs (extended) is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

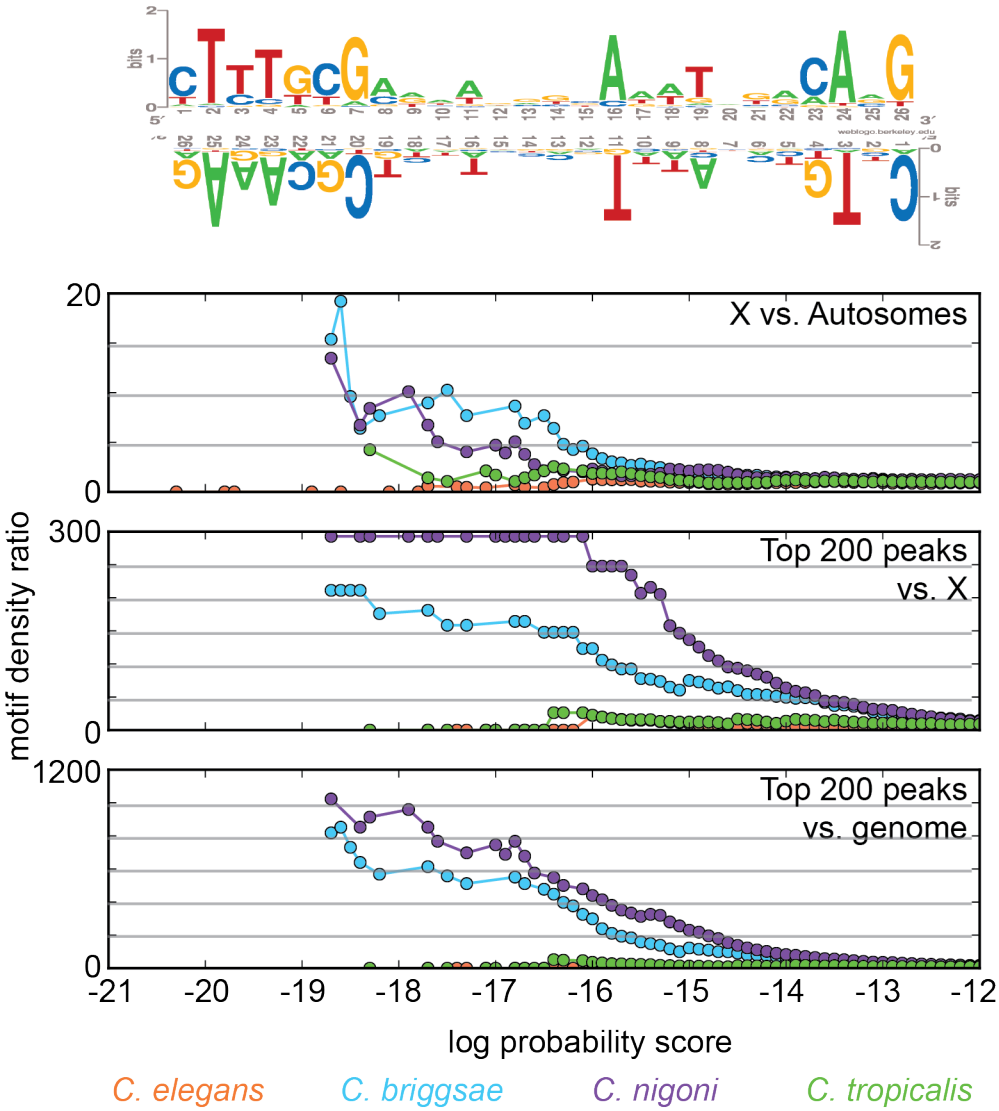


(b) **The DCC is bound at many of the top 50 average of 8 MEX-like motifs (extended) on the X chromosome.** The ChIP-seq signal is plotted at the top 50 average of 8 MEX-like motifs (extended) on the X chromosome in four species.

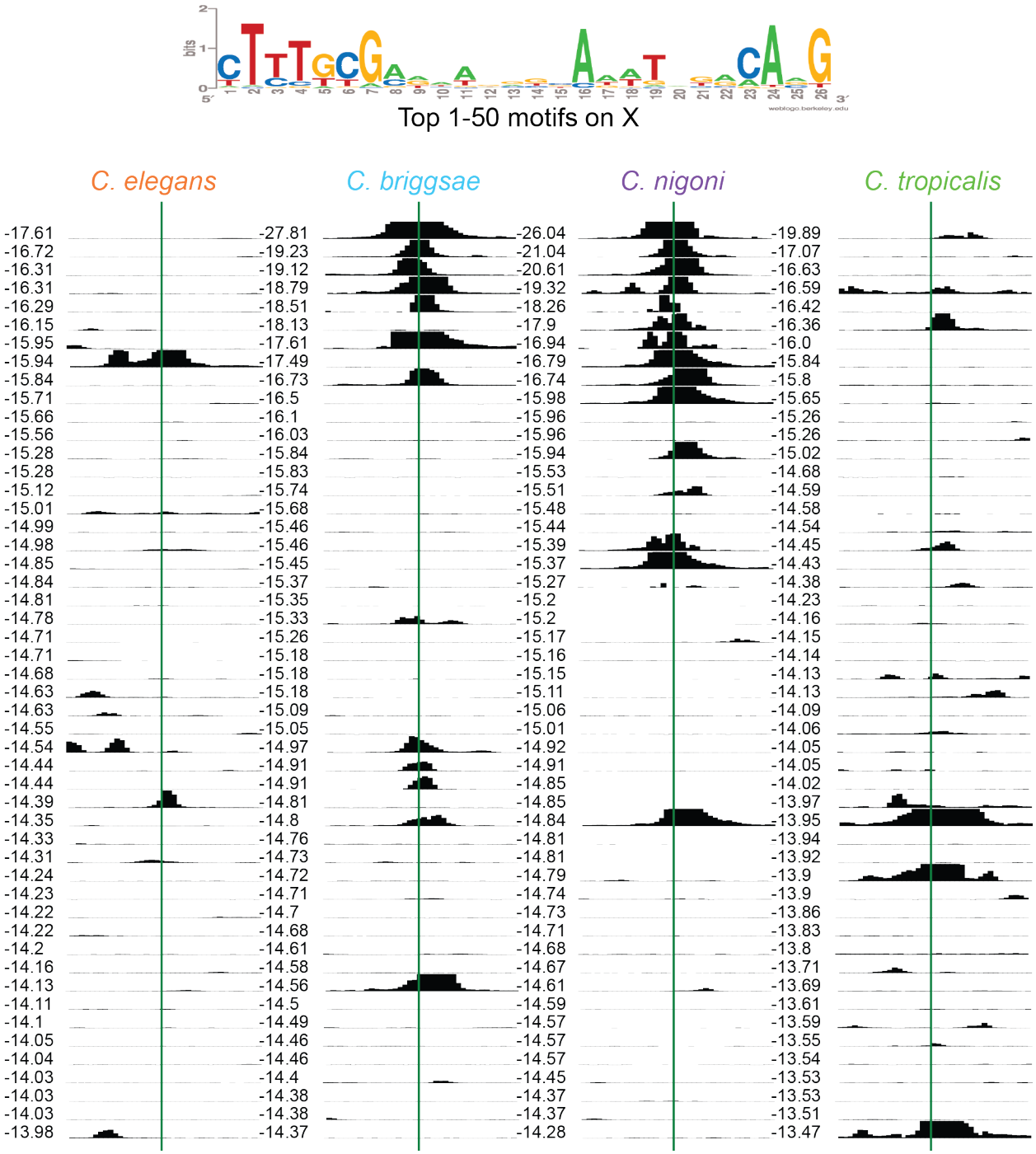


(c) **The DCC is also bound at some of the top 51-100 average of 8 MEX-like motifs (extended) on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 average of 8 MEX-like motifs (extended) on the X chromosome in four species.

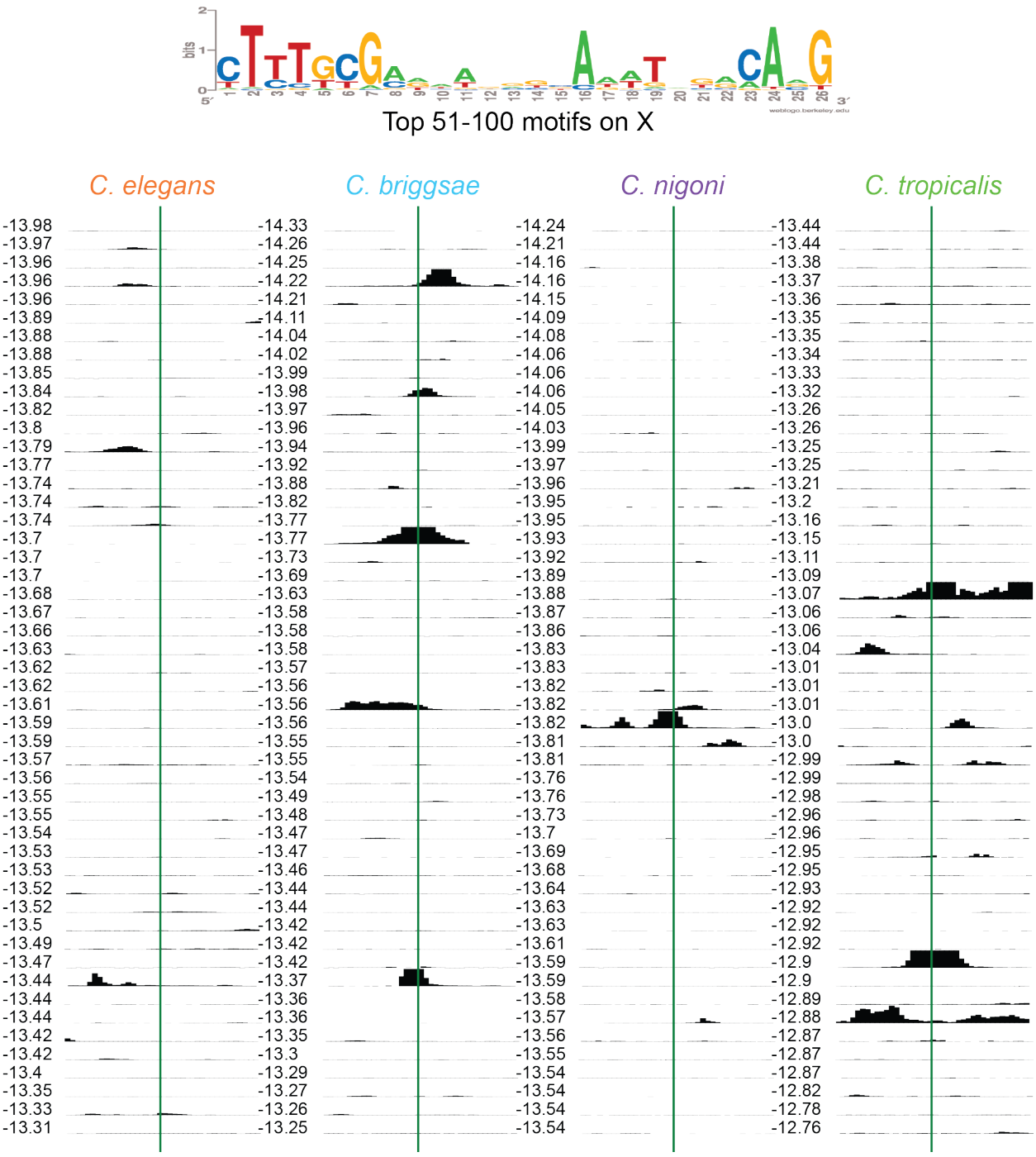
Figure G.22: Averaged Cbr30bp and 5 *C. nigoni* motifs a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The average of 6 Cbr-MEX-II-like motifs is X-enriched and peak-enriched in *C. nigoni* and *C. briggsae*.** The average of 6 Cbr-MEX-II-like motifs is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

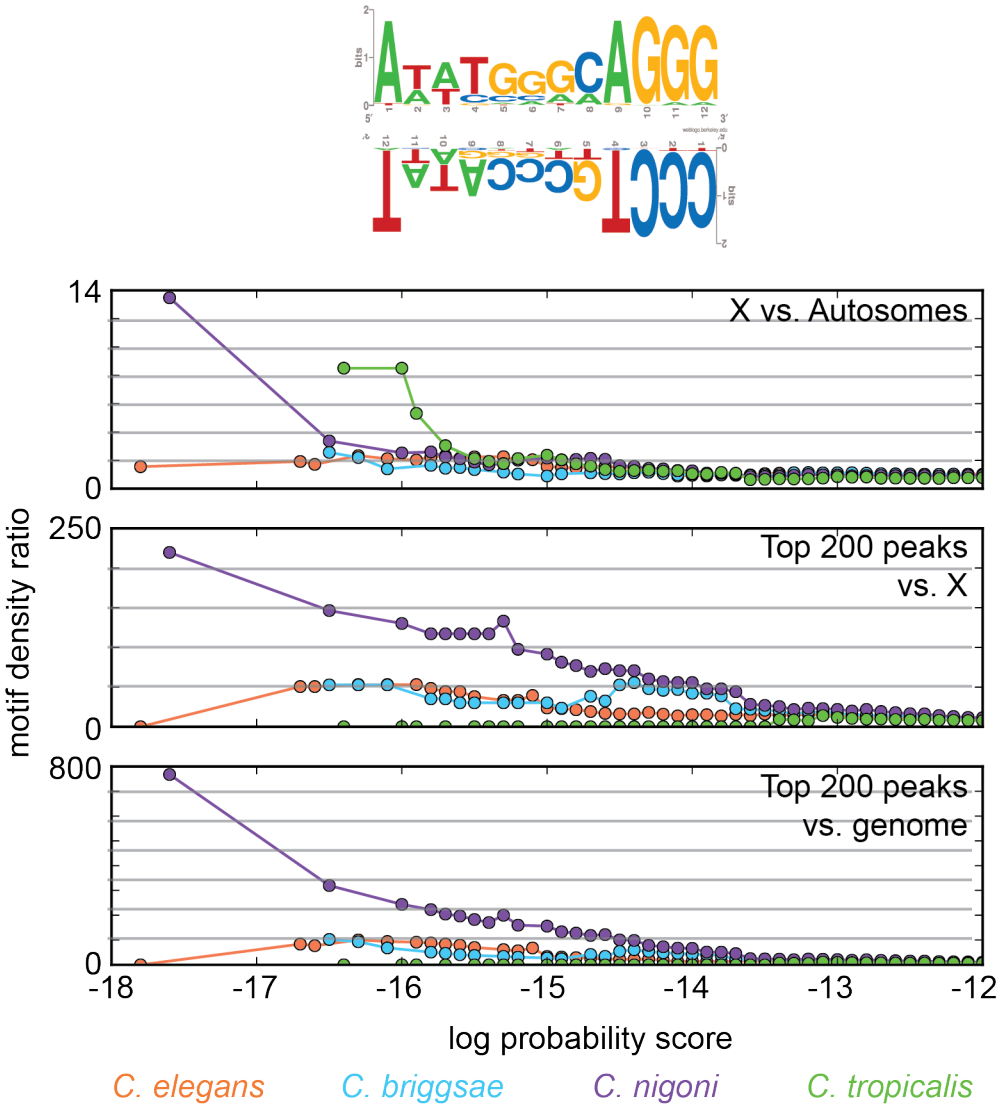


(b) **The DCC is bound at the strongest average of 6 Cbr-MEX-II-like motifs on the X chromosome in *C. briggsae* and *C. nigoni*.** The ChIP-seq signal is plotted at the top 50 average of 6 Cbr-MEX-II-like motifs on the X chromosome in four species.

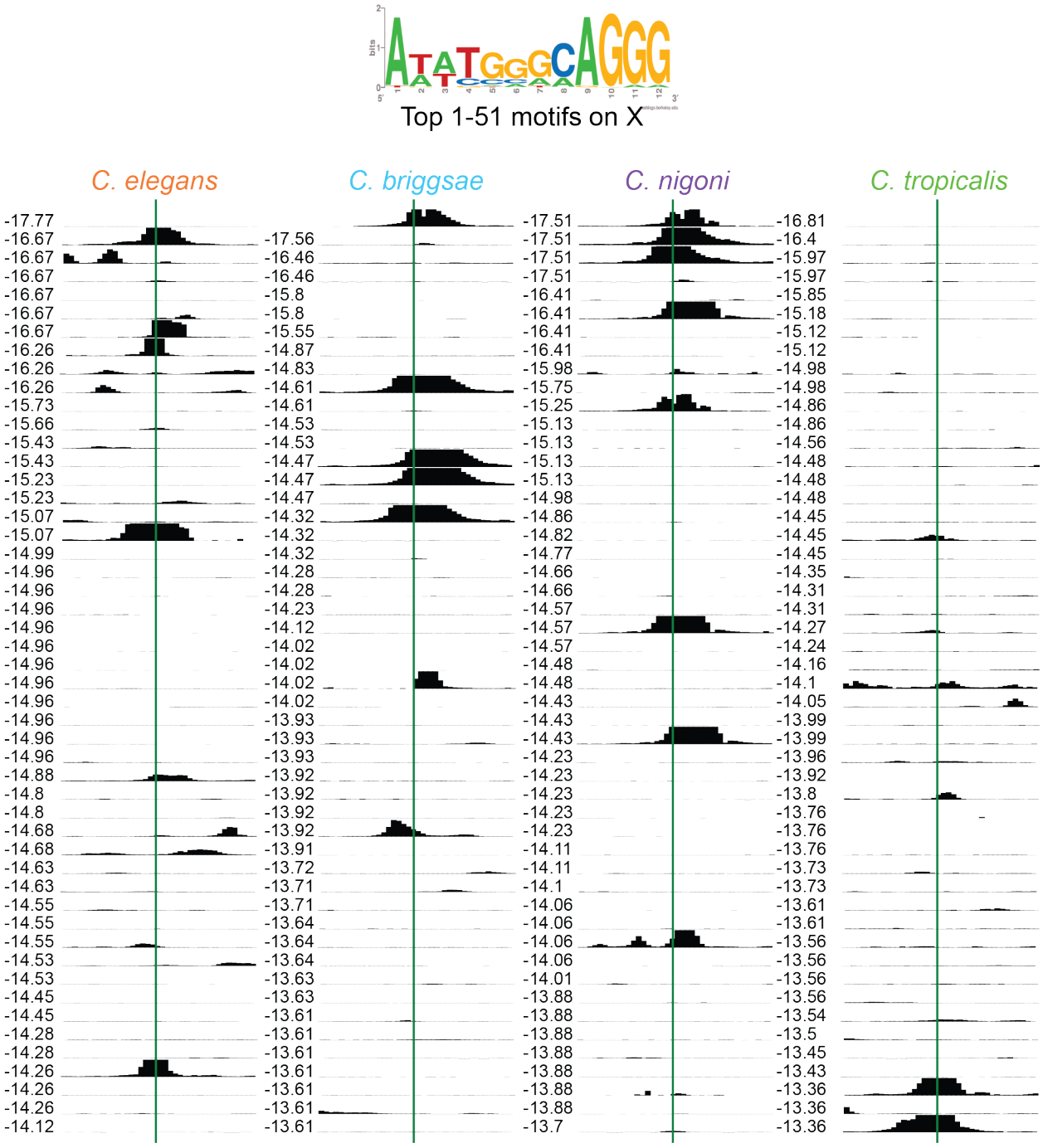


(c) **The DCC is also bound at some of the top 51-100 average of 6 Cbr-MEX-II-like motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 average of 6 Cbr-MEX-II-like motifs on the X chromosome in four species.

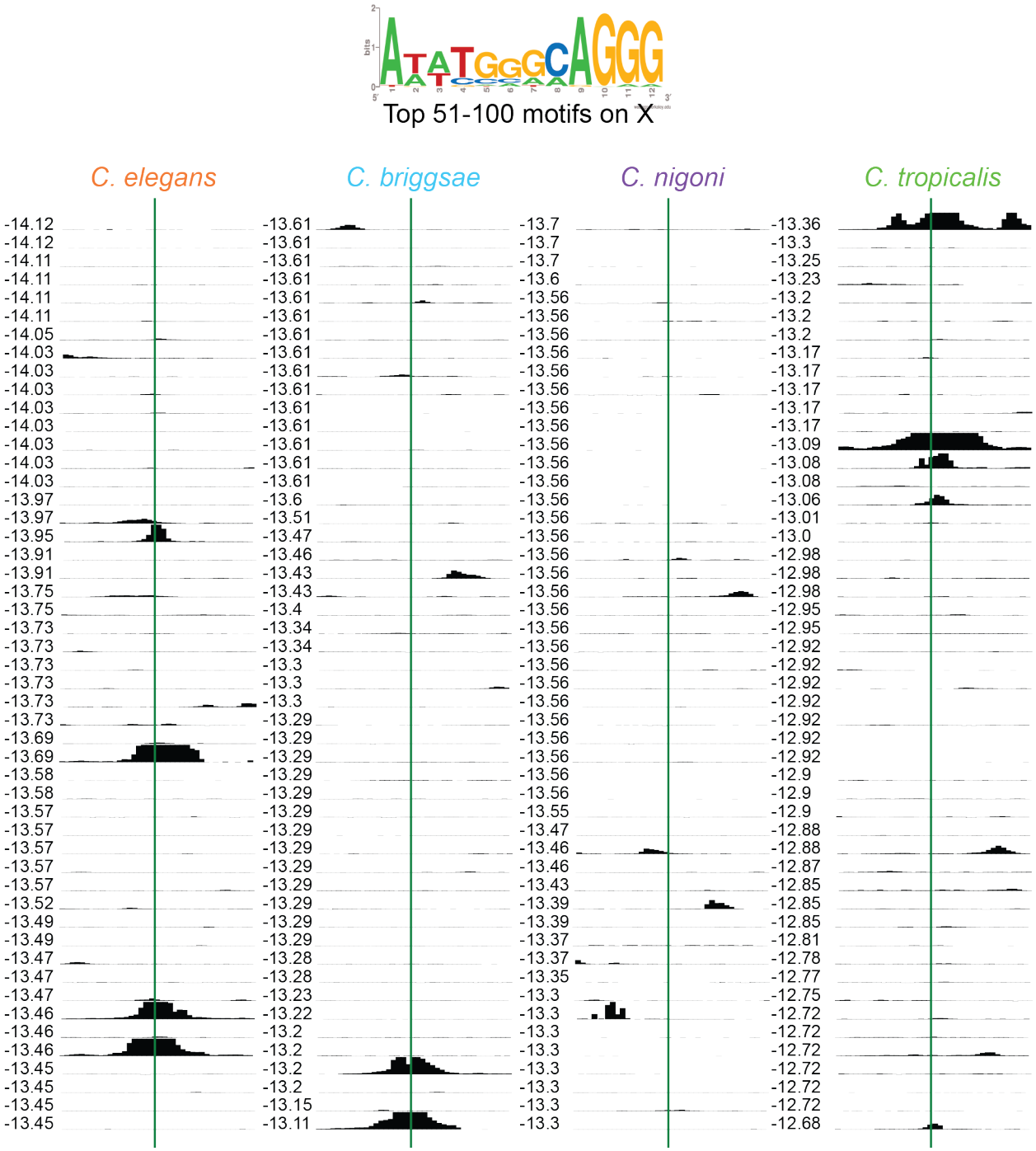
Figure G.23: Averaged 4 CbrMEX-like motifs a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The average of 4 Cbr-MEX-like motifs is X-enriched in *C. nigoni* and *C. tropicalis* and peak-enriched in *C. nigoni*.** The average of 4 Cbr-MEX-like motifs is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.



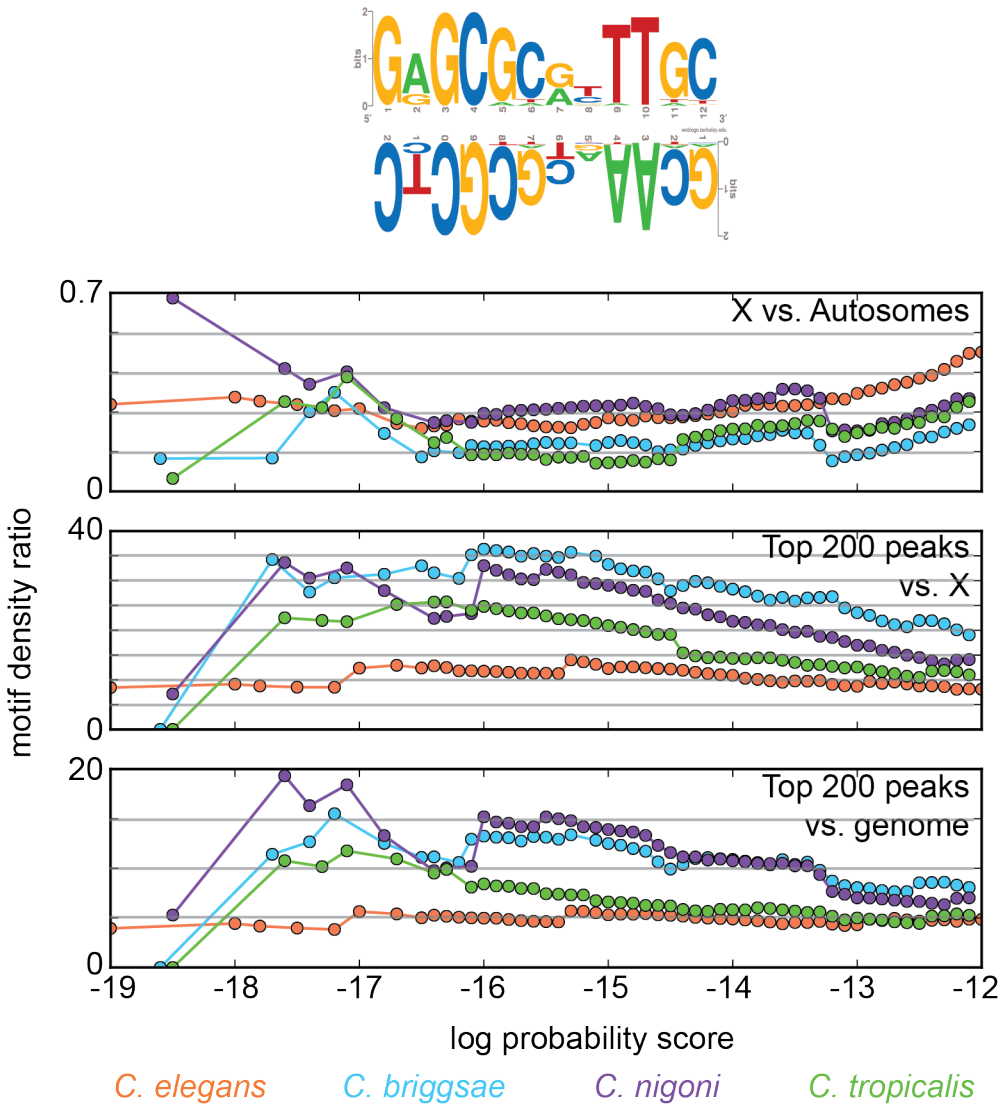
(b) **The DCC is bound at many of the top 50 average of 4 Cbr-MEX-like motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 50 average of 4 Cbr-MEX-like motifs on the X chromosome in four species.



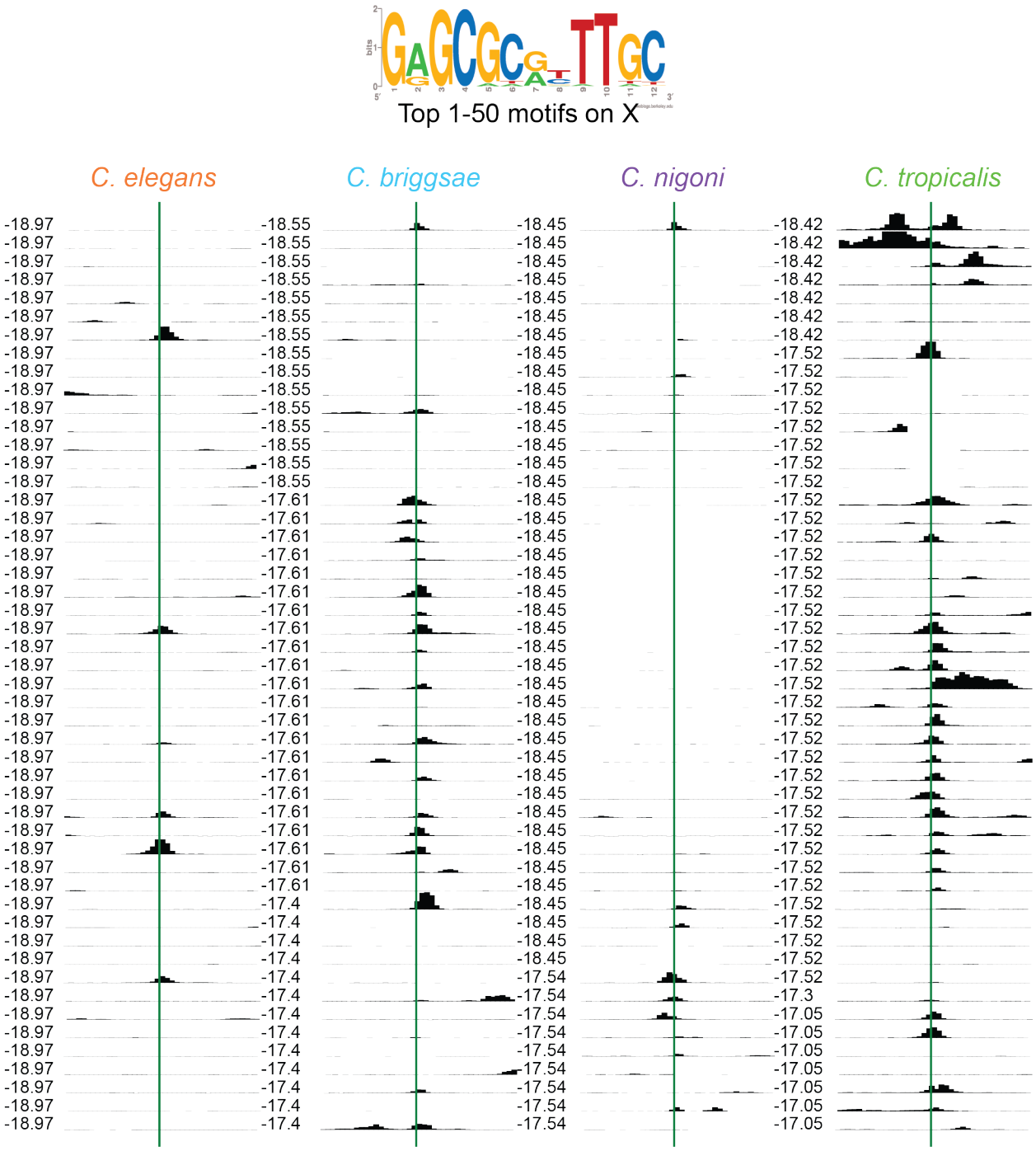
(c) **The DCC is also bound at some of the top 51-100 average of 4 Cbr-MEX-like motifs on the X chromosome.** The ChIP-seq signal is plotted at the top 51-100 average of 4 Cbr-MEX-like motifs on the X chromosome in four species.

[averaged 4 CbrMEX-like]averaged 4 CbrMEX-like

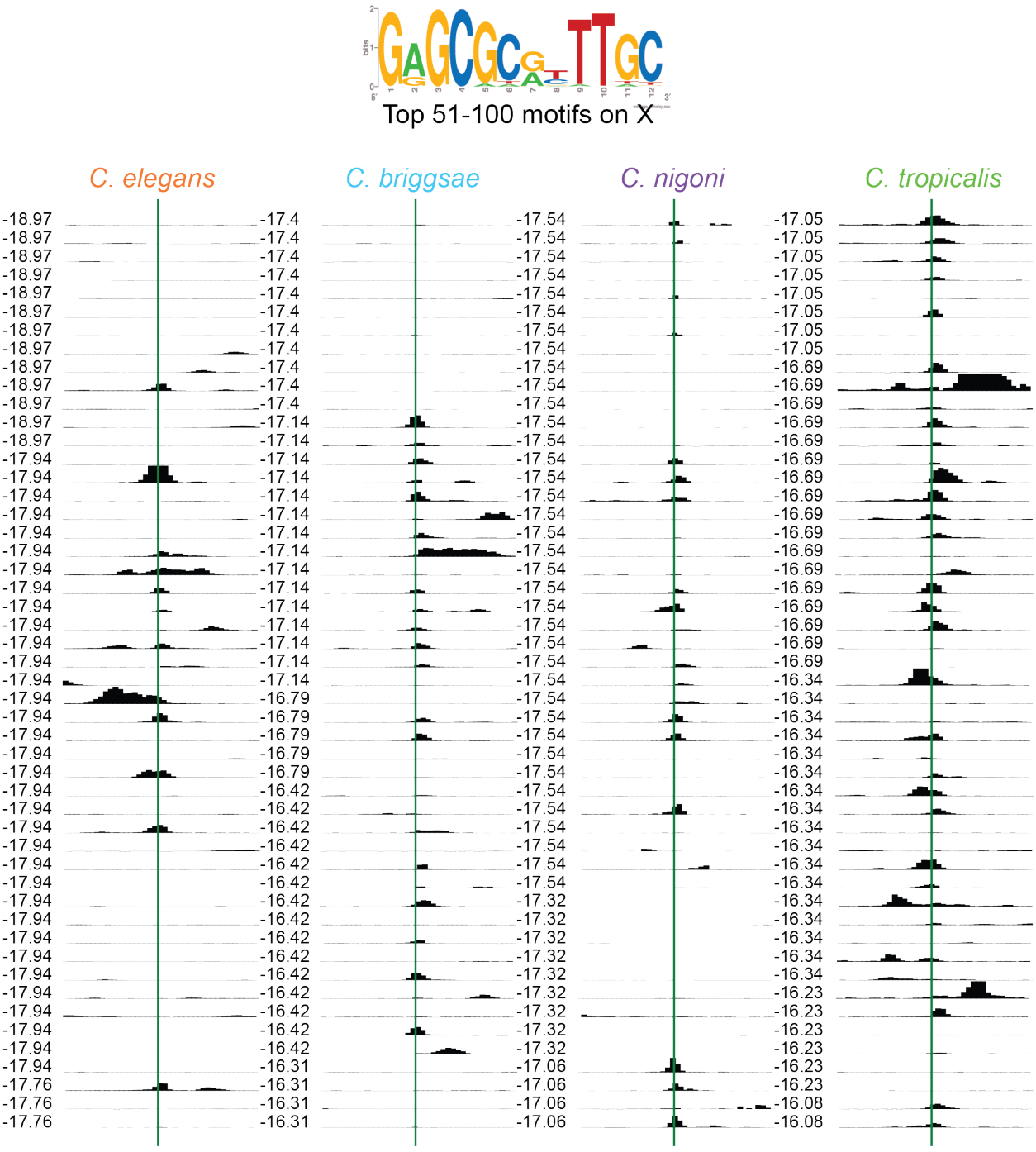
Figure G.24: Averaged 3 top600 motifs (core) a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The average of 3 "top 600 core" motifs is not X-enriched or peak-enriched.** The average of 3 "top 600 core" motifs is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.

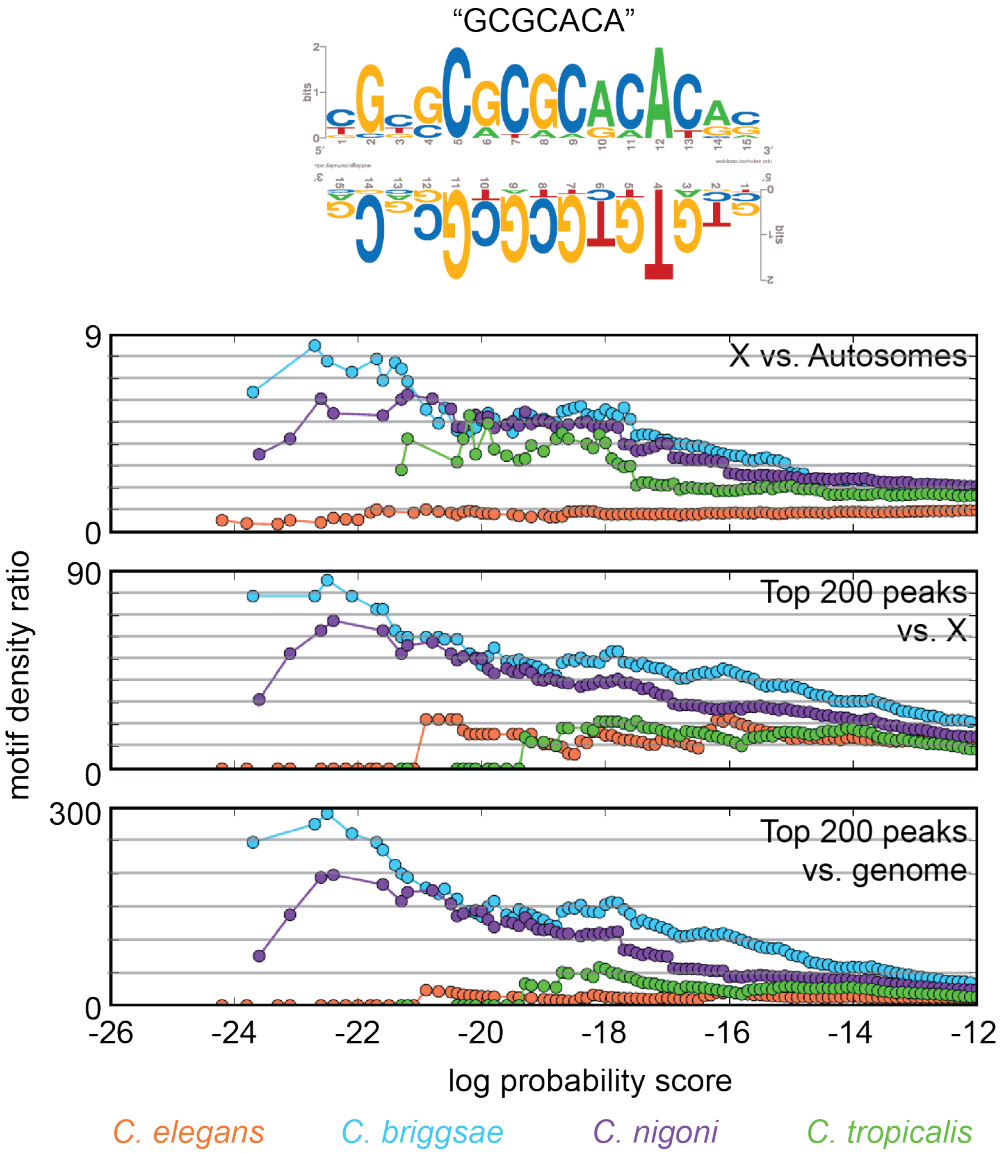


(b) **The DCC is bound at most of the top 50 average of 3 "top 600 core" motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 50 "top 600 core" motifs on the X chromosome in four species.

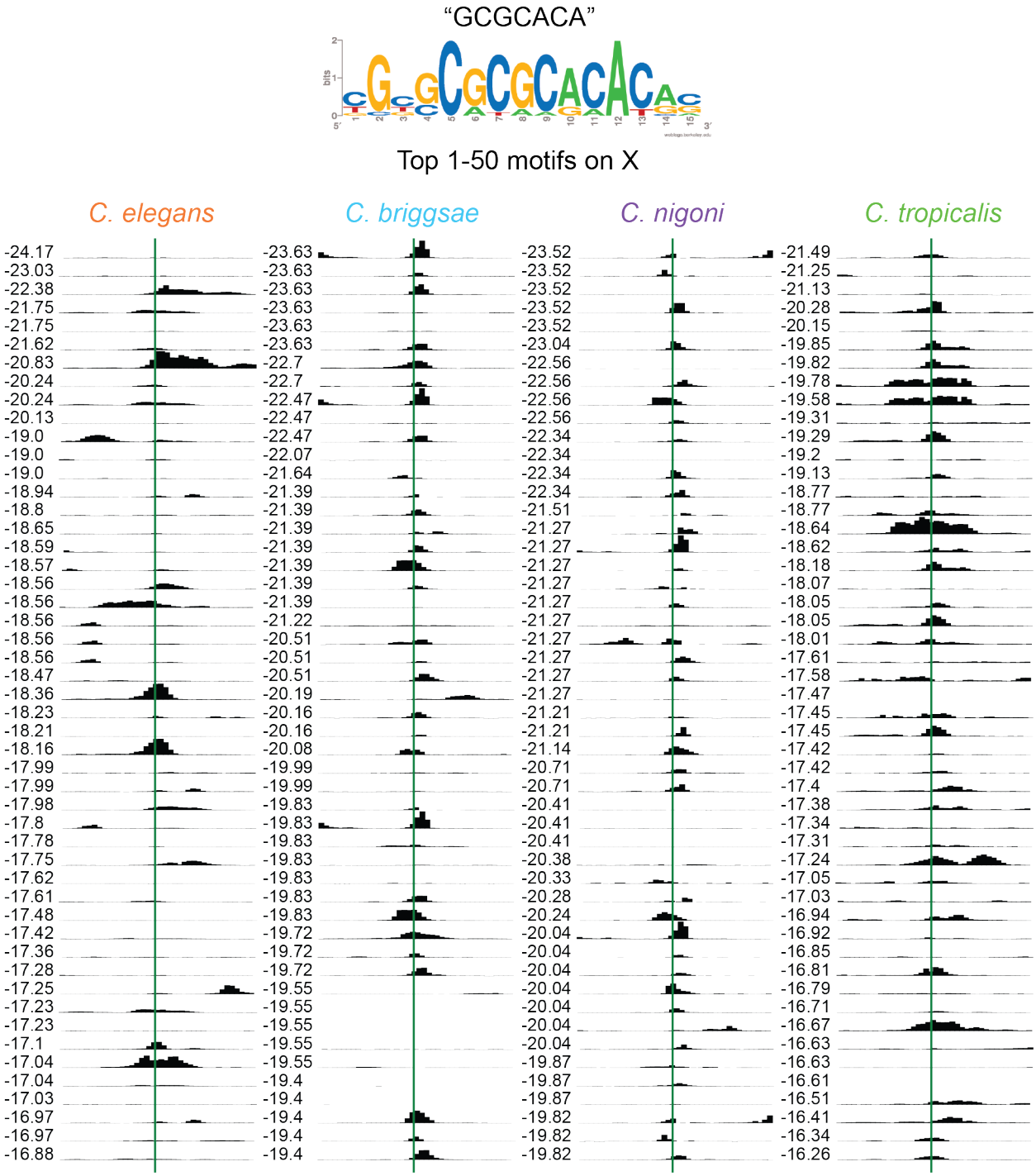


(c) **The DCC is also bound at most of the top 51-100 average of 3 "top 600 core" motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 51-100 "top 600 core" motifs on the X chromosome in four species.

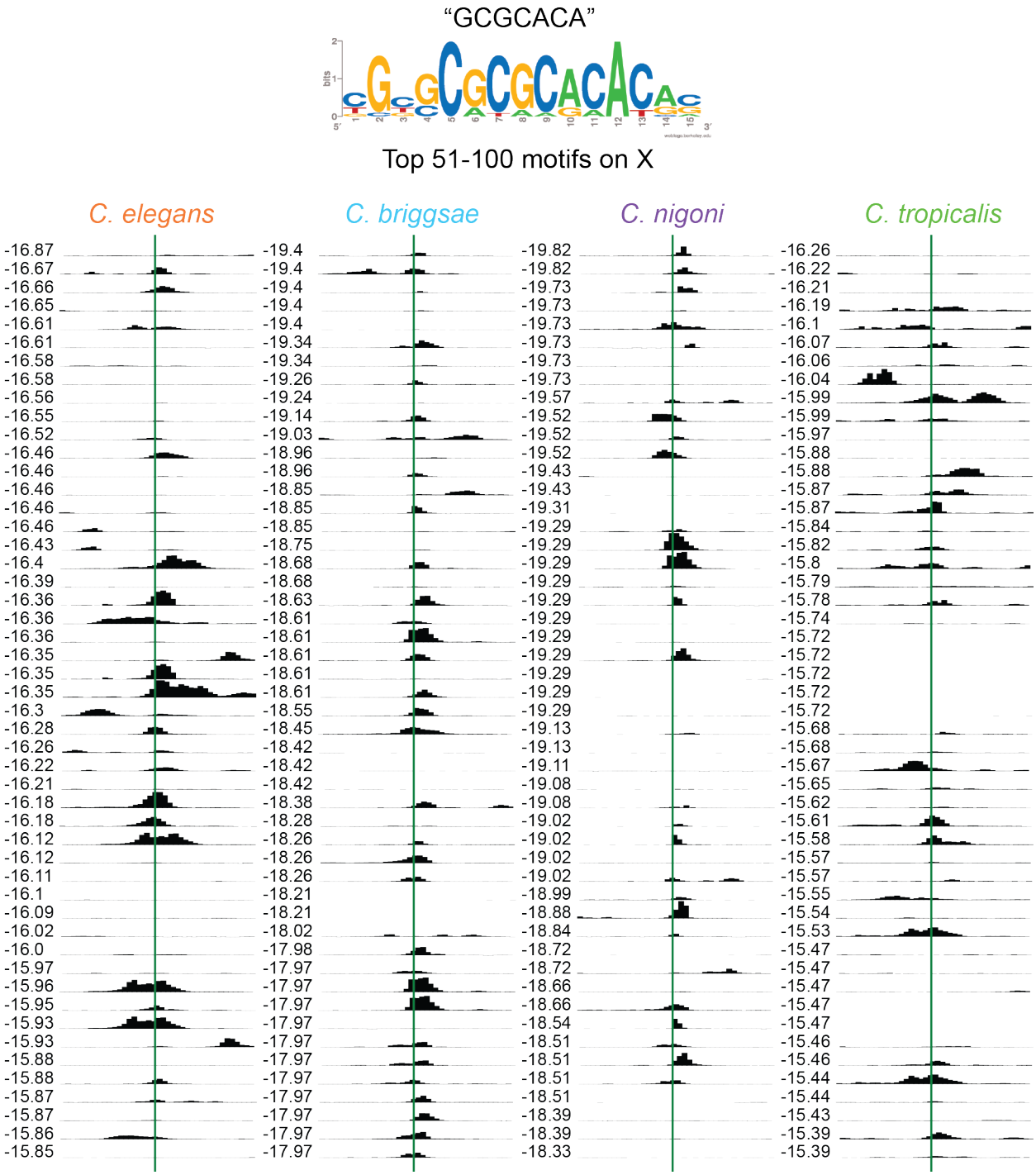
Figure G.25: CS179Cni15-1 a) X-enrichment, peak enrichment, and b-c) ChIP-seq signal at top 100 motifs on the X chromosome.



(a) **The CS179Cni15-1 is X-enriched in *C. briggsae*, *C. nigoni*, and *C. tropicalis* and peak-enriched in *C. briggsae* and *C. nigoni*.** The CS179Cni15-1 is shown in both orientations. The three plots show cumulative motif density ratios for three comparisons: X vs. autosomes, the top 200 peaks vs. the X chromosome, and the top 200 peaks vs. the genome. The motif score is represented as the natural log of the probability a sequence matches the consensus matrix, given the overall GC content of the genome.



(b) **The DCC is bound at most of the top 50 CS179Cni15-1 motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 50 CS179Cni15-1 motifs on the X chromosome in four species.



(c) **The DCC is also bound at most of the top 51-100 CS179Cni15-1 motifs on the X chromosome in each species.** The ChIP-seq signal is plotted at the top 51-100 CS179Cni15-1 motifs on the X chromosome in four species.

Appendix H

Pairwise comparison of ChIP-seq peak calls across libraries

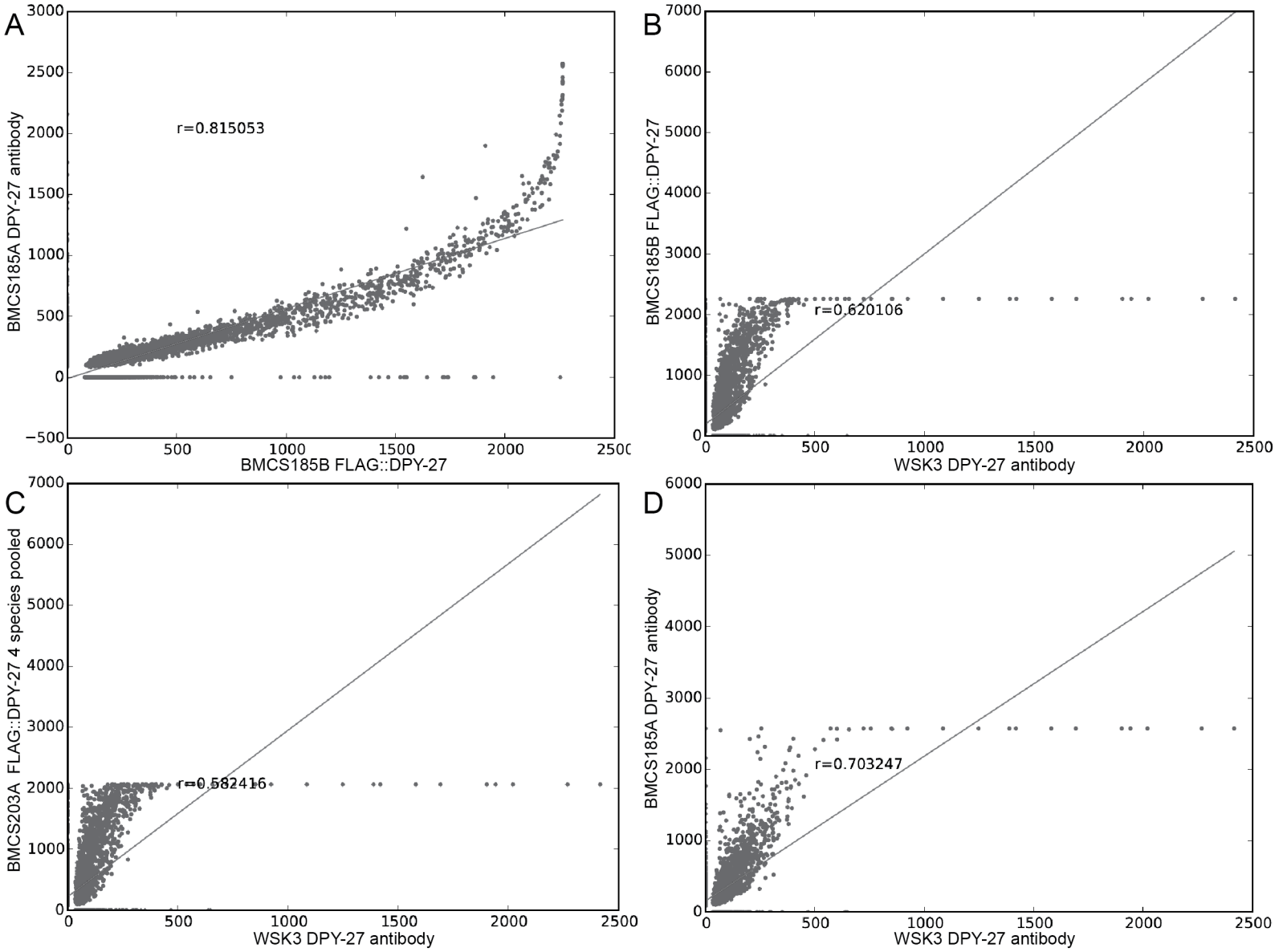


Figure H.1: ***C. elegans*** **pairwise ChIP-seq peak comparisons**

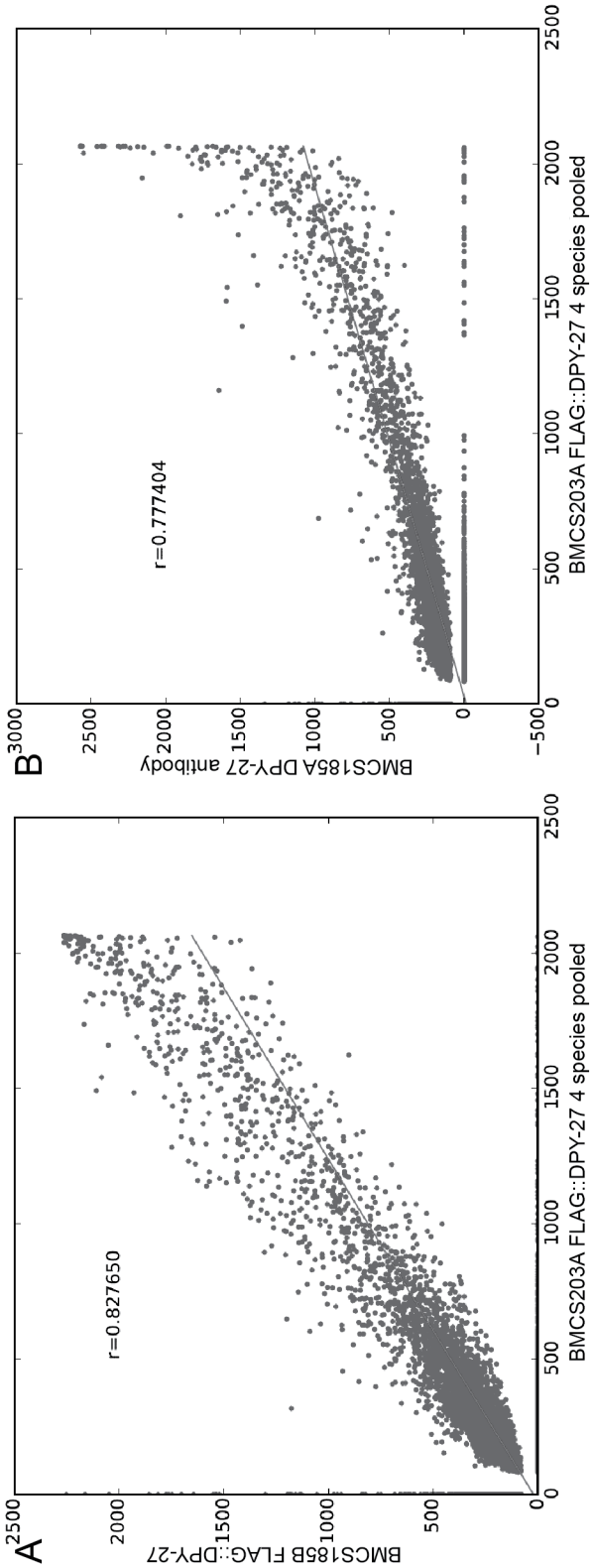


Figure H.2: ***C. elegans*** **pairwise ChIP-seq peak comparisons**

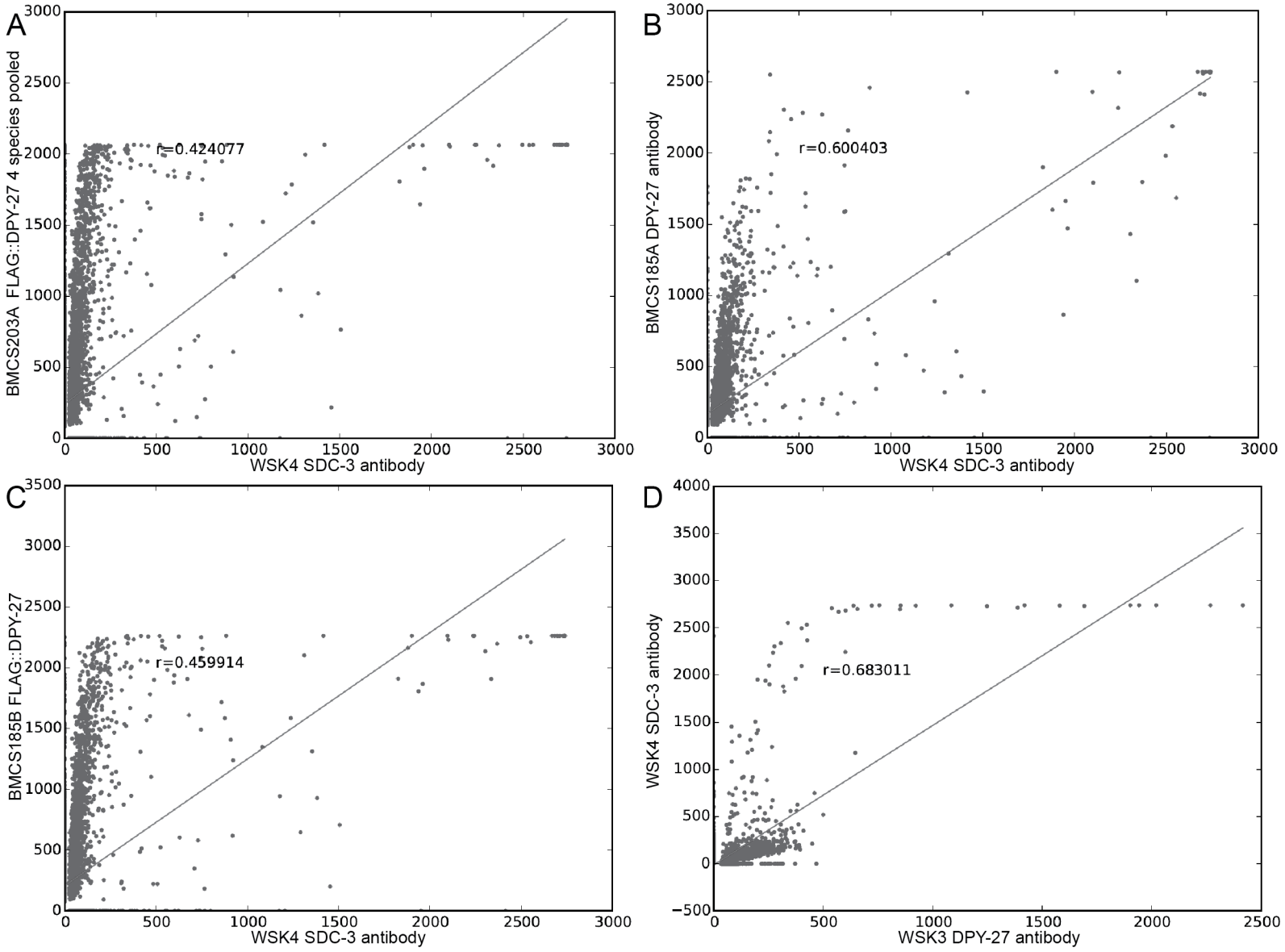


Figure H.3: ***C. elegans*** **pairwise ChIP-seq peak comparisons**

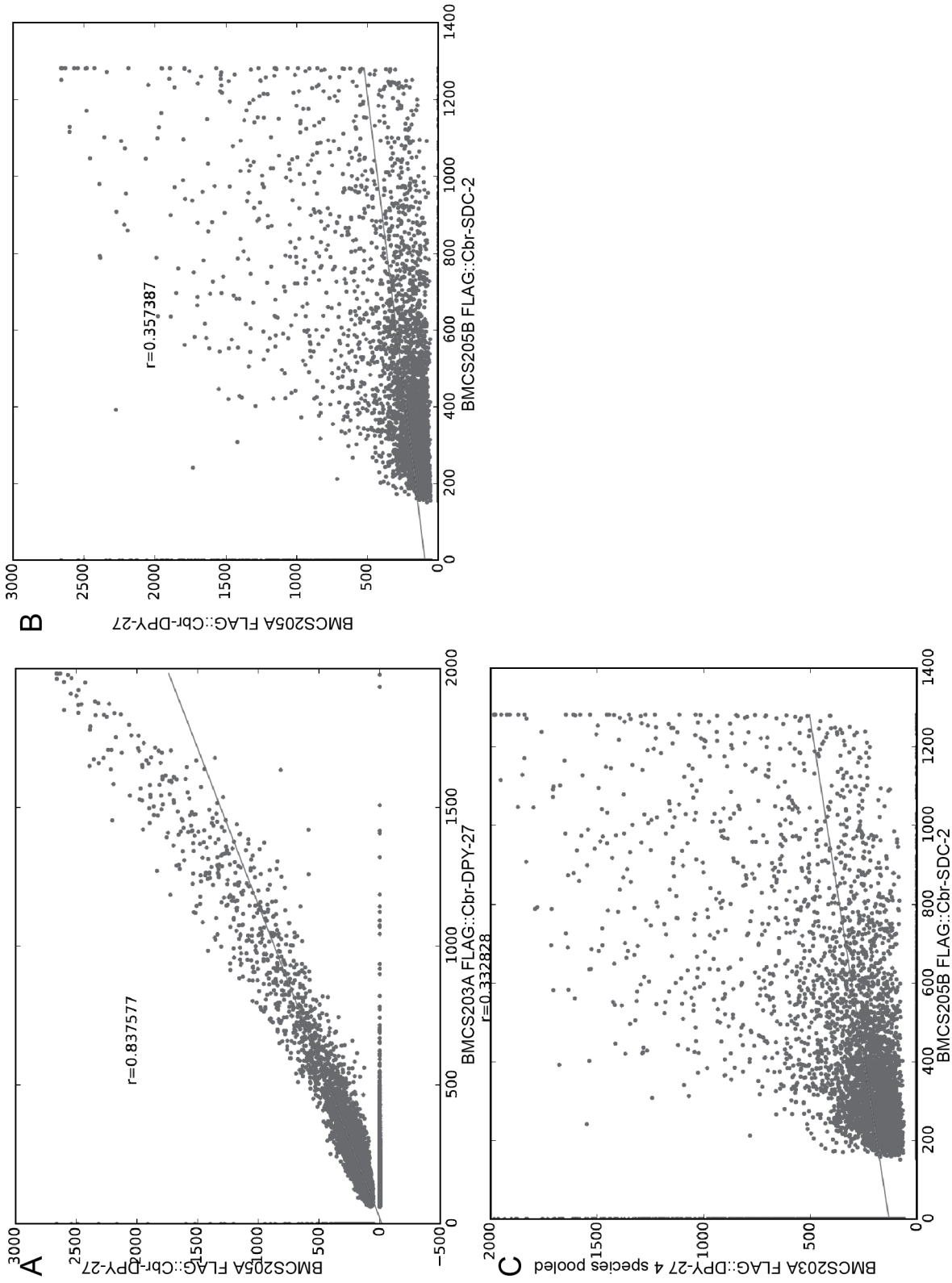


Figure H.4: ***C. briggsae*** **pairwise ChIP-seq peak comparisons**

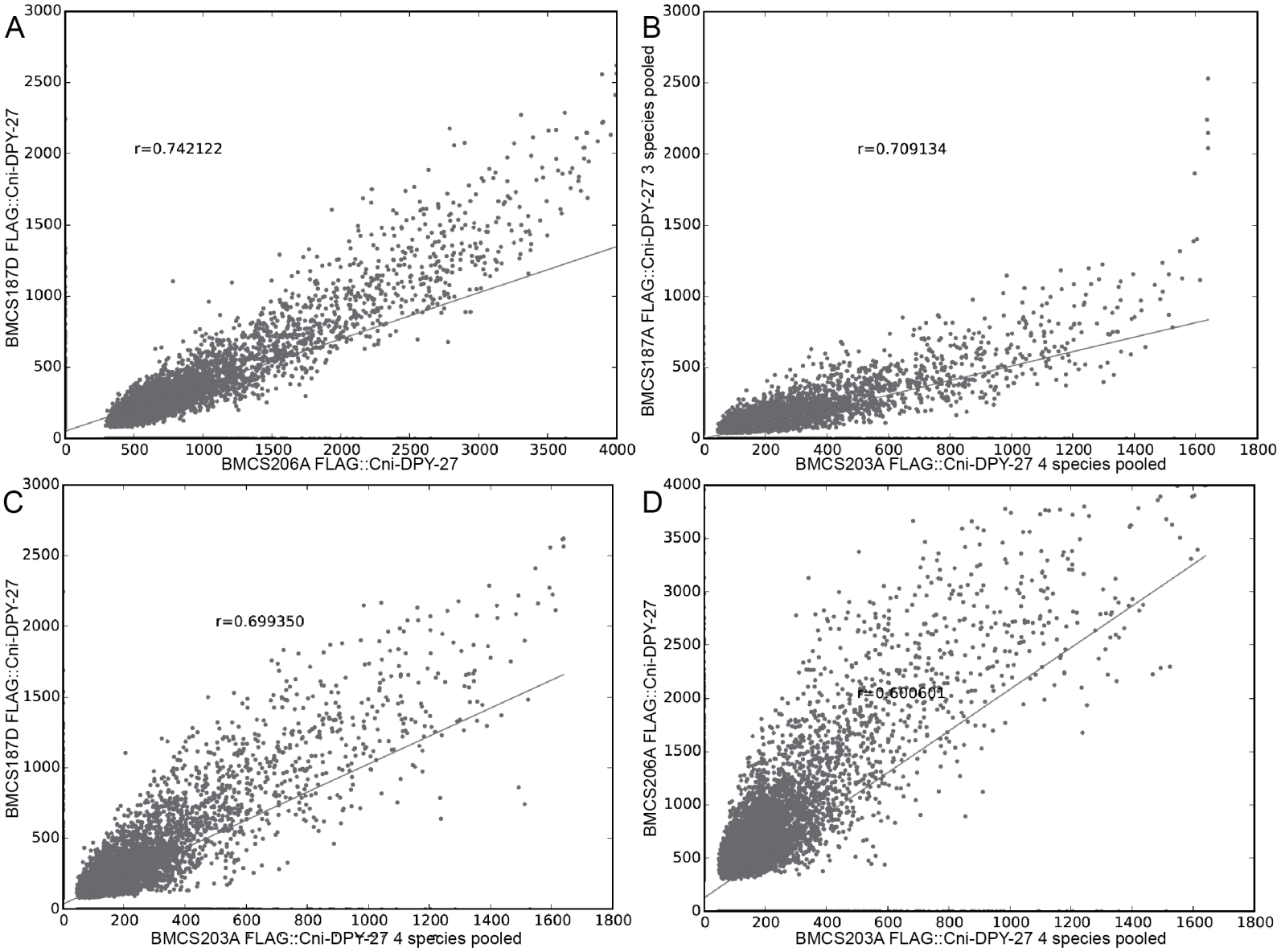


Figure H.5: ***C. nigoni*** **pairwise ChIP-seq peak comparisons**

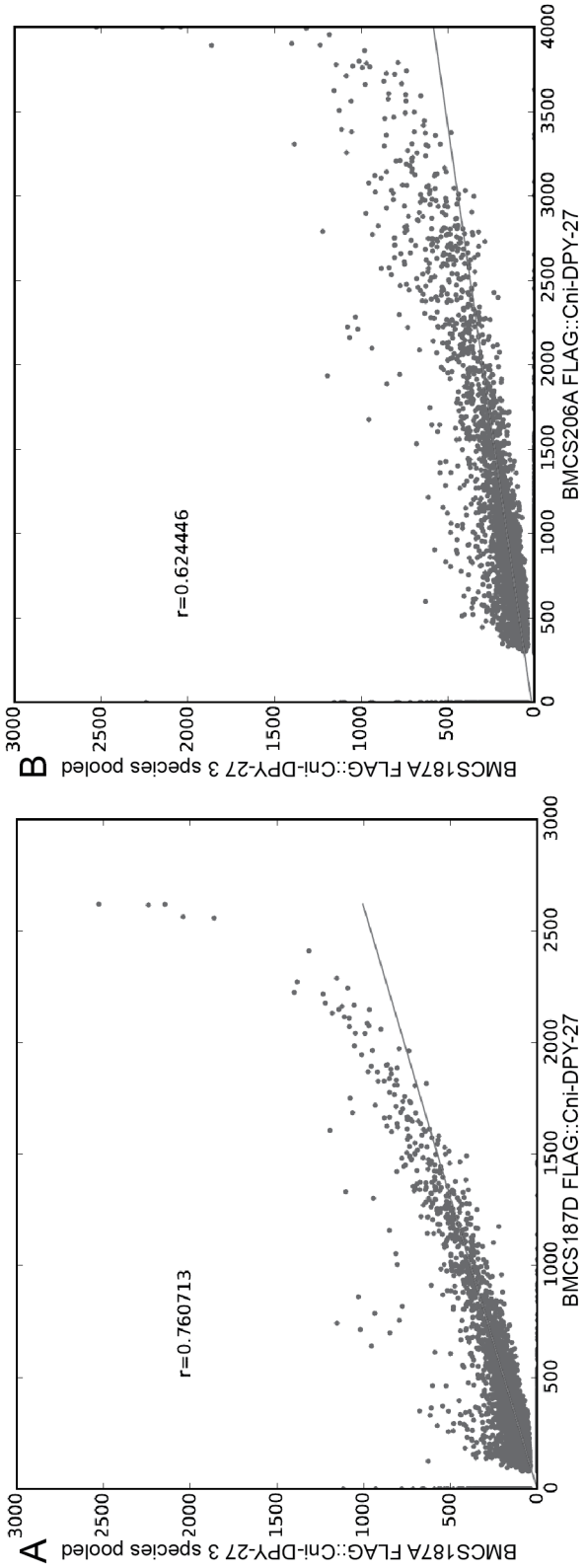


Figure H.6: ***C. nigoni*** **pairwise ChIP-seq peak comparisons**

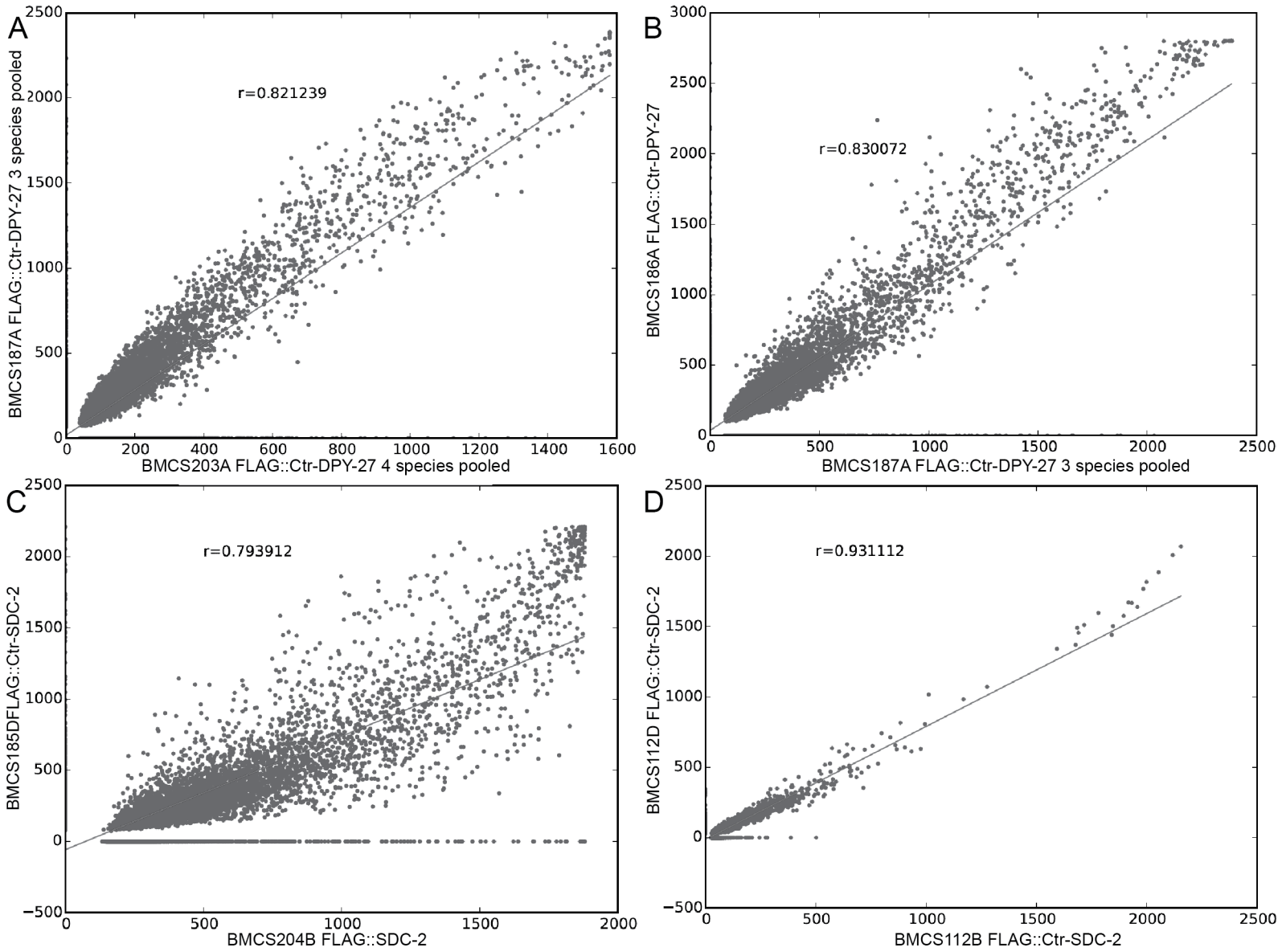


Figure H.7: ***C. tropicalis*** **pairwise ChIP-seq peak comparisons**

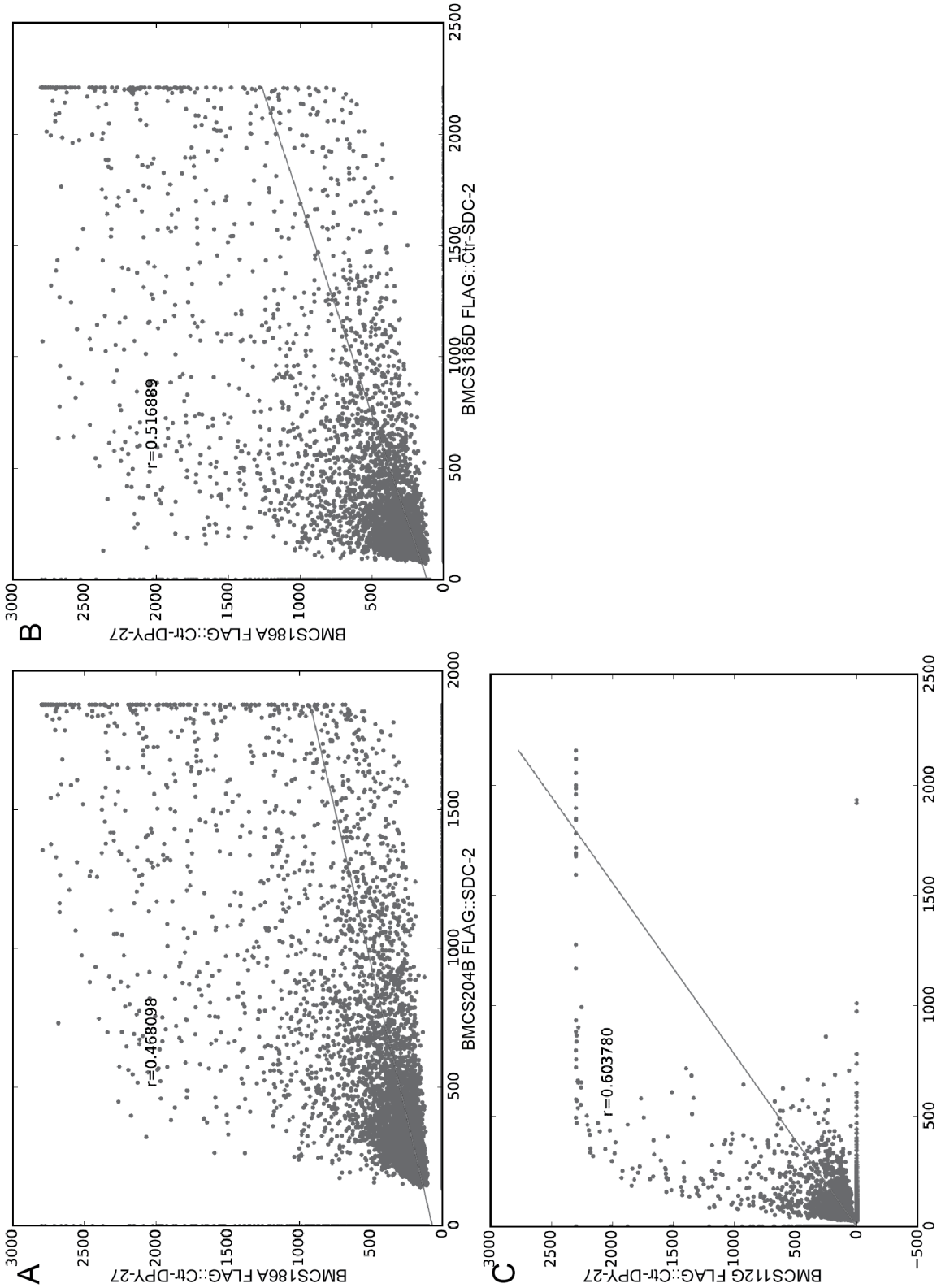


Figure H.8: ***C. tropicalis*** **pairwise ChIP-seq peak comparisons**

Appendix I

Position weight matrices

Cel-MEX

A	C	G	T
10	0	0	21
2	25	4	0
0	0	31	0
2	29	0	0
0	0	31	0
13	16	2	0
31	0	0	0
5	0	26	0
1	0	30	0
0	0	31	0
21	0	4	6
2	8	21	0

Cel-MEX-II

A	C	G	T
4	7	0	2
1	8	2	2
0	12	1	0
0	7	1	5
0	1	0	12
0	2	7	4
0	5	4	4
4	7	0	2
1	9	1	2
5	5	1	2
8	0	0	5
7	1	0	5
1	1	1	10
7	0	0	6
7	6	0	0
4	3	1	5
2	3	2	6
1	0	2	10
0	1	0	12
8	3	2	0
0	0	2	11
0	13	0	0
0	13	0	0
13	0	0	0
1	12	0	0
7	2	4	0

Cbr-MEX

A	C	G	T
0	0.263158	0.210526	0.526316
1	0	0	0
0.315789	0	0	0.684211
0.631579	0	0	0.368421
0	0.263158	0.157895	0.578947
0	0.263158	0.736842	0
0.210526	0.157895	0.578947	0.052632
0.368421	0	0.578947	0.052632
0.157895	0.789474	0.052632	0
1	0	0	0
0.052632	0	0.947368	0
0.052632	0	0.947368	0
0	0	1	0

Cbr-MEX-II			
A	C	G	T
0.230769	0.384615	0.384615	0
0.153846	0.153846	0.384615	0.307692
0.076923	0.076923	0.461538	0.384615
0	0.923077	0	0.076923
0.076923	0.307692	0.230769	0.384615
0.076923	0	0	0.923077
0	0	0.769231	0.230769
0	0.230769	0.076923	0.692308
0.307692	0.538462	0	0.153846
0.076923	0.307692	0	0.615385
0.846154	0.153846	0	0
0.076923	0.153846	0.076923	0.692308
0.230769	0.307692	0	0.461538
0	0	0.153846	0.846154
0.153846	0.307692	0.461538	0.076923
0.153846	0.615385	0.230769	0
0.153846	0.769231	0.076923	0
0.153846	0.384615	0.461538	0
0.307692	0.153846	0.076923	0.461538
0.153846	0.153846	0.307692	0.384615
0.076923	0.461538	0	0.461538
0	0	0.307692	0.692308
0	0.923077	0	0.076923
0.384615	0	0.615385	0
0.230769	0.692308	0	0.076923
0.692308	0.153846	0.153846	0
0.538462	0	0.461538	0
1	0	0	0
0.153846	0.076923	0.769231	0
0.461538	0	0.076923	0.461538

Cni-MEX			
A	C	G	T
0.147059	0.029412	0.823529	0
0.676471	0.088235	0.176471	0.058824
0.411765	0.117647	0.411765	0.058824
0.823529	0	0.176471	0
0.264706	0.529412	0.088235	0.117647
0.470588	0.029412	0.411765	0.088235
0.176471	0.588235	0.088235	0.147059
0.529412	0.029412	0.264706	0.176471
0.029412	0.088235	0.882353	0
0.941176	0.029412	0.029412	0
0.411765	0.117647	0.470588	0
0.911765	0.029412	0.058824	0
0.970588	0	0	0.029412
0.794118	0	0.205882	0
0.617647	0.382353	0	0
0.529412	0.088235	0.382353	0
0.764706	0.029412	0.176471	0.029412
0.470588	0.117647	0.323529	0.088235
0.294118	0	0.588235	0.117647

Cni-MEX-II			
A	C	G	T
0.6	0	0	0.4
0.1	0.8	0	0.1
0	0	0	1
0	0.2	0	0.8
0	0	0	1
0	0	0.5	0.5
0	0.8	0.1	0.1
0.1	0	0.9	0
0.8	0.1	0.1	0
0.4	0.1	0.2	0.3
0.4	0	0.3	0.3
0.7	0.1	0.1	0.1
0.4	0.1	0.5	0
0.6	0.1	0.3	0
0.2	0.3	0.1	0.4
0.1	0.1	0.5	0.3
1	0	0	0
0.7	0	0.2	0.1
0.3	0	0.1	0.6
0.1	0.1	0.2	0.6
0.2	0.6	0.1	0.1
0.1	0.2	0.2	0.5
0.5	0.3	0.1	0.1
0.4	0.6	0	0
1	0	0	0
0.3	0.4	0.3	0
0	0	0.8	0.2
0.2	0	0.5	0.3

Ctr-MEX			
A	C	G	T
0.230769	0	0.615385	0.153846
0	0.230769	0.769231	0
1	0	0	0
0.307692	0	0	0.692308
0.846154	0	0	0.153846
0	0	0.307692	0.692308
0.307692	0	0.692308	0
0	0	1	0
0.153846	0	0.615385	0.230769
0	0.923077	0.076923	0
1	0	0	0
0	0	0.692308	0.307692
0.538462	0	0	0.461538
0.076923	0	0.923077	0
0.384615	0.384615	0.153846	0.076923
0.153846	0.384615	0.461538	0
0.384615	0.384615	0.230769	0
0.153846	0.076923	0.769231	0
0.230769	0.153846	0.461538	0.153846
0.153846	0.153846	0.384615	0.307692
0	0.076923	0.846154	0.076923

"Top 600"			
A	C	G	T
0.053333333	0.04	0.46	0.446666667
0.1	0.173333333	0.213333333	0.513333333
0.066666667	0.133333333	0.36	0.44
0.02	0.9	0	0.08
0.633333333	0.053333333	0.286666667	0.026666667
0.493333333	0.033333333	0.386666667	0.086666667
0.02	0.233333333	0.006666667	0.74
0.453333333	0.02	0.52	0.006666667
0	0.006666667	0.993333333	0
0.766666667	0.013333333	0.22	0
0	0	1	0
0	1	0	0
0.053333333	0	0.946666667	0
0.053333333	0.886666667	0	0.06
0.406666667	0.013333333	0.58	0
0.066666667	0.326666667	0.08	0.526666667
0.033333333	0	0.006666667	0.96
0.006666667	0.006666667	0	0.986666667
0.053333333	0	0.886666667	0.06
0.013333333	0.913333333	0.02	0.053333333

CS181 Cel-MEX			
A	C	G	T
0.145833	0.25	0.125	0.479167
0.083333	0.3125	0.333333	0.270833
0.0625	0.208333	0.0625	0.666667
0.020833	0.958333	0	0.020833
0	0.9375	0	0.0625
0	0.979167	0	0.020833
0	0.020833	0	0.979167
0	0.083333	0.625	0.291667
0	0.833333	0	0.166667
0	0.229167	0.770833	0
0.083333	0.875	0.041667	0
0.166667	0.1875	0.520833	0.125
0.708333	0.041667	0	0.25
0.375	0	0.041667	0.583333

Cel-bMEX-13bp			
A	C	G	T
0	0.066666667	0.066666667	0.866666667
0	0.933333333	0	0.066666667
0	0.933333333	0	0.066666667
0	1	0	0
0	0	0	1
0	0.066666667	0.866666667	0.066666667
0	1	0	0
0	0	1	0
0	1	0	0
0.066666667	0.066666667	0.866666667	0
1	0	0	0
0.2	0	0	0.8
0.666666667	0	0.2	0.133333333

Cel-bMEX-31bp			
A	C	G	T
0	0.538461538	0.384615385	0.076923077
0.076923077	0.769230769	0.153846154	0
0.307692308	0.230769231	0.230769231	0.230769231
0.230769231	0.692307692	0.076923077	0
0.538461538	0.384615385	0	0.076923077
0.384615385	0.230769231	0.307692308	0.076923077
0.076923077	0.076923077	0.384615385	0.461538462
0	0.230769231	0.769230769	0
0.230769231	0.615384615	0	0.153846154
0.076923077	0.461538462	0.076923077	0.384615385
0.153846154	0.076923077	0.769230769	0
0.230769231	0.384615385	0	0.384615385
0.230769231	0.384615385	0	0.384615385
0.307692308	0.153846154	0.307692308	0.230769231
0.461538462	0.230769231	0.076923077	0.230769231
0.384615385	0.153846154	0.153846154	0.307692308
0	0.538461538	0.076923077	0.384615385
0.153846154	0.461538462	0.307692308	0.076923077
0	0.384615385	0	0.615384615
0	0.846153846	0	0.153846154
0	0.846153846	0	0.153846154
0.076923077	0.923076923	0	0
0	0.230769231	0	0.769230769
0	0.153846154	0.692307692	0.153846154
0.076923077	0.846153846	0	0.076923077
0.076923077	0.230769231	0.692307692	0
0.230769231	0.769230769	0	0
0.307692308	0.153846154	0.461538462	0.076923077
0.846153846	0.153846154	0	0
0.076923077	0.076923077	0	0.846153846
0.615384615	0.230769231	0.076923077	0.076923077

Cbr20-Cni14-MEME2			
A	C	G	T
0.147059	0.029412	0.823529	0
0.676471	0.088235	0.176471	0.058824
0.411765	0.117647	0.411765	0.058824
0.823529	0	0.176471	0
0.264706	0.529412	0.088235	0.117647
0.470588	0.029412	0.411765	0.088235
0.176471	0.588235	0.088235	0.147059
0.529412	0.029412	0.264706	0.176471
0.029412	0.088235	0.882353	0
0.941176	0.029412	0.029412	0
0.411765	0.117647	0.470588	0
0.911765	0.029412	0.058824	0
0.970588	0	0	0.029412
0.794118	0	0.205882	0
0.617647	0.382353	0	0
0.529412	0.088235	0.382353	0
0.764706	0.029412	0.176471	0.029412
0.470588	0.117647	0.323529	0.088235
0.294118	0	0.588235	0.117647

Cbr20-Cni14-MEME1			
A	C	G	T
0	1	0	0
0.090909	0.409091	0.090909	0.409091
0.136364	0	0	0.863636
0	0.045455	0.727273	0.227273
0	0.363636	0.090909	0.545455
0.227273	0.681818	0	0.090909
0.090909	0.272727	0.045455	0.590909
0.772727	0.090909	0.090909	0.045455
0.090909	0.136364	0.090909	0.681818
0.227273	0.227273	0.045455	0.5
0	0	0.181818	0.818182
0.227273	0.227273	0.454545	0.090909
0.136364	0.590909	0.227273	0.045455
0.136364	0.681818	0.181818	0
0.136364	0.454545	0.363636	0.045455
0.318182	0.136364	0	0.545455
0.181818	0.045455	0.318182	0.454545
0.090909	0.454545	0	0.454545
0	0	0.409091	0.590909
0	0.863636	0	0.136364
0.318182	0	0.681818	0
0.181818	0.772727	0	0.045455
0.681818	0.136364	0.181818	0
0.545455	0	0.454545	0
0.909091	0	0.045455	0.045455
0.136364	0	0.863636	0

Cbr20-Cni14-MEME3			
A	C	G	T
0.166667	0	0.222222	0.611111
0	0.277778	0.277778	0.444444
0.055556	0.111111	0.444444	0.388889
0	0.222222	0	0.777778
0.888889	0.055556	0.055556	0
0.444444	0.055556	0.111111	0.388889
0.666667	0.222222	0.055556	0.055556
0	0	0.222222	0.777778
0	0.388889	0.611111	0
0	0.111111	0.888889	0
0	0	0.888889	0.111111
0.388889	0.611111	0	0
0.722222	0	0.277778	0
0	0	1	0
0.222222	0	0.777778	0
0.055556	0.111111	0.833333	0
0.222222	0	0.5	0.277778
0	0.333333	0.666667	0
0.5	0.277778	0.055556	0.166667
0.444444	0.166667	0.222222	0.166667
0.222222	0.277778	0.388889	0.111111
0.666667	0	0	0.333333
0	0	0.833333	0.166667
0.111111	0.111111	0.166667	0.611111
0	0.555556	0.055556	0.388889

CS179Cni15-3			
A	C	G	T
1	0	0	0
0.350202	0	0	0.649798
0.623482	0.0384615	0	0.3380565
0	0.208502	0.117409	0.674089
0.0384615	0.2469635	0.714575	0
0.105263	0.117409	0.751012	0.026316
0.1842105	0	0.751012	0.0647775
0.1558705	0.817814	0.026316	0
1	0	0	0
0.026316	0	0.973684	0
0.026316	0	0.973684	0
0.0384615	0	0.9615385	0

Averaged Cbr30bp and CniMEME3			
A	C	G	T
0.5641025	0.0384615	0	0.3974355
0	0.801282	0.0384615	0.1602565
0	0	0	1
0	0.3141025	0	0.6858975
0	0.076923	0.076923	0.846154
0.0384615	0	0.6794875	0.282051
0	0.724359	0	0.275641
0.0384615	0	0.9615385	0
0.6794875	0.2371795	0.0833335	0
0.3974355	0	0.3141025	0.2884615
0.525641	0.153846	0.076923	0.2435895
0.6474355	0.0384615	0.076923	0.2371795
0.25	0.230769	0.4423075	0.076923
0.25	0.205128	0.467949	0.076923
0	0.1153845	0.474359	0.4102565
0.0384615	0.3974355	0.403846	0.1602565
0.923077	0.076923	0	0
0.5641025	0	0.2371795	0.198718
0.5128205	0.0384615	0.076923	0.371795
0	0	0.1602565	0.8397435
0.474359	0.3333335	0.153846	0.0384615
0.076923	0.1666665	0.3525645	0.403846
0.596154	0.121795	0.198718	0.0833335
0.1153845	0.8846155	0	0
0.9615385	0	0	0.0384615
0.4423075	0.3653845	0.153846	0.0384615
0.0384615	0	0.878205	0.0833335

Cni-MEME1			
A	C	G	T
0.428571	0	0	0.571429
0	0.142857	0.857143	0
0.142857	0	0.285714	0.571429
1	0	0	0
0.857143	0.142857	0	0
0.428571	0	0	0.571429
0	0	0.142857	0.857143
0.285714	0.142857	0.428571	0.142857
0.142857	0	0.857143	0
0.142857	0.285714	0.571429	0
0.428571	0.285714	0.285714	0
1	0	0	0
0	0.142857	0.857143	0
0	0	1	0
0	0	1	0
0	0	1	0
0.142857	0.428571	0.428571	0
0.857143	0	0.142857	0
0	0.428571	0.571429	0
0.571429	0	0.428571	0

Cni-MEME2			
A	C	G	T
0	1	0	0
0	0.857143	0	0.142857
0	1	0	0
0	0	0	1
0	0	0.714286	0.285714
0	0.571429	0	0.428571
0	0.285714	0.428571	0.285714
0	1	0	0
0.714286	0	0.285714	0
0.428571	0	0	0.571429
0.857143	0.142857	0	0
0.142857	0	0	0.857143
0.428571	0	0.571429	0
0.285714	0.285714	0.142857	0.285714
0.142857	0.857143	0	0

Cni-MEME3			
A	C	G	T
0.666667	0	0	0.333333
0	0.833333	0	0.166667
0	0	0	1
0	0.166667	0	0.833333
0	0	0	1
0	0	0.666667	0.333333
0	0.833333	0	0.166667
0	0	1	0
0.666667	0.166667	0.166667	0
0.333333	0	0.166667	0.5
0.666667	0	0	0.333333
0.833333	0	0	0.166667
0.5	0	0.5	0
0.5	0.333333	0.166667	0
0	0	0.333333	0.666667
0	0.333333	0.5	0.166667
1	0	0	0
0.666667	0	0.166667	0.166667
0.333333	0	0	0.666667
0	0	0.166667	0.833333
0.333333	0.666667	0	0
0	0.333333	0.166667	0.5
0.5	0.166667	0.166667	0.166667
0	1	0	0
1	0	0	0
0.5	0.5	0	0
0	0	0.833333	0.166667

Cni-top10-MEME1			
A	C	G	T
0	0.222222	0.777778	0
0	0.333333	0.222222	0.444444
0.888889	0	0.111111	0
0.888889	0	0	0.111111
0.444444	0	0	0.555556
0	0	0.111111	0.888889
0.111111	0.222222	0.666667	0
0	0	1	0
0.333333	0	0.555556	0.111111
0.444444	0.555556	0	0
1	0	0	0
0.222222	0	0.777778	0
0.111111	0.111111	0.666667	0.111111
0.111111	0	0.777778	0.111111
0	0	1	0
0	0.444444	0.555556	0
0.888889	0	0.111111	0
0	0.555556	0.444444	0

Cni-top10-MEME2			
A	C	G	T
0.75	0.125	0.125	0
0.75	0.25	0	0
0.25	0.125	0	0.625
0	0.125	0.125	0.75
0	0.25	0.5	0.25
0	0	0.5	0.5
0.5	0	0.125	0.375
0	0.625	0	0.375
0.125	0	0	0.875
0	0.125	0.125	0.75
0	0.125	0	0.875
0.125	0	0.875	0
0.125	0.75	0.125	0
0	0.125	0.875	0
0.5	0.375	0	0.125
0.25	0	0.5	0.25
0.25	0.125	0.125	0.5
0.5	0.125	0	0.375
0	0	0.375	0.625
0	0.5	0.375	0.125
0.125	0	0.875	0
0.125	0.5	0	0.375
0.875	0.125	0	0
0.25	0.125	0.125	0.5
0.5	0.125	0	0.375
0.125	0	0.125	0.75
0.125	0	0.5	0.375
0	0.125	0.75	0.125
0.375	0	0.625	0
0	0.875	0.125	0
0.875	0	0.125	0
0.375	0.375	0.25	0
0	0	1	0

Cni-top10-MEME3			
A	C	G	T
0.6	0	0	0.4
0.1	0.8	0	0.1
0	0	0	1
0	0.2	0	0.8
0	0	0	1
0	0	0.5	0.5
0	0.8	0.1	0.1
0.1	0	0.9	0
0.8	0.1	0.1	0
0.4	0.1	0.2	0.3
0.4	0	0.3	0.3
0.7	0.1	0.1	0.1
0.4	0.1	0.5	0
0.6	0.1	0.3	0
0.2	0.3	0.1	0.4
0.1	0.1	0.5	0.3
1	0	0	0
0.7	0	0.2	0.1
0.3	0	0.1	0.6
0.1	0.1	0.2	0.6
0.2	0.6	0.1	0.1
0.1	0.2	0.2	0.5
0.5	0.3	0.1	0.1
0.4	0.6	0	0
1	0	0	0
0.3	0.4	0.3	0
0	0	0.8	0.2
0.2	0	0.5	0.3

CS179Cni10-1			
A	C	G	T
0.222222	0.518519	0.074074	0.185185
0	0.111111	0	0.888889
0.037037	0.148148	0.148148	0.666667
0.037037	0.148148	0	0.814815
0.148148	0	0.666667	0.185185
0	0.777778	0	0.222222
0.111111	0	0.888889	0
0.481481	0.259259	0.222222	0.037037
0.296296	0.037037	0.407407	0.259259
0.444444	0.259259	0.074074	0.222222
0.703704	0.037037	0.037037	0.222222
0.259259	0.111111	0.333333	0.296296
0.407407	0.111111	0.37037	0.111111
0.074074	0.296296	0.37037	0.259259
0.111111	0.185185	0.259259	0.444444
0.888889	0.074074	0	0.037037
0.407407	0.074074	0.111111	0.407407
0.555556	0	0	0.444444
0.074074	0.111111	0.074074	0.740741
0.222222	0.259259	0.407407	0.111111
0.148148	0.074074	0.592593	0.185185
0.37037	0.111111	0.444444	0.074074
0.185185	0.666667	0.148148	0
0.962963	0	0	0.037037
0.222222	0.148148	0.62963	0
0	0	0.925926	0.074074
0.222222	0.037037	0.555556	0.185185
0	0.185185	0.444444	0.37037

Cni14-MEME1			
A	C	G	T
0	0	0.181818	0.818182
0.090909	0.090909	0.818182	0
0.272727	0	0.636364	0.090909
0.090909	0.090909	0.727273	0.090909
0.181818	0.727273	0.090909	0
0.909091	0	0.090909	0
0	0.090909	0.909091	0
0	0	0.909091	0.090909
0	0	1	0
0.090909	0	0.636364	0.272727
0	0.272727	0.363636	0.363636
0.272727	0.272727	0.363636	0.090909
0.090909	0	0.454545	0.454545
0	0.818182	0.090909	0.090909
0.636364	0	0	0.363636
0	0	1	0
0	0.363636	0.181818	0.454545
0	0.545455	0	0.454545

Cni14-MEME2			
A	C	G	T
1	0	0	0
0.75	0	0	0.25
0.625	0	0	0.375
0	0	0	1
0.125	0.125	0.625	0.125
0	0	1	0
0	0	0.875	0.125
0.25	0.75	0	0
0.875	0	0.125	0
0	0	1	0
0	0	1	0
0.125	0	0.875	0
0.25	0.125	0.625	0
0	0.25	0.75	0
1	0	0	0

Ctr-top17			
A	C	G	T
0.059	0.824	0.118	0.000
0.353	0.412	0.118	0.118
0.176	0.529	0.118	0.176
0.000	0.706	0.118	0.176
0.000	0.176	0.471	0.353
0.000	0.529	0.353	0.118
0.059	0.294	0.353	0.294
0.000	0.765	0.059	0.176
0.412	0.000	0.000	0.588
0.294	0.706	0.000	0.000
0.000	0.000	0.000	1.000
0.000	0.235	0.765	0.000
0.176	0.529	0.118	0.176
0.000	1.000	0.000	0.000
0.059	0.765	0.000	0.176
0.824	0.176	0.000	0.000
0.118	0.000	0.000	0.882
0.529	0.059	0.059	0.353
0.000	0.059	0.000	0.941
0.000	0.882	0.118	0.000
0.118	0.706	0.059	0.118

CS179Cni15-3			
A	C	G	T
0	0.692308	0.153846	0.153846
0.153846	0.615385	0.153846	0.076923
0	0.923077	0	0.076923
0	1	0	0
0	1	0	0
0	0	0	1
0	0	0.846154	0.153846
0.076923	0.923077	0	0
0	0.923077	0.076923	0
0	0.692308	0.230769	0.076923
0.769231	0.076923	0.153846	0
0.307692	0	0.076923	0.615385
0.615385	0	0	0.384615
0	0	0	1
0.615385	0.153846	0.230769	0

Averaged 8 MEX-like			
A	C	G	T
0.947845857	0.007936571	0.023809571	0.020408143
0.520125714	0.028344714	0.036281143	0.415248429
0.569010714	0.042735	0.007936571	0.380317857
0	0.08783975	0.11160325	0.800557
0.086207125	0.1829755	0.697335125	0.033482125
0.113978	0.0968125	0.77126675	0.017942625
0.170511375	0.047077875	0.71145	0.07096075
0.286397125	0.659946	0.053656875	0
0.938289125	0	0.061710875	0
0.03435675	0.02922075	0.9364225	0
0.06610275	0.013888875	0.894755875	0.0252525
0.04607375	0.013888875	0.9261485	0.013888875

Averaged 8 MEX-like, long			
A	C	G	T
0.166667	0	0.222222	0.611111
0.142857	0.092592667	0.378307	0.386243333
0.0853175	0.15476175	0.59126975	0.16865075
0.0238095	0.270151833	0.145384667	0.560653833
0.947845857	0.007936571	0.023809571	0.020408143
0.520125714	0.028344714	0.036281143	0.415248429
0.569010714	0.042735	0.007936571	0.380317857
0	0.08783975	0.11160325	0.800557
0.086207125	0.1829755	0.697335125	0.033482125
0.113978	0.0968125	0.77126675	0.017942625
0.170511375	0.047077875	0.71145	0.07096075
0.286397125	0.659946	0.053656875	0
0.938289125	0	0.061710875	0
0.03435675	0.02922075	0.9364225	0
0.06610275	0.013888875	0.894755875	0.0252525
0.04607375	0.013888875	0.9261485	0.013888875
0.106675667	0.046474333	0.729458167	0.117391833
0.0494505	0.313820167	0.576123	0.060606
0.7037518	0.110101	0.134632	0.0515152
0.13383825	0.2876985	0.42316	0.155303
0.264550333	0.36532	0.302789667	0.06734
0.6515155	0	0	0.3484845
0	0	0.9166665	0.0833335
0.0555555	0.2373735	0.1742425	0.532828
0	0.5505055	0.027778	0.421717

Averaged 5 Cbr-MEX-II-like			
A	C	G	T
0.053703667	0.734953167	0.025166167	0.186177
0.028409167	0.026094333	0	0.945496667
0.006172833	0.259316333	0.045524667	0.688986167
0.006172833	0.101468667	0.048368333	0.843990167
0.065921	0	0.6955615	0.2385175
0.020833333	0.743052333	0.0375	0.198614333
0.070733	0.020833333	0.908433667	0
0.621894167	0.269618167	0.0814815	0.027006167
0.365952	0.0228395	0.365026167	0.246181833
0.4333785	0.1683555	0.116395833	0.281869833
0.624005	0.056493333	0.071207833	0.248293833
0.200785667	0.172714167	0.424582167	0.201917667
0.2512345	0.2171975	0.443847667	0.087720167
0.074088167	0.175723	0.480832833	0.269356
0.0839905	0.339100167	0.299037333	0.277871667
0.904704167	0.089123	0	0.006172833
0.497602	0.040754833	0.189623833	0.272019333
0.5105025	0.048805333	0.065035	0.375657167
0.0574215	0.050336667	0.135082667	0.757159167
0.347808167	0.261896833	0.264637667	0.125657167
0.0821505	0.122067833	0.488256667	0.307525
0.497188833	0.124268333	0.321752667	0.056790167
0.173871167	0.7730285	0.0531005	0
0.937446	0	0.020833333	0.041720667
0.365154667	0.290804333	0.316068833	0.027972
0.0128205	0	0.913722667	0.073456833

CS179Cni15-1			
A	C	G	T
0	0.333333	0.533333	0.133333
0	0.266667	0.066667	0.666667
0.133333	0	0.866667	0
0	0	0	1
0	0	0.866667	0.133333
0	0.2	0	0.8
0	0	0.933333	0.066667
0	0.866667	0	0.133333
0.066667	0	0.933333	0
0	0.8	0	0.2
0	0	1	0
0	0.733333	0.266667	0
0.266667	0.2	0.533333	0
0	0.933333	0.066667	0
0.266667	0.133333	0.6	0

Averaged 4 Cbr-MEX-like			
A	C	G	T
0.96428575	0	0	0.03571425
0.362601	0	0.03571425	0.60168475
0.61084825	0.01923075	0	0.369921
0	0.1756795	0.0587045	0.765616
0.05048075	0.15473175	0.7635375	0.03125
0.12406	0.16584725	0.6969345	0.013158
0.199248	0	0.73711325	0.06363875
0.21186375	0.7749785	0.013158	0
0.96875	0	0.03125	0
0.013158	0	0.986842	0
0.04887225	0	0.95112775	0
0.05048075	0	0.94951925	0

Averaged "top 600" core			
A	C	G	T
0	0.006666667	0.993333333	0
0.766666667	0.013333333	0.22	0
0	0	1	0
0	1	0	0
0.053333333	0	0.946666667	0
0.053333333	0.886666667	0	0.06
0.406666667	0.013333333	0.58	0
0.066666667	0.326666667	0.08	0.526666667
0.033333333	0	0.006666667	0.96
0.006666667	0.006666667	0	0.986666667
0.053333333	0	0.886666667	0.06
0.013333333	0.913333333	0.02	0.053333333

Cel-top600			
A	C	G	T
0.22	0.08	0.12	0.58
0.04	0.06	0.4	0.5
0.1	0.2	0.34	0.36
0.04	0.16	0.38	0.42
0.04	0.88	0	0.08
0.66	0.04	0.3	0
0.54	0	0.34	0.12
0.04	0.16	0	0.8
0.48	0	0.52	0
0	0.02	0.98	0
0.94	0	0.06	0
0	0	1	0
0	1	0	0
0.02	0	0.98	0
0.04	0.82	0	0.14
0.42	0.02	0.56	0
0.1	0.36	0.14	0.4
0.02	0	0	0.98
0	0.02	0	0.98
0.08	0	0.88	0.04
0.02	0.92	0.02	0.04

Cbr-top600			
A	C	G	T
0.12	0	0.52	0.36
0.06	0.16	0.1	0.68
0.08	0.14	0.28	0.5
0	0.94	0	0.06
0.62	0.06	0.28	0.04
0.46	0	0.48	0.06
0	0.4	0.02	0.58
0.36	0.02	0.62	0
0	0	1	0
0.58	0.02	0.4	0
0	0	1	0
0	1	0	0
0.04	0	0.96	0
0.1	0.9	0	0
0.4	0.02	0.58	0
0.04	0.28	0.06	0.62
0.04	0	0.02	0.94
0	0	0	1
0.06	0	0.88	0.06
0	0.94	0	0.06
0.46	0.04	0.24	0.26

Ctr-top600			
A	C	G	T
0.2	0.12	0.22	0.46
0	0.06	0.46	0.48
0.14	0.16	0.2	0.5
0.08	0.1	0.42	0.4
0.02	0.88	0	0.1
0.62	0.06	0.28	0.04
0.48	0.1	0.34	0.08
0.02	0.14	0	0.84
0.52	0.04	0.42	0.02
0	0	1	0
0.78	0.02	0.2	0
0	0	1	0
0	1	0	0
0.1	0	0.9	0
0.02	0.94	0	0.04
0.4	0	0.6	0
0.06	0.34	0.04	0.56
0.04	0	0	0.96
0.02	0	0	0.98
0.02	0	0.9	0.08
0.02	0.88	0.04	0.06

Appendix J

X2A scripts

J.1 Motif enrichment on X and in peaks

```
#!/usr/bin/python

#modified X2A ratio script from Michael Eisen

#command line: python program_name matrix_filename patsin_filename
#to run on all motif files in a folder:
#for i in *.mtf; do echo $i && /usr/bin/python X2A_nigoni_final.py
   $i nigoni_2015.12.01_Mauve_plus_XXvsXO_final.patsin; done

import os
import sys
import math

#C. briggsae WS230
lengths = {
 "chrI":14998623, "chrII":16060615, "chrIII":14170909,
 "chrIV":16944347, "chrV":19015108, "chrX":21119894
 }
A_length = (lengths["chrI"] + lengths["chrII"] + lengths["chrIII"] +
            lengths["chrIV"] + lengths["chrV"])
X_length = lengths["chrX"]

#nigoni_2015.12.01
#lengths = {
# "chrI":13914462, "chrII":16218367, "chrIII":23829918,
# "chrIV":20194008, "chrV":24469482, "chrX":28695628,
```

```
# "X_random":588998, "chrun":1528089
# }
#A_length = (lengths["chrI"] + lengths["chrII"] + lengths["chrIII"] +
#            lengths["chrIV"] + lengths["chrV"])
#X_length = lengths["chrX"] + lengths["X_random"]

#C. tropicalis 33x assembly
#lengths = {"chrA":66960462, "chrX":15750803, "chrun":151263}
#A_length = lengths["chrA"]
#X_length = lengths["chrX"]

#C. elegans WS230
#lengths = {
# "chrI":15072425, "chrII":15279347, "chrIII":13783702,
# "chrIV":17493795, "chrV":20924151, "chrX":17718868,
# }

#A_length = (lengths["chrI"] + lengths["chrII"] + lengths["chrIII"] +
#            lengths["chrIV"] + lengths["chrV"])
#X_length = lengths["chrX"]

#patser output gff
motif_gff_filename = sys.argv[1]

#macs output gff
peak_gff_filename = sys.argv[2]

peak_length = 200 * 500
cumulative_count = 0

def motif_in_peak(chromosome, lower_bound, upper_bound):
  #arguments: chromosome, start, end
  for i in peak_list[0:200]:
    peak_chromosome = i[0]
    peak_lower_bound = i[1]
    peak_upper_bound = i[2]
    assert peak_lower_bound < peak_upper_bound
    if chromosome == peak_chromosome and upper_bound >= peak_lower_bound
                     and lower_bound <= peak_upper_bound:
      return True
  return False
```

```
peak_list = []
summit = 0
with open(peak_gff_filename, "r") as peak_gff_file:
  for line in peak_gff_file:
    if "#" in line:
      continue
    else:
      line = line.strip().split()
      summit = int(line[13][0:-1])
      peak_list.append((line[0], summit - 250, summit + 250, float(line[5])))

peak_list.sort(key=lambda tup: tup[3], reverse=True)

count_dictionary = {}
with open(motif_gff_filename, "r") as motif_gff_file:
  for line in motif_gff_file:
    if "#" in line or "radom" in line or "chru" in line:
      continue
    else:
      line = line.split()
      chromosome = line[0]
      start = int(line[3])
      end = int(line[4])
      lnp_value_bin = int(math.floor(float(line[5]) * 10))
      if lnp_value_bin not in count_dictionary:
        count_dictionary[lnp_value_bin] = {"A": 0, "X": 0, "un": 0,
                                           "peak": 0}
      count_dictionary[lnp_value_bin].setdefault(chromosome, 0)
      count_dictionary[lnp_value_bin][chromosome] += 1
      if motif_in_peak(chromosome, start, end):
        count_dictionary[lnp_value_bin]["peak"] += 1
      if chromosome == "chrX" or chromosome == "X_random":
        count_dictionary[lnp_value_bin]["X"] += 1
      elif chromosome.startswith("chru"):
        count_dictionary[lnp_value_bin]["un"] += 1
        #change this if chrun is to be included in the autosomal count
      elif chromosome.startswith("chr"):
        count_dictionary[lnp_value_bin]["A"] += 1
      else:
        raise Exception("unexpected chromosome name")

print >>sys.stderr, chromosome
```

```
print >>sys.stderr, start
print >>sys.stderr, end

#Next calculate the cumulative counts and the X:A ratios

print
"lnp_value_bin\tXtoA_ratio\tpeaktoX_ratio\tpeaktototal_ratio\tcumulative_
counts_A\tcumulative_counts_X\tcumulative_counts_un\tcumulative_counts_
peak\tpeak_density\ttotal_density"
cumulative_counts_A = 0
cumulative_counts_X = 0
cumulative_counts_un = 0
cumulative_counts_peak = 0
for k,v in sorted(count_dictionary.iteritems(), key=lambda item: item[0]):
  cumulative_counts_A += v["A"]
  cumulative_counts_X += v["X"]
  cumulative_counts_un += v["un"]
  cumulative_counts_peak += v["peak"]
  X_density = float(cumulative_counts_X) / float(X_length)
  A_density = float(cumulative_counts_A) / float(A_length)
  peak_density = float(cumulative_counts_peak) / float(peak_length)
  total_density = (float(cumulative_counts_A) + float(cumulative_counts_X) +
float(cumulative_counts_un)) / float(X_length + A_length)
  peak_to_total_ratio = float(peak_density) / float(total_density)
  if A_density > 0:
    XtoA_ratio = float(X_density) / float(A_density)
  else:
    XtoA_ratio = None
  if X_density > 0:
    peaktoX_ratio = float(peak_density) / float(X_density)
  else:
    peaktoX_ratio = None
  print "%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s" % (
    k, XtoA_ratio, peaktoX_ratio, peak_to_total_ratio, cumulative_counts_A,
    cumulative_counts_X, cumulative_counts_un, cumulative_counts_peak,
    peak_density, total_density)
```

J.2 Plot motif enrichment on X and in peaks

```
#!/usr/bin/python

#Plot motif enrichment on the X chromosome, in the top 200 peaks vs. the X
  chromosome, and in the top 200 peaks vs. the entire genome

import sys
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
import numpy
import pylab
import os

lines = []
line_names = []
colors = ["#FF7B4F", "#37CCFF", "#8F35DA", "#2FED2F"]
species = []
counter = 0

title = sys.argv[1]

f, (ax1, ax2, ax3) = plt.subplots(3, sharex = True)
for i in sys.argv[2:]:
  x1 = []
  x2 = []
  x3 = []
  X2A = []
  peak2X = []
  peak2total = []
  fh = open(i, "r")
  headings = fh.readline().split("\t")
  for line in fh:
    line = line.strip().split()
    try:
      X2A.append(float(line[1]))
      x1.append(float(line[0]) / 10)
    except:
      continue
    try:
```

```
      peak2X.append(float(line[2]))
      x2.append(float(line[0]) / 10)
    except:
      continue
    try:
      peak2total.append(float(line[3]))
      x3.append(float(line[0]) / 10)
    except:
      continue
  ax1.plot(x1, X2A, "-o", color=colors[counter])
  ax2.plot(x2, peak2X, "-o", color=colors[counter])
  ax3.plot(x3, peak2total, "-o", color=colors[counter])
  #lines.append((scores, X2A, "-o"))
  #species.append(species_list[counter])
  counter += 1

plt.xlabel("log probability score")
plt.ylabel("density ratio")
plt.savefig("\%s.pdf" % title)
```