a. Significance

Sex determination in *C. elegans*: Like mammals and flies, *C. elegans* use a pair of sex chromosomes to determine sex. *C. elegans* utilize an XO system which measures the ratio of X chromosomes to autosomes (Madl and Herman, 1979). Males have one X chromosome (XO;AA), and hermaphrodites have two X chromosomes (XX;AA). Worms count X chromosomes using trans-acting X-signal elements (XSEs) that are transcribed from the X. The XSEs act in a dose-dependent manner to repress transcription of *xol-1* (XO-lethal). Countering this, autosomal signal elements (ASEs) are trans-acting activators of *xol-1*. In a ratio of 1X:2A, *xol-1* is activated and the worm develops as a male. In a ratio of 2X:2A, *xol-1* is repressed and the worm adopts the hermaphroditic fate (Powel et al, 2005).

X chromosome imbalance: A consequence of chromosomal sex determination is a difference in X chromosome copy number between the sexes. This copy number difference causes two forms of X chromosome imbalance. The first is X dosage imbalance between the sexes. The second is male monosomy of the X: males have one copy of the X chromosome to two copies of every autosome. X chromosome imbalance is not unique to *C. elegans*. The XY sex determination system used by mammals and flies also results in X copy number difference between males (XY) and females (XX). Mechanisms that act to restore X dosage balance are known as dosage compensation. Most prevalently studied are those mechanisms that balance the X chromosome between the sexes (Vicoso and Bachtrog, 2009).

Sex specific dosage compensation: In all organisms where it has been studied, dosage compensation between the sexes is an essential process that involves the targeting of a chromatin-regulating complex to the X chromosome in one of the sexes (Straub and Becker, 2007). In *Drosophila,* a dosage compensation complex is active in males. It targets the single X chromosome and hyperacetylates histone H4 at lysine 16 (Conrad and Akhtar, 2012). This hyperacetylation leads to an average two-fold upregulation of the single male X to match transcriptional output of the XX females (Smith et al, 2001). In mammals, females inactivate most of the genes on one of their two X chromosomes. X inactivation is specified by a cis-acting RNA (Xist) that is transcribed from one of the X chromsomes and results in formation of the largely heterochromatic Barr body (Hall and Lawrence, 2011). In *C. elegans*, male to hermaphrodite X balance is achieved by the Dosage Compensation Complex (DCC). The DCC specifically binds to both X chromosomes in XX hermaphrodites and reduces transcription of each by an average of one half (Meyer and Casson, 1986). DCC regulates X chromosome transcription only in somatic tissues. The mechanism of transcriptional repression by the DCC remains unclear.

Condensin regulation of transcription: The DCC is composed of a modified condensin complex that interacts with at least four other non-condensin subunits in order to reduce hermaphrodite X transcription. *C. elegans* have three condensin complexes (fig 1). Condensins contain two Structural Maintenance of Chromosomes (SMC) proteins and three non-SMC subunits. All metazoans contain two condensins (I and II) that share the same SMC subunits and are distinguished by a distinct set of three regulatory proteins: CAPG, CAPD, and CAPH (Ono et al, 2003). The DCC contains condensin I^{DC}, which differs from condensin I only by the substitution of DPY-27 for SMC-4. Condensins I and II are essential for chromosome condensation during cell division (Hudson et al, 2003; Ono et al, 2003; Vagnarelli et al, 2006). They are required for both the establishment and the stability of condensed chromosome structure. Additionally, condensins in *D. melanogaster* and *S. cerevisiae* have been implicated in the regulation of transcription during interphase (Bhalla et al, 2002; Cobbe et al, 2006; Dej et al, 2004; Hartl et al, 2008; Lupo et al, 2001). The mechanism of transcriptional regulation by condensins is yet unknown.

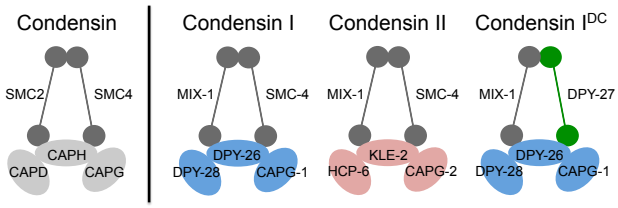

Figure 1 The left panel illustrates the subunit composition of a generic condensin. The right panel shows the *C. elegans* condensins. Condensin I^{DC} differs from Condensin I by only one SMC subunit, DPY-27.

DCC binding on the X chromosome: DCC specifically binds to both X chromosomes only in hermaphrodites (Meyer, 2005). DCC is first recruited to *rex* sites (recruitment elements on X) that contain clusters of a 12bp DNA sequence motif (Ercan et al, 2007; Jans et al 2009; McDonel et al, 2006). The non-condensin subunits SDC-1,2,3 and DPY-30 are required for the initial recruitment of DCC to *rex* sites. The DCC then spreads preferentially to active promoters. Promoter regions are not able to recruit the DCC on their own, but are bound

by the DCC in the context of the natural chromosome. Binding at promoters directly and dynamically correlates with transcriptional activity (Ercan et al, 2009). Finally, condensin I^{DC} spreads to coat the X chromosome (Ercan et al, 2007; 2009, Jans et al, 2009; Pferdehirt et al, 2011; Petty et al, 2009). Unlike recruitment, DCC spreading does not depend on the X chromosome sequence (Ercan et al, 2009; Pferdehirt et al, 2011).

DCC regulation of transcription: Upon DCC mutation, roughly 40% of all active X-linked genes show an increase in expression yet only half of these genes are bound by DCC (Jans et al, 2009). On a gene-to-gene basis, DCC binding at the promoter is neither necessary nor sufficient to induce compensation of that gene. This has lead to the hypothesis that the DCC works at a distance to regulate X transcription. How this regulation is accomplished and the range over which DCC can act have not yet been determined. Mutation of the DCC also leads to decreased expression of roughly 25 percent of active autosomal genes (Jans et al, 2009). This result indicates that the DCC might indirectly regulate autosomal transcription. A model explaining decreased autosomal expression postulates that the DCC affects the X and autosomal distribution of a genome-wide transcriptional activator that is present in limiting quantities. This model is detailed in Aim 1. Examples of titration-based regulation of X chromosome transcription exist in the *C. elegans* germline. In the germline, the DRM complex predominantly binds to autosomes, but its loss results primarily in decreased expression of X-linked genes (Tabuchi et al, 2011). Similarly, the histone-methyltransferase, MES-4, is autosomally associated but required for X chromosome silencing in the *C. elegans* germline (Bender et al, 2006). DCC regulation of autosomal expression serves as a model for complexes that paradoxically bind to one chromosome and regulate transcription of another.

X upregulation and Ohno's Hypothesis: DCC reduction of hermaphrodite X transcription balances X expression between the sexes but potentially gives hermaphrodites the same problem faced by males: functional monosomy of the X chromosome (Gupta et al, 2006). Evolutionary theory notes that there is no benefit to hermaphrodites to halving their X expression: because selection acts on individuals, an evolutionary pressure to balance X dose between the sexes does not easily explain how mechanisms of X downregulation evolved (Vicoso and Bachtrog, 2009). It has been proposed that, in order to overcome male monosomy, a mechanism evolved to upregulate X expression (Ohno, 1967) (fig 2b). This proposed mechanism is not specific to males but predicted to operate on the X in both sexes. Mathematically it has been shown that X upregulation in both sexes can evolve if the fitness benefit in males outweighs the fitness deficit (due to X overexpression) in females (Engelstädter and Haig, 2008). This is supported by work in *D. melanogaster* showing that deleterious effects of reducing a gene's dose are more pronounced than those caused by increasing dose (Lindsley et al, 1972). Following upregulation of the X in both sexes, XX individuals had to secondarily evolve mechanisms to repress X expression in order to restore balance with the autosomes (Charlesworth 1978; 1996) (fig 2c). Recent genome-wide expression studies indicate that overall expression levels are roughly equal between the X chromosome and the autosomes in both sexes (Adler et al, 1997, Deng et al, 2011; Gupta et al, 2006; Nguyen, D.K. & Disteche, 2006). These findings are consistent with the idea of X upregulation. The mechanism of X upregulation in both sexes is unknown.

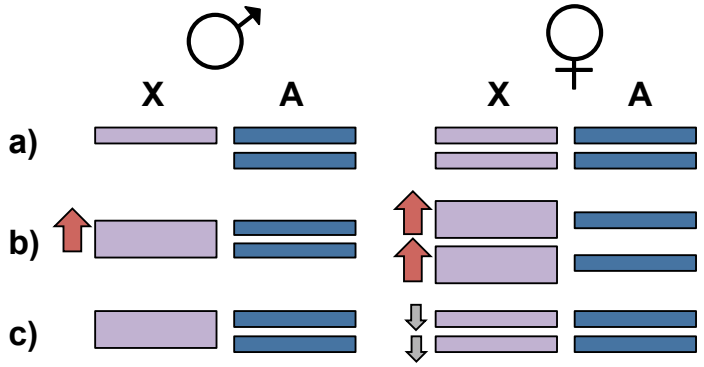

Figure 2 a) X chromosome sex determination results in two types of dosage imbalance: 1) the X is in a 1:2 ratio between the sexes and 2) males have one X chromosome to every two autosomes. **b)** To restore X:A balance in males, the X chromosome is upregulated in **both** sexes (Ohno's Hypothesis). **c)** As a consequence of X upregulation, XX hermaphrodites decrease expression of each of their X chromosomes. Together, the two dosage compensation mechanisms restore X balance between the sexes and the autosomes.

Summary: We use dosage compensation of the X chromosome as a model of chromosome-wide transcriptional regulation. In *C. elegans,* we know that DCC is necessary for reduced X expression in hermaphrodites. However, we do not know the mechanism of DCC-mediated X repression, nor how this repression is linked to autosomal regulation. Further, most work on dosage compensation has focused only on the balance of X expression between the sexes. Here, we propose an investigation of dosage compensation

evolution as a means both to determine the extent of X-upregulation as well as to define sequence changes associated with dosage compensation mechanisms. *C. elegans* is an excellent organism in which to study dosage compensation for several reasons. First, dosage compensation has been studied extensively in the worm, providing background to guide our research. Second, because the DCC is composed of a modified condensin complex we have the unique opportunity to investigate the mechanisms by which condensins mediate transcriptional regulation. Third, we have access to many genetically characterized strains, all of which are amenable to the high-throughput genetic and genomic experiments outlined in this proposal. Lastly, comparative genomics studies using different nematode species make the worm a good model for the evolutionary development of dosage compensation mechanisms.

Relevance to Human Disease: Chromosomal imbalances are often associated with disease. The consequences of aneuplodies are thought to be caused by altered gene dose of an entire chromosome. In humans, all autosomal monosomies are lethal. A few cases of monosomy 21 that have reportedly survived to term but all died shortly after birth (Fisher et al, 2012). Autosomal trisomies are less severe but often trigger miscarriage or, if the child survives to term, physical malformations and learning disabilities (Fitzpatrick, 2005). Cancers are also frequently linked to genomic instability and chromosomal aneuploidy (Gordon et al, 2012). Trisomy of chromosome 8 is associated with 10 to 20 percent of acute myeloid leukemias and Ewing sarcomas (Barnard et al, 1996; Maurici et al, 1998) Organisms, including humans, have developed mechanisms to handle the natural aneuploidies of the X chromosome. Consequently, humans with monosomies (Turner Syndrome) and trisomies (Triple X Syndrome) of the X survive with few physical defects (Otter et al, 2010; Reindollar, 2011). Work to understand how altered X chromosome dose is tolerated and how X chromosome expression is regulated will be important to developing treatments of other chromosomal imbalances.

b. Innovation

The research proposed is innovative because it considers the evolutionary history of dosage compensation in *C. elegans*. It explores the proposed mechanism of X upregulation in both sexes as it relates to hermaphrodite-specific X repression. Previous work has focused on comparing X expression between XO males, XX hermaphrodites, and XX DCC-mutant hermaphrodites and has typically measured overall X and autosomal expression. Our approach will focus on orthologous gene expression when the ortholog is X located in one species of worm and autosomally located in another. This will allow direct comparison of X and autosomal expression between species and between sexes.

c. Approach

Aim 1: Determine how X chromosome dosage compensation affects autosomal transcription

Background and Rationale: By immunofluorescence and by ChIP, the DCC was found to bind primarily and specifically to both X chromosomes in *C. elegans* hermaphrodites. ChIP-chip experiments also revealed the existence of discrete autosomal DCC binding sites (Ercan et al, 2009; Jans et al, 2009). Similar to the X chromosome, DCC binding to autosomes shows a strong bias toward promoters of expressed genes. Also like the X chromosome, DCC binding to an autosomal gene does not appear to correlate directly with that gene's transcriptional regulation. Upon DCC mutation, ~40 percent of active X genes show an increase in expression while ~25 percent of all autosomal genes show a decrease in expression (Jans et al, 2009). (For comparison, ~3 percent of X genes show a decrease and ~7 percent of autosomal genes show an increase in expression (Jans et al, 2009)). The effect of DCC mutation on genome-wide transcription can be explained by a Titration Model, which proposes that DCC acts in hermaphrodites to repel a limiting transcriptional activator. Another possibility is that the DCC attracts a limiting transcriptional repressor to the X chromosome. For the rest of the proposal we will focus on repelling an activator as this should be similar to attracting a repressor

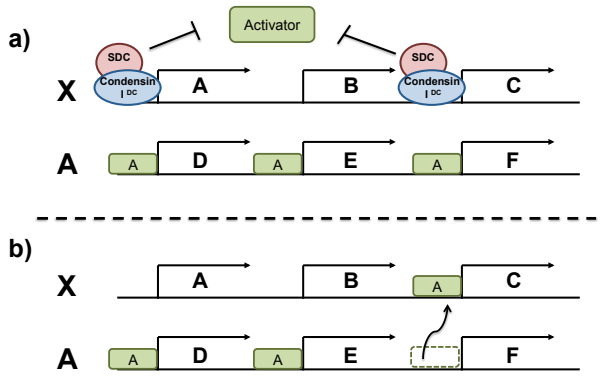

Figure 3 Model for autosomal regulation by the DCC. Only one X and one autosome are depicted. **a)** In a WT hermaphrodite, the DCC binds to the X chromosome and prevents activator activity. **b)** In a DCC mutant, activator is titrated away from autosomal genes.

with respect to titration. In a wild type hermaphrodite, the activator is prevented from acting on the X chromosome and so primarily works to activate autosomal transcription (fig 3a). The model predicts that in the absence of DCC, X chromosome genes are able to titrate the activator away from autosomal genes (fig 3b). We will test this model by determining transcriptional changes in worms carrying partial chromosome duplications predicted to titrate the activator (Aim 1.1). We will also determine the effect of altering the dose of several X-linked transcription factors contained within the duplications, either using single copy MosSCI insertion or available CGC strains carrying an extra GFP-tagged copy of the TF, and measuring transcriptional change by RNAseq (Aim 1.2).

Aim 1.1: Determine how altering X chromosome dose affects autosomal transcription.

Experimental Rationale and Design: We will test the Titration Model by assessing the ability of partial X chromosome duplications to affect autosomal transcription regardless of their individual genetic content. The strains we will work with are available from the CGC. The strains carrying X chromosome duplications vary in their ability to recruit the DCC (Csankovszki et al, 2004) (fig 4). As determined by FISH microscopy, three of the duplications strongly recruit the DCC, two recruit weakly, and two are unable to recruit the DCC on their own. This allows us to compare titration effect with or without DCC binding. All of the duplications have been roughly mapped (Figure 5): we know which transcription factors are contained within these regions. We will determine which genes are affected by the presence of a duplication and will quantify this change in expression. Because most of the autosomal genes affected in a DCC mutant show decreased expression (Jans et al, 2009) we expect the partial duplications to predominantly cause reduced autosomal transcription.

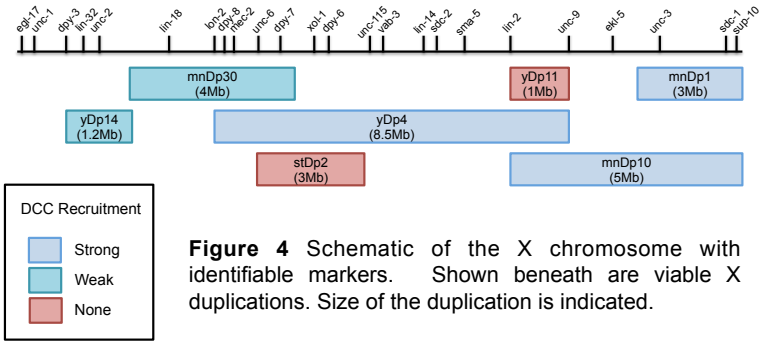

Figure 4 Schematic of the X chromosome with identifiable markers. Shown beneath are viable X duplications. Size of the duplication is indicated.

Controls to be used: Because they cannot recruit DCC, autosomal duplications will serve as a control (Table 1). When viable, males carrying X and autosome duplications will also serve as a control because they lack DCC activity and naturally have only one copy of the X chromosome.

Experimental Methods: As a first step, we will map the duplicated regions at a higher resolution using DNA-seq and copy number variation analysis, as we performed to map X-linked genes (Aim 3.1) We will perform RNA seq analysis comparing duplication strain data to each other and to the wild type strain (N2). We will perform our analysis in mixed stage embryos and in L2 larval stage animals. In these two developmental stages, most of the tissue is somatic, and we expect dosage compensation to be active.

We will use Illumina HiSeq2000 for the high-throughput sequencing of RNA. We will multiplex up to 8 libraries per lane. The sequences will be mapped to the genome using Tophat (Langmead et al, 2009; Trapnell et al, 2012). Differential gene analysis will be performed and significant expression changes will be identified using Cufflinks (Trapnell et al, 2012) and our RNA seq analysis pipeline.

Expected Outcomes and Interpretations:

X duplications: Under the Titration Model, X chromosome duplications of the same size and same ability to recruit DCC should cause equivalent decreased expression of the same set of autosomal genes, regardless of the genetic content of the duplication. We predict that larger duplications will have a greater effect on autosomal transcription. When possible, we will quantify autosomal expression changes in males bearing the duplication to evaluate the effect of duplication on transcription with or without the action of the DCC. In contrast to X chromosome duplications, autosomal duplications should have the same effect in both sexes.

Autosomal duplications: If the titratable activator is contained within a duplicated region, we would expect to see large increases in expression throughout the genome. Since we have many non-overlapping duplications, we will assume that the activator itself is not contained in most of our duplications. Then, under the Titration

Model we expect that autosomal duplications will be able to titrate the activator. This will allow us to determine if adding additional chromosomal content, regardless of DCC activity, is sufficient to titrate an activator. We may find autosomal duplications cause different effects on autosomal gene expression than X duplications. We posit two explanations for the potential difference in effects. In the first, different sequences (X vs autosomal) may be able to titrate the activator with differing abilities. In the second, it is specifically DCC action on the X duplications that results in different expression changes. We will distinguish these two possibilities by performing DPY-27 RNAi in X duplications. If the difference is due to the DCC, DPY-27 RNAi in X duplication strains should result in similar expression changes as observed in the autosomal duplication strains.

DCC titration: X chromosome duplications that strongly recruit DCC have been shown to deplete the complex off of the X chromosomes, indicating that quantities of DCC are limiting (Csankovszki et al, 2004). The transcriptional consequence of DCC depletion from the X has not yet been evaluated and will have to be taken into account during the course of our analysis. However, given that the X chromosome partial duplications are viable, and that most duplicate less than 10 percent of the total diploid X sequence, we do not expect limiting concentrations of DCC to interfere with our analysis. As a direct result of their increased copy number, we expect increased expression of genes contained within the duplication.

Table 1

Strain Name	Duplicated Chr	Duplication Name	Recruitment of DCC	Size (Mb)	Attached to Chr	Homozygous Viable	Males Viable
SP262	X	mnDp1	strong	3Mb	V	no	NA
SP219	X	mnDp1	strong	3Mb	V	no	yes
TY1909	X	yDp4	strong	8.5Mb	autosome	no	yes
SP117	X	mnDp10	strong	5Mb	I	yes	yes
SP957	X	mnDp30	weak	4Mb	free		NA
TY2025	X	yDp14	weak	1.2Mb	I	no	yes
RW2551	X	stDp2	none	3Mb	II	no	yes
TY1916	X	yDp11	none	1Mb	IV	yes	yes
DR1786	IV	mDp4		8Mb	autosome	no	yes
CB3439	III	mnDp37		6.8Mb	free		NA
BA609	III	eDp6		2.3Mb	free		NA
EJ255	I	sDp2		5Mb	free		NA

Potential Problems and Alternative Approaches: We do not expect major technical difficulties in Aim 1.1. We also do not expect problems with *dpy-27* RNAi as we have reproducibly seen excellent knockdown in N2 worms. However, RNA feeding may not be as efficient in the duplication strains. In this case, an alternative is to cross the duplication strains with a *dpy-21* mutant strain (CB428).

Aim 1.2: Determine how altering X-linked TF dose affects autosomal transcription.

Experimental Rationale and Design: We offer a second model to explain DCC autosomal regulation. We propose that in a DCC mutant, X-linked Transcription Factors (TFs) will increase in expression. If these TFs predominantly act to negatively regulate downstream autosomal targets, this would explain the observed decreased expression of a subset of autosomal genes in a DCC mutant. We call this model the Indirect Effect Model. If this model is correct, then different duplications of the same size should affect a different subset of autosomal genes. Further, some duplications may contain TFs that predominantly upregulate autosomal targets. Here, depending on the gene content of

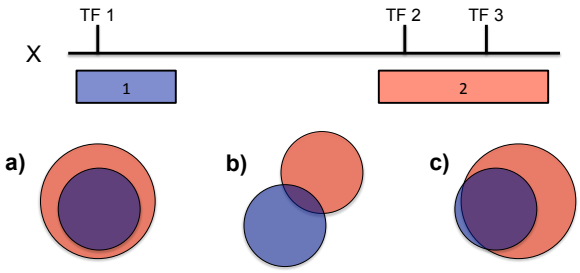

Figure 5 Two duplications of the X chromosome. Duplication1 contains TF1, Duplication2 contains TF2 and TF3. **a)** Titration Model. More autosomal genes are effected with increased size of duplication. **b)** Indirect Effect Model. Different subsets of autosomal genes are effected depending on TFs contained within the duplication. **c)** Both models. Most of the autosomal effect is due to titration. Some TF-specific targets are affected.

the duplication, autosomal expression could be either repressed or enhanced. We note that the two models are not exclusive. It could be that it is both a titration effect as well as increased action by X-linked TFs that explains changed autosomal expression in a DCC mutant. Using this duplication analysis, we can distinguish all three scenarios (fig 5). If it is just the Titration Model, the same subset of autosomal genes will be affected by all duplications. If it is just the Indirect Effect Model, different duplications will alter expression of a different subset of autosomal genes, with little overlap. If both models are correct, all duplications should affect the same core subset of autosomal genes with some variation due to altered dose of different TFs. These variations due to TF action are expected to show no bias towards up or down regulation of the X and autosomes.

Experimental Methods:. In our Indirect Effect Model, altered X chromosome dose in a DCC mutant leads to altered autosomal expression due to increased X-linked TF expression. The duplication strains we proposed in our Aim 1.1 have been characterized, allowing us to define those TFs that fall within the duplicated region. We currently have a small list of candidate X-linked TFs that have been identified as being at the top of the *C. elegans* TF hierarchy (Table 2) (Gerstein et al, 2010). That is, they interact with and regulate other downstream TFs. Based on network analyses in yeast and *E. coli* (Yu and Gerstein, 2005), these TFs are predicted to be overall modulators of global transcription. Also in our list of candidates are TFs that have been previously shown to be highly expressed in somatic tissue (where DCC is active) (McGhee et al, 2007).

To determine if the altered autosomal transcription in duplications is due to altered dose of a candidate TF, we will insert an additional copy of the TF using MosSCI insertion (Frøkjær-Jensen et al, 2008; 2012). To parallel copy number increase by partial chromosomal duplication, the transgenic TFs will be expressed under their native promoters and 3'UTRs. When possible, we will use availble strains that carry an extra GFP-tagged copy of the TF. The GFP tagged TF strains were previously used by the ModENCODE project for ChIP seq experiments. These strains were generated using microparticle bombardment of fosmids containing the native TF sequence with the GFP in frame at the 3' end. We will measure increased TF levels by q-RT-PCR and determine effects on genome-wide transcribtion by RNAseq.

Controls to be used: Autosomal TFs whose downstream targets are well known will serve as controls (see Table 2).

Table 2

TF	Chr	Description	Duplications Containing the TF	Dosage Compensated	ModENCODE ChIP datasets	Available CGC strains
ceh-14	X	hox	yDp4 ; stDp2 ; mnDp30	unknown	734	OP73
elt-3	X	GATA factor	mnDp10	unknown	2614 ; 3337	OP75
mdl-1	X	similar to vertebrate MAD	yDp4; mnDp30	unknown	2601 ; 3346	OP106
lin-15b	X	equivalent of mammalian Rb	mnDp10 ; mnDp1	no	3078 ; 2610	OP184
alr-1	X	hox	yDp4	yes	3156 ; 3330	OP200
nhr-32*	X	nuclear hormone receptor	mnDp10	unknown		
nhr-14*	X	nuclear hormone receptor	yDP4 ; stDp2	unknown		
nhr-173*	X	steroid zinc finger	yDP4 ; mnDp30	unknown		
elt-2*	X	GATA factor	yDP4	unknown		
pes-1	IV	HNF-3 Family	mDp4		3157 ; 3348	OP87
pqm-1	II	C2H2 zinc finger			2623 ; 3355	OP201
unc-62*	V	homeobox			3222 ; 3254 ; 3357	OP600

Transcription factors marked with an asterisk were selected because they have been shown to be highly expressed in somatic tissue (McGhee et al, 2007)

Expected Outcomes and Interpretations: We propose that increasing expression of X-linked TFs that primarily target autosomes may explain the decrease in autosomal expression seen in DCC mutants (Indirect Effect Model). If the autosomal effect is due only to increased TFs, then we should see decreased expression of a subset of autosomal genes in strains carrying an extra copy of a TF. Because our candidate TFs are predicted to be at the top of regulatory hierarchies, adding an additional copy should affect a large subset of the autosomal genes also affected by their containing duplication (fig 6). The two models are not exclusive. It may be that the titration effect due to adding extra X sequence predominates while effects of adding TFs is minor. If we see only a small subset of autosomal genes that are affected by duplication are also affected by adding a copy of the TF contained within the duplication, then we will conclude that the Titration Model predominates.

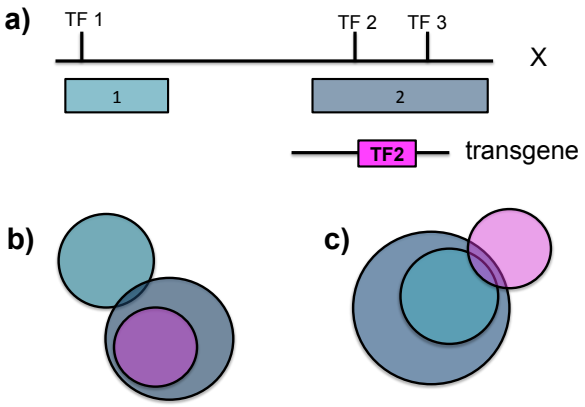

Figure 6 a) Schematic showing the X chromosome and two duplications (top) and additional transgenic copy of TF2 contained in duplication 2. **b)** In the Indirect Effect Model, duplications show little overlap and adding an extra copy of a TF has a large overlap. **c)** In the Titration Model, duplications show large overlap and adding an extra TF has a small overlap.

Potential Problems and Alternative Approaches: If we cannot get MosSCI to work for our candidate TFs or if single copy insertion is not sufficient to produce significant altered transcription of downstream targets, we can alternatively increase copy number of TFs by creating extrachromosomal arrays.

Aim 2: Determine the range over which DCC binding repress transcription.

Background and Rationale: The DCC is composed of a modified condensin complex (condensin I^{DC}) and is specifically targeted to a set of X-chromosome recruitment sites by non-condensin recruiters (Ercan et al, 2007; 2009). In *C. elegans* hermaphrodites, both X chromosomes are bound by DCC and each experiences, on average, two-fold transcriptional repression. Because DCC binds active promoters (Ercan et al, 2007) and because condensins I and I^{DC} share four out of five subunits (Csankovszki et al, 2009), it was predicted that DCC might alter chromatin structure to repress the genes that it binds to. However, current data shows little correlation between DCC binding and DCC-mediated repression (Jans et al, 2009). This has led to the hypothesis that *C. elegans* DCC does not necessarily repress the genes that it binds to but instead works over a distance to mediate transcriptional downregulation. Work in mammals has shown the existence of chromosomal "neighborhoods" formed through non-random chromosome folding in the 3D nucleus (Nora et al, 2012). If the *C. elegans* X chromosome folds into DCC-regulated neighborhoods, this might explain how binding of the DCC at one promoter affects transcription of other nearby genes. This work supports a hypothesis where DCC can act at a distance. We will test this hypothesis in two ways. First, we will determine which X-linked genes show an increase in expression when a known DCC-binding site is deleted. This will allow us also to define the longest distance over which DCC binding is necessary to reduce transcription of an X-linked gene. Second, we will use strains bearing X:Autosome fusion chromosomes, where DCC has been shown to spread ectopically to autosomal loci, to determine if nearby autosomal genes are transcriptionally repressed.

Aim 2.1: Determine if DCC binding is necessary for transcriptional repression of nearby genes.

Experimental Rationale and Design:

Deletion of DCC Binding Peaks: We will perform RNA-seq analysis in strains homozygous for deletions of X chromosome sequences known to be strongly bound by the DCC. We have a list of genes on the X that are known to be dosage compensated (Jans et al, 2009). We will determine which of these dosage compensated genes show increased expression when a single DCC binding site is deleted. We will also determine expression changes in strains carrying knockouts of autosomal sequences known to be bound by DCC. We chose available knockout strains (Table 3) whose deletion overlaps either a strong binding peak or, in one

case, a known *rex* site (Strain VC100). The deletions are homozygous viable, and predominantly delete the promoter and part of a single gene.

Table 3

Strain Name	Chr	Deleted Gene	DPY-27 Peak	SDC-2 Peak	tRNA gene	Size (kb)	Comment
VC783	X	tbc-12	N	N	chrX.tRNA17	1.4kb	deletes a X-linked tRNA
MT13653	X	mir-237	N	N	chrX.tRNA52	0.8kb	deletes an X-linked tRNA
VC1625	X	dylt-2		N	N	1.5kb	deletes a region ~ 20bp downstream of X- linked tRNA71
CB2842	IV	osm-9	N	N	chrIV.tRNA82	1.9kb	deletes an autosomal tRNA
VC100	X	many	Y	Y	N	52kb	deletes a known *rex* site and a SCD-2 peak on the X
RB1460	X	F45E1.7	N	N	N	3.2kb	deletes a region ~4kb upstream of the VC100 deletion; no known DCC binding
RB2256	X	F55A4.1	Y	N	N	0.9kb	deletes a known DPY-27 binding site on the X
RB2339	X	F53H8.3	N	N	N	0.8kb	deletes a region ~9kb upstream of the RB2256 deletion; no known DCC binding
VC1169	X	ent-2	Y	N	N	1.9kb	deletes a known DPY-27 binding site on the X
RB790	X	atf-5	Y	N	N	0.6kb	deletes a known DPY-27 binding site on the X
RB778	IV	F32E10.7	Y	N	N	3.2kb	deletes a known DPY-27 binding site on an autosome
MT16060	V	F44E7.15	N	Y	N	1.4kb	deletes a known SDC-2 binding site on an autosome
VC1255	II	vab-1	N	N	N	1.3kb	autosomal deletion; no DCC binding
RB1108	V	pmp-3	N	N	N	2kb	autosomal deletion; no DCC binding

Range of DCC regulation: In addition to genome-wide effects of deleting a single DCC binding site, we will analyze local changes in more detail. We will quantify the change of gene expression nearby and calculate the physical distance between the deleted DCC binding site and the affected genes.

tRNA deletions: We are also looking at strains carrying deletions X-linked tRNA genes. VC783 deletes chrX.tRNA17-IleAAT; MT13653 deletes chrX.tRNA52-SerAGA. ChIP-chip experiments in budding yeast have implicated RNA Pol III and tRNA gene sequences in the recruitment of mitotic condensins (D'Ambrosio, 2008). Consistent with the idea that tRNA genes mediate condensin targeting, 76 percent of all *C. elegans* X-linked tRNA genes (208/274) show a DCC binding peak within 1kB of the transcription start site (Jans et al, 2009). Further, we have recently identified a putative interaction between the condensin SMC-4 subunit and the large RNA Pol III subunit (RPC-1) by co-immunoprecipitation and western blot (Lara Winterkorn, data not shown). We predict a potential role of tRNA genes and Pol III in the recruitment of condensins to chromosomes. The X chromosome contains 44 percent of all tRNA genes in the genome. It is possible that the tRNA genes have a role in the DCC binding or function. We will perform DPY-27 ChIP-seq to determine the effect of tRNA deletion on DCC binding.

Controls to be used: To distinguish the effects of deleting the DCC binding peak and deletion of the overlapping gene, we will knockdown the gene product using RNAi and compare the resulting expression changes. Where available, we will also examine strains carrying X chromosome deletions of nearby sequence that do not contain DCC peaks (see comments, Table 3). This will distinguish deletion of a DCC binding peak from deletion of X sequence. Autosomal deletions of similar size (roughly 1kb) that lack DCC binding will be used to determine the level of background expression changes associated with small deletions.

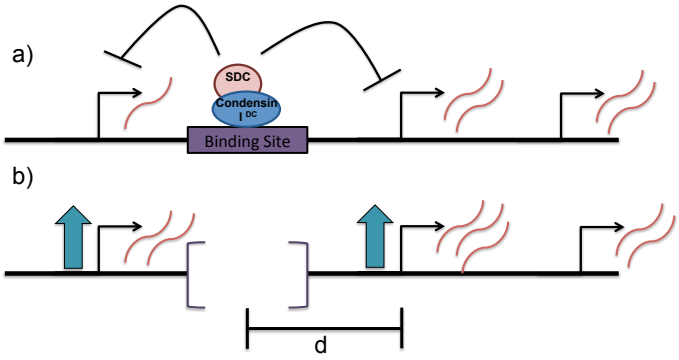

Figure 7 Predicted outcome of deleting an X chromosome DCC binding site. **a)** In the WT hermaphrodite, DCC binds to specific sites on the X chromosome and regulates transcription of nearby genes. **b)** When the binding site is deleted, transcription from nearby genes increases. **d** indicates the largest distance over which DCC is necessary for reduced X expression.

Expected Outcomes and Interpretations:

X chromosome DCC site deletions: We expect genes located near the deleted DCC binding sites will show

increased expression (fig 7). We will examine windows of various size (5kb to 100kb) to determine if dosage compensated genes contained within significantly trend towards upregulation. Dosage compensated and non-dosage compensated genes are interspersed along the X chromosome and neither group shows bias for being DCC-bound (Jans et al, 2009). Thus, we have no predictions for the distance over which DCC can act to regulate X transcription. Using our window approach, we will be able to determine the farthest distance over which DCC binding is necessary for dosage compensation. We will also compare DPY-27 ChIP seq and RNA seq to determine expression changes associated with any observed DCC mislocalization. The outcome of Aim 2.1 will reveal if DCC binding locally regulates transcription.

tRNA deletions: Because experiments in budding yeast have implicated RNA Pol III and tRNA gene sequences in the recruitment of mitotic condensins (D'Ambrosio, 2008), we predict that deletion of tRNA genes may alter the ability of DCC to properly target the X chromosome. If DCC recruitment to the X requires tRNA loci, we expect DPY-27 ChIP seq in these animals to reveal altered DCC binding pattern.

Autosomal DCC site deletions: DCC may act locally at bound autosomal loci to regulate transcription of nearby genes in a fashion similar to its function on the X. If the DCC works on autosomes as it does on the X, then we would expect to see similar expression changes as in X chromosome DCC site deletions.

Potential Problems and Alternative Approaches: It may be that DCC-mediated dosage compensation not only works at a distance but is also combinatorial. That is, multiple bound DCC may be necessary for the transcriptional regulation of any one X chromosome gene. In this case, deletion of a single DCC binding peak may not be sufficient to significantly alter transcription of dosage compensated genes. Similarly, deletion of a single tRNA gene may not be sufficient to significantly alter nearby DCC recruitment and binding. As an alternative we would cross our deletions with the *dpy-21* mutant strain (CB428). In a sensitized hypomorphic DCC background, expression changes due to deletion of a single DCC binding site may be easier to detect.

Aim 2.2: Determine if DCC binding is sufficient to induce ectopic transcriptional repression of nearby genes.

Experimental Rationale and Design: ChIP-chip analyses using two X:Autosome fusion strains indicated that, unlike X specific recruitment, spreading of the DCC occurs independent of X chromosome sequence (Ercan et al, 2009). In the fusion strains, the right end of chromosome X is fused to either the right end of chromosome V (X:V; YPT47) or to the left end of chromosome II (X:II; YPT41). In both strains, ChIP replicates gave a highly reproducible pattern of ectopic DCC autosomal binding and showed spreading as far as 3.5Mb into an autosome (fig 8) (Ercan et al, 2009). It is not known whether the spreading onto the autosomal loci result in repression of the genes at these loci. Initial experiments used qRT-PCR to measure levels of six genes close to the fusion site. Though none were significantly downregulated when compared to N2, most nearby genes showed decreased expression (Ercan et al, 2009). However, the experiment was performed in only one developmental stage and looked at only six genes. We will perform RNAseq on two X:A fusion strains (YPT41 X:II and YPT47 X:V) to determine if ectopic autosomal DCC binding leads to changed expression of nearby autosomal genes. We will analyze mixed developmental stage worms, and a mixture of L1, L2 and L3 larva. Taken together with Aim 2.1, we will have determined if DCC binding is both necessary and sufficient to reduce expression from a distance.

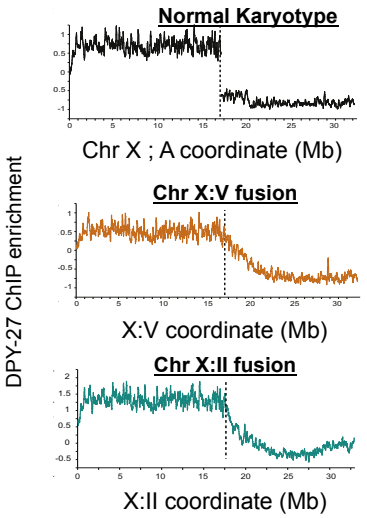

Figure 8 DPY-27 ChIP-chip in the normal karyotype and in the X:V and X:II fusions. DCC binds autosomes near the X:A fusion.

Controls to be used: The fusion strains are not directly comparable to N2. We will need to distinguish autosomal expression changes due to fusion to the X from those due to ectopic DCC binding. We will use RNAi to knockdown DPY-27 in the fusion strains. We will consider dosage compensated genes to be those that show increased expression in the DPY-27 knockdowns.

Expected Outcomes and Interpretations: As described above, genes nearby to ectopic autosomal binding were shown to have slightly decreased expression in the fusion strains when compared to N2. As such, we expect our RNA seq data to reveal decreased expression of autosomal genes near the fusion site. DCC binding of ectopic autosomal loci decreases as a function of distance from the nearest X chromosome recruitment site (fig 8). We will determine if the strength of transcriptional downregulation at these loci also decreases as a function of distance. If we see no decrease in autosomal expression in the fusion lines, and if there is no effect on local transcription when we delete a normal autosomal binding locus (see Aim 2.1), then we will conclude that DCC-mediated two-fold downregulation is specific to the X.

Aim 3: Determine the effect of X chromosome location on a gene's transcriptional regulation.

Background and Rationale: In *C. elegans*, the condensin-like DCC mediates two-fold reduction of X expression in hermaphrodites, balancing X expression between XX hermaphrodites and XO males. However, both sexes now have only one functional dose of the X chromosome compared to two copies of every autosome. A second mechanism of dosage compensation (proposed to have evolved prior to female/hermaphrodite-specific compensatory mechanisms (Ohno, 1967)) predicts upregulation of the X chromosome to balance X and autosomal transcription. Microarray and RNA seq data, have shown overall levels of X transcripts to be comparable to autosomal transcripts and support the existence of such a mechanism.

However, owing to biased gene content and tissue-specific regulation of the X, direct comparison of X and autosomal transcription is difficult. The X chromosome of mammals shows tissue specific transcription and highly skewed gene content. It is highly enriched for reproduction-specific genes that are not expressed in somatic tissues (Deng et al, 2011; Khil et al, 2004; Mueller et al, 2008). RNA seq experiments in mice revealed the X chromosome to be enriched for genes that are not expressed (Deng et al, 2011). In *C. elegans*, the X chromosome is silenced in most of the meiotic germline (Kelly et al, 2002). In an adult worm, with roughly 1000 somatic cells and 2000 germ cells, X chromosome expression can be grossly underestimated.

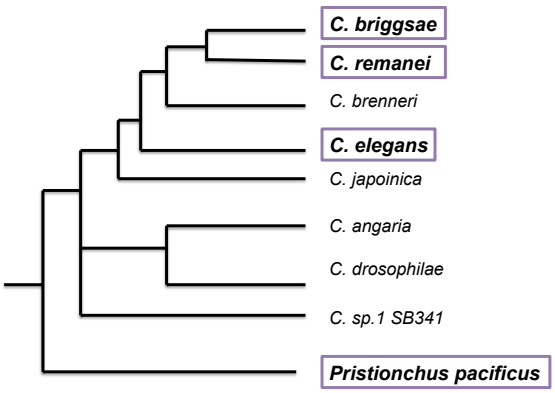

Figure 9

In order to directly compare X and autosomal transcription we propose to look at expression of 1:1 orthologs that are differentially located on the X or an autosome between two nematode species. We will work with four species: *C. elegans, C. briggsae, C. remanei,* and *Pristionchus pacificus* (fig 9). *C. elegans* and *C. briggsae* were the first pair of sister species genomes to be sequenced (Stein et al, 2003). They are predicted to have split roughly 100mya (Gupta et al, 2007; Stein et al, 2003). Recent work with *P. pacificus* predicts its split from the *Caenorhabditis* genus 200 to 300mya (Gutierrez and Sommer, 2004; Dieterich et al, 2008). We note, however, that time estimates of divergence are exceptionally difficult in the nematode lineages as they lack a fossil record and experience variable generation times due to ability to enter dauer stage (K. Kiontke, personal communication; Srinivassan and Sternberg, 2008). The *C. elegans* and *C.* briggsae genomes are well assembled and annotated. The genomes of *C. remanei* and *P. pacificus* have been previously sequenced, but are not yet fully annotated. Therefore, it was not previously known which genes are located on the X chromosome in these species.

Gene movement both within and between chromosomes is seen frequently in all eukaryotes (Bhutkar et al, 2007; Coghlan and Wolfe, 2002; Drouin et al, 2002; Emerson et al, 2004). In many organisms, there is a clear bias towards gene movements involving the X chromosome. X to Autosome movements are 3.5X higher than expected in mammals and roughly 2X higher in *Drosophila* (Betran et al, 2002; Emerson et al, 2004; Moyle et al, 2010). The bias towards X-A movement has not been documented in *C. elegans*. A comparative genomic study between *C. elegans* and *C. briggsae* identified 312 1:1 orthologs that have changed chromosomal locations. Of these, only 39 (12 percent) involved the X chromosome. This is less than the ~17 percent expected if gene translocation events were random. The authors further noted that conservation of synteny was highest for the X chromosome: 97% of orthologs retain X chromosome location in the two species (Hillier

et al, 2007). It has been proposed that X-involved translocations may be deleterious in *C. elegans* as they interrupt mechanisms of dosage compensation (Coughlan, 2005).

Preliminary Data:

X assignment: To map genes to the X chromosome, we took a copy-number-variation approach. We performed genomic DNA-seq in males and females/hermaphrodites of *C. remanei* and *P. pacificus*. Genes located on the X chromosome were expected to have a 1:2 ratio of sequencing coverage between males and hermaphrodites (X:XX) and all autosomal genes a 1:1 ratio. We used *C. briggsae* as a blind control in our analysis (fig 10). Compared to the WormBase cb3 annotation, we correctly assigned 3760/3775 (99.6%) X-linked and 17435/17480 (99.7%) autosomal genes. We misassigned only 6 genes to an autosome. These 6 genes are all contained in a single contig (cb25.fpc2301b) which is currently assigned to an un-assembled region of chromosome X. It is worth noting that contigs cb25.fpc2301a and cb25.fpc2301c are both currently assigned to chromosome V. It is, therefore, likely that our autosomal assignment of these 6 genes is correct. There were 54 genes that we could not unambiguously assign to either X or autosomes. We are currently collaborating with the Dieterich lab and preparing SNP mapping libraries to further fine-tune the *P. pacificus* assignment.

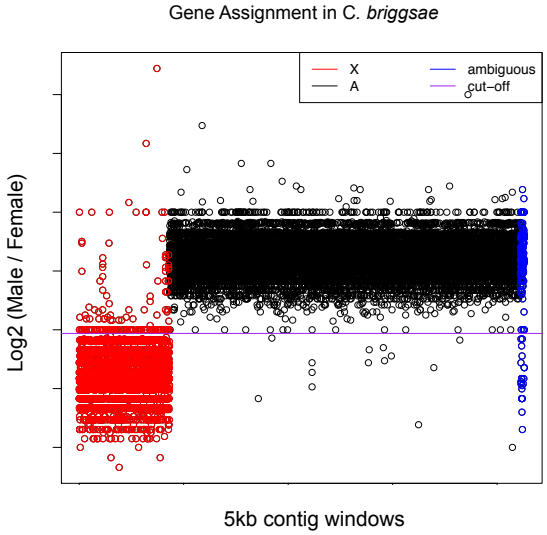

Figure 10 Each point represents a gene. 5kb windows are plotted along the X axis. Male to Hermaphrodite coverage ratios are plotted on the Y axis. Windows have been grouped by their assignment. Red = X ; Black = Autosome; Blue = Ambiguous

One-to-one ortholog identification: Using our gene assignments and Compara orthology predictions (Vilella et al, 2009), we have identified orthologs that have moved between X and autosomes. Between any two species, the number of such orthologs ranged from 28 to 474 (Table 4). There are roughly 400 orthologs that are on an autosome in *P. pacificus* and an X in all of our *Caenorhabditis* species. Our analysis confirms a predicted translocation event whereby one arm of *P. pacificus* chromosome I moved to the X in *Caenorhabditis* (Sommer, 2006).

Aim 3.1: Determine if orthologs on the X chromosome are dosage compensated.

Experimental Rationale and Design: We will perform RNA-seq to directly compare expression of 1:1 orthologous genes that are differentially located on the X and an autosome. Using the "ortholog functional conservation hypothesis", we will assume that 1:1 orthologs should retain roughly the same function and expression between two species (Altenhoff and Dessimoz, 2009; Forslund et all, 2011; Thomas et al, 2012). We will compare expression of these differentially located 1:1 orthologs both between species (to determine if expression is compensated between the X and the autosome) and between sexes (to determine if expression is compensated between the sexes). We will perform an initial analysis of orthologous expression in handpicked populations of young adult males and females/hermaphrodites. Here, we will focus on those genes known to be expressed only in the soma (355 genes in XO males, 420 genes in XX hermaphrodites (Reinke et al, 2003)) because in the germline, the X chromosome is repressed independent of somatic dosage compensation. A subsequent goal of this project is to perform single embryo and single L2-stage RNA seq. This will abrogate all potential germline contamination problems and allow for expression

		Autosomes			
		C elegans	C briggsae	C remanei	P pacificus
X chromosome	C elegans	--	25	12	410
	C briggsae	19	--	21	414
	C remanei	16	21	--	399
	P pacificus	58	60	46	--

Table 4
Pairwise comparisons of orthologous gene locations.

comparisons between all identified orthologs.
Controls to be used: We will compare expression of 1:1 orthologs that have retained their chromosomal locations (either autosomal or X) between two species. This internal control will show the extent to which orthologous genes show similar expression in our different species. Of particular interest will be comparing expression of orthologs that have moved between autosomes in *C. elegans* and *C. briggsae*.

Expected Outcomes and Interpretations: We can predict four scenarios of X regulation when comparing orthologous expression pairwise between sexes and species (fig 11). In the first, there are no mechanisms of dosage compensation acting on the X-linked ortholog. If the ortholog is on the X in species A (sp.A) and on an autosome in species B (sp.B), then sp.A hermaphrodites and both sexes of sp.B will have two copies of the gene while males of spA will have a single copy. With no X-specific regulation, we would expect to see an expression ratio of 1:1 between hermaphrodites and 1:2 between males. In the second scenario, only DCC-mediated X downregulation acts on the ortholog. This might represent an intermediate evolutionary step where the X is downregulated specifically in hermaphrodites and the ortholog must still evolve sequence changes to allow for up-regulation. Here, we expect the expression ratio between hermaphrodites and between males to be 1:2. In the third scenario, the X chromosome is upregulated in both sexes, but not repressed in hermaphrodites. This might represent an intermediate evolutionary step where the X chromosome is always upregulated and must evolve sequence changes to allow for hermaphrodite-specific DCC binding and downregulation. With only X upregulations, we would expect the spA:spB expression ratio to be 2:1 between hermaphrodites and 1:1 between males. Finally, if there is both DCC-mediated downregulation and X specific upregulation then we should see 1:1 expression ratios for both males and hermaphrodites. In all four scenarios, we will also determine the male:hermaphrodite expression ratio. We expect that dosage compensation should balance expression between the sexes for all orthologs that are not sex-specific.

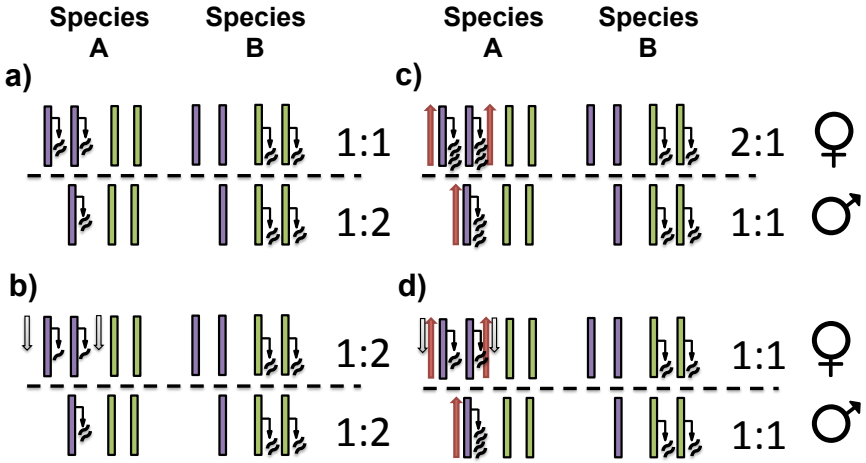

Figure 11 Four possible mechanisms of X dosage compensation acting on a 1:1 ortholog. X chromosomes are shown in purple, autosomes are in green. The ortholog is on the X in species A and an autosome in species B. **a)** No dosage compensation. **b)** Hermaphrodite-specific X downregulation. **c)** X-specific upregulation. **d)** X-specific upregulation coupled with hermahphrodite-specific X downregulation.

Potential Problems and Alternative Approaches: We are able to obtain ~ 1ug of quality RNA out of 300 worms from all four species and have experience making Illumina RNA seq libraries and analyzing the data. For single worm comparisons, it is possible that we may not be able to linearly amplify RNA from a single nematode embryo. In this case, we will first try pooling multiple RNA collections. We will collect RNA from single embryos, determine sex by RT-PCR, and pool male and female collections. In the unlikely case that pooling will not give us enough RNA to prepare sequencing libraries, we will limit our analysis to those orthologs that are not expressed in the germline. Further, we will make immediate use of *C. elegans* and *C. remanei* male and female/hermaphrodite larval and young adult RNAseq datasets available through ModENCODE (SRA Accession No. PRJNA33023 and PRJNA75295) and begin to compare expression ratios of somatically expressed 1:1 orthologs.

Aim 3.2: Determine the DNA sequence changes associated with X chromosome dosage compensation.

Experimental Rationale and Design: Mechanisms of X chromosome dosage compensation can lead to sterility of hybrid embryos due to misexpression of X-linked genes (Orr, 1989; Moyle et al, 2010). Further, any gene that moves to the X chromosome, unless it is sex-specific or dosage-independent, must become dosage compensated. An ortholog moving to the X chromosome would find itself in only single copy in males. It has been proposed that selection may act on these genes to favor promoters or enhancers with that give increased

expression (Nguyen and Disteche, 2006). Increased X expression would then need to be balanced by hermaphrodite-specific downregulation. In *C. elegans,* we hypothesize that sequences evolve to promote both upregulation (in both sexes) and recruitment of the DCC (active only in hermaphrodites to produce two-fold reduction in expression). We will test this hypothesis by analyzing sequence changes associated with X-located orthologs. From our previous work (see above) we have identified X-linked sequences in all four species. Additionally, we have a working list of all 1:1 orthologs between any two species. Following our Aim 3.1, we will also have identified X-linked orthologs that are and are not dosage compensated.

Expected Outcomes and Interpretations: To identify DNA sequence changes outside of coding regions, we will look 1Kb upstream and downstream of the translocated orthologs and identify shared DNA sequence motifs using MEME (Bailey and Elkan, 1994) that are enriched when the ortholog is on the X. As a control, we will compare regulatory sequences for orthologs that are on the X chromosome in both spA and spB to sequences for orthologs that are on an autosome in both species. We expect that any regulatory motifs enriched only in X-linked orthologs should have evolved to confer increased gene expression and/or recruitment of a DCC. In *C. elegans*, X chromosome promoters show higher GC content when compared to autosomal promoters (Ercan et al, 2011). We will determine if promoters of genes that moved to the X show this same bias.

Analyses in both *Drosophila* and *C. elegans* have shown the X chromosome to have a have a high codon use bias (CUB) (Singh et al, 2005). This bias is predicted to increase the fidelity and/or the efficiency of translation (Akashi et al, 1998; Bulmer et al, 1991; Marais et al. 2001, 2003). Two hypotheses have been proposed to explain the X chromosome CUB. The first is that the increased effect of positive selection acting on the X. Because the X is in single copy in males, X linked genes can evolve more rapidly, which could allow for increased CUB (Charlesworth et al, 1987; Singh et al, 2005). The second explanation for the observed bias is that CUB helps to compensate for the X chromosome imbalance. In *Drosophila*, it has been proposed that CUB increases efficiency of translation and helps to reduce any deleterious effects that may be caused by reduced X dosage in males (Singh et al, 2005). We will determine the frequency of optimum codons (Duret and Mouchiroud, 1999) in 1:1 orthologs that are differentially located on the X chromosome. We will compare this to the frequency of optimum codon use on the X chromosome as a whole. If the movement of an ortholog to the X chromosome is a more recent event, we expect that the CUB may be lower than the X chromosome average. We will also determine if any orthologs that "escape" dosage compensation (ie show no expression change in a DCC mutant) show strong CUB. This may indicate that dosage compensation of these orthologs occurs post-transcriptionally.

Aims and Activities	Year 0.5	Year 1	Year 1.5	Year 2	Year 2.5	Year 3
Specific Aim 1						
1.1: Test Titration Model: RNAseq in duplication strains.						
1.2: Test Indirect Effect Model : MosSCI of TFs						
Specific Aim 2						
2.1: Determine if DCC is necessary locally: deletions of DCC binding peaks						
2.2: Determine if DCC is sufficient to reduce autosomal transcription: ectopic autosomal DCC binding in X:A fusions						
Specific Aim 3						
3.1: Determine if X orthologs are dosage compensated: Compare expression between 4 nematode species						
3.2: Determine sequence changes associated with X dosage compensation: Analysis of DNA seq data (already generated)						