ABSTRACT

SNYDER, DANIEL WATCHORN. Analysis of Genes Expressed by *Meloidogyne incognita* Males: Generation of ESTs and Comparative Genomics. (Under the Direction of David McKenzie Bird)

Over the last decade genomic and molecular research in model systems including *Ceanorahbditis elegans* have generated a body of knowledge that has revolutionized the study of plant parasitic nematode biology and disease. *Meloidogyne incognita* is one of the most devastating plant parasitic nematodes worldwide and has proven to be a prime example of the utility of comparative genomics in the investigation of plant diseases. The role of males and sex determination in this species presents a particularly interesting evolutionary and biological issue but is poorly understood at the molecular level. A method for generating and collecting male *M. incognita* tissue that is free of other life stages or females and other contaminating organisms was developed. The method confirmed previous results that host-pruning stress stimulates male sex differentiation and that the effects of this stimulus are limited to the time period of early second stage larvae feeding. Approximately 5,000 ESTs were generated by 5' sequencing of cDNAs from four male *M. incognita* libraries constructed from the tissue obtained above. Two cDNA libraries were constructed by utilizing a nematode splice leader (SL1) and two libraries were constructed with a universal template switch method. The ESTs from these libraries were clustered for each library and as a complete set from all four libraries providing insight to the composition and abundance of genes expressed by these males. Searching the public databases for homologous sequences providing a second degree of clustering

and provided the putative identification of genes identified by these clusters. Approximately 89% of the clusters had significant homology to the public databases and approximately 31% of these had homology to *C. elegans*. Additional annotation of these genes was conducted using several public resources providing further insight to the molecular basis of male *M. incognita* biology and comparative analyses to other stages and nematode species. Protein phosphatases, transthyretin-like families and major sperm proteins are some of the most abundant sequences in these male libraries and are also highly expressed in *C. elegans* males. These results indicate that not only are these *M. incognita* libraries representative of male gene expression, but that male gene expression profiles may be similar across families.

ANALYSIS OF GENES EXPRESSED BY *MELOIDOGYNE INCOGNITA* MALES: GENERATION OF ESTS AND COMPARATIVE GENOMICS

by
DANIEL WATCHORN SNYDER

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Masters of Science

PLANT PATHOLOGY

Raleigh

2004

APPROVED BY:

_____ _____

_____

Chair of Advisory Committee

BIOGRAPHY

Daniel Watchorn Snyder was born on November 21, 1972 in Ft. Lauderdale, Florida, United States of America. He is the son of Johanna Luther Snyder and Philip Melancthon Snyder III. He obtained his elementary education in Chapel Hill, NC and graduated high school from Virginia Episcopal School in Lynchburg, VA in 1991. He then enrolled at Guilford College in Greensboro, NC where he studied Physics, Mathematics and Biology obtaining his B.S. in Biology in May of 1995. The following year he was hired by George Miksch Sutton Avian Research Center where he studied breeding behavior and population biology of neotropical migrant passerine species of the tall-grass prairies of North America. On year later he joined the Community Ecology Lab of Jim Clarke at Duke University, Durham, NC to study the effects of elevated CO_2 on understory growth response in temperate evergreen forests. While working at Duke University he also volunteered at the Duke Primate Center helping to plan and coordinate a prerelease behavioral study and conditioning of black and white-ruffed lemurs as part of the first ever reintroduction effort into protected national parks of Madagascar. After two years he was employed by the Zoology Department of Duke University to manage the DNA Sequencing Facility serving the Botany, Zoology and Biological Anthropology and Anatomy Departments. After working for one year in the Department of Plant Pathology at North Carolina State University for Dr. David McKenzie Bird he enrolled as a graduate student in 2000 where he has been pursuing his M.S. in Plant Pathology as a member of the Center for the Biology of Nematode Parasitism.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The members of the *Meloidogyne* genus (root-knot nematodes) are responsible for billions of dollars of economic loss to worldwide commercial crop production each year on over 5,000 host species (Sasser and Freckman, 1987). One of the most significant factors that determines the amount and severity of disease caused by these nematodes is the ratio of males to females in the adult population. Females contribute more to root-knot nematode disease than males for two primary reasons. Females feed more vigorously and for a longer period than males, presumably due to the increased energetic demands of producing and laying eggs. Secondly, the growth of the population from generation to generation is directly determined by the number of females. The sex ratio is controlled in all gonochoristic species (sexes in separate individuals) through the processes of sex determination, differentiation and development which result in the production of two distinct sexual forms. Each of these systems is therefore critical for understanding how the sex ratio of root-knot nematodes is controlled and even potentially manipulated. One obvious application of this knowledge is the development of biological control strategies based on the manipulation of any of these processes by shifting the ratio toward the production of males. The efficacy of this strategy depends on the mode of reproduction of the species to be controlled and the availability of responsive factors in the process of sex development. In the potato-cyst nematodes *Globodera rostochiensis* the sex ratio is shifted towards male development in response to the H_1 resistance gene (Forrest et al., 1986). H_1 resistance is utilized for control of potato cyst nematode worldwide and provides complete control, on its own, against pathovars in North America (Cook and Evans, 1987). The rapid selection of resistance breaking pathovars in Europe has been attributed to the sexual reproduction of these species, making this strategy more effective in clonal species

such as *M. incognita*, *M. javanica* and *M. arenearia*. The current knowledge base for these processes in root-knot nematodes is primarily limited to field studies, anatomy and cytogenetics. These studies provide insight as to the biological basis for sex and development, but little is known about the genetic or molecular basis of this biology. In several model systems such as the fruit fly (*Drosophila melanogaster*), the nematode (*Caenorhabditis elegans*), and humans the genetics and molecular basis of sex have not only generated a nearly complete picture of how these systems operate in metazoans, but also provide an indispensable resource of data for the investigation of these systems in other organisms like root-knot nematode. By integrating information from these highly annotated model systems and the growing body of data from root-knot nematodes through comparative genomics a greater understanding of the role and evolution of sex in these organisms will be obtained.

Evolutionary Significance of Reproductive Strategies

There has been considerable argument about the supposed advantages of sexual reproduction that explain its ubiquity among metazoan life (see Barton and Chalresworth, 1998 for a review). Indeed the role of sex in the evolution of species is a central one with particular relevance in the evolution of parasitism. It is commonly believed that sexual reproduction provides increased genetic variability for adaptation by natural selection and thus relatively more rapid evolution than possible through asexual reproduction. Within root-knot nematodes individual species reproduce by obligate amphimixis (sexual reproduction), facultative amphimixis, and obligate apomixes (mitotic parthenogenesis) (Triantaphyllou, 1973). Many root-knot nematodes including the most devastating species, *M. incognita*, *M. hapla* (race B) *M. arenaria* and *M. javanica*, reproduce by mitotic parthenogenesis, and

amphimixis is thought to be the ancestral mode of reproduction from which mitotic parthenogenesis arose (Triantaphyllou, 1985). It could be argued that asexual reproduction under uniform environmental conditions such as those found in monoculture cropping is optimal for population growth with no loss of energy to the production of males. Conversely, the advantages of sexual reproduction under stressful or changing environmental conditions may justify the expense of producing the male phenotype for the chance to generate a new genotype better suited to the current environment. Therefore it might make sense that asexuality may have arisen in response to these uniform host environments. It has also been shown that sexual species are more susceptible to invasion by asexual variants than the reverse case (Smith, 1971; Smith, 1978), but models for maintenance of sexual reproduction are harder to validate (Reik and Surani, 1997). In the special case of host parasite interactions there may be sufficient selective advantage for sexual reproduction of the parasite for continued adaptations in the cycling between host resistance and parasite virulence (Otto and Michalakis, 1998; Nee, 1989; Hamilton, 1980). Phylogenetic analysis of nematodes supports this evolutionary model by placing predominantly amphimictic genera including *Heterodera* and *Globodera* closer to the ancestral origin or root-knot nematodes (De Ley et al., 2002), although it is known that parthenogenesis has arisen independently several times throughout plant parasitic nematode evolution (Blaxter et al., 1998). The arrival of these parthenogenetic species is accompanied by increased variation in ploidy including polyploidy and aneuploidy (Triantaphyllou, 1985). It is possible that one or more components needed for development of functional males was lost in the lines that became asexual but it is also possible that the gain of facultative independence from sexual reproduction has allowed these seemingly asexual species to arise. With further genomic and molecular analysis in root-knot nematodes

it will be possible to dissect these questions and further understand the role of reproduction in this genus and its implications for disease control.

Conservation and Variation in Sex Determination

Sex determination is fundamentally the process of triggering sexually dimorphic development of male and female phenotypes and establishing the primary sex ratio of the population. An interesting trend is emerging from molecular and comparative studies of sex determination indicating that basal components of sex determination are highly conserved across kingdoms at the pathway and even molecular level, whereas the primary signals and mechanisms have diverged considerably (Stothard and Pilgrim, 2003). The similarities observed between sex determination mechanisms may reflect conserved elements of these pathways, but in many cases it seems the function is the only thing conserved and the elements of these systems have diverged considerably. The evolutionary flexibility observed in sexual differentiation is greater than any other developmental pathway. This is particularly surprising, due to the fact that sex determination is no less complex a regulatory system than other developmental regulation systems. These trends not only provide a guide for the investigation of sex differentiation in on other nematodes, but also demonstrate the uniquely adaptive nature of these processes.

Mammalian sex determination such as exhibited in the human and mouse models probably represent the most derived and complex systems known in such detail and thus provide an extreme for comparison to the nematode system. Although greater similarity is seen in the more closely related fly and worm models, the trends in evolution of this pathway would not be apparent without our knowledge of mammalian systems. In vertebrates, sex determination is divided into two stages: a primary stage, which is concerned with the

determination of gonad sex, and a secondary stage, which comprises the determination of somatic sex. This division is primarily due to the hormonal nature of vertebrate sex determination which is consistent between mammals, birds, retiles, amphibians and fish. There are two distinct forms of primary sex determination seen in vertebrates. These are the genotypic determination employed by all mammals and environmental determination seen in some species of reptiles, amphibians and fish. As discussed later, nematodes also utilize both environmental and genotypic modes of sex determination, sometimes even within the same species. The somatic sex is usually determined by the presence of either the X or Y sex chromosome where XX is female and XY is male, yet birds are the exception with the matched chromosome complement determining the male phenotype (Smith and Sinclair, 2001). Hormones expressed during the bipotential (indifferent) stage of gonad development determine the sexual fate of the remainder of the adult soma. Without the Y chromosome ovaries develop and produce estrogen which drives feminization of the rest of the human soma including the development of the uterus and oviducts from the Millürian duct (Couse and Kourach, 2001).

A number of genes required for normal sexual differentiation in humans have been identified through clinical studies of individuals carrying mutations of these genes that often result in sterility. Many of the primary elements of this pathway are transcription factors including sry, sox9, SR-1, wt1, and dax-1, whereas the remainder of sex determination is controlled by hormones secreted by the sex-determined gonads. The first gene in the vertebrate pathway is the well characterized gene *sry* (sex-determining region of the Y chromosome) which is the major gene for normal testis formation. In fact, a homolog to this gene has been identified in *D. melanogaster* and plays a similar role in fly sex determination.

The gene encodes a 223 amino acid protein that probably functions as a transcription factor as it contains the HMG (high-mobility group) box, which has a DNA-binding domain (Sinclair et al., 1990). The HMG domain has been shown to bind to DNA and induce a bend in the DNA at the position where it is bound (Giese et al., 1992). A second gene, the autosomal gene *sox9* has been found in all vertebrates and triggers testis formation in the absence of *sry* and can apparently replace *sry* for male development. Sry may act to trigger sox9 and initiate a conserved evolutionary pathway since *sry* is found only in mammals (Pask and Graves, 1999). Sox9 can act as both a transcriptional regulator and splicing factor. It binds to the promoter of the gene for anti-Millürian hormone to switch on the male developmental pathway (de Santa Barabara et al., 2000) and has been shown to exhibit RNA splicing functions (Ohe et al., 2002). Sex-specific splicing is a regulatory mechanism utilized in all metazoan sex determination systems and likely demonstrates how conservation of function exists even though the elements are highly diverged. Fgf9 (fibroblast growth factor 9) is a chemotactic factor that is induced by either sry or sox9 to permit the migration of mesonephric cells into the XY gonad and induce testis formation (Colvin et al., 2001). The levels of AMH (anti-Müllerian hormone) and testosterone are regulated by a transcription factor SF1 (steroidogenic factor 1) for the normal development of testis by elevating the expression of these genes in the XY gonad (Arango et al., 1999). It is believed that sry directly regulates the level of *SF1* expression, which regulates the levels of AMH and testosterone. Recent studies in mice indicate the necessity of the insulin signaling pathway for normal masculinization of gonads into testis (Nef et al., 2003). This is particularly interesting in that this pathway is present in *D. melanogaster* as well as *C. elegans* and may constitute a conserved mechanism for sexual development in all animals, presumably

including root-knot nematodes. In 1994 Bardoni and her colleagues described a potential testes-suppressing gene located on the X chromosome named *dax1*. It was discovered in a duplicated region of the X chromosome in two females who were genetically XY. It is believed that dax1 competes with sry as they are expressed in the same cells in mice and humans at the same time in males and females, but by day14 is only seen in XX gonad. After being cloned it was found that *dax1* encodes a member of the nuclear hormone receptors and is an antagonist of sry and sox9 as well as a negative regulator of *SF1* (Zanaria, 1994; Swain et al., 1998). In contrast to dax1, a potential ovary-determining factor called wnt4 is necessary for normal ovary development and may interact directly with the RNA polymerase subunit $TAF_{II}105$ (Frieman et al., 2002). *Wnt4* may be repressed by sry in XY individuals to form testis or indirectly by sox9.

Sex determination in the fruit fly is determined by sex chromosomes but in a different way from that seen in mammals. Here the ratio of the X sex chromosome to autosomes is the primary signal and the Y chromosome plays no role in sex determination at all. Instead, the Y chromosome carries genes for maleness that are not expressed until late in development of the adult, such as sperm formation. Therefore, an XX individual becomes female and an XY individual becomes a male, because the X:A ratios are 1:1 and 1:2 respectively (Bridges, 1925). This ratio is weighed between transcriptional activators on X and suppressors on the autosomes that regulate the expression of the gene s*ex-lethal* (*sxl*). These transcriptional regulators compete for the early promoter of the *sxl* gene, which is autoactivated by the binding of the sxl protein to its own late promoter (Bell et al., 1991). The genes encoded on the autosomes encode proteins such as deadpan and extramacrochaetae which block the binding of the transcriptional activators encoded by the X chromosome such as sisterless-a

and sisterless-b (Van Doren, 1991; Younger- Shepard, 1992). The sxl protein is detected in the female embryo only 2 hrs after fertilization as it plays the pivotal role in fly sex determination (Salz et al., 1989). The late promoter of *sxl* is turned on shortly after the early promoter and two different *sxl* mRNAs are produced from this late promoter in males in females, but only the female form is functional producing the sxl protein (Bell et al., 1988). Female sxl protein may even assist its own proper splicing, whereas males have no sxl protien and thus their mRNA is spliced in the male-specific manner and never produces functional sxl protein (Keyes, 1992). If *sxl* is active it triggers the alternative splicing of the female-determining double-sex protein (dsx), which activates genes involved in development of the female phenotype. There is also a male form of dsx that is formed by the 'normal' splicing of the *dsx* pre-mRNA, which then triggers the expression of masculinizing genes. The male form of the sxl protein includes an extra exon that contains a termination codon responsible for the early termination of translation making the protein non-functional. This premature termination results in the loss of RNA binding domains essential to its function in downstream sex regulation. One of the RNA targets for sxl is the pre-mRNA of the *transformer* (*tra*) gene which is the next gene in the sex determination pathway (Nagoshi et al., 1988). *Tra* is also alternatively spliced in the two sexes with the male transcript again containing an exon with an early termination codon. In this case it is the second codon that is alternatively spliced and not the third exon as in the case of *sxl*. The real switch for sex determination is made with the final element of this regulatory cascade, the *doublesex* gene (*dsx*). The tra protein present only in females works with the tra-2 protein found in both sexes to process the *dsx* primary transcript in a sex specific fashion. The male-specific mRNA contains two terminal exons which are not present in the female-specific *dsx* due to

deactivation of the 3'splice site from exon 4 by tra and tra-2 (Ryner and Baker, 1991). It is thought that the male-specific form is the default form of *dsx* as it does not require the activity of these upstream elements and thus may be the ancestral signal for male development in *D. melanogaster*. The *doublesex* gene from *D. melanogaster* shows striking similarity to a gene called *mab-3* in *C. elegans*, both in sequence and function. These two genes occupy downstream positions in their respective pathways and are the only case of commonality in the two systems (Raymond et al., 1998).This scenario further bolsters the theory of retrograde evolution where shared pathways with common basal elements diverge by the alteration of upstream components (Wilkins, 2002).

There are a number of striking similarities between the sex determination system seen in the fly and that seen in *C. elegans*. In *C. elegans,* sex is determined by a regulatory network that begins with a primary X-linked signal that responds to the X:A chromasome ratio, as in the fly system. The primary signal in *C. elegans* is called *xol-1* (*XO lethal*) and is upregulated in males and downregulated in the hermaphrodite. Dosage compensation and sex determination are coordinately controlled in *C. elegans* by a set of hermaphrodite-specific genes called *sdc* (*sex and dosage compensation*) (Villeneuve and Meyer, 1987; Nusbaum and Meyer, 1989). A second set of genes called *dosage compensation dumpy (dpy*) function to downregulate expression of genes on both the X chromosomes in XX hermaphrodites. A pair of genes on the X chromosome, *feminizing factor on X* (*fox-1*) and *signal element on X* (*sex-1*) serve as the signal elements by which the X:A ratio is measured, yet no autosomal elements have been identified for the other half of this ratio. These two genes function to reduce the expression level of *xol-1* and initiate female sex determination (Nicoll et al., 1997; Carmi et al., 1998). *Fox-1* encodes an RNA binding protein, but has no effect on the

transcription of *xol-1* and may exert its regulatory effect in a post transcriptional fashion. The much stronger regulator sex-1 has demonstrated transcriptional suppressive effects on *xol-1* and encodes a nuclear hormone receptor. Thus if there are two copies of the X chromosome the dose of these two X-chromasome elements is strong enough to suppress *xol-1* expression, whereas one X-chromosome is insufficient for *xol-1* suppression. At this point the embryo is still in the 28-cell stage before gastrulation has even begun and the sexual fate of the nematode has already been determined. The level of *xol-1* expression in response to the dose of *sex-1* and *fox-1* now serves to regulate the expression of the three *sdc* genes. If *xol-1* has been effectively suppressed by fox-1 and sex-1 in the XX embryo hermaphrodite determination continues with the activation of the *sdc* genes, but they are suppressed in the XO embryo by xol-1 activity. As their name implies, the *sdc* genes at this point in development not only act to perpetuate the sex determination signal, but they begin the developmental process of dosage compensation. In the nematode the dosage of X-linked gene expression is compensated by down regulation of genes on both X chromosomes by one half in XX individuals. This is in contrast to the human system where an entire X chromosome is inactivated or the fly system where the dose of X-linked genes in males is doubled. This method of dosage compensation may constrain variation in this system relative to that seen in the fly where the sexual fate of each cell is independent and mutations are rarely lethal. From here the level *sdc* gene expression in the XX individual is sufficient to suppress the expression of her-1 which is necessary for activating the transmembrane gene *tra-2*. Tra-2 in turn suppresses the *fem* genes and allows tra-1 to trigger feminization of the nematode soma concluding the regulatory process of sex determination in the hermaphrodite. In the case of the XO embryo her-1 is sufficient to suppress *tra-2* allowing the *fem* genes to

suppress *tra-1* in which case the male developmental process is initiated and determination of the somatic sex for the male is concluded. Additional evidence has shown that the regulation of *tra-2* requires the activity of additional elements in both the XX and the XO scenario. In XO individuals, *tra-2* suppression is enhanced by laf-1 and in the XX case *tra-2* is activated by tra-3.

Nematode Sex Determination

In nematodes, *C. elegans* is clearly the most well studied sex differentiation system, yet it represents only one of the myriad systems utilized by nematodes. Sex determination is typically XX/XO chromosomal as in *C. elegans* (Walton, 1940), but XX/XY chromosomal is also found in parasitic species such as *Ascaris lumbricoides*, *Brugia malayi*, *Onchocerca volvulus* and *Baylisascaris transfuga* (Goldstine, 1981; Sakaguchi et al., 1983; Harai et al. 1985; Mutafova, 1995; Underwood and Bianco, 1999). An interesting scenario is presented in *Stongyloides ratti* which alternates between amphimixis as free-living adults and mitotic parthenogenesis as parasitic females (Viney et al., 1993; Viney, 1994). Sex determination occurs twice in both forms of reproduction triggered by environmental and host conditions, with the first determination for male/female development and the second for parasitic parthenogenetic or free-living amphimictic females (Harvey, 2000). In most amphimictic species of root-knot nematodes such as *M. graminis* the sex ratio is less sensitive to environmental factors than in parthenogenetic species (Triantaphyllou, 1973). Sex determination in *M. incognita* and *M. javanica* is tightly linked to host and environmental conditions which may be paradigm for parthenogenetic species. A wide array of environmental and host conditions have been demonstrated to drive sex determination towards male biased differentiation. These conditions include crowding, soil temperature,

application of foliar herbicides on the host (Davide, 1965), and the presence of natural (Moura, 1993) and synthetic (Williamson and Hussey, 1996) resistance in the host. Additionally, the timing of sex determination during feeding by the second stage larvae (L2) implicates the signaling between the host and the L2 as being primary cues for male versus female development (Davide and Triantaphyllou, 1968). During a normal infection, L2 establish a fixed feeding site in the vascular cylinder, which it transforms into a cluster of 4-9 highly modified plant cells termed giant cells. Most of the nutrition needed for the nematode to develop into an adult is derived through these giant cells during the L2 stage (Agrios, 1997). Under favorable feeding conditions females develop from infective larvae and the differentiation of males is considered to be a rare event. Under less favorable feeding conditions a larger number of males develop and may comprise as much as 90% of the adult population. Females are 700 μm long, sedentary and globose in shape, while males molt from the forth larval (L4) stage as a 1.4 mm long mobile worms (Eisenbach, 1985) and immediately migrate out of the host root (Figure 1). The males of these parthenogenetic species appear to play no role in the disease cycle (Taylor and Sasser, 78) in that they are not needed for fertilization nor do they contribute to the production of eggs. The sperm of all forms of *M. incognita* males are capable of fusion with eggs, but degenerate in the oocyte without fertilization (Trianatphyllou, 1981). A more likely role for males is a survival role, where males are produced in response to poor conditions for egg laying. This form of developmental response to poor conditions is not uncommon among nematodes and may have been sequestered in root-knot nematodes for sex determination and male development.

However, although males do not contribute directly to disease development, their role in parasitism and adaptation to parasitic life may be critical, but the genetic basis of this role

is poorly understood. The biochemical signals controlling sex determination and male development in root-knot nematode are clearly coupled to host resistance and general host conditions yet little is known about the biological role of these males. The trends of rapid evolution in sex–related genes illustrated above demonstrate that little similarity at the sequence level is to be found between orthologous genes. By implementing comparative genomic and bioinformatic analyses of genes expressed in male root-knot nematodes, the molecular basis of male biology was investigated and several tools for further research in this system were developed.

II. A Method for Generating Male Root-Knot Nematodes

BACKGROUND

The primary goal of this project was to generate male tissue in sufficient quantities and purity for the generation of cDNA libraries. It is of primary concern for cDNA library construction that the source of starting material be well characterized for the purposes of interpreting the results of subsequent analyses of these libraries. Two major factors were considered in developing a method for large-scale production of root-knot nematode males, namely triggering the development of adult males and isolating these males in large numbers. Considerations for purity included the ability to produce males without producing females, eggs or other non-male forms as well as minimizing contamination by other organisms.

The effects of a number of distinct environmental factors on the sex determination of root-knot nematodes including crowding, host resistance, pruning, foliar herbicide application, host nutrition, gamma radiation, temperature and age of the host plant have been well characterized. A comprehensive review of the effect of these environmental factors on root-knot sex determination has been published (Trianataphyllou, 1973). In studies investigating the effects of the age of the host plant on the sex determination of root-knot nematodes, no effect on rate of development or sex ratios of *M. incognita* or *M. javanica* were observed, presumably due to the consistent physiology of root vascular tissue targeted for feeding by L2 regardless of host age. The only effect of temperature on sex ratios was the production of more males at the low temperature of 15 ºC in *M. incognita* and above 20 ºC in *M. javanica*. Crowding of the host roots has been shown to increase the male to female sex ratio and slow the rate of adult development when high inoculum densities of *M. incognita*

and *M. javanica* were applied to tomato plants (Davide and Triantaphyllou, 1967). Defoliation or pruning of the host reduces the rate of development and increase the male to female sex ratio in *M. incognita* and *M. javanica* (Triantaphyllou, 1960). The sex ratios of *M. incognita* populations grown on resistant varieties of soybean were heavily male-biased compared to adult populations on susceptible soybean varieties (Moura, 1993). Gamma irradiation resulted in a higher percentage of males when applied to early L2 and half grown larvae, but the later application time yielded *M. javanica* males with intersex characteristics presumably by male sex determination after the female differentiation had been triggered. The application of two foliar herbicides, morphactin (Peacock, 1960; Orion, 1971) and maleic acid hydrazide (Davide and Triantaphyllou, 1968; Ganguly and Dasgupta, 1984) have been shown to both decrease the rate of development of *M. incognita* and *M. javanica* as well as yield a higher male to female ratio in adult populations. The highest percentages of males (over 90%) were obtained when maleic acid hydrazide was applied 0-2 days post infection (dpi) with L2, and when morphactin was applied 2 dpi, more than 60% of the adults were male, suggesting that feeding and not penetration is effected by these herbicides.

The results of these studies indicate that the nature of the sex determination signal is a) related to the condition of the host and availability of nutrients for nematode development and b) that the timing of sex determination signaling is localized in early L2 feeding shortly after or at the time giant cells are initiated. The generality of this response indicates that the signal for male differentiation is common to each of these host stresses and some forms of resistance, and thus allows any of theses stresses to serve as a means of shifting the sex ratio of the population towards male development.

The obligate endoparasitic biology of root-knot nematodes inherently makes pure culturing of this parasite near impossible, yet numerous attempts have been made to develop effective methods (Koening and Barker, 1985). The biology of adult males in particular provides a number of simplifications in the process of pure tissue isolation. The numerous forms of host stress known to trigger male sex determination and development provide several mechanisms for potentially triggering male development. Several of these host stresses can be induced at a precise time and have their effects localized within a very narrow time range, such as early L2 feeding. Male root-knot nematodes are known to naturally emigrate out of host roots shortly after molting from the L4 stage into the motile vermiform adult, whereas females are sedentary and their locomotive muscles have degenerated (Eisenback, 1985). A hydroponic method for producing root-knot nematode larvae (Lambert et al., 1992) is available with the potential to be readily adapted to the purpose of collecting the males resulting from such a timed stimulus potentially free of contamination. By incorporating the natural male development response to timed host stress and their coincident emigration from host roots with a clean hydroponic collection system a method for male tissue generation was developed.

RESULTS AND DISCUSSION

The method developed here was capable of producing large numbers of males in response to pruning stress compared to a no-stress control (Figure 2). The contents of collections made from hydroponic tanks were essentially pure male *M. incognita* and minimal effort was needed to remove contaminants. The appearance of males in the soil under most conditions occurs within 10 dpi (days post infection) whereas eggs are generally not produced until 13

dpi. The timing of these events is further delayed under various forms of adverse environmental conditions and host stress (Davide and Triantaphyllou, 1967). Males were first detected 18 dpi and were found in collections made as late as 45 dpi, although 90 % of males were found in the collections made between 18 dpi and 30 dpi. These results confirm the previous conclusions that, not only does pruning stress induce male sexual development, but also that the response of this stress is limited to the first few days after infection. As a result the emergence of males from host roots could be isolated in time from the emergence of L2 from eggs produced by females of the same generation.

Approximately 14% of the L2 used for inoculum were recovered as adult males under pruning stress conditions compared to 0.18 % on no-stress controls. These yields of males are similar to those previously reported under nutrient deficiency host stress. The nature of the stress induced by pruning is thought to closely resemble nutrient deficiency with regards to this host parasite interaction. In both conditions the development of the L2 into an adult is slowed and the ratio of males to females is increased due to a switch in the developmental fate of the L2 and not due to differential death of females. There are several potential physiological and biochemical differences between plants under nutrient deprivation and pruning stress, but the signal for root-knot nematode sex determination is common to both conditions. It is still not clear whether this signal is a single molecule produced by the host or a more complex cue like the quantity or quality of nutrition the nematode is able to extract, but the later seems more plausible for several reasons. Since each of the previously mentioned forms of stress including various forms of resistance leads to inhibition of nutrient availability to the L2, it seems reasonable to describe the signal as a general food signal. In *C. elegans* L2, dauer development is triggered by a general food signal to allow prolonged

survival and environmental resistance in the absence of an adequate food supply. This food signal is perceived as a ratio of food abundance to nematode concentration. The nematode concentration is perceived as the concentration of a pheromone that is secreted by *C. elegans* into the environment. In fact, this ratio is measured twice during development, demonstrating the functionality of this developmental plasticity throughout development. *M. incognita* and *C. elegans* L2 both express the entire complement of glyoxylate pathway genes needed for lipid metabolism characteristic of nematode survival stages (McCarter et al., 2003). Reentry into the L3 stage from the dauer is also triggered by a general food signal, in this case the reappearance of an adequate food source. Recent evidence demonstrates that *C. elegans* L2 also respond to a general food signal for sex determination, where more males develop when grown in log-phase versus stationary-phase *E. coli* (Prahlad et al., 2003). Not only do these findings indicate that the signal for sex determination is a general food signal, but also that the *M. incognita* L2 is capable of altering subsequent development based on the status of this food signal.

Several classes of males including intersex and true males have been described (Papadapoulou and Triantaphyllou, 1982) as distinct forms based on gonad morphology. True males are thought to arise in response to host signals perceived within the first five days of larval feeding. Based on morphological and cytogenetic evidence from previous studies, the majority of males developing on hosts pruned shortly after inoculation are true males possessing one gonad and all the morphological characteristics of true males (Papadopoulou and Triantaphyllou, 1982). Microscopic analysis indicates that at least 90% of the males generated by this method had only one gonad and spicules indicative of true males (Figure 3). In previous studies on crowding in *M. incognita* only about 10% of males had two testes

apparently as a result of post-sex determination development By combining aspects of root-knot nematode sex determination, male behavior and a hydroponic system for producing L2s, a new method for the controlled production of male root-knot nematodes was developed.

MATERIALS AND METHODS

A population of *M. incognita* maintained on Rutgers tomato (*Lycopersicon esculentum*) was used to harvest eggs from which L2 were hatched using standard protocols. L2 larvae were hatched after 4 days at 16 ºC to synchronize hatching as previously described (Hussey et al., 1973). Filter sterilized and KDF filtered water was used for all watering and nutrient solutions to minimize contamination of hydroponic systems described later. Rutgers tomato seeds were sterilized by washing for 10 minutes in 6% sodium hypochlorite 2 minutes in 75% ethanol and rinsing 5 times in sterile water. Four seeds were planted per $5cm^3$ cell in a 24-cell tray using a sterile 4:1 sand to soil mix. Seedlings were bottom watered daily to maintain field capacity and fertilized weekly with NPK 20:20:20. After four weeks each seedling was inoculated with 1,000 L2 at a 2 cm depth in 5 locations around the root base to achieve uniform inoculation of 5,000 L2 per plant. Lower leaves of inoculated plants were pruned either immediately after inoculation, or one, two or five days later with a sterile scalpel leaving only one apical leaf and the shoot apex. Two weeks after inoculation plants were transferred to a hydroponic growth system (Figure 4). Soil was washed from the roots with filter sterilized water and the stem was wrapped at the soil line with a foam plug covered in plastic wrap to prevent water soaking. A 2 cm diameter hole was cut in the lid of 500 ml plastic container through which the plant was carefully inserted until the foam plug filled the hole in the lid. The container was filled with 400 ml of 0.5X Hoagland's #2 salt

solution. A 2 mm aquarium tube was inserted through a slit in the lid until it reached the bottom of the container and the internal end was fitted with a pumice aerator. The containers were covered in aluminum foil to block out light and the nutrient solution was changed every three days. The plants were grown at 60% humidity 25 ºC under artificial light with minimal aeration supplied by an electric air pump. Males were collected every four days by filtering the hydroponic tank contents through a No. 60 sieve stacked on a No. 325 sieve. Males collected on the No. 325 sieve were counted and snap frozen in liquid nitrogen then stored at -80 ºC.

III. Analysis of Genes Expressed by *Meloidogyne incognita* Males.

BACKGROUND

During recent years tremendous efforts in the molecular and biochemical characterization have yielded a great body of information about the biology and genetics of nematodes. Paramount among these accomplishments is the complete genome sequencing of *Ceanorhabditis elegans* and the continuing annotation of this genome. The public availability of this data through Wormbase (http://www.wormbase.org/) and other resources provides tools for comparative genomic analysis of root-knot nematodes that is invaluable for a number of reasons (Bird and Opperman, 1998; Bird et al., 1999). This complete genome allows for questions about the presence or absence of homologous genes in closely related species to be addressed that could not be answered with an incomplete genome. If a set of genes from a parasitic species such as *M. incognita* have no significant match to *C. elegans*, it is likely they have arisen since the species diverged and potentially have a specific function for parasitism. The ability to back-cross and outcross mutant strains in *C. elegans* has allowed extensive functional annotation by classical genetics revealing entire developmental pathways such as vulva formation, dauer formation and even programmed cell death. RNAi experimentation has further provided functional information to this genome and also serves as a tool for functional gene experimentation in vivo. The complete genomic map of *C. elegans* also provides information about organization of genes on the chromosomes illuminating regulatory mechanisms and evolutionary characteristics of genome structure, such as the trans-splicing seen with the tubulin gene family. All of this data is available through Wormbase, as well as tools for comparative analyses with other organisms such as

parasitic nematodes. With the genome of *C. elegans* and the numerous tools for comparative analysis, the ability to study the molecular and genetic bases of root-knot nematode biology is greatly enhanced.

The obligate nature of plant parasitic nematodes and the asexual nature of many species including *M. incognita* make classical genetics impossible. EST (expressed sequence tag) generation is rapidly proving to be an extremely effective solution to these problems in parasitic species when combined with the *C. elegans* resources described above. EST generation is fundamentally the process of sequencing cDNA libraries of genes being expressed in a given tissue. With this method only expressed genes are sequenced and the function of many of the genes can be inferred by sequence comparison to public databases such as Wormbase. This not only increases the rate of gene discovery over complete genome sequencing but also reduces the cost of gene discovery considerably as well. EST sequencing only sequences genes that are expressed, eliminating the need to sequence non-transcribed 'spacer' DNA that decreases gene discovery rates. Although EST sequencing includes redundant sequencing of abundant or over sampled mRNAs, several forms of normalization and subtraction can be used to reduce this redundancy. The ability to generate EST data from specific tissues also allows questions about the time and location of gene expression to addressed, such as what genes are expressed in males verses L2 or other stages. Among the shortcoming of EST projects is the lack of regulatory sequences obtained due to the mRNA source of starting material for this method. Only partial sequencing of genes are generated when read lengths are short and often both forward and reverse sequencing or even primer walking may be necessary for full length gene sequences. When starting material is limited reliance on PCR amplification introduces significant error in the resulting sequences, but post

sequencing analysis and multiple library construction can reduce these problems. Clustering of ESTs utilizes sequence identity to assemble sequences from the same gene into a consensus and greatly improving the usefulness of the data. Not only does this analysis reduce redundant nucleotides in the dataset by generating a unique set of sequences each defining a single gene, but the redundant sequences now serve to increase the length of each consensus and reduce the rate of sequencing error by averaging the nucletiode value at each redundant position. Clustering may also reveal information about expression levels of individual genes, since the random nature of EST generation will generate clusters with more member for genes which are more abundantly expressed. Care must be taken in interpreting the relevance of cluster size and expression levels, since a number of factors such as PCR amplification, fractionation, and mRNA structure may bias the randomness of sequencing as discussed later.

Public databases now contain over 400,000 nematode ESTs, over half of which are from animal and plant parasites including six root-knot nematode species (Table 1). Among this growing body of knowledge are EST projects which have elucidated many aspects of the biology of parasitic nematodes including many plant parasites. A subtractive cDNA cloning project was also used to generate ESTs from tomato genes induced by root-knot nematode infection (Wilson et al., 1994), showing how powerful this tool is for the study of plant parasite interactions. EST analysis has demonstrated that the shift to parasitism in the nematode *Haemonchus contourtus* is accomplished by dramatic changes in gene expression (Hoekstra et al., 2000). A similar study of 1,000 ESTs from *Globodera* species identified a variety of genes potentially involved in parasitism specifically those that are secreted at the onset of feeding (Popeijus et al., 2000). EST sequencing projects in a number of plant

parasites have been thoroughly reviewed (McCarter et al., 2000) demonstrating the utility of this approach. Recently a set of 5,700 ESTs of *Meloidogyne incognita* L2 was analyzed with a number of publicly available annotation tools and the potential function of the genes identified (McCarter et al., 2003). This was the largest set of *Meloidogyne* ESTs submitted to date and the first set clustered using methods similar to those reported here. From this dataset 1,625 clusters were formed and the Gene Ontology (GO) and KEGG databases were used to assign functions to these genes. From this study all the genes of the glyoxylate pathway were identified suggesting that *M. incognita* L2 metabolized lipid stores while searching for a host. This study also indicated the potential for independent cDNA libraries from the same tissue to yield different representations of genes due to variations in 5' processivity of mRNAs, which is also confirmed by the results from these male libraries. Finally, various classes of genes likely to be involved in plant parasitism were identified including nematode and *Tylenchida*-specific genes as well as a number of horizontal gene transfer candidates. The analyses presented in this work provide the groundwork for similar studies in other organisms as well as other stages of *M. incognita* and demonstrate the effectiveness of EST sequencing as a means of gene discovery. In the analysis presented here, many of the same tools are utilized as well as direct comparison with this data from the L2 stage when possible. Ultimately a greater understanding of the genetic and biochemical bases of male biology in *M. incognita* and the role of males in this species is investigated.

RESULTS AND DISCUSSION

Sequencing and Clustering

A total of 4,992 traces were generated by 5' sequencing of cDNA clones from all four male *M. incognita* libraries. The overall success rate of sequencing based on Phred analysis (Ewing et al., 1998; Ewing and Green, 1998) was 85%; the remaining 15% of runs failed by Phred due to sequencing failure or otherwise poor trace quality were removed (Figure 5). The successful 3,736 sequences were submitted to StackPack v.2.2 (Electric Genetics Corp., Reston, VA.). These sequences were clustered with Stackpack as separate libraries and as a complete set. Separate clusterings were conducted, since the sequences from the SL1 and SMART methods should represent distinct subsets of messages. In theory only full length ESTs are generated by the SL1 method, since the 2nd strand amplification depends on the presence of the SL1 sequence on the 5' end of the message. In the SMART method there is no dependence on the presence of the SL1 sequence in the message. As a result SMART ESTs may or may not contain the SL1 sequence depending on the actual splicing of the message and the amount of degredation of mRNA prior to amplification. Approximately 70% of messages from *C. elegans* (Blumenthal, 1997) are transpliced with the SL1 leader, 80% in *Ascaris summ* (Nilsen, 1993) and 60% in *Globodera rostochiensis* (Dautova, 2001). Manual searches for the presence of the SL1 sequence in each library reveal its presence in several ESTs of both SMART libraries and almost all ESTs of both SL1 libraries. It might also be expected that the average length of the ESTs from SL1 libraries to be longer than those from SMART libraries, due to selection of full length transcripts. Other steps in the cloning process are likely to normalize this difference, such as fractionation and ligation. The average insert size of the SL1a and SL1b libraries were 256 bp and 267 bp respectively, as

opposed to 207 and 208 for SMARTa bp and SMARTb bp respectively, thus indicating that more full length cDNAs may be generated by the SL1 method. Differences in sequence redundancy between libraries were also detected, which indicate the SMART libraries contain more redundancy than the SL1 libraries (Table 2). One of the primary functions of trans-splicing in nematodes is thought to be the coordinate regulation of genes involved in a single biological process. In the *C. elegans* genome sets of genes are arranged in operons that are transcribed as a polycistronic pre-mRNA which is then processed into monocistronic mRNAs by SL trans-splicing (Spieth et al., 1993; Zorio et al., 1994). Approximately 25% of *C. elegans* genes are organized into operons with the first gene either not trans-spliced or trans-spliced with SL1 and all downstream genes are trans-spilced with either SL1 or a second splice leader called SL2 (Spieth et al., 1993 and Zorio et al., 1994). So far, SL2-like sequences have only been detected in a few cases outside *C. elegans* (Evans et al., 1997 and Redmond and Knox, 2001), whereas SL1 has been detected in all nematodes with little variation in sequence (Davis, 1995). The detection of the SL1 sequence in many of the transcripts reported here may be an indicator of coregulation via trans-splicing, and since the presence of SL2 was not investigated, this remains speculative. In fact, many SL1 trans-spliced genes found here are known to be involved in the shared biological processes, such as the actin genes and major sperm proteins. Trans-splicing is likely to play the same role in *Meloidogyne* species as it does in other nematodes, but until a genetic map is completed this theory will remain inconclusive.

The clustering of the SL1a data yielded 68 clusters containing 86 contigs and 555 ESTs leaving 464 as singletons. The clustering of the SL1b data yielded 89 clusters containing 103 contigs and 653 ESTs leaving 363 as singletons. The clustering of the

SMARTa data yielded 120 clusters containing 142 contigs and 384 ESTs leaving 460 as singletons. The clustering of the SMARTb data yielded 105 clusters containing 121 contigs and 345 ESTs leaving 512 as singletons. When all the ESTs were clustered as a complete set, 374 clusters containing 2,255 ESTs were formed leaving 1481 singletons. In general the more sequencing done the less likely it is to generate a new cluster and thus the gene discovery rate drops. This is evidenced by the fact that 484 clusters and 1,799 singletons from the individual libraries are reduced to 374 clusters and 1481 singletons in the complete set. This phenomenon is expected in such a random sampling procedure and thus a theoretical limit to the efficacy of continued sequencing can be derived at which the probability of new gene discovery becomes prohibitively low.

From the clustering of all ESTs it was found that 374 clusters contained 455 contigs, with 48 clusters containing multiple contigs. The multiple contigs within these 48 clusters are likely to represent allelic variants, splice variants or possibly psuedogenes that are still being expressed as mRNA, but may not form a functional protein. The fact that 7.8% of clusters had multiple contigs could indicate a number of situations such as a high proportion of pseudogenes, or alternative splicing both of which are likely present in the transcriptome of this nematode. These explanations are complicated by the inherent error rate of EST data, the polyploidy of this species and thus further analysis is needed to fully understand these results. As shown below, an additional analysis of these clusters through similarity searches within public databases provides further insight to the nature of these clusters.

Transcript Abundance in Relation to Expression

Clustering ESTs serves several independent functions, including reduction of sequence redundancy, generation of a Unigene dataset, identification of splice isoforms, allelic

variants, psudogenes and potentially inference of gene expression levels. Initially sequences are assembled into contigs which theoretically identify identical transcripts generated from repeated sequencing of that transcript. Due to the random sampling nature of EST sequencing, it is expected that abundant transcripts in the tissue being investigated are likely to be cloned more frequently. Additional factors such as G/C content and mRNA secondary structure may exert a technical influence over the selection of transcripts and rate of amplification biasing this sampling in a non-random fashion. Secondary structures such as hairpins in the mRNA may limit the ability of reverse transcriptase to generate a first strand product and prevent or inhibit the sampling of certain transcripts. In cases where starting material is limited, the use of PCR amplification may be needed to generate sufficient cDNA for cloning, but the exponential amplification of PCR rapidly introduces a size sensitive bias where shorter fragments are amplified more frequently than long ones. Although size selection procedures reduce the resampling of partial amplification products and short sequences which are amplified more frequently in PCR, it also prevents or inhibits the sampling of full-length transcripts which fall below the threshold of selection. It is therefore important to carefully choose a size range appropriate for the transcriptome being sampled and a method such as fractionation which will most effectively correct this sampling bias. As a result, the fewer cycles used in amplification the less likely it is that biases will be incurred and the more likely it is that the abundance will be representative of expression levels.

Although there is no method or analysis available to determine the extent of bias in these libraries, there are several results from sequence similarity and frequency distribution analysis indicating that these libraries are representative of male *M. incognita* gene expression. Due to the random sampling of EST sequencing, the distribution of cluster sizes

verses the number of ESTs of each cluster size should approximate a Poisson Distribution with a long tail or progressively smaller clusters. If the library contains only a few sequences due to technical biases, then the EST sampling will generate a few very large clusters and a non-normal cluster size distribution will be apparent. The distribution of each of the four male libraries when plotted independently approximate a Poisson Distribution (Figures 6-9) and when ESTs from all four libraries were clustered together the approximation is even closer (Figure 10). There is also some information about what genes are likely to be highly expressed in *M. incognita* males that can indicate how representative these libraries may be. Genes associated with core biological functions such as RNA processing and protein degradation are generally more highly expressed than genes involved in more unique functions like odorant receptors (McCarter et al., 2000). From microarray analyses (Jiang et al., 2001; Reinke et al., 2000) several classes of genes from *C. elegans* males are known to be highly expressed including protein kinases, protein phosphotases and major sperm proteins. In fact many of the largest clusters from these male *M. incognita* libraries (Table 3) are most similar to these abundantly expressed genes from *C. elegans* males. Not only do these results indicate that the abundance of transcripts in these libraries is indicative of *M. incognita* male gene expression levels, but also that expression profiles are similar in males of these two species. These results also lend support to conclusions about other genes being expressed at higher or lower levels relative to other stages and other organisms where such expression level data is available.

Database Similarities

One of the primary purposes of database searches is to suggest a functional annotation to the genes revealed by EST sequencing. Considerable caution must be taken when interpreting

the results of such comparisons, due to the nature of annotations present in these databases. With this in mind, only the most conservative interpretations of these comparisons are presented. Results from querying the NCBI non-redundant protein database and the EST database (dbEST) with the 1,474 clustered sequences indicate that 1,306 (89%) have significant (E-value < $1.00x10^{-5}$) similarity to previously published sequences. Of the remaining 168 sequences, 61 are shorter than 100bp possibly resulting in a less than significant match. The remaining 107 clusters have an average length of 188bp and are good candidates as novel sequences by these analyses. A query of these 1,474 sequences against the *C. elegans* protein database indicates that 407 (28%) and 451 (31%) have significant match at the E-value < 10^{-10} and 10^{-5} levels respectively. Significant (E-value < $1x10^{-4}$) matches were found for 950 (64%) of these clusters in a query of the nucleotide databases of all other *Meloidogyne* species. These sequences were all generated from stages other than males, including females, eggs and larvae. This comparison serves as the first screen used to identify potentially male-specific transcripts discussed in further detail below. Among these 950 genes with a match to other stages of *Meloidogyne* are a number of matches to calponin genes from L2 which are known to also be highly expressed in males (Catagnone-Serreno et al., 2001). Calponins are part of the actin-linked muscle complex used for L2 migration into the host as well as male migration out of the host and seem to be absent in females. Finding 5 different calponins in these male libraries is a further indication of their role in male biology as well as the ability of cluster analysis to identify gene variants.

Of the roughly 950 genes which are expressed in males and other stages such as egg and L2, there are a number of core metabolic enzymes, ribosomal proteins, transcription factors and also a number of genes with interesting nematode functions. Similarities to FaRPs

(FMRFamide-related peptides) have been identified in other plant parasitic species and have known expression as neuropeptides throughout the nematode nervous system (Kimber et al., 2002). Cluster cn378 has a BLASTX probability of $3.8e^{-118}$ to a FaRP indicating the role of this class of neuropeptide in male biology, yet the variability of FaRPs in nematodes obscures further functional annotation. A great deal more is known about the role of the beta-1, 4-endoglucanase (cn244, 1.00e-76) in plant parasitic nematodes. This gene has been identified in a number of plant parasitc nematodes (Smant et al., 1998; Rosso et al., 1999) and is known to be secreted by the dorsal esophageal gland of infective L2 to facilitate migration into the host (Goellner et al., 2001). The identification of this gene in males serves to bolster previous hypotheses that males may also use them for emigration out of the host roots. In many plant parasites, evasion of host defenses and protection against these defenses are primary factors of pathogenicity that allow for viable reproduction in susceptible host. Among the genes identified here are a superoxide dismutase (10bE06, $5.00e^{-55}$) and several mucins, which may be male-specific and are discussed in greater detail below. Peroxidases and catylases are thought to be expressed in the host response to root-knot nematode infection as part of a generalized host defense (Niebel et al., 1995; Vercauteren et al., 2001). Superoxide dismutase produced by the nematode may serve to protect males from oxidative attack of the host as it emigrates from the roots.

Several genes in the glyoxylate pathway have been identified, indicating that males may metabolize lipids in the same manner as *M. incognita* pre-infective L2 and dauer L2 of *C. elegans*. In this case it is likely that males use this pathway for survival between the cessation of feeding and the location of females. This process requires considerable energy consumption since the male must emigrate back out the host root, which has grown around it,

but it must also locate a female at a potentially long distance and mate. Among these glyoxylate genes are two genes shared with the citrate cycle, citrate synthase (011cB01, BLASTX probability of $7.70e^{-77}$) and malate dehydrogenase (012bA06, $8.50e^{-59}$) and one that is unique to the glyoxylate pathway, malate synthase (011cD03, $1.10e^{-30}$). Each of these genes has been identified in *M. incognita* L2, whereas a homolog for the only other gene unique to glyoxylate metabolism, isoscitrate lyase, has only been found in *M. javanica* and *M. hapla* (McCarter et al., 2003). Interstingly no other genes of the citrate cycle were identified, suggesting that the glyoxylate pathway may even be the primary metabolic pathway of *M. incognita* males. It is generally the case that the glyoxylate and citrate cycle work in tandem with coordinate regulation, with succinate and aspartate as intermediates transported between glyoxysomes and mitochondria. Although DnaJ domains are common in eukaryotic genes, they may have a clear role in root-knot nematodes. DnaJ proteins have been localized to the cytosolic membrane of glyoxysomes (Diefenbach and Kindl, 2000) and some even have testis specific role (Meccariello et al., 2004). Additionally Hsp70a is a testis specific chaperone in mammals known to interact with DnaJ proteins (Hafizur et al., 2004). The identification of six DnaJ genes (12aE07, $1.2e^{-34}$; 04bF10, $5.9e^{-79}$; 04aE07, $7.4e^{-14}$; 11bD01, $6.3e^{-19}$; SMARTa20, $1.7e^{-61}$; 08aB12, $8.4e^{-38}$) and a Hsp70 gene (cn390, $1.7e^{-95}$) implicate a role for these chaparones in glyoxysome stability and possibly for glyoxylate metabolism in root-knot nematode sperm.

A similarity search of these clusters against themselves was conducted to further characterize the clusters that may contain splice isoforms, allelic variants and psuedogenes. This comparison at the nucleotide level identified 36 similarities that were not identified by the Phrap algorithm, involving 30 singletons and 53 clusters, 16 of which only contain a

single sequence. Each of these similarities was confirmed by visual inspection of the sequence alignment using Clustal W with consideration to cluster support when available. Support for 9 of these 36 similarities was found when comparing the best matches to the non-redundant amino acid and dbEST. In 8 cases quite different matches to nr and dbEST were assigned to sequences showing significant self similarity possibly indicating splice variations. By sorting the results of nr and dbEST matches based on Genbank accession numbers, yet another form of cluster annotation was obtained. From this analysis 26 clusters shared a top hit, 8 of which were shown to have sequence similarity in the analysis above. In one case four clusters (cn18, cn19, cn21 and cn25) all have similarity to a *C. elegans* protein phosphotase, but the alignments are different indicating splice variation (Figure 11). The longest match is for cn25 (2-137bp), with cn18 and cn19 matching only the first 75bp and cn21 matching only the last 80bp. Each of the four clusters contains several ESTs supporting the regions with intercluster conflict.

Male Specific Transcript Analysis

One of the primary questions posed in this research was which, if any, genes of this nematode are only expressed by males. From DNA microarray studies in *C. elegans* it is known that approximately 12% of genes are sex-regulated, of which approximately 9% are male-enriched genes (Jiang et al., 2001). In a related experiment using a mutant that only produces sperm and no oocytes, 650 (5.4%) of genes were found to be sperm-enriched (Reinke et al., 2000). Although a complete genome is not available in the genus *Meloidogyne*, more than 60,000 ESTs, all of which are from stages other than males, are currently available in the NCBI public database. This substantial dataset is not considered exhaustive for the genus but provides an effective tool for identifying transcripts from males

that might be expressed only in males. Significant (E-value < $1x10^{-4}$) matches were found for 950 (64%) of these clusters in a query of the nucleotide databases of all other *Meloidogyne* species. Of the remaining 524 sequences, 83 sequences had no significant match to the GenBank nr and dbEST databases, and 217 were either less than 50bp or had significant alignments to sequences known to be expressed in other stages of *Meloidogyne*. From these conservative screens 225 (15%) clusters were determined to be genes which are potentially expressed only in males. The 83 clusters with no significant matches to nr or dbEST are also potentially male-specific genes, but the lack of information about these sequences prevents further annotation at this time.

The identification of many of these potentially male specific genes is not only informative to their characterization as male-specific but also their potential role in male biology. Among the 225 potentially male specific genes, several have similarity to genes known to be expressed only in males of other organisms. These include similarity to major sperm proteins (MSP) from *Litomosoides sigmodontis* (cn43, $3.40e^{-27}$; cn6, $6.50E^{-24}$), *Ascaris suum* (cn48, $8.90E^{-49}$) *Parastrongyloides trichosuri* (10dB02, $2.40E^{-104}$), two testis germinal zone genes from *Ascaris suum* (cn111, $6.10e^{-42}$; cn196, $3.10e^{-36}$). Major sperm proteins are central to sperm motility by the formation of filament networks that extend the leading edge of the sperm psuedopod. Sperm from *M. incognita* males move by extension of a fibrous psuedopod, where these MSPs most likely play a conserved role in all nematodes. Mucins are a diverse class of glycosylated proteins secreted as part of the glycocalyx (surface coat), which is regenerated after each molt often with a stage specific composition. Several mucins were identified, all of which are expressed in males, but not other stages (cn373, $9.20e^{-37}$, 11aH09, $1.00e^{-11}$ and 09dA11, $1.20e^{-10}$). The mucins in animal parasitic nematodes are

though to play a role in avoiding host response, which may be the case for root-knot males emigrating from the host. The fourth stage molt is a major developmental transformation in which longitudinal somatic muscle regeneration and change in body shape mark the transition from the sedentary sacate L4 into the vermiform motile adult male. A number of neuromuscular genes including the GABA transporter *unc-47* (08dG11, $6.10e^{-36}$), the innexin *unc-7* family (04cH02, $3.30e^{-22}$; 10aE12, $1.90e^{-28}$) and the actin depolymerizing factor (ADF) *unc-60* (11cC04, $1.00e^{-66}$) were identified in males with no similarity in other stages. Although unc-47 protein is found in all GABAergic neurons of all stages in *C. elegans*, the regeneration of neurons in *M. incognita* males may require a specific role for unc-47. The role of GABA neurons in *C. elegans* adult males is reversed from its role in larval males, thus demonstrating a stage specific role of unc-47 in male nematodes (Reiner and Thomas, 1995).

Additionally several classes of genes known to be highly expressed in *C. elegans* males are also highly abundant in these *M. incognita* male libraries. These classes include serine/threonine protein phosphotases, protein kinases and major sperm proteins. The maturation of non-motile spermatids to motile spermatozoa in *C. elegans* occurs in absence of ribosomes and thus protein synthesis, resulting in the dependence upon protein modification for this transition (Reinke et al., 2000). The abundance of protein phosphotases and kinases may be the developmental mechanism for sperm maturation in these species and may be conserved among all nematodes.

GO Annotations

Two separate methods were used to apply annotation terms from the Gene Ontology Consortium to the genes identified in this analysis. The first method utilizes the NCBI

accession numbers of genes from the non-redundant amino acid database which were most similar to these male genes (Tables 4 and 5), whereas the second method uses the FASTA sequences of these genes as queries against the Interpro domain database (Table 6). GO annotations exist for 369 of the identified male sequences and 145 GO terms were identified in this analysis. A total of 1,442 matches were made: 482 to biological processes, 414 to cellular components and 527 matches to molecular function were found.

One of the most striking results from this analysis is the amount of similarity seen between male and L2 expression profiles. In the comparison with L2, relatively similar numbers of hits were found in the major categories of biological process, cellular component and molecular function, with 28% more hits for the L2 dataset on average. The percent representation of these hits in subcategories reveals a certain degree of information about gene expression profiles, where higher percentages indicate larger allocation to that category. This is different from the earlier abundance data, because here we are looking at much broader categories than the single gene abundance data discussed earlier. One possible explanation for the nearly identical distributions seen is that the majority of hits are to GO categories describing broad biological categories, thus the results lack the detail to resolve expression differences between life stages. Another way of looking at these similarities is that the distribution seen here is conserved among stages of root-knot nematode and may represent a general gene expression profile for this nematode. One of the largest differences seen here is the greater percentage of structural molecule activity in males than in L2, most of which are ribosomal and cytoskeletal molecules. From looking at the Interpro results discussed below it turns out many of these structural molecules are major sperm proteins.

There is also a higher percentage of phosphotase and kinase genes in males than L2, further indicating there upregulation in males as indicated earlier by their transcript abundance.

Several interesting categories were found to be represented in the male set and not in the L2 data including, excretion, secretion, glycosaminoglycan binding, phosphotase regulator, toxin activity and antibiotic resistance/susceptibility. Although the cluster IDs are not available for this data at this time, the secretion product that was identified may represent a hit to the SEC-1 gene from *Globodera pallida* (cn89, $6.60e^{-100}$). This is a gene expressed in the esophageal gland cell of *M. incognita* L2, but despite myosin heavy chain domain similarities, its biological role is still unclear (Ray et al., 1994). The role of the excretion is likely to be one of osmotic regulation or defecation common to all nematodes, and again a male specific role is unclear. In the cases of the phosphotase regulator and glycosaminoglycan binding annotations, the male-specific roles for sperm protein modification and mucin production respectively support the previously described results from male-specific analysis. The toxin activity and antibiotic resistance/susceptibility annotations are both directly related to pathogenicity and thus represent strong candidates for molecules involved in the host-parasite interaction.

In Table 5 broader categories are compared across larger evolutionary distances revealing some interesting differences. In the distribution among the two biological processes a shift is seen from cell growth and maintenance to cell communication is seen in males, which may be counterintuitive for these two stages. Two subcategories of cell communication are host-pathogen interaction and response to external stimulus, and the majority of Interpro annotations in theses categories are signal transduction genes. A higher percentage of cell localized annotations is seen in all but the *M. incognita* L2 data, where the

extracellular percentage is much higher. This large percentage of secreted molecules in the L2 indicates a much more environmentally interactive role of the L2, as to be expected in the many host-parasite interactions of this stage. The shift in molecular functions from structural molecules in males and other species to ligand binding and carrier functions in L2s may also be indicative of environmental sensitivity. These again are merely general trends due to the broad nature of these comparisons, yet they serve to indicate these broad differences in the allocation of gene function among different stages of root-knot nematode biology.

Interproscan Annotation

A total of 1,718 comparisons were made between the 1,474 input Fasta sequences and the Interpo domain database. The database was queried with several algorithms including HMMSmart, HMMTigr, HMMPfam, BlastProDom and Profilescan to an Interpro domain. Of the 389 which matched known Interpro domains, 257 had a significant (E-value < $1.00x10^{-5}$) match, with the remaining comparisons to regions with coil domains or particular residue enrichment only. A total of 273 matches were made to domains with 62 unique GO annotations, constituting only 16% of the comparisons made with this method. Ultimately, 130 genes (~10%) had hits to domains with GO annotation in this analysis, thus each gene has on average two GO annotations. Unfortunately only 4 of the same GO terms were identified in both analyses preventing substantial comparison between methods. It is important to note that the two methods provide much more information than either alone, and thus complement each other well. In addition to a profile of GO terms and their abundance, this Interpro analysis retains cluster names for each annotation.

The results of this analysis provide a number of insights including support for previous results in this study, attribution of GO annotations to male clusters and also

identification of new GO annotations not obtained in the analysis above. One of the most significant results is the motif identities of the putative male-specific class of genes described above. The vast majority of these genes contained MSP, protein phosphotase or protein kinase domains, further supporting the earlier results that these genes are either male-specific or highly expressed in males. In addition to these annotations, putative male-specific genes were assigned proton transport, nucleic acid binding and nuclear horomone receptor GO annotations. These genes may be involved in signal transduction pathways unique to males. The identification of a Rho-family GTPase binding receptor is particularly interesting in this regard, due to its known role in regulation of cytoskeletal reorganization and cell motility in *C. elegans* (Spencer et al., 2001). This role may be conserved in *Meloidogyne* and be part of the sperm motility regulation mechanism.

The Interproscan GO results presented here serve a number of important roles and ultimately illustrate the growing universal application of comparative functional genomics for the study of plant parasitic nematodes and metazoan biology in general. Without the tremendous collaboration of protein functional analysis and gene sequencing that has built the public databases and computational tools such as GO and Interproscan, this level of functional annotation would take years. These public resources are therefore the keystone to the utility of computational and bioinformatics analyases and provide not only the raw data for gene function, but a way to interpret that data for any organism studied. This resource is constantly growing with more specific annotations in many groups of specialized life most relevantly are pathogen specific GO annotations which enable researches to identify specific pathogenic functions of unknown sequences. These also have the ability to identify new functions for genes apart from similarity searches like BLASTs which only provide the

annotations provided by individuals and may often be inaccurate. The functional domain identification of Interproscan coupled with the GO terminology annotations for those domains provides a second interpretation that can easily complement other forms of annotation. It is ultimately the ability of such analyses to provide comparative profiles for closely related organisms with different biologies that provide some of the most interesting results including the ability to identify genes which play species –specific or as in this case phenotypic roles. The utility of these tools will certainly grow with new functional analyses and provide new ways to interpret the roles of novel and even well documented genes in the future.

MATERIALS AND METHODS

Library Construction

Males were generated by methods described above with the following modifications. Each 4 week old tomato plant was inoculated with 10,000 L2 and males were collected every two days by filtering the hydroponic tank contents through a No. 60 sieve stacked on a No. 325 sieve. The contents of the No. 325 sieve were decanted in a small volume and applied to a 40% sucrose cushion in a 50 ml conical tube and centrifuged at 2,000 rpm for 10 minutes. Males banded at the sucrose/water interface were removed by pipette and rinsed on a No. 500 sieve with water and transferred to a Petri dish for counting and collection. Microscopic inspection was used to insure only males were transferred for collection into 1.5 ml centrifuge tubes by manual pipetting. Males were spun briefly to concentrate the tissue and excess water was removed with a micropipette. Males were immediately snap frozen in

liquid nitrogen and stored at -80ºC. 4 pellets of male tissue were combined ranging from 3 weeks after inoculation to 5 weeks after inoculation.

Total RNA was extracted from the male tissue using the Trizol method (Life Technologies, Gaithersburg, MD) with a yield of 6 µg from 100 µl of tissue. The RNA was enriched for transcripts bearing a poly (A) tail using the Dynabead mRNA Collection Kit (Dynal Biotech ASA, Oslo, Norway). Synthesis of cDNA was conducted using the SMART cDNA Synthesis Kit (Clontech, Palo Alto, CA) with the following modifications. First strand cDNA was synthesized with the 5' primer AAGCAGTGGTATCAACGCAGAGTATATAGCGGCCGCr(GGG) and the 3' primer ATTCTAGAGGCCGAGCTCGAGGACATG(T)30VN. The last three nucleotides of the 5' primer are RNA and in the 3' primer the V = A, G, C and the N = A, T, C, G. Second strand cDNA was amplified for 12-14 cycles of PCR with the 3' primer ATTCTAGAGGCCGAGCTCGAGGACATG and either GATAAGAATGCGGCC GCAGGTTTAATTACCCAAGTTTGAG as the 5' primer for SL1 libraries or AAGCAGTGGTATCAACGCAGAGT as the 5' primer for SMART libraries. The cDNA was double digested with *Not* I and *Xho* I and then fractionated on Chroma-spin 400 columns (Clontech, Palo Alto, CA) according to manufactures recommendations. Inserts from four independent pools of cDNA (named SL1a, SL1b, SMARTa and SMARTb) were directionally cloned into the plasmid vector pGEM-11 zf(+) (Promega, Madison, WI). Libraries a and b were generated from the same 1st strand cDNA but came from two different 2nd strand amplifications and remained separated through sequencing. 25ng of each ligation were transformed into DH10B *E. coli* (Invitrogen, Carlesbad, CA) and plated on LB ampicillin selective agar plates with X-gal and IPTG for blue/white screening. The resulting

cloning efficiencies of these libraries was: SL1a = ($4.8x10^5$ cfu/µg), SL1b = ($1.7x10^5$ cfu/µg), SMARTa = ($5.4x10^5$ cfu/µg), SMARTb = ($3.6x10^5$ cfu/µg).

Sequence Production and Processing

Individual positive bacterial colonies were picked into 384-well plates containing LB ampicillin broth with 7% glycerol using a Q-bot robotic colony picker (Genetix, Christchurch, UK). The parameters for picking were selected based on a per plate manual assessment to minimize selection of clones with small or no inserts. Plates were incubated overnight at 37 ºC overnight and stored at -80 ºC, after plates were filled and replicated by hand. 96-deep well blocks containing 1.6 ml per well of LB ampicillin broth were inoculated from the 384-well culture plates for generation of plasmid template for DNA sequencing. Plasmid DNA was extracted using 9600 Biobots and R.E.A.L. Prep 96 Plasmid Kits (Quiagen, Valencia, CA). Cycle sequencing with the 5' SP6 primer and BigDye Terminator Mix 3.1 (Applied Biosystems, Foster City, CA) was performed on 96-well thermocyclers (MJ Research, Waltham, MA). Reactions were analyzed on ABI 3730 96-capillary sequencers (Applied Biosystems). Individual sequencing reads were assigned names based on their plate names and well locations within those plates. Four trial plates were generated, one from each library, named SL1a, SL1b, SMARTa and SMARTb and sequences generated from these plates are named platename_capillary# and a capillary to well mapping is provided in Appendix A. Sequences from the remaining plates were named platenumber(offset)(direction)(run)_wellID_capillary#, thus "011cf1_C04_028.ab1" is from well row C column 04 offset c from plate 11 and was read in the forward direction through capillary 028 on the first run of that plate.

Sequence Analysis and Clustering

Raw chromatograms were basecalled and quality trimmed to mask low quality sequence using Phred (Ewing et al., 1998; Ewing and Green, 1998) with default settings (low quality cutoff = 20) and the screen option. The screen option in Phred generates a modified sequence file where low quality basecalls are replaced with "n"s. Vector sequence was masked using Crossmatch and the output was manually inspected to insure the sequences were trimmed properly. A second form of vector screen was applied to a small subset of these sequences as a confirmation to the fidelity of these methods. This test trimming was conducted using the SeqManII application from DNAStar (DNAStar, Inc. Madison, WI). Some slight discrepancies were detected between the two methods, yet the differences were predictable and resolved by refining the definition of vector and trimming parameters in the SeqManII analysis. Ultimately the output from Crossmatch was utilized due to its accuracy in locating the primer vector boundary and suitability to pipeline construction for high throughput analysis. This output was then manually inspected and trimmed again to remove primer sequences and polyA tails using PERL script 1 (Appendix A). This script was written specifically to deal with the dataset generated with this protocol as the readthrough sequences contained artificial sequences from the primers used for cDNA amplification. These primer sequences along with polyA sequences must be removed before these sequences are clustered to avoid false cluster formation based on these universal sequences. The untrimmed data for these sequences was archived separately and is available for further analysis as intact raw sequence chromatograms. Primers used for cDNA amplification from the 5' end of mRNAs were not detected in the majority of cases and when detected were incomplete and insufficient to cause false cluster assemblies as discussed below. This is a common situation in automated sequencing and is mainly due to poor sequence quality at the beginning of each

trace and proximity of the sequencing primer to the 5' end of transcripts. The trimmed sequences were then clustered using Phrap (Ewing and Green, 2000) via Stackpack [Electric Genetics, Reston, VA]. Clustering ESTs serves several independent functions, but ultimately was used here for generating a set of unique sequences each defining a single gene. Initially sequences are assembled into contigs which theoretically identify identical transcripts generated from repeated sequencing of that transcript. Due to the random sampling nature of EST sequencing, this is to be expected as abundant transcripts in the tissue being investigated are likely to be amplified more frequently. Additional factors such as transcript sequence content and structure may influence the rate of amplification and further bias this sampling in a non-random fashion. Contig assembly therefore not only reduces this data redundancy, but also provides a measure of transcript abundance discussed earlier in the results. All contigs were visually inspected in the Stackpack output and chimeric sequences were manually edited and resubmitted for reclustering. These contigs are then further assembled into clusters which theoretically represent individual genes, but may contain variations due to allelic variance, alternative splicing or psuedogene drift. Clusters were manually inspected and the output FASTA files of clusters and singletons were trimmed with a second PERL script to remove contaminating and short sequences from the Stackpack output [Appendix A]. Output from clustering is available in MSF files, CRAW files, or FASTA files of non-redundant singletons, or only clustered data. Contig names begin with the ct prefix, cluster consensuses begin with the cn prefix and singletons retain their plate and well identifications from the raw sequencing naming convention as described above. The third script reported here serves only to rename sequences in a fashion that allows them to be processed as batch sequences in subsequent analyses.

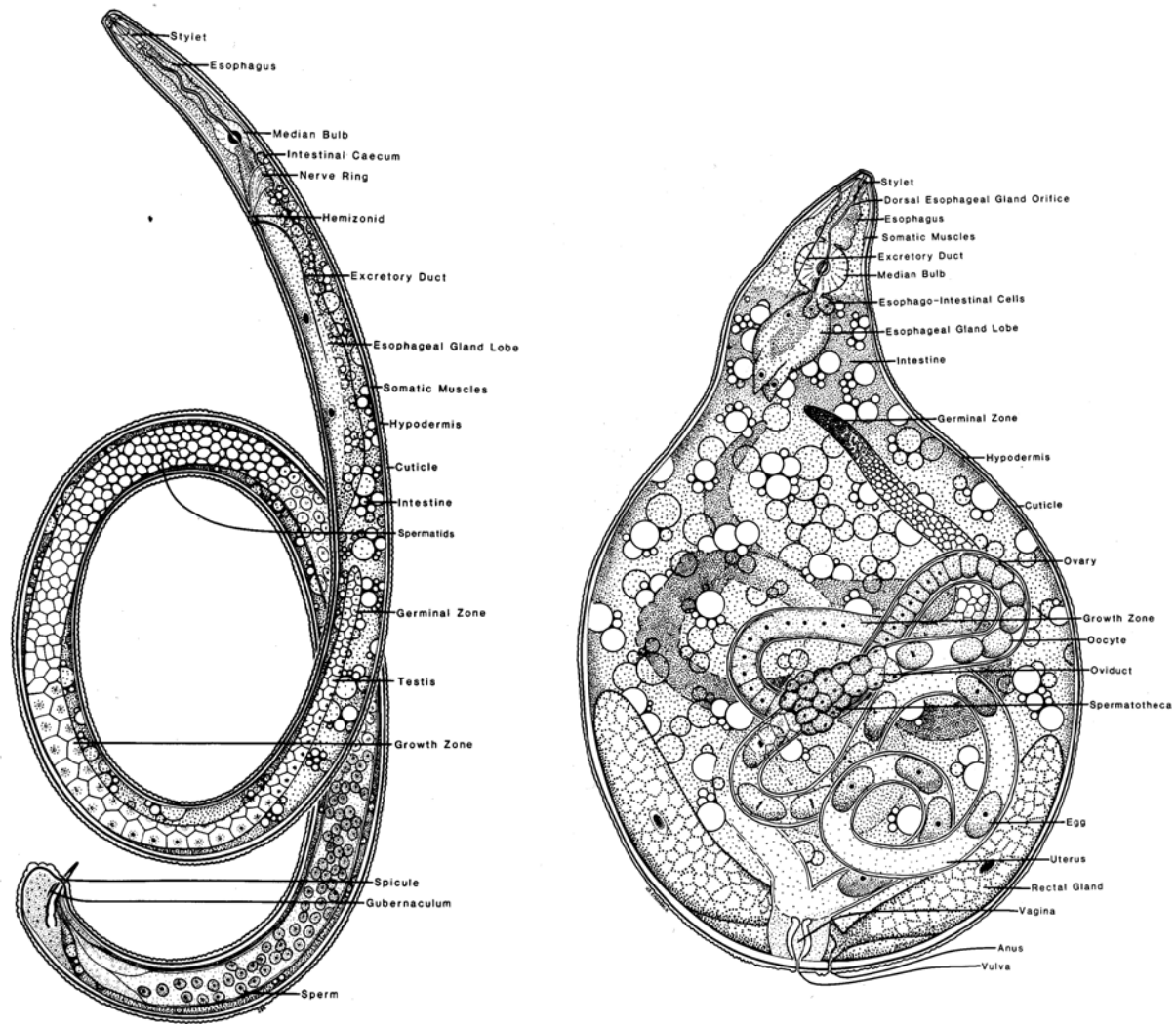
Database Comparisons and GO Annotation

Several public databases were employed as primary annotation tools for this dataset. These databases were downloaded from NCBI onto a local BLAST server (Altchul et al., 1990) and utilized for batch analysis of the 1,474 clustered sequences generated as described above. In several cases these databases were also used to analyze all 3,736 of the unclustered ESTs generated as input for clustering. In all comparisons the following default parameters were utilized unless otherwise indicated; open penalty = -12, extend penalty = -2, frame penalty = 0, scale factor = 1, matrix = blossom 62. The Smith-Watermann algorithm was used for nucleotide to amino acid comparisons. The databases utilized are as follows; 14,251 *M. incognita* nucleotides from April 9th 2003, 3,119 *M. javanica* nucleotides from April 9th 2003, 2,066 *M. arenaria* nucleotides from April 9th 2003, 1,283 *M. hapla* nucleotides from April 11th 2003, 839 *M. chitwoodi* nucleotides from April 15th 2003, the non-redundant amino acid database from June 27th 2003, the 'human', 'mouse' and 'other' EST databases from July 2003, and the *C. elegans* amino acid database from February 2003. A self-similarity search TBlastX was conducted with the 1,474 clustered sequences against themselves to detect any residual redundancy. 83 sequences showed significant similarity to a different cluster (E-value < 1.00×10^{-5}) comprising 5.6% residual redundancy after Phrap clustering. These similarities were all confirmed by manual inspection of alignments in the Stackpack and BLAST output.

Two separate methods were used to apply annotation terms from the Gene Ontology Consortium to the genes identified in this analysis. The first method utilizes the NCBI accession numbers of genes from the non-redundant amino acid database which were most similar these genes as input, whereas the second method uses the FASTA sequences of these

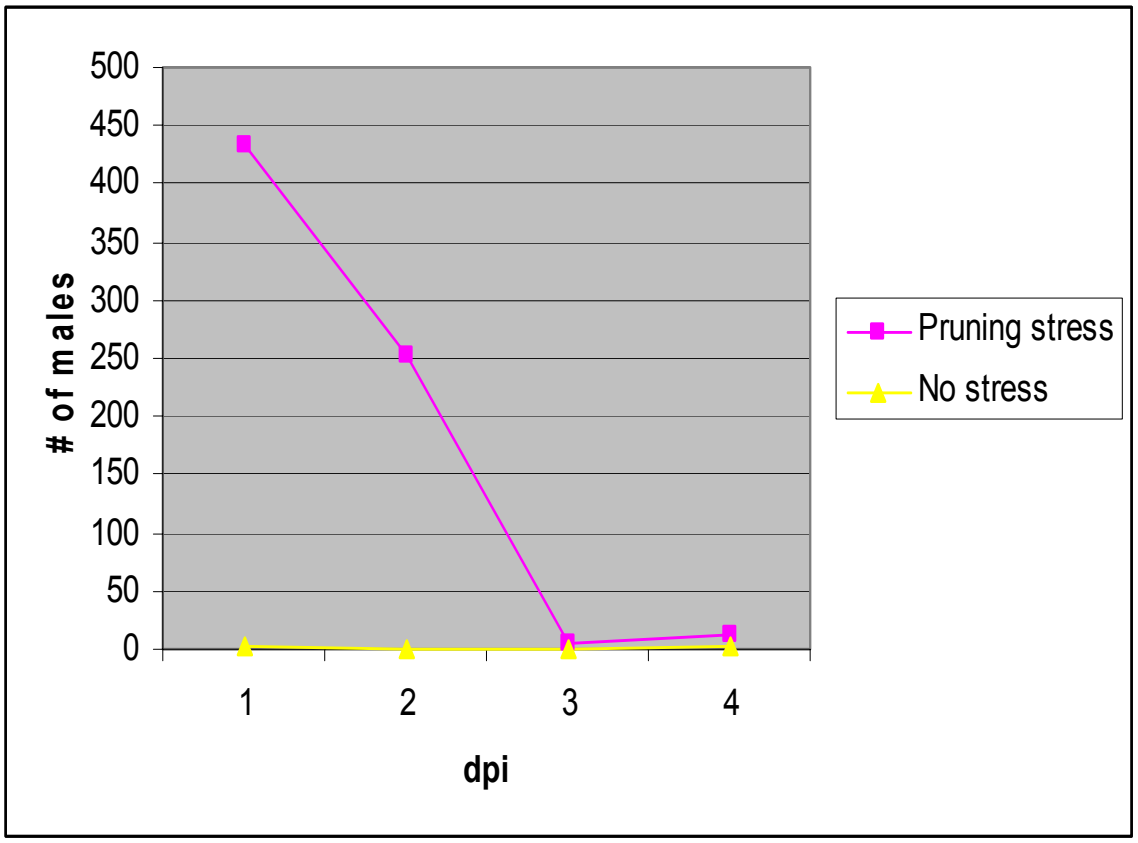
genes as queries against the Interpro domain database. In the first method 735 accession numbers were retrieved as the best protein hits (E-value < $1.00x10^{-5}$) to nr from a six frame translation of these male *M. incognita* clusters. In the second method all 1,474 FASTA sequences of these *M.incognita* male clusters were used as input for Interpro analysis.

Figure 1. Male and female root-knot nematode morphology.



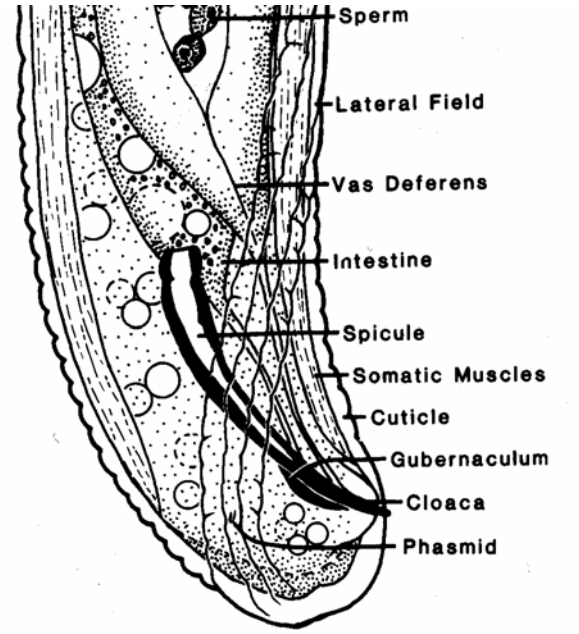
From Eisenback, J.D. 1985. (Not to scale) The male (left) is vermiform, 1400um long, motile, with one gonad, spicules and longitudinal muscles, whereas the female (right) is globose, 800um long, sedentary, with two gonads and no spicules or longitudinal muscles.

Figure 2. The effect of pruning stress and time of stress on sex differentiation ratios.



Data points for all treatments (stress and no stress) are averages for five replicates per treatment. The x-axis represents the number of days post infection (dpi) at which time stress was applied by pruning. The y-axis represents the total number of males collected between 14 and 42 dpi for each treatment.

Figure 3. Tail morphology of *M.incognita* male.

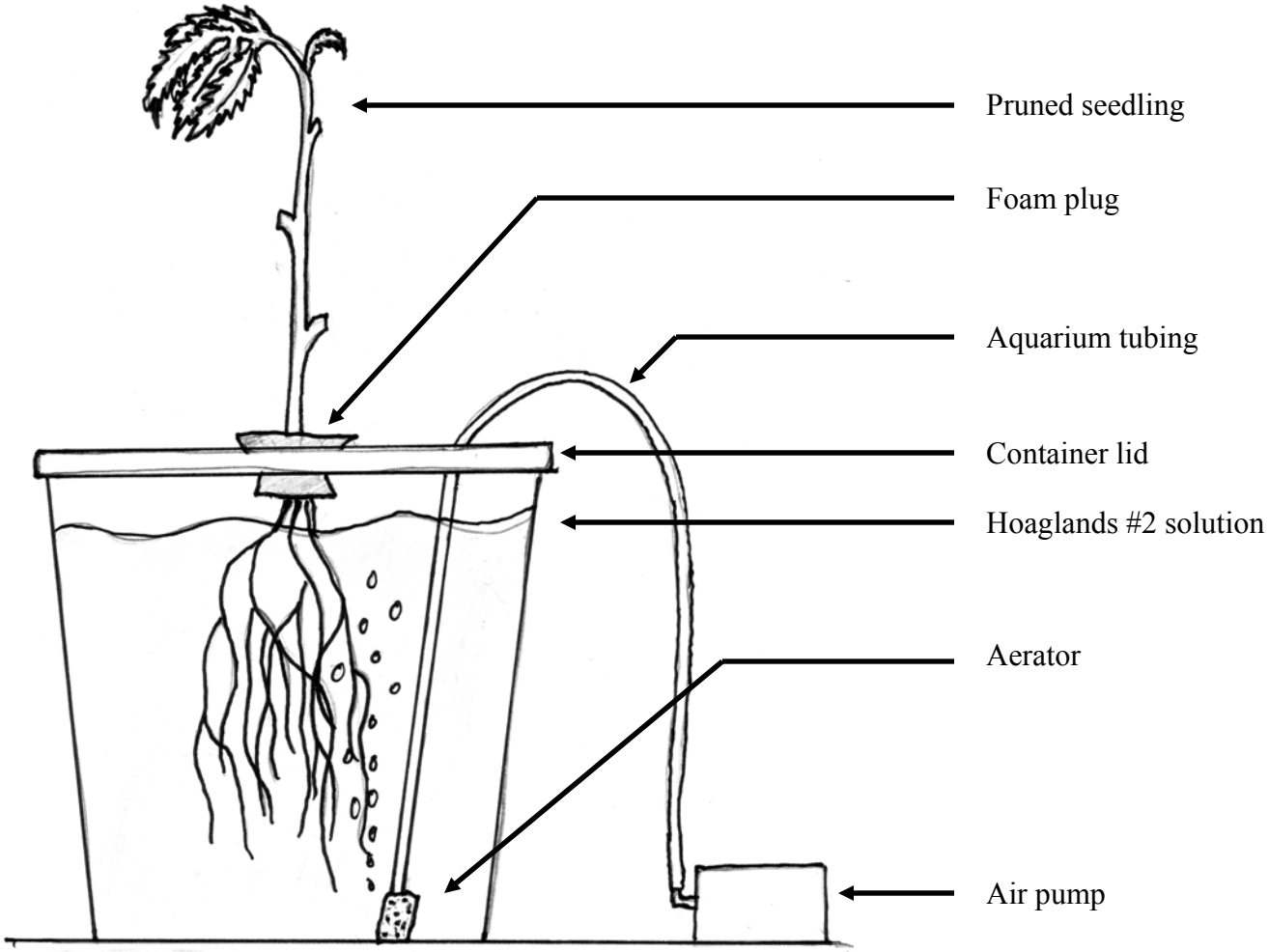


From Eisenback, J.D. 1985. Presence or absence of spicules is a distinction easily made for determination of true versus intersex morphology.

Table 1. **Current EST totals for nematodes in Genbank**. These data were obtained from Genbank nucleotide browser on 4-27-04. The numbers shown are for ESTs only and do not include genomic or other submitted sequences. The ESTs from within the genus *Meloidogyne* were used for putative male-specific transcript identification.

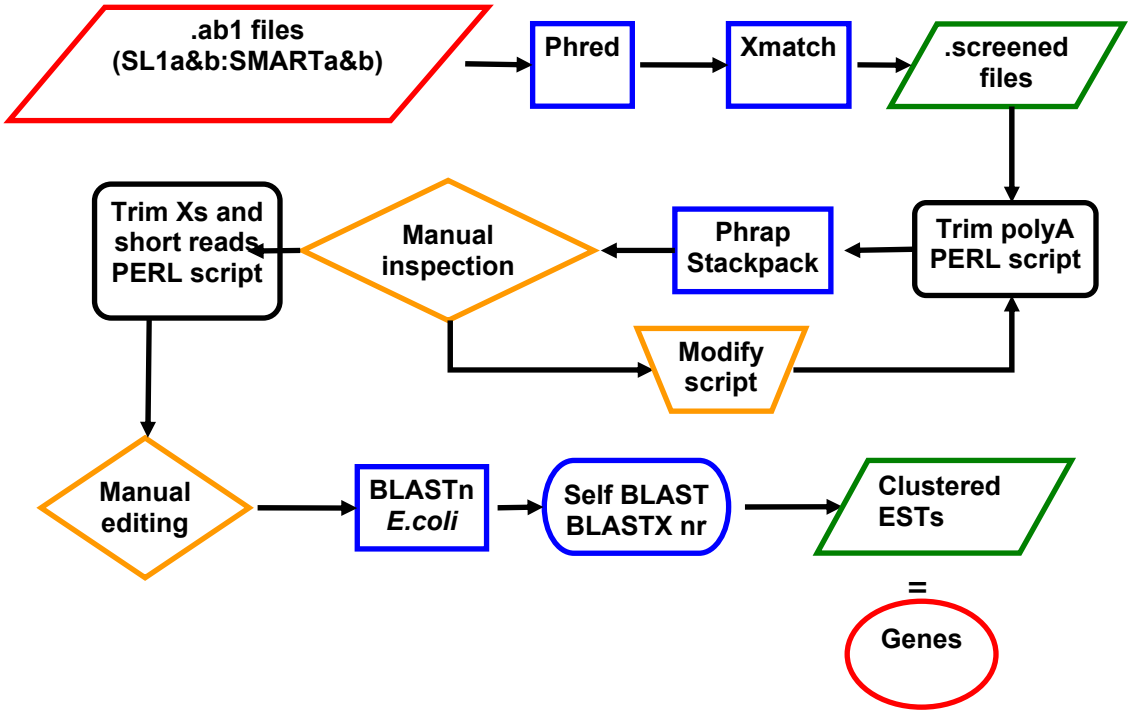
Taxa	**Sequences**
Nematoda	517,518
Tylenchida	103,233
Heteroderidae	101,971
Globodera	7,972
Heterodera	27,412
Meloidogyne	66,577
M. incognita	19,534
M. hapla	18,323
M. chitwoodii	12,255
M. javanica	7,622
M. arenaria	5,065

Figure 4. **Schematic design of hydroponic apparatus.**



Only one tank is shown here for simplicity, yet the system is designed to accommodate multiple tanks by incorporating splitter valves and a high capacity air pump. The tanks were also wrapped in tin foil to prevent light penetration to the roots.

Figure 5. Schematic of the pipeline used for EST analysis.

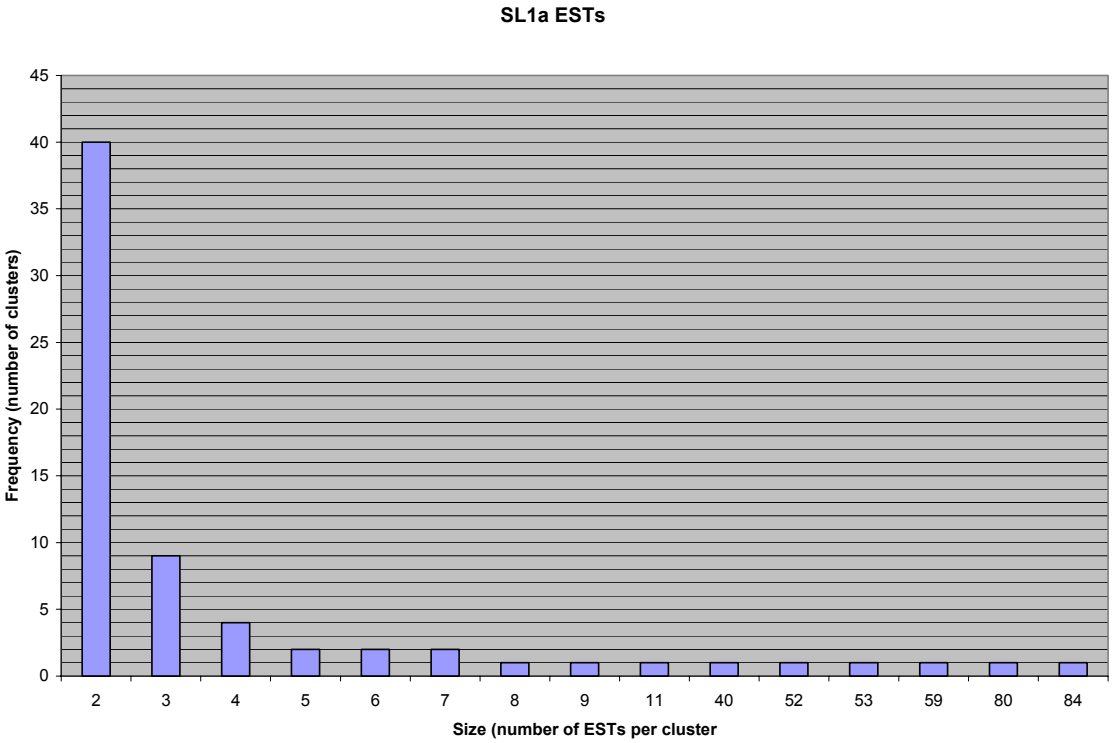


This diagram represents a number of commonly used computer programs (blue) that were linked by PERL scripts (black) and manual modifications (orange) to automate the analysis process. The PERL scripts used were developed based on the analysis of the output at each step and are specific to this dataset.

Table 2. Cluster results for each male *M. incognita* library. The SL1 libraries were generated using the trans-splice leader sequence from *C. elegans* and the SMART libraries incorporated a universal priming method. The set of clusters from all ESTs includes all the ESTs from the four independent libraries and thus the cluster totals are not addative.

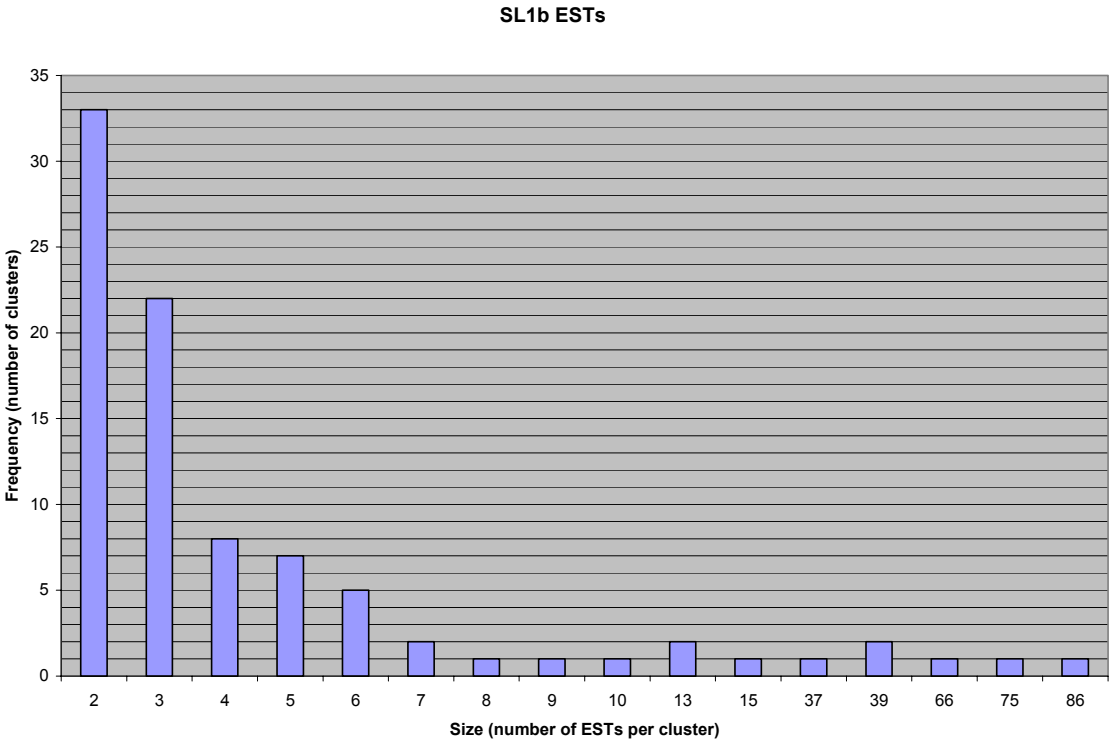
Library	# of clusters	# of singletons	# of total ESTs
SL1a	120	460	844
SL1b	105	512	857
SMARTa	68	464	1,019
SMARTb	89	363	1,016
All ESTs	374	1,481	3,736

Figure 6. Cluster distribution of 844 ESTs from the SL1a library only.



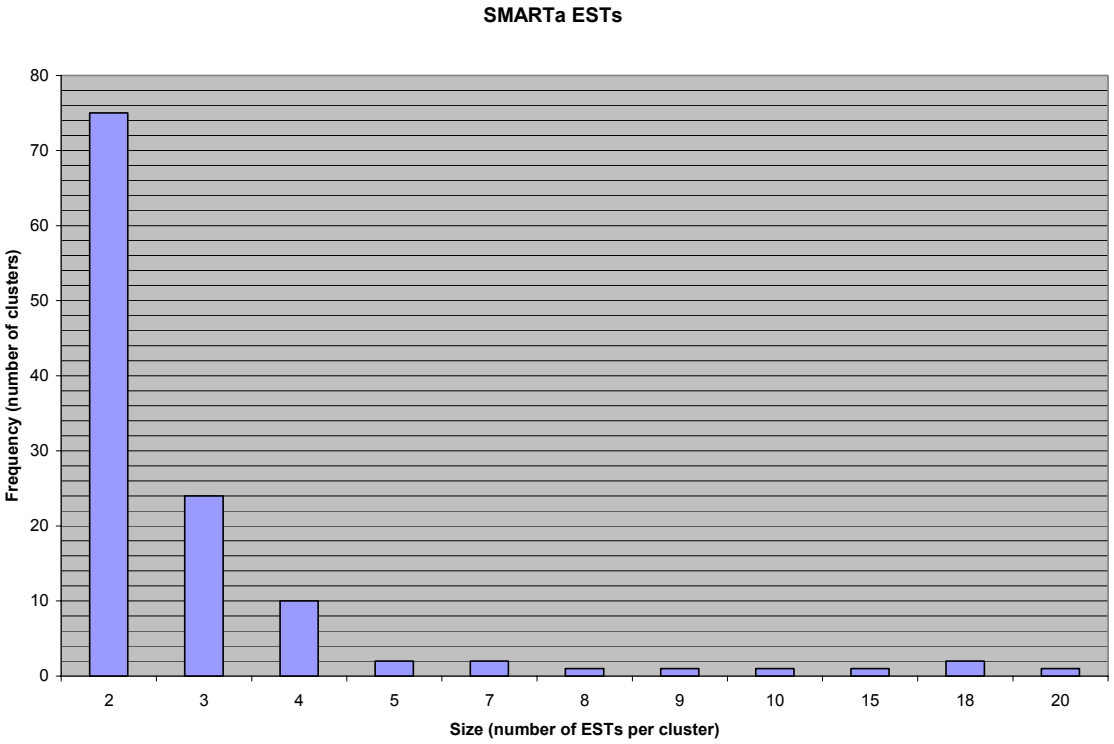
464 singletons are not shown for the purposes of graphical scaling. The x-axis is scaled to the data range and is thus non-linear.

Figure 7. Cluster distribution of 857 ESTs from the SL1b library only.



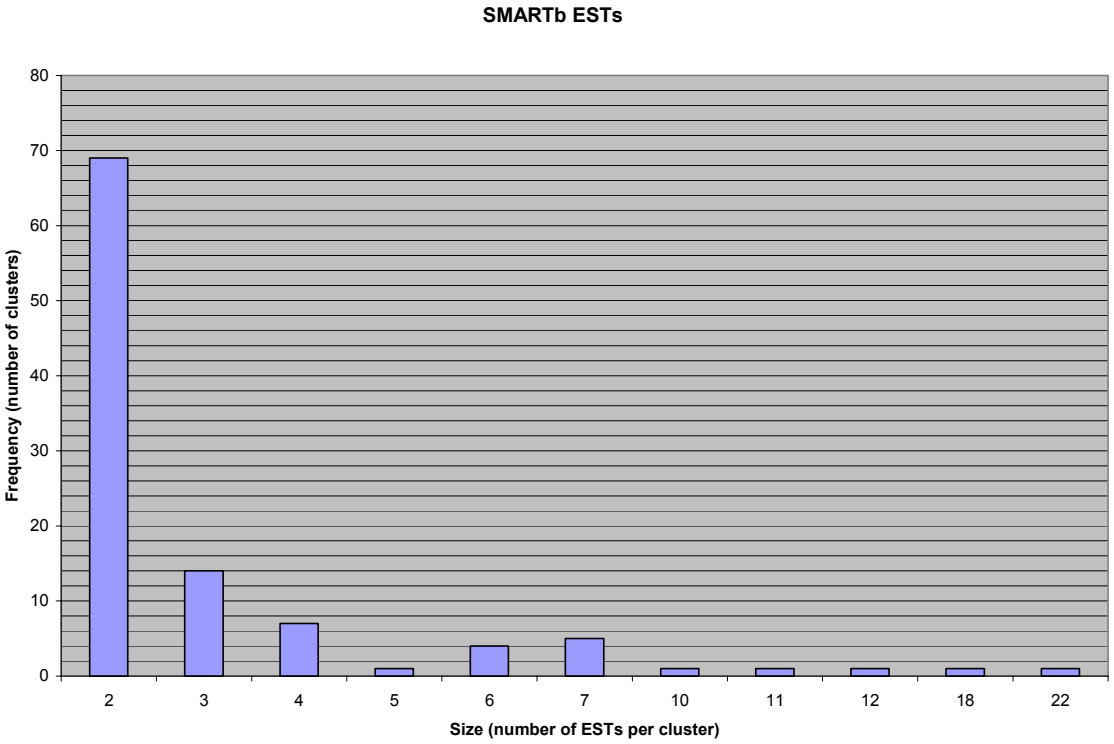
363 singletons are not shown for the purposes of graphical scaling. The x-axis is scaled to the data range and is thus non-linear.

Figure 8. Cluster distribution of 1,019 ESTs from the SMARTa library only.



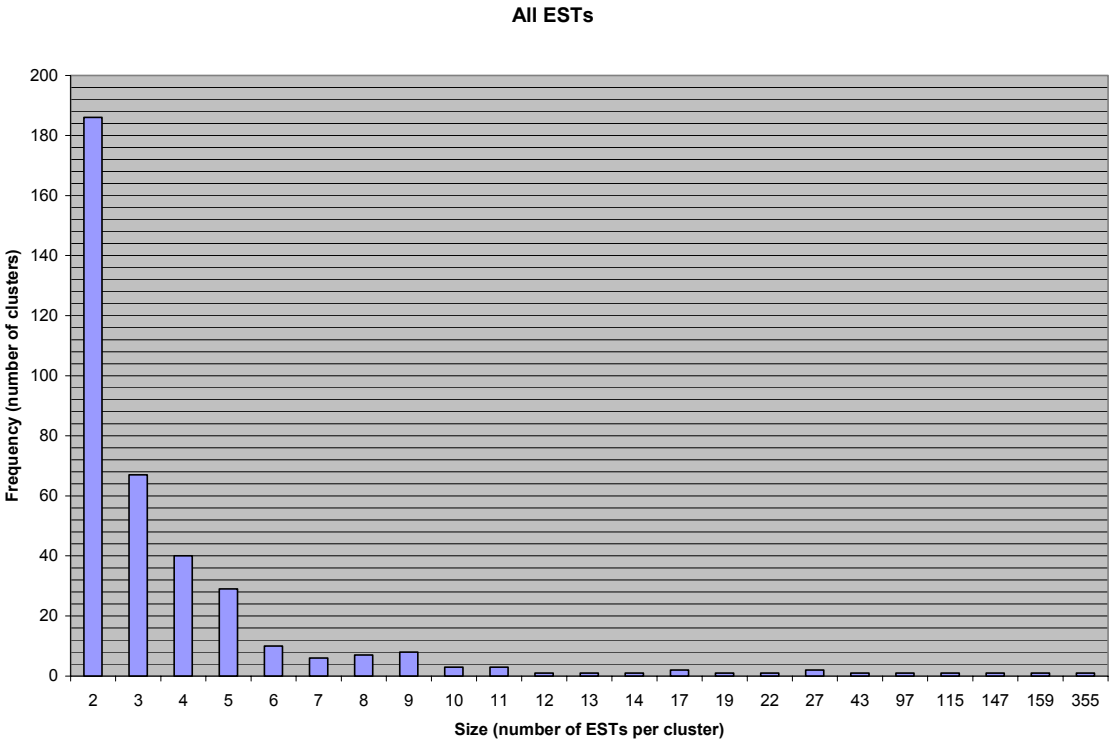
460 singletons are not shown for the purposes of graphical scaling. The x-axis is scaled to the data range and is thus non-linear.

Figure 9. Cluster distribution of 1,016 ESTs from the SMARTb library only.



512 singletons are not shown for the purposes of graphical scaling. The x-axis is scaled to the data range and is thus non-linear.

Figure 10. Cluster distribution of all 3,736 ESTs.



1481 singletons are not shown for the purposes of graphical scaling. The data plotted here represent clusters formed from combining all ESTs from the four different libraries. The x-axis is scaled to the data range and is thus non-linear.

Table 3. The 20 largest contigs representing the most highly abundant transcripts. These clusters are the 20 largest based on the number of ESTs per cluster and not on the length of the cluster. They are representative of the most abundant transcripts in the combined clustering of all ESTs from the four cDNA libraries genenerated in this study.

Contig	length	# ESTs	Description
cn12	899	135	*Meloidogyne incognita* L2 cDNA 5'
cn5	620	80	*Meloidogyne hapla* egg: similar to MSR1 repetitive element ;
cn25	717	55	Similar to serine/threonine protein phosphatase PP1; *C. elegans*.
cn9	185	50	17.5 days embryo whole body, *Mus musculus*.
cn8	232	48	Homo sapiens cDNA clone IMAGE:30403573 5'
cn4	299	38	Homo sapiens cDNA:contains Alu repetitive element;contains element TAR1 repetitive element.
cn21	711	34	Similar to serine/threonine protein phosphatase PP1; *C. elegans*
cn176	536	33	ra61d10.y1 Bird-Rao *Meloidogyne incognita* L2 *Meloidogyne incognita* cDNA 5'
cn48	660	27	Major sperm protein, *C.elegans*
cn3	626	26	*C. elegans* gene product
cn47	900	25	Major sperm protein, *C.elegans*
cn46	337	22	*Meloidogyne incognita* egg cDNA.
cn68	593	22	Major sperm protein, *C.elegans*
cn268	458	19	*Meloidogyne incognita* L2 cDNA
cn30	323	17	(synonym: hfbr1) *Homo sapiens* cDNA
cn35	296	16	*Homo sapiens* cDNA
cn45	414	16	Major sperm protein, *C.elegans*
cn44	500	15	Major sperm protein, *C.elegans*
cn2	474	14	Zebrafish shield stage whole embryo cDNA
cn97	563	13	predicted using Genefinder; similar to Transthyretin-like family

Figure 11. Alignments of clusters representing likely splice isoforms of a protein phosphotase.

The regions in red at the 5' end of these clusters may represent alternatively spliced exons of protein phosphotases transcribed from a single gene. These four clusters are each supported by multiple EST sequences throught the regions of discrepancy. The fact that cn21 and cn25 share the same 5' concensus boundry further supports this theory.

Table 4. Gene Ontology profiles from males and L2s. 369 male sequences had hits with GO annotations. Percentages are based on (a) 482 male matches and 670 L2 matches (b) 414 male matches and 557 L2 matches (c) 527 male matches and 756 L2 matches. The male GO annotations are for all clusters from the combined clustering of all ESTs from the four male libraries.

Gene Ontology Mappings

(a) Biological Process

Categories and Subcategories	Representation in males		Representation in L2s	
Physiological process	363	(75%)	496	(74%)
Metabolism	248	(51%)	354	(53%)
Response to external stimulus	13	(3%)	17	(3%)
Response to stress	3	(0.6%)	10	(1%)
Circulation	2	(0.4%)	1	(0.2%)
Death	2	(0.4%)	7	(1%)
Response to endogenous stimulus	1	(0.2%)	6	(0.9%)
Excretion	1	(0.2%)	-	-
Extracellular matrix organization	1	(0.2%)	-	-
Secretion	1	(0.2%)	-	-
Cellular process	157	(33%)	223	(33%)
Cell growth/ maintenance	110	(23%)	138	(21%)
Cell communication	29	(6%)	59	(9%)
Cell motility	13	(3%)	15	(2%)
Cell differentiation	3	(0.6%)	4	(0.4%)
Cell death	2	(0.4%)	7	(1%)
Development	42	(9%)	67	(10%)
Morphogenesis	17	(4%)	26	(4%)
Growth	7	(1%)	13	(2%)
Regulation (epigenetic)	6	(1%)	14	(2%)
Reproduction	6	(1%)	7	(1%)
Obsolete biological process	29	(6%)	21	(3%)
Embryogenesis/ morphogenesis	11	(2%)	8	(1%)
mRNA splicing	6	(1%)	5	(0.7%)
Cytokinesis	5	(1%)	2	(0.3%)
Peroxidase activity	4	(1%)	1	(0.2%)
Antibiotic susceptibility/ resistance	1	(0.2%)	-	-
Heavy metal ion transport	1	(0.2%)	2	(0.3%)
Histogenesis and organogenesis	1	(0.2%)	1	(0.2%)
Behavior	3	(1%)	7	(1%)
Locomotor	2	(0.4%)	1	(0.2%)
Learning/memory	1	(0.2%)	4	(0.4%)

(b) Cellular Component

Categories and subcategories	Representation in males		Representation in L2s	
Cell	372	(89%)	489	(88%)
Intracellular	274	(66%)	351	(63%)
Chromatin	5	(1%)	5	(0.9%)
Nucleosome	4	(1%)	5	(0.9%)
Proteasome	2	(0.5%)	7	(1%)
Importin	1	(0.2%)	1	(0.2%)
Kinetochore	1	(0.2%)	-	-
Mitotic chromosome	1	(0.2%)	-	-
Membrane	115	(28%)	143	(26%)
Extracellular	13	(3%)	32	(6%)
Extracellular space	2	(0.5%)	2	(0.4%)
Extracellular matrix	1	(0.2%)	7	(1%)
Unlocalized	10	(2%)	12	(2%)
Protein serine/ threonine phosphotase	7	(2%)	2	(0.4%)
Protein kinase CK2 complex	2	(0.5%)	1	(0.2%)
cAMP-dependant protein kinase complex	1	(0.2%)	3	(0.5%)
Cell fraction	8	(2%)	12	(2%)
Virion	5	(1%)	5	(0.9%)
Viral capsid	3	(0.7%)	4	(0.7%)

Table 4. (continued)

(c) Molecular Function

Categories and subcategories	Representation in males		Representation in L2s	
Catalytic activity	182	(34%)	303	(40%)
Hydrolase	67	(13%)	100	(13%)
Transferase	50	(12%)	91	(12%)
Oxidoreductase	41	(10%)	54	(7%)
Kinase	26	(6%)	33	(4%)
Ligase	8	(2%)	20	(3%)
Isomerase	6	(1%)	14	(2%)
Helicase	2	(0.4%)	9	(1%)
Lyase	2	(0.4%)	10	(1%)
Small protein conjugating enzyme	1	(0.2%)	1	(0.1%)
Binding	171	(34%)	267	(35%)
Nucleic acid	71	(13%)	91	(12%)
Nucleotide	35	(7%)	70	(9%)
Protein	27	(6%)	46	(6%)
Metal ion	18	(3%)	38	(5%)
Lipid	7	(1%)	8	(1%)
Receptor	3	(0.6%)	4	(0.5%)
Carbohydrate	1	(0.2%)	-	-
Glycosaminoglycan	1	(0.2%)	-	-
Structural molecule activity	69	(17%)	49	(6%)
Ribosome	46	(9%)	33	(4%)
Cytoskeleton	6	(1%)	5	(0.7%)
Cell wall	4	(1%)	1	(0.1%)
Extracellular matrix	2	(0.4%)	1	(0.1%)
Muscle	1	(0.2%)	5	(0.7%)
Transporter activity	57	(11%)	65	(9%)
Carrier activity	24	(5%)	25	(3%)
Ion	17	(3%)	14	(2%)
Electron	13	(2%)	11	(1%)
Protein	5	(1%)	7	(0.9%)
Channel/ pore class	4	(1%)	9	(1%)
Organic acid	3	(0.6%)	2	(0.3%)
Amine/ polyamine	2	(0.4%)	1	(0.1%)
Carbohydrate	1	(0.2%)	4	(0.4%)
Lipid	1	(0.2%)	2	(0.3%)
Signal transducer activity	15	(4%)	34	(4%)
Transcription regulator activity	11	(2%)	23	(3%)
Transcription factor	7	(1%)	9	(1%)
Transcription cofactor	2	(0.4%)	9	(1%)
RNA polymerase II transcription factor	1	(0.2%)	4	(0.4%)
Transcriptional activator	1	(0.2%)	-	-
Chaperone activity	8	(2%)	14	(2%)
Heat shock protein	3	(0.6%)	5	(0.7%)
Co-chaperone	1	(0.2%)	-	-
Hsp70/ Hsp90 organizing protein	1	(0.2%)	1	(0.1%)
Enzyme regulator	5	(1%)	8	(1%)
Enzyme activator	2	(0.4%)	1	(0.1%)
Kinase regulator	2	(0.4%)	4	(0.4%)
Enzyme inhibitor	1	(0.2%)	2	(0.2%)
GTPase regulator	1	(0.2%)	3	(0.4%)
Phoshotase regulator	1	(0.2%)	-	-
Translation regulator activity	3	(0.6%)	9	(1%)
Antioxidant activity	2	(0.4%)	1	(0.1%)
Peroxidase	1	(0.2%)	-	-
Cell adhesion molecule activity	2	(0.4%)	1	(0.1%)
Apoptosis regulatory activity	1	(0.2%)	1	(0.1%)
Apoptosis inhibitor	1	(0.2%)	1	(0.1%)
Motor activity	1	(0.2%)	4	(2%)
Protein tagging activity	1	(0.2%)	1	(0.1%)
Protein degradation tagging	1	(0.2%)	1	(0.1%)
Toxin activity	1	(0.2%)	-	-

Table 5. Gene ontology comparisons of males to other nematodes and *M. incognita* L2s.
GO mappings for *M.incognita* L2s, *C. elegans*, *B. malayi* and *O. volvulus* were obtained from [4].

		% Representation				
		M. *incognita*		*C.elegans*	*B. malayi*	*O. volvulus*
Gene Ontology	Categories and subcategories	males	L2s	all	all	all
Biological process						
	Cell growth and maintainance	70	88	68	91	93
	Cell communication	18	12	16	3	4
Cellular component						
	Cell	89	81	96	99	98
	Extracellular	3	15	2	-	-
	Unlocalized	2	3	0.6	-	1
Molecular function						
	Ligand binding / carrier	32	52	28	24	28
	Enzyme	35	39	39	33	31
	Transporter	11	5	13	6	13
	Signal transducer	3	3	7	2	3
	Structural molecule	13	2	5	17	15
	Enzyme regulator	9	2	1	2	-
	Cell adhesion	0.4	0.4	0.3	-	-
	Motor	0.2	0.4	1	2	3
	Transcription regulator	2	0.4	4	1	1

Table 6. InterproScan GO annotations for *M. incognita* males. Results are for all clusters, from the clustering of all ESTs, that had a significant (1.00e-5) match to the non-redundant protein database at NCBI.

Molecular Function	
structural constituent of ribosome (GO:0003735)	48
calcium ion binding (GO:0005509)	29
nucleic acid binding (GO:0003676)	25
ATP binding (GO:0005524)	22
structural molecule activity (GO:0005198)	18
DNA binding (GO:0003677)	13
molecular_function unknown (GO:0005554)	13
protein kinase activity (GO:0004672)	13
protein tyrosine phosphatase activity (GO:0004725)	10
hydrolase activity (GO:0016787)	6
electron transporter activity (GO:0005489)	5
GTP binding (GO:0005525)	5
protein serine/threonine kinase activity (GO:0004674)	5
RNA binding (GO:0003723)	5
actin binding (GO:0003779)	4
binding (GO:0005488)	4
endopeptidase activity (GO:0004175)	4
ubiquitin conjugating enzyme activity (GO:0004840)	4
calcium-dependent phospholipid binding (GO:0005544)	3
cytochrome c oxidase activity (GO:0004129)	3
kinase activity (GO:0016301)	3
protein domain specific binding (GO:0019904)	3
transferase activity, transferring phosphorus-containing groups (GO:0016772)	3
fructose-bisphosphate aldolase activity (GO:0004332)	2
inositol/phosphatidylinositol phosphatase activity (GO:0004437)	2
lactoylglutathione lyase activity (GO:0004462)	2
lyase activity (GO:0016829)	2
methionine adenosyltransferase activity (GO:0004478)	2
microtubule motor activity (GO:0003777)	2
NADH dehydrogenase (ubiquinone) activity (GO:0008137)	2
pre-mRNA splicing factor activity (GO:0008248)	2
protein phosphatase activity (GO:0004721)	2
protein transporter activity (GO:0008565)	2
structural constituent of cytoskeleton (GO:0005200)	2
translation initiation factor activity (GO:0003743)	2
catalytic activity (GO:0003824)	1
glyceraldehyde 3-phosphate dehydrogenase (phosphorylating) activity (GO:0004365)	1
metal ion binding (GO:0046872)	1
protein tyrosine kinase activity (GO:0004713)	1
RAB small monomeric GTPase activity (GO:0003928)	1
RAS small monomeric GTPase activity (GO:0003930)	1
Rho small monomeric GTPase activity (GO:0003931)	1
S-adenosylmethionine-dependent methyltransferase activity (GO:0008757)	1
small monomeric GTPase activity (GO:0003925)	1
transcription co-activator activity (GO:0003713)	1
transcription factor activity (GO:0003700)	1

Table 6. (continued)

translation elongation factor activity (GO:0003746)	1
transporter activity (GO:0005215)	1
ubiquinol-cytochrome c reductase activity (GO:0008121)	1
Biological Process	
protein biosynthesis (GO:0006412)	47
protein amino acid phosphorylation (GO:0006468)	19
regulation of transcription, DNA-dependent (GO:0006355)	12
protein amino acid dephosphorylation (GO:0006470)	10
proton transport (GO:0015992)	10
electron transport (GO:0006118)	9
intracellular signaling cascade (GO:0007242)	7
ATP biosynthesis (GO:0006754)	5
transport (GO:0006810)	5
protein modification (GO:0006464)	4
small GTPase mediated signal transduction (GO:0007264)	4
ubiquitin cycle (GO:0006512)	4
ubiquitin-dependent protein catabolism (GO:0006511)	4
glycolysis (GO:0006096)	3
amino acid metabolism (GO:0006520)	2
carbohydrate metabolism (GO:0005975)	2
cytoskeleton organization and biogenesis (GO:0007010)	2
dUTP metabolism (GO:0046080)	2
microtubule-based process (GO:0007017)	2
mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	2
mRNA splicing (GO:0006371)	2
one-carbon compound metabolism (GO:0006730)	2
protein-nucleus import (GO:0006606)	2
translational initiation (GO:0006413)	2
chromosome organization and biogenesis (sensu Eukarya) (GO:0007001)	1
nucleosome assembly (GO:0006334)	1
protein transport (GO:0015031)	1
translational elongation (GO:0006414)	1
Cellular Component	
intracellular (GO:0005622)	50
ribosome (GO:0005840)	45
nucleus (GO:0005634)	7
integral to membrane (GO:0016021)	6
extracellular (GO:0005576)	5
membrane (GO:0016020)	5
actin cytoskeleton (GO:0015629)	4
mitochondrial inner membrane (GO:0005743)	4
proteasome core complex (sensu Eukarya) (GO:0005839)	4
cytoplasm (GO:0005737)	2
microtubule associated complex (GO:0005875)	2
mitochondrial electron transport chain (GO:0005746)	2
mitochondrial membrane (GO:0005740)	2
small nucleolar ribonucleoprotein complex (GO:0005732)	2
small ribosomal subunit (GO:0015935)	2
nucleosome (GO:0000786)	1

APPENDIX

Script 1. PERL script used to trim polyA tails before clustering.

```
!/usr/bin/perl
use strict;
use warnings;
use Getopt::Std;

my $filename = undef;

while ($filename = <SMARTb/screened/*.screen>)
{
        my $seq = `less $filename`; # reads .screen files into $seq
        print "The original seq looks like this: \n$seq\n";
        #removes polyA, internal 'X's and 'n's
        $seq =~s/[A\n]{10,400}[ACGT\n]{0,10}[A\n]{0,100}[Xn\n]{0,1000}
        [ATCG]{0,100}$//g;
        print "After trimming: \n$seq\n";
        open (OUT, ">$filename.out"); # stores cleaned data in new file .out
        print OUT "$seq\n";
}
```

Script 2. PERL script used to trim Xs and remove short sequences after clustering.

```
!/usr/bin/perl
use strict;
use warnings;
use Getopt::Std;

my $filename = undef;
while ($filename = <Fasta/*.new>)
{
        my $seq = `less $filename`; # reads .screen files into $seq
        print "The original seq looks like this: \n$seq\n";
        $seq =~ s/(X\n)//g;$seq =~ s/X//g; # removes Xs
        #removes polyA, internal 'X's and 'n's
        $seq =~ s/[A\n]{10,400}[ACGT\n]{0,50}[A\n]{0,100}$//g;

        my @seq = split/./,$seq;
        if (scalar(@seq) < 20) # sorts long and shorter than 20bp reads after triming
                {
                $seq =~ s/\n|.//g; # completely removes short reads
                print "Short sequence.\n$seq\n";
                }
        else
                {
                print "Long sequence.";
                }
  print "After trimming: \n$seq\n";
  open (OUT, ">$filename.out");# stores cleaned data in new file .out
  print OUT "$seq\n";
}
```

Script 3. PERL script used to truncate names in a concatenated FASTA file.

```
!/usr/bin/perl
use strict;
use warnings;
use Getopt::Std;


#opens a .fasta file, trims the header and generates a .new file.

my $filename = undef;

while ($filename = <Fasta/*.fasta>)
{
	open (OUT, ">$filename.new");
	my $seq = `less $filename`;
	my @seq = split/\n/,$seq;
	foreach my $line(@seq)
	{
		if ($line =~ m/^>/) #line of header must start with >
			{
			$line =~ s/ .*//; #leaves only the fisrt 'name' after the >
			}
		print OUT "$line\n";
	}
}
```

Table 1. **Capillary to well mapping.** This matrix can be used to locate the well id of clones from the four test plates named SMARTa, SMARTb, SL1a and SL1b. The column and row headers are the plate well addresses and the capillary numbers in the matrix match the numbers at the end of the clone names. Thus clone name SMARTb_36 is found in well G06 of the SMARTb test plate.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	15	16	31	32	47	48	63	64	79	80	95	96
B	13	14	29	30	45	46	61	62	77	78	93	94
C	11	12	27	28	43	44	59	60	75	76	91	92
D	9	10	25	26	41	42	57	58	73	74	89	90
E	7	8	23	24	39	40	55	56	71	72	87	88
F	5	6	21	22	37	38	53	54	69	70	85	86
G	3	4	19	20	35	36	51	52	67	68	83	84
H	1	2	17	18	33	34	49	50	65	66	81	82

LITERATURE CITED

Agrios, G.N. (1997). Plant Diseases Caused by Nematodes, In: Agrios G.N. (Ed). *Plant Pathology*. New York, NY, USA: Academic Press, pp. 565 - 597.

Altchul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215: 403 - 410.

Arango, N.A., Lovell-Badge, R. & Behringer, R.E. (1999). Targetet mutagenesis of the the endogenous mouse *Mis* gene promoter: In vivo definition of genetic pathways of sexual development. *Cell* 99: 409 - 419.

Bardoni, B. et al. (1994). A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal. *Nature Genetics* 7: 497 – 501.

Bell, L.R. Maine, E.M., Schedl, P. and Cline, T.W. (1988). *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* 55: 1037 - 1046.

Bell, L.R., Horabin, J.I., Schedl, P. and Cline, T.W. (1991). Positive autoregulaiton of *sex-lethal* by alternative splicing maintains the female determined state in *Drosophila*. *Cell* 65: 229 - 239.

Bird, D. McK. and Opperman, C.H. (1998). *Caenorhabditis elegans*: a genetic guide to parasitic nematode biology. *Journal of Nematology* 30: 299 - 308.

Bird, D. McK., Opperman, C.H., Jones, S.J.M. & Bailie, D.L. (1999). The *Caenorhabditis elegans* genome: a guide in the post genomic age. *Annual Review of Phytopathology* 37: 247 – 265.

Blaxter, M.L., De Lay, P., Garey, J.R., Liu, L.X., Scheideman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T. and Thomas W. K.

(1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* 392: 71 – 75.

Blumenthal, T. and Steward, K. (1997). RNA processing and gene structure, In: Riddle, D.L., Blumentstiel, B., Meyer, B.J., Priess, J.R. (Eds). *C. elegans II*. Plainveiw, NY, USA: Cold Spring Harbor Laboratory Press, pp.117 - 145.

Bridges, C.B. (1925). Sex in relation to chromosomes and genes. *American Naturalist* 59: 127 – 137.

Carmi, I., Kopczynski, J.B. and Meyer, B.J. (1998). The nuclear hormone receptor SEX-1 is an X-chromasome signal that determines nematode sex. *Nature* 396: 168-173.

Castagnone-Sereno, P., Leroy, F. and Abad, P. (2001). cDNA cloning and expression analysis of a calponin gene from the plant-parasitic nematode *Meloidogyne incognita*. *Molecular & Biochemical Parasitology* 112: 149–152.

Colvin, J., Green, R.P., Schmal, J., Capel, B. & Ornitz, D.M. (2001). Male-to-female sex reversal in mice lacking fibroblast factor 9. *Cell* 104: 875-889.

Cook, R. and Evans, K. (1987) Resistance and tolerance. In: Brown, R.H. and Kerry, B.R. (Eds.) *Principles and practice of nematode control in crops*. Orlando, FL, USA: Academic Press, pp. 179 - 231.

Couse, J.F. and Kourach, K.S. (2001). Contrasting phenotypes in reproductive tissue of estrogen receptor null mice. *Annals of the New York Academy of Science* 948: 1-8.

Dautova, M., Rosso, M., Abad, P., Gommers, F., Bakker, J. & Smant, G. (2001). Single pass cDNA sequencing – a powerful tool to analyse gene expression in preparasitic juveniles of the southern root-knot nematode *Meloidogyne incognita*. *Nematology* 3: 129 - 139.

Davide, R.G. (1965). Influence of the environment on development and sex differentiation of *Meloidogyne incognita* and *M. javanica*. PhD thesis, North Carolina State University, Raleigh, NC, USA.

Davide, R.G. & Triantaphyllou, A.C. (1967). Influence of the environment on development and sex differentiation of root-knot nematodes I. Effect of infection density, age of the host plant and soil temperature. *Nematologica* 13: 102 - 110.

Davide, R.G. & Triantaphyllou, A.C. (1967). Influence of the environment on development and sex differentiation of root-knot nematodes II. Effect of host nutrition. *Nematologica* 13: 111 - 117.

Davide, R.G. & Triantaphyllou, A.C. (1968). Influence of the environment on development and sex differentiation of root-knot nematodes III. Effect of foliar application of maleic hydrazide. *Nematologica* 14: 37 - 46.

Davis, R. E. (1996). Spliced leader RNA trans-splicing in metazoa. *Parasitology Today* 12: 33 – 40.

De Ley, I.T., De Ley, P. Vierstraete, A., Karssen, G., Moens, M. and Vanfleteren, J. (2002). Phylogenetic analyses of *Meloidogyne* small subunit rDNA. *Journal of Nematology* 34: 319 - 327.

de Santa Barabara, P., Monoit, B., poulat, F., Berta, P. (2000). Expression and subcellular localization of SF-1, SOX9, WT1, and AMH proteins during early human testicular development. *Developmental Dynamics* 217: 293 - 298.

Diefenbach, J. and Kindl, H. (2000). The membrane-bound DnaJ protein located at the cytosolic site of glyoxysomes specifically binds the cytosolic isoform 1 of Hsp70 but not other Hsp70 species. *European Journal of Biochemistry* 267: 746 - 754.

Eisenback, J.D. (1985). Detailed morphology and anatomy of second-stage juveniles, males, and females of the genus *Meloidogyne* (root-knot nematodes), In: Sasser, J.N. & Carter, C.C. (Eds). *An Advanced Treatise on Meloidogyne, Volume II: Biology and Control*. Raleigh, NC, USA: North Carolina State University Graphics, pp. 47 - 78.

Evans, D., Zorio, D., MacMorris, M., Winter, C.E., Lea, K. and Blumenthal, T. (1997). Operons and SL2 trans-splicing exist in nematodes outside the genus Caenorhabditits. *Proceedings of the National Academy of Sciences* 94: 9751 – 9756.

Ewing, B., Hillier, L., Wendl, M.C. & Green, P. (1998). Base-calling of automated sequence traces using Phred. I. Accuracy assessment. *Genome Research* 8: 175 - 185.

Ewing, B. & Green, P. (1998). Base-calling of automated sequence traces using Phred. II. Error probabilities. *Genome Research* 8: 186 – 194.

Ewing, B. and Green, P. (2000). Analysis of expressed sequence tags indicates 35,000 human genes. *Nature Genetics* 25: 232 - 234.

Forrest, J.M.S., Trudgill, D.L. and Cotes, L. M. (1986) The fate of juveniles of *Globodera rostochiensis* and *G. pallida* in roots of susceptible and resistant potato cultivars with gene H1. *Nematologica* 32: 106 – 114.

Frieman, R.N., Albright, S., Zehng, S., Sha, W.C., Hammer, R.E. and Tjian, R. (2002). Requirements of tissue specific TBP-associated factor TAFII105 in ovarian development. *Science* 293: 2084 - 2087.

Ganguly, A. and Dasgupta, D. (1984). Effect of maleic hydrazide, a plant growth inhibitor, on the biology of root-knot nematode, *Meloidogyne incognita*. *Nematologica Mediterania* 12: 23 - 29.

Giese, K., Cox, J. and Grosschedl. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates the assembly of functional nucleoprotein structures. *Cell* 69: 185 - 195.

Goellner, M., Wang, X., Davis, E.L. (2001). Endo-beta-1,4-glucanase expression in compatible plant-nematode interactions. *Plant Cell*. 10:2241 - 2255.

Goldstein, P. (1981). Sex determination in nematodes, In: Zuckerman B.M. and Rhode, R.A. (Eds). *Plant Parasitic Nematodes Vol. III*. New York, NY, USA: Academic Press, pp. 37 - 60.

Hafizur, R.M., Yano, M., Gotoh, T., Mori, M., Terada, K. (2004). Modulation of chaperone activities of Hsp70 and Hsp70-2 by a mammalian DnaJ/Hsp40 homolog, DjA4. *Journal of Biochemistry* (Tokyo) 135: 193 – 200.

Hamilton, W.D. (1980) Sex versus non-sex versus parasite. *Oikos*, 35: 282 - 290.

Harvey, S.C., Gemmill, A.W., Read, A.F. and Viney, M.E. (2000). The control of morph development in the parasitic nematode *Strongyloides ratti*. *Proceedings of the Royal Society of London Series B* 267: 2057 - 2063.

Hirai, H., Sakaguchi, Y. and Tada, I. (1985). Chromasomes of *Onchocerca volvulus* and *O. gutturosa*. *Zeitschrift für Parasitenkun Parasitoloy Research* 71: 135 - 139.

Hoekstra, R., Visser, A., Myrthe, O., Tibben, J., Lenstra, J.A. & Roos, M.H. (2000). EST sequencing of the parasitic nematode *Haemonchus contortus* suggests a shift in gene expression during the transition to the parasitic stages. *Molecular and Biochemical Parasitology* 110: 53 – 68.

Hussey, R.S. and Barker, K.R. (1973). A comparison of methods of collecting inocula of *Meloidogyne spp*., including a new technique. *Plant Disease Reporter* 57: 1025 - 1028.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V. and Kim, S.K. (2001). Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proceedings of the National Acadamy of Sciences (USA)* 98: 218 - 223.

Keyes, L.N., Cline, T.W. and Schedl, P. (1992). The primary sex determination signal of *Drosophila* acts at the level of transcription. *Cell* 68: 933 - 943.

Kimber, M.J., Fleming, C.C., Prior, A., Jones, J.T., Halton, D.W. and Maule, A.G. (2002). Localisation of *Globodera pallida* FMRFamide-related peptide encoding genes using in situ hybridisation. *International Journal of Parasitology*. 32: 1095 - 105.

Koening, S.R. and Barker, K.R. (1985) Gnotobiotic techniques and plant-paerasitic nematodes, In: Barker, K.R., Carter, C.C. and Sasser, J.N. (Eds). *An Advanced Treatise on Meloidogyne, Volume II: Methodology*. Raleigh, NC, USA: North Carolina State University Graphics, pp. 49 - 68.

Lambert, K.N., Tedford, E.C., Caswell, E.P. and Williamson, V.M. (1992). A system for continuous production of root-knot nematode juveniles in hydroponic culture. *Phytopathology* 82: 512 - 515.

McCarter, J., Miterva, M., Martin, J., Dante, M., Wylie, T., Rao, U., Pape, D., Bowers, Y., THeising, B., Murphey, C., Kloek, A., Chiapelli, B., Clifton, S., Bird, D. McK., Waterston, R. (2003). Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*. *Genome Biology* 4: R26.1 - R26.19.

McCarter, J., Abad, P., Jones, J.& Bird, D. (2000). Rapid gene discovery in plant parasitic nematodes via Expressed Sequence Tags. *Nematology* 2: 719 - 731.

Meccariello, R., Cobellis, G., Scarpa, D., Fienga, G., Pierantoni, R., Fasano, S. (2004). Detection of msj-1 gene expression in the frog, *Rana esculenta* testis, brain, and spinal cord. *Molecular and Reproductive Development* 68: 149 - 158.

Moura , R.M., Davis E.L., Luzzi, B.M., Boerma, H.R. & Hussey, R.S. (1993). Post-infectional development of Meloidogyne incognita on susceptible and resistant soybean genotypes. Nematropica 23: 7 - 13.

Mutafova, T. (1995). Meiosis and some aspects of cytological mechanisms of chromosomal sex determination in nematode species. *International Journal of Parasitology* 25: 453 – 462.

Nagoshi, R.N., McKeown, M., Burtis, K.C., Belote, J.M., and Baker, B.S. (1988). The control of alternative splicing at genes regulating sex differentiation in *D. melanogaster*. Cell 53: 229 - 236.

Nee, S. (1989). Antogonistic co-evolution and the evolution of genotypic randomization. *Journal of Theoretical Biology* 140: 499 - 518.

Nef, S., Verma-Kurvari, S., Merenmies, J., Vassalli, J.D., Efstratiadis, A., Accili, D., Parada, L.F. (2003). Testis determination requires insulin receptor family function in mice. *Nature* 426: 291 – 295.

Nicoll, M., Akerib, C.C. and Meyer, B.J. (1997). X-chromasome counting mechanisms that determine nematode sex. Nature 388: 200 - 204.

Niebel, A., Heungens, K., Barthels, N., Inze, D., Van Montagu, M., Gheysen, G. (1995). Characterization of a pathogen-induced potato catalase and its systemic expression upon nematode and bacterial infection. *Molecular Plant Microbe Interactions* 8: 371 - 378.

Nilsen, T.W. (1993). Trans-splicing of nematode premessenger RNA. *Annual Review of Microbiology* 47: 413 - 440.

Nusbaum, C. and Meyer, B.J. (1989). The *Caenorhabditis elegans* gene *sdc-2* controls sex determination and dosage compensation in XX animals. *Genetics* 122: 579 – 593.

Ohe, K., Lalli, E. and Sassone-Corsi, P. (2002). A direct role of SRY and SOX proteins in pre-mRNA splicing. *Proceedings of the National Academy of Sciences*. 99: 1146- 1 151.

Orion, D. and Minz, D. (1971). The influence of morphactin on the root-knot nematode *Meloidogyne javanica* and its galls. *Nematologica* 19: 107 - 112.

Otto, S.P. and Michalakis, Y. (1998). The evolution of recombination in changing environments. *Trends in Ecology and Evolution* 13: 145 - 151.

Papadopoulou, J. and Triantaphyllou, A.C. (1982). Sex differentiation in *Meloidogyne incognita* and anatomical evidence of sex reversal. *Journal of Nematology* 14: 549 - 566.

Pask, A. and Graves, J.A. (1999). Sex chromosomes and sex-determining genes: Insights from marsupials and monotremes. *Cellular and Molecular Life Science* 55: 864 - 875.

Peacock, F. (1960). Inhibition of root-knot development on tomato by systemic compounds. *Nematologica* 5: 219 - 227.

Popeijus H., Blok, V.C., Cardle, L., Bakker, E., Phillips, M.S., Helder, J., Smant, G. and Jones, J.T. (2000). Analysis of genes expressed in second stage juveniles of the potato cyst nematode *Globodera rostochiensis* and *G. pallida* using the expressed sequence tag approach. *Nematology* 2: 567 – 574.

Ray, C., Abbott, A.G. and Hussey, R.S. (1994). Trans-splicing of a *Meloidogyne incognita* mRNA encoding a putative esophageal gland protein. *Molecular and Biochemical Parasitology* 68: 93 - 101.

Redmond, D.L. and Knox, D.P. (2001). *Haemonchus contortus* SL2 trans-spliced RNA leader sequence. *Molecular and Biochemical Parasitology* 117: 107 -110.

Raymond, C.S., Shamu, C.E., Shen, M.M., Seifert, K.J., Hirch, B., Hodjkin, B. and Zarkower, D. (1998). Evidence for evolutionary conservation of sex-determining genes. *Nature* 391: 691 - 695.

Reik, W. and Surani, M. (1997). *Frontiers in Molecular Biology: Genomic Imprinting*. Oxford, UK: Oxford University Press.

Reiner, D.J. and Thomas, J.H. (1995). Reversal of a muscle response to GABA during *C. elegans* male development. *Journal of Neuroscience*. 15: 6094 - 6102.

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

Rosso, M.N., Favery, B., Piotte, C., Arthaud, L., De Boer, J.M., Hussey, R.S., Bakker, J., Baum, T.J., Abad, P. (1999). Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. *Molecular Plant Microbe Interactions* 12: 585 - 91.

Ryner, L.C. and Baker, B.S. (1991). Regulation of *doublesex* pre-mRNA processing occurs by 3' splice site activation. *Genes and Development* 5: 2071 - 2085.

Sakaguchi, Y., Tada, I., Ash, L.R. and Aoki, Y. (1983). Karyotypes of *Brugia pahangi* and *Brugia malayi* (Nematoda: Filarioidea). *Journal of Parasitology* 69: 1090 – 1093.

Salz, H.K., Maine, M.E., Keys, L.N., Samuels, M.E., Cline, T.W. and Schedl, P. (1989). The *Drosophila* sex-determination gene, *Sex-lethal* has stage-, tissue- and sex-specific RNAs suggesting multiple modes of regulation. *Genes and Development* 3: 708 - 719.

Sasser, J.N. and Freckman, D.W. (1987). A World Perspective on Nematology: The Role of the Society, In: Veech, J.A., Dickson, D.W. (Eds). *Vistas on Nematology*. Hyattsville, MD, USA: Society of Nematology, pp. 7 - 14.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A., Lovell-Badge, R. & Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346: 240 - 244.

Smant, G., Stokkermans, J.P., Yan, Y., de Boer, J.M., Baum, T.J., Wang, X., Hussey, R.S., Gommers, F.J., Henrissat, B., Davis, E.L., Helder, J., Schots, A., Bakker, J. (1998). Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences* 95: 4906 - 11.

Smith, C.A. and Sinclair, A.H. (2001). Sex-determination in the chicken embryo. *Journal of Experimental Zoology* 290: 691 - 699.

Smith, J.M. (1971) Evolution in sexual and asexual populations. In: Williams, G.C. (Ed). *Group Selection*, Chicago, IL, USA: Aldine-Atherton, pp. 128 - 177.

Smith, J.M. (1978). *The evolution of sex.* Cambridge, MA, USA: Cambridge University Press.

Spencer, A.G., Orita, S., Malone, C.J.and Han, M. (2001). A RHO GTPase-mediated pathway is required during P cell migration in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* 98: 13132 – 13137.

Spieth, J., Brooke, G., Kuersten, S., Lea, K. & Blumenthal, T. (1993). Operons in *C. elegans*-polycistronic messenger-RNA precursors are processed by splicing of SL2 to downstream coding regions. *Cell* 73**:** 521 – 532.

Stothard, P. and Pilgrim, D. (2003). Sex-determination gene and pathway evolution in nematodes. *BioEssays* 25: 221 – 231.

Swain, A., Narvaez, V., Burgoyne, P., Camerino, G. and Lovell-Badge, R. (1998). Dax1 antagonizes Sry action in mammalian sex determination. *Nature* 391: 761 - 767.

Taylor, A.L. and Sasser, J.N. (1978). Biology, identification and control of root-knot nematodes (*Meloidogyne species*). Raleigh, NC, USA: United States Agency for International Development.

Triantaphyllou, A.C. (1985). Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes, In: Sasser, J.N. and Carter, C.C. (Eds). *An Advanced Treatise on Meloidogyne, Volume II: Biology and Control*. Raleigh, NC, USA: North Carolina State University Graphics, pp. 113 - 126.

Triantaphyllou, A.C. (1973). Environmental sex differentiation of nematodes in relation to pest management. *Annual Revue of Phytopathology* 11: 441 - 462.

Underwood, A.P. and Bianco, A.E. (1999). Identification of a molecular marker for the Y chromosome of *Brugia malayi*. *Molecular and Biochemical Parasitology* 99: 1 – 10.

Van Doren, M., Ellis, H.M. and Posakony, J.W. (1991). The *Drosophila* extramacrochaetae protein antagonizes sequence-specific DNA binding by daughterless/achaete-scute protein complexes. *Development* 113: 245 - 255.

Vercauteren, I., Van Der Schueren, E., Van Montagu, M., Gheysen, G. (2001). *Arabidopsis thaliana* genes expressed in the early compatible interaction with root-knot nematodes. *Molecular Plant Microbe Interactions* 14: 288 - 299.

Villeneuve, A.M. and Meyer, B.J. (1987). *sdc-1*: A link between sex determination and dosage compensation in *C. elegans*. *Cell* 48: 25 – 37.

Viney, M.E. (1994). A genetic analysis of reproduction in *Strongyloides ratti*. *Parasitology* 109: 511 – 515.

Viney, M.E., Matthews, B.E. and Walliker, D. (1993). Mating in the nematode parasite *Strongyloides ratti*: proof of genetic exchange. *Proceedings of the Royal Society of London Series B* 254: 45 – 52.

Walton, A.C. (1940). Gametogenesis. In: *An Introduction to Nematology*, Chitwood, G.B. and Chitwood, M.B. (Eds). New York, NY, USA: Babylon Press, pp. 205 - 215.

Wilkins, A.S. (2002). *The Evolution of Developmental Pathways*, Sunderland, MA, USA: Sinauer Associates, pp. 173 - 204.

Williamson, V.M. and Hussey, R.S. (1996) Nematode pathogenesis and resistance in plants. *The Plant Cell* 8: 1735 - 1745.

Wilson, M.A., Bird, D.McK. and van der Knap, E. (1994). A comprehensive subtractive cDNA cloning approach to identify nematode-induced transcripts in tomato. *Phytopathology* 84: 299 – 303.

Younger-Shepherd, S., Vaessin, H., Bier, E., Jan, L.Y. and Jan, Y.N. (1992). *deadpan*, an essential pan-neural gene encoding an HLH protein, acts as a denominator in *Drosophila* sex determination. *Cell* 70: 911 - 922.

Zanaria, E. and 13 others. (1994). An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature* 372, 635-641.

Zorio, D. A. R., Cheng, N. N., Blumenthal, T. & Spieth, J. (1994) Operons as a common form of chromosomal organization in *C. elegans*. *Nature* 372: 270 – 272.