COMPARISON OF FOSMID LIBRARIES MADE FROM TWO GEOGRAPHIC ISOLATES OF *CAENORHABDITIS ELEGANS*

by

Jaryn Daniel Perkins

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA

(VANCOUVER)

February 2011

Abstract

To fill a need in the *Caenorhabditis elegans* community for genomic DNA held in manageably sized clones for complementation assays, a fosmid library was made from the N2 strain. These fosmid clones were aligned to the canonical sequence and cover 80% of the genome, but there were 396 gaps in contiguous coverage spread over the worm's six chromosomes. In an attempt to fill in some of these gaps in the original fosmid clones' sequence, we made another library from the Hawaiian geographic isolate CB4856. Our hope was that the divergence, inherent in the deletions containing 517 genes, between the two genomes would aid in the capture of previously gapped regions. This hope was justified. This thesis outlines the production and comparison of the two *C. elegans* fosmid libraries made from N2 and CB4856 and provides evidence that the way genomic libraries are made can affect the sequences packaged. Combining the two libraries, we now have a total coverage of 92.8% of genes and 90.43% of sequence in relation to the N2 canonical genome.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

#	number
%	percent
λ	lambda
μl	microlitre
°C	degrees Celsius
8P	8 times peptone media
BAC	bacterial artificial chromosome
Blast	basic local alignment search tool
bp	base pairs
CA	cytosine adenine
CGC	Caenorhabditis genetics center
CHCl3	chloroform
contig	contiguous overlapping DNA
CsCl	cesium chloride
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol
F1	first generation
GFP	green fluorescent protein
HAE	HEPES acetate EDTA buffer
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

hr	hour
I	chromosome I
i.e.	*id est.*
IGEPAL	tert-octylphenoxy poly(oxyethylene)ethanol
II	chromosome II
III	chromosome III
indels	insertions and deletions
IV	chromosome IV
kb	kilo base pairs
KCl	Potassium Chloride
kD	kilodalton
L1	Larval stage 1
M9	Minimal Media Salt Solution 9
mg	milligram
MgCl2	Magnesium Chloride
ml	millilitre
mm	millimeter
mM	millimolar
NaOAc	Sodium Acetate
ng/µl	nanograms per microlitre
NGM	Nematode Growth Media
pH	power of hydrogen
RNAi	ribonucleic acid interference

rpm	revolutions per minute
SCODA	synchronus coefficient of drag alteration
SNP	single nucleotide polymorphism
TBE	tris borate EDTA buffer
TE	tris(hydroxymethyl)aminomethane hydrochloride EDTA buffer
TG	thymine guanine
Tris HCl	tris(hydroxymethyl)aminomethane hydrochloride
µl	microlitre
V	chromosome V
X	chromosome X
x	times
YAC	yeast artificial chromosome

Acknowledgements

I would like to thank past and present students Jason Maydan, Ryan Viveiros and Adam Warner for endless discussions and providing ideas to circumvent problems that arose. Special thanks are owed to Stephane Flibotte and Jeff Magnusson, whose guidance and bioinformatics prowess kept me from going blind staring at the reams of data necessary to complete this work. I would also like to thank Don Moerman who gave me the opportunity to work and study in his lab. Finally to Jenb, thank you for all of your support, compassion and understanding when I was most challenged.

1 Introduction

1.1 *Caenorhabditis elegans*

The suitability of *Caenorhabditis elegans* as a model organism for studies in developmental biology and exploration of genetic interaction networks was first described by Sydney Brenner (1974). The hermaphroditic nematode he portrayed is capable of self-fertilization, or cross-fertilization with males formed by genetic non-disjunction. This fertilization plasticity provides for simple stock maintenance and the ability to construct strains with desired genotypes, or to perform complementation tests. The organisms 3.5 day lifecycle at 20°C and large brood size (250-300) enhance the benefit that their clear body wall and eggshell provides for the discernment of structural variation from the outside of the worm. The easy and inexpensive strain maintenance, including freezing organisms for long-term storage, was also attractive for laboratory research (Brenner 1974). The discovery of an invariant cell lineage both post-embryonically (Sulston and Horvitz 1977) and embryonically (Sulston *et al.* 1983) provided further utility. This detailed description of the cell lineage led directly to the discovery of programmed cell death (Ellis and Horvitz 1986) and makes possible genetic analysis of lineage commitment in development. Gene transfer technology (Fire and Waterston 1989; Mello *et al.* 1991) allowing extragenic

complementation analysis and the introduction of artificial constructs added to a growing list of experimental benefits the worm provides. The power that gene transfer technology furnished was later exploited by Martin Chalfie and colleagues to demonstrate the usefulness of green fluorescent protein (GFP) to monitor gene expression patterns in vivo (Chalfie *et al.* 1994).

An ongoing goal to create mutations in every gene within the worm is underway (reviewed in Moerman and Barstead 2008). Of the 20,000 expected genes, described in Moerman and Barstead 2008, 7,000 mutations in 5,500 genes have been produced. Mutations in over 4,000 genes were produced by the Barstead, Moerman and Mitani international knockout consortium. The production of an 11,984 member library of coding sequence in a transferable vector allowed for whole genome analysis of protein interactions (Li *et al.* 2004; Vaglio *et al.* 2003) and of expression studies (Huang *et al.* 2003; Luan *et al.* 2004). Finally, the discovery of RNAi (Fire *et al.* 1998), and the subsequent understanding that bacteria expressing the double stranded DNA could be fed to worms and interfere with protein expression (Timmons and Fire 1998), allowed the production of a simple genome wide knockdown library (Fraser *et al.* 2000). These resources have allowed *C. elegans* to be one of the models at the frontline of genomics research on multicellular organisms. These studies were all made possible because C. elegans was the first multicellular organism with a completed genome. The sequence provided a framework upon which they could be produced and relied on the production of a physical map (Coulson *et al.* 1986) using cosmids and Yeast Artificial Chromosomes (YACs) (Consortium 1998).

With these genomic resources available, as well as the growing number of reverse genetic mutants, a comprehensive library of genomic DNA would provide a valuable addition as a method for complementation analysis and a system for producing functional fusions with native promoter regions. A genomic library could find utility in allowing subcloning of regions of the genome that, due to size or repetitive elements, may be difficult to clone. The ideal library would be made of clones in a size that allows the containment of a majority of the regulatory elements necessary for native expression (Dolphin and Hope 2006). As the YAC and cosmid clones produced for the sequencing project can rearrange and have been lost, it became necessary to construct a library using a more stable vector.

1.2 Fosmids

Rather than relying on a library containing genomic DNA between 100 to 3,000 kb, such as Bacteria Artificial Chromosome (BAC), making a 40-kb insert clone library was considered more useful and practical. The larger insert YAC and BAC clones are not trivial to manipulate and due to the large number of genes associated with them often do not provide adequate genetic resolution (Bauchwitz and Costantini 1998; Giraldo and Montoliu 2001). As well, some inserts in these large vectors display instability (Neil *et al.* 1990; Song *et al.* 2001; Yokobata *et al.* 1991). We chose to make a fosmid rather than a cosmid library (Kim *et al.* 1992) to contain the 40-kb

inserts. Cosmid libraries have been excellent tools, but many clones within the libraries have been prone to rearrangements or excision, resulting in a loss of viability due to their large size (Yokobata *et al.* 1991). In order to reduce the occurrence of such rearrangements, the fosmid vector pCC1FOS from Epicentre was used to maintain clones at low copy number until induced (Kim *et al.* 1992). The use of fosmids as backbones allows for the maintenance of large pieces of DNA (around 40 kb) in limited number (1-5) per bacterial host.

In 2005, I produced a fosmid library that was end-sequenced and mapped to the current genome with five-fold coverage (Perkins *et al.*, unpublished results). This library has been used by many laboratories to study individual genes and in several whole genome projects (Dolphin and Hope 2006; Tursun *et al.* 2009; Zhang *et al.* 2008). The fosmids are currently being used by the ModEncode project, headed by Robert Waterston at the University of Washington, to determine the genome wide binding sites for transcription factors within *C. elegans* (Celniker *et al.* 2009). The 2005 library encompasses 80% of the genome and 84% of the genes within 15,784 clones. A library containing complete genetic coverage and similarly sized inserts, allowing even tiling across the genome, would allow constructs with similar sizes and backbones to minimize possible spurious vector effect differences and allow more complete analysis of the genome. It was decided that using a genetic variant of N2 might allow sequences to be cloned which have been previously unattainable in large insert bacterial vectors. Regions, which have inhibited packaging or propagation, may

be sufficiently different in a divergent genome to allow production and replication of clones. The Hawaiian geographic isolate of *C. elegans* was chosen for this purpose.

1.3 Hawaiian Strain

The Hawaiian isolate of *C. elegans*, CB4856, is the most divergent of all the geographic isolates from the canonical Bristol, N2. This divergence was determined initially by Single Nucleotide Polymorphism (SNP) analysis (Denver *et al.* 2003; Swan *et al.* 2002; Wicks *et al.* 2001) and later by copy number variation. CB4856 contains deletions in 517 genes compared to N2 (Maydan *et al.* 2007; Maydan *et al.* 2010). The divergence within the Hawaiian isolate has found utility, within the worm community, in the mapping of gene mutations using SNP variation between the N2 and CB4856 (Flibotte *et al.* 2009; Maydan *et al.* 2007; Swan *et al.* 2002; Wicks *et al.* 2001). This use has prompted the re-sequencing of the CB4856 genome using Solexa sequencing (Marra, Moerman and Waterston, unpublished). The divergence is so great between the two strains that a genetic incompatibility was noticed when crossing geographic isolates. The incompatibility is caused by a gene deletion in the Hawaiian isolate that causes a paternally associated embryonic arrest and lethality of a specific diplotype from a mating of the F1 generation (Seidel *et al.* 2008). This result implies a form of sex-linked gene silencing due to a persistent chromosomal or gene difference or alteration that allows for an abnormal pattern of inheritance.

Genetic incompatibility is found in all of the known geographic isolates. Most isolates are compatible with either N2 or CB4856; only one isolate, from Roxel, Germany (Seidel *et al.* 2008), can mate successfully with both the Bristol and Hawaiian strains. The incompatibility results from balancing selection and allows for the maintenance of differing haplotypes within mixed populations. The haplotypes may be lost due to genetic drift favouring maintenance of non-detrimental alleles within a hermaphroditic population. With the number of gene deletions residing in the Hawaiian isolate, it has been suggested that a better understanding of the differences in genomic architecture between strains may possibly provide clearer insight into the molecular mechanisms underlying this phenomenon (Seidel *et al.* 2008). This unique evolutionary system may provide an opportunity to explore how the genetic diversity within hermaphroditic populations are developed and maintained.

1.4 Cloning Issues Leading to Gaps

20 % of the *C. elegans* genome was determined to be unclonable in *Escherichia coli* with cosmids (Waterston and Sulston 1995). The remaining gaps could only be covered using large insert YACs (Consortium 1998). Unclonable DNA has been previously associated with the presence of repetitive elements, palindromic sequences, Z DNA forming sequences and/or methylated DNA causing deletion or rearrangements in bacterial hosts (Yokobata *et al.* 1991). The presence of kinkable elements in DNA, made up of TG and CA dinucleotide repeats at specific intervals,

can cause local unwinding of the double helix. The resultant single stranded DNA is more likely to be part of molecular repair excision (Mcnamara *et al.* 1990; Razin *et al.* 2001) and may also result in unclonable regions due to the inability to process the secondary structures produced.

The regions that form these unclonable elements do not necessarily account for some of the inserts that have been seen during paired-end sequencing. Unconventional inserts are those that appear to align with ends on two different chromosomes, aligned in the same direction, aligned in the opposite direction away from one another an aligned with insert sizes calculated too large or small for packaging by the Lambda phage packaging extract. Termed discordant (Tuzun *et al.* 2005) or invalid (Bashir *et al.* 2008; Volik *et al.* 2003) these unconventional inserts may show differences in chromosomal architecture between similar genomes (Blakesley *et al.* 2010; Flicek and Birney 2009; Tuzun *et al.* 2005; Volik *et al.* 2003). However, these larger chromosomal rearrangements may be less common than would account for the quantities of unconventional inserts seen in the N2 library and the difference may be due to the misalignment of sequence.

Repetitive elements in a genome, when cloned, are frequently associated with inserts, which do not align conventionally (Razin *et al.* 2001; Yokobata *et al.* 1991). These regions in general provide unique difficulties for sequencing technologies. Multiple repeats of single, di, or trinucleotides can produce polymerase slippage (Murray *et al.* 1993), or template switching (Odelberg *et al.* 1995) that may cause sequencing

artifacts in these areas. Repeated elements can also produce problems with alignment due to small variations in sequence. The variations can create difficulties discerning the elements from one another and can be problematic for positioning (Flicek and Birney 2009; Metzker 2010) even without considering the issues caused by mis-sequencing. Without proper alignment, these regions cannot be tiled and gaps will occur in the contiguous sequence, leading to an incomplete genomic resource.

1.5 A New Fosmid Library

The production of the N2 library has garnered a lot of attention and with it inquiries and requests to produce more clones with the hopes of filling in the gapped regions. The Moerman lab has also received requests for the production of a Hawaiian strain library. With the understanding that repetitive sequence may lead to misalignment and unclonable regions within bacterial cells, we thought a new library might be warranted if the genome to be sequenced differed significantly from the N2 DNA used to produce the first fosmid set. The Hawaiian strain is divergent enough to possibly alter repetitive sequences within the genome thereby allowing some alignment not previously attainable. Even if this is not the case, the Hawaiian divergence (Maydan *et al.* 2010) is significant enough that the library could be a useful tool for exploring the architectural reasons for this genome difference. It could also provide a resource for individuals to further explore the genomic cause and effect of balancing selection, and its evolutionary role, as described previously. As well, the newly re-sequenced

Hawaiian DNA provides the technical ability to explore the CB4856 genome. It will possibly allow better resolution of cloned sequences that align unconventionally to N2. This option is not available for other isolates at this time.

The focus of my thesis is the production and analysis of the Hawaiian geographic isolate of *C. elegans*. The work was undertaken in the hopes that we could fill in some of the gaps found in the N2 library and possibly provide the community with a resource to explore genomic architecture and its evolutionary cause. We decided that a five-fold coverage, the same as produced for the N2 library, would create the most likely chance of complementation for a reasonable cost. The resulting WRMHS library was end sequenced and aligned to the 210th release of the Wormbase *C. elegans* genome WS210. The N2 fosmids were realigned to this release to provide a comparison for the CB4856 library and to determine the level of library overlap occurring between the sets of fosmids.

2 Methods

2.1 Fosmid Library Production

2.1.1 Strains Used

Natural isolates sampled from Bristol, England (VC196, an N2 subculture received from the Caenorhabditis stock center) and Hawaii, U.S.A (CB4856, an HA-8 subculture isolated from a pineapple field in 1972 (Caenorhabditis genetics center website 2007) were used as the source of genetic material for this study.

2.1.2 Growth

Single animals were placed on a 60mm plate of Nematode Growth Media (NGM) growing a lawn of *Escherichia coli* OP50 and allowed to grow until the F1 generation was laden with eggs. These were washed from the plate, with M9 buffer (22mM KH_2PO_4, 43mM Na_2HPO_4, 86mM NaCl, 1mM $MgSO_4$) containing 1% TritonX-100, and pelleted in a 15ml polypropylene tube. The pellet was treated with 5ml egg-prep buffer (20% household bleach, 5% 10N KOH, 75% ddH_2O) followed by vigorous

shaking. The eggs were rinsed with M9 buffer three times when they were released from the carcasses. The washed egg pellet was re-suspended in 10 ml M9 buffer and distributed on to 40 plates (150mm), containing 8P media on which a lawn of χ1666 *E. coli* was spread. These were allowed to grow until the F1 generation became gravid. The plates were washed into four 50 ml conical tubes with M9 buffer containing 1% TritonX-100. These worms were rinsed with M9 buffer a minimum of three times. The number of rinses varied with bacterial content and turbidity of supernatant. The pellets were then exposed to 25 ml of ice-cold egg-prep buffer and shaken vigorously until no visible worm carcasses were present. The egg-prep buffer was refreshed if the reaction went for more than eight minutes. The eggs were then washed 6 times in M9 buffer and left in 10 ml of M9 overnight, on a rotating platform, to hatch. The L1 worms were then washed 3 times with M9 buffer followed by one time in TE (10mM TrisHCl pH 8.0, 1mM EDTA) buffer. The packed pellet was then frozen (-80°C).

2.1.3 DNA Preparations

DNA was prepared in two ways and subsequently mixed equally for all downstream applications.

2.1.3.1 Phenol Purification

Frozen pellets were placed on ice. Proteinase K lysis buffer (8.8 mM TrisHCl pH 8.3, 44 mM KCl, 22 mM $MgCl_2$, 0.4% Tween-20, 0.4% IGEPAL, 300ug/ml Proteinase K all in ddH_2O) was added to the pellets in a volume equal to the pellet. These were incubated at 60-65°C for 1-2 hr, until there were no visible worm carcasses in the lysate. These reactions were inverted every 15 minutes during incubation. The lysate was immediately extracted with Phenol: $CHCl_3$: isoamyl alcohol (25:25:1). Equal volumes of lysate and phenol solution were inverted gently 10 times and placed in a centrifuge at 13,000 rpm for 5 minutes to separate the phases. The aqueous phase was removed and re-extracted. This was followed by one $CHCl_3$: isoamyl alcohol (25:1) extraction of equal volume done similarly to the previous description. The aqueous portion was again removed and to it was added 0.1 volumes of NaOAc followed by 2 volumes of 100% EtOH. These reactions were placed at -20°C for a minimum of 30 minutes. The DNA was precipitated from solution by centrifugation at 13,000 rpm for a minimum of 30 minutes at 4°C. The pelleted DNA was washed with 70% EtOH. The supernatant was removed and the precipitate was dried for 5 minutes at room temperature open to the air. The DNA was then resuspended in 500µl of TE buffer (10mM tris-HCl pH8.0, 1mM EDTA) containing 1 mg of RNase A and incubated at 37°C for one hour, inverting the tubes 10 times every 15 minutes. These reactions were re-extracted and precipitated with the procedure listed above. The DNA concentration was checked and the final volume made up to maintain a concentration of 500ng/µl.

2.1.3.2 Purgene Kit

Pellets not treated to phenol purifications were treated with the Puregene tissue extraction kit from Gentra according to the manufacturer's instructions. The final DNA was diluted to a concentration of 500ng/μl in TE buffer.

2.1.3.3 Further Purification

The combined DNA was further purified in two ways. The first aliquot was treated to isopycnic centrifugation in a CsCl gradient. This was followed by butanol extraction to remove DNA bound dye and buffer exchange (using amicon ultra centrifugal filter with a 10kD cutoff and microcon YM-30 centrifugal concentrators) to remove CsCl and concentrate the DNA solution. Another aliquot of DNA was purified further using synchronous coefficient of drag alteration (SCODA), a method employing differential electrical fields to separate DNA from other molecules in a matrix, (Pel *et al.* 2009) with electrophoretic washing.

2.1.4 Fosmid Preparation

Fosmids were produced using the Epicenter CopyControl Fosmid Library Production Kit. DNA was mechanically sheared to 25-50 kb using a 50 µl Hamilton syringe. The DNA was end repaired, with Epicenters End-It enzyme, to ensure blunt ends. The DNA was sized using pulsed field agarose gel electrophoresis (PFGE). The CsCl purified DNA was sized on 1X HAE (10mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 10 mM NaOAc, 0.5mM EDTA, made up and pH to 7.4 with NaOH) buffer (Kinscherf *et al.* 2009). The SCODAphoresis purified sample was sized in 0.5 X TBE (45 mM Tris base, 45 mM Boric acid, 2 mM EDTA pH 8.0). In both cases, DNA with a minimum size of 25 kb and a maximum of 42 kb (as compared to DNA markers) was cut from the agarose gel. The agarose plug was digested with agarase and the DNA was buffer exchanged using TE and concentrated using the Amicon and Microcon centrifugal filters described previously. These sized fragments were ligated to the pCC2Fos vector and packaged into a λ phage packaging extract, with a titre of 150000 clones/ml, which was used to infect EPI300-T1^{R} *E. coli*. Individual clones were plated on agarose and 15,360 colonies were picked and plated in a 384-well format and sequenced from both ends.

2.2 Fosmid Sequencing

Fosmids were grown in culture over night by the Canada's Michael Smith Genome Sciences Center. These were then spun down and DNA was purified from bacteria using Qiagen spin columns in 96-well formats. The DNA was sequenced using the ABI 3100 and ABI 3730 capillary sequencers with the M13 forward and reverse primers. The sequence was trimmed for vector DNA.

2.3 Fosmid Mapping

The end sequences, of the CB4856 strain made for the work done in this thesis and N2 strain made in 2005, were BLAST aligned to the N2 reference genome version WS210 (Wormbase WS210 2010). The best blast matches were selected, based on bitscore and query length, for each input. These individual ends were paired to their respective alternate side. Single sides, with no pair, were discarded. All of the clones discarded from further analysis beyond this point are believed to have unconventional alignments. The orientations of the inputs were determined. Those fosmids whose inserts had two ends aligned to different chromosomes, or found on the same strand were discarded. The inserts aligning with ends facing opposite directions away from one another were also discarded. The clones were then organized by size and those smaller than 15 kb and larger than 55 kb were discarded, as packaging by the λ extract is unlikely outside of this range due to size restrictions imposed by the capsid head.

The remaining clones were ordered and analyzed for completeness of genomic sequence. Both libraries were treated similarly to simplify comparison.

Gaps were determined by looking at the areas of non-contiguous clones. Both libraries were tiled together to determine the holes in the overlapping coverage for the combined fosmid sets. The total sequence not contained within the fosmid library was calculated from the combined non-contiguous regions within each library.

2.4 Protein and Yeast Artificial Chromosome (YAC) Coverage of Libraries

A list of all genes coding for proteins was pulled from Wormbase in the genomic data freeze WS210; a permanently frozen database available from Wormbase. The gaps in contiguous sequence were compared to the coordinates of each gene within the chromosome in question. Genes falling within the gaps were considered to be missing from the library. Total coverage was the difference between the missing genes and the total coding regions in the genome. Similarly, the Yeast Artificial Chromosome (YAC) coverage was determined by selecting those gene designations originally derived from YAC sources.

A χ^2 distribution was calculated for the two libraries protein coding gene coverage. For it a two by two contingency table was used to separate the Yeast and cosmid

derived genes contained in the WRMHS and WRM06 libraries independently. The c^2 value was compared to a 99.999% confidence interval with 1 degree of freedom.

2.5 Repetitive Element Gap Co-occurrence

The repetitive element gap co-occurrence was carried out in two ways. Determining the Yeast derived clone sequence co-occurrence with repetitive elements required static addresses to allow for the variable overlap seen with cosmids and other near by YACs. The protein coding sequences were used as the static addresses. To increase the likelihood a fair representation of the surrounding sequence was taken, 40 kb upstream and downstream from each of the protein-coding sequences was used to compare to the lists of repetitive elements. This was chosen to place the outer limit at the size of a fosmid insert away from the protein-coding region. The YAC coding regions were compared to the non-YAC derived protein coding regions for repetitive element distribution. For the gapped regions of the libraries, the gap endpoints were used as static positions to compare to the list of repetitive elements. These positions were compared to a theoretical similar sized region defined as if the elements were equally distributed throughout the genome.

2.6 Insertion and Deletion Determinations

The known insertions and deletions held in the VC196 laboratory N2 (Maydan *et al.* 2007) strain and the CB4856 Hawaiian geographic isolate (Maydan *et al.* 2010) were compared to the list of WRMHS library held inserts. The clones in the WRMHS library containing the known insertions and deletions were quantified and those complementing the deletion in our N2 strain are presented here and described in further detail.

2.7 Misalignment Analysis

After fosmids were mapped to the genome and separated if inserts were unconventional, those clones having end sequences aligning to two different chromosomes were looked at more closely. A PERL program designed by me, and written by Jeff Magnusson, was used to explore the alignments. The multiple blast outputs were categorized for the subgroup discarded due to alignment of paired ends on separate chromosomes. Those pairs of end alignments, from each clone, determined to be on the same chromosome were output in permutations allowing for conventional insertion only (i.e. those which are facing in opposite directions towards one another). The sets of pairs, which fell into this class, were calculated based on subject length and only those falling between 10 kb and 60 kb were outputted. The output was

quantified and separated by chromosome and by presence or absence of pairs aligning unconventionally (described previously in chapter 1.4).

3 Results

3.1 Construction and Description of a Fosmid Library for the Hawaiian Strain CB4856

I constructed the CB4856 library as was described in chapter 2. It was labeled WRMHS. The WRM designation ties it to the original N2 library made with the HS classification due to its use of DNA from the Hawaiian geographic isolate. Sequencing of 15,360 clones, performed at the Michael Smith Genome Sciences centre, produced 28,630 end sequences with 2,090 ends missing due to low quality or missing sequence. I Blast aligned the supplied sequences to the WS210 freeze. I found that 1,274 of the sequences did not have a matching pair. Other paired end sequences not used included 237 clones with ends facing the same direction, 479 with ends paired in opposite directions (facing away from each other) and 436 with ends aligned with two different chromosomes. 935 clones were calculated to have insert sizes larger than 55 kb or smaller than 15 kb, based on alignment. This left 11,358 clones for further analysis (Table 1). These clones comprising the CB4856 library covered 77.5 % of chromosome I in 1,498 clones, 88% of chromosome II in 1,879 clones, 78% of chromosome III in 1,445 clones, 83% of chromosome IV in 1755 clones, 87% of

chromosome V in 2459 clones and 96% of the X chromosome in 2322 clones. In total, 85% of the genome is covered by the library.

The library coverage can be seen in further detail in Table 1. The table shows a striking difference between the CB4856 and the N2 library. The WRMHS library shows fewer clones with a greater mean size. The theoretical and actual coverage are both greater for almost all chromosomes in WRMHS. Coverage was calculated as the quotient of the total sequence in each library with fosmid coverage for the actual value. Theoretical coverage was calculated as the quotient of the total sequence in each library and the total sequence in each chromosome. The exception to the theoretical and actual coverage being better in WRMHS is Chromosome II in which the percentage of sequence covered is still higher in the Hawaiian library. Only Chromosome X shows better sequence coverage as well as theoretical and actual coverage in the N2 library. Sequence coverage refers to the sum of genome regions covered by the library. The standard deviation of means calculated for the libraries show that the distribution of fosmid sizes is greater in the WRMHS library for almost all chromosomes and suggests a cleaner size selection during production.

Table 1: Fosmid library coverage broken up by chromosome for WRM06 and WRMHS separately and combined.

WRM06	Breakdown (#Clones)	Mean Size (bp)	Standard Deviation (bp)	Total Sequence in Libray (bp)	Theoretical *Coverage (X)	±Actual Coverage (X)	Sequence Coverage (bp)	% of bp Covered	Chromosome Size (bp)
I	1,417	33,378	3,994	47,295,973	3.14	4.37	10,829,403	58.16	15,072,421
II	1,996	33,923	3,711	67,710,646	4.43	5.46	12,409,885	81.22	15,279,324
III	1,351	33,640	3,820	45,413,564	3.29	4.74	9,587,372	69.56	13,783,685
IV	1,687	33,676	3,837	56,845,804	3.25	4.21	13,493,564	77.13	17,493,793
V	2,708	33,714	3,750	91,298,681	4.36	5.16	17,705,635	84.62	20,924,143
X	3,330	34,035	3,753	113,335,132	6.40	6.57	17,253,419	97.32	17,718,854
Total	12,489	33,782	3,797	421,899,800	4.21	5.19	81,279,278	78.92	100,272,220
WRMHS	Breakdown (#Clones)	Mean Size (bp)	Standard Deviation (bp)	Total Sequence in Libray (bp)	Theoretical *Coverage (X)	±Actual Coverage (X)	Sequence Coverage (bp)	% of bp Covered	Chromosome Size (bp)
I	1,498	34,162	4,579	51,141,145	3.39	4.38	11,677,217	77.47	15,072,421
II	1,879	34,691	4,823	65,218,492	4.27	4.86	13,407,032	87.75	15,279,324
III	1,445	34,674	4,265	50,138,428	3.64	4.64	10,812,288	78.44	13,783,685
IV	1,755	34,178	4,666	60,051,468	3.43	4.14	14,521,499	83.01	17,493,793
V	2,459	34,545	4,872	84,946,569	4.06	4.69	18,099,563	86.50	20,924,143
X	2,322	34,614	4,534	80,407,857	4.54	4.72	17,032,937	96.13	17,718,854
Total	11,358	34,495	4,653	391,789,540	3.91	4.58	85,550,536	85.32	100,272,220
Combined	Breakdown (#Clones)	Mean Size (bp)	Standard Deviation (bp)	Total Sequence in Libray (bp)	Theoretical *Coverage (X)	±Actual Coverage (X)	Sequence Coverage (bp)	% of bp Covered	Chromosome Size (bp)
I	2,915	33,778	4,323	98,462,716	6.53	7.85	12,542,308	83.21	15,072,421
II	3,875	34,298	4,301	132,903,540	8.70	9.39	14,149,826	92.61	15,279,324
III	2,796	34,176	4,083	95,556,673	6.93	8.40	11,379,949	82.56	13,783,685
IV	3,442	33,937	4,283	116,810,335	6.68	7.47	15,636,289	89.38	17,493,793
V	5,167	34,110	4,340	176,245,250	8.42	9.06	19,448,251	92.95	20,924,143
X	5,652	34,273	4,102	193,710,826	10.93	11.06	17,519,113	98.87	17,718,854
Total	23,847	34,121	4,241	813,689,340	8.11	8.97	90,675,736	90.43	100,272,220

*Theoretical coverage of the library relating to the number of times any one sequence would be covered by a fosmid in the genome

±actual coverage of the library relating to the number of times the sequence held in fosmids would be covered.

3.2 A Description of the N2 Library From WS210 Alignment

Previously, I made a Bristol N2 fosmid library. It was produced in a similar fashion to the Hawaiian library. The major difference, in packaging, between the libraries is the vector backbone. In the N2 library, the 8.1 kb pCC1Fos vector, from Epicenter, was used. This was completed four years ago. It was labeled WRM06 and the initial alignment of the library was made to the WS140 genomic data freeze (Wormbase WS140 2005); a permanently frozen database available from Wormbase. For this study, the WRM06 fosmids were realigned to the WS210 data freeze (Wormbase WS210 2010) enabling comparison between the two geographic isolate libraries.

The alignment of the Bristol N2 library to the WS210 genome showed a similar pattern of fosmids to the WS140 alignment. From the initial 15,744 fosmids, 653 clones did not align or provided imperfect sequence. 501 clones had only one arm align, while 548 clones were aligned displaying arms on two different chromosomes. 140 clones had arms aligned in the same direction and 294 clones aligned with arms facing opposite directions, away from one another. These clones were separated for ease of analysis. The remaining 13,568 clones, all in proper orientation according to the reference genome, were used to construct a revised fosmid map for the N2 isolate. The inserts' sizes were calculated according to their paired end alignment. Those

below 15 kb and above 55 kb were removed leaving 12,489 fosmids to tile onto the genome.

The N2 clones were mapped to the chromosomes as follows. Chromosome I has 1,417 total clones covering 58 % of the chromosome sequence. Chromosome II has 1,996 fosmids containing 81% of sequence, while chromosome III has 1,351 clones accounting for 70% of the chromosome. Chromosome IV has 1,687 fosmids aligning to 77% of the chromosome and chromosome V has 2,708 clones containing 85% of the chromosome sequence. The X chromosome has 3,330 fosmids accounting for 97% of sequence.

3.3 Gaps in Fosmid Library Sequence

The WRM06 and WRMHS libraries were sequenced to a depth calculated to provide five fold coverage of the genome, based on the size of fosmid inserts. As can be seen in Table 2, complete coverage for neither genome was obtained. This is also evident in the difference between the actual and theoretical coverage of the library. WRM06 displays lower than expected coverage for both calculations for all chromosomes except for X. For this chromosome, there was a slight discrepancy between its actual and theoretical value (6.57x and 6.40x, respectively) and areas of non-coverage within the chromosome were evident. The coverage difference displayed between X and all other chromosomes points to the unequal allocation of fosmids within this library

compared to that produced from the Hawaiian strain. The following sections describe the gaps observed in the coverage of fosmid contiguous sequence.

Table 2: Gaps displayed by the WRMHS and WRM06 libraries separately and combined, separated by chromosome.

WRM06	Number of gaps	Total gap size (bp)	Mean Gap (bp)	Standard Deviation	% of gaps	% of genome	Chromosome size (bp)
I	74	4243018	57338	63441	18.09	28.15	15072421
II	72	2869439	39853	49298	17.60	18.78	15279324
III	59	4196313	71124	81446	14.43	30.44	13783685
IV	102	4000229	39218	52364	24.94	22.87	17493793
V	77	3218508	41799	57937	18.83	15.38	20924143
X	25	465435	18617	33613	6.11	2.63	17718854
Total	409	18992942	46438	60319		18.94	100272220
WRMHS	Number of gaps	Total gap size (bp)	Mean Gap (bp)	Standard Deviation	% of gaps	% of genome	Chromosome size (bp)
I	75	3395204	45269	54102	16.52	22.53	15072421
II	74	1872292	25301	33529	16.30	12.25	15279324
III	65	2971397	45714	48036	14.32	21.56	13783685
IV	97	2972285	30642	33734	21.37	16.99	17493793
V	95	2824580	29732	39785	20.93	13.50	20924143
X	48	685917	14290	23543	10.57	3.87	17718854
Total	454	14721675	32427	41357		14.68	100272220
Combined	Number of gaps	Total gap size (bp)	Mean Gap (bp)	Standard Deviation	% of gaps	% of genome	Chromosome size (bp)
I	61	2530113	41477	40983	20.75	16.79	15072421
II	51	1129498	22147	20751	17.35	7.39	15279324
III	54	2403736	44514	42368	18.37	17.44	13783685
IV	65	1857504	28577	30552	22.11	10.62	17493793
V	55	1475892	26834	29044	18.71	7.05	20924143
X	8	199741	24968	51658	2.72	1.13	17718854
Total	294	9596484	32641	35189		9.57	100272220

3.3.1 Gaps in the WRMHS Library from the WS210 Alignment

The 11,358 clones of WRMHS library were tiled on the WS210 genomic data freeze. The total breakdown is detailed in Table 2. From the tiling, 75 gaps were found in the contiguous sequence on chromosome I amounting to 3,395,204 bp missing. Chromosome II was missing 1,872,292 bp in 74 gaps. III and IV had 2,971,397 bp in 65 gaps and 2,972,285 bp in 97 gaps, respectively. V had 2,824,580 bp unaccounted for in 95 gaps and X was missing 685,917 bp in 45 gaps. The total sequence covered by this library is 85.3% of the genome with 85,550,545 bp included (displayed in appendix A).

3.3.2 Gaps in WRM06 Library from Both Alignments

The initial alignment of the library to the WS140 genome showed 73 gaps, in contiguous sequence, on Chromosome I accounting for 4,227,363 missing base pairs. On Chromosome II 68 gaps in the library removed 2,850,021 bp. Chromosome III coverage included 57 gaps and 4,133,578 bp missing. Chromosome IV fosmids were missing 100 gaps including 3,887,566 bp of sequence. Chromosomes V and X showed 76 gaps and 3,200,103 bp and 22 gaps including 479,779 bp missing from the library, respectively.

The 12,489 fosmids, which were conventionally aligned to the WS210 genome, and were within the proper size range, showed a similar pattern of contiguous sequence to the WS140 alignment. The WRM06 fosmids left 409 gaps in contiguous sequence, when tiled on the WS210 genome. The gapped sequence can be compared to WRMHS in Table 2. The coverage of the library was 81.1% of the genome with 81,279,278 bp included. The gaps seen in the library coverage are split between the chromosomes with 75 gaps of mean size 84,084 bp found on I, 72 with mean size 39,853 bp on II, 59 with mean size 71,124 bp on III, 102 with mean size 39 218 bp on IV, 77 with mean size 41,799 bp on V and 18 with mean size 18,617 bp on X (displayed in appendix A).

3.3.3 Gaps Determined after Combining the WRMHS and WRM06 Fosmid Libraries

By combining the libraries, complementing areas within gapped sequence were made visible. Even though the libraries had similar numbers of gaps, they were not placed in identical locations. After combining the libraries, the number of total gaps in contiguous coverage dropped to 294 from 409 seen in the N2 library and 454 detected in the CB4856 library. These gaps are spread across all the chromosomes with 2,530,113 bp missing in 65 gaps on chromosome I, 1,129,498 bp missing in 51 gaps on chromosome II, 57 gaps missing in 4 133 578 bp chromosome III 2,403,736 bp missing bases with 54 gaps on chromosomes IV, and 1,857,504 bp in 55 gaps on

chromosome V. Of interest is the fact the X chromosome is only missing 199,741 bp spread over eight gaps.

Table 2 shows breaks in contiguous sequence found in the N2 and CB4856 libraries, both separately and amalgamated. Production of the WRMHS library has clearly improved the genomic coverage. This can be seen when the alignments for the two libraries are compared side by side, separately and combined. The combined library shows fewer gaps with less sequence missing for each chromosome than either WRMHS, or WRM06 separately. The layout of gaps over each chromosome can be seen in Figure 1. The independent libraries can be viewed for comparison in appendix a. The gaps found in the combined libraries are clustered in the arms of each chromosome, with larger holes found at the ends. Almost none are seen on the X chromosome. The standard deviation and mean size of gaps are also smaller for nearly every chromosome. This is seen in Figure 2. The combined libraries show a small decrease in mean gap size from the WRMHS library as a whole, even though the only chromosome displaying a decrease in mean is X. The libraries do not completely overlap, however, and the combined libraries still leave 9.6% of the genome with no fosmid coverage.

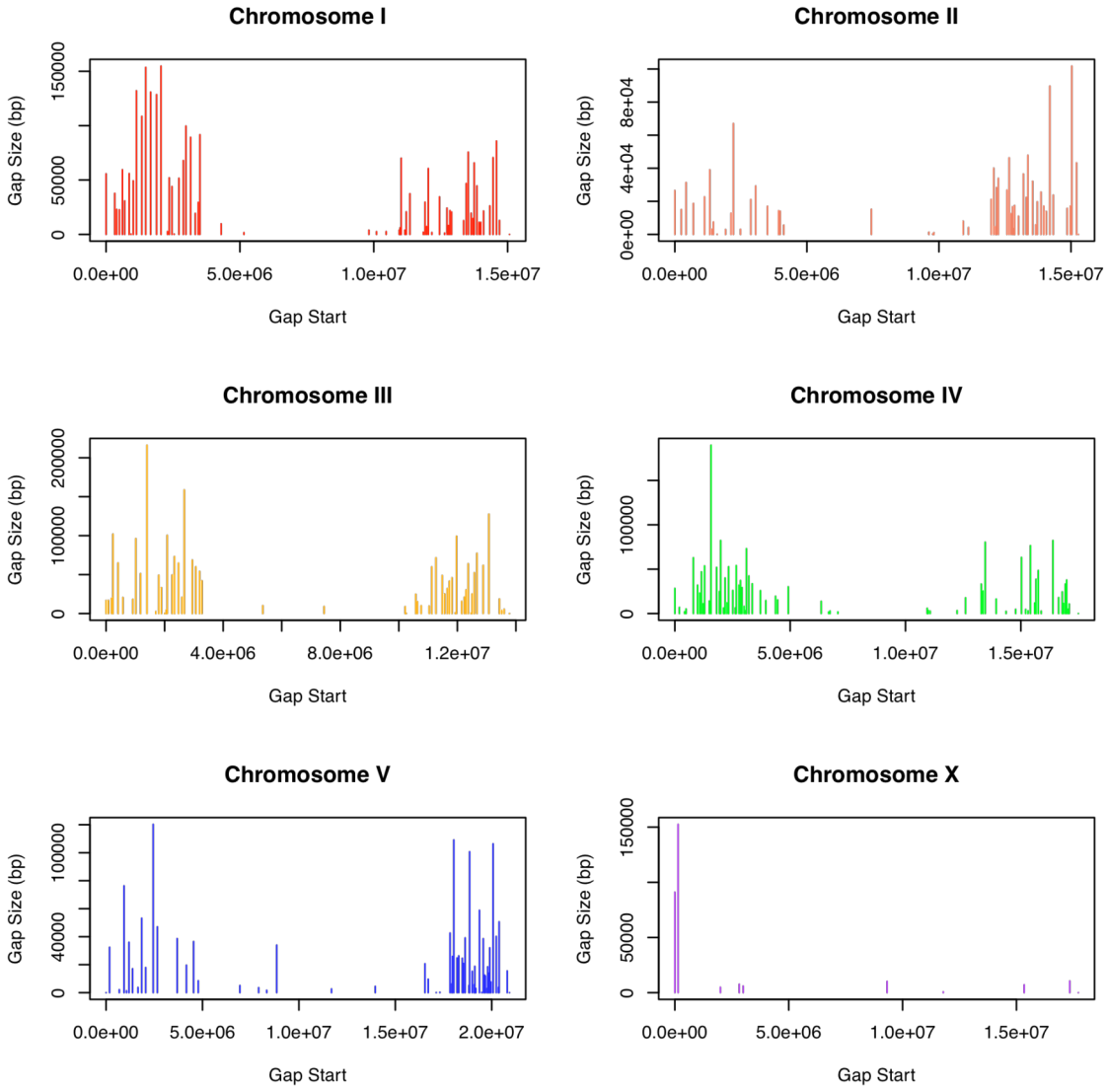

Figure 1: Graphic depiction of the gap distribution across the chromosomes for the combined coverage of the WRM06 and the WRMHS libraries. Top left in red shows chromosome I; top right in coral is chromosome II; middle left in yellow is chromosome II; middle right in green is chromosome IV; bottom left is chromosome V; bottom right in purple is chromosome X. X axis displays position of the gap on the chromosome. Y axis shows size of each gap.

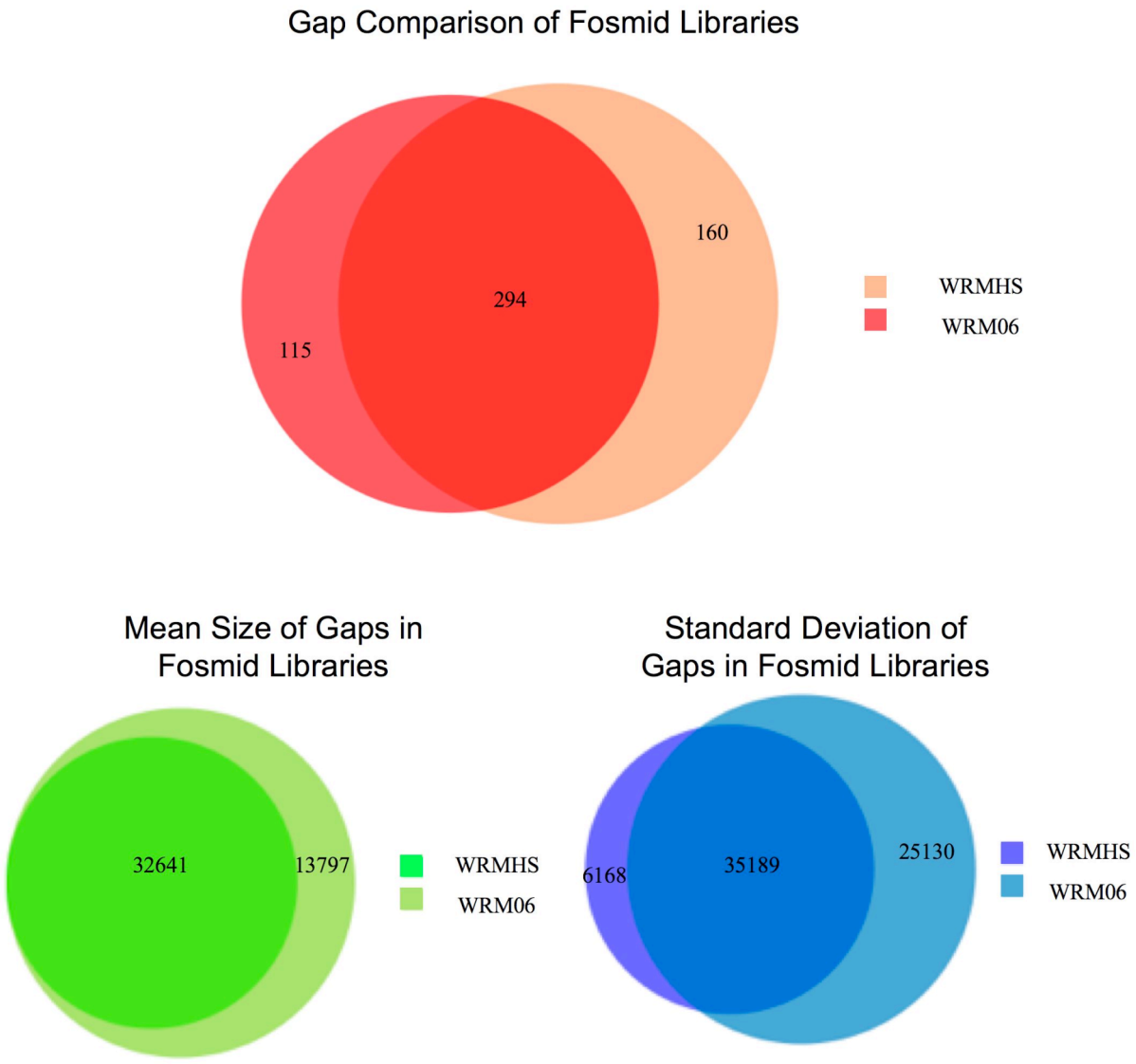

Figure 2: Venn diagrams depicting overlap of the WRMHS and WRM06 libraries. The mean size (green) and standard deviation (blue) of gaps in the combined libraries is also presented.

Table 2 shows the trends in the contiguous sequence gaps, for each library. It also shows the WRMHS library is generally better on all measured levels described. The one exception is the number of gaps. The Hawaiian library shows more gaps, however, the mean size of the gaps is smaller for each chromosome and the total gapped area is smaller for every chromosome except for X. The gaps on X also account for the biggest discrepancy in the number of gaps. The distribution of the gaps across the chromosomes is more even and the standard deviation of gap sizes is smaller for the WRMHS library.

The number of gaps seen in both of the libraries may be somewhat misleading. Closer inspection of the unconventional clones show that there may be inserts which are associated with some of the gapped regions. These are most readily seen in the fosmids determined to be too large for encapsulation by the λ phage packaging extract (Feiss *et al.* 1977). There are several instances with multiple clones (Tuzun *et al.* 2005) having ends align on either side of a gap in contiguous sequence. The fosmids, which are too large, are likely not the only subcategory of unconventional inserts to affect the gaps layout.

3.4 Protein Coding Genes Covered by the Fosmid Libraries

Based on the WS210 data freeze the WRMHS library showed fewer missing genes than did the WRM06 library and was consistent for all chromosomes except for X and V. The best coverage of coding regions was on the X chromosomes with 96.9 % and 95.0% coverage for the N2 and CB4856 libraries, respectively. The other chromosomes did not show as complete coverage with the poorest examples being III with 76.0% coverage for the N2 library and I with 82.0% coverage for CB4856. The overall coding sequence coverage contains 85.0% of all genes for N2 and 86.8% for CB4856 when the libraries are not combined. Table 3 displays the quantity of protein coding sequence missing from the libraries separately and combined.

Table 3: Protein coding sequences falling in gapped regions of the libraries broken up by chromosome

	Total Genes	WRM06	% of Total	WRMHS	% of Total	Combined	% of Total
I	3154	707	22.4%	587	18.6%	389	12.3%
II	3805	580	15.2%	454	11.9%	213	5.6%
III	2914	712	24.4%	494	17.0%	373	12.8%
IV	3735	695	18.6%	571	15.3%	300	8.0%
V	5800	736	12.7%	723	12.5%	312	5.4%
X	3213	99	3.1%	165	5.1%	49	1.5%
Total	22621	3529	15.6%	2986	13.2%	1636	7.2%

The combined coverage of the libraries overlaps in coding region gaps to cover half the number of genes missing from either library individually. The result is that of 3,392 and 2,986 genes missing from the N2 and CB4856 libraries, respectively, only 1,636 of the 22,621 genes present in the WS210 data freeze are still missing (7.2%), when the libraries are combined. Table 3 shows that the overlap is significant on each chromosome, but not complete.

Producing a CB4856 library had additional benefits. The overlap of coverage in previously tiled areas is more useful than initially expected. Genes that are not fully contained within a single fosmid or cosmid in the two N2 libraries are represented within the Hawaiian Library. One example of this is the *unc-119* gene. When originally sequenced this gene was not contained within a single contig, but split between two partially complementary cosmids. This gene is also not present in the WRM06 library and resides in a single gap on chromosome I. However, the WRMHS library has two fosmids containing the region of interest with at least 5 kb of upstream sequence. This is likely a result of the increase in depth of coverage and may therefore be observed for other genes.

3.5 Alignment of Libraries to YAC Sequence in Genome

To see if the fosmid libraries' cloning patterns are consistent with the original cosmid library, I compared their gaps in contiguous sequence to those regions sequenced only with yeast clones. The YACs were used to complete regions believed to be unclonable in bacteria during the sequencing project. I accomplished this by comparing the sequence unaccounted for in the fosmid libraries with the cosmid designation originally given during sequencing of the genome to determine vector origin. Those cosmids originally designated with the first letter Y were derived from YACs.

Table 4 shows the proportion of coding regions originally sequenced in YACs to the total coding regions falling inside gaps for each library independently and combined. It shows there is a marked increase in the proportion of the YAC-derived coding sequences lying within gaps in the libraries compared to total coding genes. The gaps in fosmid coverage show 62% and 58% YAC-derived genes for the WRM06 and WRMHS libraries, respectively. This is increased even further to 81 % for the gaps remaining after the two libraries have been combined. The co-occurrence of gaps in the fosmid libraries with YAC derived genes, representing sequence unclonable in the original cosmid library, provides evidence of the preferential exclusion of these regions from the libraries.

Table 4: Number of protein coding sequences, from the WS210 annotation, determined to be originally derived from Yeast cloning vectors and the total protein coding sequences not covered, by chromosome

	WRM06 Yeast derived	WRM06 Total protein coding	% of genes from yeast derived sequence	WRMHS Yeast derived	WRMHS Total protein coding	% of genes from yeast derived sequence	Combined Yeast derived	Combined total protein coding	% of genes from yeast derived sequence	Yeast derived vectors	Total vectors protein coding	% of genes from yeast derived sequence
I	453	707	64.1%	415	587	70.7%	343	389	88.2%	695	3,154	22.0%
II	307	580	52.9%	181	454	39.9%	160	213	75.1%	522	3,805	13.7%
III	502	712	70.5%	371	494	75.1%	332	373	89.0%	691	2,914	23.7%
IV	492	695	70.8%	411	571	72.0%	265	300	88.3%	1039	3,735	27.82%
V	389	736	52.9%	322	723	44.5%	206	312	66.0%	861	5,800	14.8%
X	47	99	47%	48	165	29.1%	22	49	45%	158	3,213	4.92%
Total	2,190	3,529	62.06%	1,748	2,994	58.4%	1,328	1,636	81.17%	3,966	22,621	17.53%

The alignment of the libraries' gapped regions with sequence cloned originally only in Yeast could suggest a mechanism inhibiting these regions from being packaged or propagated in bacteria. As repetitive sequence may cause both misalignment and challenges for packaging, I wanted to check whether there was a co-occurrence between these elements and the original cosmid library's unclonable sequence. Looking at the regions only captured in YACs and comparing them to gaps found in the recent fosmid libraries might provide insight into the reasons these regions could not be cloned or propagated.

By comparing the percentage of repetitive elements falling in Yeast derived clones to the percentage of protein coding sequences in YACs, I came up with a proportion value representative of quantity of elements held in yeast derived sequence as a proportion of the percentage of protein coding sequences. By definition this assumes every gene is associated with the same number of repetitive elements and any deviation for a subgroup of elements will show up as a factor of 1. Table 5 shows that the YAC derived sequence contains between 1.44x and 1.94x (chromosomes IV and II, respectively) the number of repeat elements per unit length that would be expected if regions were evenly distributed. For all Yeast derived sequences in the genome there is 1.85x the repetitive elements, which would be expected for the same number of coding elements contained anywhere in the genome. These values do not reflect the complete lack of regions, around cosmid-cloned genes, with an increased proportion of repetitive elements. However, they do show the regions are more frequent in YAC

derived sequence, with 79.5% of regions around genes displaying an increased amount of repetitive elements and only 29.8% of cosmid derived sequence showing a similar pattern.

Table 5: Proportion of repetitive elements that would be expected for a similar sized area, if elements were distributed evenly over the genome

	YAC (x)	WRMHS (x)	WRM06 (x)	Combined (x)
I	1.81	1.91	1.84	2.23
II	1.94	1.80	2.02	2.35
III	1.86	1.88	1.77	2.02
IV	1.44	1.75	1.82	2.16
V	1.82	1.86	2.16	2.55
X	1.71	2.13	2.60	2.91
Total	1.85	2.04	2.11	2.55

The increased density of repetitive elements in these stretches believed to be unclonable, brought forth the question of whether the gapped regions of the fosmid libraries contained a similar concentration of repeat sequence proportional to that seen in YAC derived regions for the original cosmid library. To explore the abundance of repetitive elements over the gapped regions, the quotient of the proportion of elements falling into the gaps and the percentage of genome not contained within fosmids was calculated. By definition this will assume every region is associated with the same number of repetitive elements and any deviation for a subgroup of regions' elements will present as a factor of 1. Table 5 shows the side-by-side comparison of the Yeast derived sequence to both the WRM06 library and WRMHS library as well as the combined libraries. A strong trend towards increasing quantities of repetitive elements was seen from the YAC derived sequences to the WRMHS (2.04x) and the WRM06 (2.11x) libraries. The gaps seen in the libraries show less repetitive elements for the WRMHS library as a whole compared to the WRM06. This is true for most chromosomes except for I and III. The gaps remaining after the fosmid libraries are combined showed 2.55x the repetitive elements expected for the same length of DNA.

To test whether the distribution of genes produced in the Hawaiian library would be likely if the Bristol fosmids were sequenced to a depth equal to the combined libraries, a Pearson chi squared test was performed. It was designed with a null hypothesis that there is no relationship between the two libraries' differences and the distribution of

YAC or cosmid vector derived protein-coding genes displayed. The critical value was determined to be 33.97, which is above a value necessary to achieve a confidence interval of 99.999%. The null hypothesis can therefore be rejected providing support for the alternate hypothesis that states the ratio of YAC to cosmid derived genes is dependant on the two fosmid libraries' differences.

3.6 Known Insertion and Deletion Coverage of the Libraries

As stated previously, the Hawaiian isolate CB4856 is known to contain DNA sequence insertions and deletions (indels) relative to the N2 strain. To determine if clones were produced covering these indels, I compared the list of clones from the CB4856 library WRMHS to the list of indels described (Maydan *et al.* 2010). Table 6 shows the breakdown of the search. 429 fosmids span the regions in which 289 genes associated with 143 of the 181 CB4859 indels are found. In the 38 remaining indels, 334 genes are contained. The indels from the entire list show a mean size of 7,343 bp. The 38 indels from the list of non-covered genes have a mean size of 71,568 bp. The large size of the indels may have contributed to the underrepresentation of these particular sequences in the fosmid library, especially due to the majority being deleted sequence and larger clones being dismissed (see Discussion).

To explore the possibility of clones containing the larger deletions being overlooked due to size, the clones aligning to regions too large for packaging were investigated. 10 of the 38 indels have 34 individual fosmids showing inserts calculated to be larger than the 55kb cutoff and flanking the individual indels' breakpoints. A prime example of two fosmids aligning with indels as well as gaps can be seen in appendix B labeled in red (page 104). The end sequences of these clones cover a region with a 4,500 bp deletion of coding sequence, as well as a gap in contiguous sequence. All of the indels in the CB4856 genome, with the exception of eight, are associated with fosmid ends aligning in unconventional ways. The unconventional inserts may have ends found on different chromosomes, or ends found on the same strand and pointing in the same direction. Other unconventional clones have ends mapping to opposite strands and pointing away from one another or were those found to be out of the selected size range.

Table 6: Insertions and deletions seen in the CB4856 Hawaiian strain, compared to canonical N2 sequence, by chromosome

	# genes in indels	#genes in indels uncovered	#genes in indels covered	Total indels in CB4856 (#)	#indels covered by fosmids	#fosmids covering indels
I	28	9	19	20	14	39
II	217	95	122	58	51	161
III	21	1	20	15	12	26
IV	70	56	14	16	10	25
V	279	168	111	68	54	180
X	8	5	3	4	2	8
Total	623	334	289	181	143	439

The lab strain VC196 used to produce the Bristol isolate fosmid library is also known to carry a deletion in a single gene (Maydan *et al.* 2007). This is a 1,788-bp deletion affecting exons 5 and 6 of *alh-2*. This gene is found on chromosome V in the 3,352 kb area from position 1,644,378 to 1,647,729. Two fosmids produced from the Hawaiian strain complement this region. Both WRMHS24L15 and WRMHS02O10 contain the coding region as well as at least 5kb upstream and downstream from the start and stop coordinates.

3.7 A Comment on End Alignment Procedures

The fosmid libraries produced were aligned to the WS210 data freeze of the *C. elegans* genome with BLAST alignment. BLAST alignment will forego precision for speed, providing a list of possible matches to the sequences given. The BLAST alignment performed provided a list of output matches of 1,398,885 individual blast hits, to the query search numbering 28,630 ends. To streamline the process the best hits were chosen according to length of alignment and the bit score, an algorithm describing the quality of alignments, with better scores created by increased length and fewer gaps or mismatches. These best hits were aligned as pairs and further manipulations were made with the assumption that these pairs were true pairs.

Using pairs of end sequences to align ends on the genome allows for identification of unconventional inserts. Unconventional inserts are those seen aligning with both ends pointing in the same direction, with both ends pointing in opposite directions away from each other, to two different chromosomes, or with sizes that could not be packaged within the λ phage extract. All of these are seen within both the WRMHS and WRM06 libraries. These problem inserts may arise due to repetitive sequence alignment and due to misalignment by the blast algorithm, placing higher scores for unlikely candidates.

To determine if this misalignment was the cause of some of the unconventional inserts found in the WRMHS library, several of these inserts were analyzed to examine the permutations of blast hits produced to see if there are other more likely candidates. A PERL program was designed with the help of Jeff Magnusson to list all hits where complementary ends were found on the same chromosome. The program then discounted those hits where the ends were facing the same direction or in opposite directions away from one another. Finally, it discounted all hits which had ends more than 60 kb and less than 10 kb from one another.

Table 7 represents the output produced using a program to align and match all blast hits for each clone discarded due to ends aligning to two separate chromosomes. 361 of the 436 clones show multiple conventional alignments. Most of the clones with multiple conventional alignments each show hits in multiple chromosomes.

Alignments for the 361 clones had an average of 2.5 chromosomes hits per clone. The multiple hits produced the 901 possible chromosome alignments for the set. The alignments determined to have conventional orientations, with ends on opposite strands facing each other, amounted to 10,949,876 placements. Even when these were trimmed down to take size into consideration, the clones aligning to between 10 and 60 kb segments amounted to 123,268 addresses. The 75 clones remaining, from the 436 initially tested, have no alternate single chromosome address and are more likely correct in their initial alignment.

Table 7: Permutations calculated for the blast hits produced by paired ends in the unconventional subgroup determined to align to two chromosomes

	Permutations for 436 clones' different chromosomes	Permutations of paired ends aligning to same chromosome	Permutations of paired ends aligning to same chromosome in size range	No valid permutation
I	145	1,893,136	2,849	
II	160	1,003,828	6,879	
III	134	1,189,720	977	
IV	155	955,206	3,621	
V	173	777,269	108,408	
X	134	5,130,717	534	
Total	901	10,949,876	123,268	75

Many of these alignment possibilities have a similar alignment score for one or both sides of the pair suggesting that these clones are most likely not due to translocations, but a failure in selecting the most likely placement. However, there are 75 clones, which do not align with both pairs on a single chromosome within the selection size possible. These may be interesting candidates to examine more closely for larger chromosomal rearrangements in these areas.

Two such clones are seen in the group of 75, which may show a rearrangement. Both have ends that fall within 83 kb of one another on chromosome IV and 137 kb on chromosome II. The ends on chromosome IV are nestled between two gapped regions of 18.5 kb and 45.5 kb. Chromosome II has two ends falling into gapped regions in the fosmid library. Both ends on chromosome II fall within 5 kb and 20 kb of known indels and may be caused by an unbalanced translocation event resulting in the deletions seen. These associated deletions are encompassed in individual fosmids.

4 Discussion

Two libraries have been produced. The first, WRM06, is a fosmid genomic library using the canonical N2 Bristol strain of *C. elegans*. The second, WRMHS, is a Hawaiian geographic isolate derived fosmid library. The CB4856 strain used has been determined to be the most divergent *C. elegans* isolate from N2 and represents a genome from the other half of a balancing selection evolved within the worm. The N2 library was created in 2006. The WRMHS library was made in 2010 as part of my Master's project.

Initial production of the WRM06 fosmid library was done to provide the worm community with an alternative vector source to the cosmids produced in the worm genome-sequencing project, as many of these original clones have lost viability or rearranged over time. Besides their use in complementation assays, fosmids have garnered interest from those studying the whole genome. One application provides the opportunity to study gene function through recombineering (Dolphin and Hope 2006). The large size of the fosmids allows studies of native transcription/ translation with regulatory regions in the promoter and intergenic regions, as well as those possibly found downstream, for all but the few largest genes in *C. elegans*.

The N2 library was designed to obtain 5x coverage of the genome. The actual coverage was somewhat less due to misalignment and unclonable regions with gaps left in the library's contiguous sequence. By using the Hawaiian strain CB4856 to produce another library, it was hoped that the fosmids would overlap to complement the areas and provide contiguous sequence over the entire genome. As there are significant differences in genome structure, which may relate to altered distribution of repeat elements between the two strains, this idea seemed feasible. With genomic variation providing sequence difference, allowing proper addressing and tiling, and molecular changes providing a different environment, on which bacterial machinery may be able to function, we hoped the gaps could be filled. The misalignment and molecular challenges impacting microbial DNA replication machinery, due to repetitive regions, are thought to cause the gaps in the imbricate sequence. This rationale appears justified as the combination of the two libraries did significantly reduce the number of gaps, seen in Table 2, and did increase the number of genes covered by fosmids, described in Table 3. These two libraries may find increased use in genomic studies and provide tools to determine the genomic structural changes that produced the differences between the Bristol and Hawaiian geographic isolates.

4.1 Analysis of Clones Produced in the Two Geographic Isolate Libraries

As stated in the results, clones from both libraries were end sequenced. The end sequences were blast aligned to the WS210 library and the positional data of the best-fit outputs were used to calculate the orientation and length of the inserts. The clones displaying unconventional inserts were separated and the fosmids containing conventional sequence were positioned on chromosomes. The separation of the unconventional inserts was necessary for tiling of the clones, but it may be premature to dismiss them outright as they could show structural differences from the annotated genome displayed as the canonical N2 sequence.

All of the different unconventional inserts may be caused by structural variation within the genome. This could be due to repetitive sequences, chromosomal rearrangements or possibly a combination of the two. Figure 3 illustrates possible structural mechanisms that could cause unconventional inserts. For example, a sequence tandemly repeated with repetitive elements bookending it might contain an insertion that represents the clone's size smaller than it actually is (Figure 3B). Palindromic sequence and inverted repeats can cause clones to look as though the ends are pointing in the same direction or are in opposite directions away from one another (Figure 3d and 3e). The clones displaying inserts with ends aligning on two different

chromosomes are not likely caused by tandem repeats, but may be misalignments caused by repetitive sequence or result from translocations that have occurred within the isolate in question (Figure 3a). Fosmids with insert sizes larger than would be expected are less likely to be caused directly by tandemly repeated sequences within the genome. However, the lack of a tandemly repeated region, when aligned to a genome that contains the repetitive element, may falsely represent the size of a clone. Such a clone would appear larger than is possible given the size limitations that λ phage packaging extract imposes on length (Feiss *et al.* 1977; Figure 3c).

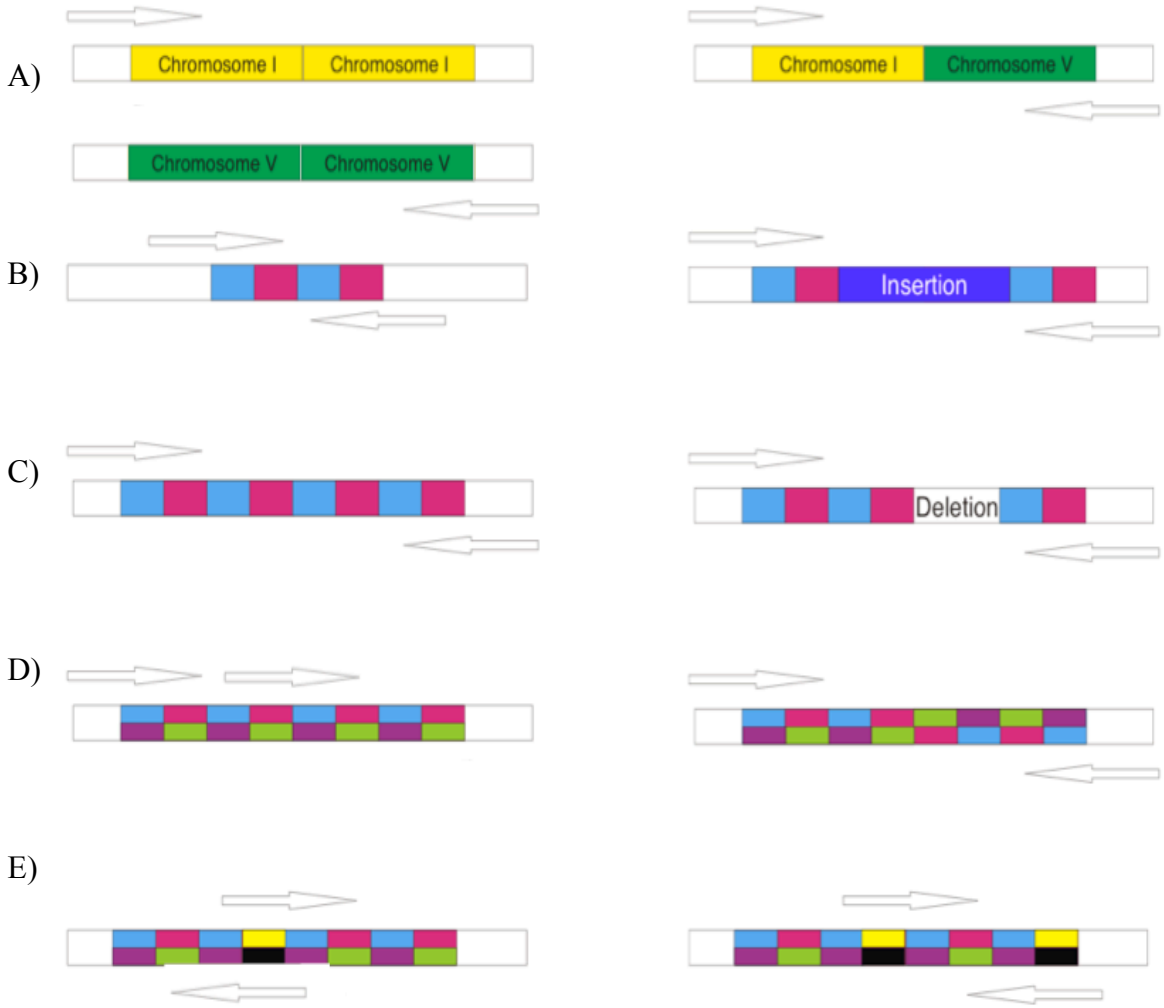

Figure 3: Representation of the five unconventional inserts and the rearrangements that could produce them. Each line represents a chromosome depicting the end sequences on the left followed by the rearrangement on the right. Like coloured fragments represent like sequence. A) Shows a translocation between two chromosomes producing paired ends aligning to different chromosomes. B) Insertion creating a larger sized fosmid sequence than expected. C) A deletion in a tandem repetitive region, which would show up as a smaller insert than is expected. D) Either palindromic sequence or an inverted repeat causing both ends to appear to be on the same chromosome. E) A sequence with a spontaneous tandem duplication showing ends to be on opposite strands facing away from one another.

Clones displaying a size under the minimum allowed by the packaging extract are likely caused by misalignment of repetitive sequence, or by ambiguous alignment of repetitive sequence causing difficulty defining the true location. The undersized inserts may also be caused by an inserted element into the chromosome making the end points appear to be far closer than would be allowed by the λ phage packaging extract (Feiss *et al.* 1977).

The fosmid clones with inserts calculated as too small to be packaged may be examples of ambiguous repetitive alignment. Of the 868 clones with aligned sizes calculating to less than 15 kb all but 56 fell into one of 27 groups of fosmids containing more than 4 members, in which the insert ends were aligned within 10 kb of one another (Appendix B). The largest of these groups consisted of 280 vectors with ends between 15,056,645 to 15,072,421 on chromosome I. This is a region with repetitive sequence, which makes even the current chromosome organization questionable in this area (personal communication Robert Waterston). The N2 library contains a similar group in the same area containing 576 clones and 21 other groups with over 4 members. Each one of these 27 groups in WRMHS and 21 groups in WRM06 are likely to provide evidence of regions of highly repetitive sequence within these genomes.

Discarding the clones that do not conventionally align simplifies the analysis of the library. It should be pointed out, however, that misaligned clones might fill in some of

the gapped sequence. Evidence for this possibility can be found in the number of fosmids aligning to regions directly beside a gap. Of the 56 individual fosmids and the 27 groups of clones, 37 fosmids aligned directly beside a gap. Some of the groups had multiple clone alignments beside gaps. The analysis necessary to evaluate the correct positions of these fosmids within the genome would be time consuming and could cost as much as the sequencing in the original library. For these reasons, the non-conforming inserts were not analyzed further. However, making clones available to interested parties will allow the discernment of the alignment of some of these unconventionally aligned inserts and the regions that produced the fragments.

With both libraries produced and end sequenced the overall quality of the fosmid resources could be compared to one another. It is clear that the library produced from the Hawaiian geographic isolate is superior in most metrics displayed. Table 1and table 2 detail the individual libraries' fosmids and their coverage of the canonical genome. There are only a few instances in which the N2 library produces more favourable results. This was not expected. The slightly smaller size (384 fewer clones produced and sequenced) of the library would alone suggest this is improbable. This incongruity is strengthened by the sequencing being somewhat less successful in the WRMHS library (28,326 ends) in comparison to the WRM06 (31,397 ends) resulting in a 9.1 % smaller aligned Hawaiian library. The sequence divergence of the Hawaiian strain from N2 decreases the likelihood of a better library being produced from the

CB4856 isolate. The unlikelihood of a better quality library being produced from the alignment to a more divergent canonical genome is unexpected.

The regions of the genome displaying gaps in contiguous sequence within the libraries also exhibited differences between WRMHS and WRM06. The smaller total gapped area in contiguous fosmids seen in WRMHS was able to account for the smaller percentage of total protein coding regions associated with the gaps (Table 3). The decreased percentage of YAC derived coding regions from the total exhibited in WRMHS (Table 4) cannot be explained as easily. As well the lower concentration of repetitive elements in WRMHS gaps (table 5) are also not as simply explained. The use of a divergent genome may have had some beneficial effects on library production that could account for these inequalities. The differences in difficulty seen in finishing some genomes (Blakesley *et al.* 2010) may have some equivalence within geographic isolate sequencing. The genomic divergence may contribute to the increased quality in the WRMHS library, though the relatively small disparity between the genomes would make it unlikely.

In production of the libraries, several adjustments were made to the protocol used to purify and size DNA. The procedure used to clone the DNA was largely unchanged from WRM06 to WRMHS. One protocol was used in the WRM06 library, but the WRMHS library used two different protocols to polish and purify. The increased attention to polishing and purification of the Hawaiian DNA may have provided

cleaner material, or slightly different chemical environments for the inserted molecules. Using two separate purification protocols each for half of the Hawaiian library may have also contributed to the differences in coverage and coding regions attained, which seems more likely than the divergence causing a decrease in difficulty cloning some areas. It is difficult to provide evidence to support either postulate with the present information. Further work will need to be done to distinguish between these two possibilities. Evidence may be found by producing a new library with each DNA purification protocol performed and packaged independently. These libraries could then be sequenced to a depth similar to, or greater than, the depth used in this study using next generation sequencing technology. A comparison of the output could then provide a clearer view of the role the chemical environment played in the fosmid libraries by controlling the variable of differential genome sequence and allowing side by side comparisons with a single changed variable.

Table 2 shows that coverage of both the WRM06 and the WRMHS libraries has produced contiguous fosmids for just over 90% of the annotated genome. The N2 library contains 409 gaps amounting to 18,992,942 bp of missing sequence or 19% of the genome. The CB4856 library is missing 14,721,675 bp or 15% of the genome in 454 gaps. When tiled together there are still 294 gaps remaining, but only 10% of the genome, or 9,596,484 bp remain unaccounted for. Lack of complete complementation of one library by the other is not surprising given the regions which do not have contiguous sequence. The clustering of gaps at the end of the arms of each

chromosome is visible in Figure 1. The libraries, independently and combined, have fewer gaps towards the centres of each chromosome and these gaps tend to be smaller. Likewise, very few and very small gaps are seen on the X chromosome. A similar pattern was observed when sequencing the worm's genome, when groups were working through clusters of repetitive elements (Consortium 1998).

The co-localization of the gaps found in the fosmid and original cosmid libraries' contiguous sequence with repetitive regions has previously been demonstrated in multiple vertebrate species (Blakesley *et al.* 2004). The pattern of gaps is most likely due to misalignment of arms of the inserts to repetitive elements elsewhere in the genome or to incorrect size estimation of properly addressed ends. Areas of repetitive sequence are frequently duplicated and translocated within a genome, causing larger chromosomal anomalies that can be difficult to align (Bishop and Schiestl 2000; Volik *et al.* 2003). These two reasons provide the first point of difficulty when attempting to find contiguous sequence in these areas. Sequence repetitions also allow secondary and tertiary structures to form, such as Z DNA, palindromes and kinkable dinucleotide steps (Razin *et al.* 2001), which cause difficulty in propagating and packaging the eukaryotic sequence with bacterial machinery and hosts.

By exploring repetitive sequence distribution on a chromosome, I was able to detect the co-occurrence of areas that were originally only clonable using YAC vectors and increased levels of repetitive elements. By comparing protein coding gene positions

with regions 40 kb upstream and downstream from each end to provide non-coding sequence, I was able to see a co-incidence of Yeast sourced clones and repetitive elements in the same areas. The X chromosome had 5% of genes that were initially derived from YAC clones and 5% of sequence around the genes came from the same source. However, 7% of repetitive elements, falling within 40 kb of a gene, were around a protein-coding sequence initially cloned in a YAC. This implies a 48% enrichment in the number of repetitive elements within 40 kb of the ends of YAC derived genes compared to the rest of the genome. This enrichment increases when moving from the originally YAC derived sequence to the individual gaps within libraries, and finally to the combined gaps when fosmids are tiled together (Table 6). The increasing quantity of repetitive elements in the non-covered regions suggests the clustered repeats may be preferentially excluded from the libraries. This is not surprising as repetitive regions are expected to be difficult to clone or align, and are likely the cause of the difficulty in filling these gaps.

4.2 Protein Coding Gene Coverage and Gaps

Gaps in contiguous fosmid sequences were analyzed to determine how many protein-coding genes are excluded from each library. Independently the two libraries cover about 85% of the genes (Table 3), but when combined the coverage jumps to about 93%. This increase will provide better access to DNA that was previously

unavailable, within bacterial vectors, and more complete access to genes in these regions. The increase in protein coding region coverage, when the libraries' contigs are combined, indicates that at least a portion of the sequence believed to be unclonable (Waterston and Sulston 1995) was simply mismapped or difficult to propagate in bacteria. However, the incomplete overlap suggests that there are regions of the chromosome, which may be unclonable or are, more likely, very difficult to clone.

The numbers of genes in the libraries will most likely increase due to re-annotation, as has been seen from the WS140 alignment of the WRM06 library, as well as new insights in the literature describing regions and repetitive sequence. One such example of this was described by Vergara *et. al.* (2009). Their analysis of the genomes of many different laboratory strains, as well as CB4856, noted several duplications arising within the *C. elegans* community's N2 strains. They were able to trace these events to the original labs that disseminated the N2 stocks and concluded that there were several tandem duplications within the genome of the laboratory strain used for the sequencing project. These duplications are only present in a subset originating from a single lab, providing further support for the idea that genetic drift is prevalent among laboratory N2 strains (see also Denver *et al.* 2009; Flibotte *et al.* 2010; Hillier *et al.* 2008). Interestingly, they also showed these duplications are missing from the Hawaiian strain (Vergara *et al.* 2009). Two tandem duplications differentiating the

sequenced N2 from our VC196 and CB4856 can be detected in the two fosmid libraries.

The region this duplication covers is on chromosome V from 2,347,883 to 2,562,875. The combined fosmid libraries have a gap on chromosome V for this region. Further analysis revealed seven individual clones from the Hawaiian strain that cover this interval. These clones were originally dismissed as too large as they have calculated sizes of 141 kb to 149 kb. The tandem repeat was discovered to be 106,707 bp in length, which would place these clones at a normally expected length, if the repeat is not present in this strain, as was shown by Vergara *et al.*(2009). Still further investigation showed 6 clones from this region in the N2 library, with similar lengths, suggesting that this 106 kb tandem repeat is not contained in the VC196 wild type genome.

A smaller duplication was also detected on chromosome V between 8,813,143 and 8,892,906. There are14 clones covering this region for the WRM06 library and all were excluded from alignment in contigs as they had calculated sizes between 68 kb and 74 kb. Accounting for the duplication size of 37,642 bp, these would again be deemed valid clones if the repeat were missing. The WRMHS library has 6 clones covering this region with inserts calculated from 72 to 85 kb. With inserts displaying paired ends aligning to this region, with sizes that are outside of that which is able to be packaged by λ phage (Feiss *et al.* 1977), evidence is provided that these tandem

repeats are not present in the libraries and therefore the genomes. These repeats were not seen in CB4856 in the study (Vergara *et al.* 2009). These two different tandem repeats are associated with 36 protein-coding genes. Due to the lack of these regions being repeated in these strains, the total genes expected in the libraries will be decreased by approximately the same number. These two instances are unlikely to be unique.

As the genome is re-annotated, coverage in the libraries will change and may encompass more coding sequence. Other tandem repeats described in the same paper (Vergara *et al.* 2009) were not readily visible within the libraries as gaps, most likely due to their small size. Further study of the Hawaiian clones determined to be outside of a suitable size for packaging by the λ phage particles showed other possible candidates, including either tandem repeat areas within the N2 library not held within the Hawaiian strain , or simple deletions of sequence. Chromosome IV contains an anomaly spanning 14 genes and deleting 41.9 kb of sequence (Maydan *et al.* 2010). Supportive evidence for this being a simple deletion was provided by three clones that were initially discarded as too large, measuring between 77 and 80 kb. The calculated size did not take into consideration the deletion interval originally. By factoring in the missing sequence, the clones are of the expected compatible size.

4.3 Correlation of Gaps with YAC-derived Coding Sequences

Even after combining the two libraries, there are several extant gaps. This may be due to a condition described to me in a personal communication with Robert Waterston. While sequencing the worm's genome, portions of the chromosomes were unable to be packaged or propagated within a bacterial cell and could only be cloned within Yeast Artificial Chromosomes (YAC). The incomplete overlap suggests that there are regions of the chromosome that are unclonable (Waterston and Sulston 1995) or, more likely, very difficult to clone within bacterial vectors.

During sequencing of the first N2 genome (Consortium 1998), even with a six fold redundant coverage of the genome in cosmids, non-random gaps in contiguous sequence persisted. Unsuccessful efforts were made to fill these gaps using cosmid and fosmid clones. In these cases YAC vectors were used to complete the genomic coverage. Our approach of using fosmid libraries prepared from two different geographic isolates reduced the number of gaps but still left a substantial number of holes in coverage along all the chromosomes. The apparent inability of fosmids or cosmids to cover all regions of the *C. elegans* genome remains. By exploring YAC-derived sequence co-occurrence within the libraries, it should be possible to see if the gaps in the original cosmid libraries are the same as those seen in our fosmid libraries.

By comparing the number of genes initially derived only in yeast sourced clones to all other coding regions found in library gaps it should be possible to determine if these sequences are still evading coverage in bacterial vectors. There is a marked enrichment of the coding sequences only cloned originally in YAC-derived vectors within the fosmid gaps. 81% of coding sequence genes, originally only cloned in YACs, are covered in the combined libraries (table 4). The enrichment shows the phenomenon described previously (Consortium 1998; Waterston and Sulston 1995) has persisted, using a complete fosmid clone set. When the libraries' coverage is combined, the enrichment is greater. This lends credence to the suggestion that some of these sequences, which were originally only YAC derived, may be unclonable in bacterial hosts (Waterston and Sulston 1995). What is more likely is that they were hard to clone, or tricky to propagate within the strain originally used, or were difficult to clone using the production and packaging system adopted in the sequencing project.

While the gapped sequences within the combined libraries were proportionally higher in YAC derived sequence, we were able to clone 2,638 coding regions initially believed to be unclonable in bacteria. Even alone, with a 4.2X and 3.9X theoretical coverage, the N2 and CB4856 libraries were able to capture 1,776 and 2,218 genes respectively, which were previously only represented in YAC vectors. This result shows that some genes not previously captured within cosmids are clonable, challenging previous assertions (Waterston and Sulston 1995). This provides evidence

that suggest the protocol, genome or bacterial strain and vector, used for producing a library affect the resulting coverage in that library. This also reinforces that there was likely a difficulty with cloning certain sequences or propagating them in the strain originally used in the sequencing libraries, which may also have been caused by the production or packaging systems adopted at the time.

A study performed on multiple vertebrate genomes, using BAC clones, with both copy control and multicopy vectors to limit the copy number of individual sequences within the host strain showed sequences that were difficult to package are made less problematic when maintained with fewer copies in the host (Blakesley *et al.* 2010). This is likely due to the interaction of library sequences within the host, causing secondary and tertiary structures that will stall or freeze replication machinery, and/or cause repair mechanisms to alter the sequence then propagated, or halt replication of the cell altogether. If the phenomenon were entirely due to recombination, the inserts would likely contain internal rearrangements that would remain imperceptible after end sequencing if carried in a multicopy vector. Empty vectors may also be a symptom of such a rearrangement and these were not observed. The complete lack of inserts for some areas points to an inability in packaging or propagation of those regions. The use of a single copy vector may be the reason the WRMHS and WRM06 fosmid libraries were able to capture these previously unclonable sequences. Even though fosmid libraries were used in the original sequencing, the relative proportion, 113 fosmids to 2,527 cosmids (Consortium 1998), would have made them unlikely to

have an effect on the overall genome coverage. A hint at the possibility of difficult sequences cloning better in fosmids was made when it was described that fosmids allowed a third of the gaps found in the central region of the chromosomes, but not at the ends, to be bridged (Consortium 1998). The results described in this study support this conclusion.

The possibility remains that the lack of complete coverage for either of the fosmid or the cosmid libraries is a result of a depth of sequencing issue. This may account for the differential presentation of gaps in contiguous sequence seen in the cosmids as well as the N2 and CB4856 fosmids. One piece of data provides evidence to the contrary. Gap positions on the genome remain largely unchanged from one library to another. Even distribution of clones over the genome would unlikely result in three separate libraries with regions of overlap and gaps at similar positions with little variation between them. It might be possible to cover some gapped regions more completely by producing a larger library and sequencing it to a greater depth. However, the lack of genomic coverage with a bacterial vector and the almost complete YAC based coverage, with far less clones proportionally, suggest the problem lies in the packaging or propagation in the specific host.

The increased coverage seen from the N2 to the CB4856 fosmids may have been due to the larger quantity of aligned sequences offsetting the decreased likelihood that the more difficult regions are cloned. A similar increase may have been visible if another

N2 library were produced and sequenced to the same depth as the Hawaiian library. To explore this a chi squared test was performed on the proportion of genes found in the WRM06 and WRMHS libraries. The test showed we are able to reject the null hypothesis and support the alternate, that the proportion of YAC and cosmid derived genes were dependant on the library producing them. This does not separate the genome used in the library from the method used to produce it and may easily be misconstrued as one or the other. However, decoupling the use of different genomes and different purification methods is not possible with the current study and may need support to provide conclusive evidence. By using next generation sequencers, as was described earlier, it would be possible to sequence a larger library to a much greater depth than was possible in this study. The greater depth could allow the investigator to probe the differences seen with another geographic isolate, by making and sequencing libraries produced independently with the different purification techniques and strains. This would thereby separate the variables to determine their individual contribution to the coverage.

4.4 Alignment Difficulties

Looking at the permutations created in some end sequence pairs it becomes apparent that excluding clones that align in unconventional ways may be premature. The multiply aligned end sequences are the most difficult hurdle to overcome in any

sequencing or genomic study. Repetitive sequences leading to miscalculation or improper addressing of ends are a difficult problem, especially with the smaller insert sizes common for next generation sequencing technology. The paired end sequencing performed here showed substantial issues with 500 bp ends and 30 kb to 45 kb overall length. Pairing next generation sized sequences may not minimize the issue.

It is quite clear that given the complexity of some genomes, the BLAST algorithm may not be ideal as a basis for alignment of sequence with some repetitive elements. The number of clones that initially showed ends aligning to separate chromosomes supports this. When the permutations of end alignments in this group were explored, the possibility of misalignments became more apparent. The 123,268 permutations falling between 10 kb and 60 kb on the same chromosome suggest the unlikelihood of the original alignment to separate chromosomes being correct. The end placements may, however, have been correct initially and a more in depth study of the structure of these inserts should define their actual position.

As previously described, the clones not showing multiple alignment positions are more likely correct than those with multiple calculated positions. The list of clones in this category was checked for multiple fosmids aligning to the same end positions. Two candidate clones were found. With end positions relating to chromosomes IV and II, and within close proximity to a gapped region on IV, these may indicate a translocation event in the WRMHS library. Further, analysis revealed that the region

falling between the two clones unpaired end positions aligning on chromosome IV contains a deletion, discovered by CGH (Maydan *et al.* 2010). The deletion position on chromosome IV provides evidence that the translocation is unbalanced and affects at least three genes. It seems likely this would be an insertional translocation as there is no genetic evidence from crosses between these two strains of a translocation of chromosome arms.

Repetitive sequences are problematic for alignment as well as *de novo* sequencing and lead to challenges in cloning. These issues are possibly reflected in the concentration of repetitive elements in the gaps found in the libraries. Using alternative low copy vectors has alleviated the problem, to some extent, but does not eliminate it. BLAST software is not entirely able to differentiate between like sequences and new alignment algorithms might be necessary to empower paired end sequencing with inter-sequence spacing. The intervals used also have to be larger than the largest individual sequence repeated, which can be quite large as seen in the tandem duplications apparent in the canonical sequence. This being said, the libraries produced are likely not representative of the total clonable and alignable sequence. There are still protein-coding genes, originally derived in cosmid sequences, which have not been captured within fosmids. These coding regions are represented by the 19% of genes, which were not YAC derived originally, that fall into gaps in combined coverage. The missing open reading frames suggest that at least the sequence held around these 308 protein coding genes should be clonable. The sequence may be found within the

groups of clones containing unconventional inserts, but may have been missed in these two libraries.

An example of a missed coding sequence, within the cosmid and WRM06 libraries, was seen with the *unc-119* gene. One reason these sequences may have been missed in this study is that they are unclonable within the vectors and hosts employed in this investigation. Further research on these gapped sequences may provide the answer. The quality attained in the WRMHS library suggests that any further attempts to fill in the gaps, seen in contiguous fosmid coverage, should be encouraged to vary the DNA purification and polishing protocols. This combined with using a copy control vector would likely produce the best possibility for covering the gapped regions.

In conclusion, I have created two fosmid libraries. The WRM06 library made in 2005 from the N2 variant of *C. elegans* VC196 and the WRMHS fosmids made as part of my thesis from the CB4856 geographic isolate. They were mapped to the WS210 genome and together cover 92.8% of genes. Both libraries show coverage of previously unclonable regions of *C. elegans* DNA in bacterial vectors. The remaining gaps in contiguous sequence for the original cosmid library and each of the fosmid libraries, independently and combined, show an increasing concentration of repetitive elements in the gaps. These repeat sequences may be the greatest cause for difficulty in cloning and propagation. These trends have been seen in other organisms.

Of the two libraries I produced, WRMHS was superior based on clone mean size and standard deviation as well as genomic coverage and gap size and standard deviation. The WRMHS library was able to cover 115 gaps in contiguous coverage within the WRM06 library and with them 1893 genes not contained in fosmids. I was also able to cover 2600 genes not previously captured in bacterial vectors. WRMHS differed from WRM06 in the strain used to produce DNA as well as the post purification polishing regimes performed to limit the effect that using only one type of purification might have on the resulting library. Due to the change of multiple variables it is not possible to distinguish which specific variable caused the significant increase in capture of previously uncloned DNA. Using next generation sequencing technology could possibly differentiate the relative importance of the different variables. The library has also have captured regions containing several larger chromosomal differences between N2 and CB4856 including deletions and duplications, examples of which can be seen in several fosmids. Further exploration of these regions may be performed by sequencing the clones aligning to the regions or sequencing groups of clones that are aligning unconventionally. This technique may be empowered by using the program I designed, described in section 3.7, to increase the likelihood that the unconventional alignments are not due entirely to mismapped repeat regions. This would have the added benefit of decreased sequencing cost and/or man hours for subcloning. The sequencing may also uncover some other clones containing genes of interest or regions in gaps.

References

Bashir, A., S. Volik, C. Collins, V. Bafna and B. J. Raphael, 2008 Evaluation of Paired-End Sequencing Strategies for Detection of genome Rearrangements in Cancer. PLoS Comput Biol **4:** e1000051.

Bauchwitz, R., and F. Costantini, 1998 YAC Transgenesis: A Study of Conditions to Protect YAC DNA from Breakage and a Protocol for Transfection. Biochimica Et Biophysica Acta-Molecular Cell Research **1401:** 21-37.

Bishop, A. J. R., and R. H. Schiestl, 2000 Homologous Recombination as a Mechanism for Genome Rearrangements: Environmental and Genetic Effects. Human Molecular Genetics **9:** 2427-2434.

Blakesley, R. W., N. F. Hansen, J. Gupta, J. C. McDowell, B. Maskeri *et al.*, 2010 Effort Required to Finish Shotgun-Generated Genome Sequences Differs Significantly Among Vertebrates. BMC Genomics **11:** -.

Blakesley, R. W., N. F. Hansen, J. C. Mullikin, P. J. Thomas, J. C. McDowell *et al.*, 2004 An Intermediate Grade of Finished Genomic Sequence Suitable for Comparative Analyses. Genome Res **14:** 2235-2244.

Brenner, S., 1974 Genetics of *Caenorhabditis-Elegans*. Genetics **77:** 71-94.

Caenorhabditis Genetics Center web site, https://dbw6.msi.umn.edu/cgcdb/strain.php?id=7525, Aug 30 2007

Celniker, S. E., L. A. L. Dillon, M. B. Gerstein, K. C. Gunsalus, S. Henikoff *et al.*, 2009 Unlocking the Secrets of the Genome. Nature **459:** 927-930.

Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward and D. C. Prasher, 1994 Green Fluorescent Protein as a Marker for Gene-Expression. Science **263:** 802-805.

Consortium, C. e. G. S., 1998 Genome sequence of the nematode *C. elegans*: a Platform for Investigating Biology. Science **282:** 2012-2018.

Coulson, A., J. Sulston, S. Brenner and J. Karn, 1986 Toward a Physical Map of the Genome of the Nematode *Caenorhabditis-Elegans*. Proceedings of the National Academy of Sciences of the United States of America **83:** 7821-7825.

Denver, D. R., P. C. Dolan, L. J. Wilhelm, W. Sung, J. I. Lucas-Lledo *et al.*, 2009 A Genome-Wide View of *Caenorhabditis elegans* Base-Substitution Mutation Processes. Proceedings of the National Academy of Sciences of the United States of America **106:** 16310-16314.

Denver, D. R., K. Morris and W. K. Thomas, 2003 Phylogenetics in *Caenorhabditis elegans*: An Analysis of Divergence and Outcrossing. Molecular Biology and Evolution **20:** 393-400.

Dolphin, C. T., and I. A. Hope, 2006 *Caenorhabditis elegans* Reporter Fusion Genes Generated by Seamless Modification of Large Genomic DNA Clones. Nucleic Acids Research **34:** -.

Ellis, H. M., and H. R. Horvitz, 1986 Genetic-Control of Programmed Cell-Death in the Nematode *C-Elegans*. Cell **44:** 817-829.

Feiss, M., R. A. Fisher, M. A. Crayton and C. Egner, 1977 Packaging of Bacteriophage-Lambda Chromosome - Effect of Chromosome Length. Virology **77:** 281-293.

Fire, A., and R. H. Waterston, 1989 Proper Expression of Myosin Genes in Transgenic Nematodes. Embo Journal **8:** 3419-3428.

Fire, A., S. Q. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver *et al.*, 1998 Potent and Specific Genetic Interference by Double-Stranded RNA in *Caenorhabditis elegans*. Nature **391:** 806-811.

Flibotte, S., M. L. Edgley, I. Chaudhry, J. Taylor, S. E. Neil *et al.*, 2010 Whole-Genome Profiling of Mutagenesis in *Caenorhabditis elegans*. Genetics **185:** 431-441.

Flibotte, S., M. L. Edgley, J. Maydan, J. Taylor, R. Zapf *et al.*, 2009 Rapid High Resolution Single Nucleotide Polymorphism-Comparative Genome Hybridization Mapping in *Caenorhabditis elegans*. Genetics **181:** 33-37.

Flicek, P., and E. Birney, 2009 Sense from Sequence Reads: Methods for Alignment and Assembly. Nature Methods **6:** S6-S12.

Fraser, A. G., R. S. Kamath, P. Zipperlen, M. Martinez-Campos, M. Sohrmann *et al.*, 2000 Functional Genomic Analysis of *C. elegans* Chromosome I by Systematic RNA Interference. Nature **408:** 325-330.

Giraldo, P., and L. Montoliu, 2001 Size matters: Use of YACs, BACs and PACs in Transgenic Animals. Transgenic Research **10:** 83-103.

Hillier, L. W., G. T. Marth, A. R. Quinlan, D. Dooling, G. Fewell *et al.*, 2008 Whole-Genome Sequencing and Variant Discovery in *C. elegans*. Nature Methods **5:** 183-188.

Huang, R. Y., S. J. Boulton, M. Vidal, S. C. Almo, A. R. Bresnick *et al.*, 2003 High-Throughput Expression, Purification, and Characterization of Recombinant *Caenorhabditis elegans* Proteins. Biochem Biophys Res Commun **307:** 928-934.

Kim, U. J., H. Shizuya, P. J. Dejong, B. Birren and M. I. Simon, 1992 Stable Propagation of Cosmid Sized Human DNA Insertis in an F-Factor Based Vector. Nucleic Acids Research **20:** 1083-1085.

Kinscherf, T. G., M. N. Yap, A. O. Charkowski and D. K. Willis, 2009 Chef Procedures: A Rapid High-Temperature Method for Sample Preparation, a High Voltage Hepes Buffer System and the Use of Nusieve (R) Agarose. Journal of Rapid Methods and Automation in Microbiology **17:** 9-16.

Li, S., C. M. Armstrong, N. Bertin, H. Ge, S. Milstein *et al.*, 2004 A Map of the Interactome Network of the Metazoan *C. elegans*. Science **303:** 540-543.

Luan, C. H., S. Qiu, J. B. Finley, M. Carson, R. J. Gray *et al.*, 2004 High-Throughput Expression of *C. elegans* Proteins. Genome Res **14:** 2102-2110.

Maydan, J. S., S. Flibotte, M. L. Edgley, J. Lau, R. R. Selzer *et al.*, 2007 Efficient High-Resolution Deletion Discovery in *Caenorhabditis elegans* by Array Comparative Genomic Hybridization. Genome Research **17:** 337-347.

Maydan, J. S., A. Lorch, M. L. Edgley, S. Flibotte and D. G. Moerman, 2010 Copy Number Variation in the Genomes of Twelve Natural Isolates of *Caenorhabditis elegans*. BMC Genomics **11:** 62.

Mcnamara, P. T., A. Bolshoy, E. N. Trifonov and R. E. Harrington, 1990 Sequence-Dependent Kinks Induced in Curved DNA. Journal of Biomolecular Structure & Dynamics **8:** 529-538.

Mello, C. C., J. M. Kramer, D. Stinchcomb and V. Ambros, 1991 Efficient Gene-Transfer in *C-Elegans* - Extrachromosomal Maintenance and Integration of Transforming Sequences. Embo Journal **10:** 3959-3970.

Metzker, M. L., 2010 Applications of Next-Generation Sequencing Sequencing Technologies - the Next Generation. Nature Reviews Genetics **11:** 31-46.

Moerman, D. G., and R. J. Barstead, 2008 Towards a Mutation in Every Gene in *Caenorhabditis elegans*. Brief Funct Genomic Proteomic **7:** 195-204.

Murray, V., C. Monchawin and P. R. England, 1993 The Determination of the Sequences Present in the Shadow Bands of a Dinucleotide Repeat Pcr. Nucleic Acids Research **21:** 2395-2398.

Neil, D. L., A. Villasante, R. B. Fisher, D. Vetrie, B. Cox *et al.*, 1990 Structural Instability of Human Tandemly Repeated DNA-Sequences Cloned in Yease Artificial Chromosome Vectors. Nucleic Acids Research **18:** 1421-1428.

Odelberg, S. J., R. B. Weiss, A. Hata and R. White, 1995 Template-Switching during DNA-Synthesis by Thermus-Aquaticus DNA-Polymerase-I. Nucleic Acids Research **23:** 2049-2057.

Pel, J., D. Broemeling, L. Mai, H. L. Poon, G. Tropini *et al.*, 2009 Nonlinear Electrophoretic Response Yields a Unique Parameter for Separation of

Biomolecules. Proceedings of the National Academy of Sciences of the United States of America **106:** 14796-14801.

Razin, S. V., E. S. Ioudinkova, E. N. Trifonov and K. Scherrer, 2001 Non-Clonability Correlates With Genomic Instability: A Case Study of a Unique DNA Region. Journal of Molecular Biology **307:** 481-486.

Seidel, H. S., M. V. Rockman and L. Kruglyak, 2008 Widespread Genetic Incompatibility in *C. elegans* Maintained by Balancing Selection. Science **319:** 589-594.

Song, J. Q., F. G. Dong, J. W. Lilly, R. M. Stupar and J. M. Jiang, 2001 Instability of Bacterial Artificial Chromosome (BAC) Clones Containing Tandemly Repeated DNA Sequences. Genome **44:** 463-469.

Sulston, J. E., and H. R. Horvitz, 1977 Post-Embryonic Cell Lineages of Nematode, *Caenorhabditis-Elegans*. Developmental Biology **56:** 110-156.

Sulston, J. E., E. Schierenberg, J. G. White and J. N. Thomson, 1983 The Embryonic-Cell Lineage of the Nematode *Caenorhabditis-Elegans*. Developmental Biology **100:** 64-119.

Swan, K. A., D. E. Curtis, K. B. McKusick, A. V. Voinov, F. A. Mapa *et al.*, 2002 High-Throughput Gene Mapping in *Caenorhabditis elegans*. Genome Res **12:** 1100-1105.

Timmons, L., and A. Fire, 1998 Specific Interference by Ingested dsRNA. Nature **395:** 854-854.

Tursun, B., L. Cochella, I. Carrera and O. Hobert, 2009 A Toolkit and Robust Pipeline for the Generation of Fosmid-Based Reporter Genes in *C. elegans*. Plos One **4:** -.

Tuzun, E., A. J. Sharp, J. A. Bailey, R. Kaul, V. A. Morrison *et al.*, 2005 Fine-Scale Structural Variation of the Human Genome. Nature Genetics **37:** 727-732.

Vaglio, P., P. Lamesch, J. Reboul, J. F. Rual, M. Martinez *et al.*, 2003 WorfDB: the *Caenorhabditis elegans* ORFeome Database. Nucleic Acids Research **31:** 237-240.

Vergara, I. A., A. K. Mah, J. C. Huang, M. Tarailo-Graovac, R. C. Johnsen *et al.*, 2009 Polymorphic Segmental Duplication in the Nematode *Caenorhabditis elegans*. BMC Genomics **10:** -.

Volik, S., S. Zhao, K. Chin, J. H. Brebner, D. R. Herndon *et al.*, 2003 End-Sequence Profiling: Sequence-Based Analysis of Aberrant Genomes. Proc Natl Acad Sci U S A **100:** 7696-7701.

Waterston, R., and J. Sulston, 1995 The Genome of *Caenorhabditis elegans*. Proceedings of the National Academy of Sciences of the United States of America **92:** 10836-10840.

Wicks, S. R., R. T. Yeh, W. R. Gish, R. H. Waterston and R. H. A. Plasterk, 2001 Rapid Gene Mapping in *Caenorhabditis elegans* Using a High Density Polymorphism Map. Nature Genetics **28:** 160-164.

Wormbase website, http://www.wormbase.org, release WS140, Mar 26 2005

Wormbase website, http://www.wormbase.org, release WS210, Dec 22 2009

Yokobata, K., B. Trenchak and P. J. Dejong, 1991 Rescue of Unstable Cosmids by Invitro Packaging. Nucleic Acids Research **19:** 403-404.

Zhang, Y., L. Nash and A. L. Fisher, 2008 A simplified, Robust, and Streamlined Procedure for the Production of *C. elegans* Transgenes Via Recombineering. Bmc Developmental Biology **8:** -.

Appendices

Appendix A: Gap Positions for Each Library

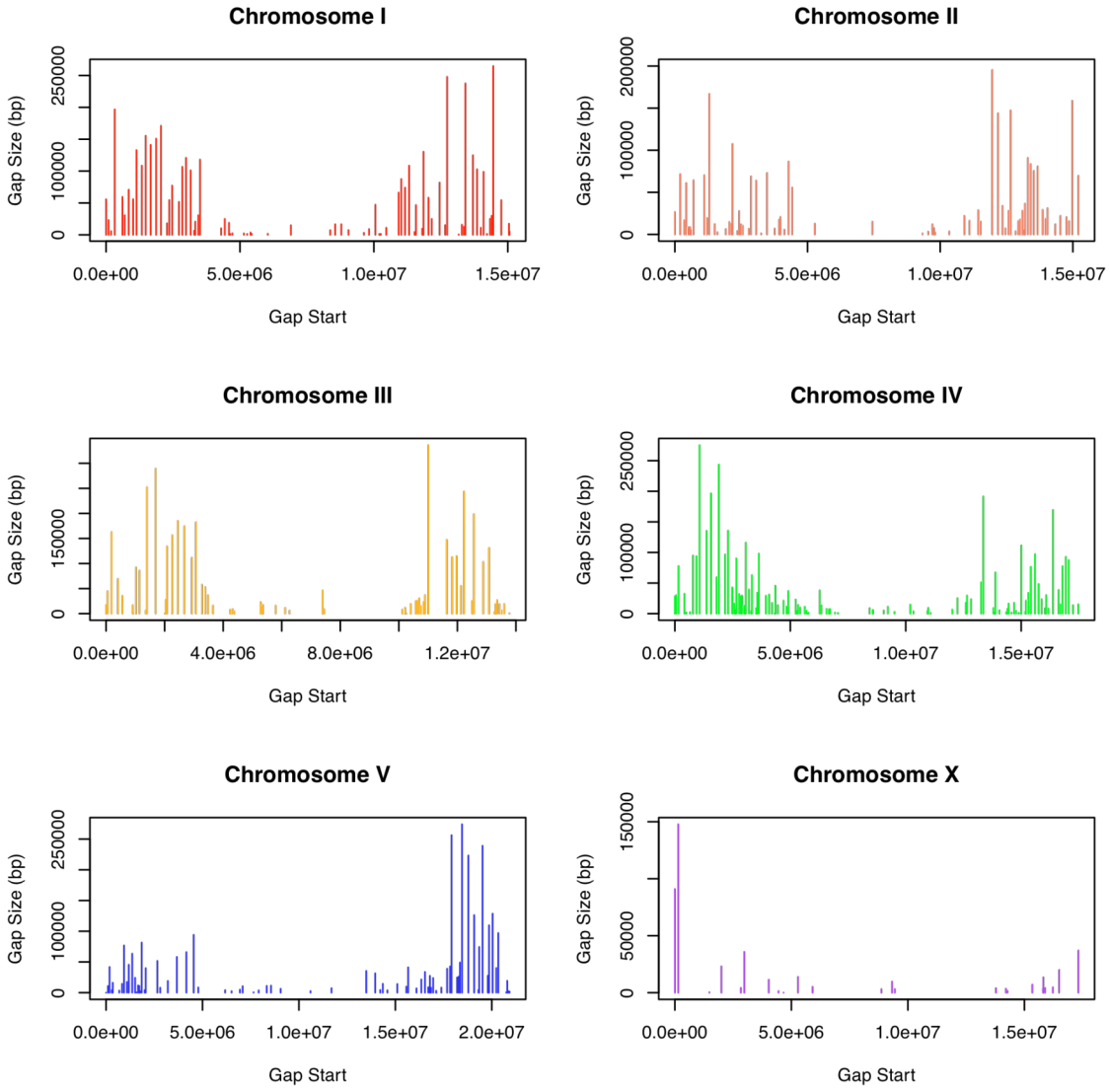

Figure A1: Graphic depiction of the gap distribution across the chromosomes for the WRM06 library. Top left in red shows chromosome I. Beside it in coral chromosome II. Middle left, in yellow, shows chromosome II. Besisde it in green is chromosome IV. The bottom left has chromosome V and beside it in purple is chromosome X. The X axis shows the position on the chromosome in which the gap lies. The Y axis shows the size of each gap.

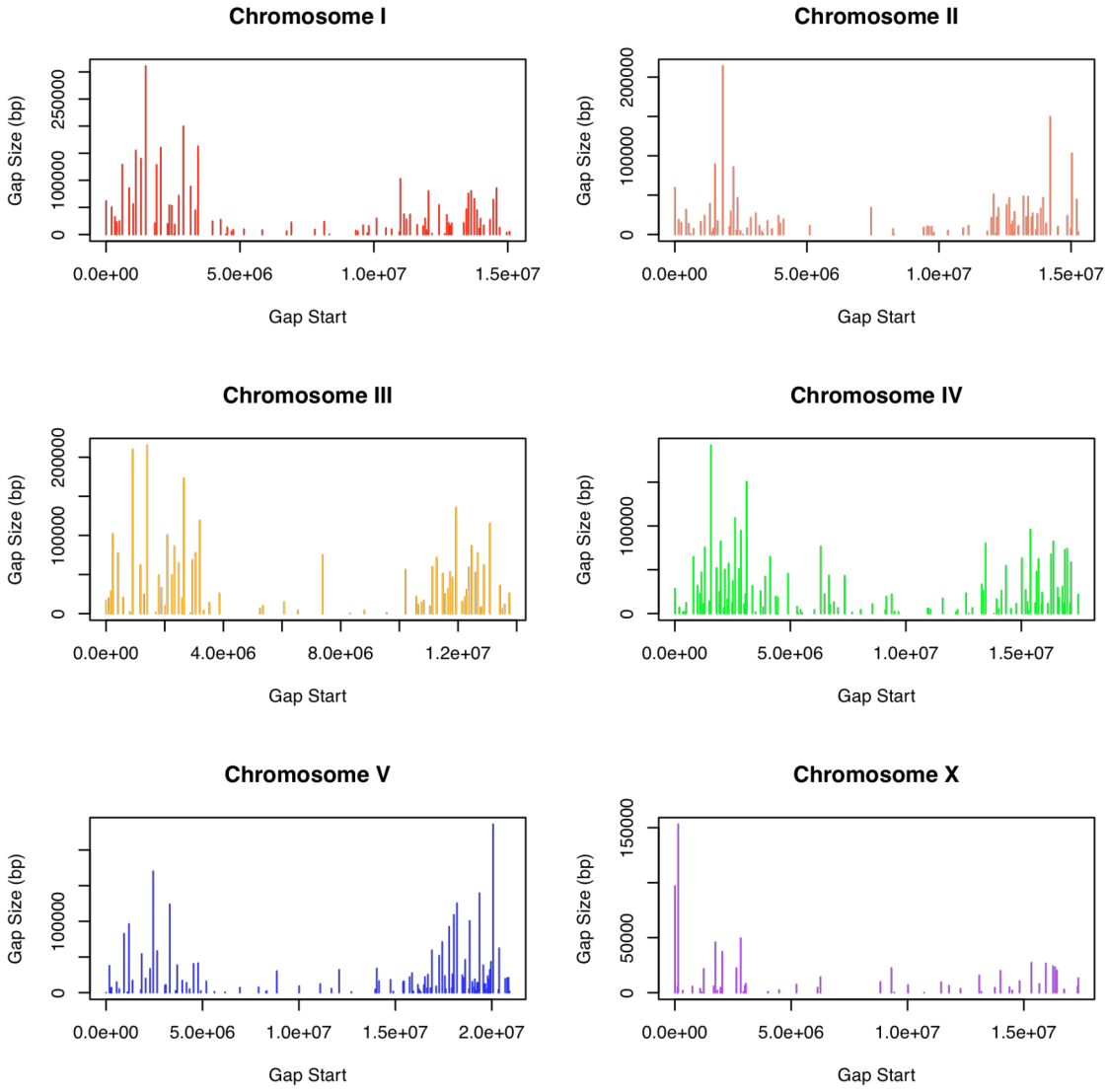

Figure A2: Graphic depiction of the gap distribution across the chromosomes for the WRMHS library. Top left in red shows chromosome I. Beside it in coral chromosome II. Middle left, in yellow, shows chromosome II. Besisde it in green is chromosome IV. The bottom left has chromosome V and beside it in purple is chromosome X. The X axis shows the position on the chromosome in which the gap lies. The Y axis shows the size of each gap.

Appendix B: Examples of Unconventional Inserts

Table B1: Fosmids with alignments too large or small to be pain reckaged by λ phage

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS30K04	933524	1242	934554	1166	I	True	1030
WRMHS28E04	933601	1053	934457	1002	I	True	856
WRMHS26B06	933661	473	934858	593	I	True	1197
WRMHS07A13	933931	444	933659	472	I	True	275
WRMHS28A07	934000	1038	934199	1027	I	True	1112
WRMHS26J01	934206	941	934221	725	I	True	976
WRMHS29O16	934281	723	933798	1105	I	True	664
WRMHS14D06	934770	660	924816	872	I	Gap	9954
WRMHS02B15	3990050	767	3998412	1258	I	Gap	8362
WRMHS01K13	4335006	1254	4341933	1421	I	True	6927
WRMHS28C22	4542593	1064	4543272	1197	I	True	821
WRMHS29B07	5152953	1247	5158235	1249	I	True	5282
WRMHS03G17	5160628	1031	5152940	1109	I	Gap	7688
WRMHS10C14	10132364	604	10118924	1386	I	Gap	13440
WRMHS12K15	10196036	1269	10209331	374	I	True	13295
WRMHS26G11	10204503	793	10212416	933	I	True	7913
WRMHS07F17	10204671	992	10215550	1375	I	True	10879
WRMHS28O16	10206735	1081	10218431	1284	I	True	11696
WRMHS29D17	10210676	1201	10208681	1230	I	True	1995
WRMHS28A08	10211621	1260	10206492	1234	I	True	5129
WRMHS26F03	10211952	1454	10212860	1068	I	True	908
WRMHS29H23	10217055	872	10208971	872	I	True	8084
WRMHS29J10	10218261	909	10220628	1013	I	True	2367
WRMHS10I15	10226103	1327	10217348	1303	I	True	8755
WRMHS29E14	10233153	1016	10218954	1411	I	True	14199
WRMHS26E04	10265632	1009	10273582	922	I	True	7950
WRMHS35J10	10267314	880	10254869	1029	I	True	12445
WRMHS09E20	10267964	811	10278891	1264	I	True	10927
WRMHS12C19	10268266	555	10281167	1363	I	True	12901
WRMHS32C01	10268595	1339	10281122	1443	I	True	12527
WRMHS27O09	10268604	1330	10277904	1537	I	True	9300
WRMHS07F05	10270526	1116	10277850	1424	I	True	7324

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS18H23	10270614	1027	10273626	1219	I	True	3012
WRMHS06M06	10274741	1345	10284918	1349	I	True	10177
WRMHS26D06	10275926	1441	10270700	857	I	True	5226
WRMHS06O02	10275979	1325	10268200	1090	I	True	7779
WRMHS29O24	10276007	1120	10268308	966	I	True	7699
WRMHS29A21	10281433	1397	10273848	1155	I	True	7585
WRMHS20L02	10289888	797	10275187	767	I	True	14701
WRMHS08P20	10946239	503	10949249	440	I	True	3010
WRMHS25G05	13244074	1068	13257861	1456	I	True	13787
WRMHS37J05	15056645	1260	15067799	1308	I	True	11154
WRMHS29B05	15059595	1256	15065105	1533	I	True	5510
WRMHS30A22	15060299	1330	15065707	1437	I	True	5408
WRMHS30I17	15060309	1565	15066762	1456	I	True	6453
WRMHS40A11	15060333	1520	15062234	1411	I	True	1901
WRMHS16G05	15060346	1471	15065780	1375	I	True	5434
WRMHS33F07	15060347	1495	15068044	1533	I	True	7697
WRMHS30I13	15060350	1192	15067650	1526	I	True	7300
WRMHS28D08	15060359	1345	15065455	1393	I	True	5096
WRMHS07P16	15060393	1236	15065202	1428	I	True	4809
WRMHS37N19	15060415	1354	15063273	1321	I	True	2858
WRMHS19I22	15060469	1238	15062893	1402	I	True	2424
WRMHS13N05	15060495	1262	15064133	1020	I	True	3638
WRMHS05J11	15060538	1264	15061229	1373	I	True	752
WRMHS19I23	15060555	1411	15065117	1541	I	True	4562
WRMHS03D10	15060560	1016	15066500	1332	I	True	5940
WRMHS34N08	15060591	1375	15063479	1402	I	True	2888
WRMHS28F20	15060603	1386	15067364	1439	I	True	6761
WRMHS32H02	15060616	1463	15062655	1452	I	True	2039
WRMHS05M17	15060637	1330	15063474	1264	I	True	2837
WRMHS26A17	15060700	1260	15065002	1301	I	True	4302
WRMHS10C18	15060703	1421	15062484	1500	I	True	1781
WRMHS26M24	15060707	1249	15061700	1375	I	True	993
WRMHS09N10	15060719	1338	15066201	1282	I	True	5482
WRMHS36D14	15060721	1352	15066913	1210	I	True	6192
WRMHS23M15	15060745	1234	15064067	1439	I	True	3322
WRMHS07L15	15060755	1236	15061134	1341	I	True	1032
WRMHS26O11	15060755	571	15063572	1314	I	True	2817
WRMHS30O20	15060771	1345	15062145	1502	I	True	1374
WRMHS19L08	15060787	1328	15065696	1572	I	True	4909
WRMHS21I12	15060802	1085	15067973	1280	I	True	7171
WRMHS27G09	15060817	1358	15064153	1321	I	True	3336
WRMHS29J14	15061006	1182	15066062	1338	I	True	5056
WRMHS28E01	15061020	1410	15063774	1397	I	True	2754
WRMHS09O06	15061031	1173	15066345	1382	I	True	5314

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS28D14	15061050	1136	15062757	1531	I	True	1707
WRMHS31N17	15061094	1539	15067942	1458	I	True	6848
WRMHS33B20	15061109	1408	15062374	1458	I	True	1265
WRMHS20C04	15061109	1053	15068111	1315	I	True	7002
WRMHS05B03	15061124	1264	15061427	1323	I	True	1111
WRMHS06L12	15061133	1269	15069034	1251	I	True	7901
WRMHS31G16	15061158	1463	15047276	1327	I	True	13882
WRMHS29H17	15061175	1360	15066262	1358	I	True	5087
WRMHS26B11	15061180	1262	15066238	1262	I	True	5058
WRMHS30O18	15061225	1387	15063509	1406	I	True	2284
WRMHS29C22	15061267	1312	15062498	1432	I	True	1231
WRMHS01P11	15061278	1262	15062502	1347	I	True	1224
WRMHS34O06	15061284	959	15065839	1474	I	True	4555
WRMHS29E04	15061355	1293	15064415	983	I	True	3060
WRMHS06L23	15061362	1275	15062412	1349	I	True	1050
WRMHS29G01	15061380	723	15062444	1456	I	True	1064
WRMHS11B09	15061380	377	15062778	1411	I	True	1398
WRMHS36D19	15061381	1448	15067158	1419	I	True	5777
WRMHS16E22	15061434	1352	15064763	1471	I	True	3329
WRMHS29H22	15061539	1369	15060582	1295	I	True	957
WRMHS09P22	15061634	1247	15064768	1288	I	True	3134
WRMHS29B11	15061663	1146	15066665	1384	I	True	5002
WRMHS26B14	15061681	1443	15060855	1391	I	True	826
WRMHS13F08	15061752	1227	15060466	1210	I	True	1286
WRMHS31H02	15061766	1522	15062358	1467	I	True	1027
WRMHS17B02	15061774	1363	15067180	1435	I	True	5406
WRMHS30J22	15061844	1445	15063891	1391	I	True	2047
WRMHS26L07	15061990	1535	15064764	1134	I	True	2774
WRMHS27A02	15061994	1472	15063734	1445	I	True	1740
WRMHS08H11	15061998	1249	15067938	1363	I	True	5940
WRMHS09F14	15062005	1123	15067857	1295	I	True	5852
WRMHS32M09	15062011	1591	15065363	1450	I	True	3352
WRMHS30O06	15062014	1607	15062669	1493	I	True	1029
WRMHS11E15	15062033	1133	15067245	1546	I	True	5212
WRMHS29M07	15062054	1397	15064641	1574	I	True	2587
WRMHS01I18	15062072	1098	15061098	1365	I	True	974
WRMHS09E18	15062110	1194	15065420	1267	I	True	3310
WRMHS29H01	15062138	1175	15062977	1472	I	True	839
WRMHS29C03	15062156	1448	15066994	1482	I	True	4838
WRMHS25J05	15062176	1358	15064281	1393	I	True	2105
WRMHS37H03	15062186	1236	15062201	1304	I	True	1369
WRMHS04M06	15062260	1293	15060961	1201	I	True	1299
WRMHS30H12	15062268	1341	15066152	1454	I	True	3884
WRMHS06O06	15062288	1325	15062752	1086	I	True	845

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS32A07	15062297	1465	15063582	1461	I	True	1285
WRMHS22P12	15062307	1083	15064039	1142	I	True	1732
WRMHS36I20	15062308	1506	15060618	1306	I	True	1690
WRMHS09F11	15062353	1064	15066062	1400	I	True	3709
WRMHS40B06	15062355	1423	15060464	1092	I	True	1891
WRMHS03G07	15062459	1443	15065589	1443	I	True	3130
WRMHS20B05	15062476	1182	15064976	1369	I	True	2500
WRMHS08K13	15062484	1238	15067903	1410	I	True	5419
WRMHS07C10	15062485	1376	15061259	1574	I	Gap	1226
WRMHS25E13	15062591	1347	15064131	1504	I	True	1540
WRMHS33P16	15062666	1014	15067826	1085	I	True	5160
WRMHS05I17	15062684	1441	15061785	1376	I	True	899
WRMHS10B20	15062698	983	15065968	1391	I	True	3270
WRMHS21G16	15062704	1419	15061091	1190	I	True	1613
WRMHS17N05	15062733	1419	15065707	1354	I	True	2974
WRMHS28M07	15062776	1349	15064252	1434	I	True	1476
WRMHS30M24	15062776	1371	15067954	1301	I	True	5178
WRMHS29O01	15062791	1323	15068154	1476	I	True	5363
WRMHS28D07	15062892	1306	15060495	1312	I	True	2397
WRMHS28P24	15062924	1242	15061856	1310	I	True	1068
WRMHS30P01	15062929	1496	15061508	1297	I	True	1421
WRMHS06F23	15062958	1293	15063933	1288	I	True	975
WRMHS26C16	15062968	931	15062259	1206	I	True	709
WRMHS27F15	15062990	1262	15065968	1400	I	True	2978
WRMHS30E04	15063007	1496	15063951	1448	I	True	944
WRMHS28D03	15063049	1256	15065214	1445	I	True	2165
WRMHS30I03	15063056	1544	15066615	1483	I	True	3559
WRMHS02J08	15063060	1314	15066492	1384	I	True	3432
WRMHS30L10	15063124	1037	15064617	1432	I	True	1493
WRMHS12P19	15063171	1271	15065636	1465	I	True	2465
WRMHS05P11	15063182	1223	15065518	1432	I	True	2336
WRMHS30G10	15063246	1447	15061358	1404	I	True	1888
WRMHS40H06	15063271	1395	15067733	830	I	True	4462
WRMHS26K05	15063274	708	15064030	1295	I	True	756
WRMHS27D24	15063317	1328	15061549	1312	I	True	1768
WRMHS39K05	15063331	534	15065056	1467	I	True	1725
WRMHS27E09	15063365	1496	15061153	1179	I	True	2212
WRMHS26J16	15063444	1347	15062135	1000	I	True	1309
WRMHS27N04	15063484	1432	15065621	1279	I	True	2137
WRMHS06A10	15063519	1242	15067298	1399	I	True	3779
WRMHS30N23	15063561	1378	15067280	1338	I	True	3719
WRMHS27H01	15063706	1465	15063737	1445	I	True	1551
WRMHS26N23	15063737	1321	15064451	1221	I	True	714
WRMHS28L02	15063756	1297	15064747	1456	I	True	991

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS23O07	15063832	1437	15065684	1461	I	True	1852
WRMHS29G22	15063892	1404	15061381	1116	I	True	2511
WRMHS11P20	15063911	1284	15061931	1238	I	True	1980
WRMHS26D16	15063912	1362	15066205	1214	I	True	2293
WRMHS40E23	15063919	1437	15062221	1315	I	True	1698
WRMHS26B03	15063933	1402	15065332	1395	I	True	1399
WRMHS06F04	15063966	1400	15065904	1469	I	True	1938
WRMHS30F06	15063971	1399	15065413	1404	I	True	1442
WRMHS21J14	15063974	1371	15067320	1358	I	True	3346
WRMHS21I04	15063992	1487	15065905	1352	I	True	1913
WRMHS27J19	15064016	1273	15062949	1072	I	True	1067
WRMHS26O24	15064102	1389	15063325	1158	I	True	777
WRMHS28N10	15064122	1266	15062087	1360	I	True	2035
WRMHS04O07	15064156	1000	15063402	1157	I	True	754
WRMHS29N08	15064156	1306	15064183	1195	I	True	1337
WRMHS28A01	15064170	1371	15063394	837	I	True	776
WRMHS19F17	15064199	1199	15062110	1254	I	True	2089
WRMHS06I06	15064206	1476	15060759	1389	I	True	3447
WRMHS40C08	15064217	1498	15061578	1216	I	True	2639
WRMHS02I11	15064260	1391	15066263	1251	I	True	2003
WRMHS16G02	15064397	1367	15061476	1423	I	True	2921
WRMHS30I09	15064406	1397	15061621	1456	I	True	2785
WRMHS37H20	15064431	1273	15067027	1254	I	True	2596
WRMHS14H10	15064469	1214	15066870	1254	I	True	2401
WRMHS30B12	15064489	616	15064833	621	I	True	344
WRMHS29L07	15064499	1260	15060606	1317	I	True	3893
WRMHS32C18	15064536	1495	15062676	1530	I	True	1860
WRMHS30P11	15064561	1480	15066768	1347	I	True	2207
WRMHS33O10	15064647	1535	15066902	1511	I	True	2255
WRMHS07J15	15064709	1280	15065888	1487	I	True	1179
WRMHS14F20	15064728	1328	15061123	1347	I	True	3605
WRMHS10N18	15064747	771	15064915	1319	I	True	961
WRMHS29N01	15064777	1406	15061414	1520	I	True	3363
WRMHS01C04	15064780	1404	15061444	1363	I	True	3336
WRMHS01F14	15064801	1027	15066635	1369	I	True	1834
WRMHS16B07	15064828	1574	15063469	1461	I	True	1359
WRMHS26B10	15064879	1458	15061119	1027	I	True	3760
WRMHS26A19	15064954	1541	15065468	1341	I	True	1055
WRMHS40O13	15064974	1519	15067905	1445	I	True	2931
WRMHS31P18	15065026	1561	15063948	1282	I	True	1078
WRMHS11F24	15065064	1090	15060337	1367	I	True	4727
WRMHS28G20	15065067	1458	15060739	1391	I	True	4328
WRMHS28P15	15065081	1151	15067415	1428	I	True	2334
WRMHS37M16	15065120	1334	15061057	1247	I	True	4063

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS13M08	15065135	1236	15066716	1371	I	True	1581
WRMHS28C08	15065148	1362	15065947	1434	I	True	799
WRMHS30G21	15065156	1177	15064334	1386	I	True	822
WRMHS26J02	15065169	1297	15062117	1158	I	True	3052
WRMHS02H05	15065216	1391	15063262	1146	I	True	1954
WRMHS28L11	15065223	1448	15063051	1371	I	True	2172
WRMHS24G23	15065225	1448	15062560	1371	I	True	2665
WRMHS29B13	15065243	1232	15067011	1537	I	True	1768
WRMHS36K21	15065270	1363	15060312	1445	I	True	4958
WRMHS28G16	15065320	1315	15068109	1391	I	True	2789
WRMHS18P05	15065335	1256	15063494	1323	I	True	1841
WRMHS36K15	15065361	1341	15063657	1435	I	True	1704
WRMHS04C08	15065367	1472	15065930	1240	I	True	915
WRMHS27E02	15065394	1581	15062484	1474	I	True	2910
WRMHS35E06	15065417	1445	15064514	1487	I	True	903
WRMHS38B06	15065429	1434	15060659	1277	I	True	4770
WRMHS03F17	15065432	693	15060597	1107	I	True	4835
WRMHS29A19	15065500	1419	15060648	1155	I	True	4852
WRMHS12J24	15065524	1271	15062222	1288	I	True	3302
WRMHS33E09	15065541	1314	15063974	1570	I	True	1567
WRMHS36A03	15065582	1417	15067634	1386	I	True	2052
WRMHS40F19	15065582	368	15067634	798	I	True	2052
WRMHS26B24	15065595	1345	15064123	1347	I	True	1472
WRMHS10M19	15065664	1485	15064561	1290	I	True	1103
WRMHS26O14	15065691	592	15064815	1216	I	True	876
WRMHS35F06	15065694	1314	15063566	1400	I	True	2128
WRMHS16I24	15065703	1352	15062844	1450	I	True	2859
WRMHS32I12	15065750	1349	15067905	1526	I	True	2155
WRMHS31L19	15065877	1483	15067767	1358	I	True	1890
WRMHS09N07	15065879	1267	15063035	1249	I	True	2844
WRMHS17I17	15065883	1406	15064704	1290	I	True	1179
WRMHS34I22	15065931	1338	15068027	1471	I	True	2096
WRMHS28N15	15065963	1367	15064630	1448	I	True	1333
WRMHS40A17	15065995	1482	15064346	527	I	True	1649
WRMHS39K22	15066041	1369	15067297	1284	I	True	1256
WRMHS35P07	15066064	1262	15065341	1391	I	True	726
WRMHS30J09	15066067	1548	15061078	1345	I	True	4989
WRMHS02P20	15066068	1267	15060624	1284	I	True	5444
WRMHS29B23	15066078	1238	15062353	1295	I	True	3725
WRMHS40P03	15066114	1447	15068023	1330	I	Gap	1909
WRMHS05H19	15066128	1380	15063017	1166	I	True	3111
WRMHS28H15	15066138	1399	15062245	1382	I	True	3893
WRMHS05E23	15066166	1487	15065051	1258	I	True	1115
WRMHS10M20	15066172	1461	15068175	1491	I	True	2003

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS26I12	15066199	1323	15060827	889	I	True	5372
WRMHS28D22	15066201	1279	15063848	1507	I	True	2353
WRMHS21F22	15066226	1356	15060752	1195	I	True	5474
WRMHS15J19	15066253	1448	15064767	1245	I	True	1486
WRMHS35D04	15066328	1229	15064463	965	I	True	1865
WRMHS26L05	15066354	1445	15066474	1467	I	True	1473
WRMHS26F18	15066434	1423	15065363	1005	I	True	1071
WRMHS29A09	15066455	1526	15061991	1221	I	True	4464
WRMHS28E23	15066475	1314	15065382	1269	I	True	1093
WRMHS28K16	15066497	1310	15065675	1506	I	True	822
WRMHS40M22	15066499	1448	15062484	1330	I	True	4015
WRMHS27A09	15066601	952	15061151	784	I	True	5450
WRMHS34C18	15066656	1635	15065090	1454	I	True	1566
WRMHS15B18	15066675	1480	15063612	1245	I	True	3063
WRMHS01E13	15066729	1496	15062602	1439	I	True	4127
WRMHS29M22	15066772	1369	15063229	1360	I	True	3543
WRMHS01E18	15066784	876	15067170	1275	I	True	803
WRMHS25L09	15066803	1352	15062759	1334	I	True	4044
WRMHS21D07	15066807	1445	15061351	1114	I	True	5456
WRMHS35P03	15066811	1395	15064312	1443	I	True	2499
WRMHS08C18	15066814	1430	15063457	1288	I	True	3357
WRMHS04H15	15066831	1188	15062096	1218	I	True	4735
WRMHS02G10	15066865	1378	15063491	1112	I	True	3374
WRMHS25N24	15066938	1365	15061220	959	I	True	5718
WRMHS39E08	15066940	573	15061887	1432	I	True	5053
WRMHS33B10	15066948	1602	15061090	1432	I	True	5858
WRMHS13C23	15066970	1328	15064082	1415	I	True	2888
WRMHS23C20	15067025	1415	15061901	1419	I	True	5124
WRMHS09L19	15067085	1236	15062923	1243	I	True	4162
WRMHS27K21	15067184	1535	15061075	1387	I	True	6109
WRMHS16M01	15067188	1591	15062584	1513	I	True	4604
WRMHS28F14	15067223	1197	15065344	1415	I	True	1879
WRMHS27G23	15067224	117	15061790	1332	I	True	5434
WRMHS04B23	15067226	1323	15061572	826	I	True	5654
WRMHS19I08	15067247	1406	15064845	1472	I	True	2402
WRMHS32H16	15067255	1260	15061781	1474	I	True	5474
WRMHS26B23	15067292	1051	15067869	1068	I	Gap	577
WRMHS07N21	15067327	1423	15065637	1397	I	True	1690
WRMHS27D20	15067347	1519	15063053	1035	I	True	4294
WRMHS26A14	15067359	1345	15064062	1140	I	True	3297
WRMHS03J10	15067362	1199	15066493	1251	I	True	869
WRMHS29J23	15067399	1304	15063001	1419	I	True	4398
WRMHS12J19	15067441	1408	15060302	1321	I	True	7139
WRMHS09B22	15067453	1384	15063049	1232	I	True	4404

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS29N21	15067458	1279	15060481	1269	I	True	6977
WRMHS21F14	15067469	1387	15060770	1133	I	True	6699
WRMHS13M20	15067497	1365	15066741	1437	I	True	759
WRMHS11P04	15067599	1304	15066554	1229	I	True	1045
WRMHS12N06	15067639	1410	15063315	1310	I	True	4324
WRMHS25F06	15067659	1280	15060618	1308	I	True	7041
WRMHS36C11	15067671	1583	15066214	1500	I	True	1457
WRMHS35E23	15067671	1290	15066214	1295	I	True	1457
WRMHS05N23	15067732	1463	15064847	1277	I	True	2885
WRMHS27N02	15067741	1489	15065216	1458	I	True	2525
WRMHS30H17	15067750	1474	15062063	1496	I	True	5687
WRMHS28B20	15067766	1406	15066878	1432	I	True	888
WRMHS09B10	15067770	1179	15063551	1419	I	True	4219
WRMHS09H05	15067840	994	15062436	1328	I	True	5404
WRMHS28H21	15067843	1397	15067064	1421	I	True	779
WRMHS30D14	15067942	1452	15065994	1437	I	True	1948
WRMHS29P17	15067979	1258	15064128	1277	I	True	3851
WRMHS12A07	15067980	1701	15061755	1413	I	True	6225
WRMHS27M03	15068007	1330	15061056	1445	I	True	6951
WRMHS26L22	15068106	1356	15060741	1168	I	True	7365
WRMHS29B24	15068175	1304	15065938	1397	I	True	2237
WRMHS28J04	15068187	1308	15063026	1415	I	True	5161
WRMHS27I21	15068240	1633	15065068	1334	I	True	3172
WRMHS34H11	15068250	1410	15060735	1426	I	True	7515
WRMHS24N14	15069208	1544	15062113	1221	I	True	7095
WRMHS27C11	1019832	285	1029825	1441	II	True	9993
WRMHS05C14	1583274	1031	1591824	1208	II	Gap	8550
WRMHS33F09	2423729	1389	2414808	372	II	Gap	8921
WRMHS02O24	5134323	479	5147172	497	II	True	12849
WRMHS27L18	8276628	1277	8289451	846	II	True	12823
WRMHS28E11	8288567	1061	8291822	1112	II	True	3255
WRMHS07C14	8288776	1096	8289207	785	II	True	710
WRMHS28L01	8288901	1031	8289806	1088	II	True	905
WRMHS33O01	8288929	1358	8292752	1177	II	True	3823
WRMHS27E07	8289025	1099	8289248	939	II	True	1010
WRMHS25H03	8289113	979	8288369	1109	II	True	744
WRMHS26J06	8289284	769	8290670	745	II	True	1386
WRMHS36N14	8289310	1062	8291003	872	II	True	1693
WRMHS27A18	8289491	1173	8292160	1109	II	True	2669
WRMHS09G21	8289661	950	8288776	1074	II	True	885
WRMHS05O04	8289710	721	8289156	1050	II	True	537
WRMHS20G06	8290294	619	8288413	970	II	True	1881
WRMHS28F02	8290361	1140	8288655	1125	II	True	1706
WRMHS27K17	8290922	1138	8290151	1086	II	Gap	771

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS14C19	8291254	1051	8291991	1101	II	True	737
WRMHS18A15	8291281	798	8302902	1417	II	True	11621
WRMHS36G15	8291578	1079	8291784	1014	II	True	1125
WRMHS32O02	8291616	1103	8292472	1096	II	True	856
WRMHS30J14	8291671	1072	8292436	1219	II	True	765
WRMHS25J22	8291746	928	8289085	979	II	True	2661
WRMHS14J07	8291917	870	8292717	1072	II	True	800
WRMHS30N02	8292125	1251	8289547	1136	II	True	2578
WRMHS27N01	8292422	1127	8288850	1166	II	True	3572
WRMHS26A22	8292563	1351	8285140	1363	II	True	7423
WRMHS25N02	9810445	204	9807784	667	II	True	2661
WRMHS29P08	9815520	809	9805447	1179	II	True	10073
WRMHS04D12	9819575	987	9830375	1184	II	True	10800
WRMHS04L18	12853740	1373	12845411	948	II	True	8329
WRMHS14C24	12859687	1000	12845434	917	II	True	14253
WRMHS09N15	13999151	1044	13985246	747	II	Gap	13905
WRMHS07G06	925560	1325	931569	693	III	Gap	6009
WRMHS12B22	941198	414	940921	414	III	True	277
WRMHS27D15	1016830	743	1019970	1081	III	Gap	3140
WRMHS29O17	1019605	1007	1018505	1094	III	True	1100
WRMHS28F16	3334340	935	3335287	1275	III	True	947
WRMHS10F07	3334408	496	3336707	1044	III	True	2299
WRMHS30M08	3334497	627	3334925	638	III	True	428
WRMHS29O06	3334560	881	3336331	961	III	True	1771
WRMHS30O15	3334570	1144	3336558	1262	III	True	1988
WRMHS26D19	3334698	933	3335264	817	III	True	544
WRMHS10J06	3334754	789	3336423	966	III	True	1669
WRMHS25N16	3335314	826	3334422	941	III	True	892
WRMHS01J10	3335399	717	3335912	809	III	Gap	513
WRMHS27O10	3335423	1103	3336189	1133	III	True	766
WRMHS01A22	3335506	950	3336162	1024	III	True	656
WRMHS28C02	3335989	929	3321012	1376	III	True	14977
WRMHS28B15	3336357	1079	3335532	1107	III	True	825
WRMHS28I09	5345583	1175	5355089	1011	III	True	9506
WRMHS28I06	5349346	1314	5355537	1230	III	True	6191
WRMHS04O13	7403795	1168	7411651	1035	III	True	7856
WRMHS18A02	7408508	1151	7400640	1273	III	Gap	7868
WRMHS27G11	10224629	1107	10225367	1070	III	True	738
WRMHS29I09	10224659	965	10225435	1146	III	True	776
WRMHS26K12	10224688	924	10225183	782	III	True	587
WRMHS29J15	10224699	1055	10225403	592	III	True	704
WRMHS05D22	10224749	1014	10227299	1147	III	True	2550
WRMHS26M08	10224752	1194	10237144	1160	III	True	12392
WRMHS30J24	10224757	1192	10225356	1212	III	True	856

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS28B02	10224817	737	10225320	1072	III	True	503
WRMHS09O20	10224839	641	10227078	1037	III	True	2239
WRMHS25E02	10224910	739	10225361	1153	III	True	766
WRMHS10F12	10225056	880	10225390	1098	III	True	937
WRMHS26O09	10225090	501	10225546	913	III	True	456
WRMHS26M09	10225345	477	10224706	819	III	True	639
WRMHS29F21	10225360	1055	10224651	1157	III	True	675
WRMHS07B13	10225370	1062	10224616	1225	III	True	660
WRMHS28D21	10225380	1018	10224677	1323	III	True	743
WRMHS27B11	10225384	1199	10224617	1122	III	True	767
WRMHS29E10	10225386	928	10224536	1223	III	Gap	850
WRMHS28K14	10225415	1085	10224665	1208	III	True	730
WRMHS31J24	10225471	1275	10224701	1053	III	True	770
WRMHS27P12	10225489	1188	10224687	1037	III	True	802
WRMHS30D01	10225594	1399	10224620	1230	III	True	974
WRMHS28F18	10225998	983	10224766	1155	III	True	1232
WRMHS23P03	13586681	274	13571871	1397	III	True	14810
WRMHS30G17	2828801	1110	2829975	1253	IV	True	1174
WRMHS03O06	2828877	856	2829851	1068	IV	True	974
WRMHS27H16	2829261	1182	2829951	592	IV	True	445
WRMHS27G05	2829327	1273	2830520	1234	IV	True	1193
WRMHS26K14	2829453	989	2829851	961	IV	True	856
WRMHS05K24	2829819	928	2828909	1037	IV	True	910
WRMHS29L16	2829960	813	2830504	1005	IV	True	544
WRMHS26K13	2829977	850	2829339	1031	IV	True	638
WRMHS29B20	2830085	1014	2828859	1081	IV	True	1226
WRMHS26G19	2830158	1208	2828804	422	IV	True	1354
WRMHS25P16	2830263	861	2828950	979	IV	True	1313
WRMHS28O12	2830276	1037	2828912	1024	IV	True	1364
WRMHS29O14	2830528	1035	2828800	1105	IV	Gap	1728
WRMHS30E10	2830537	1216	2829425	1101	IV	True	1112
WRMHS07B12	3203573	453	3208551	1245	IV	Gap	4978
WRMHS27F22	3207894	920	3211817	1116	IV	True	3923
WRMHS35E01	3212543	1003	3207639	976	IV	True	4904
WRMHS37N07	3213005	1118	3203655	1040	IV	True	9350
WRMHS19L11	3218674	1212	3207817	1055	IV	True	10857
WRMHS22J02	4406770	826	4419436	839	IV	True	12666
WRMHS27C01	4416454	1417	4417388	1393	IV	True	934
WRMHS09I08	4416462	1059	4420206	1133	IV	True	3744
WRMHS27J09	4416760	1351	4427259	1014	IV	True	10499
WRMHS29G24	4417519	1037	4419859	1005	IV	True	2340
WRMHS11D07	4418883	830	4427899	1491	IV	True	9016
WRMHS30P10	4419952	1245	4417825	1160	IV	True	2127
WRMHS28I24	4426473	1218	4419564	1173	IV	True	6909

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS26C03	4427142	787	4417746	481	IV	True	9396
WRMHS13B24	4432064	1138	4419291	1090	IV	True	12773
WRMHS33O12	4432392	1483	4417858	1284	IV	True	14534
WRMHS26G01	6676544	497	6683317	795	IV	Gap	6773
WRMHS11N12	8566672	1092	8577448	555	IV	True	10776
WRMHS01F24	8577296	695	8580973	867	IV	True	3677
WRMHS26G18	8577318	785	8583057	508	IV	True	5739
WRMHS03J11	8577977	652	8575493	979	IV	True	2484
WRMHS23N18	8579689	752	8589359	1197	IV	True	9670
WRMHS07A08	8581318	861	8575518	941	IV	True	5800
WRMHS01C05	8586990	1301	8575670	1033	IV	True	11320
WRMHS01C11	9046016	1077	9047779	983	IV	True	1763
WRMHS05D04	9046948	880	9047686	985	IV	True	738
WRMHS30D17	9046949	1138	9052865	1443	IV	True	5916
WRMHS30N22	9046961	1190	9047703	1144	IV	True	791
WRMHS09H13	9047501	835	9045921	900	IV	True	1580
WRMHS40N10	9047702	1122	9035931	950	IV	True	11771
WRMHS12K21	9047742	920	9047089	1040	IV	True	624
WRMHS27B15	9047779	1086	9045999	1214	IV	True	1780
WRMHS26G04	11072934	734	11080781	1236	IV	True	7847
WRMHS28L12	11074244	852	11075679	695	IV	True	1435
WRMHS08M24	11074298	433	11072764	905	IV	True	1534
WRMHS29H21	11074300	418	11072060	905	IV	True	2240
WRMHS27I17	11074691	909	11072805	1005	IV	True	1886
WRMHS07B04	11074770	811	11072972	695	IV	True	1798
WRMHS09L16	11074901	926	11072945	542	IV	True	1956
WRMHS20D20	11074906	941	11072516	739	IV	True	2390
WRMHS29K23	11075476	483	11072807	662	IV	True	2669
WRMHS27J05	11075592	1007	11072509	606	IV	True	3083
WRMHS11O15	11078091	665	11072505	351	IV	True	5586
WRMHS21P14	11081059	1533	11072938	741	IV	True	8121
WRMHS09K06	12161244	1038	12170138	1225	IV	True	8894
WRMHS12K16	12322053	1284	12309209	1360	IV	True	12844
WRMHS10F23	12731613	856	12730795	811	IV	True	818
WRMHS11B11	13361904	894	13359934	1306	IV	Gap	1970
WRMHS26N21	13549405	180	13549507	180	IV	True	102
WRMHS27E03	13549405	174	13549507	174	IV	True	102
WRMHS28D12	13549405	180	13549507	180	IV	True	102
WRMHS30G19	13549405	180	13549507	180	IV	True	102
WRMHS26C08	13549405	178	13549507	180	IV	True	102
WRMHS28K11	13549507	180	13549405	180	IV	True	102
WRMHS30F05	13549507	180	13549405	180	IV	True	102
WRMHS05D17	16963438	534	16976159	843	IV	True	12721
WRMHS08F14	16963446	843	16973114	1251	IV	True	9668

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS16A17	16976924	1382	16963095	1127	IV	Gap	13829
WRMHS35N22	253541	1303	265927	1168	V	True	12386
WRMHS27M21	265076	1338	267244	1304	V	True	2168
WRMHS12K18	265199	1168	266803	1062	V	True	1604
WRMHS26I18	265230	1099	267148	708	V	True	1918
WRMHS30K17	265580	1240	266363	1083	V	True	830
WRMHS27I09	265652	909	266498	449	V	True	846
WRMHS08P13	265690	1068	267008	1142	V	True	1318
WRMHS11A23	265725	846	266240	1279	V	True	859
WRMHS28K18	265846	1101	266288	1199	V	True	1017
WRMHS05J21	266023	922	266808	1053	V	True	785
WRMHS26M20	266075	1284	266870	924	V	True	603
WRMHS30D22	266091	1234	267134	1199	V	True	1043
WRMHS01C16	266133	1158	274687	1245	V	True	8554
WRMHS25K21	266170	869	266439	1158	V	True	1014
WRMHS28P01	266171	1127	266805	955	V	True	720
WRMHS29O09	266181	1057	267066	1088	V	True	885
WRMHS26L11	266204	1136	266560	632	V	True	764
WRMHS11I07	266302	1053	266827	1151	V	True	880
WRMHS30P22	266406	1079	266803	1086	V	True	1015
WRMHS29A22	266462	1123	265591	1175	V	True	871
WRMHS04K15	266579	963	267228	918	V	True	649
WRMHS30J04	266761	595	266897	1118	V	True	976
WRMHS28I08	266764	1011	266104	1042	V	True	660
WRMHS06F24	266766	893	265730	1035	V	True	1036
WRMHS29D20	266776	1107	265308	1086	V	True	1468
WRMHS27G24	266806	1140	265881	1081	V	True	925
WRMHS27M06	266827	1096	266118	1114	V	True	691
WRMHS28H13	266848	1022	266163	1144	V	True	660
WRMHS30C13	266867	1109	266111	1212	V	True	810
WRMHS26M16	266911	1155	265920	850	V	True	991
WRMHS12G17	267007	1358	266067	1136	V	True	940
WRMHS08G22	267191	957	266362	1155	V	True	829
WRMHS29I08	267234	977	266380	1206	V	True	854
WRMHS26P10	1103175	444	1105054	531	V	True	1879
WRMHS26B05	1104561	669	1105573	1142	V	True	1012
WRMHS29D07	1104563	621	1104954	636	V	True	395
WRMHS04F04	1104592	678	1105048	702	V	True	456
WRMHS05E04	1104632	610	1105054	747	V	True	422
WRMHS04K11	1104701	420	1105034	697	V	True	333
WRMHS03G15	1104736	424	1105053	773	V	True	317
WRMHS27A23	1104921	1254	1102436	828	V	True	2485
WRMHS26B13	1105022	715	1104561	743	V	True	476
WRMHS04M07	1107008	966	1104565	730	V	True	2443

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS03E19	1112848	521	1104561	712	V	True	8287
WRMHS16D06	2216373	1496	2205888	1321	V	True	10485
WRMHS29A06	3095779	1155	3108191	1166	V	True	12412
WRMHS08P06	3099004	294	3108419	1249	V	True	9415
WRMHS04H16	3101488	1199	3092204	894	V	True	9284
WRMHS29H06	5074561	795	5073913	1177	V	True	573
WRMHS39I18	5269669	1489	5284321	898	V	True	14652
WRMHS31J16	5272798	1465	5284092	763	V	True	11294
WRMHS28L06	5282864	865	5285622	1009	V	True	2758
WRMHS27O21	5283316	1325	5281267	1266	V	True	2049
WRMHS30M09	5283600	1127	5291013	1443	V	True	7413
WRMHS09L01	5283833	588	5287702	1236	V	True	3869
WRMHS30B07	5283834	1101	5284540	1077	V	True	796
WRMHS27B23	5283892	1066	5290680	1271	V	True	6788
WRMHS01E17	5283932	859	5296939	1099	V	True	13007
WRMHS12F15	5284471	918	5270602	1358	V	True	13869
WRMHS27K09	5284892	878	5286618	965	V	Gap	1726
WRMHS09I15	5285715	182	5274840	1157	V	True	10875
WRMHS04J23	6174365	935	6183415	1199	V	True	9050
WRMHS26N19	6181556	1203	6174319	821	V	True	7237
WRMHS04D03	6186920	1291	6174349	942	V	True	12571
WRMHS29M05	6189114	896	6174357	972	V	True	14757
WRMHS07G19	6190135	1454	6175519	833	V	True	14616
WRMHS25H15	6939200	972	6939527	1014	V	True	802
WRMHS27E18	6939473	1072	6936934	1131	V	True	2539
WRMHS09F01	6939527	970	6938067	1253	V	True	1460
WRMHS15N24	10595141	1417	10608706	1075	V	True	13565
WRMHS05N15	10602831	1027	10608692	1055	V	True	5861
WRMHS28C15	10604666	1363	10607655	1232	V	True	2989
WRMHS27N12	10605999	235	10605873	265	V	True	142
WRMHS04B17	10606652	719	10615167	512	V	Gap	8515
WRMHS28E20	10606751	1267	10607691	1315	V	True	940
WRMHS25M08	10606759	1118	10609591	1166	V	True	2832
WRMHS25O06	10607273	835	10607410	1218	V	True	1083
WRMHS29J12	10614200	1101	10606800	1184	V	True	7400
WRMHS26M07	10619057	1495	10605996	1219	V	True	13061
WRMHS17F06	10620243	1445	10608031	891	V	True	12212
WRMHS37J02	14870620	1358	14880314	523	V	True	9694
WRMHS38N21	14873486	1543	14879183	1251	V	True	5697
WRMHS37G01	14880476	1382	14883133	1291	V	True	2657
WRMHS06F14	14881987	1402	14879026	1434	V	True	2961
WRMHS08C07	14882506	1266	14892224	1417	V	True	9718
WRMHS16J09	14891515	785	14883152	1216	V	True	8363
WRMHS23A01	17120467	309	17122924	926	V	Gap	2457

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS01E09	17122329	1369	17123879	1286	V	True	1550
WRMHS37I14	17122414	1064	17124620	1171	V	True	2206
WRMHS36B04	17122586	1247	17123693	1197	V	True	1107
WRMHS15J03	17122616	1260	17123319	1206	V	True	703
WRMHS08N08	17122891	1188	17123155	1164	V	True	1041
WRMHS35F02	17123231	1175	17124343	1205	V	True	1112
WRMHS30J23	17123348	1378	17124302	1310	V	True	954
WRMHS32K09	17123592	1430	17122467	1421	V	True	1125
WRMHS01B16	17123722	863	17123226	761	V	True	496
WRMHS36F10	17123873	959	17123320	1042	V	True	544
WRMHS03M12	17123899	1149	17123121	1334	V	True	778
WRMHS14G05	17124020	1122	17122460	1194	V	True	1560
WRMHS15L11	17124526	1362	17122470	1262	V	True	2056
WRMHS02D14	17130274	1254	17137195	1286	V	Gap	6921
WRMHS25A23	18242128	935	18252831	440	V	True	10703
WRMHS07M16	102328	645	114186	1005	X	True	11858
WRMHS30A05	109734	143	114000	1096	X	True	4266
WRMHS26A11	109861	1040	111216	455	X	True	1355
WRMHS28K02	110151	470	110419	475	X	True	268
WRMHS08B22	1105719	1086	113657	966	X	True	3078
WRMHS29E03	110767	929	113859	952	X	True	3092
WRMHS27D12	110794	1147	107537	989	X	True	3257
WRMHS27N06	110852	987	114048	874	X	True	3196
WRMHS10L04	110928	630	113550	872	X	True	2622
WRMHS12J05	111462	800	110997	800	X	True	465
WRMHS01I16	111623	658	112947	1026	X	Gap	1324
WRMHS05I06	111940	651	112447	671	X	True	507
WRMHS29K04	112256	902	113118	1171	X	True	862
WRMHS29M01	112319	555	113377	1205	X	True	1058
WRMHS04B20	112398	1271	113559	911	X	True	1161
WRMHS28M13	112657	1099	114038	1166	X	True	1381
WRMHS01A17	112792	1075	113834	1210	X	True	1042
WRMHS29O11	112803	765	113973	911	X	True	1170
WRMHS28J03	112828	1090	112871	1064	X	True	1298
WRMHS06L24	113259	1221	113964	1118	X	True	681
WRMHS08N07	113368	549	112979	381	X	True	389
WRMHS29L01	113444	774	112769	784	X	True	675
WRMHS28K06	113449	904	114211	826	X	True	762
WRMHS28H10	113453	905	114233	885	X	True	780
WRMHS30D18	113784	946	111645	1085	X	True	2139
WRMHS01M01	113793	1218	110533	979	X	True	3260
WRMHS27K02	113930	959	113229	1013	X	True	720
WRMHS01C18	113948	625	110295	1003	X	True	3653
WRMHS29P15	113968	913	112754	885	X	True	1214

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS26F06	113973	586	113082	540	X	True	891
WRMHS27M17	113973	1044	111665	1227	X	True	2308
WRMHS26N06	114103	878	113509	824	X	True	594
WRMHS30M16	114160	1059	112603	1085	X	True	1557
WRMHS30L09	114223	918	113223	931	X	True	1000
WRMHS30C10	114223	1214	112067	987	X	True	2156
WRMHS11O04	114226	959	112765	782	X	True	1461
WRMHS28K21	118493	1214	113432	1327	X	True	5061
WRMHS28E22	290240	1173	294505	1229	X	Gap	4265
WRMHS30D15	292764	1158	294084	1426	X	True	1320
WRMHS27J21	292766	1310	293879	1334	X	True	1113
WRMHS09J03	293000	1059	294590	1177	X	True	1590
WRMHS35A02	293024	1194	294506	1253	X	True	1482
WRMHS29K11	293178	1092	293567	1295	X	True	957
WRMHS02P02	293483	435	292734	477	X	True	749
WRMHS30H02	293671	1400	294595	1175	X	True	924
WRMHS29C06	293899	1229	295191	1338	X	True	1292
WRMHS05K20	293931	1105	294828	809	X	True	897
WRMHS09D15	294013	1011	294792	1236	X	True	779
WRMHS05E06	294038	963	294750	1151	X	True	712
WRMHS29L22	294121	1110	293423	966	X	True	698
WRMHS30H10	294175	1358	305320	1456	X	True	11145
WRMHS09L21	294245	961	294823	1155	X	True	578
WRMHS30N21	294484	1114	293076	1260	X	True	1408
WRMHS12P18	294603	1195	293111	1358	X	True	1492
WRMHS29D10	294861	1040	292074	1325	X	True	2787
WRMHS30C16	1631699	1312	1636850	1201	X	True	5151
WRMHS28L23	1633167	1142	1637467	1249	X	True	4300
WRMHS26G12	1633184	1273	1637226	427	X	True	4042
WRMHS07B17	1634659	1050	1637127	1190	X	True	2468
WRMHS28A02	1636522	968	1637425	1122	X	True	903
WRMHS10D11	1636612	893	1646541	1380	X	True	9929
WRMHS30E03	1636704	1175	1636917	1158	X	True	1171
WRMHS29E20	1636965	846	1641608	1367	X	True	4643
WRMHS01P07	1637308	1005	1624757	1310	X	True	12551
WRMHS29C20	1637471	1062	1634662	1291	X	True	2809
WRMHS16J06	1637483	990	1628890	1454	X	True	8593
WRMHS05I19	1638044	795	1645055	1155	X	True	7011
WRMHS26C06	1638394	1347	1640095	1391	X	True	1701
WRMHS29J19	1638719	874	1636673	1085	X	True	2046
WRMHS39G13	1654843	1382	1641638	1170	X	True	13205
WRMHS25I03	3484059	1371	3481743	739	X	True	2316
WRMHS12J20	5056582	1295	5067254	1356	X	True	10672
WRMHS29B01	7069760	1295	7078393	1009	X	True	8633

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS26C22	7076564	1476	7079148	928	X	True	2584
WRMHS26L18	7077464	612	7077813	339	X	True	331
WRMHS25H06	7077473	466	7079038	747	X	True	1565
WRMHS30I08	7077475	699	7079185	1064	X	True	1710
WRMHS28E18	7077476	802	7078541	952	X	True	1065
WRMHS28O18	7077476	715	7079029	952	X	True	1553
WRMHS30E07	7077476	854	7079109	713	X	True	1633
WRMHS30D20	7077477	660	7079153	942	X	True	1676
WRMHS29H12	7077488	361	7078649	804	X	True	1161
WRMHS26I02	7077513	446	7078944	436	X	True	1431
WRMHS08M22	7077516	385	7078473	617	X	True	957
WRMHS26O12	7077516	420	7079116	726	X	True	1600
WRMHS05N10	7077544	460	7079185	758	X	True	1641
WRMHS29A08	7077549	307	7079145	974	X	True	1596
WRMHS26I09	7077556	643	7078349	715	X	True	793
WRMHS09L13	7077561	691	7078468	922	X	True	907
WRMHS05C19	7077563	575	7078524	734	X	True	961
WRMHS30O12	7077563	876	7078621	1016	X	True	1058
WRMHS30H01	7077567	887	7079045	961	X	True	1478
WRMHS08F12	7077568	449	7078094	592	X	True	526
WRMHS28P18	7077568	725	7079161	1026	X	True	1593
WRMHS29D02	7077570	621	7078286	872	X	True	716
WRMHS07J02	7077576	486	7079216	896	X	True	1640
WRMHS05E14	7077584	464	7078613	769	X	True	1029
WRMHS27G20	7077586	813	7079208	970	X	True	1622
WRMHS30K11	7077597	414	7078380	918	X	True	783
WRMHS02N16	7077616	625	7078655	651	X	True	1039
WRMHS09A06	7077616	375	7079129	837	X	True	1513
WRMHS29G21	7077620	789	7078120	854	X	True	749
WRMHS28M11	7077621	640	7078384	905	X	True	763
WRMHS12C22	7077623	481	7079029	446	X	True	1406
WRMHS03G08	7077629	372	7078228	388	X	True	599
WRMHS28C12	7077634	791	7079201	1040	X	True	1567
WRMHS30C05	7077644	907	7079029	929	X	True	1385
WRMHS27N03	7077676	878	7078285	876	X	True	698
WRMHS29O04	7077689	682	7079090	959	X	True	1401
WRMHS30K23	7077689	1040	7079121	981	X	True	1432
WRMHS07K01	7077691	850	7078843	972	X	True	1152
WRMHS03A08	7077694	815	7078769	656	X	True	1075
WRMHS27D17	7077700	652	7078286	843	X	True	586
WRMHS28L24	7077710	717	7078309	922	X	True	599
WRMHS33E15	7077734	784	7078275	1003	X	True	858
WRMHS28M12	7077817	761	7079161	974	X	True	1344
WRMHS29O21	7077841	898	7078908	1011	X	True	1067

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS27C06	7077844	902	7078545	907	X	True	701
WRMHS27P08	7077859	704	7078092	878	X	True	908
WRMHS29I23	7077890	867	7079062	1055	X	True	1172
WRMHS06H02	7077952	845	7078523	905	X	True	773
WRMHS05M24	7077952	654	7079209	876	X	True	1257
WRMHS26O13	7077991	374	7078265	379	X	True	274
WRMHS27D14	7078014	676	7079183	773	X	True	1169
WRMHS21D06	7078033	523	7083641	1042	X	True	5608
WRMHS25M03	7078072	839	7078368	1098	X	True	1177
WRMHS30N11	7078114	843	7078973	1048	X	True	859
WRMHS06L08	7078136	776	7079024	739	X	True	888
WRMHS30H03	7078145	721	7081590	981	X	True	3445
WRMHS27J12	7078155	315	7078375	326	X	True	220
WRMHS28H14	7078172	756	7079205	937	X	True	1033
WRMHS26H17	7078184	832	7079062	560	X	True	878
WRMHS29F01	7078191	822	7079048	1007	X	True	857
WRMHS30E09	7078192	848	7078206	909	X	True	1358
WRMHS29O10	7078207	636	7079130	776	X	True	923
WRMHS07O14	7078212	728	7078341	1059	X	True	1221
WRMHS40G24	7078215	763	7077548	399	X	True	667
WRMHS30N20	7078226	464	7078364	989	X	True	986
WRMHS29G13	7078241	863	7078372	865	X	True	1216
WRMHS29B03	7078268	612	7078448	833	X	True	907
WRMHS04G13	7078293	571	7077532	401	X	True	761
WRMHS30I20	7078331	998	7079183	1018	X	Gap	852
WRMHS29G09	7078358	662	7079173	981	X	True	815
WRMHS30A10	7078361	1138	7077476	1009	X	True	885
WRMHS27F21	7078375	955	7077612	712	X	True	763
WRMHS29C23	7078380	857	7077488	795	X	True	892
WRMHS26G15	7078386	913	7077536	470	X	True	850
WRMHS29C13	7078393	939	7079216	1000	X	True	823
WRMHS29B09	7078395	857	7079021	977	X	True	755
WRMHS28N11	7078420	737	7079162	981	X	True	742
WRMHS27H04	7078452	833	7078769	883	X	True	1032
WRMHS28D16	7078456	885	7079084	902	X	True	628
WRMHS08C09	7078486	368	7078785	793	X	True	299
WRMHS29C09	7078490	891	7078689	972	X	True	1196
WRMHS30H04	7078494	758	7079121	926	X	True	627
WRMHS27J02	7078507	926	7079109	863	X	True	685
WRMHS27C12	7078552	902	7078955	728	X	True	791
WRMHS30F14	7078572	512	7078689	917	X	True	1089
WRMHS28P02	7078599	688	7079066	946	X	True	780
WRMHS08I13	7078630	828	7077605	388	X	True	1025
WRMHS28A22	7078653	721	7078985	1050	X	True	989

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS26B17	7078656	551	7079046	556	X	True	390
WRMHS29G07	7078690	468	7079136	939	X	True	446
WRMHS27L20	7078729	970	7077756	761	X	True	973
WRMHS08G14	7078759	494	7079105	479	X	True	404
WRMHS25K14	7078801	787	7077680	778	X	True	1121
WRMHS27C15	7078809	952	7077532	909	X	True	1277
WRMHS03D17	7078811	725	7077490	865	X	True	1321
WRMHS26I07	7078832	436	7077625	669	X	True	1207
WRMHS29H03	7078833	361	7079090	381	X	True	257
WRMHS26P13	7078852	424	7079213	564	X	True	361
WRMHS08O08	7078856	593	7077986	472	X	True	870
WRMHS05H04	7078862	375	7078591	387	X	True	271
WRMHS28L22	7078896	793	7077823	730	X	True	1073
WRMHS26B18	7078912	952	7077584	472	X	True	1328
WRMHS30B15	7078920	1085	7078001	959	X	True	919
WRMHS29B02	7078941	296	7077952	778	X	True	989
WRMHS27L02	7078951	946	7077584	712	X	True	1367
WRMHS27C09	7078955	924	7078012	728	X	True	943
WRMHS30F24	7078955	1026	7077858	891	X	True	1097
WRMHS04H18	7078973	752	7077487	448	X	True	1486
WRMHS10B15	7079049	872	7077584	645	X	True	1465
WRMHS26B20	7079051	1099	7077984	804	X	True	1067
WRMHS19B04	7079063	813	7078506	756	X	True	557
WRMHS26J14	7079084	756	7077616	327	X	True	1468
WRMHS26K15	7079097	1079	7077730	802	X	True	1367
WRMHS29A04	7079102	351	7077700	405	X	True	1402
WRMHS11P23	7079105	926	7078353	647	X	True	752
WRMHS29H05	7079108	518	7078012	765	X	True	1096
WRMHS28J02	7079111	747	7077476	616	X	True	1635
WRMHS28J24	7079136	710	7077570	523	X	True	1566
WRMHS29I20	7079141	809	7078192	865	X	True	949
WRMHS09F13	7079185	664	7078633	691	X	Gap	552
WRMHS30P17	7079185	977	7078274	922	X	True	911
WRMHS27F20	7079195	902	7078141	549	X	True	1054
WRMHS10K03	7079197	745	7075583	1426	X	True	3614
WRMHS28J16	7079216	918	7078450	972	X	True	766
WRMHS34G24	7082353	1330	7077585	641	X	True	4768
WRMHS28K13	7082823	1096	7077535	957	X	True	5288
WRMHS27G08	7086042	555	7077680	148	X	True	8362
WRMHS35M05	7092173	804	7077556	220	X	True	14617
WRMHS08N23	9177559	1330	9186961	715	X	True	9402
WRMHS29G23	9179382	1378	9186261	1162	X	True	6879
WRMHS12P17	9180690	1256	9188463	667	X	True	7773
WRMHS28C16	9185102	1011	9186693	1166	X	True	1591

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS28B07	9185512	1083	9186525	948	X	True	1013
WRMHS10C11	9185870	1158	9199097	1424	X	True	13227
WRMHS29E07	9186487	950	9175909	1517	X	True	10578
WRMHS29D22	9186525	941	9176725	1341	X	True	9800
WRMHS05G04	9186832	924	9180904	1312	X	True	5928
WRMHS28C01	9186867	1103	9185096	1367	X	True	1771
WRMHS27G14	9186893	1066	9179506	1136	X	True	7387
WRMHS05M07	9186900	782	9186035	1018	X	True	865
WRMHS03M15	9188495	603	9185889	1009	X	True	2606
WRMHS26D21	9188521	1214	9186670	959	X	True	1851
WRMHS09A21	9188542	802	9186270	1048	X	True	2272
WRMHS05I02	9188597	547	9174816	1208	X	True	13781
WRMHS21E21	9189786	963	9187945	992	X	True	1841
WRMHS10K08	9201031	1219	9186055	1074	X	True	14976
WRMHS30P08	11440660	1112	11443423	1053	X	True	2763
WRMHS26M06	11442711	556	11443120	571	X	True	409
WRMHS26N05	11442776	1079	11449635	1118	X	True	6859
WRMHS27M20	11442991	983	11446910	1247	X	True	3919
WRMHS28N09	11443569	1112	11433821	1421	X	True	9748
WRMHS27E10	11449622	1321	11442968	832	X	True	6654
WRMHS29M23	11774695	1338	11787198	453	X	True	12503
WRMHS09A01	11782637	1410	11786833	1037	X	True	4196
WRMHS30P16	11784130	1393	11787157	1188	X	True	3027
WRMHS07E13	11786499	1096	11789170	728	X	True	2671
WRMHS01L16	11786520	1099	11788900	894	X	True	2380
WRMHS01P05	11786703	1090	11789146	992	X	True	2443
WRMHS29J20	11786906	636	11781289	1351	X	True	5617
WRMHS07H23	11787322	1136	11774354	1168	X	True	12968
WRMHS29N07	11788323	924	11792461	1088	X	True	4138
WRMHS25G19	11788999	1038	11788146	1160	X	True	853
WRMHS30B08	11789080	1273	11786155	1312	X	True	2925
WRMHS10K13	11789205	1149	11786492	1256	X	True	2713
WRMHS28F06	11789471	981	11788463	985	X	True	1008
WRMHS10I16	11880024	1149	11888107	1136	X	True	8083
WRMHS26L17	11887402	1079	11887851	682	X	True	702
WRMHS27N08	11887468	891	11887914	992	X	True	722
WRMHS25I19	11887766	455	11879940	1376	X	True	7826
WRMHS01B04	12277629	549	12278100	385	X	True	334
WRMHS30L22	14385908	1339	14394760	1354	X	True	8852
WRMHS11O08	14897726	1133	14909123	736	X	True	11397
WRMHS28G08	14908076	880	14908848	1057	X	True	772
WRMHS29F22	14908470	1020	14908746	898	X	True	1100
WRMHS04P12	14909340	896	14908095	905	X	True	1245
WRMHS29J24	14909764	828	14908255	595	X	True	1509

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS26L19	14910973	1365	14908120	769	X	True	2853
WRMHS29L21	14914609	1223	14908208	1027	X	True	6401
WRMHS29O12	15807564	737	15810300	955	X	True	2736
WRMHS28P11	15807607	898	15808639	1018	X	True	1032
WRMHS20J18	15807671	717	15809565	695	X	True	1894
WRMHS30F07	15807906	791	15808465	785	X	True	559
WRMHS12P15	15808075	848	15809972	1210	X	True	1897
WRMHS27B04	15808153	959	15811087	767	X	True	2934
WRMHS27O08	15808220	641	15808611	647	X	True	391
WRMHS30N19	15808253	1083	15808960	1007	X	True	707
WRMHS30B06	15808303	1055	15811090	994	X	True	2787
WRMHS30B14	15808324	1094	15808726	1114	X	True	1008
WRMHS30B16	15808342	390	15811026	433	X	True	2684
WRMHS06A16	15808379	1066	15808676	1134	X	True	1159
WRMHS29F07	15808410	691	15810946	990	X	True	2536
WRMHS27H13	15808547	1016	15807557	721	X	True	990
WRMHS10C24	15808583	1099	15808986	1216	X	True	1086
WRMHS29L23	15808646	654	15807885	1053	X	True	761
WRMHS28O21	15808687	1075	15807772	874	X	True	915
WRMHS01C20	15808731	361	15808296	752	X	True	255
WRMHS12B18	15808946	994	15807952	994	X	True	994
WRMHS02M08	15808997	898	15807660	900	X	True	1337
WRMHS28L03	15808999	859	15808288	1000	X	True	711
WRMHS27B16	15809021	920	15808310	808	X	True	711
WRMHS28H18	15809124	1040	15808361	1177	X	True	763
WRMHS28C21	15809139	1042	15808126	1127	X	True	1013
WRMHS28N05	15809145	968	15808429	1048	X	True	716
WRMHS30G24	15809527	983	15807574	992	X	True	1953
WRMHS26M23	15809571	1267	15807938	1110	X	True	1633
WRMHS27K10	15809712	833	15808566	987	X	True	1146
WRMHS30O11	15810128	1240	15809166	1282	X	True	962
WRMHS19N20	15810499	702	15811029	937	X	True	530
WRMHS26B12	15810687	477	15810974	472	X	True	287
WRMHS28E02	15810753	424	15811035	440	X	True	282
WRMHS30L07	15810898	1035	15809661	1199	X	True	1237
WRMHS05L08	15810913	682	15808292	739	X	True	2621
WRMHS29K20	15811012	603	15808488	684	X	True	2524
WRMHS30F01	15811030	1048	15810266	1050	X	True	764
WRMHS28M19	15811040	215	15810909	231	X	True	131
WRMHS01M03	15811049	664	15808413	691	X	True	2636
WRMHS06G08	15811082	1131	15808420	1000	X	True	2662
WRMHS29H10	15811092	828	15810304	1011	X	True	788
WRMHS29B19	17317027	1338	17305399	1367	X	True	11628

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS24E10	11425095	1264	11489948	556	I	True	64853
WRMHS18I10	11428940	1304	11492586	981	I	True	63646
WRMHS14E15	11429691	1321	11486236	1269	I	True	56545
WRMHS08C19	1822301	867	1940449	1365	II	Gap	118148
WRMHS30H08	1017641	108	10225476	416	III	True	9207835
WRMHS28B17	1018398	743	10225375	1254	III	True	9206977
WRMHS27J24	1019269	1107	10225235	913	III	True	9205966
WRMHS25K06	7444692	311	926690	1415	III	True	6518002
WRMHS21F19	4110160	1550	4187837	497	IV	True	77677
WRMHS28I03	4112258	1301	4189456	1315	IV	True	77198
WRMHS34N02	4201075	1386	4121802	1360	IV	True	79273
WRMHS37H02	9065943	802	14248025	1002	IV	True	5182082
WRMHS23K12	10321619	1275	10950551	1179	IV	True	628932
WRMHS04J20	13509925	791	3401166	1358	IV	True	10108759
WRMHS36L17	15927960	1408	17171723	351	IV	Gap	1243763
WRMHS13A22	2177671	1332	18429864	819	V	True	16252193
WRMHS04B11	2379160	961	2108446	294	V	True	270714
WRMHS37O08	2423815	531	2569999	1234	V	True	146184
WRMHS35D22	2435713	1258	2578303	1336	V	True	142590
WRMHS08B20	2444577	1206	2586241	392	V	True	141664
WRMHS21B11	2579350	183	2429956	582	V	True	149394
WRMHS36O17	2579713	1384	2431166	435	V	True	148547
WRMHS24K17	2593743	1251	2445406	532	V	True	148337
WRMHS31I09	2598479	1367	2451497	1195	V	True	146982
WRMHS02N11	3294263	1376	3401684	412	V	True	107421
WRMHS08B04	3329209	1452	5663665	1386	V	True	2334456
WRMHS07P22	3330634	1212	3265189	933	V	True	65445
WRMHS39D23	3412397	1031	766562	1363	V	True	2645835
WRMHS14L12	3620906	351	7673722	1179	V	True	4052816
WRMHS06C05	3889516	1424	17814520	496	V	True	13925004
WRMHS21L14	8823814	1465	8907151	1066	V	True	83337
WRMHS25G15	8832890	1463	8907816	1358	V	True	74926
WRMHS06H10	8851961	1014	8927018	1197	V	True	75057
WRMHS28E14	8900602	1426	8823932	1406	V	True	76670
WRMHS37I22	8905261	1304	8819862	1334	V	True	85399
WRMHS12N14	8918536	1194	8846586	1315	V	True	71950
WRMHS17J20	14778352	150	15137122	444	V	True	358770
WRMHS27C08	15899332	1362	17170843	159	V	True	1271511
WRMHS32P16	16169358	1411	16226844	983	V	Gap	57486
WRMHS11L24	16922909	909	16978077	1399	V	True	55168
WRMHS17K18	16977321	1443	16922172	647	V	Gap	55149
WRMHS10F09	17123120	1164	17428106	1330	V	True	304986

Fosmid Name	Sequence Start	Bit Score	Sequence End	Bit Score	Chromosome	Gap*	Size
WRMHS32P01	17405279	695	17501053	1459	V	True	95774
WRMHS30G09	17413321	893	17509035	918	V	True	95714
WRMHS07M20	17423961	1330	17520732	1447	V	True	96771
WRMHS22J12	17428220	1264	17122853	911	V	True	305367
WRMHS30D09	17433187	1513	17529165	1393	V	True	95978
WRMHS25P09	17505742	584	17429242	708	V	True	76500
WRMHS36C21	17507209	429	17262329	636	V	True	244880
WRMHS22O21	17514137	1254	17421325	1153	V	True	92812
WRMHS07G03	17524664	1397	17426227	1166	V	True	98437
WRMHS17K20	17525688	1256	17426996	1234	V	True	98692
WRMHS35B04	17530231	1334	17427209	1404	V	True	103022
WRMHS21K01	17551073	1057	17293592	508	V	True	257481
WRMHS04K04	17787850	569	17848799	60.2	V	Gap	60949
WRMHS21L15	18165232	1432	18262933	1264	V	True	97701
WRMHS25B02	18177783	1293	18268914	1450	V	True	91131
WRMHS18K20	18265458	1214	18175693	889	V	True	89765
WRMHS18C13	19294299	628	19350053	1356	V	Gap	55754
WRMHS02D11	19356576	313	19301218	1098	V	True	55358
WRMHS18P17	1736033	1400	1808657	1218	X	True	72624
WRMHS10A21	1744670	1199	1811070	1487	X	True	66400
WRMHS15C05	1744733	1467	1815043	1362	X	True	70310
WRMHS40P12	1789255	1443	1728594	1290	X	True	60661
WRMHS14E12	1806153	1378	1737317	1286	X	True	68836
WRMHS01C19	1902224	1290	1960431	1482	X	True	58207
WRMHS32N06	1961105	1301	1905094	1186	X	True	56011
WRMHS14A13	10285262	1517	9072085	1511	X	True	1213177

* Gap displays whether the fosmid associated aligned next to a Gapped region, when tiled, or whether the sequence surrounding was a true contig.