



## Examination of Exhaustive Cloning Attempts Reveals that *C. elegans* piRNAs, Transposons, and Repeat Sequences are Efficiently Cloned in Yeast, but not in Bacteria

Or Sagy, Ron Shamir and Oded Rechavi

Journal Name:	Frontiers in Genetics
ISSN:	1664-8021
Article type:	Original Research Article
Received on:	28 May 2014
Accepted on:	26 Jul 2014
Provisional PDF published on:	26 Jul 2014
www.frontiersin.org:	<a href="http://www.frontiersin.org">www.frontiersin.org</a>
Citation:	Sagy O, Shamir R and Rechavi O(2014) Examination of Exhaustive Cloning Attempts Reveals that <i>C. elegans</i> piRNAs, Transposons, and Repeat Sequences are Efficiently Cloned in Yeast, but not in Bacteria. <i>Front. Genet.</i> 5:275. doi:10.3389/fgene.2014.00275
Copyright statement:	© 2014 Sagy, Shamir and Rechavi. This is an open-access article distributed under the terms of the <a href="https://creativecommons.org/licenses/by/4.0/">Creative Commons Attribution License (CC BY)</a> . The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

# Examination of Exhaustive Cloning Attempts Reveals that *C. elegans* piRNAs, Transposons, and Repeat Sequences are Efficiently Cloned in Yeast, but not in Bacteria

Or Sagy<sup>1</sup>, Ron Shamir<sup>2</sup> & Oded Rechavi<sup>1</sup>

1. Department of Neurobiology, Wise Faculty of Life Sciences & Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 69978, Israel.

2. Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel.

Correspondence to: odedrechavi@gmail.com

## Abstract

Genome sequencing requires insertion of random fragments of the sequenced organism's DNA into a unicellular host, most often *E. coli* bacteria. This manipulation was found in the past to be analogous to naturally occurring horizontal gene transfer, and moreover has proved valuable to understanding toxicity of foreign genetic elements to *E. coli*. Sequencing of the *C. elegans* genome was similarly achieved via DNA transformation into *E. coli*. However, numerous attempts have proven a significant percentage of the genome unclonable using bacteria, although clonable via yeast. We examined the genomic segments that were not in bacteria but in yeast, and observed that, in line with previous hypotheses, such sequences are more repetitive on average compared with the entire *C. elegans* genome. In addition, we found that these gap-sequences encode significantly more for DNA transposons. Surprisingly, we discovered that although the vast majority of the *C. elegans* genome is in bacteria (77.5%), almost all the thousands of sequences that encode for PIWI-interacting small RNAs, or 21U-RNAs (91.6%) were only in yeast. These results might help understanding why most piRNAs in *C. elegans* are physically clustered on particular loci on chromosome IV. In worms and in a large number of other organisms, piRNAs serve to distinguish "Self" from "Non-Self" sequences, and thus to protect the integrity of the genome against foreign genetic elements, such as transposons. We discuss the possible implications of these discoveries.

## Introduction

During the 1990s and early 2000s, there was a race to sequence the human genome (*Collins et. al, 2003*). The required techniques were developed in a rapid pace, and in parallel utilized for sequencing of other multicellular organism's genomes. The genome of the nematode *Caenorhabditis elegans* (*C. elegans*) was the first one completed (*The C. elegans Sequencing Consortium, 1998*). An unexpected result of the development of the different DNA cloning techniques was the accumulation of negative results, failed cloning attempts that allow gaining insight regarding barriers for genomic information transfer between organisms.

As part of sequencing the *C. elegans* genome, the whole genome of the worm was randomly broken into overlapping fragments, which were transformed into *E. coli* bacteria through the use of very large cloning vectors termed cosmids and fosmids (Coulson *et. al.*, 1986; Kim *et. al.*, 1992; Perkins *et. al.*, 2005). With sufficiently high coverage, sequencing of multiple overlapping fragments should allow in theory the assembly of the corresponding full genome. During the project, researchers soon realized that certain parts of the worm's genome could not be cloned in the bacteria, leaving gaps in the resultant genome (Coulson *et. al.*, 1991). Even though there have been numerous attempts over the past 20 years at filling the gaps using cosmids and fosmids (including a consortium dedicated to creating a library of fosmids that would cover the entire genome (Perkins *et. al.*, 2005)), none have been successful in covering these gaps. In total, about 20% of the genome could not be cloned in bacteria in spite of these repeated efforts. Throughout this manuscript we will refer to such sequences as "gap sequences".

In the 1990s, in order to close the gaps and obtain a complete sequence of the *C. elegans* genome, researchers turned to cloning in yeast, by using YACs (Yeast artificial chromosomes (Coulson *et. al.*, 1991)). The gap sequences quickly proved clonable by yeast (*S. cerevisiae*), and thus YACs that contain sequences covering the gaps were rapidly obtained and facilitated the completion of the worm genome (Wilson, 1999). Even so, the reason for the non-cloneability of these gaps in bacteria largely remained a mystery.

We found this incompatibility of worm genome fragments with bacterial cloning a valuable resource. *C. elegans* nematodes feed on bacteria both in real-life and in lab conditions, and this close contact facilitates exchange of genetic information between the nematodes and their resident microbes. For example, bacterially-expressed double stranded RNA transfers from the bacteria to the worm and silences endogenous *C. elegans* genes (Mello and Conte, 2004; Timmons *et. al.*, 2001; Liu *et. al.*, 2012; Sarkies and Miska, 2013). Although lateral gene transfer has been most extensively described in microorganisms, recent studies suggest that the process occurs more frequently than previously appreciated in eukaryotes as well (Bruto, *et al.*, 2014). DNA sequences were also shown to be laterally-exchanged between nematodes and symbionts (Haegeman *et. al.*, 2011). This is consistent with the various findings of gene exchange between organisms at all evolutionary distances (Koonin, 2010; Boto, 2010). Thus, a side benefit of the artificial cloning procedure used for sequencing the *C. elegans* genome is that it provides a keyhole into limitations of natural genome transfer between worm and bacteria.

Given that the aforementioned gaps have not been clonable in *E. coli*, a prokaryote, but have been clonable in the eukaryote *S. cerevisiae*, we hypothesized that there would be a genetic factor in these regions causing the adopting prokaryotic host to die. This notion was inspired by Sorek *et al.*'s experimental research on inhibition of gene transfer between different bacterial genomes to *E. coli* (Sorek *et. al.*, 2007). Such analysis of clone insertions and cloning gaps has proved valuable to understanding toxicity of bacterially-derived genetic elements to *E. coli* in the past (Sberro *et. al.*, 2013).

It has been believed that the cause of the lack of cloneability of various regions of the *C. elegans* genome was their assumed high repetitiveness (Wilson, 1999). We set out to explore whether there was another underlying property of these regions that, possibly in conjunction with repetitiveness, was actively inhibiting the prokaryotic hosts of these genome sequences from thriving.

## Methods

### Precise alignment of the clones to the genome

The sequences of *C. elegans* clones were downloaded from NCBI BioProject PRJNA13758, and the precise location of each of the clones was determined by BLASTing the sequence to the *C. elegans* genome (version WBcel235; this step was required as the locations in the NCBI record did not fit the genome version that we utilized). We used clone sequences from WormBase, retrieved from NCBI on April 4<sup>th</sup>, 2013. The sequences were categorized as “Cosmid”, “Fosmid” or “YAC” clones.

The YAC sequences that we analyzed tended to bridge across gap sequences, with both YAC ends occasionally overlapping with finished cosmid/fosmid sequences. The overlap varied between 0% and 100% (3 YAC sequences were found to be covered fully by cosmids/fosmids as well - Y119C1C, Y110A2B, Y70C5B). On average ~7% of the YAC sequence's ends overlapped with cosmids/fosmids. To prevent this overlap from influencing our results, we removed the overlapping segments from the YAC sequences before running any analyses. The resulting "YAC-exclusive" sequences were used in our study (hereby referred to as “YACs”).

### RepeatMasker analysis

As a first step, and in accordance with previous notions of repetitiveness as the factor limiting cloning of the YAC sequences, we ran RepeatMasker separately on YACs, cosmids and fosmids. This was conducted for each of the individual clones. We compared the average, median, minimal, maximal and non-zero results of each RepeatMasker parameter for every clone type (RepeatMasker was run with the species set as “*C. elegans*”). RepeatMasker lists retroelements (including sub-classifications), DNA transposons (including sub-classifications), rolling-circles, other unclassified interspersed repeats, small RNA, satellites, simple repeats and low complexity regions (*Smit et. al, 1996-2010*). As a control test for each clone type we also generated a random library with the same number of clones taken from random segments of the *C. elegans* genome, (hereby "random clones"), of similar sizes as those comprising that clone type (based on the average length and length standard deviation). These libraries were examined using the same RepeatMasker analysis. This analysis was used to assess the significance of our results, in comparison to enrichments detected by chance.

### Comparison by genomic location

It has been observed that the distribution of repeat elements along chromosomes is highly uneven (*The C. elegans Sequencing Consortium, 1998*). We therefore evaluated the difference in repeat element distributions between YACs and cosmids, by taking into account the location of the clones. The purpose of this analysis was to reduce bias caused by different location distributions of the types of clones. It may be hypothesized that these differences were causing the disparity that we found. In this, and the next couple of analyses, we compared YACs to cosmids and ignored fosmids, as there were much fewer fosmids compared to cosmids (107 vs. 2,520). To measure the effect of the clone's location along the chromosome and to obtain robust statistical analysis, we partitioned each chromosome into

four equal size quarters and summarized the statistics of the two extreme quarters together and of the two middle ones together.

### **Comparison by clone length**

Another factor that we took into consideration was the clone's length. Although YACs are generally much longer than cosmids, the YAC segments that were evaluated here were shorter, as they were trimmed (as mentioned above) to contain only the gaps used to bridge between cosmids. The YAC-exclusive segment median length was roughly 23.7kb, while for the cosmids the median length was approximately 32.5kb. In this analysis we divided each clone type into two groups – one of clones below the median length of the clone type and the other of clones over the median length. Statistics were collected for each group.

### **Comparison of genes encoded in the clones**

We categorized the genes in the clones based on the type of product that they encode for (or do not encode for): Coding Sequences (CDS), Non Coding RNAs (ncRNA), tRNAs and rRNAs. This categorization was determined based on the NCBI classification of each gene.

## **Results**

By comparing the sequences that were cloned in YACs, cosmids and fosmids, we observed, consistently with previous reports, that YAC sequences were overall more than twice as repetitive as the sequences cloned in cosmids. On average, 21.42% of the YAC nucleotides vs. 10.15% of cosmid nucleotides were masked by RepeatMasker. The most abundant repeat category, and an interesting result by itself, was enrichment in DNA transposons of various types in the YAC sequences (See Table 1). On average, 11.44% of YAC sequences were found to be DNA transposons vs. 5.69% in cosmids. Of the transposons, the most major difference was in PiggyBac transposons – 1.53% in YACs on average vs. 0.48% in cosmids. Other than the transposons, significant differences were found in unclassified repeats (4.65% vs 1.58%) and satellites (2.52% vs. 0.51%). The content of Simple repeats also varied – 1.65% vs 1.06%. The total interspersed repeats (all the repeats that were identified by RepeatMasker, excluding small RNA, satellites, simple repeats and low complexity regions) composed, on average, 17.04% of a YAC sequence vs. 8.27% of a cosmid sequence. Contrary to what might be expected, low complexity repeats did not show a significant difference (0.31% vs 0.3%).

Comparison with the results that were obtained with the randomly-generated clones showed that the differences between the clone types did not originate from their size. For all repeat types the results of the "random YACs" were very similar to those for "random cosmids" and "random fosmids", unlike the results that were obtained with the real clones (Table 2).

### **The relatively high repetitiveness of the sequences in YACs persists even when controlling for the chromosomal location.**

As previously reported, we observed a significant difference in repetitiveness and transposon-frequency between the different chromosome regions (both  $p < 0.0001$ , t-test). The extreme quarters of chromosomes were more repetitive than the middle quarters of the chromosomes (bases masked – 18.69% vs 6.99%), and were made up more of transposons (10.12% vs

3.96%). We also found a noteworthy difference between chromosomal location distributions of the clone types. The vast majority of the YACs, a little over three quarters (396) were in the extreme quarters and the minority in the middle quarters (119), while the cosmids were mostly in the middle quarters (979 vs 1,540 – a ratio of ~2:3). We then ran RepeatMasker while taking into account this difference – comparing the differentiated groups between YACs and cosmids and within each clone type. Histograms of the locations of YACs and cosmids and the locations of repeats of any type are shown in Figures 1-2. A similar histogram for transposons, portraying the deviation more finely can be found in Figure 3. To test the effect of the chromosomal location, we reran RepeatMasker separately on the extreme and on the middle quarters. In both cases the conclusions that were drawn from the previous analysis persisted. YACs-contained sequences were more repetitive, and were composed more of transposons than cosmids, although clones in the first and last quarters were overall more repetitive than their second and third quarter counterparts (YACs – 12.78% and 7.01%, cosmids – 8.49% and 3.91%).

### **Differences in the length of YACs and cosmids do not account for their repetitiveness difference**

There were no outstanding differences in repeat frequencies between the clones from the same type above and below the median length. Hence, the differences in repeat frequency between the clone types could not explained by length variation.

### **18 YACs had no repetitive elements**

We examined the 18 YACs for which RepeatMasker did not find any repetitive elements (as well as 3 YACs that were completely overlapped by cosmids/fosmids). Out of these YACs, after removing overlaps with cosmids and fosmids, 12 were found to be smaller than 1kb in length, 3 between 1 and 1.5kb, and only 3 over 1.5kb in length - Y68G5A (3.4kb, after removing a 0.2kb overlap; no genes in the non-overlapping portion), Y2C2A (8.5kb, after removing an 8.5 kb overlap; part of the gene tag-80 in the non-overlapping portion) and Y53C12C (23.4kb, after removing a 0.2kb overlap; the gene *eyg-1* is enclosed in the non-overlapping portion).

### **YACs are significantly enriched for 21U piRNAs**

An outstanding difference between the densities of CDS genes in YACs and cosmids was not found (a slightly higher density was found in cosmids: 182 genes per MB in YACs vs. 238 genes per MB in cosmids). tRNAs and rRNA are overrepresented in YACs – 94 YAC tRNA genes vs. 114 cosmid tRNA genes (4.2 vs. 1.5 genes per MB), and 10 YAC rRNA genes vs. 12 cosmid rRNA genes (0.45 vs. 0.15 genes per MB). Results for the CDS and ncRNA genes can be found in Table 3.

The most striking result of our analysis was that ncRNAs were dramatically overrepresented in YACs. A roughly similar absolute amount of ncRNA was found in YACs and cosmids despite the fact that YACs take up about a quarter of the length that cosmids take in the genome (10,799 vs. 11,108 ncRNA genes, respectively – corresponding with densities of 484 vs. 143 ncRNAs/MB). Further investigation revealed that the vast majority of the YAC ncRNAs are PIWI-interacting small RNAs, or 21U-RNAs (about 91%), while they are less than half of the cosmid ncRNAs (about 42.4%). Moreover, in terms of absolute quantities, YACs include more than twice as many 21U-RNAs as cosmids – 4,713 vs. 9,879.

The majority of the DNA sequences that give rise to 21U piRNAs reside in two regions in chromosome IV (and also in another small region on chromosome IV) that are called “piRNA clusters” (Ruby *et. al.*, 2006). We analyzed these regions for repetitiveness and they turned out to be in overall less repetitive compared to other regions, both on chromosome IV and on other chromosomes (the two major piRNA clusters had 8.15% and 11.51% of the bases masked as repetitive and the third cluster had 8.11% of the bases masked, compared to 11.89% for chromosome IV as a whole and 14.9% for chromosome I for example). Since the regions that encode for abundant YAC-specific piRNAs are not repetitive, it raises the possibility that they are unclonable in *E. coli* for other reasons, perhaps due to toxicity (see Discussion). The longest four YACs, and seven of the longest ten YACs, were on chromosome IV (see Table 4). The longest four YACs alone hold 3,754 ncRNAs (roughly 35% of all ncRNA genes found in YACs), with the 6<sup>th</sup> longest holding 1,083 more. The longest 4 YACs also mapped specifically to one of the two piRNA clusters on chromosome IV.

While chromosome IV contains the vast majority of the YAC ncRNAs (93.3%, vs. 44.1% of cosmid ncRNAs), it is not over-represented for YACs overall (21.8% of the YAC exclusive bases are on chromosome IV, which constitutes 17.4% of the nematode genome). When removing chromosome IV from the analysis, the discrepancy between YACs and cosmids on ncRNA content was flipped: instead of the ncRNAs being divided roughly equally between YACs and cosmids in terms of absolute numbers, when excluding chromosome IV, there were about 8.8 as many ncRNAs in cosmids as in YACs (6,182 vs 702).

## Discussion

We used a bioinformatic approach to probe DNA sequences that could be cloned in a eukaryote (yeast), but not in a prokaryote (bacteria). We found several major characteristics in such sequences, including high repetitiveness and enrichment for DNA transposons and 21U piRNAs. If these characteristics are typical to sequences of other organisms, it could in theory shed light on the barriers to genetic information transfer that exist between prokaryotes and eukaryotes, and should be interesting to validate experimentally.

The inability to efficiently clone repetitive sequences in bacteria is perhaps not biologically relevant, as it could simply stem from technical considerations. On the other hand, and while our results do not prove it in any way, it is possible that the inability to clone DNA transposons or piRNAs in bacteria has physiological importance.

Acquisition of vectors (cosmids or fosmids in this case) that carry mobile elements could be detrimental, if such genomic parasites are allowed to colonize or disrupt the genome of the host. It would be interesting to examine whether the type of transposons that we found to be enriched in YACs “jump” in *E. coli*, but not in *S. cerevisiae*, and if transposition, should it be shown to occur, compromises the bacteria’s viability. Transposons that are not dependent on host factors could in theory transpose and confer damage, even upon lateral gene transfer.

piRNA-mediated RNA interference has a role in genome surveillance against foreign sequences in multiple organisms across the tree of life (animals and protists). One fascinating scenario that could be envisioned based on our data is that piRNAs might act *in trans*, to eliminate bacteria, *E. coli* in this case, with which *C. elegans* intimately interacts. Although

*C.elegans* worms are not natural hosts of *E.coli*, the two organisms have been grown together in the lab for half a century. *C.elegans* and *E.coli* could in theory adapt to interact by these mechanisms, as piRNAs and piRNA pathway genes evolve rapidly, due to an “arms race” with transposons (Yi *et al.*, 2014), and since piRNAs target rapidly evolving genes (Weick *et al.*, 2014). The promoters of most piRNAs, which are autonomous transcriptional units in *C.elegans*, contain a consensus motif for binding of transcription factors. Thus in theory, possession of just one appropriate transcription factor could allow a host (such as *E. coli*) to transcribe most piRNAs. In flies, piRNA clusters act as “traps” in which transposons land (Malone and Hannon, 2009). It is not understood why most piRNAs in *C.elegans* are clustered together on chromosome IV. Perhaps the inability to transfer a large cluster of piRNAs from *C.elegans* to *E.coli* is one of the reasons.

Characterization of DNA sequences that are lethal to bacteria but not to yeast could be applicative (Sorek *et al.*, 2007). Given further research and validation, it could be possible in the future to combine, for example, piRNA-encoding DNA sequences, and thus to affect bacteria, while sparing the eukaryote hosts with which they interact.

### Acknowledgments

We thank all members of the Rechavi lab for their helpful comments, and Gregory Minevich for discussions and help with the development of the idea. Oded Rechavi and Ron Shamir were supported in part by the Israel Science Foundation (OR grant 1101/13, RS grant 317/13)

### References

- Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proceedings. Biological Sciences / The Royal Society*, 277, 819–827. doi:10.1098/rspb.2009.1679
- Bruto, M., Prigent-Combaret C., Luis P., Moëgne-Loccoz Y., & Muller D. (2014). Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes. *Proceedings. Biological Sciences / The Royal Society*, 281. doi: 10.1098/rspb.2014.0848
- Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. *Science (New York, N.Y.)*, 300, 286–90. doi:10.1126/science.1084564
- Coulson, A., Kozono, Y., Lutterbach, B., Shownkeen, R., Sulston, J., & Waterston, R. (1991). YACs and the *C. elegans* genome. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, 13, 413–417. doi:10.1002/bies.950130809
- Coulson, A., Sulston, J., Brenner, S., & Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, 83, 7821–7825. doi:10.1073/pnas.83.20.7821

- Haegeman, A., Jones, J. T., & Danchin, E. G. J. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Molecular Plant-Microbe Interactions : MPMI*, *24*, 879–887. doi:10.1094/MPMI-03-11-0055
- Kim, U.-J., Shizuya, H., de Jong, P. J., Birren, B., & Simon, M. I. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Research*, *20*, 1083–1085. doi:10.1093/nar/20.5.1083
- Kimelman A, Levy A, Sberro H, Kidron S, Leavitt A, Amitai G, et al. (2012). A vast collection of microbial genes that are toxic to bacteria. *Genome Research*, *22*, 802-9. doi: 10.1101/gr.133850.111
- Koonin, E. V. (2010). The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biology*, *11*, 209. doi:10.1186/gb-2010-11-5-209
- Liu, H., Wang, X., Wang, H.-D., Wu, J., Ren, J., Meng, L., et al. (2012). Escherichia coli noncoding RNAs can affect gene expression and physiology of *Caenorhabditis elegans*. *Nature Communications*, *3*, 1073. doi:10.1038/ncomms2071
- Malone, C. D., & Hannon, G. J. (2009). Small RNAs as Guardians of the Genome. *Cell*. doi:10.1016/j.cell.2009.01.045
- Mello, C. C., & Conte, D. (2004). Revealing the world of RNA interference. *Nature*, *431*, 338–42. doi:10.1038/nature02872
- Perkins, J., Wong, K., Warren, R., Schein, J., Stott, J., Holt, R., et al. (2005). A *Caenorhabditis elegans* fosmid library. *International Worm Meeting*.
- Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., et al. (2006). Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in *C. elegans*. *Cell*, *127*, 1193-1207. doi: 10.1016/j.cell.2006.10.040
- Sarkies, P., & Miska, E. a. (2013). Is there social RNA? *Science*, *341*, 467–468. doi:10.1126/science.1243175
- Sberro H, Leavitt A, Kiro R, Koh E, Peleg Y, Qimron U, Sorek R. (2013). Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. *Molecular Cell*, *50*, 136-48. doi: 10.1016/j.molcel.2013.02.002
- Smit A, Hubley R, Green P. RepeatMasker Open-3.0. (1996–2010). Available from: [www.repeatmasker.org](http://www.repeatmasker.org)
- Sorek, R., Zhu, Y., Creevey, C. J., Francino, M. P., Bork, P., & Rubin, E. M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science (New York, N.Y.)*, *318*, 1449–52. doi:10.1126/science.1147112

The C. elegans Sequencing Consortium. (1998). Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science*, 282, 2012–2018. doi:10.1126/science.282.5396.2012

Timmons, L., Court, D. L., & Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene*, 263, 103–112. doi:10.1016/S0378-1119(00)00579-5

Weick E. M., Sarkies P., Silva N., Chen R. A., Moss S. M., Cording A. C. et al. (2014). PRDE-1 is a nuclear factor essential for the biogenesis of Ruby motif-dependent piRNAs in *C. elegans*. *Genes & Development*, 28, 783-96. doi: 10.1101/gad.238105.114

Wilson, R. K. (1999). How the worm was won: the *C. elegans* genome sequencing project. *Trends in Genetics*, 15, 51–58. doi:10.1016/S0168-9525(98)01666-7

Yi M., Chen F., Luo M., Cheng Y., Zhao H., Cheng H., et al. (2014). Rapid Evolution of piRNA Pathway in the Teleost Fish: Implication for an Adaptation to Transposon Diversity. *Genome Biology and Evolution*, 6 (6), 1393-1407. doi: 10.1093/gbe/evu105

## Figure Legends

### Figure 1:

**Cosmid and YAC Locations:** Start locations of cosmids and YACs, averaged across chromosomes, portrayed by portion of Cosmids and YACs starting at each percent of the chromosome (A) and starting from each percent on (B)

### Figure 2:

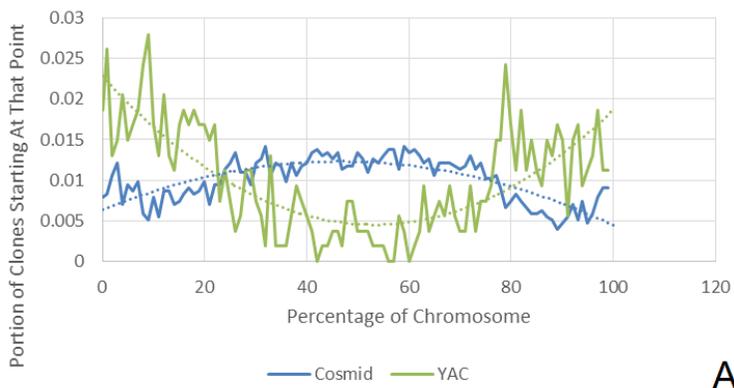
**Repeat locations:** Start locations of repeats identified by RepeatMasker, averaged across chromosomes, portrayed by portion of repeats starting at each percent of the chromosome (A) and starting from each percent on (B)

### Figure 3:

**DNA transposon locations:** Start locations of DNA transposons, averaged across chromosomes, portrayed by portion of transposons starting at each percent of the chromosome (A) and starting from each percent on (B)

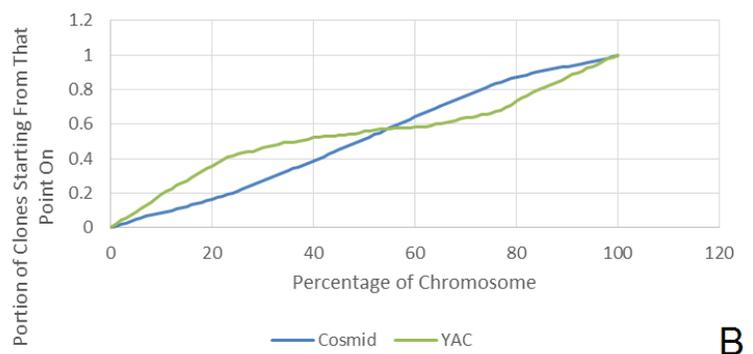
Figure 1.TIFF

Location Histogram of YACs and Cosmids



A

Cumulative Location Histogram of YACs and Cosmids



B

Figure 2.TIFF

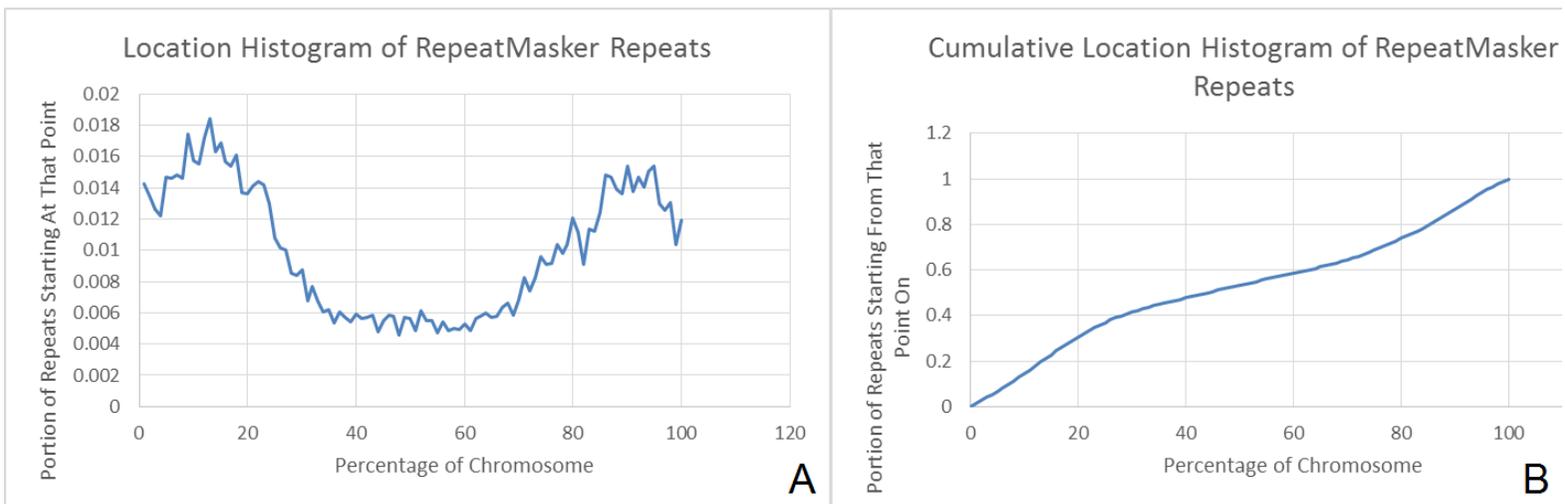
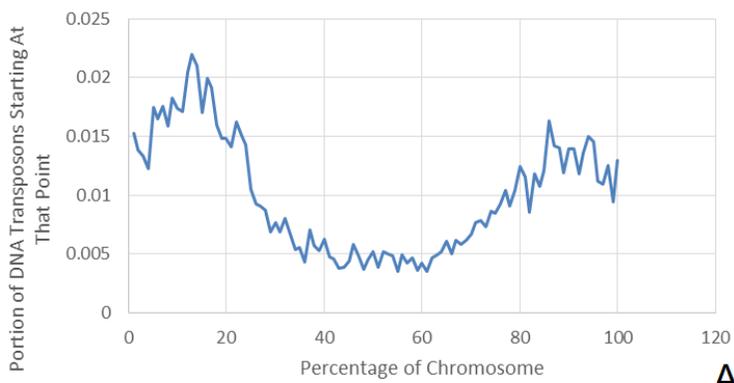


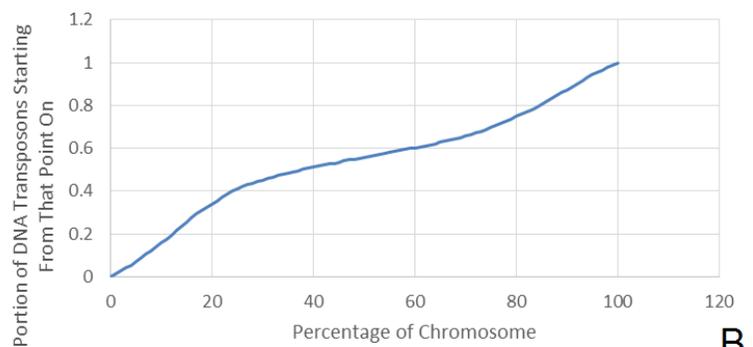
Figure 3.TIFF

Location Histogram of DNA Transposons



A

Cumulative Location Histogram of DNA Transposons



B

Figure 4.TIFF

### Cumulative Location Histogram of RepeatMasker Repeats

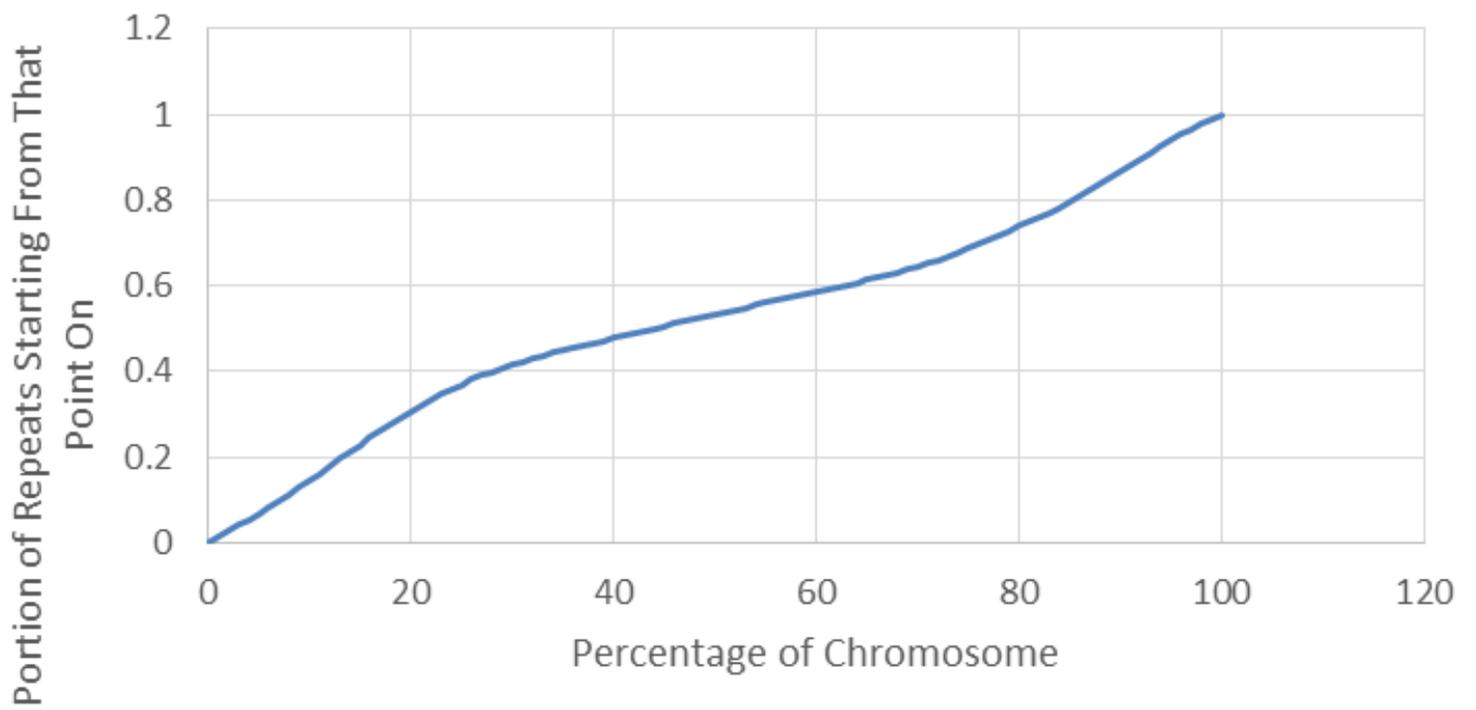


Figure 5.TIFF

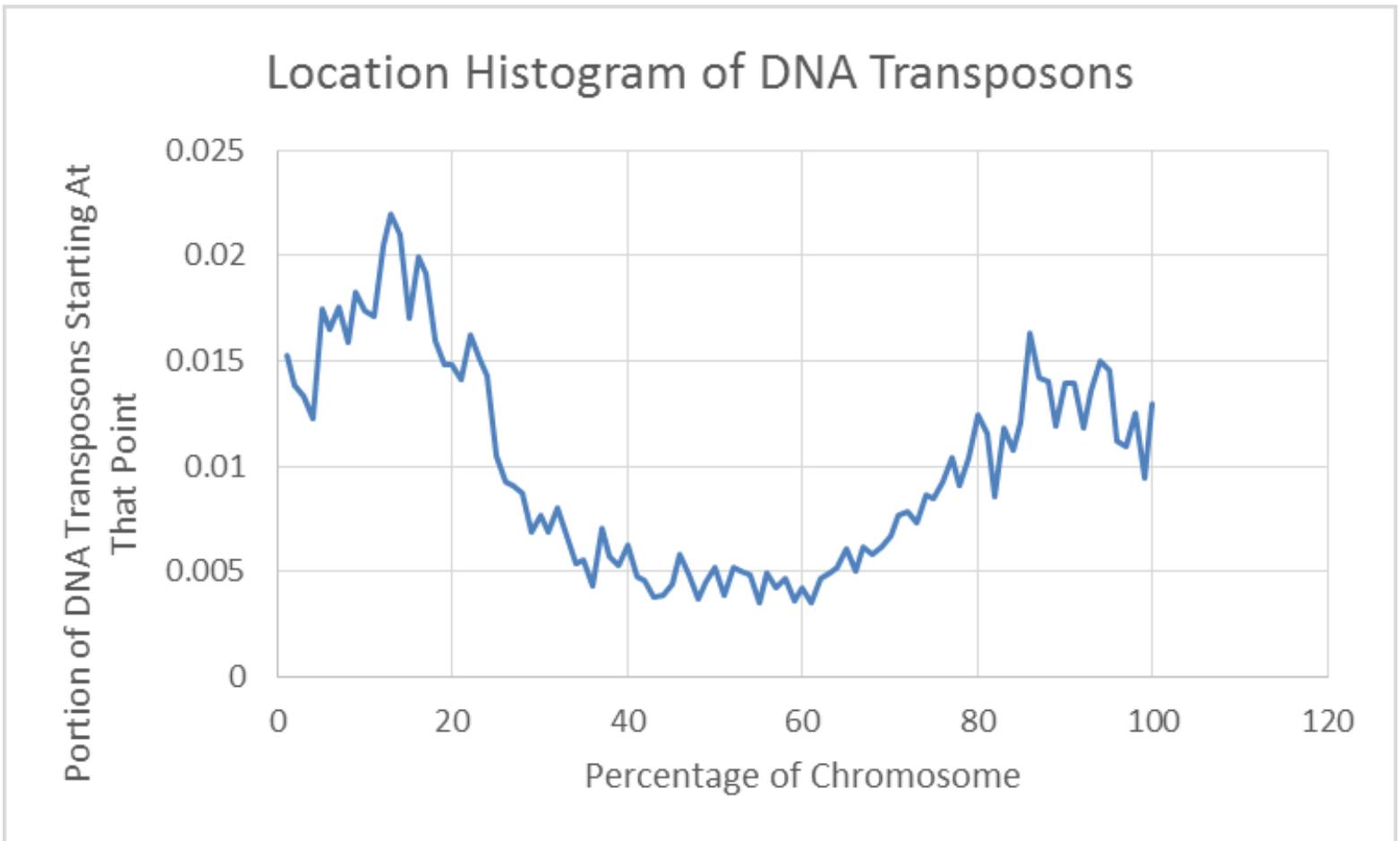


Figure 6.TIFF

