2011

Shifting patterns of natural variation in the nuclear genome of *caenorhabditis elegans*

Eleanne Solorzano  
*University of New Hampshire*

Kazufusa Okamoto  
*University of New Hampshire*

Pushpa Datla  
*University of New Hampshire*

Way Sung  
*University of New Hampshire*

R. D. Bergeron  
*University of New Hampshire*

See next page for additional authors

Follow this and additional works at: [https://scholars.unh.edu/hubbard](https://scholars.unh.edu/hubbard)

Part of the Genomics Commons

**Recommended Citation**

Solorzono et al.: Shifting patterns of natural variation in the nuclear genome of *caenorhabditis elegans*. BMC Evolutionary Biology 2011 11:168.

This Article is brought to you for free and open access by the Research Institutes, Centers and Programs at University of New Hampshire Scholars’ Repository. It has been accepted for inclusion in Hubbard Center for Genome Studies (HCGS) by an authorized administrator of University of New Hampshire Scholars’ Repository. For more information, please contact Scholarly.Communication@unh.edu.
Shifting patterns of natural variation in the nuclear genome of *caenorhabditis elegans*

Eleanne Solorzano1,2,4*, Kazufusa Okamoto1,2, Pushpa Datla1,3, Way Sung1,2, RD Bergeron1,3 and WK Thomas1,2

Abstract

**Background:** Genome wide analysis of variation within a species can reveal the evolution of fundamental biological processes such as mutation, recombination, and natural selection. We compare genome wide sequence differences between two independent isolates of the nematode *Caenorhabditis elegans* (CB4856 and CB4858) and the reference genome (N2).

**Results:** The base substitution pattern when comparing N2 against CB4858 reveals a transition over transversion bias (1.32:1) that is not present in CB4856. In CB4856, there is a significant bias in the direction of base substitution. The frequency of A or T bases in N2 that are G or C bases in CB4856 outnumber the opposite frequencies for transitions as well as transversions. These differences were not observed in the N2/CB4858 comparison. Similarly, we observed a strong bias for deletions over insertions in CB4856 (1.44:1) that is not present in CB4858. In both CB4856 and CB4858, there is a significant correlation between SNP rate and recombination rate on the autosomes but not on the X chromosome. Furthermore, we identified numerous significant hotspots of variation in the CB4856-N2 comparison.

In both CB4856 and CB4858, based on a measure of the strength of selection (k_v/k_s) all the chromosomes are under negative selection and in CB4856, there is no difference in the strength of natural selection in either the autosomes versus X or between any of the chromosomes. By contrast, in CB4858, k_v/k_s values are smaller in the X chromosome in comparison to the autosomes. In addition, in CB4858, k_v/k_s values differ between chromosomes.

**Conclusions:** The clear bias of deletions over insertions in CB4856 suggests that either the CB4856 genome is becoming smaller or the N2 genome is getting larger. We hypothesize the hotspots found represent alleles that are shared between CB4856 and CB4858 but not N2. Because the k_v/k_s ratio in the X chromosome is higher than the autosomes on average in CB4858, purifying selection is reduced on the X chromosome.

Background

All variation found in natural populations is ultimately derived from evolutionary forces acting upon mutation and recombination. By studying genome-wide patterns of variation among closely related species, it is possible to make inferences about the evolutionary process of that organism. Although the model organism *Caenorhabditis elegans* has been well studied in biology, our knowledge of the evolution of natural populations of *C. elegans* is limited [1]. In the past, molecular variation in *C. elegans* has been studied using assays involving single mitochondrial and nuclear genes [2], transposons [3,4], amplified fragment length polymorphisms (aflps), and microsatellites [5-8].

In this study, we investigate genetic variation in three strains of *C. elegans*: N2, CB4856, and CB4858 using large-scale DNA sequence datasets [9-11]. N2 (from Bristol, England) is the most commonly used lab strain. It is the strain for which the genome sequence has been completely determined. CB4856 is a strain of *C. elegans* from Hawaii while CB4858 is a strain of *C. elegans* from Pasadena, CA. Strains CB4856 and CB4858 were isolated in 1972 and 1971, respectively. Since then, they have been cultured in the lab. In their *C. elegans* population genomics study, Rockman and Kruglyak [12] distinguished 41 haplotypes among 125 wild isolates in their genotype data from 1460 N2-CB4856 SNPs. They indicated that CB4858 shares haplotype 20 with strains

* Correspondence: eleannes@cisunix.unh.edu
1Hubbard Center for Genome Studies, University of New Hampshire, Durham, NH, US
Full list of author information is available at the end of the article
from other localities, including AB2-4, from Adelaide, Australia, and CB4855, from Palo Alto, California. The genotypic similarity among CB4858, CB4855, and AB2-4 suggests they may share an ancestor in a laboratory.

Previous analysis of multiple *Caenorhabditis* lines shows extensive variation in nuclear single-nucleotide polymorphism (SNP) density [9], with the Hawaiian strain CB4856 having the highest SNP density and most unique SNPs [13]. Additional polymorphisms in CB4856 were found by Swan et al. [13], and analysis of the distribution of these polymorphisms by Cutter et al. [14] revealed that the polymorphisms occurred more frequently in the arms than in the core in chromosomes one to five, but are randomly distributed in chromosome X.

The patterns of natural variation found in a genome are the result of mutational events that are influenced by both drift and selection. Under neutral theory, a majority of mutations that arise spontaneously are neutral, and their fate (loss or fixation) is primarily determined by drift. In rare cases, a spontaneous mutation will confer a selective advantage (positive selection) or disadvantage (purifying selection). If by chance this mutation is not lost by drift, in consequent generations the mutation will spread throughout the population, and variation at linked sites will either accumulate or be purged. The combination of neutral mutations fixed through drift and the mutations that are accumulated or purged determine the genomic landscape of the organism [15]. By surveying the patterns of natural variation between the closely related *Caenorhabditis* strains CB4856 and CB4858, it is possible to investigate the evolutionary processes of drift and selection occurring in this lineage.

Although previous studies provide insights into *C. elegans* SNP patterns and densities in the natural isolates, a large-scale genomic analysis of the *C. elegans* polymorphic spectrum has not been accomplished until now. Other studies of *C. elegans* have not focused on examining the patterns of variation that we present here, but have instead focused on important patterns. For example, Weber et al. [16] reported the genome sequence of the LSJ1 (Bristol) strain and compared it to the reference wild type N2 (Bristol). LSJ1 is a sibling of the reference wild type N2. The authors indicated that comparing the genomes of N2 and LSJ1 highlight genetic changes associated with lab domestication and errors in the *C. elegans* reference genome. They confirmed that LSJ1 and N2 are derived from the same Bristol isolate. In addition, they identified 3246 predicted differences between the LSJ1 and N2 genome sequences. Finally, by sequencing two recently isolated wild *C. elegans* strains, CB4856 (Hawaii) and ED3054 (Kenya), they inferred the ancestral state of the original Bristol isolate. All of this makes CB4856 an extremely valuable resource not only for *C. elegans* gene mapping [10] but also for the study of the distribution and spectrum of *Caenorhabditis* SNPs.

Graustein et al. [17] compared the levels of DNA sequence polymorphism among three closely related species in the genus *Caenorhabditis*: two self-fertilizing species, *C. elegans* and *C. briggsae*, and one cross-fertilizing species, *C. remanei*. Jovelin et al. [18] analyzed nucleotide variation within and between species across two olfactory pathways in *Caenorhabditis* to examine the relationship between evolutionary rate and gene function and their position within the pathway. The authors found that transcription factors specifying chemosensory neuron subtype identity exhibit higher levels of nucleotide divergence and polymorphism than the structural genes in the signaling pathways. They found much higher levels of polymorphism within *C. remanei* than within the related species *C. elegans* and *C. briggsae*. Their findings suggest behavioral differences between species in response to olfactory stimuli are likely to involve differences in gene regulation due to the divergence of regulatory genes.

We report the entire spectrum of events critical for understanding the fundamental divergence patterns of these natural isolates. In addition, we compare our results to the N2 MA line results from Denver et al. [19] because we wish to compare the differences in mutation patterns when selection is present (in our study) versus when it is nearly absent (in MA lines).

**Results and Discussion**

**Overview of SNP variation**

We identified a total of 13360 putative SNPs in our analysis of the CB4856 genome compared to N2. Direct resequencing of 55 of these putative SNPs confirmed 51, suggesting our rate of confirmation is 93%. The 55 SNPs chosen for resequencing were representative of all the SNPs in this study. The level of SNP differences between N2 and CB4856 is $28.6 \times 10^{-4}$ SNP per base with insertion deletion events (indels) contributing slightly more of the variation than base substitutions and inversions (Table 1). This is consistent with the known mutation spectrum described in Denver et al. [19] that reports slightly more indels than SNPs in the *C. elegans* MA lines. In contrast, a recent genome wide comparison of N2 and the strain CB4858 [11] revealed a much lower genome wide SNP level ($5.89 \times 10^{-4}$ SNPs per base) and those differences were dominated by base substitutions. The difference in the ratio of base substitutions to indels in the two comparisons could be attributed to the differences in sequencing approaches used to assay variation. The CB4856 data was based on longer Sanger reads (average trimmed length was 130 bp) while
the CB4858 data was based on short read sequencing. Reads with more than one SNP are difficult to assemble when using short read technology. Consequently, in CB4858, reads with more than 2 differences are incompatible with this analysis [11]. In that scenario, sequential or highly clustered SNPs may not be detected in the CB4858 dataset.

The total number of bases assayed in CB4856 was 4658922 (4.6465% of the C. elegans CB4858 dataset. Potential or highly clustered SNPs may not be detected in the reference genome) [11]. Even though 4.6465% is a small portion of the assay of 77% of the C. elegans reference genome [Hiller et al., 2008], the base substitution/indel ratio is 0.6883:1 (5402/7848) for CB4856 and 7.3369:1 (37132/5061) for CB4858 (Table 1). Two observations suggest that the differences in the observed base substitution/indel ratios in CB4846 vs CB4858 do not stem from read length differences in the two assays. First, clusters of SNPs (see CB4856 Statistically Significant Hotspots) do not contribute to a large fraction of the total variation. Therefore, sequential indels which may be undetectable using short-read technology do not greatly influence the base substitution/indel ratio. Second, when we limit the comparisons to single base substitutions and indels which are fully assayable using short read technology in CB4858, we find that the discrepancy between these two natural isolates remains (Table 1). Based on the above observations, the CB4856 SNPs have a higher proportion of indels than the CB4858 SNPs. These observations are consistent with those reported by Denver et al. [19] in which C. elegans MA lines showed a base substitution/indel ratio of 0.7059:1 (12/17).

Either the CB4856 genome is getting smaller or the N2 genome is getting larger

Among the SNPs predicted in the comparison of CB4856 and N2, we observed a strong and significant bias for deletions (4627) over insertions (3221). This bias was reflected in single event indels as well as larger indels (see Table 1 and Additional File 1). Although the direction in which the indel occurred is unknown, this pattern differs from the CB4858 ratio (1:1), suggesting two possibilities: (1) the CB4856 genome is becoming smaller or (2) the N2 genome is getting larger because it fails to purge insertions. This result differs from the baseline mutation spectrum in C. elegans MA lines in Denver et al. [19]. They observed a high frequency of indels where the mutations appeared to be biased toward insertions. There are two possible explanations for this observation: (1) there is a change in the mutation spectrum between N2 and CB4856 or (2) selection constraints between N2 and CB4856 have changed.

Sequential vs single SNPs

Sequential SNPs can arise from single mutational events involving multiple base pairs or as a consequence of individual events at adjacent positions. There is a larger proportion of sequential SNPs in CB4856 compared to CB4858 (Table 1). The high proportions of observed sequential base substitutions [3.6% (476/13360)] and indels [24% (3206/13360)] in CB4856 are not consistent with expectations based on a hypothesis of independent adjacent polymorphisms. This suggests a large fraction of the adjacent SNPs result from mutations involving multiple adjacent base pairs or that the adjacent SNPs are not independent. The distribution of the sizes for sequential SNPs, however, suggests that the occurrence of sequential SNPs is negatively correlated with the number of bases involved.

Due to the limitations of short read sequence alignment, the analysis between N2 and CB4858 is limited to sequential base substitutions, indels, and putative inversions with a length of 2 bp. In particular, inversions are a special subclass of sequential changes that occur through a single event. We identified all size classes of inversions in CB4856 and in CB4858 (Additional File 2).
These account for only a small fraction [0.82% (110/13260) in CB4856 and 0.18% (76/42269) in CB4858] of all SNPs and do not affect the overall conclusions of our study. We note that in CB4858, the inversions may be downwardly biased due to the CB4858 sequencing method.

### Shifting patterns of Base substitution

The base substitution pattern when comparing N2 against the CB4858 isolate reveals a *C. elegans* transition over transversion bias of 1.32:1 (21145/15993; Table 2). However, when comparing N2 against CB4856, there is no transition over transversion bias (1:1 ratio; Table 2). Furthermore, in CB4856, there is a significant bias in the direction of base substitution. When comparing N2 against CB4856, the frequency of A or T bases in N2 that are G or C bases in CB4856 greatly outnumber the opposite frequencies for transitions as well as transversions. At equilibrium, the numbers of G:C to A:T differences between two genomes should be equal to the number of A:T to G:C differences. The shift in base substitution results in a significant detectable shift in base composition between N2 (35.44% GC) and CB4856 (37.77% GC). In particular, there is a significant difference between the following rates in the N2/CB4856 comparison: G → A lower than A → G; C → T lower than T → C; C → A lower than A → C; G → T lower than T → G. None of these differences were observed in the N2/CB4858 comparison (see Table 3).

The results shown here differ from the expectation of a significant mutation bias from G:C to A:T and G:C to T:A based on mutation accumulation lines [20]. Because the base substitution pattern is consistent with the N2/CB4858 comparison, but different in CB4856, the change is likely either a change in the mutational mechanisms or shifting selective forces. CB4856 is a prime example of base compositional differences (percentage of the CB4856 genome that consists of the various nucleotides A, T, C, G) that can occur within a species.

### CB4856/CB4858 SNP rate of chromosomes

Both CB4856 and CB4858 exhibit significant differences in SNP rates among the chromosomes (Table 4). In particular, chromosomes I to X do not have a uniform SNP rate in either CB4856 (p < 2.2e-16) or CB4858 (p < 2.2e-16). In addition, the SNP rate is higher in the autosomes than in chromosome X in CB4856 (p < 2.2e-16). In CB4856, chromosome III has the smallest SNP rate while chromosome V has the largest SNP rate. Although previous observations of this disparity have been attributed to the mutagenic effect of recombination [21], the random distribution of mutations across all chromosomes in mutation accumulation lines [20] makes it more likely that selective sweeps are purging polymorphisms from different regions of the genome. A comparison of SNP rates between chromosomes in CB4858 is not appropriate because of the stochastic nature of the SNPs found in the CB4858-N2 comparison. There are large stretches of DNA that are shared by N2 and CB4858.

### CB4856/CB4858 arms versus core SNP rate

We compared the SNP rate in the arms vs the cores in different chromosomes in CB4856 and CB4858 using chi-squared tests. In both strains, 1) the SNP rate in the arms is significantly higher than in the core in all of the chromosomes; 2) the rates of SNPs in the arms are different across the chromosomes; 3) the rates of SNPs in the core are different across the chromosomes (Table 5). There are many possible reasons why the SNP rates differ between the arms and the core in both CB4856 and CB4858. The first possibility is that there are more genes in the core. Since purifying selection removes

---

**Table 2 Base substitutions (bs) in CB4856 and CB4858 relative to the N2 reference genome***

| N2>CB4856 | Base Sub | N2>CB4858 | Base Sub | N2>CB4858 | Base Sub | N2>CB4858 | Base Sub |
|-----------|----------|-----------|----------|-----------|----------|-----------|----------|
| **Transitions** | | | | | | |
| G>A | 621 (0.11) | A>G | 745 (0.14) | G>A | 5266 (0.14) | A>G | 5375 (0.14) |
| C>T | 639 (0.12) | T>C | 722 (0.13) | C>T | 5163 (0.14) | T>C | 5341 (0.14) |
| **Transversions** | | | | | | |
| C>A | 278 (0.05) | A>C | 374 (0.07) | C>A | 1871 (0.05) | A>C | 1945 (0.05) |
| G>T | 276 (0.05) | T>G | 345 (0.06) | G>T | 1906 (0.05) | T>G | 1893 (0.05) |
| T>A | 480 (0.09) | A>T | 484 (0.09) | T>A | 3185 (0.09) | A>T | 3240 (0.09) |
| G>C | 240 (0.04) | C>G | 232 (0.04) | G>C | 964 (0.03) | C>G | 989 (0.03) |

*Number (Proportions out of all base substitutions in each strain: 5436 base substitutions in CB4856 and 37138 base substitutions in CB4858)
mutations from coding regions, if the gene density is higher in the core, fewer SNPs will occur there [9,20,22]. The second possibility is that there may be more recombination away from the core. If recombination is mutagenic [21], higher rates of recombination will increase SNP density. On the other hand, recombination interrupts linkage disequilibrium reducing the efficacy of selection. As stated in Cutter et al. [23], chromosome arms have moderate to high rates of crossover recombination and low gene density and cores have strikingly little recombination despite comprising nearly half of each autosome. Rockman and Kruglyak [12] and Barnes et al. [21] provided evidence of these observations in their studies. Finally, there may be an inherent difference in mutation rates in the arms and the core. As observed in this paper and by Stein et al. [1], there is a difference in the rate of SNPs in the arms vs the core. Observations by Denver et al [20] have suggested that the baseline rate of mutation in MA lines is not increased in the arms of autosomes. This suggests that selection, not mutation, plays a larger role in the biased physical distribution of polymorphisms. We then conclude that the reason the SNP rate differs in the arms and core regions is primarily due to natural selection.

Recombination and SNP rate
In both CB4856 and CB4858 (Figure 1), the distribution of SNP variation within the C. elegans chromosomes shows there is a significant relationship between SNP frequency and recombination rate on the autosomes. However, there is no significant linear relationship between SNP frequency and recombination rate on the X chromosome in either isolate (Table Table 6).

The discrepancy with the X chromosome can be explained by the X chromosomes’ many functional differences from the autosomes [24]. Recombination is almost non-existent in the right terminal arm of the X chromosome.

### Table 3 Bayesian Intervals for difference in rates between different types of transitions and transversions in CB4856 and CB4858

| CB4856 Base sub | N2->CB4856 | Base Sub | N2->CB4856 | Base Sub | 95% Bayesian Interval for difference using MCMC |
|----------------|------------|----------|------------|----------|-----------------------------------------------|
| CB4856 Transitions | G->A | 621 | A->G | 745 | (-0.031, -0.008) |
| | C->T | 639 | T->C | 722 | (-0.025, -0.003) |
| CB4856 Transversions | C->A | 278 | A->C | 374 | (-0.024, -0.008) |
| | G->T | 276 | T->G | 345 | (-0.020, -0.005) |
| | T->A | 480 | A->T | 484 | no sign difference |
| | G->C | 240 | C->G | 232 | no sign difference |
| CB4858 Base sub | N2->CB4858 | Base Sub | N2->CB4858 | Base Sub | 95% Bayesian Interval for difference using MCMC |
| CB4858 Transitions | G->A | 5266 | A->G | 5375 | no sign difference |
| | C->T | 5163 | T->C | 5341 | no sign difference |
| CB4858 Transversions | C->A | 1871 | A->C | 1945 | no sign difference |
| | G->T | 1906 | T->G | 1893 | no sign difference |
| | T->A | 3185 | A->T | 3240 | no sign difference |
| | G->C | 964 | C->G | 989 | no sign difference |

### Table 4 SNP rate by chromosome in CB4856 and CB4858

| Chrome | CB4856 | CB4858 |
|--------|--------|--------|
|        | SNPs   | SNP rate | SNPs   | SNP rate |
| I      | 1752   | 0.002711 | 4584   | 2.98E-04 |
| II     | 2435   | 0.003373 | 16322  | 0.001047296 |
| III    | 1335   | 0.00207  | 8701   | 0.0006215 |
| IV     | 2093   | 0.002544 | 413    | 1.9444E-05 |
| V      | 3620   | 0.003754 | 3311   | 0.000157667 |
| X      | 2107   | 0.002466 | 8868   | 0.000492667 |

### Table 5 Comparison of arms vs core SNP rates in CB4856 and CB4858*

| SNP rate comparison | CB4856 p-value | CB4858 p-value |
|---------------------|----------------|----------------|
| overall arms vs core across chromosomes | <2.2e-16 | <2.2e-16 |
| arms across chromosomes | <2.2e-16 | <2.2e-16 |
| core across chromosomes | <2.2e-16 | <2.2e-16 |
| arms vs core in each chromosome | | |
| I | <2.2e-16 | <2.2e-16 |
| II | <2.2e-16 | <2.2e-16 |
| III | <8.918e-06 | <2.2e-16 |
| IV | <2.2e-16 | 0.5464 |
| V | 1.778e-14 | 2.2e-16 |

*Based on chi-squared tests
chromosome [21]. Furthermore, sex is determined by X chromosome dosage and this dosage is compensated in hermaphrodites [25]. Observations of correlation between SNP rate and recombination on the autosomes in both CB4856 and CB4858 have been proposed to either reflect an inherent mutagenic effect of recombination or conditions where selection results in the loss (negative selection) or fixation (positive selection) of new mutations along with all linked polymorphisms. The data in this study shows us that in the autosomes, the observed levels of polymorphisms are positively correlated with levels of recombination in both isolates.

Using the chromosomal domains in Rockman and Kruglyak [12], we found that 1) in the arms, the recombination rate is negatively correlated with chromosome size, although the correlation was not significant (p = 0.303003); 2) the two tips of all autosomes and the X chromosome have an effectively zero recombination; 3) the *C. elegans* autosomes and the X chromosome have nearly constant recombination rates within each

Table 6 Relationship between SNP rate and recombination rate in CB4856 and CB4858*

| chromosome  | CB4856 p-value | CB4858 p-value |
|-------------|----------------|----------------|
| autosomes   | 0.023          | 0.0423         |
| X chromosome| 0.6458         | 0.4321         |

*p-values found doing Spearman’s rho test for correlation

---

Figure 1 SNP rates vs. Recombination Rates for CB4856 and CB4858
chromosomal domain (the slope of the recombination rate over the chromosomal domains is 0). Findings 2) and 3) agree with the findings in Rockman and Kruglyak [12] but finding 1) is different from the finding in Rockman and Kruglyak [12].

CB4856 Hotspot analysis
The level of SNP variation across the genome in any comparison reflects two possible mechanisms. A high frequency of local polymorphisms can be a consequence of high localized mutation rate or a mutational hotspot. Alternatively, localized patterns of divergence can reflect the age of the allele and historic patterns of recombination. As a consequence of recombination and inter-breeding of the ancestors that gave rise to these isolates, each allele can have a unique history of coalescence.

To further investigate the pattern of variation across the C. elegans genome and specifically identify regions of significantly high localized polymorphisms, we identified clusters of SNPs with a significant number of polymorphisms (Additional File 3). In general, these hotspots are DNA segments that have an unusual number of polymorphisms considering the length of the DNA (in bp). A more detailed description of the hotspots is given in the Methods section. These hotspots of polymorphisms may be mutational hotspots or may indicate old alleles. We excluded the CB4858 datasets due to the inherent inability of short reads to represent loci with highly clustered polymorphisms.

The summary statistics of the hotspots per chromosome are shown in Table 7. The optimal window size for the initial cluster analysis was constrained by the average single read length in the CB4856 comparisons (571.14 bp). Statistically significant hotspot clusters ranged from 49 bp to 628 bp in length (Table 8). There were 126 clusters when comparing N2 against CB4856 and the average number of events per hotspot was 17.59. In general, the proportion of base substitutions was the largest in Chromosome V, the proportion of nonsequential indels was the largest in chromosome X, and the proportion of sequential indels was the largest in chromosome IV. There was a significant positive relationship between hotspot length (in bp) and the number of events in the hotspot (p-value = 2.2e-16). Chromosome V has the largest number of hotspots while Chromosome X had the lowest number. We did a linear regression with dummy variables to test if the location of SNPs (arms vs core) was a significant predictor of the number of hotspots. We saw no significant hotspot distributional biases with respect to chromosomal arm/core boundaries (p-value = 0.9618, Figures 2, 3 and 4).

The sizes of the hotspots vary per chromosome. We categorized the hotspots into small and large hotspots based on a cutoff of 100 bp in length. The number of small vs large hotspots is in Table 9. The hotspots with lengths less than 100 base pairs have on average 10.76 of events and 15% of the events are base substitutions while 79% of the events are nonsequential indels. On the other hand, the hotspots with lengths more than 100 base pairs have on average 18.97 events and 58% of the events are base substitutions while 30% of the events are nonsequential indels.

In some taxa, it has been shown that the rate of mutation at specific loci is strongly correlated with GC content [26-31]. To test this possibility, we assayed the GC content in all clusters and compared the %GC content in the clusters to the N2 genome as a whole (%GC).

There is sufficient evidence to indicate the %GC content in the statistical hotspots differs from the %GC content in the N2 genome as a whole (p-value <2.2e-16). The %GC content in different statistical hotspots is displayed in Additional File 4. Some of the hotspots have a higher %GC content than N2 and some of them have a lower %GC content than N2.

We also compared these hotspots of polymorphism to mutations identified in a genome wide analysis of N2 mutation accumulation lines [20]. Denver et al [20] found a random distribution of mutations in their MA lines. The 391 mutations identified among ten N2 derived MA lines did not fall within any of the hotspots identified here. Finally, in a comparison of CB4856 and CB4858, there were 37 SNPs that were shared between

Table 7 Summary Statistics of CB4856 Statistical Hotspots per Chromosome

| Chromosome | Hotspots | Number of hotspots | Average Number of bases | Average Number of Events | Proportion of Base Substitutions | Proportion of Nonsequential Indels | Proportion of Sequential Indels | Proportion of Inversions |
|------------|----------|--------------------|-------------------------|-------------------------|---------------------------------|-----------------------------------|-----------------------------|-------------------------|
| I          | 10       | 186.1              | 16                      | 0.47                    | 0.43                            | 0.09                              | 0.01                        |                         |
| II         | 32       | 223.59             | 18.03                   | 0.54                    | 0.35                            | 0.09                              | 0.01                        |                         |
| III        | 10       | 150.9              | 11.3                    | 0.36                    | 0.54                            | 0.09                              | 0.01                        |                         |
| IV         | 13       | 180.54             | 13.08                   | 0.32                    | 0.57                            | 0.11                              | 0.00                        |                         |
| V          | 52       | 292.96             | 20.67                   | 0.65                    | 0.26                            | 0.09                              | 0.01                        |                         |
| X          | 9        | 145.33             | 13.56                   | 0.19                    | 0.72                            | 0.08                              | 0.01                        |                         |
| Total      | 126      | 233.44             | 17.59                   |                         |                                 |                                   |                             |                         |
Table 8 Distribution of CB4856 Hotspots

|                        | Mean       | Standard Deviation | Minimum | Maximum |
|------------------------|------------|--------------------|---------|---------|
| Length of hotspot (bp) | 233.444444 | 138.2597607        | 49      | 628     |
| Number of Events       | 17.5952381 | 9.744067792        | 7       | 55      |
| Proportion of Base substitutions | 0.51654754 | 0.38545355        | 0       | 1       |
| Proportion of non sequential indels | 0.38553057 | 0.381410192      | 0       | 1       |
| Proportion of sequential indels | 0.08892548 | 0.095726976      | 0       | 0.705882|
| Proportion of Inversions | 0.00899641 | 0.027289781      | 0       | 0.181818|

Figure 2 SNP rate per Mb for Chromosomes I and II in CB4856 and CB4858

Solorzano et al. BMC Evolutionary Biology 2011, 11:168
http://www.biomedcentral.com/1471-2148/11/168
Page 8 of 15
CB4856 and CB4858 (Additional File 5), which is larger than the expected number of SNPs shared between CB4856 and CB4858 (approximately 7).

We hypothesize that the hotspots of polymorphisms in this study are alleles that are shared between CB4856 and CB4858 but not by N2 and are not mutational hotspots because (1) the %GC content in the hotspots differs from the %GC content in N2 and (2) because the SNPs found in CB4856 vs N2 were not in the MA lines.

Distribution of SNPs by function
To investigate the distribution of SNPs across major functional partitions in the C. elegans genome, we identified the putative functional role of each SNP position based on the annotation of the N2 reference genome (Table 10). The general variation in both CB4856 and CB4858 follows expected patterns. The functional categories of SNPs are ordered by Intergenic>Introns>UTR>Exons in CB4856 and Intergenic>UTR>Exons>Introns (Table 11).
functional categories of SNPs are ordered as expected for CB4856 but for CB4858, where the frequency of SNPs in exons is higher than in introns. CB4858 may have relaxed selection constraints on coding mutations (smaller population size).

**CB4856/CB4858 neutrality**

The $k_a/k_s$ ratio is the rate of nonsynonymous substitutions ($k_a$) to the rate of synonymous substitutions ($k_s$), which can be used as an indicator of selective pressure.

**Table 9 Total Number of Small and Large Hotspots in CB4856 per Chromosome**

| Chromosome | Small (<100 bp) | Large (> 100 bp) |
|------------|----------------|-----------------|
| I          | 2              | 8               |
| II         | 5              | 27              |
| III        | 1              | 9               |
| IV         | 3              | 10              |
| V          | 8              | 44              |
| X          | 2              | 7               |

Figure 4 SNP rate per Mb for Chromosomes V and X in CB4856 and CB4858.
acting on a protein-coding gene. A value of \(k_a/k_s > 1\) indicates positive selective pressure whereas a value of \(k_a/k_s = 1\) indicates neutral evolution. Finally, a value of \(k_a/k_s < 1\) indicates pressures to conserve amino acid sequence [32]. We test for neutrality by testing the null hypothesis of \(k_a/k_s = 1\) in both CB4856 and CB4858 [33]. Analysis of genome-wide SNP patterns reveal a \(k_a/k_s\) of less than 1 for all chromosomes in both CB4856 and CB4858 in Table 12, adjusting for multiple comparisons [34]. This result is consistent with the role that purifying selection has on SNPs in most protein coding regions.

In CB4856, the autosomes have \(k_a/k_s\) values that are not significantly different from X in either direction (p-value = 0.3924). In addition, when considering all of the chromosomes, the \(k_a/k_s\) values are not significantly different in the different chromosomes in either direction (p-value = 0.2343). However, the SNP rate is lower in chromosome X (0.002406) compared to the autosomes (0.00289) in CB4856 (p < 2.2e-16).

Differences in the length or the number of cell divisions in the male germline compared to the hermaphrodites in CB4856 may explain the lower SNP rate in the

| Table 10 Frequencies of SNPs by function and chromosome |
|---------------------------------------------------------|
|             | I       | II      | III     | IV      | V       | X       |
| Exon        |         |         |         |         |         |         |
| Insert      | 9,942   | 91      | 16,423  | 334     | 10,263  | 11679   |
| Deletion    | 13825.97| 87.05   | 17536.07| 320.82  | 13684   | 19126.7 |
| bs          | 19107.81| 4110.87 | 41335.02| 21231.21| 19126.52| 11330.5 |
| syn         | 9165.53 | 2406.68 | 19484.52| 1092    | 7775.01 | 6780.87 |
| Total       | 42876.05| 4279.99 | 75293.76| 21866.6 | 43073.55| 11679.1 |

| Intron      |         |         |         |         |         |         |
| Insert      | 26409.16| 435.8   | 20319.57| 8397.01 | 23176.5 | 5723.94 |
| Deletion    | 11806.45| 305.71  | 22268.03| 648.06  | 26590.53| 192.04  |
| bs          | 19263.15| 5281.69 | 31731.94| 14167.6 | 14306.02| 13648.8 |
| syn         | 9165.53 | 2406.68 | 19484.52| 1092    | 7775.01 | 6780.87 |
| Total       | 57478.76| 6023.21 | 74319.54| 15598.4 | 49293.56| 12172.2 |

| UTR         |         |         |         |         |         |         |
| Insert      | 11495.75| 699.99  | 21850.5  | 1107.01 | 1322.96 | 8356.29 |
| Deletion    | 24079    | 17.65   | 23325    | 1046    | 15487   | 16548   |
| bs          | 23457.55 | 9997.48 | 25747.41 | 18421.7 | 15239.02 | 2634.8 |
| syn         | 9165.53 | 2406.68 | 19484.52 | 1092    | 7775.01 | 6780.87 |
| Total       | 59032.24 | 7584.3  | 65273.15 | 50071.06| 49293.56| 12172.2 |

| IG          |         |         |         |         |         |         |
| Insert      | 23923.59| 1144.8  | 28670.08 | 3054.24 | 11351.51| 1856.42 |
| Deletion    | 59032.24| 787.05  | 44814.4  | 2771.91 | 23791.53| 15932.5 |
| bs          | 34021.21| 9997.48 | 54556.66 | 39538.3 | 32966.04| 31558   |
| syn         | 74923.59| 1144.8  | 28670.08 | 3054.24 | 11351.51| 1856.42 |
| Total       | 116977.1| 11929.33| 128041.2 | 45364.4 | 68109.09| 23870.2 |

The numbers represent isolate × 10^8 SNPs per base. *For ease in comparison of frequencies across categories and strains, we divided the number of polymorphisms by the number of analyzed sites and multiplied that number by 10^8. For example, the proportion of insertion polymorphisms in CB4856 is 9942*10^-8.

| Table 11 Frequencies SNPs by function in CB4856 and CB4858 across the genome |
|-------------------------------------------------------------|
| CB4856          | CB4858          |
| Type             | SNPs            | Type             | SNPs            |
| Exon             | 302784.89       | Exon             | 445124.9        |
| UTR              | 317184.6        | UTR              | 469319.8        |
| Intrin           | 348234.62       | Intrin           | 565948.8        |
| IG               | 745269.1        | IG               | 1152932.7       |

The numbers represent isolate × 10^8 SNPs per base. *For ease in comparison of frequencies across categories and strains, we divided the number of polymorphisms by the number of analyzed sites and multiplied that number by 10^8.

In CB4856, the autosomes have \(k_a/k_s\) values that are not significantly different from X in either direction (p-value = 0.3924). In addition, when considering all of the chromosomes, the \(k_a/k_s\) values are not significantly different in the different chromosomes in either direction (p-value = 0.2343). However, the SNP rate is lower in chromosome X (0.002406) compared to the autosomes (0.000289) in CB4856 (p < 2.2e-16).

Differences in the length or the number of cell divisions in the male germline compared to the hermaphrodites in CB4856 may explain the lower SNP rate in the

| Table 12 Mean k-values and p-values testing if k_a/k_s = 1 in each chromosome* |
|-----------------------------|------------------------|------------------------|
| CB4856                      | CB4858                 |
| Chrom                      | k_a/k_s                | p-value for k_a/k_s = 1 |
| I                          | 0.41157                | 0.225                  |
| II                         | 0.5746                 | 0.321                  |
| III                        | 0.596                  | 0.238                  |
| IV                         | 0.5443                 | 0.1745                 |
| V                          | 0.4856                 | 0.3297                 |
| X                          | 0.4261                 | 0.2984                 |

*All k_a/k_s values in both isolates are significantly smaller than 1.
X chromosome [35]. When the X chromosome is passed through the male germline, it is in a hemizygotic state and exposed to increased purifying selection [35]. In addition, the mitotic cell cycle length has been shown to be shorter in males than hermaphrodite germ cells [36]. Taken together, both of these factors can contribute to a lower SNP rate in the X chromosome than in the autosomes in CB4856. These observations are dependent upon how often males are involved in reproduction from generation to generation. While direct observations of C. elegans populations show a low frequency of males and low levels of heterozygosity, it is possible that some lineages of Caenorhabditis have historically undergone much more outbreeding involving males [37].

In CB4858, the autosomes have ka/ks values (0.2841) that are significantly smaller than X (0.2984) on the average (p-value = 0.0014). Unlike CB4856, the different chromosomes have significantly different k_s/k_a values (p-value = 7.176e-05, Table 13). It is not appropriate to consider the differences in SNP rates between chromosomes in CB4858 due to the long stretches of DNA shared by N2 and CB4858.

Because the k_s/k_a ratio in the X chromosome is higher than the autosomes on average in CB4858, purifying selection appears to be reduced on the X chromosome in this strain compared to CB4856. The relative difference in ka/ks ratios in CB4858 could reflect differences in the life history of CB4858. We suggest that males (XO) could contribute to a smaller degree in natural populations of CB4858 than in CB4856 populations, resulting in a lower rate of protein sequence evolution for X-linked mutations in the Hawaiian isolate CB4856. Previous studies have presented varying levels of the genetic contribution of males in different natural isolates of Caenorhabditis elegans [5,6,38,39].

Conclusions
This study provides a large scale genomic portrait of the C. elegans polymorphisms in CB4856 and CB4858. The patterns observed here give us insight into the evolutionary spectrum of the Caenorhabditis lineage. There is a clear bias of deletions over insertions in CB4856, suggesting two possibilities. The first one is that the CB4856 genome is becoming smaller. Other studies have argued that possible reasons that genomes could become smaller in general are gene loss, natural selection, or the occurrence of long terminal repeat (LTR) sequences of transposable elements [40]. The second possibility is that the N2 genome is getting larger because it fails to purge insertions. Interestingly, direct sequencing of C. elegans insertion deletion rates in mutation accumulation lines [19] observe a bias towards insertions. Taken together, we conclude that CB4856 is evolving under different mutation or selection pressures than N2 or CB4858.

Our results show that in both CB4856 and CB4858, SNP rates are higher in the chromosome arms than in the chromosome cores in agreement with previous studies [1]. Direct genome sequencing of C. elegans mutation accumulation lines [20] show that the mutation rate is not higher in arm regions compared to the core regions. We hypothesize SNP rates differ in the core regions of N2, CB4856, and CB4858 in our study due to higher levels of purifying selection in the core regions.

We found several statistically significant hotspots of polymorphisms in our CB4856-N2 comparison. Some of the CB4856 hotspots contained significantly higher levels of G/C when compared to the genome average, consistent with observations of higher rates of deamination at methylated cytosines [26-28,30,31]. None of these hotspots coincided with C. elegans mutations identified in Denver et al. [20], and are likely to be older alleles shared between CB4856 and CB4858 but not N2.

The ka/ks ratios in both CB4856 and CB4858 are consistent with the effect of purifying selection on SNPs in most protein coding regions. The differences in ka/ks in the autosomes vs the X chromosome in CB4858 suggest selection may be more efficient for X-linked genes. In addition, these differences may be attributed to male-driven evolution of the X chromosomes. We hypothesize that natural populations of CB4858 strains have fewer males than CB4856 populations, resulting in a lower rate of protein sequence evolution for X-linked mutations. The finding in CB4858 of a significant difference in selection on the X chromosome compared to the autosomes is consistent with a biological role for males in the lineage giving rise to CB4858, consistent with previous published results [5,6,38,39].

### Table 13 CB4858 chromosomes whose ka/ks values are significantly different and their p-values

| Chromosomes | p-value |
|-------------|---------|
| II vs IV    | 0.0004736 |
| II vs V     | 0.0036938 |
| III vs IV   | 0.00056427 |
| III vs V    | 0.033124   |
| IV vs X     | 0.032036   |

Methods
Pipeline for analysis of CB4856 data
The reads for CB4856 were trimmed for quality using the default parameters of LUCY [41], and individually assembled at 90% identity using the AMOS [42] reference assembler against the WS170 C. elegans assembly. To reduce the amount of false calls based upon paralogy, each of the polymorphic sites and 70 flanking base pairs was extracted from the reference. These sequences
were BLASTed reciprocally at E-5 (~85%) against the genome, and any position that matched multiple sites in the genome was removed from the analysis. The CB4858 base calls followed an identical procedure to remove false polymorphic calls based upon paralogy.

Due to the low coverage of CB4856 reads, it was necessary to set a stringent cutoff of 40 sanger quality score (.0001 base call error rate). CB4856 deletions were scored based on the lowest neighboring quality score. The coding context of both CB4856 and CB4858 polymorphisms were determined using the WS170 General Feature Format (gff).

**Discovery of the statistical hotspots of polymorphisms in CB4856**

**Overall procedure**

We did a three step procedure to finalize our statistical hotspots. First, we found initial clusters of SNPs. Then, we decided on an optimal window size to consider for our preliminary hotspots. Finally, we used Gibbs sampling to determine the final statistically significant hotspots.

**Initial Clusters of SNPs**

We initially considered windows of SNPs ranging from 50 bp to 600 bp in increments of 50 bps. For each window, we calculated the mean and standard deviation of the number of SNPs. We then calculated the mean + 3 standard deviations. If there were SNPs outside the window whose location was within 3 standard deviations of the mean, then those SNPs belonged to this cluster. Otherwise, the first SNP which was outside the mean + 3 standard deviations was the beginning of the next cluster. We did this by considering all of the SNPs between N2 and CB4856.

**Decision on the Size of the Window of the Initial Clusters**

A window size of 300 bp yielded the highest number of initial clusters. Assume the number of events in the initial cluster (X) is distributed as Poisson with mean λ = 2.305315 (the average number of events in the 300 bp clusters). To find the total number of events required to be a statistical hotspot, we find number xo such that P(X ≥ xo) = .0015. This signifies the number of events in the initial cluster (xo) is an outlier and a putative hotspot. The number .0015 corresponds to an event which is multiple standard deviations above the mean in our Poisson distribution. The number of events required to be a statistical hotspot (xo) in a given bp region is in Additional File 6.

**Gibbs Sampling Methodology**

Gibbs Sampling is an algorithm used to generate a sequence of samples from the joint probability distribution of multiple random variables. It is used as a means of statistical inference. In this work, we used Gibbs Sampling to find posterior confidence intervals to detect statistical hotspots and to detect differences between base substitution rates. Gilks et al. [43] contains a useful description of Gibbs Sampling.

**Gibbs Sampling Methodology to find Final Statistically Significant Hotspots**

Let X_i = total number of events in any 300 bp window i, \( \theta_i \) = the event rate for window i, and \( t_i \) = total number of bp between the first event and the last event in a 300 bp window. Then, \( X_i \sim \text{Poisson}(\theta_i t_i) \) for \( i = 1, ..., n \), where n is the total number of 300 bp windows being considered. We assume a conjugate gamma prior distribution for the event rate \( \theta_i \sim \text{Gamma}(\alpha, \beta) \) for \( i = 1, ..., n \). In addition, we assume the prior specification for the hyperparameters as \( \alpha \sim \text{Exponential}(1.0) \) and \( \beta \sim \text{Gamma}(0.08,1.0) \). We find the posterior distribution of all \( \theta_i \mid x_i, \alpha, \beta \) for \( i = 1, ..., n \) to determine the statistical hotspots and construct 95% credible intervals for the rate of events in a region (\( \lambda = \theta_i t_i \)). If the lower bound of the 95% credible interval for \( \theta_i t_i \) is above the number \( x_o \) (in Additional File 6), the initial cluster in window i is a statistical hotspot. We do this simultaneously for all initial clusters.

**Gibbs Sampling Methodology to find the Difference in the Direction of Base Substitution Rates in CB4856-N2 and CB4858-N2 comparisons**

The model to test the difference between base substitution rates is described below and is reasonable given the large number of bases assayed in both CB4856 and CB4858. Let \( \theta_i \) represent the true difference between base substitution rates and \( y_j \) represent the sample difference between base substitution rates. For example, \( \theta_i \) represents the difference between the G → A and A → G base substitution rates. Then, \( y_{ij} \mid \theta_i \sim \text{Normal}(\theta_i, \sigma^2) \) for \( j = 1,...,6 \) for the 6 possible base direction changes. The prior distribution \( \theta_i \sim \text{Normal}(\mu, \tau^2) \) for \( \theta_i \). Finally, we have an improper noninformative hyperprior distribution \( \text{p}(\mu, \tau^2) \sim \text{Normal}(\mu, \tau^2) \). We can obtain the posterior confidence intervals for the direction of base substitution rates given in Table 3. If the posterior confidence intervals do not contain 0, then we conclude there is a difference in the direction of base substitution rates in the isolate considered.

**Polymorphism Confirmation**

Polymorphic loci identified by pairwise comparisons between N2 and CB4856 were selected at random and satisfied the criteria of two or more polymorphic loci within 300-500 bp. Forward primers were designed specific to CB4856. PCR amplification of the 300-500 bp fragments containing the candidate polymorphic base(s) were directly sequenced in the forward direction using standard fluorescent sequencing technology on an ABI3130 Genetic Analyzer at the University of New Hampshire Hubbard Center for Genome Studies. We
found 51 out of 55 polymorphisms were present in the CB4856 genome. The rate of confirmation was 93%.

Additional material

Acknowledgements
ES thanks Abraham E. Tucker for his genetics insight and helpful discussions. ES also thanks Sanjuro Jogdeo for his bioinformatics insight and helpful discussions. This work was supported by NSF IGMS Grant 0527342.

Author details
1 Hubbard Center for Genome Studies, University of New Hampshire, Durham, NH, US. 2 Department of Molecular, Cellular, and Biomedical Research, University of New Hampshire, Durham, NH, US. 3 Department of Computer Science, University of New Hampshire, Durham, NH, US. 4 Department of Decision Sciences, University of New Hampshire, Durham, NH, US.

Authors’ contributions
ES was responsible for the statistical analysis. PD and WS were responsible for the computational techniques necessary to summarize the data. KO was responsible for the biological confirmation of SNPs. RDB was responsible for guiding the bioinformatics aspects of this project. WKT was responsible for guiding the genetics implications of this project. All authors read and approved the final manuscript.

Received: 10 December 2010 Accepted: 16 June 2011
Published: 16 June 2011

References
1. Stein LD, Bao Z, Blaisier D, Blumenthal T, Brint MR, Chen N, Chinwalla A, Clarke J, Clew C, Coghlan A, Coulon A, D’Eustachio P, Fitch DHA, Fulton LA, Fulton RE, Griffiths-Jones S, Harris TW, Hillier LW, Kamath R, Kuwabara PE, Mardis ER, Maia MA, Miner TL, Minx P, Mullikin JC, Plumb RW, Rogers J, Schein JE, Sohmmann M, Speth J, Stajch JE, Wilce C, Willey D, Wilson RK, Durbin R, Waterston RH. The Genome Sequence of Caenorhabditis briggsae: A Platform for Comparative Genomics. PLoS Biol 2003; (12):166-192.
2. Thomas WK, Wilson AC. Mode and tempo of molecular evolution in the nematode Caenorhabditis: cytochrome oxidase i and calmodulin sequences. Genetics 1991, 128:269-279.
3. Eglmez-NK, Ebert RH, Shmooler Reis RJ. Strain evolution in Caenorhabditis elegans: transposable elements as markers of interstrain evolutionary history. J Mol Evol 1995, 40:372-381.
4. Hodgkin J, Donnach T. Natural variation and copulatory plug formation in Caenorhabditis elegans. Genetics 1997, 146:149-164.
5. Sivasundar A, Hey J. Sampling from Natural Populations with RNAI Reveals High Outcrossing and Population Structure in Caenorhabditis elegans. Current Biology 2005, 15(17):1596-1602.

5. Sivasundar A, Hey J: Sampling from Natural Populations with RNAI Reveals High Outcrossing and Population Structure in Caenorhabditis elegans. Current Biology 2005, 15(17):1596-1602.

6. Barriere A, Felix M: High Local Genetic Diversity and Low Outcrossing Rate in Caenorhabditis elegans Natural Populations. Current Biology 2005, 15(13):1176-1184.
7. Barriere A, Felix M: Temporal Dynamics and Linkage Disequilibrium in Natural C. elegans Populations. Genetics 2007, 176:999-1011.
8. Cutter AD, Felix MA, Barriere A, Charlesworth D: Patterns of nucleotide polymorphism distinguish temperate and tropical wild isolates of Caenorhabditis briggsae. Genetics 2006, 173(4):2021-31.
9. Koch R, van Leeuwen HG, van der Horst M, Thijsen K, Plasterk RH: Single nucleotide polymorphisms in wild isolates of Caenorhabditis elegans. Genome Res 2003, 10:1690-1696.
10. Wicks SR, Yeh RT, Gosh WR, Waterston RH, Plasterk HA: Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map. Nature Genetics 2001, 28:160-164.
11. Hillier LW, Marshall GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock Jr, Hickenbotham M, Huang W, Magnini V, Rixt R, Sander SN, Stewart DA, Stromberg M, Tsung EF, Wylie T, Shedel T, Wilson RK, Mardis ER: Whole-genome sequencing and variant discovery in C. elegans. Nature Methods 2008, 5:183-188.
12. Rockman MV, Kruglyak L: Recombinational Landscape and Population Genomics of Caenorhabditis elegans. PLoS Genetics 2009, 5(3):1-16.
13. Denver DR, Lynch M, Thomas WK: High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. Letters to Nature 2004, 430:679-682.
14. Swan KA, Curtis DE, McKusick KB, Voinov AV, Mapa FA, Cancilla MR: High-Throughput Gene Mapping in Caenorhabditis elegans. Genome Research 2002, 12:1001-105.
15. Cutter AD, Poyser BA: Selection at Linked Sites in the Partial Selfer Caenorhabditis elegans. Molecular Biological Evolution 2003, 20(5):663-673.
16. Lynch M: The Origins of Genomic Architecture. Sunderland MA: Sinauer Associates Inc, 2007.
17. Weber RP, De S, Kozareva I, Turner DJ, Babu WM, DeBono M: Whole Genome Sequencing Highlights Genetic Changes Associated with Laboratory Domestication of C. elegans. PLoS One 2010, 5(11):1-10.
18. Graustein A, Gaspar JM, Walters JR, Palopoli MF: Levels of DNA Polymorphism Vary With Matting System in the Nematode Genus Caenorhabditis. Genetics 2002, 161:99-107.
19. Jovelin R, Sung FS, Phillips PC: High Nucleotide Diversity in Developing Reproductive Genes Contrasts With the Structural Elements of Offactory Pathways in Caenorhabditis. Genetics 181:1387-1397.
20. Barnes TM, Kohara Y, Coulson A, Hello S: Metic recombination, noncoding DNA and genomic organization in Caenorhabditis elegans. Genetics 1995, 141(1):158-79.
21. Denver DR, Dolan PC, Wilhelm LW, Sung W, Lucas-Lledo JI, Hower DK, Lewis SC, Okomoto K, Thomas WK, Lynch M, Baer CF: A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. PNAS 2009, 106(38):16310-16314.
22. Denver DR, Morris K, Thomas WK: Phylogenetics in Caenorhabditis elegans: An analysis of divergence and outcrossing. Mol Biol Evol 2003, 20:393-400.
23. Cutter AD, Dey A, Murray RL: Evolution of the Caenorhabditis elegans Genome. Molecular Biology and Evolution 2009, 26(6):1199-1234.
24. Kelly WJ, Schaner CE, Demburg AF, Lee M, Kim SK, Welleneere AM, Reinke V: X-chromosome silencing in the germline of C. elegans. Development 2002, 129:479-492.
25. Hodgkin J: X-chromosome dosage and gene expression in Caenorhabditis elegans: Two unusual dumpy genes. Mol Gen Genet 1983, 192:452-458.
26. Li WH: A model for the correlation of mutation rate with GC content and the origin of GC-rich isochors. J Mol Evol 1987, 24:468-475.
27. Fullerton SM, Carvalho AB, Clark AG: Local Rates of Recombination are Positively Correlated with GC Content in the Human Genome. Molecular Biology and Evolution 2001, 18(6):1139-1142.
28. Filipski J: Correlation between molecular clock ticking, codon usage fidelity of DNA repair, chromosome banding and chromatin compactness in germine cells. FEBS Lett 1987, 217:184-186.
29. Suroka N: Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci 1988, 85:2653-2657.
30. Wolfe KH, Sharp PM, Li WH: Mutation rates differ among regions of the mammalian genome. Nature 1989, 337:283-285.
31. Francino MP, Ochman H: A comparative genomics approach to DNA asymmetries. Ann NY Acad Sci 1999, 870:428-431.
32. Kosakovsky Pond SL, Frost SDW: Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Molecular Biology and Evolution 2005, 22(5):1208-1222.
33. Zhang Z, Li J, Zhao X, Wang J, Wong G, Yu J: KaKs Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. Genomics, Proteomics & Bioinformatics 2006, 4(4):259-263.
34. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series (Methodological) 1995, 57(1):289-300.
35. Vicoso B, Charlesworth B: Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet 2006, 7:645-653.
36. Morgan D, Crittenden SL, Kimble J: The C. elegans adult male germline: Stem cells and sexual dimorphism. Developmental Biology 2010, 346:204-214.
37. Felix M, Braendle C: The natural history of Caenorhabditis elegans. Current Biology 2010, 20(R965-R969).
38. Mayden JS, Lorch A, Edgley ML: Copy number variation in the genomes of twelve natural isolates of Caenorhabditis elegans. BMC Genomics 2010, 11:12.
39. Haber M, Schungel M, Putz A, Müller S, Hasert B, Schulenburg H: Evolutionary History of Caenorhabditis elegans Inferred from Microsatellites: Evidence for Spatial and Temporal Genetic Differentiation and the Occurrence of Outbreeding. Molecular Biology and Evolution 2005, 22(1):160-173.
40. Ridley M: Evolution Malden MA: Wiley-Blackwell; 2003.
41. Chou H, Holmes MH: DNA sequence quality trimming and vector removal. Bioinformatics 2001, 17(12):1093-1104.
42. Pop M, Philippsen P, Delcher A, Salzberg SL: Comparative Genome assembly. Briefings in Bioinformatics 2004, 5(3):237-248.
43. Gilks WR, Richardson S, Spiegelhalter D: Markov Chain Monte Carlo in Practice. Boca Raton, FL: Chapman and Hall/CRC, 1995.

doi:10.1186/1471-2148-11-168
Cite this article as: Solorzano et al.: Shifting patterns of natural variation in the nuclear genome of caenorhabditis elegans. BMC Evolutionary Biology 2011 11:168.