Extremely high molecular diversity within the East Asian nematode Caenorhabditis sp. 5

GUO-XIU WANG,* SHUANG REN,* YUAN REN,* HUI AI* and ASHER D. CUTTER†‡

*Hubei Key Laboratory of Genetic Regulation and Integrative Biology, HuaZhong Normal University, Wuhan, Hubei, China, †Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada, ‡Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON M5S 3B2, Canada

Abstract

Most relatives of the self-fertilizing hermaphroditic nematode model organism Caenorhabditis elegans reproduce via obligate outbreeding between males and females, which also represents the ancestral mode of reproduction within the genus. However, little is known about the scope of genetic diversity and differentiation within such gonochoristic species, especially those found outside of temperate Europe and North America. It is critical to understand the evolutionary processes operating in these species to provide a framework for deciphering the evolution of hermaphroditism and a baseline for the application of outcrossing Caenorhabditis to problems in evolutionary genetics. Here, we investigate for the first time molecular sequence variation for Caenorhabditis sp. 5, a species found commonly in eastern Asia. We identify enormous levels of standing genetic variation that approach the levels observed in the marine broadcast-spawning sea squirt, Ciona savignyi. Although we document significant isolation by distance, we demonstrate that the high polymorphism within C. sp. 5 is not because of strong differentiation among populations or to the presence of cryptic species. These findings illustrate that molecular population genetic approaches to studying obligately outbreeding species of Caenorhabditis will prove powerful in identifying and characterizing functionally and evolutionarily important features of the genome.

Keywords: Caenorhabditis, nucleotide variation, phylogeography

Received 30 June 2010; revision received 30 August 2010; accepted 3 September 2010

Introduction

The Caenorhabditis genus of nematodes is the focus of a recently burgeoning literature on comparative developmental biology, evolution and genetics (Kammenga et al. 2008; Haag 2009). Much of this work is inspired by the experimental tractability of the classic nematode animal model, C. elegans—but a protracted question surrounds the generality of inference to be made from molecular and phenotypic variation found in C. elegans. Specifically, the derived breeding system of C. elegans, in which populations primarily comprise self-fertilizing hermaphrodites, exerts profound consequences for much of the organism’s biology (Cutter et al. 2009), despite the conveniences of laboratory manipulation imparted by selfing hermaphroditism. A major unsolved issue needed to fully characterize the evolution and ecology of Caenorhabditis nematodes—and C. elegans’ place within this group—is to better understand species with the ancestral, gonochoristic (=dioecious, male-female) breeding system. At present, analysis of phylogeographical, population genetic and phenotypic variation for gonochoristic Caenorhabditis is known reasonably well only for C. remanei in this group that contains over 30 species (Kiontke & Fitch 2005; Kammenga et al. 2008; Cutter et al. 2009; K. Kiontke personal communication). In this study, we characterize for the first time molecular genetic variation in a broad collection of the gonochoristic species C. sp. 5—which appears to be restricted in distribution to eastern Asia—to help fill this gap in our understanding of outbreeding taxa that represent the ancestral mode of reproduction for Caenorhabditis.
Despite the exceptional insights into biological processes permitted by studying highly selfing species of *Caenorhabditis*, several important problems are impossible to investigate in such species. For example, sexual selection is thought to be largely absent in natural populations of *C. elegans* and *C. briggsae*, yet is probably an important evolutionary force in obligately outbreeding species (LaMunyon & Ward 1999; Cutter 2008c). Investigation of the functional and molecular evolutionary consequences of male–male sperm competition and the potential for intersexual antagonistic coevolution will be best dissected in outbreeders that currently experience such selection in nature. Moreover, data from *C. remanei* (and from *C. sp. 5* presented in this study) indicate that segregating genetic variation and effective population size (*N_e*) are dramatically greater than in selfing species (Cutter et al. 2009). This has important implications for detecting and quantifying adaptation, the evolutionary response to even very weak natural selection and other population genetic processes in the absence of complex demographic population structure and genome-wide linkage disequilibrium (LD) induced by a selfing lifestyle. A variety of molecular population genetic methods have been developed for such inference, and their statistical power is determined in-part by the extent of genetic variation and linkage, which is problematic for their application to *C. elegans* (Kreitman 2000; Boffelli et al. 2004; Nielsen 2005).

*C. sp. 5* is an obligately outbreeding species discovered in China independently by Walter Sudhaus and Marie-Anne Félix (Kiontke & Sudhaus 2006; Sudhaus & Kiontke 2007). Like other *Caenorhabditis*, *C. sp. 5* is a small (~1 mm long), transparent bactivorous nematode amenable to laboratory culture. It is closely related to *C. briggsae* and *C. sp. 9* (Kiontke et al. 2007; Cutter 2008a), and by virtue of providing the nearest outgroup for these two sister species, *C. sp. 5* holds an important position for analyses of molecular evolution and comparative phenotypic analysis based on *C. briggsae*. The shape of part of the male reproductive anatomy is distinct in *C. sp. 5* compared to other *Caenorhabditis* (K. Kiontke, personal communication), as are some aspects of cell division in the developing female vulva (Dolgin et al. 2008), making it an interesting subject for morphological analysis of the evolution of development. Until recently, only one main isolate was available (JU727), for which an expressed sequence tag library was constructed and used in analyses of molecular divergence and the evolution of biased codon usage (Cutter 2008a; Cutter et al. 2008). The evidence of selection-mediated codon bias suggests that *C. sp. 5* likely has a large effective population size, which is necessary for weak translational selection for codon bias to manifest (Bulmer 1991). However, an understanding of the extent of its species range, the level genetic variation and the diversity of environments inhabited by *C. sp. 5* has been unavailable until now.

In this study, we document the extent of nucleotide polymorphism across the X-chromosome of *C. sp. 5* and quantify the degree to which molecular genetic variation is partitioned among sampling localities and evinces deviation from neutral population genetic expectations for a single panmictic population. Our extensive sampling confirms that this species has a widespread distribution within China, but so far has been isolated elsewhere in the world only in northern Vietnam near its border with China. *C. sp. 5* harbours an exceptionally high level of polymorphism, contributing to an emerging picture of the gonochoristic species of *Caenorhabditis* being hyperdiverse in molecular population genetic variation.

**Materials and methods**

**Nematode strains**

We collected 18 isofemale lines of *C. sp. 5* from throughout eastern and central China during 2008 and 2009 according to established procedures (Barrière & Félix 2006) (Table 1). We supplemented these isolates with an additional 4 strains, also from China or from Vietnam near the Chinese border, that were gifts of Marie-Anne Félix (Table 1). Strains are available from the authors or the Félix lab (http://www.justbio.com/worms). Initial species determination was made through crosses with other known species and sequencing a portion of the small subunit rDNA. The geographical sources of collection are viewable through an online mapping resource http://www.google.com/maps/ms?msa=0&msid=117700919974655793194.00046e90c5fa3a4c5837b. We computed the distance between collection sites using the haversine formula, assuming the radius of the Earth to be 6367.5 km. The most distant pair of sites are separated by 1786 km, and on average the sites are 644 km apart.

**Molecular methods**

We used PCR primers previously developed for *C. remanei* to amplify nuclear gene fragments from the *C. sp. 5* genome (Cutter 2008b). After initial testing of a large number of primer pairs, we successfully obtained sequence from *C. sp. 5* for 7 primer sets. These primers include the p46, p47, p50, p51, p74, p87 and p94 primer sets described previously (Cutter 2008b). These loci are putatively X-linked and contain coding sequence from a single exon, representing about 4.4 kb of total sequence from each individual. X-linked loci were selected to
ease analysis, because DNA isolated from a single male (karyotype: 5 diploid autosomes and a haploid X chromosome) eliminates the potential for heterozygous base calls in sequence traces. We refer to the genes of these partial coding sequences using an extension of the nomenclature suggested by Jonathan Hodgkin for other Caenorhabditis, namely, an abbreviation representing the species identity followed by the C. elegans ortholog gene name (e.g. Csp5-spc-1). DNA used as template for PCR was isolated from single male worms of each strain and amplified with the Qiagen Repli-g mini kit as described previously (Cutter 2008b). The lack of heterozygous single nucleotide polymorphism (SNP) calls in sequence traces supports the expectation of sex-chromosome linkage for these loci. Sequencing was unsuccessful for VX0070 for primer set p94, giving a sample size of 21 for this locus in analysis. Sequencing was performed at the University of Arizona UAGC sequencing facility and by the University of Toronto CAGEF sequencing facility. New sequences have been deposited in Genbank under accessions HQ189144-H189296.

**Sequence analysis**

The programs SEQUENCER 4.10 and BioEdit 7.0.9 were used to manually edit and confirm sequence quality. We then calculated population genetic summary statistics for synonymous and replacement sites separately with DnaSP 5.10 (Librado & Rozas 2009), including polymorphism from per-site pairwise differences (π) and segregating sites (θ) with Jukes-Cantor multiple-hits correction, the number of segregating sites (S) and the site frequency spectrum (Tajima’s D). We also used DnaSP to calculate the number of haplotypes (H), the minimum number of recombination events (Rmin) and LD (r²) for each locus. Pairwise genetic distances used in a test of isolation by distance are based on Nei & Gojobori’s (1986) model for synonymous sites (Ks), also calculated in DnaSP. Measures of codon bias (Fop, frequency of preferred codons using C. elegans codon table; Nc, effective number of codons) were measured for the gene fragments with CodonW (J. Feden, http://codonw.sourceforge.net). Divergence with Caenorhabditis briggsae of homologous regions (Ks, Ks/Kw) were calculated using the Goldman and Yang method implemented in Kaks calculator (Goldman & Yang 1994; Zhang et al. 2006). The population recombination rate (ρ) was estimated with LDhat’s finite-sites model, and nonparametric tests for recombination were performed in LDhat with the ‘Lkmax’ method (McVean et al. 2002). We used nonlinear fitting in JMP 7 to fit the expected decay in r² with distance between polymorphic sites (Weir & Hill 1986) as carried out previously (Cutter et al. 2006), as well as a spline fit. A Mantel test on genetic and geographical distance matrices was implemented with custom Perl scripts, assuming independence of loci (1000 matrix permutations). We repeated this Mantel test approach using just a single,
arbitrarily selected strain from those localities for which we had sampled two or more strains; this approximates the ‘scattered sample’ paradigm of the coalescent process for many demes (Wakeley & Aliacar 2001). We made a rough calculation of effective population size (N_e) assuming demographic equilibrium (i.e., \( \pi_{\text{syn}} = 0 = 4 N_e \mu \)) and neutral mutation rate (\( \mu \)) from \textit{C. elegans} \( (2.7 \times 10^{-9} \text{mutations/site/generation}) \) or \textit{C. briggsae} \( (2\text{-fold higher than } C. \text{ elegans}) \) (Denver et al. 2009; Phillips et al. 2009).

**Results**

**Patterns of polymorphism**

Nucleotide polymorphism at synonymous sites is very high in our sample of 22 individuals of \textit{C. sp. 5}, averaging 6.84% across seven nuclear loci (Jukes–Cantor corrected; uncorrected average \( \pi_{\text{syn}} = 6.40\% \)). Polymorphism varied roughly 2.5-fold across loci, with the locus with lowest diversity still being highly variable in the population: \( \pi_{\text{syn}} = 4.36\% \) (Csp5-asp-3; Table 2). However, polymorphism at replacement sites is very low (average \( \pi_{\text{rep}} = 0.05\% \); Table 2), indicating strong constraint on changes to the amino acid sequence encoded by these genes. Indeed, the overall \( \pi_{\text{rep}}/\pi_{\text{syn}} \) ratio of 0.0093 implies that more than 99% of mutations to replacement sites are effectively eliminated by purifying selection. Across the 4.4 kb of coding sequence sampled from each individual, we identified 222 SNPs at synonymous sites and 9 at replacement sites.

We calculated the Tajima’s (1989) \( D \) summary of the variant frequency spectrum for synonymous sites and found that none of the loci deviated significantly from the neutral expectation. We note, however, that the average value of \( D \) across loci is positive (mean \( D_{\text{syn}} = 0.24 \)), which could be indicative of a modest level cryptic population structure within our sample or a relatively recent population contraction. Given the fairly broad geographical distribution of sampling sites across China, we consider cryptic population structure to be very plausible. As expected with purifying selection targeting replacement sites, Tajima’s \( D \) indicates a trend for skew towards low-frequency variants and a much lower mean value (mean \( D_{\text{rep}} = -0.27 \)) relative to synonymous sites.

We also calculated codon bias for these gene regions as well as their divergence with \textit{C. briggsae} (Table 3), because selection on codon usage has been demonstrated to affect synonymous-site divergence in \textit{C. sp. 5} (Cutter 2008a; Cutter et al. 2008). However, our sample of 7 loci is not sufficient to allow strong tests for correlations with polymorphism, and we did not adjust \( K_s \) for selection on synonymous sites.

### Linkage disequilibrium

Linkage disequilibrium decays detectably across only hundreds of nucleotides between SNP pairs, dropping by half within \(~250 \) bp on average (Fig. 1). When each locus was considered separately (mean \( 630 \) bp long), the decay of LD was significant for five of the seven loci (Table 3). Moreover, recombination was evident in all loci, with the minimum number of recombination events \((R_{\text{min}})\) ranging from 2 to 13 within each locus (Table 3). Estimating the population recombination parameter \((\rho = 4 N_e c, \text{ effective population size } N_e \text{ and recombination rate per generation } c)\) with the finite-sites method implemented in LDhat, we again confirm extensive recombination with mean \( \rho = 0.0125 \) per site. Consequently, the ratio \( \rho/\pi_{\text{syn}} \) averages 0.199, indicating that there is abundant nucleotide variation relative to the level of recombination.

### Table 2 Summary of molecular genetic variation statistics†

<table>
<thead>
<tr>
<th>gene (primer)</th>
<th>( \pi_{\text{syn}}^a )</th>
<th>( \pi_{\text{rep}}^a )</th>
<th>( \theta_{\text{syn}} )</th>
<th>( \theta_{\text{rep}} )</th>
<th>( S_{\text{syn}} )</th>
<th>( S_{\text{rep}} )</th>
<th>( D_{\text{syn}} )</th>
<th>( D_{\text{rep}} )</th>
<th>( H )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Csp5-asp-3 (p46)</td>
<td>0.0436</td>
<td>0.00112</td>
<td>0.0390</td>
<td>0.00074</td>
<td>16</td>
<td>1</td>
<td>0.280</td>
<td>0.895</td>
<td>11</td>
</tr>
<tr>
<td>Csp5-spc-1 (p47)</td>
<td>0.1131</td>
<td>0</td>
<td>0.0846</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>0.839</td>
<td>n.a.</td>
<td>10</td>
</tr>
<tr>
<td>Csp5-col-41 (p50)</td>
<td>0.0677</td>
<td>0</td>
<td>0.0697</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>-0.320</td>
<td>n.a.</td>
<td>12</td>
</tr>
<tr>
<td>Csp5-myo-2 (p51)</td>
<td>0.0599</td>
<td>0.00020</td>
<td>0.0480</td>
<td>0.00059</td>
<td>20</td>
<td>1</td>
<td>0.662</td>
<td>-1.162</td>
<td>12</td>
</tr>
<tr>
<td>Csp5-pgp-14 (p74)</td>
<td>0.0891</td>
<td>0.00089</td>
<td>0.0710</td>
<td>0.00050</td>
<td>42</td>
<td>1</td>
<td>0.638</td>
<td>1.334</td>
<td>13</td>
</tr>
<tr>
<td>Csp5-W07E11.1 (p87)</td>
<td>0.0480</td>
<td>0.00035</td>
<td>0.0430</td>
<td>0.00056</td>
<td>24</td>
<td>1</td>
<td>0.241</td>
<td>-0.641</td>
<td>11</td>
</tr>
<tr>
<td>Csp5-E01G6.1 (p94)</td>
<td>0.0573</td>
<td>0.00107</td>
<td>0.0650</td>
<td>0.00265</td>
<td>36</td>
<td>5</td>
<td>-0.629</td>
<td>-1.795</td>
<td>15</td>
</tr>
<tr>
<td>average</td>
<td>0.0684</td>
<td>0.00052</td>
<td>0.0600</td>
<td>0.00072</td>
<td>31.7</td>
<td>1.3</td>
<td>0.244</td>
<td>-0.274</td>
<td>12.0</td>
</tr>
</tbody>
</table>

†Average pairwise nucleotide polymorphism at synonymous (\( \pi_{\text{syn}} \)) and replacement-sites (\( \pi_{\text{rep}} \)), number of segregating sites at synonymous (total observed \( S_{\text{syn}} \) scaled per site \( \theta_{\text{syn}} \)) and replacement sites (\( S_{\text{rep}} \), \( \theta_{\text{rep}} \)), Tajima’s (1989) metric of the site frequency spectrum at synonymous (\( D_{\text{syn}} \)) and replacement sites (\( D_{\text{rep}} \)), observed number of haplotypes (\( H \)).

© 2010 Blackwell Publishing Ltd
Table 3 Summary of molecular evolution statistics†

<table>
<thead>
<tr>
<th>Gene (primer)</th>
<th>Lengtha</th>
<th>$R_{\min}$</th>
<th>$\rho^b$</th>
<th>$F_{op}^c$</th>
<th>$N_{C}^d$</th>
<th>$K_{s}^d$</th>
<th>$K_{s}^d$</th>
<th>$K_s/K_a^d$</th>
<th>$\pi_{rep}/\pi_{syn}^w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Csp5-asp-3 (p46)</td>
<td>485</td>
<td>2</td>
<td>0.0153**</td>
<td>0.383</td>
<td>52.0</td>
<td>0.0238</td>
<td>0.369</td>
<td>0.0644</td>
<td>0.257</td>
</tr>
<tr>
<td>Csp5-spc-1 (p47)</td>
<td>670</td>
<td>7</td>
<td>0.0046**</td>
<td>0.721</td>
<td>33.1</td>
<td>0.0020</td>
<td>0.523</td>
<td>0.0039</td>
<td>0.0006</td>
</tr>
<tr>
<td>Csp5-col-41 (p50)</td>
<td>657</td>
<td>6</td>
<td>0.0088**</td>
<td>0.322</td>
<td>51.0</td>
<td>0.0355</td>
<td>0.417</td>
<td>0.0850</td>
<td>0.0000</td>
</tr>
<tr>
<td>Csp5-mnp-2 (p51)</td>
<td>579</td>
<td>3</td>
<td>0.0079**</td>
<td>0.347</td>
<td>44.7</td>
<td>0.0130</td>
<td>0.503</td>
<td>0.0259</td>
<td>0.0033</td>
</tr>
<tr>
<td>Csp5-pgp-14 (p74)</td>
<td>709</td>
<td>13</td>
<td>0.0295**</td>
<td>0.284</td>
<td>51.2</td>
<td>0.0261</td>
<td>0.885</td>
<td>0.0295</td>
<td>0.0100</td>
</tr>
<tr>
<td>Csp5-W07E11.1 (p87)</td>
<td>645</td>
<td>3</td>
<td>0.104**</td>
<td>0.366</td>
<td>50.2</td>
<td>0.0244</td>
<td>1.430</td>
<td>0.0171</td>
<td>0.0073</td>
</tr>
<tr>
<td>Csp5-E01G6.1 (p94)</td>
<td>680</td>
<td>4</td>
<td>0.0110**</td>
<td>0.333</td>
<td>48.9</td>
<td>0.0230</td>
<td>0.927</td>
<td>0.0248</td>
<td>0.0187</td>
</tr>
<tr>
<td>average</td>
<td>632.1</td>
<td>5.4</td>
<td>0.0125</td>
<td>0.394</td>
<td>47.3</td>
<td>0.0211</td>
<td>0.722</td>
<td>0.0358</td>
<td>0.0093</td>
</tr>
</tbody>
</table>

†minimum number of recombination events ($R_{\min}$), population recombination parameter per site ($\rho$), frequency of optimal codons ($F_{op}$), effective number of codons ($N_{C}$), divergence with C. briggsae at replacement sites ($K_{s}$) and synonymous sites ($K_{a}$), ratio of pairwise nucleotide polymorphism at replacement and synonymous sites ($\pi_{rep}/\pi_{syn}$); ‡gaps excluded; §significance of recombination

\[ P < 0.001 \]. This finding of isolation by distance further supports the notion that there is some level of population structure among the sampled individuals, as suggested by the slightly positive values of Tajima’s $D$. However, the fact that we have just a single iso-female isolate (strain) from most sampling sites precludes further tests of substructure among sampling localities within China (e.g., $F_{ST}$).

**Discussion**

Our sample of individuals for C. sp. 5 identifies substantial nucleotide polymorphism, averaging $\pi_{syn} = 0.068$, which implies that there is about 1 SNP of every 15 silent sites between a random pair of sequences. We find that C. sp. 5 exhibits weak, but significant, genetic isolation by distance, with individuals from more geographically distant localities tending to have more divergent haplotypes. However, there is pervasive recombination within loci, causing LD to decay rapidly over spans of just a few hundred nucleotides. Thus, the extremely high diversity observed within C. sp. 5 does not appear to be because of divergence between differentiated lineages or subpopulations that might comprise cryptic species. Instead, we propose that the effective population size of C. sp. 5 is very large. We expect that large effective population size could be maintained by virtue of the abundant food resource for C. sp. 5 (bacteria in rotting vegetable matter), an apparently high dispersal ability and its obligately outbreeding mode of reproduction. A rough calculation of long-term $N_e$ for C. sp. 5 is $3–6 \times 10^6$, assuming demographic equilibrium and that the estimates of mutation rate from C. elegans and C. briggsae based on mutation accumulation experiments can extend to C. sp. 5 (Denver et al. 2009; Phillips et al. 2009).
The signature of genetic isolation by distance within C. sp. 5 suggests that this species harbours some degree of population structure across its range in east Asia. With available sampling, which is currently limited to one or a few isofemale strains from any given locality, it is not yet possible to detail the extent of subpopulation differentiation. By considering the total sample in the face of some degree of structure, our analyses likely give slightly biased estimates of some population genetic parameters. For example, within any single locality, we should expect polymorphism, LD and the deficit of low-frequency variants to be somewhat lower than the values reported here. However, the high number of pairwise differences between sequences expected even at a geographical distance of 0 km (Fig. 2) as well as the rapid decay of LD, even in the total sample (Fig. 1), indicate that detailed investigation of individual localities should corroborate the main findings we report here.

C. remanei populations also have high levels of molecular variation at synonymous sites, albeit substantially lower than C. sp. 5 (Graustein et al. 2002; Jovelin et al. 2003, 2009; Cutter et al. 2006; Cutter 2008b; Jovelin 2009). Homologous loci in C. remanei from Ohio, USA show $\pi_{syn} = 0.022$, although much larger samples of loci in that species give an average of $\pi_{syn} \sim 0.04$ (Cutter 2008b; Jovelin 2009). Thus, high nucleotide polymorphism appears to be a common feature of obligately outbreeding species of Caenorhabditis—and should exemplify the state of those ancestral populations that gave rise to the derived, extant species that are comprised of self-fertile hermaphrodites (C. elegans, C. briggsae, C. sp. 11). C. sp. 5, in particular, harbours enormous levels of polymorphism in absolute terms, rivalling that of the sea squirt Ciona savignyi (average $\pi = 0.08$ at 4-fold synonymous sites), which is the metazoan with highest documented molecular diversity across its genome to date (Small et al. 2007).

The level of polymorphism seen in C. sp. 5 and C. remanei greatly exceeds the levels observed in self-fertilizing taxa like C. elegans and C. briggsae, differing by a factor of at least 20 (reviewed in Cutter et al. 2009). Moreover, this disparity between the obligate outbreeding species and the self-fertile species is much greater than the roughly 2-fold difference expected from selfing alone. Presuming that polymorphism in these outbreeding species is representative of the values present in the outbreeding ancestors of C. elegans and C. briggsae, this indicates that other effects beyond selfing alone must have reduced disproportionately the amount of variation in self-fertile lineages (Graustein et al. 2002; Sivasundar & Hey 2003; Barrière & Félix 2007; Cutter et al. 2009). Such diversity-reducing factors potentially include potent extinction–recolonization metapopulation dynamics, a recent origin of high selfing rate in each species and the consequences of selection at linked sites. These factors are not mutually exclusive, and there is support in the literature from molecular evolution and population genetic data for each of them individually in C. elegans and/or C. briggsae (reviewed in Cutter et al. 2009).

C. sp. 5 is found only in East Asia but in both tropical and temperate latitudes, whereas C. remanei appears to be found around the globe but geographically restricted to northerly latitudes (Sudhaus & Kiontke 2007). Other gonochoric species have more tropical known distributions (e.g. C. brenneri) (Sudhaus & Kiontke 2007). We isolated C. sp. 5 from a variety of sample types that are typical of habitats in which Caenorhabditis have been found previously (Kiontke & Sudhaus 2006; Sudhaus & Kiontke 2007), including in phoretic association with isopods and snails, in rotting fruit and vegetation, in compost and in agriculture-associated soil (Table 1). We also identified C. remanei and C. briggsae at some localities where we also found C. sp. 5, indicating that the geographical ranges of these three species overlap at a fairly fine scale within China. Further sampling of Caenorhabditis species is necessary to fully document the extent of their species ranges, habitat types, potential niche partitioning and phylogenetic diversity.

One locus, Csp5-spc-1, revealed unusual patterns of genetic variation in several respects (Tables 2 and 3). Csp5-spc-1 had the fewest haplotypes, and yet exhibited the highest synonymous-site diversity among the loci we analysed, despite being the only locus with strongly skewed codon usage that is indicative of selection acting on synonymous sites. Fifteen of 22 samples fall into three differentiated haplotypes for the 670-bp portion of
Csp5-spc-1 we have analysed, one of which includes individuals derived only from localities in southeastern China. This locus also had the most skewed variant frequency spectrum (Tajima’s $D = +0.84$) and was one of the two loci for which we could not detect significant decay in LD along its length. These patterns suggest that Csp5-spc-1, and its genomic neighbourhood, might experience stronger subpopulation differentiation than other loci. The protein sequence for Csp5-spc-1 is by far the most highly constrained of the loci in our sample, based on low interspecific divergence, and indeed we observed no replacement-site polymorphisms within C. sp. 5. Notably, the ortholog of this locus in C. remanei also stood out in population genetic analysis (Cutter 2008b), albeit in a qualitatively different manner (significant negative skew of the variant frequency spectrum with $D = -1.9$, with below-average synonymous-site polymorphism). In C. elegans, the alpha spectrin SPC-1 is a long 2427 amino acid protein involved in body morphogenesis, including body-wall muscle development, and is expressed at high levels throughout embryogenesis (Norman & Moorman 2002). Further investigation of the genomic region surrounding Csp5-spc-1 in a more extensive collection of population samples from different localities would help determine the causes of the unusual patterns of polymorphism at this locus.

Despite the streamlined morphology that is largely devoid of obvious differences among Caenorhabditis species, it is becoming clear that individual species harbour enormous genetic variation at the molecular level—in addition to the well-known high molecular divergence between species (Thomas & Wilson 1991; Stein et al. 2003; Cutter 2008a). This molecular population genetic biodiversity provides unprecedented opportunity to combine the laboratory experimental tractability of these organisms to characterize organismal function and evolution at the genetic level. In particular, ‘SNP shadowing’—an intraspecific analog to ‘phylogenetic footprinting’—should prove especially valuable in C. sp. 5 and related outbreeding taxa for characterizing the evolution and function of regulatory and other sequence features as a precursor to further experimental work (Boffelli et al. 2004).

Acknowledgements

We are grateful to Marie-Anne Félix and Josselin Milloz for collecting and sharing strains with us. We thank Cecilia Chan, Alivia Dey, Richard Jovelin and Joanna Kozlowska for discussion and assistance. This research was supported by the National Natural Science Foundation of China (NSFC #30873350, 31071998) and Huazhong Normal University (G.-X.W.) and by the Natural Sciences and Engineering Research Council of Canada (A.D.C).

References


Research in Asher Cutter’s laboratory addresses genome evolution and the evolution of breeding systems, reproductive isolation, and adaptation in nematodes, focusing on population genetic and molecular evolutionary approaches. Guo-xiu Wang’s lab studies population genetics and molecular evolution in *Caenorhabditis* and entomopathogenic Mermithid nematodes. Shuang Ren and Yuan Ren are graduate researchers working with Guo-xiu Wang. Hui Ai’s research interests include studying the biodiversity and phylogenetics of nematodes in China.