Available online at www.sciencedirect.com

SCIENCE DIRECT®

Gene xx (2005) xxx–xxx

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

Intraspecific variation in *fem-3* and *tra-2*, two rapidly coevolving nematode sex-determining genes

Eric S. Haag*, Amanda D. Ackerman

Department of Biology, University of Maryland, College Park, MD 20742, USA

Received 18 August 2004; received in revised form 16 December 2004; accepted 27 December 2004

Received by D. Fitch

Abstract

The sex determination gene *fem-3* encodes one of the most divergent proteins yet described in the terrestrial nematode *Caenorhabditis*. Despite this rapid sequence change, however, FEM-3 is essential for male development in the three species surveyed thus far. It also participates in conserved protein–protein complexes with the transmembrane receptor TRA-2 and the phosphatase FEM-2 in these species. These interactions show strong species specificity, indicating that conserved residues are not sufficient for function and that compensatory evolution between binding partners is important. To shed further light on the nature of this coevolution, and to discern the extent of amino acid polymorphism allowed in FEM-3 and the domain of TRA-2 that binds it, we have examined intraspecific variation in the gonochoristic species *Caenorhabditis remanei*. Ten new complete *Cr-fem-3* alleles from three regions of the United States are described. We also obtained sequences for the FEM-3-binding domain of TRA-2 for 9 of the same strains. These alleles were compared with each other, with the European founder alleles, and with the orthologous sequences from the congeners *Caenorhabditis elegans* and *C. briggsae*. We find that FEM-3 harbors abundant amino acid polymorphisms along its entire length. The majority (but not all) of these occur in nonconserved residues, and in at least one domain there is evidence for diversifying selection. The FEM-3-binding domain of TRA-2 is less polymorphic than FEM-3. Amino acids neither polymorphic nor conserved between species are candidates for residues mediating species-specific interaction of FEM-3 with its binding partners.

Keywords: *C. remanei*; Polymorphism; Sex determination; Coevolution

1. Introduction

The regulation of sexual dimorphism is an important process in most animals, but the mechanisms by which it is accomplished are not conserved among phyla. For instance, sex chromosomes or environmental cues are two different ways in which animals may distinguish sexes. Even systems of genetic sex determination that are superficially similar may be completely different at the gene level (Cline and Meyer, 1996). Over smaller phylogenetic distances where conserved genes can be recognized, their sequences evolve rapidly compared to other genes in a diverse group of organisms (e.g., Tucker and Lundrigan, 1993; de Bono and Hodgkin, 1996; Civetta and Singh, 1998; Haag et al., 2002). Numerous studies in both *Drosophila* and in mammals have addressed whether this change is driven by adaptive evolution or by unconstrained neutral processes. Several of these studies have found that neutral evolution cannot be rejected despite the higher than usual substitution rates (O'Neil and Belote, 1992; McAllister and McVean, 2000; Moreira, 2002). Where positive selection has been inferred, it often fails to account for the bulk of sequence evolution (Pamilo and O'Neill, 1997; Jansa et al., 2003). Similar evaluation of evolutionary dynamics in nematode sex determination genes is hampered by the large genetic distance between presumably closely related species (Kiontke et al., 2004), including the mutational saturation of

Abbreviations: LD, linkage disequilibrium; PCR, polymerase chain reaction; EGF, epidermal growth factor; LNR, LIN-12/Notch repeat.

* Corresponding author. Tel.: +1 301 405 8534; fax: +1 301 314 9358.
E-mail address: ehaag@wam.umd.edu (E.S. Haag).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.12.051

silent sites between congeners (Stein et al., 2003). As a result, many standard statistical methods to detect positive selection are not applicable.

In addition to selection pressures, the rapid evolution of sex determination genes also raises interesting questions about the relationship between primary protein sequences and biochemical function, and how molecular variation affects the developmental genetic network controlling sexual dimorphism. Here the nematode system fares better due to the extensive molecular characterization of the *C. elegans* sex determination gene network (Goodwin and Ellis, 2002) and recent interspecific comparisons of sex determination gene function (de Bono and Hodgkin, 1996; Haag and Kimble, 2000; Chen et al., 2001; Stothard et al., 2002; Haag et al., 2002; Pires-da Silva and Sommer, 2004). This study extends this work by examining within-species variation in the rapidly evolving nematode sex determination gene, *fem-3,* which is essential for all male somatic cell fates in *C. elegans*, *C. briggsae*, and *C. remanei* (Hodgkin, 1986; Haag et al., 2002). It encodes a novel cytoplasmic protein that acts genetically between *tra-2*, which encodes a membrane protein, and *tra-1*, which encodes a Gli-related zinc-finger transcription factor. FEM-3 interacts physically with both the cytoplasmic domain of TRA-2 (TRA-2c) and with another cytoplasmic protein, the PP2c-type phosphatase FEM-2 (Pilgrim et al., 1995; Chin-Sang and Spence, 1996; Mehra et al., 1999). In males, TRA-2A is inhibited by the secreted ligand HER-1, allowing FEM-3, along with FEM-1 and FEM-2, to repress *tra-1* activity and promote male development (reviewed by Kuwabara and Perry, 2001).

The physical interaction and epistatic relationship between *fem-3* and *tra-2* are conserved and necessary for proper male somatic development in the sister species *C. remanei* and *C. briggsae*. However, the FEM-3-binding domain (F3BD) of TRA-2c is hyperdivergent between species (Kuwabara, 1996; Mehra et al., 1999; Haag and Kimble, 2000), and only when FEM-3 and TRA-2c are from the same species can they interact successfully in yeast (Haag et al., 2002), implying the existence of compensatory coevolution. This phenomenon led us to examine the extent and location of allowed intraspecific polymorphisms in FEM-3 and the TRA-2c F3BD. If coevolution is characterized by bouts of strong selection, then little variation in either is expected to exist. Conversely, coevolution may be an essentially neutral process facilitated by a large amount of standing variation in the interacting proteins.

We also saw the potential to discern whether selection was driving divergence by comparing intraspecific variants from distinct locales. The gonochoristic species *C. remanei* was chosen as the organism of study because past research has shown that there is little genomic variation among populations of *C. elegans* (Graustein et al., 2002; Sivasundar and Hey, 2003). Sequencing of two *Ce-fem-3* alleles confirmed low sequence variation between populations (L. Phillips and E. Haag, unpublished data). The original *Cr-tra-2* and *Cr-fem-3* sequences (Haag and Kimble, 2000; Haag et al., 2002) were from SB146, the type strain isolated in Germany (Sudhaus and Kiontke, 1996). We have determined the entire sequence of 10 new alleles of *Cr-fem-3*, as well as 9 new alleles of the F3BD-encoding portion of *Cr-tra-2*, from populations in three regions of the United States. They reveal the existence of abundant polymorphic amino acids scattered along the length of *Cr*-FEM-3, and in one region we find evidence for their excess relative to the neutral expectation. The TRA-2c F3BD was considerably less variable. We further examine how the nonsynonymous polymorphisms relate to interspecies conservation and to previous functional studies of FEM-3. Silent changes are also abundant, including some large insertion/deletion polymorphisms.

2. Methods

C. remanei strains CR1014, CR1415, CR2124, EM464, PB205, PB206, PB212, PB219, PB228, and PB229 were obtained with the kind assistance of M. Palopoli, Bowdoin College, and cultured under standard *C. elegans* conditions (Wood, 1988) with the exception of increasing the agar content of plates to 2.5% to discourage burrowing. For each strain, genomic DNA was isolated as described in Haag and Kimble (2000) and used as template for the polymerase chain reaction (PCR) with primers annealing to the 5′ and 3′ untranslated portions of the terminal exons of *Cr-fem-3*, based on the sequence from strain SB146 (Haag et al., 2002). For strains refractory to this approach, *Cr-fem-3* was isolated as two overlapping fragments that together amplified the entire gene (Table 1). To minimize the chances of creating chimeric alleles, clones were assembled that were identical at polymorphic sites lying in the region of fragment overlap. When possible the entire *fem-3* coding region was amplified with Pfu and Taq polymerases.

Direct sequencing of the *Cr-fem-3* amplicons failed in many cases due to heterozygosity of indel polymorphisms, so most of the data were obtained from cloned material. Alignments of partial sequences from multiple clones from a given strain revealed multiple polymorphisms, similar to that reported in other studies (Haag and Kimble, 2000; Graustein et al., 2002). We therefore chose a single haplotype from each strain for complete characterization. Products larger than 900 base pairs were cloned into either the pCR-script (Stratagene) or the pCR 2.1 TOPO-TA (Invitrogen) plasmid vectors. Each gene was sequenced with a combination of direct sequencing of PCR products and sequencing of plasmid clones with a combination of GPS-1 transposon insertions (New England Biolabs), gene-specific primers, and flanking vector primers. Sequence reads were assembled into contigs with the GCG Wisconsin Package. Completed *C. remanei fem-3* alleles were aligned against

Table 1
Primers used to amplify *Cr-fem-3* and *Cr-tra-2* F3BD

Strain	No. of PCR fragments	Primers
CR1014	2	EH7/RF3R2 (5′ end); EH20/EH4 (3′ end)
CR1415	2	EH7/RF3R2 (5′ end); RF3F1/EH4 (3′ end)
CR2124	2	EH7/RF3R2 (5′ end); EH20/EH4 (3′ end)
EM464	1	EH7/EH4
PB205	1	EH8/EH4
PB206	2	EH8/Rf3R1 (5′ end); EH11/EH4 (3′ end)
Pb212	1	EH7/EH4
PB219	1	EH8/EH4
PB228	1	EH8/EH4
PB229	1	EH8/EH4

Primer	Sequence
EH4	CATCGACCTCTGGCACAGATT
EH7	GGCGCAATGAAAATGACTTGCG
EH8	CGGCACGCTTCTATTCAAC
EH11	CGATGGATCATTTCCAGTTC
EH20	CATGAAGCACTTGGTTGCCCGC
RF3F1	GAAGCAAAGTACTTCCGAGAAAG
RF3R1	CCAAGTCAAACAAGGGATC
RF3R2	GCGGGCAACCAAGTGCTTCATG
EH27	GGCAAGTGTGTCATATGGCTCAG
EH28	TTCGGATCATCGAAATTGGGTGC

each other and a single *C. briggsae* sequence (AF16) using Vector NTI software, in the interspecies case with manual refinement based on the published protein alignment (Haag et al., 2002). The DnaSP (Rozas and Rozas, 1999) package (version 3.99) was then used to analyze sequence variation.

Representative *Cr-tra-2* FEM-3-binding domain (F3BD) sequences from the above strains were determined by a different strategy. Given its small size (526 nt including introns), it was amplified by PCR using single virgin adult female worms as template as previously described (Haag et al., 2002). This was preceded by three or four rounds of single female, full-sib mating to obtain PCR products largely free of heterozygosity for each strain. In 9 of 10 cases the sequence of agarose gel-purified products were successfully determined for both strands using the amplification primers, the exception being PB205.

3. Results and discussion

3.1. Nucleotide variation in Cr-fem-3

We used PCR to isolate new *Cr-fem-3* alleles from 10 strains of *C. remanei* from the USA: PB205, PB206, PB212, PB219, PB228, and PB229 were collected in Dayton, Ohio, EM464 in Brooklyn, New York, and CR1014, CR1415, and CR2124 in Gloucester, Massachusetts. The *Cr-fem-3* sequence from the German strain SB146 was previously described (Haag et al., 2002). The most striking variants included two alleles that harbored large noncoding insertions relative to the other nine (Fig. 1A). CR2124 was found to have a 633 nt segment in place of an 86 nt segment in the reference strain, SB146, between exons 1 and 2a. The resulting length variation region thus cannot be explained by simple insertion or deletion event. CR1014 has two simple insertions, one of 220 nt in intron 1 and another of 120 nt in intron 2. Outside of these large indels the remaining polymorphisms were single nucleotide polymorphisms or small indels of 3 nt or less.

We calculated the value of π, a measure of nucleotide diversity (Nei, 1987), over both introns and exons of the *Cr-fem-3* alleles using a sliding window (Fig. 1B). This analysis revealed a higher degree of polymorphism in introns, although exons 1, 4, and 5/6 have localized patches of comparably high polymorphism. The value for π across all sites, 0.027, is higher than, but comparable to, those reported for portions of *Cr-tra-2* (0.018) and *Cr-glp-1* (0.010) by Graustein et al. (2002). Given the uncertain

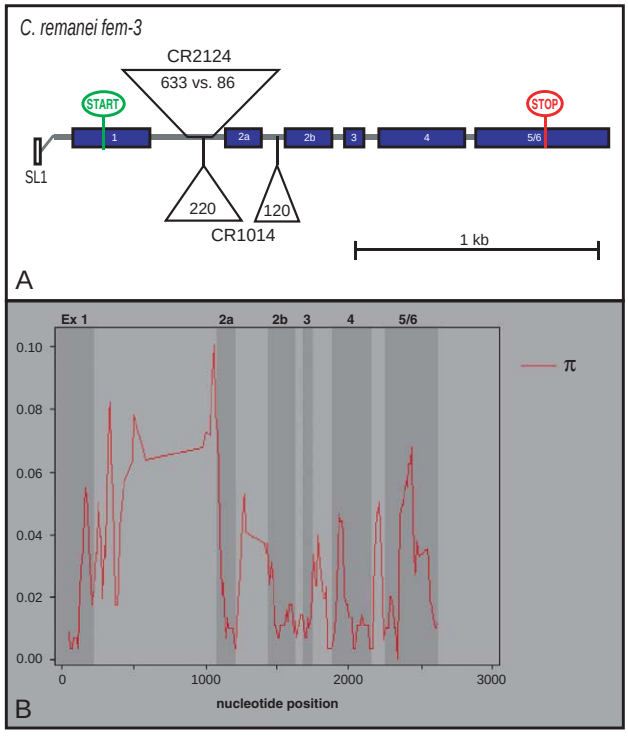

Fig. 1. Summary of nucleic acid variation in *C. remanei fem-3*. (A) Genomic structure, based on the SB146 reference strain, with large insertion polymorphisms in introns 1 and 2 indicated. The CR2124 strain harbors a 633 nt sequence that replaces an 86 nt stretch conserved in all other strains, while CR1014 has two smaller insertions that are both bounded by conserved sequences. Exons are numbered to indicate homology to *C. elegans* (Haag et al., 2002), and the splice site to SL1 leader is indicated. (B) Plot of nucleotide diversity (π) across the length of *Cr-fem-3*, using a 50 nt window slid in 10 nt increments. Exons are indicated by shading, and are labeled above the graph with numbers. Note that the plot begins just 5′ of the start codon, and finishes at the stop codon, so not all exonic sequences are included here. Relative length discrepancies in introns 1 and 2 between (a) and (b) are due to the insertional polymorphisms. Sequences of all alleles have been deposited with GenBank as accessions AY769072–AY769081.

population structure for *C. remanei*, we caution that these values of π are for not suitable for estimation of population genetic parameters, such as effective population size (N_e).

The extent to which gene flow and recombination exists between *C. remanei* populations remains unclear. To determine if *Cr-fem-3* polymorphisms are segregating independently, or are linked into a smaller number of haplotypes, linkage disequilibrium (LD) was estimated for all 13,695 pairs of segregating sites. Using Fisher's Exact Test, 126 were in significant disequilibrium ($p<0.05$), but with the Bonferroni correction for multiple tests no pairs appeared to be significant. However, Bonferroni correction is probably overly conservative due to non-independence of the comparisons. Indeed, there is a strong distance bias in the above 126 sites: 63 were less than 100 nt apart, and all but seven of the remainder were less than 500 nt apart. The lack of extensive LD further suggests that the polymorphisms that do exist recombine freely, even in widely separated populations. How the gene flow necessary for this to occur is unknown. Unassisted, *C. remanei* individuals would not be expected to be able to migrate significant distances, but their association with isopods or other invertebrates may facilitate movement (Baird, 1999).

3.2. Cr-FEM-3 amino acid polymorphism

A total of 37 polymorphic amino acids were found (Fig. 2) in the 439 total *Cr*-FEM-3 sites. Of these, 16 variants were seen in only a single haplotype sequence and 21 were found at least twice. The polymorphic residues lie along the entire length of the FEM-3 sequence, indicating that no major domain is completely constrained in its amino acid sequence. However, conservation between species varies dramatically on the scale

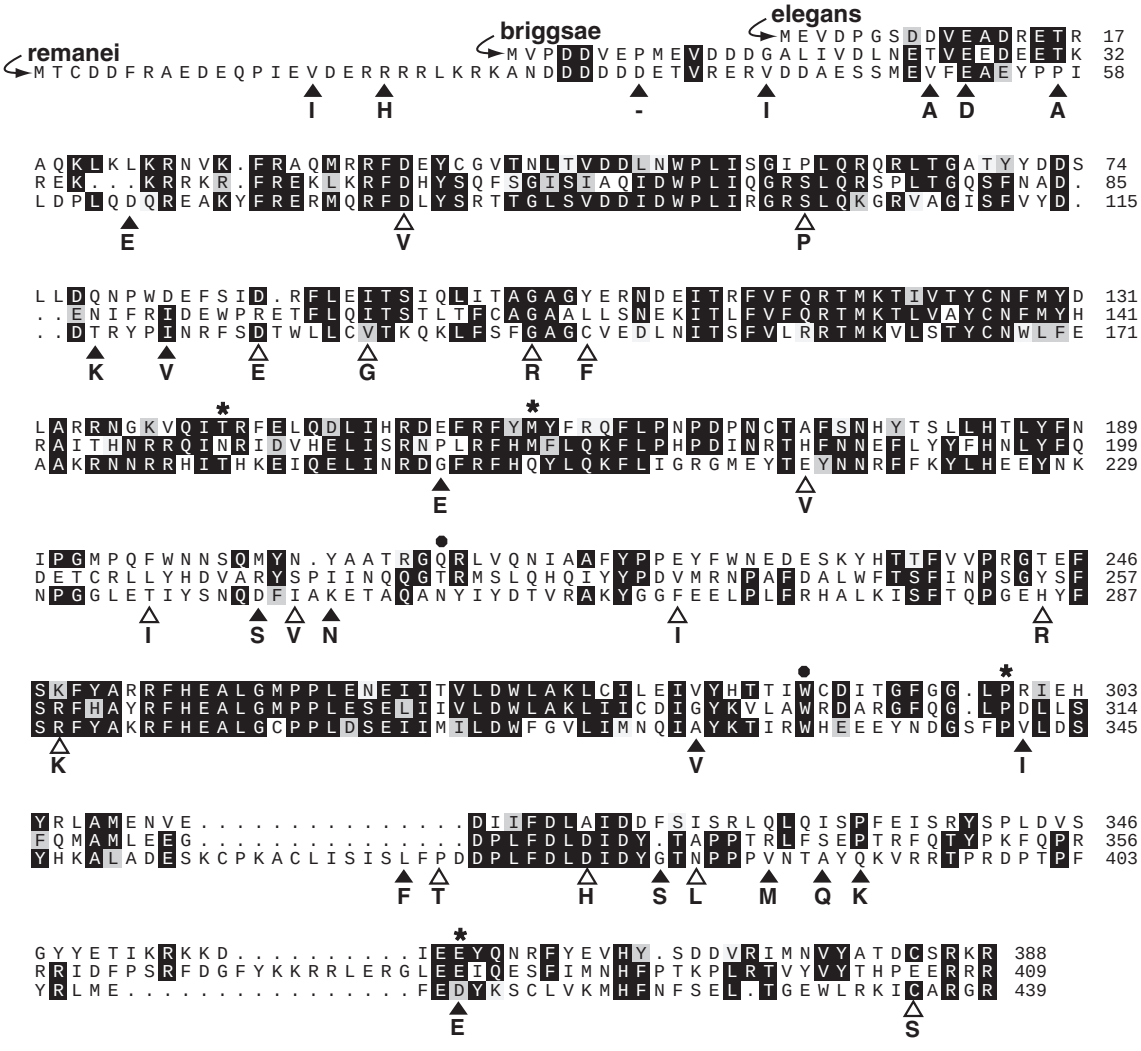

Fig. 2. Amino acid polymorphisms in *C. remanei* FEM-3 and their relation to interspecies conservation. The alignment is identical to that from Haag et al. (2002). Asterisks and octagons above the alignment indicate missense and premature stop mutations, respectively, reported from *C. elegans fem-3* loss-of-function mutants (Ahringer et al., 1992). Residues found to be polymorphic in *Cr*-FEM-3 are indicated below the alignment, with open triangles representing singleton variants and closed triangles those found in multiple alleles. Amino acids indicated below the triangles denote the alternate form in non-SB146 strains described in this paper. The minus sign under residue 34 of *Cr*-FEM-3 indicates an indel polymorphism.

of 5 to 10 amino acids, raising the possibility that small, highly conserved regions may behave differently than more divergent regions. To examine this, we broke the *Cr*-FEM-3 sites down into those conserved in the reference allele (SB146), *C. briggsae* and *C. elegans* (83, or 19%), those conserved in any two of the three species (144, or 33%), and species-specific sites that are either completely divergent in the three species or are unique to *Cr*-FEM-3 (212, or 48%), as inferred from the alignment of Haag et al. (2002). Of the 37 polymorphisms, 3 (8%) were in universally conserved sites, 10 (27%) fell in sites conserved in two species, and 24 (65%) were in divergent or unique sites. This distribution thus shows an excess of polymorphisms in divergent sites and a lack of them at conserved sites, and this trend is significant (χ^2=7.41, 2 *df*=2, *p*=0.025). Conserved amino acids are therefore indeed less flexible than divergent ones. It makes intuitive sense that tolerance of polymorphism at a particular site may predispose that site to undergo differential fixation after speciation, even in the absence of divergent selection.

3.3. Coding sequence variation

Using the method of Nei and Gojobori (1986) we estimate K_s from the *Cb-fem-3/Cr-fem-3* coding sequence alignment to be 3.02, and with the maximum likelihood method of Yang (1997) it climbs to a spectacular 12.98. This confirms the mutational saturation of silent sites in even the closest interspecies comparisons within *Caenorhabditis*. To learn whether the *Cr-fem-3* coding sequence variation bears hallmarks of intraspecific positive selection, we first estimated the relative frequencies of nonsynonymous and synonymous substitutions (K_a and K_s, respectively) for the entire codon set in all pairwise comparisons (Fig. 3A). In all cases K_a/K_s is well below one, and the overall measure of K_a/K_s across all pairwise comparisons, π_a/π_s, is 0.23**.** There is also no positive association at between geographic distance of the isolates and replacement change. For example, two strains from the same park in Dayton, Ohio (PB228 and PB229), have the highest K_a/K_s ratio (0.53), and a European/North American pair (SB146 and EM464) has the lowest (0.050). Both the spatial distribution of π and the K_a/K_s data are consistent with a dominant role for purifying selection to limit amino acid change, and are not consistent with extensive adaptive divergence of even widely separated *C. remanei* populations.

The above analysis is a highly conservative test for the detection of selection because it averages across all sites. As the interspecies FEM-3 sequence conservation is patchy and highly variable on the scale of 5–20 amino acids, we augmented whole sequence pairwise comparisons with a sliding window analysis that calculated π_a/π_s for 100 nt subsets of the *Cr-fem-3* open reading frame (Fig. 3B). This revealed marked variation in local synonymous/nonsynonymous ratios. For the most extreme window, centered on coding position 413 (codon 138), π_a/π_s=1.37, and three other windows have ratios >0.8. Five noncoding polymorphisms lie in the 413-centered window (codons 121–154), a region of especially low sequence interspecies conservation. It is possible that these high-scoring windows do contain one or more adaptive polymorphisms. However, all four peaks in the sliding window analysis fall in relatively low variability regions (the plot of π overall in Fig. 3B), and are thus produced by small numbers of polymorphic sites. In the case of the 413 window, the extreme π_a/π_s value is due to synonymous variation that is fivefold lower relative to the average across all codons, rather than to elevated nonsynonymous variation. The biological significance of these polymorphisms will therefore be unclear without functional studies.

3.4. Variation in the Cr-tra-2 F3BD domain

The TRA-2c F3BD, as defined by studies in *C. elegans* (Mehra et al., 1999), is 141 amino acids long in *C. remanei* (Haag and Kimble, 2000). As a complement to the FEM-3 data, we sequenced the genomic DNA encoding the F3BD from 9 of the 10 strains used in the *Cr*-FEM-3 study, including two short introns that lie within it. Unlike *Cr*-FEM-3, very little amino acid polymorphism exists in the F3BD (Fig. 4A). There are only 4 variable residues, all of which fall in the first third of the domain. Interestingly, one of these polymorphisms is a 3 nt indel that removes part of two adjacent codons relative to SB146. As a result, a glutamine–threonine pair is replaced by a single serine. Polymorphism at the residue just N-terminal to this site covaries perfectly with the indel, such that "KQT" and "IS" haplotypes exist. We found multiple copies of each haplotype in our sample, and in PB219 an individual heterozygous for each was found, indicating they are not mutually exclusive in a single population.

The cluster of missense mutations described above produces a local spike in the sliding window value of π_a/π_s that is nearly unity (Fig. 4B). Across the entire F3BD π_a/π_s is 0.131, about half that of *Cr-fem-3*. The overall nucleotide diversity, π, is 0.013, also about half the value for *Cr*-FEM-3. Sliding window analysis of π (Fig. 4B) shows rather uniform variation over the both coding and intronic parts of the F3BD region, except for the portion encoded by Exon 23, which is nearly invariant. Variation in the *Cr*-TRA-2c F3BD can thus be summarized as localized to the N-terminal/5′ end, and overall less abundant than that seen in *Cr-fem-3*.

3.5. Comparison of Cr-fem-3 and Cr-tra-2 F3BD with other C. remanei coding sequences

To help clarify whether intraspecific variation in *fem-3* and the *tra-2* F3BD are unusual or typical, we compared them to previously published sequence data for the N-terminal extracellular domain of *tra-2* and a portion of *glp-*

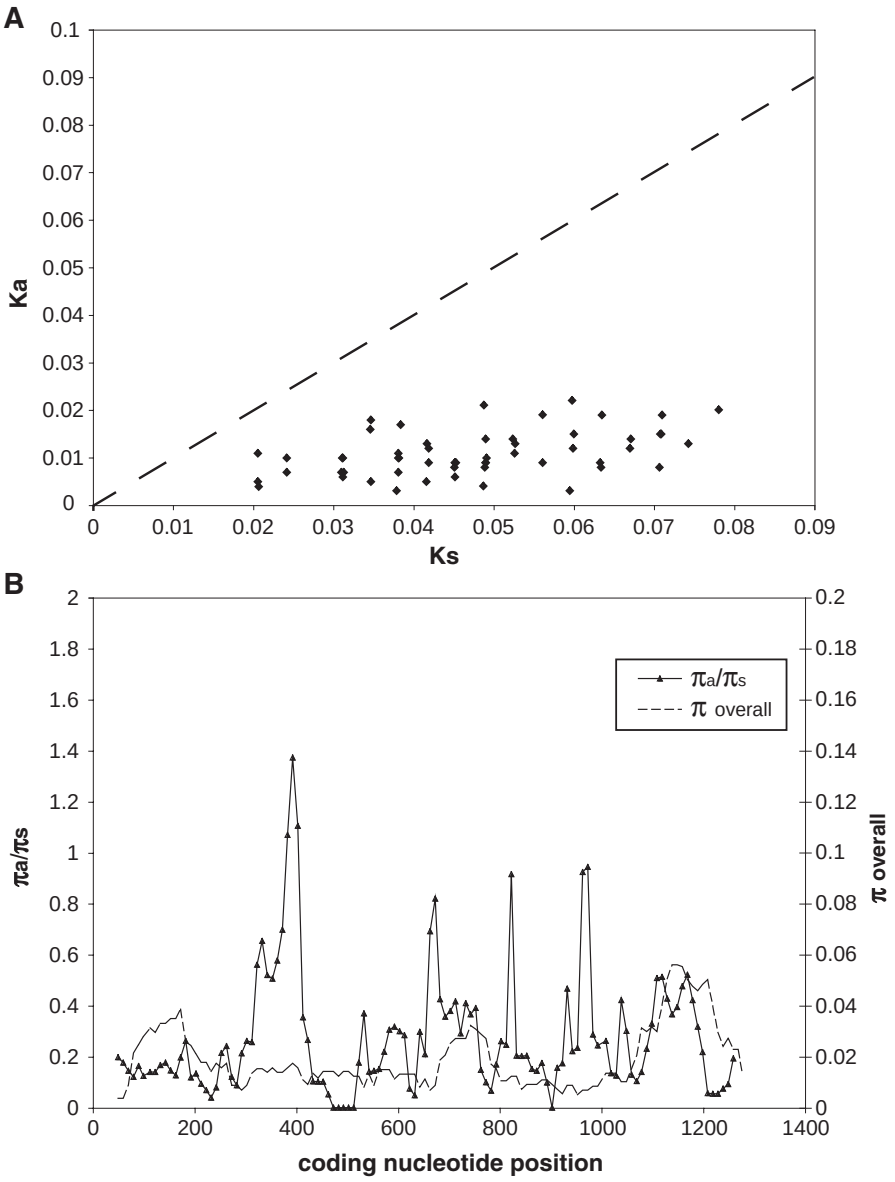

Fig. 3. (A) Plot of synonymous change per site (K_s) versus nonynonymous change per site (K_a) over all codons of *Cr-fem-3* for all pairwise comparisons between haplotypes. The dashed diagonal represents the purely neutral expectation of equal rates. (B) Sliding window estimation of the average pairwise ratio of nonsynonymous (π_a) and synonymous (π_s) polymorphisms (solid line) and overall nucleotide diversity (π, dashed line) for the concatenated coding sequences of the 11 *Cr-fem-3* alleles, corrected with by the Jukes–Cantor method. The window is 100 nt long, and is slid in steps of 10 nt.

1, the two other nuclear sequences available for the same *C. remanei* strains (Graustein et al., 2002). Using DnaSP, we estimated that the extracellular domain of *Cr-tra*-2 has experienced 3 synonymous and 1 nonsynonymous mutations, and the EGF/LNR region of *Cr-glp-1* has 13 synonymous and 2 nonsynonymous changes. Similarly, the *Cr-tra-2* F3BD has an excess of synonymous mutation, 8, versus 3 nonsynonymous. Thus, despite its remarkable interspecies divergence (26% identity for conserved sites and 6 indels compared with *C. briggsae*), *Cr*-TRA-2c F3BD is not unusually polymorphic or divergent between distant populations.

In contrast to these other coding sequences, *Cr*-FEM-3 has experienced 35 synonymous and 30 nonsynonymous changes. The extent to which intraspecific polymorphism is biased towards silent sites is significantly different (χ^2=5.47, *df*=1, *p*=0.019) between *Cr-glp-1* and *Cr-fem-3*, and this is largely explained by an elevated amount of nonsynonymous polymorphism in *fem-3*. Relative to FEM-3, GLP-1 shows roughly twofold higher overall sequence conservation between *C. remanei* and *C. briggsae* (Rudel and Kimble, 2001). *Cr-fem-3* therefore has elevated levels of intraspecies polymorphism that parallels its extreme interspecies divergence, but *Cr-tra-2* does not. This difference raises the interesting possibility that the rate of *fem-3* interspecies divergence is relatively uniform and mostly due to neutral processes, while that of *tra-2* may be more influenced by episodic selection.

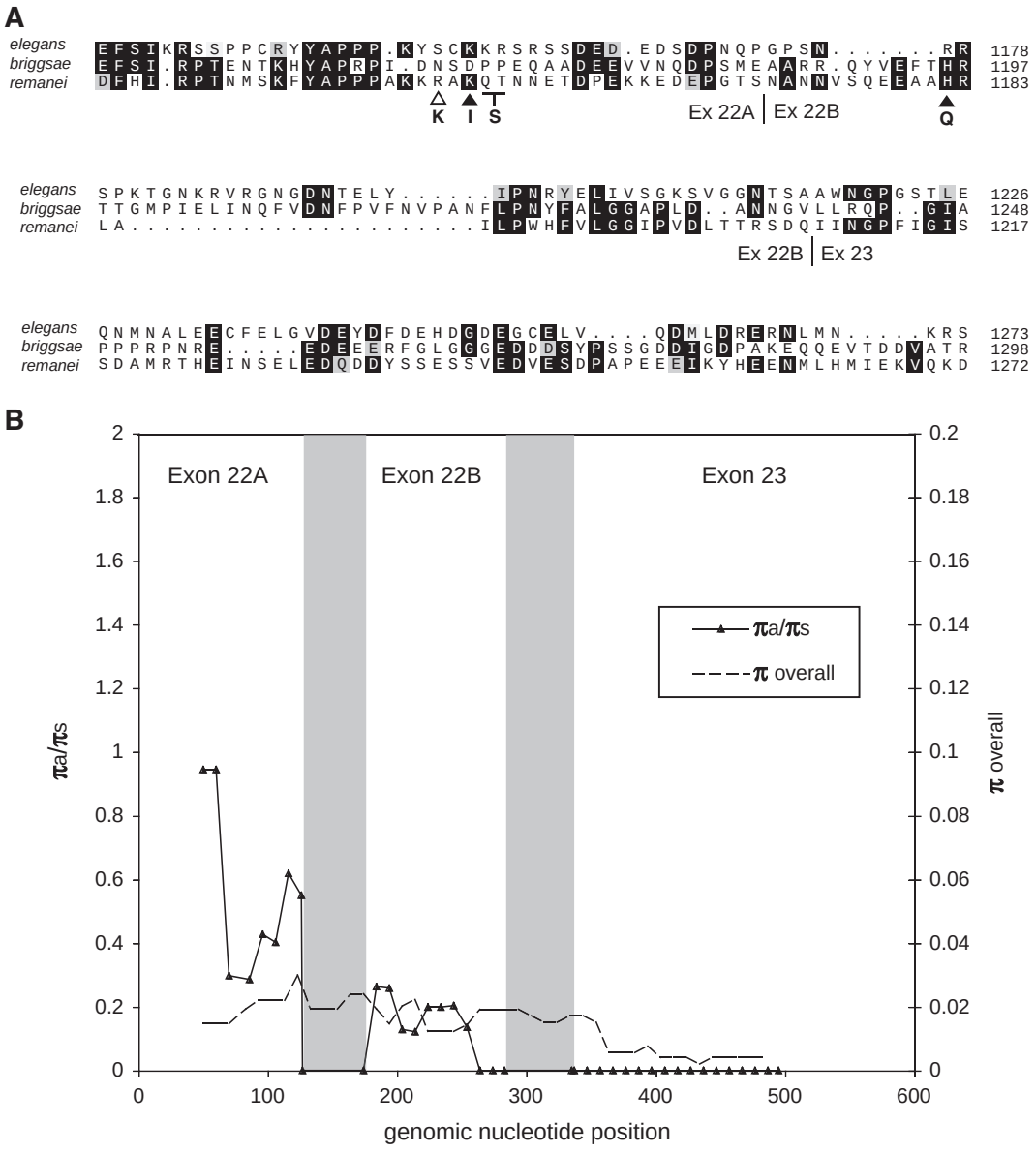

Fig. 4. Amino acid polymorphisms in *C. remanei* TRA-2c F3BD and their relation to interspecies conservation. (A) Interspecies amino acid alignment (from Haag et al., 2002) with *C. remanei* polymorphisms indicated below. As in Fig. 2, open triangles represent singleton variants and closed triangles those found in multiple alleles. The mark below *C. remanei* residues 1153–54 indicates that these are replaced by a single serine in 6 of the 10 alleles. Each TRA-2c fragment is numbered at the right of each row with respect to the complete TRA-2A proteins. The positions of two *Cr-tra-2* introns are indicated below the alignment with vertical bars, flanked by the relevant exon numbers. (B) Sliding window analysis of nucleotide variation across the *Cr*-TRA-2c F3BD, with both the π_a/π_s ratio and overall π plotted as in Fig. 3B. Gray bars represent the extent of intron sequences, with the flanking exons numbered above. Sequences of all alleles have been deposited with GenBank as accessions AY841138–AY841146.

3.6. Conclusions

That *Cr*-FEM-3 has a large amount of amino acid polymorphism is somewhat surprising given that this average sized protein makes critical protein–protein contacts with both FEM-2 and TRA-2c (Chin-Sang and Spence, 1996; Mehra et al., 1999; Haag et al., 2002). This is even more surprising because even small N- or C-terminal truncations of FEM-3 eliminate TRA-2c binding, indicating that little of its length is dispensable (Mehra et al., 1999). The species specificity seen in the FEM-3/TRA-2c (Haag et al., 2002) and FEM-3/FEM-2 (Stothard and Pilgrim, personal communication) complexes indicates that FEM-3 evolution is constrained, but in no way stopped, by its interactions.

Species specificity further implies that species-specific amino acids are crucial for these interactions. The data presented here define a set of nonconserved but non-polymorphic sites that may be good candidates for these species-specific contact points. However, in some cases polymorphic residues will be involved in key contacts because each variant works equally well. These alternative

residues would present distinct contexts for new mutations that arise, perhaps in some cases with distinct phenotypic consequences. A recent modeling study suggests that neutral polymorphisms may indeed be an essential precursor for some apparently compensatory changes in interacting proteins (Haag and Molla, manuscript submitted). No structural data currently exist for FEM-3 or its complexes, but when available it will be of great interest to examine the polymorphic and invariant residues defined by this study in their three-dimensional context.

Acknowledgements

We thank Brian Verrelli, Philip Awadalla, and Willie Swanson for their helpful analysis suggestions, several anonymous reviewers for their comments on the manuscript, and Michael Palopoli for the strains. This work was supported by startup funds from the University of Maryland and an Undergraduate Biological Sciences Education Program fellowship from the Howard Hughes Medical Institute to ADA.

References

Ahringer, J., Rosenquist, T.A., Lawson, D.N., Kimble, J., 1992. The *Caenorhabditis elegans* sex determining gene *fem-3* is regulated post-transcriptionally. EMBO J. 11, 2303–2310.

Baird, S.E., 1999. Natural and experimental associations of *Caenorhabditis remanei* with *Trachelipus rathkii* and other terrestrial isopods. Nematology 1, 471–475.

Chen, P., Cho, S., Jin, S., Ellis, R., 2001. Specification of germ cell fates by FOG-3 has been conserved during nematode evolution. Genetics 158, 1513–1525.

Chin-Sang, I.D., Spence, A.M., 1996. *Caenorhabditis elegans* sex-determining protein FEM-2 is a protein phosphatase that promotes male development and interacts directly with FEM-3. Genes Dev. 10, 2314–2325.

Civetta, A., Singh, R., 1998. Sex-related genes, directional sexual selection, and speciation. Mol. Biol. Evol. 15, 901–909.

Cline, T., Meyer, B., 1996. Vive la difference: males vs. females in flies vs. worms. Annu. Rev. Genet. 30, 637–702.

de Bono, M., Hodgkin, J., 1996. Evolution of sex determination in *Caenorhabditis*: unusually high divergence of *tra-1* and its functional consequences. Genetics 144, 587–595.

Goodwin, E., Ellis, R., 2002. Turning clustering loops: sex determination in *Caenorhabditis elegans*. Curr. Biol. 12, R111–R120.

Graustein, A., Gaspar, J.M., Walters, J.R., Palopoli, M.F., 2002. Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. Genetics 161, 99–107.

Haag, E., Kimble, J., 2000. Regulatory elements required for development of *C. elegans* hermaphrodites are conserved in the *tra-2* homologue of *C. remanei*, a male/female sister species. Genetics 155, 105–116.

Haag, E.S., Wang, S.P., Kimble, J., 2002. Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*. Curr. Biol. 12, 2035–2041.

Hodgkin, J., 1986. Sex determination in the nematode *C. elegans*: analysis of tra-3 suppressors and characterization of fem genes. Genetics 114, 15–52.

Jansa, S., Lundrigan, B., Tucker, P., 2003. Tests for positive selection on immune and reproductive genes in closely related species of the murine genus *Mus*. J. Mol. Evol. 56, 294–307.

Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., Fitch, D.H.A., 2004. *Caenorhabidits* phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc. Natl. Acad. Sci. U. S. A. 101, 9003–9008.

Kuwabara, P.E., 1996. Interspecies comparison reveals evolution of control regions in the nematode sex-determining gene *tra-2*. Genetics 144, 597–607.

Kuwabara, P.E., Perry, M., 2001. It ain't over till it's ova: germline sex determination in *C. elegans*. BioEssays 23, 596–604.

McAllister, B., McVean, G.A.T., 2000. Neutral evolution of the sex-determining gene transformer in *Drosophila*. Genetics 154, 1711–1720.

Mehra, A., Gaudet, Heck, J., Kuwabara, P.E., Spence, A.M., 1999. Negative regulation of male development in *Caenorhabditis elegans* by a protein–protein interaction between TRA-2A and FEM-3. Genes Dev. 13, 1453–1463.

Moreira, M.A., 2002. SRY evolution in Cebidae (Platyrrhini: Primates). J. Mol. Evol. 55, 92–103.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia Univ. Press, New York.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

O'Neil, M., Belote, J., 1992. Interspecific comparison of the transformer gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. Genetics 131, 113–128.

Pamilo, P., O'Neill, R., 1997. Evolution of the *Sry* genes. Mol. Biol. Evol. 14, 49–55.

Pilgrim, D., McGregor, A., Jackle, P., Johnson, T., Hansen, D., 1995. The *C. elegans* sex-determining gene *fem-2* encodes a putative protein phosphatase. Mol. Biol. Cell 6, 1159–1171.

Pires-da Silva, A., Sommer, R., 2004. Conservation of the global sex determination gene *tra-1* in distantly related nematodes. Genes Dev. 18, 1198–1208.

Rozas, J., Rozas, R., 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15, 174–175.

Rudel, D., Kimble, J., 2001. Conservation of *glp-1* regulation and function in nematodes. Genetics 157, 639–654.

Sivasundar, A., Hey, J., 2003. Population genetics of *Caenorhabditis elegans*: the paradox of low polymorphism in a widespread species. Genetics 163, 147–157.

Stein, L., et al., 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1, 166–192.

Stothard, P., Hansen, D., Pilgrim, D., 2002. Evolution of the PP2C family in *Caenorhabditis*: rapid divergence of the sex-determining protein FEM-2. J. Mol. Evol. 54, 267–282.

Sudhaus, W., Kiontke, K., 1996. Phylogeny of Rhabditis subgenus *Caenorhabditis* (Rhabditidae, Nematoda). J. Zoolog. Syst. Evol. Res. 34, 217–233.

Tucker, P., Lundrigan, B., 1993. Rapid evolution of the sex determining locus in old world mice and rats. Nature 364, 715–717.

Wood, W.B., 1988. The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13, 555–556.