Characterization of the p.Q189X nonsense mutation in dpy-17 in *C. elegans*

By: Alyssa Li

Wellesley College, Department of Biological Sciences

**Abstract**

The need to characterize the genetic basis for phenotypes in *Caenorhabditis elegans* arises from its astonishingly high homology to humans and other eukaryotic organisms. In particular, aspects of the *dpy* genes’ structure and function, which encodes a flexible and resilient exoskeleton (i.e., cuticle collagen) crucial for normal body morphology in *C. elegans*, can be applied to other organisms. We report the location and characterization of a functionally significant *dpy* gene mutation in *C. elegans*. The *dpy* gene of interest was determined to be located on chromosome 3 and linked to *unc-32* via linkage testing. The *dpy* gene was found to be 14.7 map units (m.u.) from *unc-32* via mapping. Complementation analysis indicated that the gene of interest was *dpy-17*, a previously characterized gene. The mutant *dpy-17* gene was sequenced and analyzed to understand the effects of the mutation at the nucleotide and amino acid level. The nature and extent of the change in the *dpy-17* gene in the mutant vs. wild type (WT) *C. elegans* was determined to be a nonsense mutation via PCR and Sanger sequencing, resulting in an early stop codon in the mutant. The WT polypeptide sequence is nearly twice as long as the mutant peptide, as predicted by DNAstar. Our work in studying the molecular consequences of mutations in *dpy-17* in *C. elegans* is extremely valuable, as it provides insight into the gene’s significant relevance to Marfan syndrome in humans and may inform new and effective treatments in the future.

**Introduction**

There is an urgency to discover the genetic basis for functionally significant mutations in model organisms in order to elucidate the relationship between gene and function in humans. Forward genetic analysis provides an unbiased approach for characterizing the genes of important phenotypes, such as those expressed in a disease. *Caenorhabditis elegans* serves as an excellent model organism for genetic studies and provides relevant insight into human biology and disease, including Marfan syndrome. Marfan syndrome is an autosomal dominant disorder of the connective tissue in humans that affects more than 200,000 people in the US every year (Fotopoulus et al., 2015). Creating and
analyzing mutant phenotypes in *C. elegans* in forward genetic experimentation are therefore highly applicable in analyzing and possibly treating human diseases.

One of the most important structural component in nematodes, such as *C. elegans*, is collagen. *C. elegans* has over 100 genes that encode for collagen (Cox et al. 1984). Collagens also play a significant role in the extracellular matrix of all multicellular organisms. About 30 genes encode collagen in mammals and 25% of total body protein is a form of collagen (van der Rest and Garrone 1991). In *C. elegans*, the *dpy* gene family encodes cuticle collagen, a flexible and resilient exoskeleton that is crucial for normal body morphology (Brenner, 1974). The expression of cuticle collagens in *C. elegans* increases during the four larval stages, as a new cuticle is synthesized (Cox et al. 1981; Cox and Hirsh 1985). Mutations in the *dpy* gene family will thus impair normal growth and developmental processes, including hermaphrodite tail development and posterior canal cell process outgrowth (Brenner, 1974).

Previous studies have investigated the role of *dpy* genes in the interaction with other genes to determine a larger functional role. For instance, Novelli et al. suggests that loss-of-function mutations in the *dpy-17* gene in *C. elegans* can prevent proper localization of SQT-3, another essential cuticle collagen, and may thus lead to the Dpy phenotype (2006). Fotopoulus et al. focused on the significant relevance of the *dpy-17* gene in *C. elegans* to Marfan syndrome because *dpy-17* suppresses the function of *mua-3*, a homolog of mammalian fibrillin1, via interactions with *dpy-31* (Novelli et al. 2004), which is required for TGFb activation in mammals (2015). Human fibrillin1 mutants cannot suppress TGFb, and this excessive TGFb signaling contributes to Marfan syndrome (Fotopoulus et al., 2015). In this study, we report the identification of the *dpy* gene responsible for the coiled, short and fat Dpy phenotype. We have determined its linkage to *unc-32* on chromosome 3 via linkage analysis and its 14.7 m.u. distance from *unc-32* via mapping. Complementation analysis suggested that the *dpy* gene was specifically *dpy-17*. Furthermore, we discovered that a nonsense mutation in the *dpy-17* mutant led to a truncated protein almost half the size of the WT protein via PCR and Sanger sequencing to perform DNA sequencing analysis.
**Results**

*Fig. 1.* Phenotypic characteristics of L4 wild type (WT) (A) vs. Dpy (B) vs. Unc (C) hermaphrodite *C. elegans*. (A), (B), and (C) also illustrate the vulval clearing indicative of L4 stage hermaphrodite worms.

*Fig. 2.* Resulting heterozygous F1 male *C. elegans* (phenotypically WT) from the parental cross between three L4 Dpy hermaphrodites and three WT males. Males are distinguished by their characteristic fan-shaped tail.

*C. elegans* is a free-living soil nematode, widely used in investigating gene functions and mutations in order to determine the genetic basis for a certain phenotype. Notable gene mutations, in particular, the *dpy* genes in *C. elegans*, have significant relevance to human diseases. Different strains of *dpy* cause worms to be short and fat relative to WT worms, but does not affect movement (Fig. 1A and 1B), whereas *unc* impairs movement, but does not impact morphology (Fig. 1C).

**Autosomal inheritance in the *dpy* gene**

To determine whether the *dpy* gene of interest is located on the autosome (i.e., autosomal inheritance) or the X chromosome (i.e., X-linked), three L4 Dpy hermaphrodites were crossed with three WT males in the parental cross. The cross produced heterozygous hermaphrodite and male progeny that were both phenotypically WT. It can thus be concluded that the *dpy* gene is located on the autosome and not the X chromosome; an X-linked gene would be expressed in all of the male progeny since *C. elegans* males are hemizygous (i.e. only have one X chromosome) and would receive their single X chromosome with the *dpy* mutation from the *dpy* hermaphrodite parent. Since the *dpy* gene was not expressed in F1 male progeny (Fig. 2), the mode of inheritance of the *dpy* gene is autosomal recessive; it is masked by the dominant, WT form of *dpy* on the autosome.
Linkage analysis indicates *dpy* gene linked to *unc-32* on chromosome 3

To discover the location of the mutated gene in the *C. elegans* genome responsible for the Dpy worms, a set of 5 *unc* reference marker genes was used to perform linkage testing: *unc-13* on chromosome 1, *unc-104* on chromosome 2, *unc-32* on chromosome 3, *unc-17* on chromosome 4, and *unc-76* on chromosome 5. Three F1 heterozygous males were crossed with three L4 hermaphrodites that were homozygous for one of 5 known *unc* mutations. The F2 self-cross was created using one L4 heterozygous hermaphrodite worm that was phenotypically WT for each marker strain. Four replicates of each strain were created. A total of 3086 F2 worm progeny was scored randomly across all 20 self-crosses.

It was found that the *dpy* gene of interest is on the same chromosome as *unc-32*, indicating that the two genes are linked on chromosome 3. The F2 self-cross produced an observed phenotypic ratio of 8.3 WT: 3.8 Dpy: 3.9 Unc: 0.25 Dpy Unc. The Dpy Unc ratio deviates from the expected 1/16 Dpy Unc ratio in 9 WT: 3 Dpy: 3 Unc: 1 Dpy Unc, which is typically seen in a cross between heterozygous individuals (i.e. dihybrid cross) (Table 1C). The dihybrid cross illustrates the different genotypes and phenotypes that can be generated, given that the probability of generating each of the four gametes during meiosis (i.e. the process whereby haploid gametes are generated from diploid cells) is equal in both parents. Since the ratio of Dpy Unc is much lower than the ratio expected from independent assortment, it is likely that the *dpy* gene of interest is linked to *unc-32*. The Dpy Unc worms are rare because 2 recombination events must occur to give rise to a double mutant; the chances that 2 recombinant gametes will come together to form the double mutant is much smaller than the 1/16 chance in a dihybrid cross via independent assortment, especially considering that the recombination frequency is <50%.

It was also found that the *dpy* gene of interest is not on the same chromosome as the other 4 *unc* reference markers: *unc-13* (Table 1A), *unc-104* (Table 1B), *unc-17* (Table 1D), and *unc-76* (Table 1E). The observed phenotypic ratio for each of the 4 *unc* markers approaches the expected 9:3:3:1 ratio, suggesting that the *dpy* gene of interest is on a different autosome than the *unc* marker. Hence, the *dpy* gene is not on chromosome 1, 2, 4, or 5.
Mapping shows dpy gene and unc-32 are 14.7 m.u. apart

To examine how closely linked dpy-17 and unc-32 are on chromosome 3, the distance in map units (m.u.) was determined by creating a “test cross” in order to calculate recombination frequency (RF). Four L4 Dpy Unc hermaphrodites were crossed with four WT males in the parental cross in order to create male heterozygotes. The F1 male heterozygotes were “test crossed” with Dpy Unc hermaphrodites to phenotypically distinguish the progeny of parental gametes from those of recombinant gametes. A total of 1185 F1 adult worm progeny was scored randomly.

The RF was determined to be 14.7% ([number of single mutants (recombinant progeny)/total progeny] *100%). Since RF is directly proportional to m.u. (i.e., 1% RF = 1 m.u.), the distance between the dpy gene of interest and unc-32 is 14.7 m.u (Table 2).

**Table 1.** Observed vs. expected number of F2 hermaphrodite progeny of *C. elegans* with WT, Dpy, Unc, and Dpy Unc phenotypes, and the resulting observed vs. expected ratio of each phenotype in each of the 5 unc reference marker genes: *unc-13* on chromosome 1 (A), *unc-104* on chromosome 2 (B), *unc-32* on chromosome 3 (C), *unc-17* on chromosome 4 (D), and *unc-76* on chromosome 5 (E). A total of 3086 worms were randomly scored.

(A) *unc-13* on chromosome 1

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>Dpy</th>
<th>Unc</th>
<th>Dpy Unc</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># observed</td>
<td>586</td>
<td>214</td>
<td>220</td>
<td>137</td>
<td>1157</td>
</tr>
<tr>
<td># expected</td>
<td>651</td>
<td>217</td>
<td>217</td>
<td>72</td>
<td>1157</td>
</tr>
</tbody>
</table>

(B) *unc-104* on chromosome 2

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>Dpy</th>
<th>Unc</th>
<th>Dpy Unc</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># observed</td>
<td>334</td>
<td>122</td>
<td>158</td>
<td>43</td>
<td>657</td>
</tr>
<tr>
<td># expected</td>
<td>370</td>
<td>123</td>
<td>123</td>
<td>41</td>
<td>657</td>
</tr>
</tbody>
</table>

(C) *unc-32* on chromosome 3

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>Dpy</th>
<th>Unc</th>
<th>Dpy Unc</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># observed</td>
<td>197</td>
<td>90</td>
<td>93</td>
<td>6</td>
<td>386</td>
</tr>
<tr>
<td># expected</td>
<td>218</td>
<td>72</td>
<td>72</td>
<td>24</td>
<td>386</td>
</tr>
</tbody>
</table>

(D) *unc-17* on chromosome 4

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>Dpy</th>
<th>Unc</th>
<th>Dpy Unc</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># observed</td>
<td>230</td>
<td>86</td>
<td>84</td>
<td>34</td>
<td>384</td>
</tr>
<tr>
<td># expected</td>
<td>216</td>
<td>72</td>
<td>72</td>
<td>24</td>
<td>384</td>
</tr>
</tbody>
</table>

(E) *unc-76* on chromosome 5

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>Dpy</th>
<th>Unc</th>
<th>Dpy Unc</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># observed</td>
<td>287</td>
<td>97</td>
<td>85</td>
<td>33</td>
<td>502</td>
</tr>
<tr>
<td># expected</td>
<td>283</td>
<td>94</td>
<td>94</td>
<td>31</td>
<td>502</td>
</tr>
</tbody>
</table>
Complementation analysis indicates \textit{dpy} gene of interest is \textit{dpy-17}

To discover the identity of the \textit{dpy} gene of interest, a complementation test was performed by crossing the \textit{dpy} gene of interest with three other known \textit{dpy} genes on chromosome 3: \textit{dpy-1}, \textit{dpy-17}, and \textit{dpy-18}. The parental cross was created by crossing three WT males with three Dpy hermaphrodites that have the \textit{dpy} gene of interest in order to create heterozygous males. Three F1 crosses were generated by crossing three F1 heterozygous males with three L4 Dpy hermaphrodites from each of the three \textit{dpy} reference strains. The F1 male progeny was observed.

It was found that \textit{dpy-17} failed to complement the \textit{dpy} gene of interest (i.e., fail to rescue the WT phenotype), suggesting that the two genes are allelic (i.e., alleles of the same gene) and thus, fail to be masked by the dominant WT allele. The failure to complement indicates that the \textit{dpy} gene of interest is \textit{dpy-17}. In contrast, the other \textit{dpy} reference strains (\textit{dpy-1} and \textit{dpy-18}) complement the \textit{dpy} gene of interest. The complementary \textit{dpy} genes are therefore non-allelic and the \textit{dpy} mutant alleles can thus be masked by their respective WT allele, resulting in the restoration of the WT phenotype. These data suggest that the \textit{dpy} gene of interest is neither \textit{dpy-1} nor \textit{dpy-18}. Furthermore, the findings confirm the data from the linkage analysis and mapping indicating that the \textit{dpy} gene of interest is located on chromosome 3.

\textbf{DNA sequencing reveals nonsense mutation in \textit{dpy-17}}

To determine the nature and extent of the change in the \textit{dpy-17} gene in mutants vs. WT \textit{C. elegans}, the \textit{dpy-17} gene was amplified via polymerase chain reaction (PCR) and the DNA template (i.e. PCR product) was sequenced via Sanger sequencing using selective incorporation of chain-terminating dideoxynucleotides (ddNTPs) during \textit{in vitro} DNA replication. The sequence was analyzed using DNASTar (Lasergene) at the DNA and protein level to understand the molecular effects of the \textit{dpy-17} gene mutation in \textit{C. elegans}.

Analysis of the \textit{dpy-17} DNA sequence in DNASTar indicated a single base substitution (i.e., point mutation) at 617 base pair (bp) from cytosine (C) in the WT to thymine (T) in the mutant (Fig. 3A). The base alteration introduced an early stop codon when the DNA sequence was transcribed.

\begin{table}[ht]
\centering
\caption{Observed number of F1 progeny of \textit{C. elegans} with WT, Dpy, Unc, and Dpy Unc phenotypes, resulting in a RF of 14.7\% and a 14.7 m.u distance between \textit{dpy} and \textit{unc-32}.}
\begin{tabular}{|c|c|c|c|c|c|}
\hline
 & WT & Dpy & Unc & Dpy Unc & Total \\
\hline
\# observed & 490 & 91 & 83 & 521 & 1185 \\
\hline
\end{tabular}
\end{table}

\begin{align*}
RF = \frac{174}{1185} = 14.7\% \\
14.7\% \text{ RF} = 14.7 \text{ m.u.}
\end{align*}
translated to a polypeptide sequence. This point mutation is therefore a nonsense mutation, in which the CAG codon, which corresponds to glutamine (Q), was changed to TAG, which corresponds to the stop codon, resulting in a shortened protein. The WT protein sequence is 352 amino acids (aa) long while the truncated protein is 189 aa long (Fig. 3B). Thus, the mutant protein sequence was nearly half the length of the WT sequence, likely resulting in the Dpy phenotype.

\[(A) \text{dpy-17 DNA}\]

\[
\begin{array}{cccc}
\text{WT} & 5' & \ldots & \text{AAGGCT} & \text{CAGACT}\ldots & 3' \\
\text{Mutant} & 5' & \ldots & \text{AAGGCT} & \text{TAGACT}\ldots & 3'
\end{array}
\]

\[(B) \text{dpy-17 Protein}\]

\[
\begin{array}{cccc}
\text{WT} & \text{Nterm} & \ldots & K & A & T & \ldots & \text{Cterm} & \text{Length: 352 aa} \\
\text{Mutant} & \text{Nterm} & \ldots & K & A & \text{Cterm} & \text{Length: 189 aa}
\end{array}
\]

**Fig. 3.** Single base pair (bp) substitution resulting in a stop codon (i.e., nonsense mutation) in dpy-17 C. elegans mutant. (A) shows the point mutation at 617 bp from C in the WT to T in the mutant, which led to an early stop codon (DNA sequence: TAG) in the mutant instead of the Q amino acid (aa) (DNA sequence: CAG) in the WT at 189 aa (B).

**Discussion**

The genetic basis for a specific Dpy phenotype in C. elegans was characterized. The location of the functionally significant dpy gene mutation was determined via linkage testing to be on chromosome 3 and linked to unc-32. Using mapping, the dpy gene of interest and unc-32 was found to be 14.7 m.u. apart. Complementation analysis of the dpy gene of interest with three other known dpy reference markers indicated that dpy-17 was the dpy gene of interest since the genes failed to complement each other. The dpy-17 gene on chromosome 3 was sequenced and analyzed to understand the effects of the mutation at the nucleotide and amino acid level. The nature and extent of the change in the dpy-17 gene in the mutant versus WT C. elegans was determined to be a nonsense mutation at 617 bp, resulting in an early stop codon in the mutant. The WT polypeptide sequence is 352 aa in length; the mutant peptide is only 189 aa in length as predicted by DNAstar.

Mapping analysis indicated that the distance between the dpy-17 gene and marker unc-32 was 14.7 m.u., representing a recombination frequency between the two loci of 14.7%. This finding is significant in that it suggests that double recombination (i.e., generation of Dpy Unc double-mutants)
is highly unlikely. This is likely a consequence of evolution; nematodes that were phenotypically Dpy Unc were possible selected against due to their uncoordinated behavior and short morphology. Thus, there may be strong linkage between the \textit{dpy-17} mutant allele and the \textit{unc-32} WT allele; there is also strong linkage association between the \textit{dpy-17} WT allele and the mutated \textit{unc-32} WT allele, thereby reducing the possibility of Dpy Unc mutant offspring.

Previous findings from Brenner (1974) report the mapping distance between \textit{dpy-17} (e164) and \textit{unc-32} (e189) to be 2.95 (2.0 - 4.19) m.u. (n = 1136). This number deviates from our reported mapping distance of 14.7 m.u. (n = 1185). It is likely that this deviation arises from the effects of stochasticity; a much larger sample of scored worms may remove these probabilistic effects. Moreover, the precise markers used (i.e., \textit{dpy-17}, e164 and \textit{unc-32}, e189) may be different from the markers used in our study, leading to differences in the measured mapping distance.

\textit{dpy-17} is a gene that encodes cuticle collagen, an extremely flexible and resilient exoskeleton in nematodes, such as \textit{C. elegans}, and is crucial for normal body morphology, hermaphrodite tail development, and outgrowth of posterior canal cell processes (Brenner, 1974). The molecular mechanism by which this nonsense mutation causes the Dpy phenotype (Fig. 1) is not fully understood. Following transcription, the mRNA containing the nonsense mutation can either be fully transcribed to form a truncated protein or be destroyed by nonsense mediated decay, effectively “knocking out” this gene. In the first case, it is likely that the truncated protein product is either nonfunctional (i.e., loss of binding sites or catalytic domains) or deleterious to the organism (i.e., loss of regulatory domains or localization signals). In the second case, it is likely that the loss of the WT protein causes the phenotype. In order to determine how the nonsense mutation at 617 bp in \textit{dpy-17} causes the Dpy phenotype, further experiments, including Western Blot analysis or gene expression analysis by quantitative PCR, are necessary.

Most importantly, genetic studies have noted that \textit{dpy-17} genes in \textit{C. elegans} have significant relevance to Marfan syndrome, an autosomal dominant disorder of the connective tissue in humans (Fotopoulus et al., 2015). Common symptoms include vision defects, elongated extremities, scoliosis, and other skeletal defects. It is believed that monogenic defects in TGF signaling pathways, including \textit{fibrillin1}, are responsible for causing Marfan syndrome in humans. \textit{dpy-17} is highly homologous to isoform 3 of \textit{Col2A1} in humans (Sibley et al., 1994). Incidentally, in \textit{C. elegans}, \textit{dpy-17} suppresses the function of \textit{mua-3}, a homolog of mammalian \textit{fibrillin1}, via interactions with \textit{dpy-31} (Novelli et al. 2004), a BMP-1/Tolloid-like metalloprotease required for TGFb activation in mammals. Human \textit{fibrillin1}
mutants cannot suppress TGFβ, and this excessive TGFβ signaling contributes to Marfan syndrome (Fotopoulus et al., 2015).

Understanding how dpy-17 mutations interact with other gene targets has greatly enriched the knowledge of not just C. elegans, but also other species with dpy-17 homologs, including plants, flies, and humans. We have identified through sequencing and amino acid prediction, a nonsense mutation in dpy-17 in C. elegans that causes the Dpy phenotype. Mutations in dpy-17 have also been reported to have a suppressive effect on mutations in mua-3. Due to the strong homology between mua-3 and human fibrillin1, there is a unique possibility that similar mutations in Col2A1 (a homolog of dpy-17) may suppress mutant fibrillin1 activity and reduce symptoms in Marfan syndrome. It is also possible that the nonsense mutation at 617 bp in dpy-17 may serve as a potential gene therapy target in humans to treat or prevent Marfan syndrome, especially through new genome-editing techniques, including CRISPR/Cas9. Therefore, studying the molecular consequences of mutations in dpy-17 in C. elegans is extremely valuable, as it provides insight into its effects in other species, especially disease pathogenicity in human Marfan syndrome, and may inform new and effective treatments in the future.
Literature Cited


