Copyright © 2001 by the Genetics Society of America

Comparative Analysis of the *nonA* Region in Drosophila Identifies a Highly Diverged 5′ Gene That May Constrain *nonA* Promoter Evolution

Susanna Campesan,* David Chalmers,*,[1] Federica Sandrelli,† Aram Megighian,‡ Alexandre A. Peixoto,*,§ Rodolfo Costa† and Charalambos P. Kyriacou*

**Department of Genetics, University of Leicester, Leicester LE1 7RH, England, †Dipartimento di Biologia, Università di Padova, 35131 Padova, Italy, ‡Dipartimento di Anatomia e Fisiologia Umana, Università di Padova, 35131, Padova, Italy and §Departamento de Bioquimica e Biologia Molecular, Fundacao Oswaldo Cruz, Rio de Janeiro, CEP 21045-900, Brazil*

Manuscript received May 1, 2000
Accepted for publication November 7, 2000

ABSTRACT

A genomic fragment from *Drosophila virilis* that contained all the *no-on-transientA* (*nonA*) coding information, plus several kilobases of upstream material, was identified. Comparisons of *nonA* sequences and the gene *nonA-like* in *D. melanogaster*, a processed duplication of *nonA*, suggest that it arose before the split between *D. melanogaster* and *D. virilis*. In both species, another gene that lies <350 bp upstream from the *nonA* transcription starts, and that probably corresponds to the lethal gene *l(1)i19*, was identified. This gene encodes a protein that shows similarities to GPI1, which is required for the biosynthesis of glycosylphosphatidylinositol (GPI), a component for anchoring eukaryotic proteins to membranes, and so we have named it *dGpi1*. The molecular evolution of *nonA* and *dGpi1* sequences show remarkable differences, with the latter revealing a level of amino acid divergence that is as high as that of *transformer* and with extremely low levels of codon bias. Nevertheless, in *D. melanogaster* hosts, the *D. virilis* fragment rescues the lethality associated with a mutation of *l(1)i19e*, as well as the viability and visual defects produced by deletion of $nonA^{-}$. The presence of *dGpi1* sequences so close to *nonA* appears to have constrained the evolution of the *nonA* promoter.

MUTATIONS in the sex-linked, *no-on-transientA* (*nonA*) gene of *Drosophila melanogaster* produce behavioral defects in vision and in the male courtship song (Hotta and Benzer 1970; Pak *et al.* 1970; Kulkarni *et al.* 1988). The gene lies within cytological position 14C1-2, a region that also contains two lethal complementation groups situated immediately distally to *nonA*, *l(1)i19e* and *l(1)9-21* (Jones and Rubin 1990). One of these, *l(1)i19e*, appears to overlap with the 5′ half of *nonA*, but the sequences corresponding to this lethal gene have yet to be identified (Stanewsky *et al.* 1993). The *l(1)9-21* region encodes the pre-mRNA splicing factor U2AF (Kanaar *et al.* 1993) and more distally, the region 20–50 kb from *nonA* has revealed a number of cDNAs and open reading frames, the most studied being *cyclophilin-1* (Rutherford 1995). Conceptual translation of *nonA* predicts a protein of 700 amino acids (Besser *et al.* 1990; Jones and Rubin 1990). The central segment contains two tandemly repeated 80-amino-acid motifs, common to a family of proteins known for their ability to bind RNA. The RNA recognition motif, RRM, also known as RNP or RBD (Siomi and Dreyfuss 1997), is moderately conserved from yeast to humans (Bandziulis *et al.* 1989).

The original *nonA* mutants have defects in their visual system but their courtship song is unaffected, whereas the first *nonA* song mutant, *dissonance* (later renamed $nonA^{diss}$), has song pulses that appear reasonably normal at the beginning of a song burst, but become polycyclic as the burst progresses (Hotta and Benzer 1970; Pak *et al.* 1970; Kulkarni *et al.* 1988). Like other visual mutants, $nonA^{diss}$ is also defective in its abnormal electroretinogram (ERG) and optomotor response, suggesting both peripheral and central visual system lesions (Kulkarni *et al.* 1988; Rendahl *et al.* 1992, 1996; Stanewsky *et al.* 1996). Amino acid substitutions within or very close to the second RRM of NONA produce visual but not song abnormalities, whereas the $nonA^{diss}$ song mutation creates an amino acid substitution in a downstream region notable for the high proportion of charged residues (Rendahl *et al.* 1996).

The pulse structure of the songs of $nonA^{diss}$ mutants resembles, at least superficially, that of *D. virilis* (Hoikkala and Lumme 1984, 1987). Specifically, pulses of *D. virilis* are more polycyclic compared to wild-type *D. melanogaster*, but they also show the additional $nonA^{diss}$ feature of increasing the number of cycles per pulse as the song burst progresses (S. Campesan, Y. Dubrova, J. C. Hall and C. P. Kyriacou, unpublished results). These observations stimulated us to attempt to identify

Corresponding author: C. P. Kyriacou, Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom. E-mail: cpk@leicester.ac.uk

[1]*Present address:* ETS/FC, 25020 Besancon, France.

the *D. virilis nonA* orthologue and to analyze in detail the courtship song phenotypes of *D. melanogaster* transformants carrying the *virilis* transgene. Similar interspecific transformation experiments performed with the *period* (*per*) gene have revealed that *per* carries species-specific behavioral information for both circadian locomotor activity patterns and for ultradian love song cycles (Petersen *et al.* 1988; Wheeler *et al.* 1991). Although severe, the $nonA^{diss}$ mutation is not amorphic (Stanewsky *et al.* 1993), so it is preferable for the proposed study that the *D. virilis nonA* transgene be placed on a *nonA* null background in *D. melanogaster*. This can be provided by the use of the translocation *T(1;4)9e2-10*, in which both *nonA* and the distal, partially overlapping lethal locus *l(1)i19e* are deleted (Jones and Rubin 1990). The most efficient strategy would be to identify a fragment from the *D. virilis nonA* region that encodes both *nonA* and the adjacent lethal locus and to transform this fragment into the double gene deletion background provided by *T(1;4)9e2-10*.

Our comparative analysis of *D. virilis* and *D. melanogaster nonA* thus also sought to identify the genomic and cDNA sequences corresponding to *l(1)i19e*. Consequently, we have isolated a *D. virilis* genomic fragment that encodes *nonA* and have identified within its promoter a gene we call *dGpi1*, which almost certainly corresponds to *l(1)i19e*. We compare the two species *nonA* sequences with those of *D. melanogaster nonA-like*, which lies in the bithorax complex (Martin *et al.* 1995), and make further comparisons between the two species coding sequences of *dGpi1*, which reveal extremely high levels of divergence. We also examine whether the presence of the *dGpi1* gene in the promoter region of *nonA* constrains the evolution of the *nonA* regulatory region. This is particularly relevant given that the accompanying article by Sandrelli *et al.* (2001) demonstrates that the transcription unit of *dGpi1* also acts as both a silencer and enhancer of various behavioral and tissue expression phenotypes of *nonA*. Finally, we investigate whether the transformed *D. virilis* fragment is able to rescue the lethality associated with *l(1)i19e* and the *nonA* ERG and viability defects.

MATERIALS AND METHODS

Identification of *D. virilis nonA* and *l(1)i19e*: A small fragment of the *D. virilis nonA* gene was amplified by PCR using 5′ primer 5′-CGCGAGATGTTCAAGCCATA-3′ (4163–4182) and 3′ primer 5′-GCCCTCTCGATGGGACCAAA-3′ (4422–4403), based on the second exon sequence of the *D. melanogaster nonA* gene (nucleotide positions from sequence of Jones and Rubin 1990). The amplified 261-bp fragment was sequenced to confirm its homology with *D. melanogaster nonA* and used to screen an EMBL3 phage *D. virilis* genomic library constructed by Ron Blackman and kindly donated by John Belote. DNA from one putative positive clone was digested with *Sph1*, and two fragments of 6.5 and 6.0 kb, respectively, which hybridized to a *D. melanogaster nonA* probe, were subcloned into pUC18. Manual sequencing was performed on various subclones of the two fragments and coding regions were confirmed several times on both strands. Intron-exon boundaries were studied by comparing sequences from RT-PCR products with genomic sequences. In addition 5′ rapid amplification of cDNA ends (RACE) was performed to reveal the transcription start site of both *D. melanogaster* and *D. virilis nonA*. RT-PCR and 5′ and 3′ RACE were also performed on putative *D. melanogaster* and *D. virilis* transcripts from the lethal gene *l(1)i19e*, which was believed to be embedded within the 5′ and N-terminal regions of *nonA* (Jones and Rubin 1990; Stanewsky *et al.* 1993). The two positive clones were then ligated into the pw8 transformation vector to reconstitute *D. virilis nonA*, including ~3 kb of upstream and 1 kb of downstream sequence, using a number of cloning steps. The integrity of *nonA* was confirmed by sequencing.

Analysis of sequence variation in *D. melanogaster* and *D. simulans nonA* promoter fragments: Single *D. melanogaster* males were obtained from five isofemale lines established in 1994 from a natural population sampled in Lecce (Italy), and *D. simulans* males were obtained from three isofemale lines established from a natural population from Zimbabwe. Single fly genomic DNA was prepared as previously described (Gloor and Engels 1990). An 863-bp fragment for *D. melanogaster* and an 880-bp fragment for *D. simulans*, located immediately upstream of the coding region of *nonA*, were amplified by using the forward primer 5′-GCGGGTACCCAGGTCGCAC TGAGTCCC-3′ (positions 974–991 in the sequence of Jones and Rubin 1990) and the reverse primer 5′-TATGGATCCGC TACAACTCGTTGACAA-3′ (positions 1849–1866). The amplified fragments were sequenced automatically.

Computer analyses: All sequence analyses were performed using the programs available from the Genetics Computer Group (GCG) package for molecular biology. A statistical analysis of cryptic simplicity in the coding sequence DNA was performed using the SIMPLE34 program, which generates a Relative Simplicity Factor (RelSF) for each sequence (Hancock and Armstrong 1994). The PSITE program was used to search for functional motifs in the NONA proteins (Solovyev and Kolchanov 1994). SIGNAL SCAN (Prestridge 1991) and TF SEARCH (Akiyama 1995; Heinemeyer *et al.* 1998) were used to search for functional motifs in the 5′ regulatory region of *nonA*.

Transformations: *P*-element-mediated transformation was performed using standard methods with the pW8 vector that carries *w*+ as a marker (Spradling and Rubin 1982; Klemenz *et al.* 1987). Embryos microinjected were either *w; Sb e* Δ*2-3/TM6* or w^{1118}. When using the latter injectees, transposase was provided by coinjection with PUChsπ Δ2-3 (a gift from J. M. Dura). A number of independent lines were obtained and the inserts were mapped to at least the chromosomal level. Southern blotting showed that all lines contained single copy insertions. Line 112 was sex linked and mapped close to endogenous *nonA*, and line 113 integrated on the *Sb e* Δ*2-3* chromosome and was crossed off to avoid further transposition via Δ*2-3*. Because the 113 insert was homozygous lethal, it was used in a mobilization assay to generate two further hops, 168-8 and 67-4, which complemented the lethality of insert 113. Lines 72 and 297 both contained homozygous viable X chromosome inserts. The 297 insert was successfully mobilized to chromosome 3 to give line 297-6. Lines 97, 135, and 191 contained single chromosome 3 insertions, and line 75 carried the transgene on chromosome 2.

Viability: Females heterozygous for *In(1)FM7* (marked with *y w B*) and the translocation *T(1;4)9e2-10* (marked with *y cv v f car*), which carries a deletion uncovering *nonA* and *l(1)19e* (Stanewsky *et al.* 1993), were crossed to males carrying an autosomal copy of the *D. virilis nonA* fragment. Ordinarily, translocation males can survive only if the deletion is comple-

mented by the product of *l(1)19e*, and even then, viability is compromised severely due to the lack of a *nonA* encoded product (Stanewsky *et al.* 1993). Consequently, the nonlethal F_1 males from this cross, assuming the *virilis* fragment confers 100% viability, are *T(1;4)9e2-10; transgene/+* (marked with *y cv v f car*), red-eyed *In(1)FM7; transgene/+* (the transgene carries *w+*), and white-eyed *In(1)FM7; +/+* males in approximately equal proportions.

Drosophila ERGs: Cold anesthetized flies were immobilized in dental wax, and one glass microelectrode (the reference electrode) was inserted in the median head region between eyes, and the other (recording electrode) was inserted in one eye just below the cornea. After a 10-min recovery period, flies were dark adapted for 5 min and then submitted to several 2- to 5-sec light stimulations using a DC-powered lamp mounted close to the head. Between each light stimulation, flies were again dark adapted for 60 sec. Recorded signals were enhanced with an intracellular amplifier (WPI Instruments), fed to a signal conditioner (Axon Instruments, Foster City, CA), lowpass filtered (3 kHz), and then fed to a PC through an A/D converter (Axon Instruments). The output signals from the signal conditioner were also displayed on a digital oscilloscope for online evaluation. The amplitude of ON and OFF ERG transients was measured using appropriate software (PClamp 6.04, Axon Instruments).

RESULTS

Intron-exon structure of *D. virilis nonA*: The five-exon/four-intron organization found in *D. melanogaster nonA* is conserved in the *D. virilis* homologue (Figure 1). The intron-exon boundaries are also conserved as revealed by cDNA and genomic DNA comparisons (data not shown). The approximate lengths of the first two introns (~2.4 and 2.0 kb), as calculated by measuring the length of PCR products obtained by using primers annealing to the exon boundaries, are two and four times, respectively, the sizes of their *melanogaster* counterparts. The third intron is the same size in both species (~70 bp) but the length of the fourth intron is unknown.

Sequence comparisons: Dot matrix comparisons between the *D. virilis* and the *D. melanogaster nonA* coding sequences revealed an area of considerable divergence covering approximately the first half of the gene (data not shown). Plots of each of the two *nonA* sequences against itself clearly showed numerous large regions of repeated DNA, clustered especially at the beginning and the end of the gene. *D. melanogaster nonA* appeared much less repetitive than its *virilis* counterpart, and this was confirmed by computing the RelSF for the two sequences (Hancock and Armstrong 1994), which gave values of 1.552 for *D. melanogaster* and 1.898 for *D. virilis*.

Figure 2 shows an alignment of the ~700-amino-acid sequence of *D. virilis* and *D. melanogaster* NONA, together with a third *D. melanogaster* protein encoded by the *nonA-like* gene (Martin *et al.* 1995). Overall identity between the two species NONA proteins is just over 75%, while the similarity is ~83% (Table 1). Pairwise comparisons revealed corresponding values of 68 and 72% between *D. melanogaster nonA-like* and both *D. melanogaster* and *D. virilis nonA*. The N-terminal third of the NONA protein up to the RNA-binding domain is the most diverged, with <50% identity (Figure 2). These regions of divergence are constituted in large part by stretches of repeats, particularly tracts of poly(Gly), and a QN and a degenerate GNQGGX repeat found in *D. melanogaster*, which has been replaced by a QA and a very long 29-residue poly(Gly) repeat in *D. virilis*. The RNA-binding domain (RRM1 + RRM2, residues 295–453 in *D. melanogaster nonA*) is very well conserved between the two species. The RNP1 octamer and RNP2 hexamer motifs within RRM1 are perfectly conserved, but RNP1 in RRM2 has two changes.

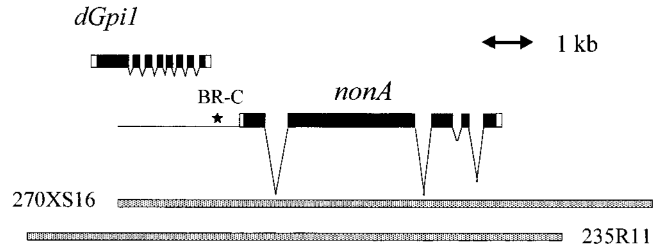

Figure 1.—The *nonA* interval. The 5 exon-4 intron structure on *nonA*, and 8 exon-7 intron arrangement of *dGpi1* are shown for both *D. melanogaster* and *D. virilis*. Translated exons are in black. In *D. melanogaster*, the first exon of *dGpi1* is initiated at −2372 bp, relative to the start codon of *nonA* (−2347 in *D. virilis*). The methionine start of *dGpi1* is at −2350 (−2257 in *D. virilis*). The exon coordinates of *D. melanogaster dGpi1* are −2372 to −1704 (1), −1649 to −1532 (2), −1470 to −1335 (3), −1273 to −1129 (4), −1076 to −1000 (5), −933 to −818 (6), −764 to −650 (7), and −582 to −487 (8). In *D. virilis* they are −2257 to −1632 (1), −1565 to −1457 (2), −1391 to −1256 (3), −1178 to −1034 (4), −967 to −891 (5), −831 to −716 (6), −653 to −539 (7), and −475 to −382 (8). In *D. melanogaster*, the *dGpi1* transcription termination is at −424, and the *nonA* transcription start is at −89, a few base pairs upstream from that found by Stanewsky *et al.* (1993). In *D. virilis*, *dGpi1* transcript termination is at −336, and *nonA* transcription start is at −84. The conserved putative Broad-Complex binding site is found at −417 in *D. melanogaster* and −335 in *D. virilis*, immediately downstream of transcription termination (see also Figure 6). Other BR-C sites are also found in this region if the search algorithms are used with a reduced stringency of 0.85 as opposed to 0.9. The genomic fragments 270XS16 (begins at −1909) and 235R11 that rescue *nonA* mutant phenotypes are shown. 235R11 also rescues the lethality associated with *l(1)i19e*, whereas 270XS16 rescues the lethality sporadically (Jones and Rubin 1990). The beginning of fragment 270XS16 corresponds to amino acid residue 148 of *dGpi1*.

The adjacent charged region (amino acids 454–568; see Figure 2) includes residue 548, in which an asparagine is substituted by cysteine in the $nonA^{diss}$ mutation in *D. melanogaster* (Rendahl *et al.* 1996). This position is conserved in both *nonA-like* and in *D. virilis*. Application of the PSITE program revealed that the sequence KRESDNE (residues 530–536) spanning the $nonA^{diss}$ site in *D. virilis* contains putative phosphorylation sites for both cyclic nucleotide-dependent and casein kinase II

```
NONA mel          MESAGKQDNN-ATQQLPQRQQRGNQQANKNLGKHNAQKQND-SADGGPAEKKQRFGGPNA   58
NONA vir          MENSVKMDNSGNSTPLPQRQRRANNQPNKNIGKLGPQKQNEGASDGGPAEKRQRFG-PNN   59
NONA-like         MEGAVK-KNSLNSSPLPQRQQRGNS-TNKNLGKPTPPKLNA-ASDGNPAEKKARLG-GNT   56
                  ** . *  *   .  *****.* *   ***.**    * *  ..**.****. * *  *

                     ▼
NONA mel          QNQNQNQNQNGGVTGGGGAVGGPNQNKNFGNNKGGFVGNRNRNNNRAGNQN--RT-FPGN   115
NONA vir          QNGGGGSVVGG--GGGGGGGGGQNQNKNFAN-KGGFGGGGNRNRNRGGNQN--RS-NFQN   113
NONA-like         QNGGGVAGGGG--TGGGGGGGG--------A-TGGVEFSRNRN--RGGNQNENRQGFQVA   103
                  ** .     .*   ****  **           **     ***  * ****  *

NONA mel          NNSNQKPNNETSKADGPNALAKNNEPATAAAGQNQANQNANKGQNQRQGQNQNQNQVHGQ   175
NONA vir          QNQNQKSTTDAPKADGGNLNDKSNE---ANN-ANQSNSNS---AAQAQAQLQAQAQAHAQ   166
NONA-like         NNSHQKQINESPKPAAGNVPAKNNELSAGGGGQNQPNHSN-----KGQGNQGDQGEQGNQ   158
                  .* .**  ... *    *   * **        ** *        . * .   * .   *

NONA mel          GNQGGPGNQGGAGNQGGQGNQGGAGNQGNGQGFRGRNAGNNQGGGFSGGPQNQQRDNRNR   235
NONA vir          A-------QAQAQAQAHAQAQAQAHAHAQNQAFRARGGGGGGGGGGGGGGGG-----GG   214
NONA-like         G--------PNFRGRGGGPNQP---NQNANQ-EQSNGYPGNQGDNKGG--QGQ-------   197
                                .     *     .  .*  .  .  .. * .  *   .

NONA mel          SGPRPGGGAGGAM-NSTNMGGGGGGG--GGGGPRG--GEDFFITQRLRSISGPTFELEPV   290
NONA vir          GGGGGGGGGGRDR-NPDRRGGGGGGGQNSGGGNNSQRGDDFFYSQRLRSISGPTHELPPI   273
NONA-like         ---RGAGGGKHQRGNRSRRSGGSGIMNSSMGGGGQ-RGEDFFIAQRLLDISGPTHELPPI   253
                        **      *     ** *      **     *.*** .***  ***** ** *.

NONA mel          EVPTETKFSGRNRLYVGNLTNDITDDELREMFKPYGEISEIFSNLDKNFTFLKVDYHPNR   350
NONA vir          EVAQETKFSGRNRLYVGNLTNDITDEELREMFKPYGEIGEIFSNLEKNFTFLKVDYHINA   333
NONA-like         ELPTDNKFVGRNRLYVGNLTSDTTDDDLREMFKPYGEIGDIFSNPEKNFTFLRLDYYQNA   313
                  *.  ..** *********** * **..*********** .**** .******..**  *

NONA mel          EKAKRALDGSMRKGRQLRVRFAPNATILRVSNLTPFVSNELLYKSFEIFGPIERASITVD   410
NONA vir          EKAKRPLDGSMRKGRHVRVRFAPNATILRVSNLTPFVSNELLYKSFEIFGPIERASITVD   393
NONA-like         EKAKRALDGSLRKGRVLRVRFAPNA-IVRVTNLNQFVSNELLHQSFEIFGPIERAVICVD   372
                  ***** ****.**** .******** *.**.**. ******* .*********** *.**

NONA mel          DRGKHMGEGIVEFAKKSSASACLRMCNEKCFFLTASLRPCLVDPMEVNDDTDGLPEKAFN   470
NONA vir          DRGKHLGEGTVEFAKKSSASACLRLCNEKCFFLTASLRPCLVEPMEVNDDNDGLPEKALN   453
NONA-like         DRGKHTGEGIVEFAKKSSASACLRLCNEKCFFLTASLRPCLVEPMEVNNDNDGLPDKTLN   432
                  ***** *** **************.*****************.***** *.****.*. *

NONA mel          KKMPDFNQERSIGPRFADPNSFEHEYGSRWKQLHNLFKTKQDALKRELKMEEDKLEAQME   530
NONA vir          KKLQEFNQERSVGPRFADLNSFEHEYGSRWKQLHDLFKSKQDALKRELKMEEEKLDAQME   513
NONA-like         KKSLEFRHERSVGPRFACLNSFEHEYGSRWKQLHDLFKSKQDSLKRELKMEEDKLEAQME   492
                  **  .* .***.*****  *************** ***.***.*********.**.****

                        ▼
NONA mel          YARYEQETELLRQELRKREVDNERKKLEWEMREKQAEEMRKREEETMRRHQTEMQSHMNR   590
NONA vir          YARYEQETELLRQELRKRESDNERKKLEWEMREKQAEEMRKREEETMRRHQTEMQSRMVR   573
NONA-like         YARYEQETELLRQELKKRELDNERMKLEWEMREKQAEEIRKREEEYMHR----YQNQLLR   548
                  ***************.*** **** *************.****** *.*     * ..  *

NONA mel          QEEDMLRRQQE-TLFMKAQQLNSLLDQQEGFGGG-GGGN----------NSTFDNFAGNS   638
NONA vir          QEEDMRRRQQENTLFMQAQQLNSLLDQQEGFGGGNGGGGGGGGGGGVGNSNFDNFGGNS   633
NONA-like         HEEDMRARQQENDLLMQAKKLNMILDQQEGFGGS---------------NSGFEHFD---   590
                  .****  ****  * *.*..** .*********                ** *..*

                          ▼                     ▼
NONA mel          NSPFEVFRGNNNNNSTMIGNNAAPNT----QDSFA-FEFGVNNMNQGGNQRGNNGGGNNV   693
NONA vir          NSPFEVFRGNNN--SSMAGNNAGPGANNQQQDSFAAFEFGVNNMNQGGNQRGNNGG-NNV   690
NONA-like         -SPFEVFGNNSN-----NSTMAGPGG--------P-----D---NSDGNQHGHIDS----   624
                   ****** .* *       . * *.                   *  ***.*.

NONA mel          PWGRRRF   700
NONA vir          PWGRRRF   697
NONA-like         -WGHRRF   630
                   **.***
```

FIGURE 2.—CLUSTAL alignment of NONA protein in *D. melanogaster* and *D. virilis* (EMBL database accession no. AJ298998) together with *D. melanogaster* NONA-like. The two adjacent 80-residue RRMs are underlined, and within these the RNP-1 octapeptides and RNP-2 hexapeptides are italicized. Inverted triangles show the position of the introns within the translation products of *D. melanogaster* and *D. virilis nonA*. The *nonA-like* gene is intronless.

protein kinases, whereas the corresponding sequence in *D. melanogaster*, KREVDNE (residues 547–553), has lost these potential modifications. All other putative post-translational modification sites in the RRM and charged regions are conserved between the two species NONA proteins (data not shown).

The 5′ regulatory region of *nonA*: Approximately 2.5 kb of upstream sequence from the *D. virilis nonA* fragment was initially obtained and compared with the upstream sequence of *D. melanogaster* (Jones and Rubin 1990). A dot matrix analysis (window, 21; stringency, 14) was performed for the *D. virilis* and the *D. melanogaster* sequence comparison to graphically highlight the regions of homology (Figure 3). Revealed are seven re-

TABLE 1

Drosophila protein identity

%	dGPI1	NONA	EN	HB	KNI	NOS	OSK	PER	RUNT	SEV	TRA
Similarity	67	83	88	89	83	78	74	78	87	78	66
Identity	59	75	83	81	75	64	59	68	81	65	50

Percentage similarity and identity between *D. melanogaster* and *D. virilis* homologous proteins was calculated using the method of Smith and Waterman (1981). The proteins compared between the two species include dGPI1, NONA, engrailed (EN), hunchback (HB), knirps (KNI), nanos (NOS), oskar (OSK), period (PER), runt, sevenless (SEV), and transformer (TRA).

gions of moderate-to-good conservation, which are interspersed with regions of complete divergence between the two sequences, while the terminal third is very diverged. As the vital gene *l(1)i19e* may be encoded within the 5′ region of *nonA* and could overlap with the N-terminal half of *nonA* coding sequences (Jones and Rubin 1990; Stanewsky *et al.* 1993), this pattern of divergence and conservation could reflect the intron-exon pattern, respectively, of *l(1)i19e*. Primers were generated based on putative conserved coding regions of both species genes, and 5′ and 3′ RACE performed. The amplified cDNA fragments were sequenced and Figure 4 shows a CLUSTAL alignment of the two putative *l(1)i19e* coding regions. The gene has an 8 exon-7 intron structure (Figure 1) and encodes a protein of 481 amino acids in *D. melanogaster* and 473 in *D. virilis*. The ends of the two transcripts, as detected by 3′ RACE, fall 424 and 336 bp before the initiating ATG codons of *nonA* in *D. melanogaster* and *D. virilis*, respectively. Overall identity between the two Drosophila proteins is 59% and similarity is 67% (Table 1). However, the N

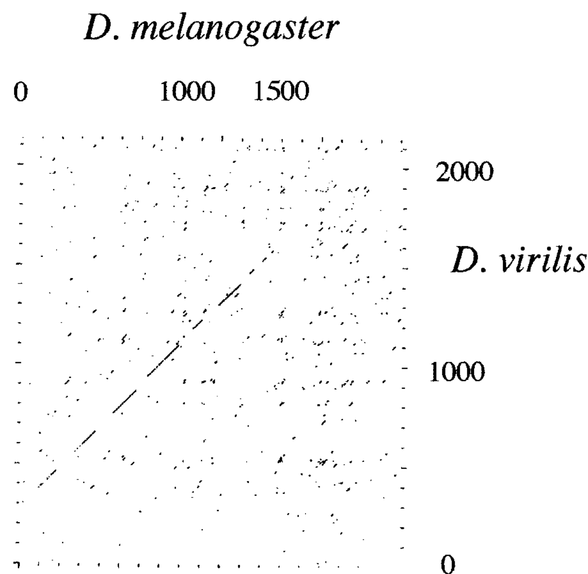

Figure 3.—Dot matrix alignment of 2.2 kb of *D. melanogaster* and *D. virilis* sequence 5′ to *nonA* transcription start. Note the seven to eight regions of identity that represent the intron-exon structure of *dGpi1*.

and C termini are much more diverged, with identities of 43% (residues 1–214) and 33% (residues 420–481), respectively, compared to the central region (residues 215–419), whose identity is 82%.

A BLAST search of the databases using both sequences revealed similarity with the *gpi1* genes of mammals, *Caenorhabditis elegans*, and yeasts. These encode a component necessary for the first step in the biosynthesis of glycosylphosphatidylinositol (GPI), which is used to anchor eukaryotic proteins to membranes. Figure 4 also shows the CLUSTAL alignments of these various GPI1 proteins. The similarity between fly and human GPI1 is 34%, whereas identity is 23%. This rises to corresponding figures of 37 and 28% when compared to *C. elegans* and falls slightly when compared to the two yeast species. The alignment shows very few conserved residues among all species, and so putative secondary structure was investigated to look for similarities between the Drosophila and other species proteins. Hydropathy analysis (Kyte and Doolittle 1982; Engelman *et al.* 1986) reveals that in spite of their low overall identity, there is extensive similarity between the *D. melanogaster* and *Saccharomyces cerevisiae* proteins in both the number and spatial patterning of hydrophobic regions that may represent transmembrane domains (Figure 5). Consequently, this analysis suggests that they may be homologous proteins, and we suggest naming this Drosophila gene *dGpi1*. Whether this corresponds to *l(1)i19e* will be discussed below.

Comparisons among several *D. virilis* and *D. melanogaster* homologous proteins revealed identities ranging from 50 to 83% (Table 1). dGPI1 has similarity and identity scores very similar to those of *transformer*, making it one of the most diverged genes known in Drosophila (O'Neil and Belote 1992).

Molecular evolution of *nonA* and *dGpi1*: Considerable divergence has been found in the first half of *nonA* and in the N- and C-terminal regions of the dGPI1 proteins. This could reflect a lack of functional importance and freedom from selective constraints or could serve adaptive, species-specific characteristics, particularly in the case of a "behavioral" gene such as *nonA* (*e.g.*, Wheeler *et al.* 1991). We therefore used the K_a/K_s test to examine the ratio of nonsynonymous (K_a) to synonymous (K_s) substitutions (Li and Graur 1991). A ratio greater than

```
D. melanogaster   MSIKIYLPVNYYYLNKSTN--------LYGQVQINEDNVVSYYVLEASDLEFSDKAINHE 52
D. virilis        MSIKIFLPTEYFYLKRACN--------LYGQLHISEDNVVVYYVVDAD-----NNKLDSD 47
H. sapiens        MVLKAFFPTCCVSADSGL---------LVGRWVPEQSSAVVLAVLHFPFIPIQVKQLLAQ 51
M. musculus       MVLKAFFPTCCASADSGL---------LVGRWVPGQNSAVILAVVHFPFIPIQVKELLAQ 51
C. elegans        ------------------------------------------------------------
S. pombe          MQFLSLEPLSLLLLKDS----------FINKSNPEYESMQHQQILLKK------LKLHFP 44
S. cerevisiae     MPNYIFWPYESLFENSAAQGPQVALAISFEKTHFVVLGVCEPQYLEEVSIRPPYSVVATK 60

D. melanogaster   KGNLRFLGSILN--------------EDFNADQNQFLKFN--MRLCFTYN----SSQANI 92
D. virilis        TNKMRFLGSILC--------------SDSYADSNYCRSHD--VVLCFTYS----NDAPNV 87
H. sapiens        VRQASQVGVAVLGTWCHCRQEPEESLGRFLESLGAVFPHEPWLRLCRERGGTFWSCEATH 111
M. musculus       VQKASQVQVTVLGTWCHRQQEPEESLGKFLEGLGAIFSHNPWLQLCREKGTRFWSCKATY 111
C. elegans        ------------------------------------------------------------
S. pombe          RRKENSWKRSLR--------------SGLIELLNQSFEVR-------------------- 70
S. cerevisiae     NNGAEGWNYKVADPCNVHFRIPKLKFMQFYSSDPISLIIPEKEVGLHSSVGETLNYSKLE 120

D. melanogaster   TLVYLGITPEYLKHIKVILYDKQMVRN------LFVSGE----------------SENRR 130
D. virilis        SLMFTGITPENLNQVKLILYDKQTVRS------LVIKDN----------------NTLAH 125
H. sapiens        RQAPTAPGAPGEDQVMLIFYDQRQVLLSQLHLPTVLPDRQAGATTASTGGLAAVFDTVAR 171
M. musculus       HQMSSTLDTPTEDQVMLIFYDQRKLLLSWLHPPPVLPDCQIGDSTASTGGLADIFDTVAR 171
C. elegans        ------------------------------------------------------------
S. pombe          -MLTHENNNK-KNSYVFRLFDRVSSST-------FYFFN----------------SLFAY 105
S. cerevisiae     QHPRYKRDNKKLSETLNIINLFPAYCKALNELYPFIQTSQENLRGTMLNSVAAWCSSTCI 180

D. melanogaster   SLEYNDNDSTDCDFLELSRLNQPTADN------QNKANNNANKSIQYGLTLIADSPIKIF 184
D. virilis        WTRENQIN-DECDFYMLAQLVQPNPDT------CRKSNDLWFYGSQS-LCLLANIPMQMF 177
H. sapiens        SEVLFRSDRFDEGPVRLSHWQSEGVEASILAELARRASGPICLLLASLLSLVSAVSACRV 231
M. musculus       SEVLFRNDQFDEGPVRLSHWQSEGVEASILVELAKRASGPVCLLLASLLSLISAASACRL 231
C. elegans        ------------------------------------------------------------
S. pombe          FIILLR---IINEVILLAINYRPIPLS-------YNMMD-IFVSARQVDLRLQQACFWPV 154
S. cerevisiae     YKMVAKIGFYLTFVICSIASLVSSLLN-----YSHFQLVNYSAFVQQIDLRCQQICYFPV 235

D. melanogaster   EYMAENVFINSIMVHTTIYKH-FKEWQTACDKRS--------RPAN----IVLDRILGII 231
D. virilis        QYIVGNKFINSIITHTVIYRH-YKEWQSIYTKGS--------RLSN----IMIDRALGII 224
H. sapiens        FKLWPLSFLGSKLSTCEQLRHRLEHLTLIFSTRKAENPAQLMRKANTVASVLLDVALGLM 291
M. musculus       WKLGPLAFIRSKLSTCEQLQHRLKHLSFIFSTEKAQSPMQLMRKANMLVSVLLDVALGLL 291
C. elegans        -----MDKITKNVVLIRTLEAKILLVRTFSFTRL----------------LLLDVAFSIY 39
S. pombe          QYMKLWVFRKSKRVAIEDYKEYIRFYNNLWLVAN-------------------DMIFGIT 195
S. cerevisiae     QYERINKKDNIQNVGSMVEKDNSNSQFSHSYMPSKFYP-DYILLYNTIWLIINDISFGLI 294
                                     .                                *  :.:

D. melanogaster   LMLILFS--LATQPGDFLIQ--ISHYVIDELYGLLKVLEGSPIGLKLNIHLNNFFLDCFK 287
D. virilis        LMLVLFT--LVSHPGDFLIQ--ISHIIIHQLYSLLKVLEGSPIGLKLNIHLNNFFLDCFK 280
H. sapiens        LLSWLHGRSRIGHLADALVP--VADHVAEELQHLLQWLMGAPAGLKMNRALDQVLGRFFL 349
M. musculus       LLSWLHSNNRIGQLANALVP--VADRVAEELQHLLQWLMGAPAGLKMNRALDQVLGRFFL 349
C. elegans        LWNIWTP--NWEWTVNEFWD--QTGNVADNLNGTITWLRSNPAGLKLNTPVNETLAWFFT 95
S. pombe          MSSFILEN---LHLVVKLIENITFEYAIKNVRSMVIWLVDTPAGLKLNNDICKFIMKLSV 252
S. cerevisiae     LGAILIE--NRDFLVSASHR-VLKFFLYDSLKTITETLANNPLGIKLNAELANFLSELFL 351
                  :                           ..:      * . * *:*:*  : : :

D. melanogaster   YHIEL-WSTFLD-FIEPLVR----------------QVFLAIGMIGCLGFTFQIALLVDL 329
D. virilis        YHIEL-WSTFLD-LIEPIVR----------------QVFLAIGAFGCLGFTYQIALLADL 322
H. sapiens        YHIHL-WISYIH-LMSPFVE----------------HILWHVGLSACLGLTVALSLLSDI 391
M. musculus       YHIHL-WISYIH-LMSPFIE----------------HILWHVGLSACLGLTVALSIFSDI 391
C. elegans        YHIYL-WTSMLF-CFQKFAFNFHNDFSAFIGFLRSDAFFRFIAYSLIGGISTFSAMVYDF 153
S. pombe          WVIDV-WSNFLL-HCLPWTP----------------FLVQVVAISGFGGASLMIALISDF 294
S. cerevisiae     WVIEFSYTTFIKRLIDPKTLS--------------SLLTLTIYMMFLVGFSFAVSLAIDF 397
                  : * . : . :                        .   :     * :   :: *:

D. melanogaster   ISVIGLHSHCFYIYTKVLYNVERRGLSVLWQVVRGNRYNILKGRTESHNYMNRQLYLATI 389
D. virilis        ISIVGLHAHCFYVYTKVLNNVEVKGLTVLWQVVRGNRYNILRNRIEAHNYMNRQLYLATI 382
H. sapiens        IALLTFHIYCFYVYGARLYCLKIHGLSSLWRLFRGKKWNVLRQRVDSCSYDLDQLFIGTL 451
M. musculus       IALLTFHIYCFYVYGARLYCLKIYGLSSLWRLFRGKKWNVLRQRVDSCSYDLDQLFIGTL 451
C. elegans        SQIFFLHFNCFDAYATKLCYLCYYTLTVLWSLVRGKKWNPLRERKDTVILDTRQQFLATS 213
S. pombe          LSVMTIHIHLLYLASSRMYNWQLRVIYSLLQLFRGKKRNVLRNRIDSYEYDLDQLLLGTI 354
S. cerevisiae     FAILSFPIYVFYRISSKLYHCQLNIMASLFNLFCGKKRNVLRNRIDHNYFQLDQLLLGTL 457
                    :. :    :     :       :  *  :. *:: * *: * :       *  :.*

D. melanogaster   FFSAILFLLPTTLVYYIVFAALKALTFATLSVFHFVRRKLMYLPIEVCIKRLLRGCHEID 449
D. virilis        FFSAILFLFPTTLVYYVVFATLKALTCATLATLECFRRKLLNFPIEMFLKYIKKGFYEID 442
H. sapiens        LFTILLFLLPTTALYYLVFTLLRLLVVAVQGLIHLLVDLINSLPLYSLGLRLCRPYRLAA 511
M. musculus       LFTILVFLLPTTALYYLVFTLLRLLVITVQGLIHLLVDLINSLPLYSLGLRLCRPYRLAA 511
C. elegans        LFVILLFILPTIFVYFVVFRCLRLAVSALQTVLYFFATWPFQLFALEKHLAEKYGKPADA 273
S. pombe          LFTVLIFFLPTIYVFYAAFALTRVSVMTCLAICETMLAFLNHFPLFVTMLRIKDPYRIPS 414
S. cerevisiae     LFIILVFLTPTVMAFYMSYTVLRMLTITIEIFSEAVIALINHFPLFALLLRLKDPKRLPG 517
                  :*  ::*: **   ::  :   : . :      .     :

D. melanogaster   --CIQIKDVSHHER-----AFLMHKQQKI----NVTVYKITTL------------------ 481
D. virilis        --CLRVLDIPLQKP-----LFFIHRNSKL----IIFVYKLRV------------------- 473
H. sapiens        GVKFRVLRHEASRP-----LRLLMQINPLPYSRVVHTYRLPSCGCHPKHSWGALCRKLFL 566
M. musculus       GVKFRVLEKEAGRP-----LRLLMQINPLPYSHVVHTYRLPSCGCHPKHSWGTLCRKLFF 566
C. elegans        QNEALAEKKTKSQN---------------------------------------------- 287
S. pombe          GLNFEIVSFEPLKQDGFATLYLNCNSKPMSLGSMFEHYRKLARRLISHYLSKTTLISLLV 474
S. cerevisiae     GISIELKTTVSNKH-----TTLELQNNPIKFKSMFRPYNLLLSQMRTNYFSFATVRKIVR 572
                              .

D. melanogaster   -------------------------------------------
D. virilis        -------------------------------------------
H. sapiens        GELIYPWRQRGDKQD---------------------------- 581
M. musculus       GELIYPWRQREDKQD---------------------------- 581
C. elegans        -------------------------------------------
S. pombe          GCPVPAIPAEQLYNIQYAMLPTKRISIRKLRDLLFHQKKFPYD 517
S. cerevisiae     GESIMVNRNKLYYVLYSS-LPSKPLSVKDLYKRLTIQA----- 609
```

FIGURE 4.—CLUSTAL alignment of GPI1 proteins in a variety of species. The corresponding intron positions within the two Drosophila GPI1 proteins are shown as inverted triangles. EMBL accession numbers for *D. melanogaster* and *D. virilis Gpi1* are AJ298995 and AJ298996, respectively, and for the *nonA* promoter region for *D. virilis*, AJ298997.

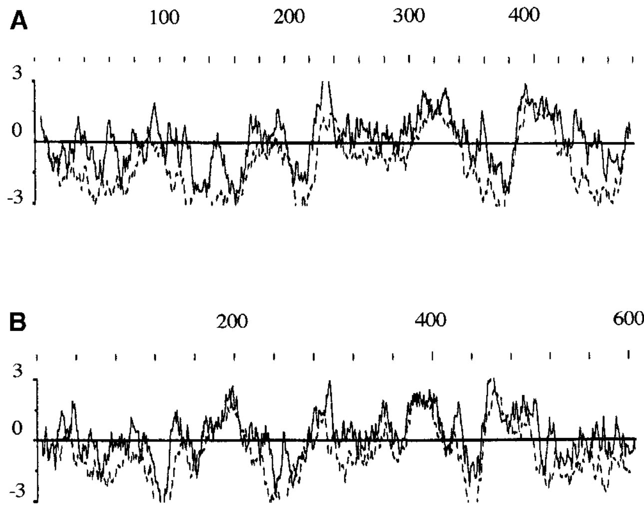

FIGURE 5.—Hydropathy plots based on KYTE and DOOLITTLE (1982, solid) and ENGELMAN *et al.* (1986, dashed) methods for GPI1 in (A) *D. melanogaster* and (B) *S. cerevisiae*. Hydrophobic areas are shown from 0 to +3 and hydrophilic from 0 to −3. The residue numbers for each protein are given on the horizontal axis.

unity indicates the action of selection in fixing nonsynonymous variation at a rate faster than the neutral mutation rate. For comparison, the test was also performed on a number of other *D. virilis* and *D. melanogaster* homologues (Table 2). In all these genes the rate of nonsynonymous substitutions is occurring at a much lower rate than the synonymous level. Even the high K_a/K_s ratio for *tra* is still far from unity. The striking feature of these results is the astonishingly high K_s value observed for *dGpi1* (3.377 sem 0.236), almost three times that of *nonA* and twice that of *tra*. This is not an artefact of poor alignment because when the *dGpi1* gene is divided into the conserved central region *vs.* the nonconserved N- and C-terminal regions together, K_s values >3 are obtained for both portions. The most diverged part of *nonA* (from nucleotide 1 to 813) gives a K_s value of 2.155, much lower than that of *dGpi1*.

One possible way to explain the high synonymous rate for *dGpi1* would be to invoke low levels of codon bias. With no selection for specific codons, the third position would be relatively free of constraints, providing an avenue for inflation of K_s values. The Relative Synonymous Codon Usage (RSCU) index was calculated for a number of different genes in *D. melanogaster* and *D. virilis* (SHARP *et al.* 1988 and see legend of Table 3). It can be seen that *dGpi1* has the lowest overall codon bias levels in both species, and this may contribute toward its high K_s levels. In addition, the substitution rates for all seven small introns in *dGpi1*, which range from 51 to 70 bp in length in both *D. melanogaster* and *D. virilis* (see Figure 1), and for the large first intron of *nonA* (1298 bp in *D. melanogaster*, 1444 in *D. virilis)*, were also calculated. This was done after first removing the canonical donor and acceptor dinucleotides, the short conserved pyrimidine tracts close to the 3′ end, and the single conserved adenine branch point from each intron. Intron nucleotide identity was 37% for *dGpi1* and 41% for *nonA*, revealing no obvious increase in the mutation rate of *dGpi1* that could explain its extraordinary K_s values.

The proximity of the two genes raises the issue of whether *dGpi* sequences act as promoter and enhancer regions for *nonA* expression. This has been studied in the accompanying article by SANDRELLI *et al.* (2001) using various deleted fragments of the *nonA* upstream regions. It is clear from these results that enhancers and silencers of *nonA* expression must overlap with *dGpi1* sequences. We therefore examined potential transcription factor binding sites in the ∼2.3-kb region immediately upstream of the transcription start of *nonA* in both species (see MATERIALS AND METHODS). Putative binding sites might suggest which *trans*-acting factors could be involved in *nonA* regulation. The most significant sites (scores ≥90) include those for Broad-Complex (BR-C), situated ∼350 and 300 bp upstream of the *melanogaster* and *virilis nonA* transcription starts, respectively (see Figures 1 and 6), and those for *Deformed* and heat-shock factors, which were found within the *dGpi1* sequences (Figure 6). Reducing the stringency of the match between the binding site consensus and the target sequences (scores ≥85) revealed two more BR-C sites in the intergenic region of both species.

The presence of *dGpi1* could thus constrain the evolution of the *nonA* regulatory region. To explore this further, we adopted a neutrality test (MCDONALD and

TABLE 2

Synonymous and nonsynonymous substitution rates

	nonA	*en*	*per*	*tim*	*tra*	*dGpi1*	*nos*
K_s	1.242	1.035	1.234	1.264	1.515	3.377	2.113
K_a	0.190	0.130	0.243	0.102	0.493	0.327	0.266
K_a/K_s	0.153	0.126	0.197	0.081	0.325	0.097	0.126

Synonymous (K_s) and nonsynonymous (K_a) substitution rates between *D. melanogaster* and *D. virilis* were calculated for *nonA*, *engrailed* (*en*), *period* (*per*), *timeless* (*tim*), *transformer* (*tra*), *dGpi1*, and *nanos* (*nos*) using the method of LI and GRAUER (1991).

TABLE 3

Codon bias in Drosophila

RSCU	*nonA*	*en*	*hb*	*kni*	*nos*	*osk*	*per*	*runt*	*sev*	*Gpi1*	*tra*
>1.5	8	7	6	5	3	3	7	3	5	3	7
<2	(4)	(7)	(9)	(6)	(9)	(6)	(12)	(6)	(5)	(4)	(4)
>2	3	6	6	4	2	4	4	8	6	1	6
	(2)	(8)	(5)	(7)	(3)	(4)	(4)	(8)	(3)	(2)	(2)

The number of codons within each gene that has an RSCU index >1.5, <2, and >2 is shown for *D. virilis* and *D. melanogaster* (parentheses). RSCU is defined as the observed number of codons divided by the expected number if all codons were used equally (Sharp *et al.* 1988). Note the extremely low overall codon bias of *dGpi1*.

Kreitman 1991) that has been used in an attempt to identify adaptive changes in putative regulatory sequences (Jenkins *et al.* 1995). Five sequences were obtained from a natural population collected in Lecce, Italy, of *D. melanogaster*, three from a natural population of *D. simulans* from Zimbabwe, plus the reference sequence from *D. melanogaster* (Jones and Rubin 1990). These sequences corresponded to the intergenic region between *dGpi1* and *nonA*, plus ∼400 bp of the 3′ half of *dGpi1* (representing exons 6, 7, and 8; 823 bp for *D. melanogaster* and 880 bp for *D. simulans*). The TF SEARCH program was used on the reference *D. melanogaster* sequence as before using a stringency of 0.9, and each nucleotide was classified as to whether it lay within or outside of a putative binding site. A total of 85 differences (substitutions, insertions, and deletions) were found among the sequences studied. Of these, 24 were fixed changes between the two species and 61 were polymorphisms. We divided the region into two: the *dGpi1* sequence up to its stop codon and the sequence downstream to the transcription start of *nonA* (Figure 6). Table 4 shows that the *dGpi1* sequences reveal no significant association between the type of change (fixed *vs.* polymorphic) and whether the sequence represents a putative binding site ($P = 1.00$). However, in the intergenic region, a significant increase is observed in the number of fixed changes within putative binding sites relative to polymorphisms ($P = 0.0398$). These results suggest the possibility of adaptive changes in the putative binding sites within the intergenic region, whereas *dGpi1* imposes additional constraints on the fixation of such changes, even in the face of high K_s values.

Rescue of *l(1)119e*⁻ and *nonA*⁻ mutant phenotypes: The extremely high levels of divergence in the *dGpi1* coding sequence, which probably corresponds to *l(1)i19e*, generates considerable doubt as to whether the 12.5-kb *D. virilis* fragment we have studied will rescue the lethality associated with *T(1;4)9e2-10*, in which both *nonA* and *l(1)i19e* are deleted. We therefore studied the viability of male progeny carrying the *virilis* transgene on a *T(1;4)9e2-10* background by crossing *T(1;4)9e2-10/In(1)FM7* females with males carrying a single balanced autosomal transgenic copy of the *D. virilis* fragment (*insert/balancer*). The viability results are shown in Table 5 and reveal considerable heterogeneity between the various lines studied ($\chi^2 = 31.05$, d.f. = 6, $P < 0.01$), yet it is clear that in all but line 191, at least one-third of the male progeny from all crosses produce *T(1;4)9e2-10/Y; insert/+* individuals. Even in line 191 there was no significant departure from the 1:2 ratio of *T(1;4)9e2-10/Y* to *In(1)FM7* males ($\chi^2 = 2.43$, d.f. = 1), further confirming that the *virilis* fragment rescues both the mutant *l(1)i19e* and *nonA* viabilities to normal levels.

Finally, we examined the ERG, a sensitive measure of *nonA* function (Rendahl *et al.* 1996; Stanewsky *et al.* 1996), of males carrying the *T(1;4)9e2-10/Y; insert/+* genotype from lines 135, 191, and 297.6 and compared them with a Canton-S wild type. Table 6 shows that all flies demonstrated both ON and OFF transients, and ANOVA revealed no significant differences between any of the genotypes in the amplitudes of either response (ON, $F = 1.39$, d.f. = 1, 61; OFF, $F = 0.90$, d.f. = 1, 69).

DISCUSSION

D. virilis shows an elevation in the amount of repetitive DNA in both the coding and upstream regulatory regions of *nonA* compared to *D. melanogaster*, mirroring similar observations that were made in comparisons between these two species involving the *hunchback* gene (Hancock *et al.* 1999). Comparative analyses of homologous genes reveal that areas of high divergence are often associated with regions of repetitive DNA, both in coding (Treier *et al.* 1989; Peixoto *et al.* 1993) and noncoding regions (Tautz *et al.* 1987; Hancock and Dover 1988). Not surprisingly, therefore, much of the divergence between *D. melanogaster* and *D. virilis* in the N-terminal regions of *nonA* involves repetitive motifs such as QN, GNQGGX, and poly(G).

Poly(G) motifs [another long poly(G) stretch is found in the C terminus of *D. virilis* NONA] are of particular interest as several known RNA-binding proteins, such as the hnRNP proteins A1 and A2 and the nucleolar pre-rRNA-binding protein, Nuclein, have auxiliary domains constituted by glycine-rich regions (Bandziulis *et al.*

1989). These auxiliary domains may be involved in protein-protein interactions, but have also been shown to be involved in the polynucleotide binding properties of RNA-binding domains (Bandziulis *et al.* 1989). Thus the long N-terminal stretch of 29 Gly residues in *D. virilis* compared to *D. melanogaster nonA* may have functional relevance. Interestingly, two-dimensional plots of NONA from the two species (using the Peptidestructure and Plotstructure programs from the GCG package) revealed a marked difference in the N-terminal regions (data not shown). In *D. virilis* NONA, a large uninterrupted domain of turns is predicted from the poly(Gly) tract and is preceded by a long α-helical conformation produced by the QA repeats. The corresponding region in *D. melanogaster* NONA has no helical conformation and very short, frequently interrupted areas of turns (data not shown).

The areas of high conservation between the two *nonA* sequences correspond to the RRMs. Mutational studies have revealed that the first RRM domain (RRM1) in *nonA* is necessary for all the known functions of NONA (Rendahl *et al.* 1996; Stanewsky *et al.* 1996). Mutations in this region not only cause severe defects in both visual and song phenotypes, but also invariably reduce the viability of the affected flies. On the contrary, mutations in the RRM2 domain either have little or no effect or produce impairments of the visual system only (Rendahl *et al.* 1996; Stanewsky *et al.* 1996). In this regard, we note that the RNP1 region of RRM2 has a lower level of conservation than RRM1 (see Figure 2). The NONA protein can therefore be roughly divided into two, the N-terminal diverged fragment and the central and C-terminal conserved regions. However, within the C terminus lies the charged region in which is located the site of the $nonA^{diss}$ song mutation (Rendahl *et al.* 1996). It was therefore of interest that a single substitution between the two species in the region including this mutant site generated additional potential post-translational modifications in *D. virilis*. If *nonA* does act as a reservoir for species-specific song information (S. Campesan, Y. Dubrova, J. C. Hall and C. P. Kyriacou, unpublished results), then perhaps this difference in sequence might be relevant, because it lies in an area of the NONA protein that has some influence on the song phenotype (Rendahl *et al.* 1996).

Comparison of the *nonA* genes with *nonA-like* revealed lower identity scores between the *nonA* and *nonA-like* proteins than between the *nonA* orthologues. The *nonA-like* gene is found within the *bithorax* complex of chromosome 3 and is unusual because it encodes a single open reading frame (Martin *et al.* 1995). This suggests that the duplication event giving rise to *nonA-like* was mediated by an RNA intermediate followed by transposition. The absence of stop codons suggests that *nonA-like* may be functional, although the available sequence is genomic only. Consequently, *nonA-like* may represent a unique example of a processed duplicated behavioral gene. On the basis of the identity scores, we can assume that the duplication event occurred before the *D. melanogaster-D. virilis* split.

Comparison of the 5′ region of *nonA* revealed the presence of *dGpi1*, which may correspond to *l(1)i19e*. The protein sequence has a low level of identity with the product of the *gpi1* gene of yeast (Leidich and Orlean 1996), which is used in GPI synthesis to anchor proteins in the endoplasmic reticulum. The biosynthesis of GPI requires sequential additions of sugar molecules to phosphatidylinositol (PI) in a number of steps (Englund 1993; McConville and Ferguson 1993). The first stage requires the synthesis of N-acetylglucosaminyl phosphatidylinositol (GlcNAc-PI) from UDP-GlcNAc and PI, catalyzed by GPI-GlcNAc transferase (GPI-GnT, Doering *et al.* 1989), and involves more than one gene. In yeast, one of these genes is *GPI1*; it encodes a 609-residue polypeptide with several hydrophobic membrane-spanning domains (Leidich and Orlean 1996). The amino acid identity between GPI1 and the Drosophila protein is low, but the hydropathy profile is very similar to that of *S. cerevisiae*. We have therefore taken the liberty of naming this fly gene *dGpi1*.

The *dGpi1* gene almost certainly corresponds to *l(1)i19e* for a number of reasons. First, it lies in a region of overlap between *nonA* and *l(1)i19e*, as predicted (Jones and Rubin 1990). Second, a *nonA* fragment with the same 5′ end as that of 270XS16 (beginning at amino acid residue 148 of dGPI1; see Figure 1), and encompassing the whole promoter region up to the *nonA* start codon, does not rescue the inviability of transformants carrying the *l(1)i19e* mutation (Sandrelli *et al.* 2001, accompanying article). In contrast, the original 270XS16 fragment that carries the 5′ and coding regions of *nonA* has been reported to rescue this lethality to some degree, suggesting that *l(1)i19e* sequences may "straddle" the 5′ end of 270XS16 (Jones and Rubin 1990). The N terminus of GPI1 proteins is so diverged (see Figure 4) that it could be that, in some transformant lines, flanking regions at the 5′ end of the insert may promote the transcription and translation of either a truncated dGPI1 or a fusion dGPI1 product that may rescue viability. It is certain that, because these rescuing (or nonrescuing) fragments begin in the first exon of *dGpi1* (Figure 1), the sequences corresponding to *l(1)i19e* cannot be encoded downstream of those for *dGpi1*. The only caveat (on the basis of our results only) is that the *D. virilis* fragment that rescues *T(1;4)9e2-10* still has ∼1 kb of unsequenced material upstream of the transcriptional start of *dGpi1*, so a gene could be encoded immediately 5′ of *dGpi1*. However, if one accepts the arguments outlined above concerning the sporadic rescue of *l(1)i19e* with 270XS16 (Jones and Rubin 1990), then *dGpi1* is *l(1)i19e*.

The divergence of the *dGpi1* gene between *D. melanogaster* and *D. virilis* is almost as high as that of *transformer* (O'Neil and Belote 1992). The central region, which

```
Drosophila melanogaster Lecce- 1     ........ .......... .......... .......... ..........
Drosophila melanogaster Lecce- 2     ........ .......... .......... .......... ..........
Drosophila melanogaster Lecce- 3     ........ .......... .......... .......... ..........
Drosophila melanogaster Lecce- 9     ........ .......... ....A..... .......... ..........
Drosophila melanogaster Lecce-13     ........ .......... .......... .......... ..........
Drosophila melanogaster ref. seq.    CAACTATA TGAATCGGCA ATTGTATCTA GCTACCATAT TTTTCTCGGC 1040
Drosophila simulans Zimbabwe-21      ........ .......... ...AA..... .......... ..........   48
Drosophila simulans Zimbabwe-72      ........ .......... C..A...... .......... ..........
Drosophila simulans Zimbabwe-19      ........ .......... C..A...... .......... ..........

.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
TATTCTGTTT TTGTTACCCA CTACCCTTGT CTACTATATA GTTTTTGCTG CTGTAAGTTT AATATATAAT CGTTGAATTT 1120
.......... .......... .C........ .......... .......... .........A ......A... ..........  128
.......... C......... .C........ .......... .......... .......... ......A... .....T....
.......... C......... .C........ .......... .......... .......... ......A... ..........

.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
....A..... .......... .......... .......... .......... .......... .......T.. ..........
.......... .......... .......... .......... .......... .......... .......... ..........
CGCAGTAATA CAACTTTCTT TATAGCTGAA GGCGCTCACA TTTGCTACTC TAAGTGTCTT TCATTTTGTG CGAAGAAAAC 1200
.AA.A..... .......... .......A.. .......... .......... .......... .G.....T.. ..........  208
.A..A..... ......A... .......... ......T... .....C.... .......... .G.....T.. ..........
.A..A..... .......... .......... .......... .......... .......... .G.....T.. ..........

.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
TGATGTATCT TCCGATTGAA GTCTGTATAA AACGGTTATT AAGAGGATGT CATGAAATAG GTAATTGTGT AAAGAATGTG 1280
.......... .......... .......... .G........ G......... .......C.. ........A. G.........  288
.......... G......... .......... .G........ ...G..G... ..C....... ........A. G..A......
.......... G......... .......... .G........ ......G... .......... ........AA ...A......

.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
......C... ..G....... .......... .......... ..C....... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
AATGCATAAA ATTATATTTA TATAATGATT TTTCTATTGA CTTGCAGATT GTATTCAAAT CAAAGATGTT TCACATCACG 1360
G......... .......... ......T... .....T.... ...T...... .......... .......... ..........  368
G......... C......... .......... .......... .......... .C........ .......... ..........
G......... C......... .......... .......... T......... .T........ .......... ..........

.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... ...*...... ..........
AAAGGGCTTT CCTGATGCAC AAACAACAGA AAATCAACGT GACAGTGTAT AAAATTACAA CTTTATAATC CTTGCCATAT 1440
.T........ .......... .....G.... .......... .......... ..T.....G. .......... ..........  448
.......... .......... .....G.... .......... .......... .......... .......... ....G.....
.......... .......... .....G.... .......... .......... .......G.. .......... ..........

.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... ..A....... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
.......... .......... .........C .......... .......... .......... .......... ..........
.......... .......... .......... .......... .......... .......... .......... ..........
ATGTAGAATA TATTTAGAAG TAAACATAAG ATGTTTTAAA AGTACATTTA CATTTTATTT ATTGAAACTT CATTTGTTCA 1520
.--....... .......... .......... ......C... .......... .......... .C..T..... ..........  526
.--....... .......... .......... ......C... .......... .......... ....T..... ..........
.--....... .......... .......G.. ......C... .......... .......... ....T..... ..........

.......... .......... .......... .......--- ---------- ---------- ---------- ....------
.......... .......... .......... .......--- ---------- ---------- ---------- ....------
.......... .......... .......... .......--- ---------- ---------- ---------- ....------
.......... .......... .......... .......--- ---------- ---------- ---------- ....------
.......... .......... .......... .......--- ---------- ---------- ---------- ....------
TTTGATG-TT TTT--ATTAC TATCAATACA AAACAAA--- ---------- ---------- ---------- TCGA------ 1558
.......C.. ...--.A..T ...T...... .......ATA GAAAATAATA AAAATTGTGG TTTTTGTGTT ....TAAATT  606
.......... ...TT....T ..CT...... .......ATA GAAAATAATA AAAATTGTGG TTTTTGTGTT ....TAAATT
.......G.. ..CTT...G. ...T...... .......ATA GAAAATAATA AAAATTGTGG TTTTTGTGTT ....TAAATT
```

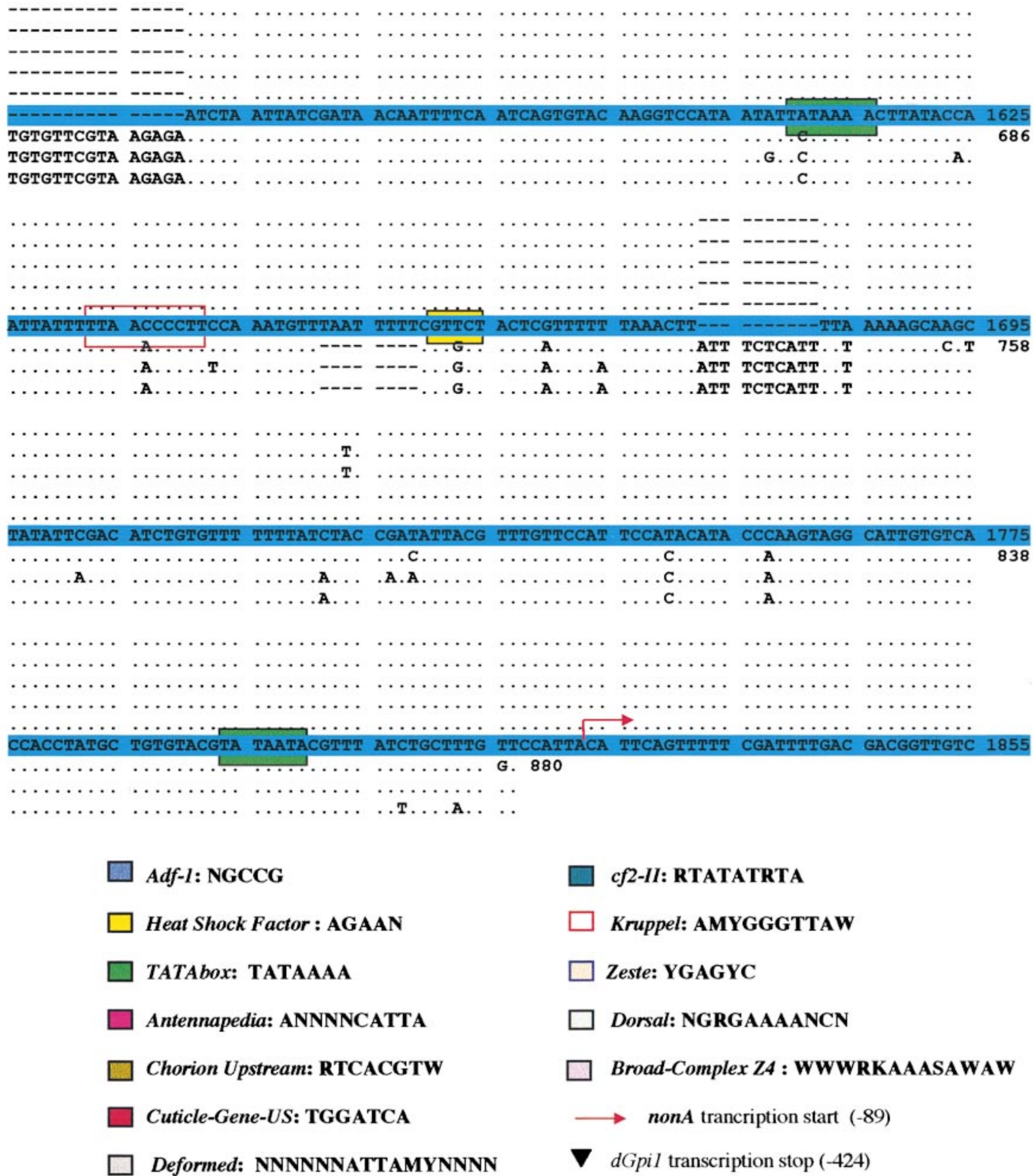

Figure 6.—Alignment of sequences for five *D. melanogaster* (863 bp) and three *D. simulans* (880 bp) natural haplotypes 5′ of the *nonA* transcriptional start. The reference sequence is that obtained by Jones and Rubin (1990) and is reported from nucleotide 992 to 1855 (GenBank accession no. M33496). This corresponds to nucleotides −917 to −54 in relation to the initiating methionine codon of *nonA* (see Figure 1). A dot represents bases that are identical to the Jones and Rubin (1990) sequence; a dash represents single base deletion; *dGpi1* introns are shown in italics. Putative binding sites (or their complementary sequences) are boxed. The long turquoise box represents sequences downstream of the *dGpi1p* stop codon, which is asterisked. The *Kpn4GAL4* construct from Sandrelli *et al.* (2001, accompanying article) includes nucleotides 992–1866 (see text). Nucleotides 992–1236 carry regulatory sequences for *nonA* expression that include silencers as well as enhancers, whereas nucleotides 1236–1686 carry enhancers only (Sandrelli *et al.* 2001, accompanying article). The transcription termination of *dGpi1* in relation to the initiating codon of *nonA* (−424 bp) is shown as an inverted arrowhead. The red arrow indicates the *nonA* transcription start found by 5′ RACE (−89 bp) and is 11 bp upstream of the one proposed by Stanewsky *et al.* (1993). Binding sites are represented with A, C, G, or T plus the IUBS code (K, G or T; M, A or C; N, any base; R, A or G; W, A or T; Y, C or T). Eighty-five differences are present among the analyzed sequences, either within *D. melanogaster* or *D. simulans.* Of these, 78 are single nucleotide substitutions and 7 involve insertions or deletions. In *D. simulans,* five insertions relative to the *D. melanogaster* sequence are present: a single nucleotide in position 1528 of Jones and Rubin's sequences, two nucleotides in position 1533, 33 nucleotides in position 1533, 21 nucleotides in position 1558, and 13 nucleotides in position 1681. In addition, in *D. simulans* there are two deletions involving two nucleotides in position 1442–3 and eight nucleotides in position 1651–9. The EMBL database accession nos. are AJ296020 for *D. melanogaster* and AJ296021 for *D. simulans.*

TABLE 4

Neutrality test on *nonA* promoter sequences

	Fixed changes	Polymorphisms
A. *dGpi1*		
Binding region	3	11
Nonbinding region	6	22
B. Intergenic region		
Binding region	5	2
Nonbinding region	10	26

Association between number of fixed and polymorphic nucleotide differences in putative binding and nonbinding sites in the *nonA* regulatory region in five wild-type haplotypes of *D. melanogaster* (823 bp) and three wild-type haplotypes of *D. simulans* (880 bp). (A) Within *dGpi1* transcription unit (445 bp in *D. melanogaster* and *D. simulans*). The interaction is not significant in a two-tailed Fisher's exact test ($P = 1.00$). (B) Within intergenic region between *dGpi1* transcription unit and start of transcription of *nonA* (*D. melanogaster*, 388 bp; *D. simulans*, 435 bp). The interaction is significant in a two-tailed Fisher's exact test ($P = 0.0398$).

may be membrane spanning, is more conserved than the N and C termini. There is no evidence for positive selection to account for this divergence, as measured by the K_a/K_s test, but the K_s value of >3.3 is extremely high and significantly higher than all other genes that we compared between the two species. Neither is there an elevated mutation rate in *dGpi1* as compared to *nonA* as measured by their respective intron divergence. However, in the estimated 40–60 million years since the divergence between *D. virilis* and *D. melanogaster*, the molecular clock may have saturated the synonymous position with mutation and elevated K_s (Schlotterer *et al.* 1994). If so, why do the other genes (except *nos*) have significantly lower K_s values?

TABLE 5

Viability of transformants

Transgenic line	A *T(1;4)9e2-10/Y;* ∇/+	B *FM7/Y;* ∇/+ *FM7/Y;* +/+	C % A
113	92	102	47.4
67-4	64	64	50.0
168-8	85	101	45.7
135	109	122	47.2
97	78	64	54.9
191	41	109	27.3
297-6	54	94	36.5

The transformant line number is shown in the left-hand column. The genotypes of F_1 males from the cross *T(1;4)9e2-10/In(1)FM7* females × +/*Y;* ∇/+ males (insert indicated with ∇/+ denotes second or third chromosome balancer) carrying an autosomal copy of the *virilis* transgene are shown in columns A and B. (A) Number of viable males carrying *T(1;4)9e2-10* and the *virilis* insert. (B) Number of males with *FM7* and the insert or males with *In(1)FM7* only. (C) Percentage of A males.

TABLE 6

Electroretinogram transient amplitudes

Genotype	*N*	ON	OFF
Canton-S	4	3.058 ± 1.035 (19)	2.663 ± 1.258 (22)
135	4	3.355 ± 0.725 (18)	2.282 ± 0.633 (24)
191	5	2.762 ± 0.886 (14)	2.552 ± 1.356 (14)
297-6	4	3.002 ± 0.964 (12)	2.900 ± 1.202 (11)

Mean ± SD of amplitudes (mV) of ON and OFF ERG transients in Canton-S and transformants. *N* is number of flies examined; in parentheses is the number of ON and OFF transients measured.

The *dGpi1* codon usage (and, to a lesser extent, that of *nos*) is very unusual compared to the other genes in that it shows extremely low codon bias in both species of Drosophila. This is in spite of the fact that it is a relatively small gene of <500 residues, and smaller coding sequences tend to have high levels of bias (Powell and Moriyama 1997). Thus the low codon bias is expected to contribute to the higher K_s value of *dGpi1*, as there would be little or no constraint on the synonymous position. K_s values are positively correlated with K_a values in a number of organisms, including Drosophila, and so the high levels of divergence in *dGpi1* are consistent with the high K_s value, although this argument cannot be applied to *tra* (Comeron and Kreitman 1998). Finally, codon bias is positively correlated with expression levels, suggesting that *dGpi1* may be expressed at low levels (Sharp *et al.* 1988). This could be at odds with its function of anchoring various cell-surface proteins to the phospholipid bilayer, which requires the dGPI1 protein to be expressed in many, if not all, cell types (Kinoshita *et al.* 1995).

Sequences within the *dGpi1* transcription unit, perhaps the coding regions themselves, contain both positive and negative elements that regulate *nonA* expression (Sandrelli *et al.* 2001, accompanying article). Although overlapping genes are common, particularly in prokaryotes, examples in the literature for eukaryotic coding sequences acting as regulatory regions for neighboring genes are rare (discussed in Sandrelli *et al.* 2001, accompanying article). These unusual observations might shed some light on the high *dGpi1* K_s value, but in addition they imply that the coding regions of *dGpi1* may constrain the evolution of the *nonA* promoter. To test this we first predicted conserved binding sites in *dGpi1* for *trans*-acting factors in *D. melanogaster* and *D. virilis*. The most stringent application of the algorithms revealed a BR-C binding site at the same position in *D. melanogaster* and *D. virilis*, immediately after the end of the *dGpi1* transcription unit. Broad-Complex is a member of the BTB or POZ domain family of zinc-finger proteins and its role is to transmit the

ecdysone signal to downstream genes during metamorphosis (DiBello *et al.* 1991). The amino-terminal BTB motif is embedded within the BR-C core and is alternatively spliced to give four isoforms, Z1 to Z4, each carrying different pairs of zinc fingers and each having slightly different, but overlapping, functions (Bayer *et al.* 1997; Sandstrom *et al.* 1997). One of these functions is to regulate the proper attachment of the thoracic musculature, and the Z1 and Z4 isoforms are able to rescue the muscle defects in BR-C mutants (Sandstrom *et al.* 1997). Note that in Figure 6 it is the Z4 BR-C binding site that is conserved (also in *D. simulans*). The relationship between thoracic musculature and the courtship song provides a possible rationale for the presence of the BR-C site on a "song gene" such as *nonA* (Ewing 1977).

We applied a modified version of the McDonald-Kreitman test to inspect variation within and outside these putative binding sites in a number of *D. melanogaster* and *D. simulans* sequences (Jenkins *et al.* 1995). These sequences correspond almost exactly to those that are used to drive expression of GAL4 in the *Kpn4GAL4* construct used by Sandrelli *et al.* (2001, accompanying article; Figure 6). The distal fragment (from −917 to −673 bp upstream of the initiating *nonA* methionine codon or nucleotides 992 to 1236 in Figure 6) contains *dGpi1* coding sequences and carries silencers for *nonA* expression in the thoracic muscles and enhancers for the visual optomotor response (Sandrelli *et al.* 2001, accompanying article). The proximal fragment from −673 to −223 (nucleotides 1236–1686 in Figure 6) encodes the C-terminal sequences of *dGpi1*, plus the intergenic spacer between it and *nonA*, and carries enhancers for all tissue-specific *nonA* expression and the optomotor response (Sandrelli *et al.* 2001, accompanying article).

In fulfilling these roles in *nonA* regulation, the transcription unit of *dGpi1* might be under different selective constraints compared to the intergenic spacer. Indeed, a significant excess of fixed changes relative to polymorphisms in the binding regions of the intergenic, as opposed to *dGpi1*, sequences was observed, so the presence of *dGpi1* is placing constraints on the fixation of adaptive changes in the putative binding sites. The high levels of polymorphism in the binding regions within *dGpi1* are consistent with the low codon bias in this gene and would serve as a barrier against adaptive fixation. We realize that this analysis is speculative and will rely on future work to show that these binding sites are biologically relevant. Nevertheless, if the algorithms we used were identifying completely nonfunctional sites, it is difficult to understand why a significant result would be obtained at all with this neutrality test, let alone in the intergenic region only.

Finally, in spite of extensive divergence, the *dGpi1* sequence within the *D. virilis* fragment is nevertheless able to rescue the lethality associated with *T(1;4)9e2-10*. The *nonA* sequences contained in the same fragment also rescue the *nonA*-associated ERG defect caused by the translocation and can be used to study whether *nonA* carries species-specific song information in these transformants (S. Campesan, Y. Dubrova, J. C. Hall and C. P. Kyriacou, unpublished results). In conclusion, the comparative analysis of *nonA* has clarified the molecular genetics of this genomic region and revealed some interesting and unusual evolutionary dynamics. These appear to reflect the unique regulatory relationships between *dGpi1* and *nonA* that are identified in the accompanying article (Sandrelli *et al.* 2001).

S.C. thanks the European Community for a predoctoral fellowship. This work was supported by a Human Frontiers Science Programme and Biotechnology and Biological Sciences Research Council (BBSRC) grant to C.P.K., Ministero per Universitá e la Ricerca Scientifica e Tecnologica (MURST) grants to R.C. and A.M., a MURST studentship for a "dottorato di ricerca" to F.S., a CNPq fellowship to A.A.P., and a Wellcome Trust International Research Development Award to A.A.P. and C.P.K.

LITERATURE CITED

Akiyama, Y., 1995 TFSEARCH: Searching Transcription Factor Binding Sites, http://pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html.

Bandziulis, R. J., M. S. Swanson and G. Dreyfuss, 1989 RNA-binding proteins as developmental regulators. Genes Dev. **3:** 431–437.

Bayer, C. A., L. von Kalm and J. W. Fristrom, 1997 Relationships between protein isoforms and genetic functions demostrate functional redundancy at the *Broad-Complex* during *Drosophila* metamorphosis. Dev. Biol. **187:** 267–282.

Besser, H. V., P. Schnabel, C. Wieland, E. Fritz, R. Stanewsky *et al.*, 1990 The puff-specific *Drosophila* protein Bj6, encoded by the gene *no-on-transientA*, shows homology to RNA-binding proteins. Chromosoma **100:** 37–47.

Comeron, J. M., and M. Kreitman, 1998 The correlation between synonymous and nonsynonymous substitutions in *Drosophila*: Mutation, selection or relaxed constraints? Genetics **150:** 767–775.

DiBello, P. R., D. A. Withers, C. B. Bayer, J. W. Fristrom and G. M. Guild, 1991 The *Drosophila Broad-Complex* encodes a family of related proteins containing zinc fingers. Genetics **129:** 385–397.

Doering, T. L., W. J. Masterson, P. T. Englund and G. W. Hart, 1989 Biosynthesis of the glycosyl phosphatidylinositol membrane anchor of the trypanosome variant surface glycoprotein: origin of the nonacetylated glycosamine. J. Biol. Chem. **264:** 11168–11173.

Engelman, D. M., T. A. Steitz and A. Goldman, 1986 Identifying nonpolar transbilayer helices in amino-acid sequences of membrane proteins. Annu. Rev. Biophys. Biol. **15:** 321–353.

Englund, P. T., 1993 The structure and biosyntheis of glycosyl phosphatidylinositol protein anchors. Annu. Rev. Biochem. **62:** 121–138.

Ewing, A. W., 1977 The neuromuscular basis of courtship song in *Drosophila*: the role of the indirect flight muscles. J. Comp. Physiol. A **119:** 249–266.

Gloor, G., and W. Engels, 1990 Single-fly DNA preps for PCR. *Drosophila* Inf. Newsl. 1.

Hancock, J. M., and J. S. Armstrong, 1994 SIMPLE34: an improved and enhanced implementation for VAX and SUN computers of the simple algorithm for analysis of clustered repetitive motifs in nucleotide-sequences. Comput. Appl. Biosci. **10:** 67–70.

Hancock, J. M., and G. A. Dover, 1988 Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. Mol. Biol. Evol. **5:** 377–391.

Hancock, J. M., P. J. Shaw, F. Bonneton and G. A. Dover, 1999 High sequence turnover in the regulatory regions of the developmental gene *hunchback* in insects. Mol. Biol. Evol. **16:** 253–265.

Heinemeyer, T., E. Wingender, I. Reuter, H. Hermjakob, A. E. Kel *et al.*, 1998 Databases on transcriptional regulation: TRANSFAC, TRRD, and COMPEL. Nucleic Acids Res. **26:** 364–370.

Hoikkala, A., and J. Lumme, 1984 Genetic control of the difference in male courtship sound between *Drosophila virilis* and *Drosophila lummei*. Behav. Genet. **14:** 257–268.

Hoikkala, A., and J. Lumme, 1987 The genetic-basis of evolution of the male courtship sounds in the *Drosophila virilis* group. Evolution **41:** 827–845.

Hotta, Y., and S. Benzer, 1970 Genetic dissection of the *Drosophila* nervous system by means of mosaics. Proc. Natl. Acad. Sci. USA **67:** 1156–1163.

Jenkins, D. L., C. A. Ortori and J. F. Y. Brookfield, 1995 A test for adaptive change in DNA sequences controlling transcription. Proc. R. Soc. Lond. Ser. B Biol. **261:** 203–207.

Jones, K. R., and G. M. Rubin, 1990 Molecular analysis of *no-on-transient A*, a gene required for normal vision in *Drosophila*. Neuron **4:** 711–723.

Kanaar, R., S. E. Roche, E. L. Beall, M. R. Green and D. C. Rio, 1993 The conserved pre-mRNA splicing factor U2AF from *Drosophila*: requirement for viability. Science **262:** 569–573.

Kinoshita, T., N. Inoue and J. Takeda, 1995 Defective glycosyl phosphatidylinositol anchor synthesis and paroxysmal nocturnal hemoglobinuria. Adv. Immunol. **60:** 57–103.

Klemenz, R., U. Weber and W. J. Gehring, 1987 The *white* gene as a marker in a new *P*-element vector for gene transfer in *Drosophila*. Nucleic Acids Res. **15:** 3947–3959.

Kulkarni, S. J., A. F. Steinlauf and J. C. Hall, 1988 The *dissonance* mutant of courtship song in *Drosophila melanogaster*: isolation, behavior and cytogenetics. Genetics **118:** 267–285.

Kyte, J., and R. F. Doolittle, 1982 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 105–132.

Leidich, S. D., and P. Orlean, 1996 Gpi1, a *Saccharomyces cerevisiae* protein that participates in the first step in glycosylphosphatidylinositol anchor synthesis. J. Biol. Chem. **271:** 27829–27837.

Li, W-H., and D. Graur, 1991 *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, MA.

Martin, C. H., C. A. Mayeda, C. A. Davis, C. L. Ericsson, J. D. Knafels *et al.*, 1995 Complete sequence of the bithorax complex of *Drosophila*. Proc. Natl. Acad. Sci. USA **92:** 8398–8402.

McConville, M. J., and M. A. J. Ferguson, 1993 The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. Biochem. J. **294:** 305–324.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

O'Neil, M. T., and J. M. Belote, 1992 Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. Genetics **131:** 113–128.

Pak, W. L., J. Grossfield and K. S. Arnold, 1970 Mutants of the visual pathway of *Drosophila melanogaster*. Nature **227:** 518–520.

Peixoto, A. A., S. Campsesan, R. Costa and C. P. Kyriacou, 1993 Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. Mol. Biol. Evol. **10:** 127–139.

Petersen, G., J. C. Hall and M. Rosbash, 1988 The *period* gene of *Drosophila* carries species-specific behavioral instructions. EMBO J. **12:** 3939–3947.

Powell, J. R., and E. N. Moriyama, 1997 Evolution of codon bias in *Drosophila*. Proc. Natl. Acad. Sci. USA **94:** 7784–7790.

Prestridge, D. S., 1991 SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. Comput. Appl. Biosci. **7:** 203–206.

Rendahl, K. G., K. R. Jones, S. J. Kulkarni, S. H. Bagully and J. C. Hall, 1992 The *dissonance* mutation at the *no-on-transient A* locus of *D. melanogaster*: genetic control of courtship song and visual behaviors by a protein with putative RNA-binding motifs. J. Neurosci. **12:** 390–407.

Rendahl, K. G., N. Gaukhshteyn, D. A. Wheeler, T. A. Fry and J. C. Hall, 1996 Defects in courtship and vision caused by amino acid substitutions in a putative RNA-binding protein encoded by the *no-on-transient A* (*nonA*) gene of *Drosophila*. J. Neurosci. **15:** 1511–1522.

Rutherford, S. L., 1995 The genetic and biochemical characterization of the major cyclophilin isoform in *Drosophila melanogaster*. Ph.D. Thesis. University of California, San Diego.

Sandrelli, F., S. Campsesan, M. G. Rossetto, C. Benna, E. Zieger *et al.*, 2001 Molecular dissection of the 5′ region of *no-on-transientA* of *Drosophila melanogaster* reveals *cis*-regulation by adjacent *dGpi1* sequences. Genetics **157:** 765–775.

Sandstrom, D. J., C. A. Bayer, J. W. Fristrom and L.L. Restifo, 1997 *Broad-Complex* transcription factors regulate thoracic muscle attachment in *Drosophila*. Dev. Biol. **181:** 168–185.

Schlotterer, C., M. T. Hauser, A. von Haeseler and D. Tautz, 1994 Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol. Biol. Evol. **11:** 513–522.

Sharp, P. M., E. Cowe, D. G. Higgins, D. C. Shields, K. H. Wolfe *et al.*, 1988 Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*: a review of the considerable within-species diversity. Nucleic Acids Res. **16:** 8207–8211.

Siomi, H., and G. Dreyfuss, 1997 RNA-binding proteins as regulators of gene expression. Curr. Opin. Genet. Dev. **7:** 345–353.

Smith, T. F., and M. S. Waterman, 1981 Identification of common molecular subsequences. J. Mol. Biol. **147:** 195–197.

Solovyev, V. V., and N. A. Kolchanov, 1994 Search for functional sites using consensus, pp. 16–21 in *Computer Analysis of Genetic Macromolecules*, edited by N. A. Kolchanov and H. A. Lim. World Scientific, Singapore.

Spradling, A. C., and G. M. Rubin, 1982 Transposition of cloned P elements into *Drosophila* germ line chromosomes. Science **218:** 341–347.

Stanewsky, R., K. G. Rendahl, M. Dill and H. Saumweber, 1993 Genetic and molecular analysis of the X chromosomal region 14B17-14C4 in *Drosophila melanogaster*: loss of function in *NON-A*, a nuclear protein common to many cell types, results in specific physiological and behavioral defects. Genetics **135:** 419–442.

Stanewsky, R., T. A. Fry, I. Reim, H. Saumweber and J. C. Hall, 1996 Bioassaying putative RNA-binding motifs in a protein encoded by a gene that influences courtship and visually mediated behavior in *Drosophila*: *in vitro* mutagenesis of *nonA*. Genetics **143:** 259–275.

Tautz, D., C. Tautz, D. Webb and G. A. Dover, 1987 Evolutionary divergence of promoters and spacers in the rDNA family of 4 *Drosophila* species: implications for molecular coevolution in multigene families. J. Mol. Biol. **195:** 525–542.

Treier, M., C. Pfeifle and D. Tautz, 1989 Comparison of the gap segmentation gene *hunchback* between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. EMBO J. **8:** 1517–1525.

Wheeler, D. A., C. P. Kyriacou, M. L. Greenacre, Q. Yu, J. E. Rutila *et al.*, 1991 Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. Science **251:** 1082–1085.

Communicating editor: J. J. Loros