

**Molecular and behavioural analysis of the
no-on-transientA (*nonA*) gene
of *Drosophila virilis*.**

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

**Susanna Campesan
Department of Genetics
University of Leicester**

MARCH 1998

UMI Number: U104137

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U104137

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

*"A fly got into the transmitter pod with me that first time...
My teleporter turned into a gene splicer, a very good one...
You look so pretty...Help me, please, help me."*

-BRUNDLEFLY

(Seth Brundle to Veronica in *The Fly*, D.Cronenberg, dir.
-Brooksfilms Toronto, 1986)

ACKNOWLEDGEMENTS

First of all I wish to thank everybody in the lab for being fun to work with and bearing with me all these years. I am terribly grateful (but not that grateful) to Dave Chalmers who taught me all the basic molecular techniques and scrupulously supervised me at the beginning of my work. I am also indebted to him for feeding me the first year I came to Leicester and for giving me driving lessons! He was a friend I could always rely on. Alexandre Peixoto and Valeria Ribeiro were also two of the best friends I had in Leicester, and I missed them when they left. Martin Couchman helped me a lot with my constant "computer problems" and saved me from years of hell by writing a macro for analysing *Drosophila* songs. Anthony Rogers and Ian Townsen were also tormented by my constant requests for help in dealing with computer programs, and even though they teased me about my incompetence, I am very grateful to them. Alberto Piccin and Ezio Rosato, being my compatriots and friends, were a constant source of help and good company. I know that by always talking in Italian we irritated our British colleagues, but I am actually glad that nobody could understand what those two were telling me!

I would like to thank Federica Sandrelli for performing *in-situ* hybridizations on my transformants, and for her enthusiasm and competence in developing the project we started together. A big thanks also to Rudi Costa who sent me to Leicester as an Erasmus student seven years ago, and by doing that probably changed the course of my life. I also owe him for sequencing the promoter of *nonA* in his lab, and for always caring about me. Yuri Dubrova showed an unprecedented enthusiasm for my song analysis data, and I am grateful to him for performing the multivariate analyses and helping me to understand the results he obtained. Magnus (I do not know his surname) from the Zoology Department was extremely nice for investing a great amount of time in trying to get my electroretinogram recordings to work. Even though we did not succeed with the ERGs, it was a pleasure meeting him.

Finally, I have not enough words to thank my supervisor Bambos Kyriacou, not only for his technical suggestions ("change the pH" always worked....not!), but for the constant support, encouragement, and guidance.

I also wish to thank my parents supporting me (also financially!), and my twin sister Giorgia and my best friend Rita for showing me how much more fun you can have if you avoid doing a Ph.D!

ABBREVIATIONS USED

ANOVA	Analysis of variance
A_x	Absorbance at a wavelength of x nm
bp	Base pair
BSA	Bovine serum albumin
°C	Degrees centigrade
Ci	Curie
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
DEPC	Diethyl pyrocarbonate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetraacetic acid
gr, mg, µg	Grammes, milligrammes, microgrammes
IPTG	Isopropyl-β-D-thiogalactopyranoside
Kb	Kilobase pairs
KV	Kilo Volts
L, ml, µl	Litres, millilitres, microlitres
M, mM, µM	Molar, millimolar, micromolar
min	minutes
mRNA	messenger ribonucleic acid
Myr	Million years
nm	nanometres
OD	Optical density
pmol	picomole
r.p.m.	revolutions per minute
RNA	Ribonucleic acid
RNase	Ribonuclease
s, ms	Seconds, milliseconds
SDS	Sodium dodecyl sulfate
TEMED	N,N,N',N'-tetramethyl-ethylenediamine
tRNA	Transfer ribonucleic acid
U	Unit (of enzymatic activity)
X-gal	5-bromo-4chloro-3indolyl-β-D-galactoside

ABSTRACT

The *nonA* gene of *Drosophila melanogaster* influences vision, courtship song and viability, and encodes a protein inferred to bind single-stranded nucleic acids.

This study describes the molecular and functional characterisation of the *nonA* gene of *Drosophila virilis*. The main goal of the project was to establish whether the *D.virilis nonA* gene rescued the viability and behavioural defects of *nonA*⁽⁻⁾ mutants, and whether it carried species-specific information concerning the lovesong. The *D.virilis nonA* gene was cloned and sequenced. The overall structure of the *D.virilis nonA* gene was similar to that of *D.melanogaster*. Nucleotide and amino acid sequence comparisons revealed a highly diverged and repetitive N-terminus, followed by a conserved C-terminal region. Study of the promoter region highlighted islands of homology containing putative transcription factor binding sites. In addition, the *D.virilis* gene was found to display an unusual codon usage bias in comparison to other *D.virilis* or *D.melanogaster* genes. Finally, protein secondary structure predictions revealed differences between the *D.virilis* and *D.melanogaster* polypeptides.

P element-mediated transformation was used to assess the ability of the *D.virilis nonA* clone to rescue viability, visual and song defects of *D.melanogaster* flies lacking *nonA* (*nonA*⁽⁻⁾) and an adjacent and partially overlapping lethal gene *l(1)i19e*. Complete rescue of the viability indicated that the transgene contained an entire functional *D.virilis l(1)i19e* gene. Visual studies assessing the transformant flies' optomotor response suggested that the *D.virilis* transgene completely rescued this phenotype. Lovesong analysis initially presented difficulties in distinguishing between incomplete rescue or transfer of species-specific song characteristics. However, extensive statistical analyses suggested that some *D.virilis* song traits may be conveyed to transformants by the *D.virilis nonA* gene, but the large differences between the *D.virilis* and *D.melanogaster* lovesongs are probably due to the influence of many other genes.

CONTENTS

ACKNOWLEDGMENTS	i
ABBREVIATIONS USED	ii
ABSTRACT	iii
CONTENTS	iv
CHAPTER 1: INTRODUCTION	1
Neurogenetics of sexual behaviour in <i>Drosophila</i>	3
The role of courtship song	11
Courtship song mutants	16
The <i>no-on-transientA</i> (<i>nonA</i>) gene	18
Aims of the project	28
 CHAPTER 2: MATERIALS & METHODS	
Basic procedures	30
Nucleic acids isolation	30
Genomic DNA isolation	31
Plasmid DNA isolation	31
Small scale	31
Midi prep	31
Large scale	31
Total RNA extractions	31
Polymerase Chain Reaction (PCR)	32
RT PCR	32
Gel electrophoresis and DNA recovery from gel bands	33
DNA sequencing	34
Direct sequencing of PCR products	34
Sequencing double stranded plasmid DNA	34
Southern Blotting of DNA	35
Preparation of radiolabeled DNA probe	35
Filters hybridization with ³² P probes	36
Prehybridization	36
Hybridization	37
Filters washing	37
Cloning procedures	37
Dephosphorylation of linearized plasmid DNA	37
1) Using calf intestinal alkaline phosphatase (CIP)	37

2) Using shrimp alkaline phosphatase (SAP)	38
Ligation reactions	38
Plasmid transformation (chemical method)	38
Preparation of competent cells	38
Transformation procedure	39
Transformation by electroporation	39
Cells for electroporation	39
Electroporation procedure	40
Bacterial cracking	40
Genomic DNA library screening	40
Plating and titration of the library	41
Preparation of plating bacteria	41
Plating bacteriophage λ	41
<i>In situ</i> hybridization of bacteriophage λ plaques	41
Preparation of the filters	41
Prehybridization, hybridization and washing of the filters	42
Isolation of positives and rescreening	42
Phage lysate preparation: plate method	42
Bacteriophage DNA isolation	43
<i>P</i> element transformation of <i>Drosophila</i>	43
Embryos collection and dechoriation	43
Microinjection of embryos	44
Post-injection care of embryos	44
Behaviour	45
Fly stocks and balancers	45
Stock maintenance	45
Recording and analysing <i>Drosophila</i> lovesongs	46
Recording <i>D.melanogaster</i> songs	46
Recording <i>D.virilis</i> songs	46
Song characters analysed	47
Cycles <i>per</i> pulse (CPP)	47
Pulse frequency (PF)	48
Pulse song (PS)	48
Sine song (SS)	49
Sine song frequency (SSF)	49
Sine song proportion (SSP)	49
Interpulse intervals (IPIs)	49
Walking optomotor test	50
Testing <i>D.melanogaster</i>	50
Testing <i>D.virilis</i>	50
Statistical analyses	51
Computer analyses of DNA sequences	51

CHAPTER 3: The *D.virilis nonA* gene

Introduction	52
Methods	56
<i>D.virilis</i> genomic library	56
<i>D.virilis</i> PCR fragment	56
Library screen	56
Subcloning and sequencing	58
Sequence of the <i>D.virilis nonA</i> promoter region	60
Computer analyses	60
Results & Discussion	61
Structure of the <i>D.virilis nonA</i> gene	61
The 5' regulatory region of <i>nonA</i>	62
The coding sequence of the <i>D.virilis nonA</i> gene	68
Conceptual translational start of the <i>D.virilis nonA</i> gene	71
Protein comparison	71
Protein motifs in NONA	73
NONA secondary structure	74
Molecular evolution of the <i>nonA</i> gene	76
Codon usage	81

CHAPTER 4: Transformation of *nonA*^{vir}

Introduction	86
Methods	86
Transforming <i>w; e Sb Δ2-3/TM6</i> flies	87
Transforming <i>w</i> ¹¹⁸ flies	87
Southern blotting of transformed lines	88
Mapping the inserts	88
Mobilization of X-linked inserts	89
Results	89
Lines 112 and 113	89
Lines 168-8 and 67-4	91
Lines 72, 135, and 297	93
Lines 297-6, 297-64, 297-71	93
Lines 75, 97, 191	93
Discussion	94

CHAPTER 5: Viability tests

Introduction	100
Methods	101
Placing the <i>nonA</i> ^{vir} construct on a <i>T(1;4)9e2-10</i> background	101
Viability tests	101
Results	102

Discussion	103
CHAPTER 6: Analysis of visual behaviour	
Introduction	105
Materials & Methods	107
Genotypes	108
A: Controls	108
Hemizygous <i>nonA</i> ⁽⁻⁾ [<i>nonA</i> ⁺]	108
B: hemizygous <i>nonA</i> ⁽⁻⁾ [<i>nonA</i> ^{vir}]	109
C: heterozygous <i>nonA</i> ^{vir} / <i>nonA</i> ⁺	109
D: <i>3mel</i>	110
E: <i>2mel</i>	110
F: <i>mel+2vir</i>	111
Genetic background	111
Results	112
Hemizygous <i>nonA</i> ⁽⁻⁾ [<i>nonA</i> ⁺]	114
Hemizygous <i>nonA</i> ⁽⁻⁾ [<i>nonA</i> ^{vir}]	114
Heterozygous <i>nonA</i> ^{vir} / <i>nonA</i> ⁺	115
<i>2mel</i>	116
<i>3mel</i>	116
<i>mel+2vir</i>	116
Discussion	117
CHAPTER 7: Song analysis	
Introduction	121
Materials & Methods	122
Genotypes	122
Song analysis	123
Results	123
bCPP	125
CPP	129
CPP1	130
CPP10	131
CPP distributions	132
bPF	135
PF	137
SSP	139
SSF	141
IPI	143
Multivariate analyses	146
Discussion	149
Poor rescue or species-specific effect?	150

Dosage effect?	155
CHAPTER 8: DISCUSSION	156
APPENDICES FOR CHAPTER 3	
Appendix 3.1	166
Appendix 3.2	167
Appendix 3.3	168
Appendix 3.4	171
Appendix 3.5	172
APPENDICES FOR CHAPTER 6	
Appendix 6.1	173
APPENDICES FOR CHAPTER 7	
Appendix 7.1	174
Appendix 7.2	175
Appendix 7.3	176
Appendix 7.4	177
Appendix 7.5	178
Appendix 7.6	179
Appendix 7.7	180
Appendix 7.8	181
Appendix 7.9	181
Appendix 7.10	182
Appendix 7.11	182
REFERENCES	183

CHAPTER 1

INTRODUCTION

The ideas of inheritance, particularly of human behavioural characteristics, have engaged human thought for hundreds of years. However it is only in the last century that the basics of what we now understand as the genetics of behaviour, were founded. Francis Galton, whose thinking was profoundly influenced by that of his cousin, Charles Darwin, was the first to address the question of whether human behaviour contained a hereditary component. By using statistical methods, he analysed different physical and mental characteristics in large numbers of families, including the first behavioural studies of twins (Galton, 1869). His conclusions stressed that nature prevails over nurture, yet neither he nor Darwin, were aware of the particulate theory of inheritance which had been published earlier by Mendel. If they had, both could have supported considerably their respective theories, on the inheritance of human mental capacity on one hand, and evolution by natural selection on the other.

The debate concerning the relative contributions of nature and nurture, took a downward turn in the early years of the 20th century, especially in the United States, where by the 1920's the extreme environmentalism of J.B. Watson, totally dominated Psychology. Small pockets of resistance remained, particularly in the laboratory of Tolman, who performed the first selection experiment on 'smart' and 'dumb' rats. His student, Tryon extended this early work, and generated the famous maize bright and maize dull strains of rats (Tryon; 1940). However, the work was not fully published until 20 years after the experiment, such was the prevailing atmosphere. A breath of fresh air finally came from Europe, where the students of the ethologists Lorenz and Tinbergen, began to experiment with the genetic analysis of mice and fruit-flies (reviewed in Manning, 1975). The latter species was particularly important given the enormous genetic literature that Morgan and his students had been amassing on *Drosophila* from about 1915 onwards. So by the early 1960's, the first papers on behavioural genetic analysis had appeared, and there was no turning back.

Welcome as this new development was, most behavioural genetic research focused on studying behavioural traits that could only be described as quantitative. It was unusual that any study, whether it was on the mouse, the rat or the fly, implicated single genes in determining a significant proportion of behavioural variation. This was

to change in 1967, when Benzer advocated mutagenesis as an approach to dissecting nervous system function in *Drosophila* (Benzer, 1967; 1971).

His students enthusiastically took up the challenge, and using clever screening procedures, isolated behavioural mutants which spanned the range from simple reflexes like flight, to more complex phenotypes, such as learning. With other genetic methods, such as mosaic mapping, the anatomical foci of mutant genes could be roughly localised (Benzer, 1973). This procedure, which was originally based on gynandromorphs, could also be extended to map sex-specific behaviour in both male and female flies (Hotta and Benzer, 1976).

By the early to mid-1980's, *Drosophila* behaviour genetics had far surpassed that of its main rival, the mouse. Yet it was only with the advent of transformation technology (Rubin and Spradling, 1982), that the first behavioural gene was cloned (Reddy *et al.*, 1984; Bargiello and Young., 1984), and then demonstrated to encode a behavioural protein by rescue of the mutation (Zehring *et al.*, 1984; Bargiello *et al.*, 1984). The gene in question was *period* (*per*), which remains the 'flagship' for the single gene dissection of behaviour, and which determines both circadian locomotor cycles (Konopka and Benzer, 1971), as well as fast one-minute oscillations in the male's courtship song (Kyriacou and Hall, 1980; 1989). Thus the mutate, clone and transform school had its successes, and many genes, initially identified as behavioural mutants, have provided major stories, not necessarily in the study of behaviour, but also in development (ie *sevenless*, Rubin, 1988; Simon *et al.*, 1989), and channel physiology (eg *Shaker*, Wu and Ganetzky, 1992).

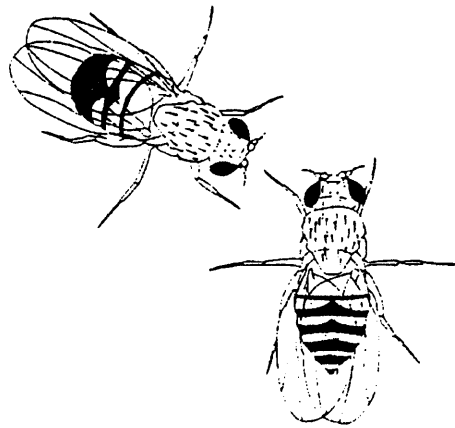
More recently, reverse genetics has been brought to bear on behaviour in the fly (Sentry *et al.*, 1994). The second generation enhancer trap using the yeast transcriptional activator GAL4 allows the simultaneous visualisation of expression patterns, and the misexpression of any gene that can be placed under GAL4 control, in the tissue of interest (Kaiser, 1993). The recent misexpression of the *transformer* (*tra*) gene to feminize male brain structures provides an elegant example of this approach (Ferveur *et al.*, 1995, O'Dell *et al.*, 1995). As can be seen from this brief review into the history of behavioural genetics, the fruitfly looms large in the story, and the sexual

behaviour of *Drosophila*, has been particularly prominent as a model phenotype (Hall, 1994). I shall limit the rest of my introduction to a discussion of sexual behaviour in the fly, and its relevance to behavioural genetics and speciation.

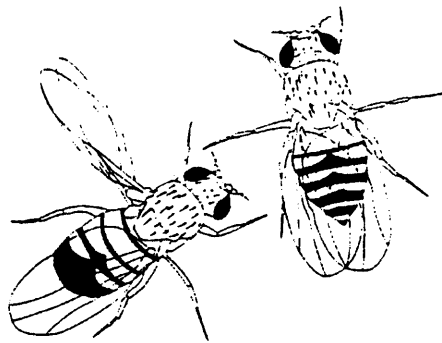
Neurogenetics of sexual behaviour in *Drosophila*.

Courtship in *Drosophila* involves a series of interactive behaviours performed by both male and female. However the male takes the more active part of the courtship which is extensively documented (figure 1.1; for a review, see Hall, 1994). When a mature *Drosophila* male perceives a female, he orients towards her and starts tapping the female's abdomen using his foreleg. If the female moves, he follows her, extends one wing and vibrates it. The wing vibration produces the "lovesong", an acoustic signal which constitutes the most effective and species-specific component of each fly's courtship. Depending on the female's receptiveness, he may go back and repeat the same actions several times. Finally, the male will extend his proboscis and lick the female's genitalia. This is always followed by his first copulation attempt. If the female rejects him, he might stop courting for few seconds. When he resumes courting, the male invariably goes back to orientation and wing vibration stages, and repeats the behavioural sequence. When a sexually mature female has been courted sufficiently, she may accept the male's copulation attempt and allow copulation which has a species-specific duration (in *D.melanogaster* it lasts about 20 minutes).

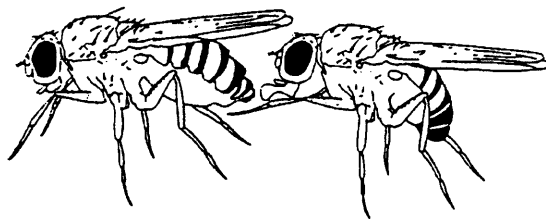
Courtship behaviours are innate: males isolated as eggs and kept singly until maturity are able to recognize a potential mate and perform effectively all the courtship steps described above (O'Dell and Kaiser, 1997). However, this does not mean that courtship is a completely stereotyped series of actions that cannot be modified. For example, previously mated females elicit relatively little courtship compared to virgin females, and block the male's copulatory attempts by extruding their ovipositor (Connolly and Cook, 1973). Males that have courted fertilized females show reduced levels of courtship when they are subsequently placed in the presence of virgin females (Siegel and Hall, 1979). It seems that the males are not inhibited by rejection



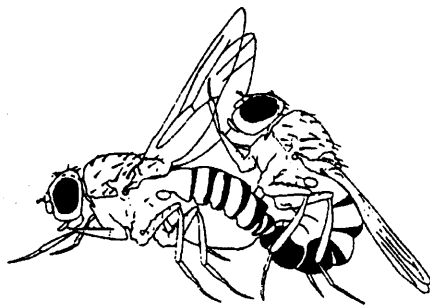
ORIENTATION



SONG



LICKING



**ATTEMPTED
COPULATION**

Figure 1.1. The principal elements of courtship in *D.melanogaster*. Original drawings by Barrie Burnet.

behaviour itself, but by aversive olfactory cues emitted by the female (Ackerman and Siegel, 1986). The decreased courtship towards virgin females after exposure to mated females results from counterconditioning, in which the male associates otherwise excitatory pheromones with aversive chemicals simultaneously emitted by the mated female (Zawistowski, 1988). Thus after exposing the male to a fertilized female, appetitive stimuli (female pheromones) either acquire aversive qualities or lose their excitatory capacity.

In a similar experiment, males were presented to virgin females in the presence of quinine, a repellent used as a negative reinforcer for conditioning of *D.melanogaster* (Ackerman and Siegel, 1986). The males exposed to virgin females in the presence of quinine revealed reduced courtship levels in a subsequent test with virgin females in the absence of quinine. However, males placed alone, without virgin females, in the presence of quinine, did not show decreased courtship in subsequent tests with virgin females, in the absence of quinine. This suggests that the exposure to quinine itself has no after-effects, and the observed decline in courtship was due to counter-conditioning by exposure to aversive (quinine) and excitatory (virgin female pheromones) stimuli together.

Studies with learning and memory mutants (for a review, see Davis, 1996) tend to support this scenario. *dunce* (*dnc*, Davis and Kiger, 1981; Byers *et al.*, 1981) and *amnesiac* (*amn*, Folkers, 1982; Feani and Quinn, 1995) mutant males, showed no reduction in courtship levels after simultaneous exposure to virgin females and quinine (Ackerman and Siegel, 1986). Mutant males such as, *amnesiac* and *rutabaga* (Gailey *et al.*, 1984; Dudai, 1985; Livingstone, 1985; Tully and Gergen, 1986; Levin *et al.*, 1992) perform inappropriately vigorous courtship immediately after being trained in the presence of mated females (Hall, 1994). The same is observed in transgenic flies carrying a specific inhibitor of the calcium-calmodulin-dependent protein kinase II (CaMKII), an enzyme known to be involved in the neuronal mechanism of learning and memory (Griffith *et al.*, 1993). Males carrying this construct could be trained in the presence of mated females, in that after one hour of exposure to mated females, they showed the same decrements in courtship shown by wild-type flies. However, the

transformant males failed to retain the effects of the training: when subsequently presented with virgin females, instead of showing inhibition of courtship, they courted vigorously.

Another kind of experience-dependent modification of behaviour can be observed in *Drosophila*. One-day old or younger males tend to elicit vigorous courtship from mature males since they produce an aphrodisiac pheromone which is different from that of virgin females (Jallon, 1984, Vaias *et al.*, 1993). However, when a male is placed together with an immature male, its courting behaviour tends to decrease after a while. Successively, the mature male will completely avoid further courtship of young males (Gailey *et al.*, 1982). It seems that habituation is the key to this type of courtship-related learning, and it is sufficient to expose mature males to the sex-stimulating substances extracted from young males to repress their attraction towards young males (Gailey *et al.*, 1986). All the learning mutants mentioned above are aberrant in this regard, and keep courting young males regardless to their previous experiences with them (Hall, 1986, Hall, 1994).

Apart from stressing the importance of learning and memory as important aspects of courtship behaviours, the experiments described above also underline the fundamental role that pheromones play in mate stimulation and discrimination. It is not surprising then that mutant males that cannot smell normally, such as *smellblind* (poor responses to a variety of volatile compounds) and *olfactoryC* (poor responses to acetates and some alcohols), are also subnormal in their courtship performance since they probably do not sense the females' sex-inducing odours (Tompkins *et al.*, 1980; Tompkins and Hall, 1981).

Male-to-female chemical communication also plays an important role during courtship. The mutation *nerd* (*nrd*, Ferveur and Jallon, 1993) dramatically reduces the levels of 7-tricosene (7-T, the male predominant sex pheromone, Ferveur and Jallon, 1993) synthesized by males. *nerd* males are subnormal in their courtship behaviour and reproductive success. It seems that females find *nerd* male's courtship insufficiently stimulating, and fail to emit the next stimulus in the courtship sequence.

Consequently, *nerd* affects the probability of a given mutant to reach the more advanced stages of courtship behaviour.

Also visual stimuli are extremely important during courtship, especially for the male, which has to be able to track the female. Blind males such as *no-receptor-potentialA* (lacking light-elicited photoreceptor potentials) and *glass* (lacking photoreceptors and having reduced compound eyes) are not able to orient properly towards the female and consequently spend less time courting than wild-type males (Tompkins, 1984). *Optomotor-blind* (*omb*, defective in their optomotor responses, Pflugfelder *et al.*, 1990; Pflugfelder *et al.*, 1992), which are insensitive to horizontally moving stimuli, behave during courtship as if they were totally blind, and court subnormally (Tompkins, 1986, Hall, 1994). Also *white* (*w*) mutants court poorly, since they lack screening pigments that shield the eye from being flooded by too much light and as a result they do not respond properly to moving stimuli (Hall, 1994). If only visual and olfactory cues provided by the females were responsible for stimulating males to court, then males that cannot see or smell would be expected not to be able to court at all. However, double mutants *smellblind; glass* display courtship if they can touch the female, showing that tactile stimuli are also important during courtship (Tompkins, 1984). Since a mature male courts vigorously virgin females, but is not attracted by other mature males, olfactory, visual and tactile stimuli are undoubtedly sex-specific.

Our understanding of the genetic control of sexual differentiation and sexual behaviour in *Drosophila* has progressed enormously since the first studies on sex-determination, started almost 90 years ago by Morgan and his colleagues (see Belote, 1992; Baker, 1989; Hodgkin, 1989; Hodgkin, 1990; Burtis, 1993; McKeown, 1994). The primary determinant of sex in *Drosophila* is the ratio of X chromosomes to sets of autosomes (X:A ratio). One of the first responses to the X:A ratio in females (X:A=1), is the activation of the *Sex-lethal* (*Sxl*) locus (Cline, 1984). *Sxl* autoregulates the splicing of its own pre-mRNA (Bell *et al.*, 1991) and activates *transformer* (*tra*) by regulating the *tra* messenger splicing (Inoue *et al.*, 1990; Boggs *et al.*, 1987; Belote *et al.*, 1989; Sosnowski *et al.*, 1989). The products of the *tra* and

the *transformer-2* (*tra-2*, Belote and Baker, 1982; Mattox and Baker, 1991; Amrein *et al.*, 1988; 1990 Goralski *et al.*, 1989) genes then regulate the splicing of the pre-mRNAs of the *doublesex* (*dsx*, Baker and Wolfner, 1988; Hedley and Maniatis, 1991) gene, leading to the production of a female-specific form of the Dsx protein (Dsx^F) (Tian and Maniatis, 1993; Hoshijima *et al.*, 1991; Hedley and Maniatis, 1991; Burtis and Baker, 1989; Ryner and Baker, 1991). In males (X:A=0.5) *Sxl* and *tra* do not make functional products, and *dsx* pre-mRNA is spliced in its default manner, resulting in a male-specific Dsx protein (Dsx^M) which differs from the female-specific protein at its carboxy-terminus. The Dsx^F and Dsx^M proteins function as sex-specific transcription factors leading respectively to female and male differentiation.

Using temperature-sensitive diplo-X mutants (*Sxl*^{M#1, fm#3}/*Sxl*^{fm#7, M#1}), which externally looked like males, it was shown that the *Sxl* gene must function normally during the pupal period or the first few days of adult life for a fly to elicit vigorous courtship (Tompkins, 1984b; 1986), suggesting that the female-specific *Sxl* transcript is essential for the production of the female pheromone, 7,11-heptacosadiene (HCD) (Tompkins *et al.*, 1980; Tompkins *et al.*, 1981; Jallon and Cobb, 1987; Ferveur and Sureau, 1996; Ferveur *et al.*, 1996), and the repression of the male inhibitory pheromone 7-tricosene (7T) synthesis (Jallon, 1984; Ferveur and Sureau, 1996; Ferveur and Jallon, 1996). Haplo-X flies carrying a gain-of-function *Sxl* mutation (which leads to ectopic expression of female-specific *Sxl* functions) elicit anomalously high levels of interest from normal males, and also perform less courtship than wild-type males, supporting the idea that *Sxl* controls the production of female pheromones (Tompkins and McRobert, 1989).

The synthesis of female-specific pheromones is regulated by *Sxl* via *tra* and *tra-2*, since *tra* and *tra-2* mutations make diplo-X flies unattractive to normal males (Baker and Ridge, 1980; McRobert and Tompkins, 1985). In males *tra* is not functional, and it seems that the gene *intersex* (*ix*) may play a role in the production of sex-specific pheromones. In fact *ix* mutations, although they have no morphological effects on flies, make haplo-X flies attractive to normal males but not as attractive as females, indicating that these mutants are either producing some HCD (stimulating pheromone)

or not producing enough 7T (inhibitory pheromone) (Baker and Ridge, 1980; McRobert and Tompkins, 1985; Jallon, 1984; Scott, 1986; Scott *et al.*, 1988; Scott and Jackson, 1988).

There are no doubts that *Sxl*, *tra* and *tra-2* control all aspects of somatic, neuronal and behavioural sex-specific differentiations. XX flies homozygous for *tra* mutations not only look like males but they also behave like males, being able to court females vigorously and to sing to them in an essentially normal manner (Kyriacou and Hall, 1980; Kulkarni *et al.*, 1988; Rendahl *et al.*, 1992; Bernstein *et al.*, 1992). A conditional mutation of *tra-2* allowed Belote and Baker (1987) to "turn off" the gene at different stages of the flies' life cycle simply by raising the temperature. It was found that inactivation of the gene in diplo -X pupae or even young adults could lead to the appearance of male courtship behaviour. This result suggested that the product of the *tra-2* gene must be functional in adults in order to maintain a female-specific behavioural programme. In addition, it indicated that the adult nervous system retains some functional plasticity in respect to the sex-specificity of courtship (Belote and Baker, 1987).

In contrast, *dsx* does not control all aspects of sex-specific somatic differentiation. For example, *tra* and *tra-2*, but not *dsx*, control the development of a pair of bilaterally symmetrical male-specific abdominal muscles called Muscles of Lawrence (MOLs, Lawrence and Johnston, 1986; Taylor, 1992). It seems that the development of these muscles depends on the sex of the neurons that innervate them (Lawrence and Johnston, 1986; Currie and Bate, 1995). In addition, constitutive expression of the male form of the Dsx protein transforms females to morphological males, but these flies do not court (Taylor *et al.*, 1994). These data suggest that a previously unknown branch of the sex-determination hierarchy immediately downstream of *tra* and *tra-2* may control sex-specific neuronal development and behaviour. It has to be noted however that *dsx* is involved in the development of one set of neurons (Taylor and Truman, 1992) and one particular aspect of the courtship song (Villella and Hall, 1996).

The best candidate for acting in this new neuro-behavioural branch of the sex-determination pathway seems to be the *fruitless (fru)* gene (Hall, 1978; Gailey *et al.*, 1991; Taylor, 1992). The original mutation at the *fru* locus caused males to court both females and males and to be courted by wild-type males. The most striking behaviour of the *fru* mutant males is that they form long "courtship-chains" in which every individual is simultaneously courting and being courted (Gill, 1963; Gailey and Hall, 1989). *fru* mutants are never able to copulate (Hall, 1978), and, in addition, they fail to develop a normal MOL (Gailey *et al.*, 1991; Taylor and Knittel, 1995). Ito *et al.* (1996) and Ryner *et al.* (1996) have recently cloned *fru*. Both groups found that *fru* is a large and transcriptionally complex locus, carrying at its 5' untranslated region three copies of the Tra-Tra-2 recognition motif. The predicted translation products include a male-specific and a female-specific form and show similarities to the BTB-ZF family of transcription factors (Albagli *et al.*, 1995). These results, and the fact that *fru* is expressed in a population of brain cells, including those of the antennal lobe, which have been suggested to be involved in determination of sexual orientation (see later), support an intriguing scenario in flies in which sexual orientation is controlled by the same hierarchy of genes that govern all other aspects of sex.

A second *tra*-dependent but *dsx*-independent locus controlling sexual behaviour has been recently isolated (Finley *et al.*, 1997). Male mutants of this gene, named *dissatisfaction (dsf)*, are bisexual and although they attempt to copulate, they do so with difficulty due to a defect in their abdominal curling mechanism. Interestingly, *dsf* mutations cause abnormal female courtship in that they resist males and fail to lay mature eggs. The defective abdominal curling in males and the inability to lay eggs in females seems to be due to abnormal differentiation of sex-specific abdominal neurons. Because of several features similar to those observed in *fru* mutants (homosexual courtship, abdominal curling defects) *dsf* could be the first *fru*-dependent gene in the putative neuro-behavioural sexual cascade. However fundamental behavioural and neurological differences exist between *dsf* and *fru*, such as the courtship and fertility abnormalities observed in *dsf* females, and the fact the *dsf* males produce normal MOLs. Therefore Finley *et al.* (1997) suggest that *dsf* and *fru* might be acting in two separate pathways.

Various areas of the nervous system have been identified as foci for determined aspects of courtship behaviour (reviewed in Greenspan , 1995 and Hall, 1994). In the 1970s, J.C. Hall was one of the first researchers that used gynandromorphs (part female-part male) flies to try to find the areas in the brain that need to be either male or female in order to achieve a particular step in the courtship pattern. He found that, in order to be able to perform the early steps of courtship, such as orientation, tapping, following and wing extension, at least one side of the dorsal posterior brain has to be haplo-X (Hall, 1977; 1979). For two of the more advanced behaviours, such as wing vibration and attempted copulation, male tissue is required in the thoracic ganglia. In particular, only the left or right ventral nervous system needs to be male for normal courtship song output from both wings (Schilcher and Hall, 1979). Another advanced courtship behaviour, licking, has its focus in the dorsal posterior brain, but in this case the focus is submissive, since both sides of this part of the brain must be haplo-X in order to achieve proboscis extension (Hall, 1979). Tompkins and Hall (1983) also identified the brain sites controlling female receptivity. They found a group of cells in the dorsal anterior brain that have to be bilaterally diplo-X in order to ensure receptivity to copulation.

More recently, Ferveur *et al.* (1995) have found the neural foci that control sexual orientation in *Drosophila*. Using the Gal4 system (Brand and Perrimon, 1993), they drove ectopic expression of the *tra* gene in order to feminize determined regions of the male brain. They discovered that males that were feminized in a portion of the antennal lobes, or in a subset of the corpora pedunculata (mushroom bodies), courted males and females indiscriminately. These two regions of the brain are known to be involved in the detection and processing of olfactory stimuli (Carlson, 1991; Stocker, 1994; Stocker and Gendre, 1989; Heisenberg *et al.*, 1985; de Belle and Heisenberg, 1994). Therefore these partially feminized strains might be lacking male-specific centres that enable them to detect antiaphrodisiac pheromones produced by normal males. Alternatively, the feminization of the above structures might enable the affected flies to detect sex-stimulating male pheromones. However, since these flies' brains are still mainly male, their behavioural output remains typically male-like.

An experiment similar to that of Ferveur *et al.* (1995) involved ectopic expression of *tra* in well defined subsets of cells within the mushroom bodies (known as Kenyon cells), and mapped with even greater accuracy several sites controlling sexual orientation (O'Dell *et al.*, 1995). Sexual orientation has also been altered by the ectopic general expression of the *white (w)* gene. Activation of a *hsp-70/miniwhite* transgene in mature males resulted in a drastic change in their sexual preference since they courted vigorously other males and mainly ignored females. The cause of this unexpected phenomenon is not known, however it could have something to do with lower levels of serotonin induced by the misexpression of the *w* gene (Zhang and Odenwald, 1995). Interestingly, it was also noted that wild-type mature males placed in a vigorous homosexual courtship scenario started to court other males too, demonstrating that also environmental factors can influence sexual orientation in flies (Zhang and Odenwald, 1995).

The role of courtship song.

Since it was noted that the wing-vibration sounds emitted by the male during courtship are species-specific (Waldron, 1964), the courtship songs of more than a hundred species of *Drosophila* have been characterized (Ewing and Bennet-Clark, 1968; 1969; Ewing, 1970; 1979; Miller *et al.*, 1975; Chang and Miller, 1978; Lakovaara and Hoikkala, 1979; Ikeda *et al.*, 1980; Cowling and Burnet, 1981; Ewing and Miyan, 1986).

The *D.melanogaster* courtship song consists of two main components: pulse song and sine song (figure 1.2). The former is composed by trains of sound pulses ranging from a minimum of two to a maximum of 50 pulses per train. Each pulse usually consists of one to three cycles (Kulkarni and Hall, 1987) and lasts about 5-10 ms. Successive pulses are separated by intervals (interpulse-intervals, IPIs) with a mean value of 35 ms (Ewing and Bennet-Clark, 1968). In 1980, Kyriacou and Hall noted that, in *D.melanogaster*, the IPI values oscillate sinusoidally between 28 ms and 40 ms, with a period of 55-60 s. Finally, the sine song is composed of humming

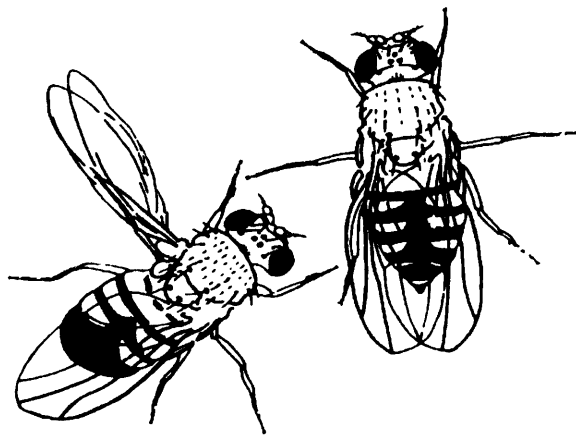
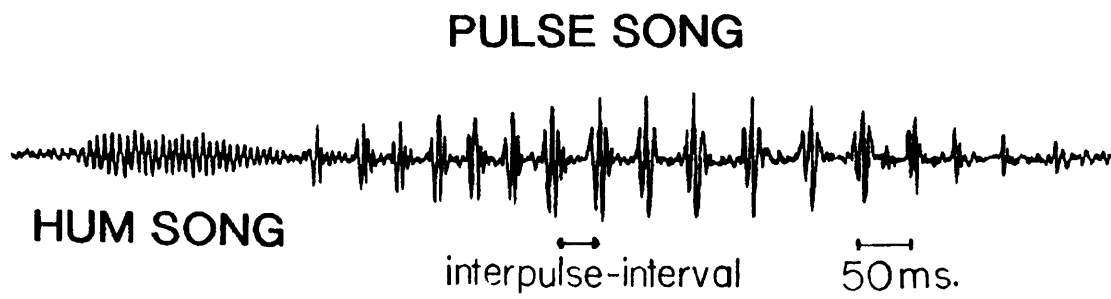


Figure 1.2. Burst of *D.melanogaster* courtship song, including sine (hum) and pulse components.

sounds of variable length with fundamental frequencies of about 160-170 Hz (Cowling and Burnet, 1981; Wheeler *et al.*, 1989). A continuous period of song constitute what Cowling and Burnet (1981) called a phrase. Phrases, which are of variable length, usually contain both pulse and sine song in no fixed order, however phrases containing only one or other of these components are often observed (Wilson *et al.*, 1976).

Cowling and Burnet (1981) analyzed and compared the courtship behaviour and songs of six sibling species in the *D.melanogaster* subgroup: *D.melanogaster*, *D.mauritiana*, *D.simulans*, *D.yakuba*, *D.teissieri* and *D.erecta*. They found clear differences both in behaviour and lovesongs characteristics between these six species, with the only exception being *D.mauritiana*. The courtship song produced by this species in fact has a mean sine song frequency (SSF) and intrapulse frequency (IPF) close to that of *D.simulans*, while its IPI values resemble those of *D.melanogaster*. The ambiguity of the *D.mauritiana* song can be explained assuming that species-specific song differences have a function in allowing recognition of conspecific flies and thus maintain sexual isolation between different species. Sexual isolation mechanisms need to be stronger when two species have the potential to interbreed, for example when they are sympatric (Fisher, 1930; Dobzhansky, 1940; Welbergen *et al.*, 1987). Reinforcement is the process by which natural selection strengthens sexual isolation between incipient species, reducing the frequency of maladaptive hybridization and hence completing reproductive isolation (Noor, 1995). Since *D.mauritiana* is found only on Mauritius, where no other species of the *D.melanogaster* subgroup are found, there is no need for its song to be very different from that of the other species. On the contrary, *D.melanogaster* and *D.simulans* live in sympatry, therefore their songs may differ considerably in order to maintain sexual isolation between the two species. Cowling and Burnet found that *D.simulans* has a pulse song with a mean IPI value of 55 ms. However, other workers have reported a value of 48 ms (Ewing and Bennet-Clark, 1968). It seems that the IPI values of *D.simulans* are in general more variable than those of *D.melanogaster* (Kawanishi and Watanabe, 1980). The IPI rhythm in this species was found to be 35-40 s (Kyriacou and Hall, 1980; 1986).

In the *D. virilis* group species it seems that differences in courtship songs might also contribute to maintaining sexual isolation between sympatric species (Hoikkala *et al.*, 1982). The *montana* phylad species (*D. kanekoi*, *D. littoralis*, *D. ezoana*, *D. borealis*, *D. flavomontana*, *D. lacicola*, *D. montana montana*, and *D. montana ovivorum*) live in sympatry and are very effectively sexually isolated (Patterston and Stone, 1952; Stone *et al.*, 1960; Throckmorton, 1982). As expected, these species emit very distinct courtship sounds. In contrast, the *virilis* phylad species (*D. americana americana*, *D. americana texana*, *D. novamericana*, *D. lummei*, and *D. virilis*) are not very effectively sexually isolated and are usually allopatric. The songs of these species are very similar to each other and consist of dense pulse trains with very short IPIs, the pulse trains of *D. virilis* being shorter than those of the rest of the species (Hoikkala *et al.*, 1982).

In their study, Cowling and Burnet (1981) tried to map the song differences found between the six species of the *D. melanogaster* subgroup by generating interspecific hybrids (when possible) between any pair of the species analyzed. The characteristics of the hybrid songs suggested that the genes controlling the IPI were located on the autosomes. Sine song frequency seems to be determined mainly by autosomal factors, but the ability to generate sine song appears to be sex-linked. *D. yakuba* males do not produce any sine component in their lovesongs, and interspecific hybrid males from a cross *D. yakuba* females x *D. mauritiana* males also show absence of this element in their courtship songs (Cowling and Burnet, 1981), while hybrid males from the reciprocal cross (females *D. mauritiana* x males *D. yakuba*) always produce sine song (Demetriades, 1997). Finally, the control of the mean intrapulse frequency in the six sibling species of the *D. melanogaster* subgroup seems to be largely autosomal (Cowling and Burnet, 1981).

Similar studies performed on the songs of *D. persimilis* and *D. pseudoobscura* revealed that the genes controlling the qualitative aspects of the songs, that is whether the overall pattern is the *pseudoobscura* or *persimilis* type, are located on the X chromosome. However in these two species the IPI length seems to be controlled by autosomal genes (Ewing, 1969). In the *D. virilis* phylad, differences in several traits of

the lovesong (such as pulse train length (PTL), number of pulses per train (PN), IPI length, pulse length (PL), number of cycles in a pulse (CPP), and length of a cycle (CL)) appear to be determined by autosomal genes (Hoikkala and Lumme, 1987). In crosses between *D.virilis* and *D.littoralis* or *D.flavomontana* (both species belonging to the *montana* phylad) reciprocal hybrids differed from each other in PTL, IPI, PL and CPP, indicating a sex-linked or cytoplasmic inheritance. In the backcrosses between *D.virilis* and *D.littoralis* the involvement of the X chromosome was found to be decisive. Hoikkala and Lumme (1987) propose therefore that a major change on the X-chromosome allowing variation in IPI occurred during the separation of the two *virilis* group phylads (*virilis* and *montana*), the long IPI allowing variation also in PL and CN. The evolution of the sounds in the *virilis* phylad seems to have gone towards longer and denser pulse trains, whereas in the *montana* phylad, the sounds have evolved in various directions.

In addition to its function as a species-discriminator, the courtship song has an undoubted role in stimulating the female to mate. Von Schilcher (1976a; b) found that *D.melanogaster* females prestimulated with sine song, before being placed with males, showed reduced copulation latencies compared to females which had not been pre-stimulated. However, a pulse song with a constant IPI of 34 ms did not have the same effect in similar experiments. Nevertheless, the same pulse song could speed mating latencies of females in the presence of mute, wingless males. Von Schilcher suggested that females might summate sine songs, increasing their receptivity to a threshold level. On the contrary, pulse song might act immediately as a trigger to mating.

Kyriacou and Hall (1982; 1986) studied the receptivity of *D.melanogaster* and *D.simulans* females which had been stimulated with artificially generated pulse songs. The pulse songs they tested were more elaborate than those used by Von Schilcher, since they incorporated all the possible combinations of species-specific 35 ms IPI, 48 ms IPI, 55 s and 35 s oscillation periods. It was found that *D.melanogaster* females responded best to stimulation with a song with *melanogaster*-like characteristics, 35 ms IPIs with a superimposed rhythm of 55 s, while *D.simulans* females preferred a *simulans*-like song with 48 ms IPI and a 35 s rhythm. Therefore pulse song can

stimulate females when both the IPI oscillation and IPI values are species-specific. Greenacre *et al.* (1993) subsequently observed that *D.melanogaster* females actively discriminated against the *D.simulans* heterospecific song: they did not necessarily prefer a 55 s song cycle, since a 80 s rhythm stimulated them as effectively. However they did reject a 35s cycle *simulans*-like song.

D.melanogaster females can also be pre-stimulated by a species-specific 55s rhythmic-35 ms IPI pulse song (Kyriacou and Hall, 1984). Pre-stimulation effects seem to involve mechanism of information storage and retrieval of information (the conspecific IPI cycles). In fact memory and learning mutants such as *dunce*, *rutabaga*, *amnesiac* and CAM-kinase depleted females show minimal or short lived effects of song pre-stimulation (Kyriacou and Hall, 1984; Griffith *et al.*, 1993).

The stimulation and pre-stimulation experiments (Kyriacou and Hall, 1982; 1984; 1986) highlight the importance of the IPI cycle component of the pulse song as a means of discriminating conspecific mates. However, it has been argued that courtship encounters in the wild might be very brief, not allowing enough time for the females to "calculate" the IPI oscillations (Ewing and Ewing, 1987). Nevertheless it is possible that females recognize the rate of variation of the IPI around a certain value. Indeed, Alt *et al.* (1998) have recently suggested that females may recognize the "nature" of a cycle from the IPIs generated in less than a complete 60 s oscillation. Alternatively, females might possess a preferred IPI value: by rhythmically changing its IPI values, a male increases his chances of producing the right IPI values for stimulating a large number of females (Kyriacou and Hall, 1982).

Thus IPI seems to be one of the most important song parameters involved in sexual stimulation and conspecific recognition during courtship. Evidence from a study of the *D.auraria* complex supports this idea. The species of the *auraria* complex are mostly sympatric and show strong premating isolation. Species-specific courtship song and failure in copulating displayed by "mute" males suggest that the courtship song is particularly important in maintaining sexual isolation in this complex. Interestingly, the IPI is the only song character which is consistently species-specific

among the several courtship elements studied in these species (Tomaru and Oguma, 1994a; 1994b).

Analysis of the songs produced by reciprocal *D.melanogaster*/*D.simulans* interspecific hybrid males allowed Kyriacou and Hall (1986) to show that mean IPI values are controlled by autosomes, whereas IPI cycles are determined by the X chromosome. This was consistent with the X chromosome *per* locus determining species-specific lovesong cycle variation, because Kyriacou and Hall (1980; 1989) noticed that three mutations of the *period* (*per*) locus which shortened (*per^s*), lengthened (*per^L*) and obliterated (*per⁰¹*) the circadian rhythms of *D.melanogaster* (Konopka and Benzer, 1971), had parallel effects on the song IPI cycles. The *per^s* mutant had an IPI rhythm of ~40 s, *per^L* had a cycle of ~80 s, and *per⁰¹* seemed not to have any evident IPI cycles. Furthermore, deletion of a repetitive Threonine-Glycine (TG) region from an otherwise wild-type *D.melanogaster per* transgene produced IPI cycles of ~ 40 s in *per⁰¹* transformant flies (Yu *et al.*, 1987). These results supported the idea that the *per* locus, and more specifically its TG region, could carry species-specific information related to the IPI cycling. This was confirmed by the work of Wheeler *et al.* (1991), who generated several chimeric constructs in which the TG repeat, together with some flanking DNA, were interchanged between *D.simulans* and *D.melanogaster per* genes. Transformant *per⁰¹* males carrying the *D.melanogaster per* gene, but a *D.simulans* TG region, produced songs with the characteristic *D.simulans* ~40 s IPI cycles. Transformant males carrying the *D.simulans per* gene, but a TG region from *D.melanogaster*, sang with *melanogaster*-like IPI cycles of ~ 60 s.

Courtship song mutants.

Only a few loci have been identified, along with *per*, which alter the male's courtship song. The *cacophony* (*cac*) mutant was isolated through mutagenesis, by identifying flies which took too long to mate (von Schilcher, 1977; Kulkarni and Hall, 1987). The characteristics of a *cac* courtship song are polycyclic pulses of high amplitude and longer than normal IPIs (on average 45 ms instead of the wild-type 35

ms; figure 1.3). In spite of such grossly different pulse song patterns, the sine song produced by *cac* mutants is essentially indistinguishable from that of wild-type males. *cac* males and females are observed to fly normally with normal wing-beat frequency, thus the mutation appears to affect only the physiology of pulse song generation (Kulkarni and Hall, 1987). However it must be noted that *cac* males are less successful in mating than wild-type flies, even after wing amputation, which suggest that *cac* mutants bear some additional defects that reduce their mating ability (Kulkarni and Hall, 1987). Such defects were subsequently mapped to a separate locus after it was noted that recombinant *sn*³ *cac* males, in spite of their abnormal polycyclic song, initiated mating much more quickly than did males of the original *cac* stock (Kulkarni and Hall, 1987).

Another genuine courtship song mutation identified, the *dissonance* mutation (Kulkarni *et al.*, 1988), turned out to be an allele of the *no-on-transientA* (*nonA*) gene (Rendahl *et al.*, 1992), and will be described in detail in the next section.

Yokokura *et al.* (1995) identified *croaker* (*cro*) mutants which initiate courtship after a long delay, if compared to wild-type males, and produce polycyclic pulse songs with prolonged IPIs, similar to those of *cac* males. Because flying ability is significantly impaired in *cro* mutants, a general defect in the motor system was suspected. However this possibility was excluded after several tests that confirmed that muscle function is normal in these mutants. Additionally, electroretinogram (ERG) recordings (see later) and gustatory behaviour essays of *cro* mutants appeared normal. These findings support the notion that *cro* flies have a defect confined to the CNS, leaving the sensory input and the motor output systems intact. The fact that the *cro* mutation also reduces female attractiveness without changing pheromonal production (D.Yamamoto and J-M Jallon, unpublished results) supports the idea that it causes a central defect.

It should be mentioned that the two mutants *fru* and *dsx*, known for their effects on sexual behaviour and sexual determination, also generate defective courtship songs. The original *fru*¹ mutants (Gill, 1963) and the secondly isolated *fru*² (Moses, 1989), have been observed to produce a pulse song with aberrantly long IPIs (Wheeler

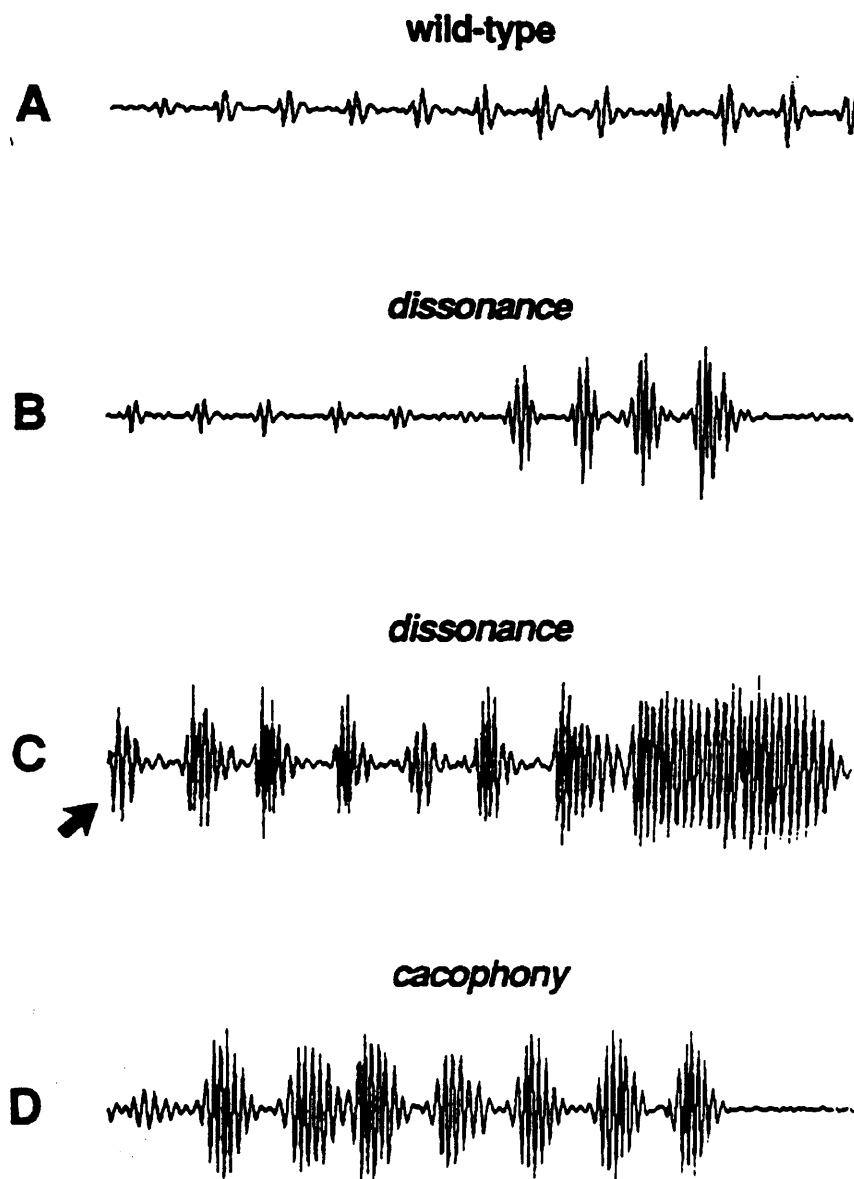


Figure 1.3. Pulse song of mutant and normal males. **A:** a wild-type (Canton S) male's typical bout of pulse song. **B:** a moderately aberrant bout produced by a *nonA^{diss}* male; note that this train starts with wild-type-like pulses. **C:** wildly aberrant burts from the same male as in B, though only the latter portion of this pulse train is shown (i.e. the arrow points to pulse No. 11; this train, like that in B, commenced with normal-appearing pulses). **D:** a typical *cac* pulse train. The total time for each plot is *ca.* 400 ms. Picture taken from Kulkarni *et al.*, (1988).

et al., 1989, Villella *et al.*, 1997). Moreover, the other two *fru* mutations discovered so far (*fru*³ and *fru*⁴, Castrillon *et al.*, 1993) have been shown to completely suppress pulse song (Villella *et al.*, 1997). *dsx* mutant males were found to produce essentially normal pulse song bouts, but to completely lack the sine song component (Villella and Hall, 1996).

The *no-on-transientA* (*nonA*) gene.

The sex-linked *no-on-transientA*^{dissonance} (*nonA*^{diss}) mutation was found directly by screening for flies that produced aberrant songs (Kulkarni *et al.*, 1988). As in the case of *per*, and the *Dmca1A*-encoding locus (of which *cac* is an allele, Smith *et al.*, 1996), it was found that the *nonA* locus actually affected more than one aspect of the fly's behaviour (Kulkarni *et al.*, 1988; Smith *et al.*, 1996; Rendahl *et al.*, 1992; Smith *et al.*, in press). *nonA* is involved in at least two distinct and apparently unrelated behaviours: the fly's vision and the male's courtship song.

The first *nonA* mutants were isolated by the group lead by Benzer (Hotta and Benzer, 1970; Pak *et al.*, 1970), in their screens for flies defective in phototaxis (the tendency to follow the light). The same mutants also produced an altered electroretinogram (ERG), which is an extracellular measure of the light-generated currents in the eyes and brain of the insect (Heisenberg, 1971b; Pak, 1975). The ERG of *nonA* mutants lacked both light-on and light-off transient spikes, whereas the maintained depolarization which is produced by the sum of the photoreceptors generation potentials, was apparently normal (figure 1.4). The on and off-transients are believed to arise from the large monopolar cells L1 and L2 in the lamina, which is the first optic ganglion postsynaptic to the photoreceptor cells (Pak, 1975; Coombe, 1986). Thus the lack of transient spikes could mean a defect in the transmission from the photoreceptors to the neurons of the lamina, or the inability of the the neurons of the lamina to respond to information from the photoreceptors. Furthermore *nonA* mutants show several other visual defects which suggest an even more central defect in the nervous system. For example, *nonA* mutants are impaired in their optomotor

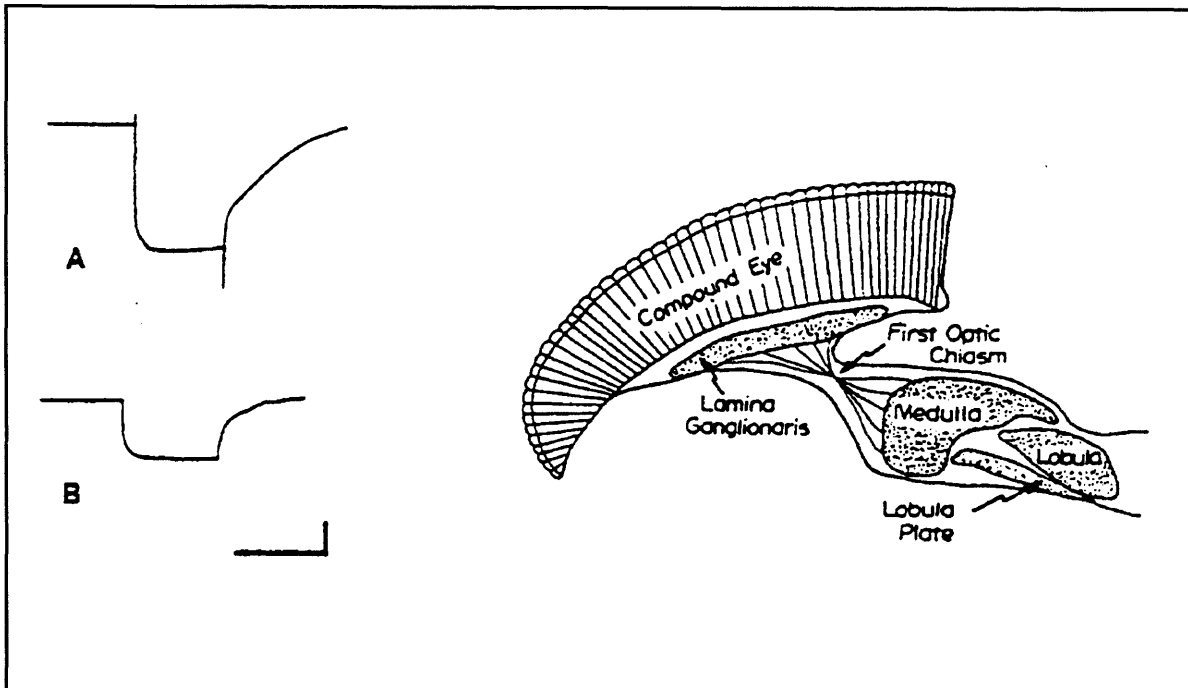


Figure 1.4. **A:** typical ERG of a wild-type fly. **B:** ERG recorded from a mutant *nonA^{diss}* fly; note the complete absence of light-on and light-off transient spikes. The vertical scale bar represents 5mV, the horizontal indicates 3s. On the right, a schematic representation of the dipteran compound eye and optic ganglia.

response (Heisenberg, 1972; Heisenberg and Gotz, 1975). The optomotor test in *Drosophila* measures the tendency of the fly to follow movement in its environment. If a fly is located in the middle of a plexiglass rotating drum which has vertical black and white stripes, the insect tends to follow the direction of the stripes. *nonA* mutants show different degrees of impairment in this test (Heisenberg, 1972). In particular, *nonA*^{H2} flies seem to respond better to regressive motion (back to front) than to progressive motion (front-to-back, Heisenberg, 1972).

The *nonA*^{diss} mutation was isolated, as mentioned above, for its effect on the male's courtship song (Kulkarni *et al.*, 1988). In *nonA*^{diss} the pulse song is specifically affected in that pulse trains start out normally, but as the train proceeds, pulses become increasingly polycyclic and very high in amplitude (figure 1.3). However, short trains in *nonA*^{diss} (containing less than 6 pulses) are usually completely normal. An extreme phenotype is displayed by *tra* mutant diplo-X flies (which look and behave as males) carrying the *nonA*^{diss} mutation over a deletion of the *nonA* locus: the pulses produced by these flies can bear up to 100 cycles per pulse (Rendahl *et al.*, 1992). Further analysis of *nonA*^{diss} mutants revealed that these flies are impaired also in their visual behaviour, and they display exactly the same defects found in all the other *nonA* visual-only mutations (Kulkarni *et al.*, 1988). As a consequence of their visual impairment, *nonA*^{diss} males frequently lose contact with the females they are courting, and sometimes are observed to sing to empty spaces. However they do not court without the presence of a virgin female, since they do not sing if isolated in fresh unused courtship chambers, nor do they produce any wing vibration when placed in mating chambers that have previously hosted virgin females (Kulkarni *et al.*, 1988). The aberrant song of *nonA*^{diss} males is not related to the fact that they continuously lose contact with the female, since blind males (such as *norpA*^{P24}) produce perfectly normal lovesongs (Kulkarni *et al.*, 1988).

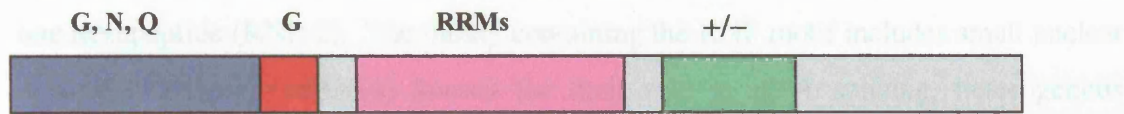
nonA alleles are all recessive and display a complex pattern of complementation. *nonA*^{diss}, when heterozygous with any of the visual-only mutations (*nonA*^{H2}, *nonA*^{P14}, *nonA*^{P49}, *nonA*^{P60}), in a *tra* genetic background (to generate diplo-X males), leads to absence of ERG transient spikes and lack of optomotor responses.

Only two of the visual-only alleles (*nonA*^{H2} and *nonA*^{P49}) can clearly complement *nonA*^{diss} song defects. Heterozygous *nonA*^{diss}/*nonA*^{P14} "pseudomales" (*tra* diplo-X flies) produce songs intermediate between the wild-type and *nonA*^{diss} (Rendahl *et al.*, 1992). On the other hand all the *nonA* alleles, except *nonA*^{diss}, produce perfectly normal courtship songs (Kulkarni *et al.*, 1988). The fact that *nonA*^{P14} males are closer to wild-type in their singing behaviour than diplo-X *nonA*^{diss}/*nonA*^{P14} heterozygotes is probably due to dosage compensation of the expression levels for X-linked genes (Jaffe and Laird, 1986). Thus one dose of the *nonA*^{P14} allele in a haplo-X fly would produce higher level of gene product than one dose of the same allele in a diplo-X individual. This observation also suggests that *nonA*^{diss} is, in terms of biological functions, more similar to a null mutation than is *nonA*^{P14}.

The *nonA* locus has been mapped to the cytological position 14C1-2 by *in-situ* hybridization (Jones and Rubin, 1990, Besser *et al.*, 1990). The gene was cloned independently and almost contemporaneously by two groups (Jones and Rubin, 1990, Besser *et al.*, 1990). Jones and Rubin (1990) used P-element mediated transformation to show that two overlapping 11 and 16 Kb fragments of genomic DNA can rescue all the *nonA* visual defects. Subsequently, Rendahl *et al.* (1992) demonstrated that the same two fragments can also rescue all the phenotypes caused by the *diss* (*nonA*^{diss}) mutation, thus undoubtedly confirming that *diss* is an allele of the *nonA* locus. Besser *et al.* (1990) identified the NONA protein among proteins extracted from polytene chromosomes of *Drosophila* larvae, and used an antibody against it to detect a *nonA* cDNA from an expression library.

The 9 Kb DNA fragment encoding the *nonA* gene has been sequenced (Jones and Rubin, 1990, Besser *et al.*, 1990). Conceptual translation of the locus predicts a single open reading frame of 700 amino acids. The NONA protein presents some interesting features (see figure 1.5): its N-terminal third is very rich in glycine, asparagine and glutamine. The C-terminal two thirds of the protein is rich in charged residues (Jones and Rubin, 1990, Besser *et al.*, 1990). The central segment of the protein contains two tandemly repeated 80 amino acid motifs common to a family of proteins known for their ability to bind RNA. This 80 amino acid sequence is named

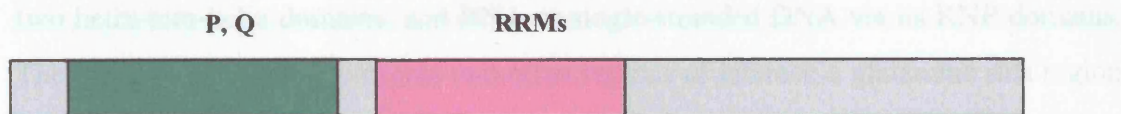
NONA



NONO



PSF



p54

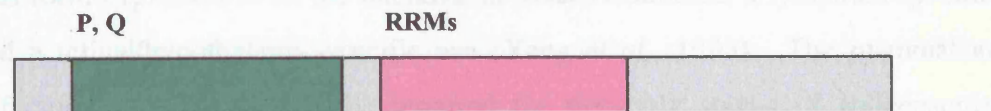


Figure 1.5. Schematic representation of the putative functional domains of the NONA, NONO, PSF and p54 proteins. RRM= RNA-binding domains (RRM1 and RRM2). +/- = charged domain. HTH = helix-turn-helix motif. G, N, Q= Glycine, Asparagine and Glutamine-rich region. G= poly-Glycine repeat. Q= Glutamine-rich region. P= Proline-rich region. P, Q= Proline and Glutamine-rich region.

RRM (RNA recognition motif, known also as RNP or RBD motif, Siomi and Dreyfuss, 1997) and is moderately conserved from yeast to man (Bandziulis, 1989). The most conserved segments within such domain are one octapeptide (RNP-1) and one hexapeptide (RNP-2). The family containing the RNP motif includes small nuclear ribonucleoproteins (snRNPs) known for their role in RNA splicing, heterogeneous ribonucleoproteins (hnRNPs), and several *Drosophila* gene products involved in the fly's development, such as *embryonic lethal abnormal visual system* (*elav*, Robinow *et al.*, 1988), or in sex determination, such as *Sex-lethal* (*Sxl*, Bell *et al.*, 1988) and *transformer-2* (*tra-2*, reviewed in Baker, 1989).

Three mammalian proteins are related to the product of *nonA* in a region containing the two RRM domains plus a downstream ~ 110 amino acid peptide sequence: the product of *non-O* (Yang *et al.*, 1993), the splicing factor PSF (Patton *et al.*, 1993), and the HeLa protein p54^{nrb} (Dong *et al.*, 1993). The NONO protein is known to be able to bind specifically to an octamer of double-stranded DNA through two helix-turn-helix domains, and RNA or single-stranded DNA via its RNP domains. The sequence of NONO presents two other regions of interest: a glutamine rich region and a proline rich one (figure 1.5). These regions, like the glycine rich domain found in NONA, are believed to be functionally important in protein-protein interactions (Patton *et al.*, 1993). So far, three different forms of NONO have been isolated: an ubiquitous form expressed in all the mammalian tissues examined, a lymphoid-specific form, and a retinal/hypothalamic-specific one (Yang *et al.*, 1993). The mammalian splicing factor PSF is known to be required for the early stages of spliceosome formation *in vitro*. This protein displays a marked preference for uridine stretches at the branch-point of mammalian introns (Patton *et al.*, 1993). The function of the HeLa protein p54^{nrb} is still unknown, however, preliminary analyses suggest that this protein could bind to RNA *in vitro* (Dong *et al.*, 1993). Interestingly, both PSF and p54^{nrb} are unusually rich in Gln and Pro at their N- termini. The ~310 amino acid sequence homologous between *nonA* and these three mammalian proteins has been named the DBHS domain, which stands for *Drosophila* behaviour and human splicing domain (Dong *et al.*, 1993). More recently, an additional gene product has been found to share homology with *nonA* in a region overlapping the DBHS domain. This protein is

the predicted translation product of a novel *Drosophila* P1 clone (M. Palazzolo, unpublished material), which shares a 99% identity with *nonA* in the DBHS region (Rendahl *et al.*, 1996).

RNA-binding proteins can perform a broad variety of functions, such as capping, polyadenylation, constitutive or alternative splicing of pre-mRNAs, transport in the cytoplasm or stabilization of mRNAs (Rendahl *et al.*, 1996; Graumann and Marahiel, 1996). Several RRM-containing proteins of *Drosophila* have been identified and their role in mRNA splicing actually demonstrated (review: Rio, 1993). Many of these proteins are hypothesized to be specifically required for mRNA splicing in the nervous system (reviewed by Yao *et al.*, 1993). Examples are the product of the gene *elav*, which is expressed at early stages of neuronal differentiation (Robinow and White, 1988; Robinow and White, 1991); *couch potato* (*cpo*), expressed in precursor and differentiated cells of the PNS (Bellen, *et al.*, 1992); *musashi* (*msi*), which is expressed in neurons of the PNS and CNS and is involved in adult organ sensory development (Nakamura *et al.*, 1994). Because of its homology to such proteins, and given its pleiotropic behavioural functions, it is believed that the product of *nonA* might be implicated in some aspects of processing of nervous system-specific mRNAs (Rendahl *et al.*, 1996).

Pre-mRNAs known to undergo regulated splicing in the *Drosophila* nervous system, such as those transcribed from the sodium, potassium, and calcium channel genes, and pre-mRNAs from loci required for wild-type visual or courtship behaviours, are all possible targets of NONA splicing functions. For example, alternatively spliced transcripts obtained from the *Shaker* (*Sh*) potassium channel locus are localized in the dorsal longitudinal muscles (DLMs) and retina of late pupae and adults (Mottes *et al.*, 1993). Several mRNAs encoding voltage-gated sodium-channel polypeptides are transcribed from the *para* gene (Thackeray and Ganetzky, 1994). The histidine-decarboxylase-encoding *hdc* gene, which is essential in the production of ERG transients, also gives rise to multiple transcripts which are found in various regions of the central brain and the thoracic ganglion (Burg *et al.*, 1993). Finally, the transcript of the *Dmca1A*-encoding locus, which gives rise to a voltage-gated calcium channel

protein, and displays a complicated pattern of alternative splicing, is a particularly appealing candidate for being regulated by NONA (Smith *et al.*, 1996). The alleles of the *Dmca1A*-encoding gene, in fact, produce an array of courtship song, visual behaviour, and viability defects similar to, although distinguishable from, those produced by *nonA* alleles. The *cac* mutation specifically affects, like the *diss* mutation, the pulse portion of the male's courtship song, rendering the pulses visibly polycyclic and high in amplitude. However *cac* mutant songs can be distinguished from *diss* since in *cac* all the pulses, independently from their position in a given train, are highly polycyclic (Kulkarni and Hall, 1987). The *night-blindA* mutation, another allele of the *Dmca1A*-encoding locus (Smith *et al.*, 1996) affects the optomotor responses and produces reduced transient spikes in ERG recordings. Finally, mutant alleles of *l(1)L13* (which also map to the *Dmca1A*-encoding locus Smith *et al.*, 1996), cause late embryonic lethality (Kulkarni and Hall, 1987). The diversity of the *Dmca1A* transcripts suggests that the complicated complementation interactions observed between *cac*, *nbA*, and *l(1)L13* mutations (Kulkarni and Hall, 1987) could be due to isoform-specific lesions (Smith *et al.*, 1996).

Besser *et al.* (1990) showed that the NONA protein is localized in several puffs on polytene chromosomes of *Drosophila* larvae, and its binding seemed correlated with active transcription within these puffs (Frash and Saumweber, 1989). This finding is not in disagreement with the idea that NONA could be a splicing factor. In fact there is evidence that splicing can occur on nascent transcripts (Dong *et al.*, 1993; Beyer and Osheim, 1988). Moreover, in *Drosophila* there are two other examples of splicing factors which can be found on polytene chromosome active puffs: the proteins B52 (Champling *et al.*, 1991) and RBP1 (Kim *et al.*, 1992). Both proteins are members of the SR family of nuclear phosphoproteins, which appear to function in constitutive and alternative splicing (Mayeda *et al.*, 1992; Zahler *et al.*, 1992; Kim *et al.*, 1992; Fu *et al.*, 1992).

The *nonA*^{P14}, *nonA*^{H2} and *nonA*^{diss} mutant alleles have been sequenced by Rendahl *et al.* (1996). They are all caused by single point mutations: *nonA*^{P14} codes for an asparagine instead of a glycine at amino acid position 400 (based on the *nonA*

sequence from Jones and Rubin, 1990); *nonA*^{H2} has an arginine instead of a glycine at position 463, while in *nonA*^{diss} a cysteine is found at position 548 instead of the wild-type arginine. The mutations in *nonA*^{P14} and *nonA*^{H2} are respectively in the centre of and just outside the second RRM repeat. The latter is close enough to this RRM that it could affect ligand binding (Kenan *et al.*, 1991). *nonA*^{diss}, which is the most severe mutation, surprisingly does not lie in the RRM domains, but it is found in the proximal charged region. Simplistically it could then be inferred that NONA visual functions are particularly sensitive to alterations in the RRM domains, with courtship song being also affected whenever the charged region is mutated. An alternative interpretation would be that the array of phenotypes observed in *nonA* mutants results from increasingly severe, non-specific alterations of the protein. Therefore, relatively mild *nonA* mutations would affect visual behaviour only, while more severe protein alterations would affect courtship song as well.

The levels of protein produced by each mutant allele were examined in Western-blot by Rendahl *et al.* (1992). It was found that all the mutants, with the exception of *nonA*^{P14}, produced normal amounts of full-length NONA protein. The fact that *nonA*^{P14} has lower protein levels is intriguing since this allele affects only vision. This observation, taken together with the fact that a *nonA* encoding transgene which yields subnormal levels of protein is still able to rescue all *nonA*-mutant phenotypes, suggest that protein levels might not be a primary factor in determining the mutant phenotypes (Stanewsky *et al.*, 1996).

The phenotypic effects caused by an amorphic allele of the *nonA* locus have been studied by generating an X-linked deletion (*T(1;4)9e2-10*) which uncovered both *nonA* and a distal, partially overlapping, lethal locus (*l(1)i19e*, Jones and Rubin, 1990). Viable *nonA*⁽⁻⁾ mutants were then generated by transforming flies carrying the (*T(1;4)9e2-10*) deletion with a clone encoding exclusively for the lethal gene's product (Stanewsky *et al.*, 1993). In hemizygous males, deletion of the *nonA* gene causes semilethality: only 10-30% of the expected number eclose from metamorphosis. If kept at a temperature below 25°C no *nonA*⁽⁻⁾ males emerge. *nonA*⁽⁻⁾ individuals are deficient in locomotor activity and flight and less than one-third of them survive to

reproduce. In visual and courship behaviour, *nonA*-deficient males show similar, although more severe defects to those of *nonA^{diss}* mutants, thus confirming that the *nonA^{diss}* mutation is indeed the allele closest to an amorphic mutation.

Several other *nonA* mutations have been generated *in vitro* in order to assess the behavioural repercussions of altering highly conserved and functionally important amino acid residues within the RNA-binding domains and the C-terminal charged region (Rendahl *et al.*, 1996; Stanewsky *et al.*, 1996). The results of these studies reveal that the first RRM domain (RRM1) in *nonA* is absolutely necessary for all the known functions of NONA. 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. In particular, mutations replacing functionally important residues in the RNP2 and RNP1 of RRM1 resulted in "worse-than-nothing" phenotypes: mutant flies were in fact much more subviable and feeble than *nonA⁽⁻⁾* flies. This suggests that these mutations could be antimorphic, in a way that the mutated form of NONA might be actually interfering with the functioning of other, similarly acting, polypeptides. In *nonA⁽⁻⁾*, other nucleic-acid-binding complexes could be partially compensating for the gene's product absence, without the interference of qualitatively altered NONA molecules. The RRM2 domain, on the contrary, seems to be much less important for the functioning of the *nonA* product. Two mutations in the RNP2 of RRM2 left unaltered all the visual, courtship and viability phenotypes tested. Other mutations at the RRM2 produced defects of the visual behaviour only, leaving the song and viability phenotypes generally unaltered. Therefore it has been suggested that RRM2 might be primarily involved in visual behaviour. Given its lower level of conservation, compared to that of RRM1, the second RRM of NONA could be specifically processing transcripts required in the nervous system for normal vision in a stringent sequence-specific recognition manner. The high stringency of such NONA-ligand interactions for normal visual output is consistent with the observation that, whenever a mutation affects NONA functions, the visual phenotype is the first to be affected. Increasingly severe mutations (such as those affecting the RRM1 of NONA) would lead to subsequent song and, finally, viability defects.

Interestingly, all the *in vitro* RRM mutated forms of NONA bind to chromatin in an essentially normal manner (Stanewsky *et al.*, 1996). This implies that domains important for chromatin binding might be different from those interacting with RNA. In previous experiments, Stanewsky *et al.* (1993) had also shown that two different C terminal-truncated forms of NONA could bind to chromatin, suggesting that this part of the protein is not responsible for NONA's presence on polytene chromosomes. Furthermore, both two C-terminal truncated NONA types were unable to rescue any behavioural mutant phenotype, indicating that all the biological functions of NONA are independent from its binding to chromatin.

The distribution of *nonA* RNA and protein during embryonic development has been studied (Rendahl *et al.*, 1992). The *nonA* transcript is already present in the fertilized egg. At a more advanced developmental stage its expression seems to decrease in all tissues except in the central nervous system (CNS) where it starts to decline only towards the end of embryonic development. The expression pattern of the NONA protein is similar, but not identical, to that of its RNA. The protein seems to be present ubiquitously in the developing oocyte and the embryonic, larval, pupal and adult stages. Interestingly, the protein is localized in the cytoplasm of the fertilized egg, but it enters the nuclei at the cellular blastoderm stage (Frash and Saumweber, 1989). It then remains nuclear in all tissues during embryonic development. For the protein there is no apparent central nervous system-specific expression as observed for the *nonA* RNA. In the adult flies, NONA is localized in the nuclei of cells of most tissues analyzed. Its expression was detected in many non-neural tissues as well as in the nervous system (Rendahl *et al.*, 1992). The expression pattern of *nonA* RNA and protein has been studied in all the *nonA* chemically-induced mutants and compared to that of wild-type flies. All the mutants show the same amount and distribution of RNA and protein as seen in the wild-type. The only exception is *nonA*^{P14} which presents a significant decrease in antibody-staining levels, consistent with the results obtained with Western Blots for this allele (Rendahl *et al.*, 1992).

It is rather surprising that a protein which is ubiquitously expressed throughout all stages of development and adulthood, would be involved in determining specific

and distinct aspects of behaviour such as vision and courtship. However, NONA is not a unique case in this respect. For example, the products of the *per* and *dnc* genes are found in most *Drosophila* tissues analyzed during all developmental and adult stages. In spite of this, deletions of *per* or *dnc* are fully viable, but mutations at either of these two loci cause multiple but specific defects in the fly's behaviour (circadian and courtship behaviour in the case of *per*, reviewed by Hall and Kyriacou, 1990; learning behaviour in the case of *dnc*, reviewed by Quinn and Greenspan, 1984; Hall, 1986; Tully, 1987).

The significance of the temporal expression of *nonA* has been dissected by Rendahl and Hall (1996) with the use of a heat-shock-inducible *nonA* construct. An *hsp-nonA*⁺ (cDNA) transgene was placed in a *nonA*^{dis} genetic background and activated either during development only, or during the first five days of adulthood only, or during both development and imaginal stages. Surprisingly, it was found that any of these three temporal patterns of *nonA* expression could produce full rescue of the visual and courtship singing defects of *nonA*^{dis} mutants. The fact that the conditional *nonA* construct was tested on a *nonA*^{dis} genetic background instead of a *nonA*⁽⁻⁾, did not allow full appreciation of *nonA* requirements during development in terms of viability and general vigour. *nonA* does have a role in development, since *nonA*⁽⁻⁾ flies often do not eclose and the gene product is known to be essential for the female germ line (Stanewsky *et al.*, 1993). The finding that *nonA* expression during development only is enough to guarantee full rescue of behaviours in the adult, could be explained if the NONA protein was very stable. In support of this, it has been noted that, in wild-type, NONA protein levels are still high in embryonic tissues 6 hours after the *nonA* mRNA levels have started to decline (Stanewsky *et al.*, 1993). Alternatively, *nonA* expression during development could turn on a cascade of downstream genes whose products then keep acting for a very long time. *nonA* is not the only case of a gene whose expression in adulthood only is enough to obtain rescue of the behaviours it controls. Expression of the *per* gene (using a *hsp-per* construct) exclusively in adulthood or even only during behavioural testing, assured rescue of the rhythmic phenotypes associated with the locus. However for *per*, continuous expression of the gene is required to obtain the relevant rhythmic behaviours, owing to the fact that PER

protein has a very short half life (Ewer *et al.*, 1988; 1990). Thus expression of *per* in pre-adult stages only is not enough to restore rhythmicity once the flies are adults.

Aims of the project.

We have seen how mutations of the *nonA* gene can affect the expression of the male's love song. Sound pulses produced by *nonA^{diss}* mutants are reminiscent of the types of pulses produced by males of other *Drosophila* species. One of these species is *D.virilis* whose song pulses are in fact polycyclic (Hoikkala *et al.*, 1982; Hoikkala and Lumme, 1987). This similarity between a mutant phenotype in *D.melanogaster* and a wild-type phenotype in another species is analogous to the case of the *per^S* mutation of *D.melanogaster*. This mutation not only shortens the circadian locomotor activity cycles of the fly (Konopka and Benzer, 1971), but it also shortens the ~60 s male's love song cycles to about 40 s (Kyriacou and Hall, 1980; 1981). Males of *D.simulans* sing with love song cycles of about 40 s (Kyriacou and Hall, 1980; 1986) and we have seen how species-specific changes in the *period* gene around the Threonine-Glycine repeat region convey the species-specific characteristics of the love song rhythms (Yu *et al.*, 1987; Wheeler *et al.*, 1991). Could it possibly be that a corresponding situation exists in the case of the *nonA* gene, i.e. that the *nonA* gene conveys species-specific courtship song information?

The *per* gene not only conveys species-specific lovesong information, but also can transmit species-specific patterns of locomotor activity (Petersen *et al.*, 1988). *per* has therefore been called a "speciation" gene by Coyne (1992) because it transmits species-specific information for at least two characters which may help to maintain species boundaries, and consequently may have important implications for the speciation process itself. How the love song may prevent interspecific matings has already been covered earlier, but locomotor activity patterns could be equally, or even more important as a pre-mating interspecific barrier, by restricting flies' movements to different phases of the day. In the case of the *nonA* gene, could it be that species-specific characteristics of the love song, particularly of the sound pulses, which are

clearly affected in *nonA^{diss}* mutants, are controlled by this single gene? A simple way to answer this question is to isolate the *nonA* gene from a different species and transform it into *D.melanogaster* flies carrying the *nonA^{diss}* mutation or a deletion of the locus (*nonA⁽⁻⁾*).

So far *per* is the only behavioural gene that has restored species-specific characters from a donor species to the host (Petersen *et al.*, 1988; Wheeler *et al.*, 1991). In both phenotypes discussed above, locomotor activity and courtship song, the transfer of the behaviour by the donor species *per* gene was complete. This in itself is surprising, because speciation theories predict that changes in behaviour, or in any other phenotype such as morphology, occur by the accumulation of many small changes at multiple loci (Coyne, 1992). The *per* gene clearly contravenes this elementary law of traditional speciation thinking. Could it be that the *nonA* gene also presents us with an example of such a complete transfer of behaviour from one species to another? The precedents have been set by the *per* gene, and my project seeks to examine whether the *nonA* gene is similarly implicated in the control of species-specific behaviour. To investigate this possibility, I have cloned the *nonA* gene from a species whose courtship song was as different from that of *D.melanogaster* as possible. All the species of the *D.melanogaster* subgroup have pulse songs that are very similar to those of *D.melanogaster*, so clearly these were not appropriate for my purposes (Cowling and Burnet, 1981). A great deal of work has been done on the courtship song of *D.virilis* (Hoikkala *et al.*, 1982; Hoikkala *et al.*, 1987; Hoikkala and Lumme, 1984), which is very different from that of *D.melanogaster*. Given the use of *D.virilis* as a comparative model system, particularly in developmental studies (Michael *et al.*, 1990; Heberlein and Rubin, 1990; O'Neil and Belote, 1992), and given the availability of genomic libraries of this species, I decided to begin my investigations by attempting to identify a *nonA* homologue in *D.virilis*. By examining the rescue of the visual optomotor behaviour and love song characteristics in a number of independent transformants carrying the *D.virilis nonA* homologue, I was able to assess whether in fact the *nonA* gene can be considered to be a "speciation gene" in the same way as *per*.

CHAPTER 2

MATERIALS & METHODS

BASIC PROCEDURES

Basic techniques such as the purification of nucleic acids with phenol-chlorophorm, ethanol precipitation, and DNA quantitation, were performed following Sambrook *et al.* (1989).

NUCLEIC ACIDS ISOLATION

Genomic DNA isolation:

TE = 10 mM Tris-HCl (pH 8.0); 1 mM EDTA (pH 8.0).

Fly genomic DNA to be used in PCR reactions was prepared using the method of Gloor and Engels (1990). Single males were ground in 50 µl of buffer (10 mM Tris-HCl pH 8.2; 1 mM EDTA; 25 mM NaCl, 20 µg/ml Proteinase K), incubated at 37 °C for 30 min, and then heated at $\geq 92^{\circ}\text{C}$ for at least 2 min. 1 µl of DNA was sufficient for each 10 µl of PCR reaction.

When better quality DNA was required (i.e. for Southern blotting), 10-50 flies were homogenized in 200-500 µl of a solution containing 0.1 M Tris-HCl (pH 9.0); 0.1 M EDTA; 1% SDS and 0.5-1% DEPC (added directly before use) and heated at 70°C for 20-30 min. The samples were incubated on ice for 30 min after adding 14 µl of Potassium Acetate for each 100 µl of homogenate. The DNA was recovered in the supernatant after spinning at 4°C for 30 min. 5 µl of RNaseA were added and the samples incubated at 37° for at least 30 min. The DNA was purified with phenol/chloroform and precipitated by adding 0.5 volumes of isopropanol and spinning for 15 min at room temperature. The DNA was then washed with 70% ethanol, dried and resuspended in 10-20 µl of TE.

Plasmid DNA isolation:

Small scale:

Solution A = 50 mM Tris-HCl (pH 8.0); 4% Triton X-100; 2.5 M LiCl; 62.5 mM EDTA.

The method by Goode and Feinstein (1992) was used to rapidly prepare plasmid DNA to be digested. 15 ml of an overnight bacterial culture was centrifuged at 13000 rpm for 30 s, and the pellet resuspended in 100 µl of solution A. The DNA was then extracted in 100 µl of phenol/chloroform and precipitated in 200 µl of cold 100% ethanol. The pellet was washed with 1 ml of 70% ethanol, dried, and then resuspended in 10 µl of TE + 100 µg/ml RNaseA.

Midi prep:

Up to 200 µg DNA from a high copy plasmid vector were isolated using the Hybaid RecoveryTM Plasmid Midi Prep Kit. This kit usually provided high quality DNA to be used as a template in sequencing reactions.

Large scale:

Large scale plasmid DNA was obtained using the MagicTM Megapreps DNA Purification System by Promega or the Qiagen Plasmid Maxi Kit. Culture volumes and procedures followed the instructions provided by each kit.

Total RNA extractions:

Total RNA was extracted from *D. virilis* using the RNAzol^{TMB} method from Biotecx Laboratories, Inc., Houston, Texas. For long term storage the RNA was kept in isopropanol at -80°C in order to minimize degradation.

POLYMERASE CHAIN REACTION (PCR)

7.4X PCR mix = 167 µl of 2 M Tris-HCl

83 µl of 1 M Ammonium Sulphate

33.5 µl of 1 M MgCl₂

3.6 µl of 2-mercaptoethanol

3.4 µl of 10 mM EDTA (pH 8)

75 µl of 100 mM dATP

75 µl of 100 mM dCTP

75 µl of 100 mM dGTP

75 µl of 100 mM dTTP

85 µl of 10mg/ml DNase free BSA.

The Polymerase Chain Reaction (PCR) was performed according to Jeffreys *et al.* (1988). Each 10 µl of reaction contained 1.35 µl of 7.4x PCR mix, 1 µl of each 10 µM primer, 0.15 µl of Taq Polymerase (5 units/µl), and H₂O to volume. Taq Polymerases from different sources were used, such as Amplitaq Polymerase from Perkin-Elmer-Cetus, Taq polymerase from Amersham, or Thermostable DNA Polymerase from Advanced Biotechnologies. The reactions were carried out in a Perkin-Elmer-Cetus thermal cycler or in a Techne Programmable PHC-1 and PHC-3. The thermal profile mostly used was 1 min at 95°C (denaturation); 1 min at 65°C (annealing); 1 min at 70°C (extention); for 30-40 cycles.

RT-PCR

Reverse transcription from total RNA was performed by adding to 1 µg of RNA to the following "RT mix":

- 4 µl 5X First Strand Buffer (250 mM Tris-Hcl pH 8.3; 375 mM KCl; 15 mM MgCl)
- 0.8 µl dNTPs (25 mM each)
- 1 µl specific primer (10 µlM)
- 1 µl Reverse Transcriptase (M-MLV RT from GIBCOBRL; 200 U/µl)
- 0.6 µl rRNasin (Ribonuclease Inhibitor from Promega; 38 U/µl)

- DEPC H₂O to a total volume of 20 µl.

DEPC H₂O= 1 ml of DEPC (diethyl pyrocarbonate, Sigma) was added to each litre of H₂O and left overnight before autoclaving.

The "RT mix" was incubated at room temperature for 15 min, then at 37°C for 65 min, and finally heated at 95°C for 10 min. 5 µl of the "RT mix" were then used as a template DNA for each PCR reaction which contained 2.7 µl of 7.4X PCR mix; 1 µl of 5' primer (10 µM); 1 µl of 3' primer (10 µM); 0.3 µl of Taq Polymerase; and 11 µl of H₂O.

GEL ELECTROPHORESIS AND DNA RECOVERY FROM GEL BANDS

1X TAE = 0.04 M Tris-acetate; 0.001 M EDTA.

6X Gel-loading buffer = 0.25% bromophenol blue; 0.25% xylene cyanol FF; 30% glycerol in water.

Agarose gel electrophoresis followed the methods of Sambrook *et al.* (1989). Gels were usually prepared at concentrations 0.7%-1.5%, according to the size of the DNA fragments to be resolved. 0.5 mg/ml of ethidium bromide was added to the melted gel before pouring it, and once set, the gel was run in 1X TAE buffer at voltages which ranged between 20-120 V. DNA markers, to be run alongside the samples, were either λ cut with *Hind*III (for fragments between 2 and 23 Kb), or φ X174 cut with *Hae*III (for fragments from 70 bp to 1.3 Kb). 1/10 volume of gel loading buffer was added to increase the density of each DNA sample and to indicate the progression of the electrophoresis. The DNA was then visualized on the gel by UV fluorescence and photographed using a video imaging system (GDS2000, UVP International). For a better quality photograph, a Polaroid MP-4 camera and Kodak negative films (T-max Professional 4052) were used. The films were then processed with Kodak LX24 developer, FX40 fixer and HX40 hardener.

DNA was recovered from agarose bands using the QIAEX and QIAEXII gel purification kits (QIAGEN Inc.). In case of very large bands (weight above 250 mg), the DNA was extracted by placing the band in a dialysis bag filled with 2 ml of TE buffer and performing electrophoresis on the bag submerged in TAE. The DNA migrated from the gel band to the TE buffer which was then transferred in microcentrifuge tubes and purified with one phenol/chlorophorm and one chlorophorm/isoamylalcohol extraction. The DNA was then ethanol-precipitated and resuspended in an appropriate volume of H₂O.

DNA SEQUENCING

Direct sequencing of PCR products:

Double stranded direct sequencing of PCR products was performed according to Bachmann *et al.* (1990) using the dideoxy chain-termination method of Sanger *et al.* (1977). The amplified DNA to be sequenced was usually recovered from an agarose gel band using the QIAEX gel purification kit (QIAGEN). The template was annealed to the specific primer by mixing 6 µl of DNA with 2 µl of sequencing buffer 5X (200 mM Tris-HCl pH 7.5; 100 mM MgCl₂; 250 mM NaCl), 1.5 µl of the specific primer (10 µM) and 0.5 µl of 10% Np40. The tube containing the mixture was then boiled for 3 min and rapidly cooled in dry ice. The sequencing reactions were performed using the Sequenase Version 2.0 Kit from United States Biochemicals and then run on a 5% denaturing polyacrilamide gel using a Sequi-Gen (Bio-Rad) nucleic acid sequencing cell. At the end of the run the gel was fixed for 15 min in a solution 10% acetic acid and 10% methanol and placed on a sheet of 3MM paper. The gel was then vacuum dried at 80°C and placed in an autoradiography cassette with a Fuji RX100 X-ray film.

Sequencing double stranded plasmid DNA:

1.5-3 µg of plasmid DNA were sequenced using the T7 Sequencing Kit from Pharmacia. The annealing of primer to the double stranded DNA was performed

following the standard protocol suggested in the kit instruction booklet.

SOUTHERN BLOTTING OF DNA (Southern, 1975).

Denaturing solution = 1.5M NaCl; 0.5 M NaOH.

Neutralizing solution = 1.5 NaCl; 1M Tris-HCl (pH 8.0).

20X SSC = 3 M NaCl; 0.3 M Na₃ citrate (pH 7.0).

The gel to be blotted was photographed after placing a ruler next to it in order to have a record of the position of the DNA markers. When large fragments of DNA (>10 Kb) needed to be transferred, the gel was depurinated for 10 min in a solution of 0.25 M HCl, rinsed with water, then placed in the denaturing solution for 30 min, and finally neutralized for 30 min. The gel was then placed on the blotting apparatus (as described in Sambrook *et al.*, 1989) and surrounded by cling film. A precut, prewet (3X SSC) piece of Hybond-N filter (Amersham International) was carefully placed on the gel, followed by one wet (3X SSC), one dry pieces of 3MM paper, and a thick pad of absorbent paper towels. DNA transfer was usually allowed to occur overnight. The filter was then rinsed in 3X SSC, air dried and fixed with a 45 s UV light exposure on a transilluminator.

Preparation of radiolabeled DNA probe:

10X Random priming buffer (RPB) = 500 mM Tris-HCl (pH 6.6); 100 mM MgCl₂; 10 mM DTT; 500 µg/ml BSA (DNase and RNase free, Sigma).

3 dNTPs = 20 mM each of dATP, dGTP and dTTP in H₂O.

A modification of the method of primer extension with random primers by Keller *et al.* (1989) was used to generate nucleic-acid probes with specific activities of 10⁹ cpm/µg. 50-200 ng of double-stranded DNA in 10 µl of H₂O were heat-denatured

by boiling for 5 min and placed immediately on ice for at least 5 min. The denatured DNA was mixed with:

- 5 µl of 10X RPB
- 5 µl of p(dN)₆ (2 mg/ml, Pharmacia)
- 1 µl of 3 dNTPs
- 2 µl of [$\alpha^{32}\text{P}$]dCTP (>3000 Ci/mmol)
- 1 µl of Klenow (5 U/µl)
- 26 µl of H₂O.

The reaction was incubated at 37°C for 30 min, and then terminated with 5 µl of 0.5M EDTA. The unincorporated nucleotides were removed by ethanol precipitation adding 1 µl of tRNA (1 µg/µl), 1/10 volume of 3 M NaAc pH 5.2, and 2 volumes of ethanol 100%. The probe was denatured by heating at 100°C for 5 min and immediately cooled on ice for 5 min before mixing to the filter-hybridization solution (see later).

Filters hybridization with ^{32}P probes:

Prehybridization:

At least 2 hours before adding the radiolabeled probe, the blotted filters were incubated at 65°C in a pre-hybridization solution composed of:

- 1 mg/ml ficoll
- 1 mg/ml polyvinyl pyrrolidone
- 5mg/ml BSA fraction V
- 0.5% SDS
- 2 mg/ml Marvel dried milk
- 6X SSC.

Hybridization:

The pre-hybridization buffer was changed with some identical fresh solution (pre-heated at 65°C) just before adding the radiolabeled probe. The filter was then incubated at 65°C overnight.

Filters washing:

Hybridized filters were washed briefly at room temperature in a solution of 4X SSC, 0.1% SDS. This wash was repeated at 65°C for 30 min. Then a series of washes of increasing stringency were performed (2X SSC, 0.1% SDS; 1X SSC, 0.1% SDS; 0.5X SSC, 0.1% SDS; 0.1X SSC, 0.1% SDS), until the filters emitted almost background levels of radioactivity. Every wash was done for 30 min at 65°C. Filters were then air dried (not completely if the filter had to be re-probed), wrapped in SaranWrap, and placed in an autoradiography cassette containing an intensifying screen and a Fuji RX100 X-ray film.

CLONING PROCEDURES

Dephosphorylation of linearized plasmid DNA:**1) Using calf intestinal alkaline phosphatase (CIP):**

10X CIP buffer = 0.5 M Tris-HCl (pH 9.0); 10 mM MgCl₂; 1 mM ZnCl₂; 10 mM spermidine.

10X STE = 100 mM Tris-HCl (pH 8.0); 1 M NaCl; 10 mM EDTA.

0.01 u/pmol ends of CIP (GIBCO BRL) were added to a 48 µl reaction containing the linearized DNA and 5 µl of 10X CIP buffer. To dephosphorylate protruding 5' termini, the mixture was incubated at 37 °C for 30 min, and for an additional 30 min after adding a second aliquot of the enzyme. To dephosphorylate DNA with blunt ends or recessed 5' termini, the reaction was incubated for 15 min at 37°C, and 15 min at 56°C, and the two incubations repeated after adding a second

aliquot of CIP. The reaction was then stopped by adding 10 µl of 10X STE, and 5 µl of 10% SDS and heating at 68°C for 15 min. The DNA was then extracted twice with phenol/chlorophorm, and twice with chorophorm.

2) Using Shrimp Alkaline Phosphatase (SAP):

10x SAP buffer = 200 mM Tris-HCl (pH 8.0); 100 mM MgCl₂.

Dephosphorylations were performed following the instructions accompanying the SAP enzyme (United States Biochemicals). The reactions were prepared in 20 µl total volume containing 2 µl of 10X SAP buffer, linearized DNA, and suggested amounts of SAP. The mixture was incubated at 37°C for 1 hour, and the reaction stopped by heating at 65°C for 15 min. After heating, the enzyme was completely inactive, and the DNA was ready to use in ligation reactions.

Ligation reactions:

Ligations were performed following Sambrook *et al.* (1989). Insert and vector DNA were usually mixed in 3:1 ratios in a 15 µl total volume of reaction containing T4 Ligase Buffer (from GIBCO BRL) at working concentration (1X), and 1 unit of T4 DNA ligase (GIBCO BRL). The ligation reactions were incubated overnight at 16°C.

Plasmid transformation (chemical method):

Preparation of competent cells:

XL1-Blue = *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^qΔ M15 TN10* (Tet^r)].

LB (Luria-Bertaini) medium (or Luria broth) = 1% (w/v) Bacto-tryptone (Difco); 0.5% (w/v) Bacto-yeast extract (Difco); 1% (w/v) NaCl.

Tfb I = 30 mM KAc; 100 mM RbCl₂; 10 mM CaCl₂; 50 mM MnCl₂; 15% (v/v) glycerol; pH 5.8.

TfbII = 10 mM MOPS; 75 mM CaCl₂; 10 mM RbCl₂; 15% (v/v) glycerol; pH 6.5.

Bacterial cells (*E. coli* strain XL1-Blue, Stratagene) were grown at 37°C in 500 ml of LB broth. When the early log phase was reached ($OD_{550} = 0.5-0.55$), the culture was chilled on ice for 5 min, and then spun in a Sorvall centrifuge (6,000 r.p.m. for 5 min at 4°C). The cells were resuspended in 200 ml of TfbI (2/5 original volume), left on ice for 5 min and centrifuged again at 6,000 r.p.m. for 5 min (at 4°C). The pellet was resuspended in 8ml of TfbII (1/25 volume), and left on ice for 15 min. Suitable volumes of cells (usually 100-200 μ l) were then aliquoted into microcentrifuge tubes, using a prechilled pipette and tubes. The tubes were snap frozen in dry ice and stored at -80°C.

Transformation procedure:

LB agar plates = LB medium + 15 g/liter Bacto-agar.

100 μ l of competent cells were added to 10 μ l of ligation and left on ice for 20 min. The DNA-cells mixture was heat shocked at 42°C for 45 s and returned to ice for 1-2 min. The transformed cells were grown at 37° for 1 hour in 1ml of Luria broth and then plated (usually in 100 μ l aliquots) on LB agar plates containing appropriate antibiotics. When blue-white selection was required, the plates were made adding X-gal and IPTG according to Sambrook *et al.* (1989).

Transformation by electroporation:

Cells for electroporation:

XL1-Blue cells were harvested in Luria broth at an A_{600} of about 0.55. A series of washes was performed in order to concentrate the cells into a dense suspension and to wash away the large amounts of electrolytes in the culture medium. The washing solutions used (either sterilized H₂O or 10% glycerol) were pre-chilled at 4°C. The washing protocol was the following:

1 L of culture $\Rightarrow$ pellet (4,000 r.p.m., 15 min, 4°C) $\Rightarrow$ resuspend in 1 L of H₂O
 $\Rightarrow$ pellet (4,000 r.p.m., 15 min, 4°C) $\Rightarrow$ resuspend in 1/2 L of H₂O

⇒ pellet (4,000 r.p.m., 15 min, 4°C) ⇒ resuspend in 20 ml of 10% glycerol

⇒ pellet (4,000 r.p.m., 15 min, 4°C) ⇒ resuspend in 2 ml of 10% glycerol.

The cells were then aliquoted (50 µl in each tube) and stored at -80°C.

Electroporation procedure:

Ligations were ethanol precipitated and resuspended in 5 µl of TE or H₂O. 1-5 µl of transforming DNA was mixed with 40 µl of washed cells (prepared as above). The mixture was transferred on a cuvette chilled on ice and placed between the electrodes of the electroporation unit. An electrical pulse of 15 KV was delivered, and the cells immediately resuspended in 1 ml of Luria broth and gently shaken at 37°C for 1 hour before plating them onto LB agar plates (see above).

Bacterial cracking:

This method was used to rapidly identify recombinant clones when blue-white selection was not possible. 1ml of overnight culture from a single colony was spun down and resuspended in 200 µl of 1X gel loading buffer. After adding 200 µl of phenol/chloroform, the cells were vortexed and spun for 2 min. 20 µl of the top layer were then loaded on an agarose gel (usually 1%) and run at 100-120 V. Insert-containing plasmid DNA runs considerably slower than non-recombinant and could be easily detected.

GENOMIC DNA LIBRARY SCREENING

A *D. virilis* genomic DNA library constructed by Ron Blackman (Harvard University, Cambridge, MA) was kindly provided by John Belote (Syracuse University, NY). The library had been prepared with genomic DNA from adult flies partially digested with the enzyme *Mbo*I. Size selected fragments (15-20 Kb) had been inserted

between the *Bam* HI sites of the lambda (λ) EMBL3 phage (Frischauf *et al.*, 1983) and packaged in vitro.

Plating and titration of the library:

Preparation of plating bacteria:

A culture of the Y1090 bacterial strain was inoculated overnight at 37°C in 200 μ l of Luria broth containing 0.2% maltose and ampicillin (50 mg). The cells were then spun down and resuspended in 10 mM MgSO₄ to an appropriate density (OD₆₀₀ = 0.5).

Plating bacteriophage λ :

SM buffer = 50mM Tris-HCl (pH 7.5); 100 mM NaCl; 8 mM MgSO₄; 0.01% gelatin.

LB top agarose = 1% (w/v) Bacto-tryptone; 0.5% (w/v) Bacto-yeast extract; 0.5% (w/v); 0.6% (w/v) agarose.

Serial dilutions of bacteriophages were prepared in 50 μ l of SM buffer. 0.2 ml of plating bacteria were then added to each dilution, and incubated at 37°C for 15 min, allowing the cells to adsorb the bacteriophage particles. 3 ml of molten (47°C) LB top agarose were added, and each preparation plated on pre-set LB agar plates. The plates were incubated at 37°C until plaques appeared. The number of plaques given by each dilution was an indication of the library titre.

***In situ* hybridization of bacteriophage λ plaques:**

Preparation of the filters:

Denaturing solution = 1.5 M NaCl; 0.5 M NaOH.

Neutralizing solution = 1.5 M NaCl; 1 M Tris-HCl (pH 7.6).

Appropriate dilutions of the library were plated in a total number of 30 LB plates and incubated at 37°C overnight. When the plaques became almost confluent, the plates were cooled at 4° for 1 hour. One marked Hybond-N (Amersham) circular

filter was placed on each plate for 5 min. The filters were lifted carefully from the plates and placed for 10 min in the denaturing solution, for 1 min in 1 M Tris-HCl (pH 7.6), for 10 min the neutralizing solution, and then rinsed in 3X SSC. When completely dry (usually after 1 hour at room temperature), the filters were sandwiched in SaranWrap, and the DNA was fixed by UV exposure for 45 s.

Prehybridization, hybridization and washing of the filters:

These techniques were performed as already described above for Southern blotting filters.

Isolation of positives and rescreening:

The position of the possible positives on the developed film was located on the plates. A "plug" of top agarose containing the plaque of interest was then lifted using a pipette tip, and placed in a microcentrifuge tube containing 500 μ l of SM buffer and 20 μ l of chlorophorm. The suspension was left at room temperature for 4-6 hours to allow the bacteriophage particles to diffuse out of the agarose. Several dilutions of the suspension were then plated and rescreened as described above. The procedure of positive picking, replating and rescreening was continued until all the plaques on a plate were positives.

Phage lysate preparation: plate method:

For each plate, about 1×10^5 pfu of bacteriophage (or 60-100 μ l of a phage "plug" left to elute 1 hour at room temperature in 1 ml of SM buffer) were plated and incubated at 37°C until the plaques became confluent. The plates were then overlaid with 2-3 ml of SM buffer, and the top agar scraped with a spatula into a high speed 30 ml centrifuge tube (Corex). After incubating at room temperature for 30 min, the suspension was centrifuged at 10,000 r.p.m. for 10 min at 4°C. The supernatant was transferred in a fresh tube, and 0.3% of chloroform was added for long term storage.

Bacteriophage DNA isolation:

Bacteriophage λ DNA was extracted from 10 ml of phage lysate using the MagicTM Lambda Preps DNA Purification System.

P-ELEMENT TRANSFORMATION OF *DROSOPHILA*

P element mediated transformation was performed according to Spradling and Rubin (1982). The transformation vector used was *pw8* (Klemenz *et al.*, 1987) which carries *white*⁺ as a marker. Embryos microinjected were either *w; $\Delta 2-3$ Sb/TM6* (Robertson *et al.*, 1988), or *w1118* (Pastink *et al.*, 1987). When using the latter injectees, transposase was provided by co-transformation of a helper plasmid *PUC_{hs} π $\Delta 2-3$* (a gift from J.M. Dura).

Embryos collection and dechorionation:

4 -7 day old females were encouraged to lay eggs by placing them on a fruit juice-based food. Eggs were collected at intervals no longer than 60 minutes to assure injection prior the time of pole cell formation. Embryos dechoronation was performed manually by placing eggs over a piece of double-sided sticky tape on a microscope slide and gently rolling them until the chorion was ruptured. The dechorionated embryos were then aligned on one side of the tape with the posterior end just hanging off the tape. About 10 minutes after dechoronation the eggs were covered with Voltalef oil (Grade 10S) to prevent further dessication.

Microinjection of embryos:

Microinjection was performed using a Nikon TMS inverted microscope and a Narishige micromanipulator. Needles were prepared by pulling microcapillary tubes (GC100TF-15, Clark Electromedical Instruments) using a Flaming Brown micropipette puller (Model P80/PC, Sutter Instruments Co.) and placed on a microelectrode holder from World Precision Instruments, Inc. A constant flow of DNA from the needle was obtained thanks to a PV830 Pneumatic Pico Pump (World Precision Instruments, Inc). Embryos were injected, according to Spradling and Rubin (1982), into their posterior tip, and the amount of DNA introduced depended on the degree of dessication of the embryo.

Post-injection care of embryos:

All improperly injected or aged embryos were destroyed, and the coverslip with the remaining embryos was placed on a petri dish containing standard fly food dyed with a red food colouring. The red colour allowed easy detection of the developing embryos. More oil was added over the embryos to protect them from dessication. The plate was incubated at 18°C. Hatched larvae were removed from the plates after a couple of days, and placed in numbers of 5-10 into sugar food vials. Further development was accelerated by placing the larvae at 25°C. Adults (G₀ generation) were individually mated to *w¹¹¹⁸* flies, and their progeny (G1 generation) screened for transformants (red-eyed flies).

BEHAVIOUR

FLY STOCKS and BALANCERS

D. melanogaster:

Canton S strain had been kept in the laboratory for several years.

*w*¹¹¹⁸: see Lindsley and Zimm, 1992.

w; *e* Δ 2-3 *Sb/TM6*: transposase providing strain (Laski *et al.*, 1986; Robertson *et al.*, 1988).

P[(ry) 235R11]D1: Jones and Rubin (1990). Transformant strain carrying an 11Kb *EcoRI* fragment (homozygous on the third chromosome) rescuing all *nonA* mutants (Jones and Rubin, 1990; Rendahl *et al.*, 1992). Donated by Jeff Hall.

T(1;4)9e2-10/FM7: Stanewsky *et al.* (1993). Strain with reciprocal translocation T(1;4), uncovering the essential *l(1)i19e* locus and the *nonA* gene.

P[(ry) 211S12]45 and *P[(ry) 211S12]36*: Stanewsky *et al.* (1993). Transgenic strains containing only the first two exons of the *nonA* gene.

P[(ry) 235R11-SacII]5: Stanewsky *et al.* (1993). Transformant stock carrying the *nonA* gene with a frame-shift mutation introduced at the *SacII* site in the first exon.

nonA^{diss}/*FM7a*; *ry*⁵⁰⁶/*MKRS*: donated by Jeff Hall.

w, *FM7*: balancer for the first chromosome (see Lindsley and Zimm, 1992).

w; *TM2/MKRS*: balancer for the third chromosome (see Lindsley and Zimm, 1992).

w; *T 2,3 ap*^[*xa*]/*SM5 TM3*: balancer for both second and third chromosome (see Lindsley and Zimm, 1992).

All balancer stocks were obtained from American or European *Drosophila* stock centres.

D. virilis: (strain 1411, Finland) was donated by Anneli Hoikkala, University of Oulu, Finland.

STOCK MAINTENANCE

Fly stocks were reared on sugar/agar medium (46.3 g of sucrose, 46.3 g of yeast, 7.1 g of agar and 2 g of Nipagen in 1 L of water) in one-third pint milk bottles or glass vials (10 cm height x 2.2cm diameter) The flies were kept in temperature-controlled rooms at either 18°C or 25 (± 1)°C.

RECORDING AND ANALYSING *DROSOPHILA* LOVESONGS

Recording *D.melanogaster* songs:

Newly eclosed males were isolated and placed individually into vials at 25°C. At the age of 2-4 days their song was recorded by placing one male and one or two virgin females (maximum 1 day old), into a small rectangular (1.5cm x 1.1cm large, 0.6cm high) perspex cell with a nylon mesh at the bottom. The flies were inserted in the cell through a small hole cut in one side of the cell. The cell was placed, mesh down, on top of a condenser microphone (custom-made in Leicester) contained in a foam box. The box could be closed to function as an anechoic chamber, and the movements of the flies observed through a magnifying view lens on top of the box. A light inside the box assured visibility. The box also contained an amplifier and filters, and the song was recorded onto a Revox 477 reel to reel tape recorder at 7.5 ips. The chamber temperature was monitored continuously with a digital thermometer, and recordings were carried out between 25-26°C.

Recording *D.virilis* songs:

Males and females were collected immediately after eclosion and kept in sugar food vials at 25°C. Males were kept individually, while females were grouped in numbers of 5 per vial. At about 15 days of age, one male and one female were placed in a perspex cell identical to that used for *D.melanogaster* flies, but which had a bigger hole for inserting the larger *virilis* flies. The recording conditions and equipment were the same as those used for *D.melanogaster*.

Each male's song was recorded for 10-15 minutes using a Revox reel-to-reel tape recorder. Long silences were common, and were avoided by switching the tape recorder off when males were not singing. The recording was filtered below 1000Hz and above 100Hz using an EF5-03 LP/HP filter (Barr & Stroud, Cambridge, U.K.), and then played into a CED 1401 (Cambridge Electronic Design) analogue-to-digital converter, in order to get a graphic representation of the song signal. The signals

could then be displayed on the screen of an IBM-PC 486, and stored on the hard disk. The SPIKE2 software developed by CED was used for primary analysis of the signals. This program allows simultaneous display and processing of waveform data and events.

Song characters analysed:

Cycles per pulse (CPP):

Figure 2.1 shows a small portion of one of the songs recorded, as it was displayed using the Spike2 LEVELS script. The LEVELS script can be used to detect events such as peaks and troughs from waveform channels. Channel 1 contains the song waveform. Above the waveform there is a second channel (channel 2) displaying the song events (shown as bars) which in this case are the peaks and the troughs found in the song's pulses. Peaks and troughs were automatically marked by the Spike2 program. For the peaks, a positive threshold rising above the song's background noise was first selected so that Spike2 marked every single event above it. All the peak events were stored and displayed on channel 2 above the song waveform. To automatically mark all the troughs in the song, a negative threshold was chosen below the background. All the troughs were recorded and displayed in channel 2 together with all the peaks. At this point the song was manually edited by moving along the song waveform and removing or adding events from channel 2 as needed. The editing mainly involved removing events caused by loud noises, and adding or removing some events from pulses in which the exact number of peaks and/or troughs had not been marked properly in the initial automatic analysis.

When the manual editing was completed, the events stored in channel 2 were imported into the Microsoft Excel 5.0 program. Each datum was expressed as the real time, in milliseconds, in which the event had occurred. A macro written by Martin Couchman permitted calculation of the number of cycles in each pulse and the regression for CPP during song bursts. For each pulse, the number of cycles was calculated as: $((\text{peaks} + \text{troughs}) - 1) / 2$. Instructions were given that allowed the

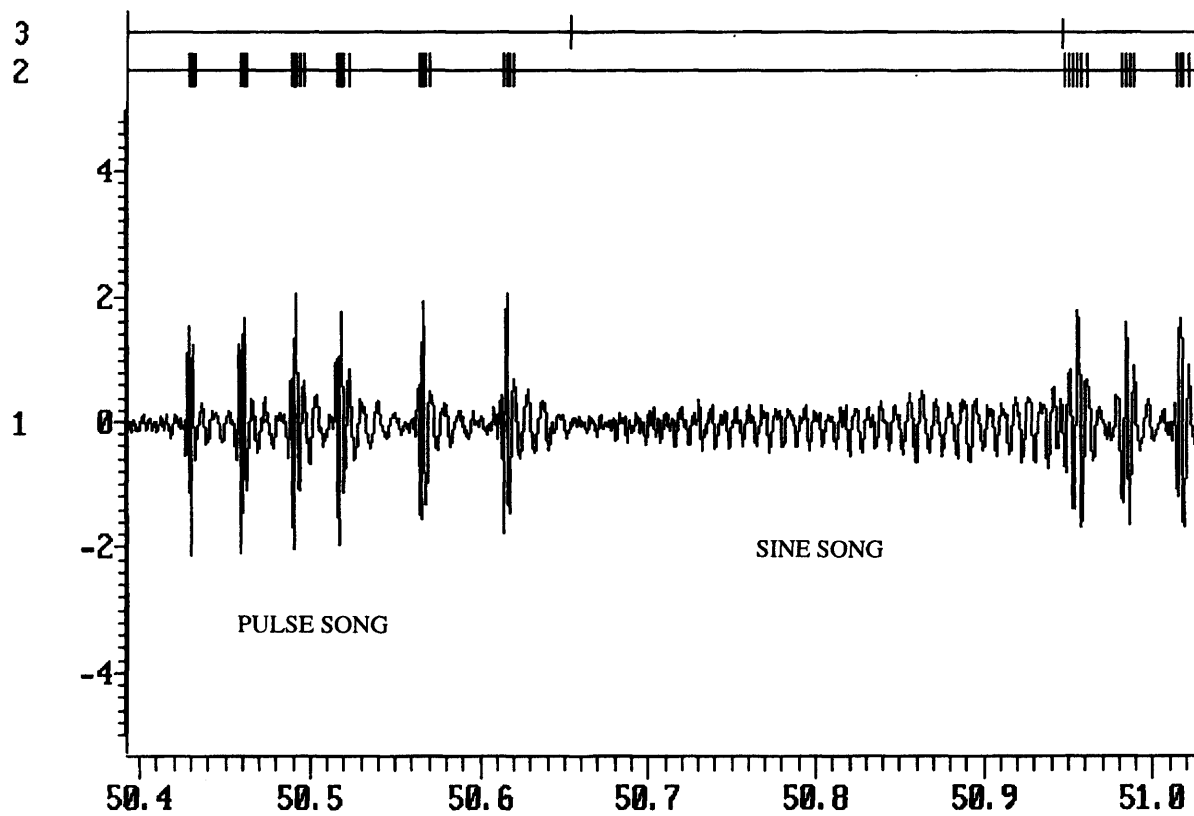


Figure 2.1. Portion of song displayed using the Spike2 LEVELS program. Channel 1 (song waveform), channel 2 (with marks for pulse peaks and troughs) and channel 3 (with marks for sine song) are indicated. Time scale in seconds.

computer to recognize events of the same pulse from events of the next pulse and pulses of one train from pulses of the next train. The minimum gap allowed between a peak and a trough in each pulse was 0.6 ms. The maximum gap between a peak and a trough in a pulse was set at 5 ms: if an event occurred more than 5 ms from the previous one then it was not be considered part of the same pulse. The minimum gap from one pulse to the next pulse of the same train was 10 ms, so an event more than 10 ms apart from the previous one was considered part of the next pulse. The minimum distance between two pulse trains was set at 70 ms: if an event was more than 70 ms apart from the previous one, then it was considered part of the first pulse in the next pulse train. CPPs were calculated only for trains that contained at least 5 pulses or more, shorter trains were not included in the analysis. In order to make sure that the macro was properly calculating the CPPs, an entire song was manually analysed and the results compared with those obtained with the macro.

For each song, the macro calculated the average CPP scores for the pulses found in each consecutive position along the pulse trains, i.e. it calculated the average CPP for all the pulses in the first position , the second position, the third position and so on. A least squares regression was performed using these averages. Averages calculated for pulse positions containing less than 5 data points were not included in the regression analysis.

Pulse frequency (PF):

From the CPP analysis, the macro calculated the average frequency of each pulse by dividing the time between the first and the last event in the pulse by the number of cycles in it and converting into Hz. A least square regression was performed using the average frequency values of pulses at consecutive positions in the pulse trains. Again, only averages calculated from at least 5 frequency values were used for the regression.

Pulse song (PS):

The macro automatically calculated the total amount, in seconds, of pulse song. For each train of pulses the time between the first event in the first pulse and the last

event in the last pulse, gave the total length, in seconds, of the train. The total amount of pulse song was calculated as the sum of all the pulse trains' lengths in the song. The length of short pulse trains (containing less than 5 pulses) was not included in the calculation.

Sine song (SS):

In every song, the beginning and the end of each burst of sine song was marked manually and the events stored on channel 3 displayed above the peaks and troughs channel (figure 2.1). The events in channel 3 were imported into Microsoft Excel 5.0 together with the events from channel 2. The macro automatically calculated the length, in seconds, of each sine song burst by using the time points that marked the beginning and the end of the burst. The total amount of sine song was calculated for each song by summing the lengths of all the sine song bursts found.

Sine song frequency (SSF):

Ten sine song bursts were selected from the terminal part of each song (which is usually more rich in sine song) in order to calculate the frequency. The Fast Fourier analysis of Spike2 was used to obtain the frequency of each selected sine song burst. The SSF for each song was calculated as the mean of the 10 selected sine song burst frequencies.

Sine song proportion (SSP):

This character was calculated automatically as the amount of sine song over the total amount of song recorded ($SS/PS+SS$).

Interpulse intervals (IPIs):

The mean IPI of each song was calculated as: $PS/(Npulses-Nbursts)$ where PS is the total amount, in ms, of pulse song recorded, Npulses is the total number of pulses in the song and Nbursts is the number of bursts.

An example of a typical page output obtained after applying the macro to an individual song is shown in figure 2.2. Both PF and CPP regressions are accompanied

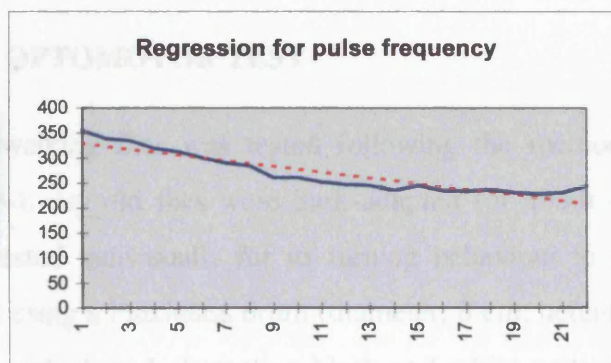
Regression of pulse frequency using data in first 22 columns:

slope (m)	-5.7751	335.6106	y intercept (b)
s. error of m	0.5429	7.1299	s. error of b
coeff. of determ	0.8498	16.1541	s. error of y. estimate
F statistic	113.1718	20.0000	degrees of freedom
regression sum o	29532.9174	5219.1297	residual sum of squares

$$y = -5.7751x + 335.6106$$

Data file: Waveform

min. gap between peaks/troughs	0.0006
max. gap between peaks/troughs	0.0050
min. gap for next pulse	0.0100
min. gap for next burst	0.0700
min. pulses per burst to record:-	5



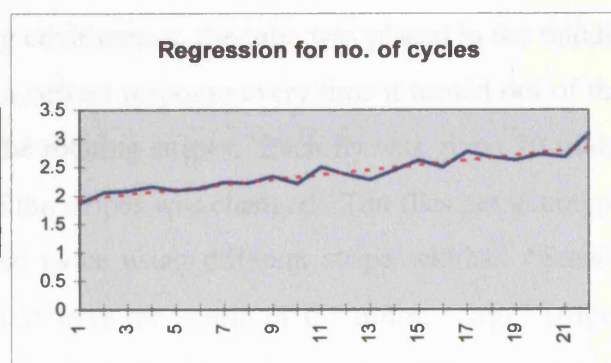
Regression of cycles per pulse using data in first 22 columns:

slope (m)	0.0426	1.8998	y intercept (b)
s. error of m	0.0035	0.0457	s. error of b
coeff. of determ	0.8826	0.1034	s. error of y. estimate
F statistic	150.3363	20.0000	degrees of freedom
regression sum o	1.6087	0.2140	residual sum of squares

$$y = 0.0426x + 1.8998$$

Data file: Waveform

min. gap between peaks/troughs	0.0006
max. gap between peaks/troughs	0.0050
min. gap for next pulse	0.0100
min. gap for next burst	0.0700
min. pulses per burst to record:-	5



Sine song: 12.1510
Pulse song: 24.6780
sine/total 0.3299
Length of pulse song excldes short bursts

Figure 2.2.

Typical page output of the macro written by Martin Couchman.

by the appropriate F-ratio which indicates the significance of the fit. Their slope and y-axes intercepts are calculated and illustrated graphically. The amount of sine song and pulse song and the proportion of sine song are also indicated at the bottom of the page.

WALKING OPTOMOTOR TEST

The optomotor response of walking flies was tested following the method described by Burnet *et al.* (1968). 3-8 day old flies were dark-adapted for about 4 hours in food vials. Each fly was tested individually for its turning behaviour in a moving visual field. This was created using a Plexiglass drum (diameter, 8 cm; height, 9 cm) rotating around a vertical axis, which had alternating black and white vertical stripes.

Testing *D.melanogaster*:

Each fly was placed in the middle arm of a T-shaped glass tube. This arm was painted black, so that the fly was forced to walk out of this opaque tube, into a choice point where it could turn into the right or the left arm. To test whether the turning behaviour corresponded to the moving environment, the tube was placed in the middle of the rotating drum. A fly produced a correct response every time it turned out of the opaque arm in the same direction of the rotating stripes. Each fly was given 20 trials, and each time the rotating direction of the stripes was changed. Ten flies per genotype were tested. The test was performed twice using different stripe widths. "Small" stripes subtended a small angle of 13.3° from the centre of the drum; "large" stripes subtended a larger angle of 36° . The drum was constantly rotated at 46 rpm. A desk lamp (60 watts) was placed above the drum in order to illuminate in a uniform way the centre of the cylindre. All tests were performed at room temperature (20-25°).

Testing *D.virilis*:

Given the larger size of *D.virilis* flies, they could not move freely in the T-shaped tube used for testing *D.melanogaster* flies. Each *D.virilis* fly was therefore placed in an empty glass vial which was positioned at the centre of the rotating

cylinder. The test was started once the fly had reached the top of the tube (given its negative geotropism). From there, the movement of the flies were scored as the drum was rotating. A correct optomotor response was obtained each time a fly performed at least five rotations in the same direction as the moving stripes. Again, two sets of stripes ("small" and "large", see above) were used, and each fly tested 20 times in each set of stripes.

STATISTICAL ANALYSES

All statistical analyses, such as one or two-way ANOVAs, χ^2 tests, planned comparisons and Newman-Keuls *a posteriori* tests (Winer, 1971), were performed using the Statistica Microsoft for Windows Software (StatSoft, Inc., 2325 East 13th Street, Tulsa, Oklahoma 74104, U.S.A.). Multivariate analysis was performed using the Systat Software (Version 7.0 for Windows, 1997 by SPSS Inc.).

COMPUTER ANALYSIS OF DNA SEQUENCES

All the computer analysis of DNA sequences were performed using the programs of the GCG package for molecular biology (Version 7; University of Wisconsin Genetics Computer Group, Madison, WI, USA; Devereux *et al.*, 1984). The various programs used will be mentioned in the relevant chapter.

CHAPTER 3

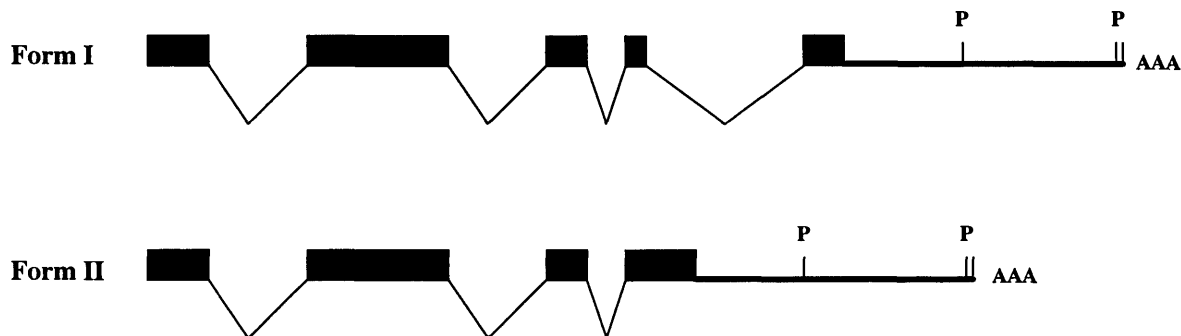
The *D.virilis nonA* gene

INTRODUCTION

The *nonA* gene was cloned and sequenced by Jones and Rubin (1990). By *P* element-mediated transformation with different genomic clones from a 20 kb interval within the 14C1,2 region of the X chromosome, they identified a *XbaI-EcoRI* 9Kb genomic DNA fragment necessary and sufficient to rescue all the *nonA* mutant phenotypes (see also Rendahl *et al.*, 1992). Simultaneously, Besser *et al.* (1990) cloned the *BJ6* gene whose product they found associated with transcriptionally active puffs on polytene chromosomes of *Drosophila* larvae. When they compared their sequence of the *BJ6* gene to that of *nonA*, they found that the two sequences were identical.

The *nonA* gene is organized in five exons separated by four introns (figure 3.1). Based on their cDNA data, Jones and Rubin (1990) suggested the existence of two differently spliced forms of the NONA protein generated by alternative splicing of the last intron. In the form I of NONA the fourth intron is spliced out and the fifth exon encodes the last 35 amino acids of the protein. In form II, the fourth intron encodes the last 33 amino acids of *nonA*, so form I and form II NONAs only differ in their carboxy terminus. However, the results of successive studies have argued against the existence of a form II protein. Firstly, only form I has been shown to carry all the *nonA* related biological functions (Jones and Rubin, 1990; Rendhal *et al.*, 1992). Secondly, attempts to identify the form II NONA in Western blots have never succeeded (Stanewsky *et al.*, 1993).

Figure 3.1. Schematic diagram of the two putative forms originating from the *nonA* gene (Jones and Rubin, 1990). Black boxes indicate the exons. Introns are indicated by thin lines. The thick line after the last exon represents the 3' untranslated region with the two polyadenylation signals and the polyA tail (Besser *et al.*, 1990).



Two mRNA transcripts of ~3 and ~4 Kb have been identified in Northern blots using probes for the *nonA* gene. The existence of two polyadenylation sites, the first found 673 nucleotides after the stop codon, the second about 1Kb after the first, could explain the occurrence of the shorter transcript as a result of alternative termination (Besser *et al.*, 1990). A third 2.8 Kb transcript was sporadically resolved from the 3 Kb species by Jones and Rubin (1990) and Rendahl *et al.* (1992).

The 700 amino acid putative NONA protein is basic and has an N-terminus rich in glycine, asparagine and glutamine, often arranged in repeats. The C-terminal domain is rich in charged amino acids, while the last 100 carboxy terminal residues again contain clusters of glycine and asparagine. Hydrophobicity analysis indicated that the protein does not contain a membrane-spanning segment and is unlikely to have an N-terminal signal sequence (Jones and Rubin, 1990; Besser *et al.*, 1990).

As mentioned in chapter 1, the central domain of the protein has two tandemly repeated motifs of 80 amino acids each, bearing homology to the RNA recognition motifs (RRMs). These motifs are found in a large family of more than 30 RNA-binding proteins including snRNPs, hnRNPs, and several proteins encoded by

Drosophila genes such as *embryonic lethal abnormal visual system (elav)* (Robinow *et al.*, 1988), *Sex-lethal (Sxl)* (Bell *et al.*, 1988) and *transformer-2 (tra-2)* (Amrein *et al.*, 1988; Bandziulis *et al.*, 1989; Besser *et al.*, 1990; Rendhal *et al.*, 1992). The most conserved RRM feature is an eight-amino acid sequence called RNP1 (Kenan *et al.*, 1991). Some RNA-binding proteins contain an additional hexapeptide known as RNP2 (Kenan *et al.*, 1991; Dong *et al.*, 1993). The RNP1 and RNP2 elements contain aromatic and basic residues, implicated in interactions between proteins and single-stranded nucleic acids (Kenan *et al.*, 1991).

Each of the two RRM found in NONA possesses an RNP1 and an RNP2 (Rendahl *et al.*, 1992; 1996). The more conserved RRM of NONA, situated at amino acid position 295-373 from the N-terminus (Jones and Rubin, 1990), contains 6 out of 6 amino acids of the RNP-2 consensus sequence (Bandziulis *et al.*, 1989) and 4 out of 8 amino acids of the RNP-1 consensus sequence (Rendahl *et al.*, 1992). Figure 3.2, taken and corrected from Rendahl *et al.* (1992) shows an alignment of the two RRM sequences (NONA 1 and NONA 2 in figure 3.2) of NONA with those of the yeast PABP (polyadenylate binding protein, Adam *et al.*, 1986; Sachs *et al.*, 1986), SXL (Bell *et al.*, 1988) and TRA-2 (Amrein *et al.*, 1988).

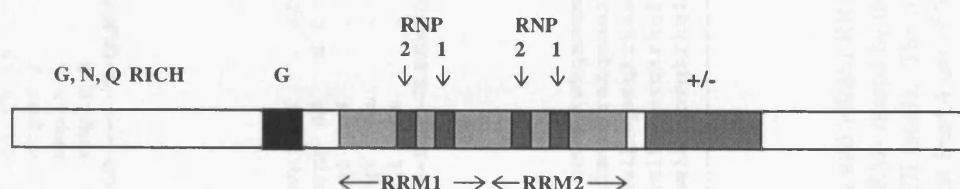
Auxiliary domains, hypothesized to mediate RNA-protein or protein-protein interactions, are also commonly found in RRM family members. These include homopolymeric glycine rich regions (Dreyfuss *et al.*, 1993), arginine-serine repeats (RS; Zamore *et al.*, 1992) and arginine-glycine-glycine repeats (RGG; Kiledjian and Dreyfuss, 1992). A glycine-rich region, present in many hnRNPs polypeptides and thought to have protein-protein annealing activity (Dreyfuss *et al.*, 1993), is found near the N-terminus of *nonA* (see figure 3.3).

Figure 3.2.

Sequence similarities involving NONA's putative RNA binding motifs. The two RNA-binding domains of NONA (1 and 2) are aligned with the four domains of the yeast polyadenylate binding protein (PABP), the two domains of *Drosophila* Sex-lethal (SXL) and the single domain of transformer-2 (TRA-2). Below each block of comparisons is the "consensus" sequence, referring to the larger family of more than 30 proteins (Bandziulis *et al.*, 1989). The RNP-1 and RNP-2 motifs are indicated in bold. Within the NONA RNPs, conserved aromatic amino acids of RNP-1 and RNP-2 are underlined, and characteristic hydrophobic residues of the RNP domains (Kenan, *et al.*, 1991) are marked (* or *), indicating the presence of hydrophobic amino acids in one or both NONA domains.

			RNP-2			
PABP	1	SQSVENSSAS	LYVGDL -EPS	VSEAHLYDIF	SP-IGSVSSI	RVCRDAITKT
PABP	2	PSLRKKGSGN	IFIKNL -HPD	IDNKALYDTF	SV-FGDILSS	KAITDEN-GK
PABP	3	LEETKAHYTN	LYVKNI -NSE	TTDEQFQELF	AK-FGPIVSA	SLEKDAD-GK
PABP	4	EKMAKYQGVN	LFVKNL -DDS	VDDEKLEEEF	AP-YGTITSA	KVMRTEN-GK
SXL	1	MNDPRASNTN	LIVNYL -PQD	MTDRELYALF	RA-IGPINTC	RIMRDYKTGY
SXL	2	PGGESIKDTN	LYVTNL -PRT	ITDDQLDTIF	GK-YGSIVQK	NILRDKLTQR
TRA-2		SREHPQASRC	IGVFGL -NTN	TSQHKVRELF	NK-YGPIERI	QMVIDAQTGR
NONA	1	TETKFSGRNR	LYVGNL -TND	ITDDELREMF	KP-YGEISEI	F-----SNL
NONA	2	-----NATI	LRVSNL -TPF	VSNELLYKSF	EI-FGPIERA	SITVDDR-GK
			* *	* **	* *	
CONSENSUS			LFVGNL	E L F	FG I	K
			IYIKG	D	Y V	R
			RNP-1			
PABP	1		S- LGYAYVNF	NDHEAGRKAI	EQLNYTPIKG	RLCRIMWSQR
PABP	2		S- KGFGFVHF	EEEGAAKEAI	DALNGMLLNG	QEIYVAPHLS
PABP	3		L- KGFGFVNY	EKHEDAVKAV	EALNDSELNG	EKLYVGRAQK
PABP	4		S- KGFGFVCF	STPEEATKAI	TEKNQQIVAG	KPLYVAIAQR
SXL	1		S- FGYAFVDF	TSEMDSQRAI	KVLNGITVRN	KRLKV-----
SXL	2		P- RGVAFVRY	NKREEAQEAI	SALNNVIPEG	GSQPLSVRLA
TRA-2			S- RGFCFIYF	EKLSDARAAK	DSCSGIEVDG	RRIRVDFSIT
NONA	1		D- KNFTFLKV	DYHPNAEKAK	RALDGSMRKG	RQLRVRFAP-
NONA	2		H- MGEGIVEF	AKKSSASACL	RMCNEKCFFL	TAS-LRPCLV
			* * *	* *		* * *
CONSENSUS			KGFGFVXF	A	L G	
			R YA Y		I	

Figure 3.3. Schematic representation of the principal protein domains in NONA. Amino acids are represented by a one-letter code. The two RRM domains are indicated, and the darker bars within represent the RNP2 and RNP1 motifs. +/- indicates the charged domain. The black box shows the position of the Glycine repeat stretch.



The more recent alignment of NONA with novel sequences from the database (Rendahl *et al.*, 1996) resulted in a larger region of homology than previously identified. A conserved region of about 320 amino acids (called DBHS, Dong *et al.*, 1993) was found between NONA, the human splicing factor PSF (Patton *et al.*, 1993), the HeLa protein p54^{nrb} (Dong *et al.*, 1993), the murine NON-POU containing Octomer binding protein or NONO (Yang *et al.*, 1993), and *nAhomo* (M. Palazzolo, unpublished material). Figure 3.4, from Rendahl *et al.* (1996), presents an alignment between NONA and the four DBHS domain-containing proteins isolated so far. Table 3.1 reports the percent identity between each pair of proteins (Rendahl *et al.*, 1996).

Table 3.1. Percent identities among NONA and related DBHS containing proteins. These were derived from the 311-amino acid region that spans the two RRMs and the adjacent region of relatively high charge density. From Rendahl *et al.* (1996)

	dr-NONA	mu-NONO	hu-p54 ^{nrb}	dr-nAhomo
hu-PSF	44	71	72	42
dr-NONA		42	43	80
mu-NONO			99	41
hu-p54 ^{nrb}				41

This chapter describes the identification, the cloning and sequencing of the *D.virilis nonA* homologue. The complete coding sequence of the gene was obtained, plus an additional ~2 kb of 5' upstream promoter sequence. Comparisons between the *D.virilis* and the *D.melanogaster* sequences was performed. Given the approximate 40 (Moriyama, 1987) to 60 (Beverley and Wilson, 1984) million years divergence between *D.virilis* and *D.melanogaster*, sequence conservation should highlight regions of possible functional importance.

METHODS

***D.virilis* genomic library.**

A *D.virilis* genomic library constructed by Ron Blackman was kindly donated by John Belote. Genomic DNA had been isolated from adult flies and partially digested with *Mbo*I. Size-selected fragments (15-20Kb) had then been inserted into the *Bam*HI sites of the EMBL3 phage vector and packaged *in vitro*.

***D.virilis* PCR fragment**

A small fragment of the *D.virilis nonA* gene was amplified by PCR after trying various combinations of primers all based on the second exon sequence of the *D.melanogaster nonA* gene (Jones and Rubin, 1990). Table 3.2 shows the sequence and position (relative to the *D.melanogaster nonA* sequence) of these primers. Only primers 5'C and 3'A worked after lowering the annealing temperature of the PCR to 50°C. PCR reactions were performed as described in chapter 2.

Table 3.2. Primers based on the second exon sequence of the *D.melanogaster nonA* gene. Their position based on the *nonA* sequence from Jones and Rubin (1990) is indicated.

PRIMER	SEQUENCE	POSITION
5'A	5'-AAGAACAATGAACCGGCAAC-3'	3617-3636
5'B	5'-CGATGAACAGCACCAATATG-3'	3948-3967
5'C	5'-CGCGAGATGTTCAAGCCATA-3'	4163-4182
3'A	5'-GCCCTCTCGATGGGACCAAAA-3'	4422-4403
3'B	5'-TCCTGGTTGAAGTCGGGCAT-3'	4644-4625
3'C	5'-CCTTGACGCAACAGTTCAGT-3'	4839-4820

The *D.virilis* 261 bp PCR fragment was gel purified using the QIAEX gel purification kit (QUIAGEN Inc.) and directly sequenced with the dideoxy-chain-termination method (see chapter 2). Analysis of the sequence strongly suggested that the fragment was part of the putative *D. virilis nonA* second exon. Figure 3.5 shows an alignment between the *D.virilis* PCR fragment and the homologous fragment from the *D.melanogaster nonA* sequence. 20 ng of the *D.virilis* PCR product were labelled with 32 P by random priming (as described in chapter 2), and used as a probe to screen the *D.virilis* genomic library.

Library screen.

The *D.virilis* library was plated using the *E.coli* strain Y1090. About 3×10^5 plaques were screened at high stringency (0.1% SDS; 0.5X SSC at 65°C) isolating 8 putative positive clones. Only two of them proved to be positive after successive rescreening. DNA was obtained from one of them (**2.2b1**) using the Promega Magic Lambda Preps DNA purification System Kit (see chapter 2).

Figure 3.5. Sequence of the *D.virilis* PCR fragment aligned to the homologous fragment of the *D.melanogaster nonA* gene. The primers used to amplify the *D.virilis* DNA fragment are indicated by the arrows. The nucleotide numbers shown for the *D.melanogaster* sequence are equivalent to those of the *nonA* sequence in Jones and Rubin (1990).

Subcloning and sequencing.

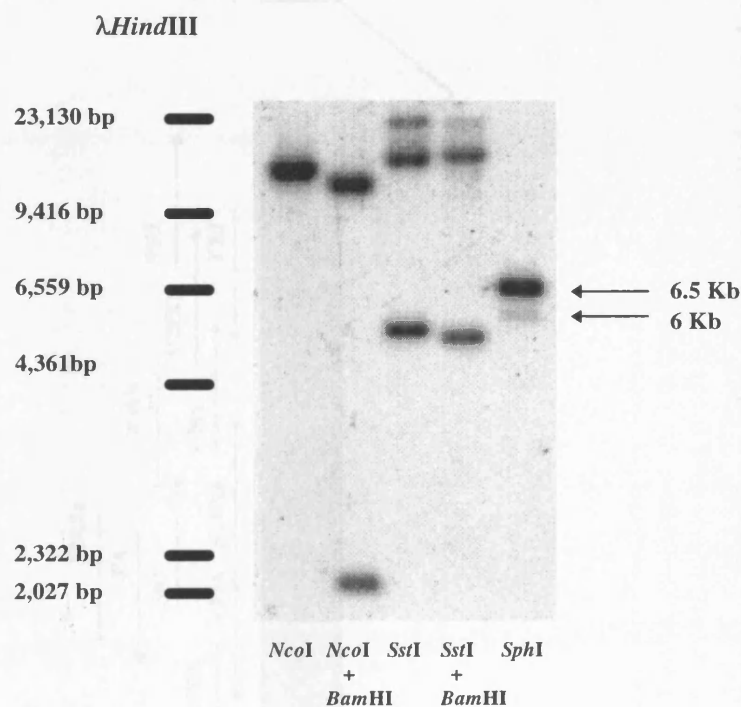
Clone **2.2b1** was cut into 5 fragments using the restriction enzyme *SphI* (data not shown), run on a 0.8% agarose gel and Southern blotted. When a 9Kb *XbaI-EcoRI* DNA fragment containing the *D.melanogaster nonA* gene (Jones and Rubin, 1990) was used as a probe, only two of the five *SphI* fragments hybridized to it. Of these two, the larger one (~6.5 Kb) hybridized very strongly to the *melanogaster* gene, while the smaller (~6 Kb) produced only a very faint band in the autoradiograph (figure 3.6). Both fragments were subcloned into the vector pUC 18 (Vieira and Messing, 1982) at the *SphI* site and were respectively named pUC39 (the 6.5 Kb fragment in pUC18) and pUC7 (the 6 Kb fragment in pUC18).

The Universal and Reverse primers, annealing respectively to the 5' and 3' side of the pUC18 polylinker, were used to sequence both ends of the pUC7 and pUC39 clones. It was found that pUC7 contained the first half of the putative *D.virilis nonA* gene, up to roughly a third of the second exon. The orientation of this insert was the same as the pUC vector. The larger pUC39 clone contained the remaining half of the gene in opposite orientation to pUC. New primers were designed based on the sequence obtained from the second exon in order to obtain its complete sequence. These primers were designed in a way to allow sequencing of overlapping regions of the exon thus assuring repeated sequencings of both DNA strands (see figure 3.7).

A strategy of cutting pUC7 and pUC39 with different enzymes, and religating the vector containing a shorter fragment of the insert was used in an attempt to obtain sequence from any of the other exons. A restriction map of the putative *D.virilis nonA* gene was also generated (figure 3.8). After cutting pUC7 with *EcoRI*, three fragments were produced with the approximate sizes of 1 Kb, 1.5 Kb and 6.5 Kb. The 6.5 Kb fragment was composed of the 3 Kb pUC+a 3.5 Kb insert. This 6.5 Kb fragment was recircularized using the T4 Ligase enzyme (for ligation techniques see chapter 2) and used to transform *E.coli* cells (XL1-Blue, Stratagene) as described in chapter 2. Plasmid DNA was isolated as described in chapter 2. The Reverse primer, annealing in 3' of the pUC polylinker was used to sequence the 3' end of the 3.5 Kb insert in pUC. The sequence obtained was compared to the *D.melanogaster nonA* sequence and

Figure 3.6.

Southern blot with the DNA from the positive λ clone **2.2b1** hybridized to the ~9 Kb *D.melanogaster nonA* gene. Each lane represents different digestions of the DNA and the enzymes used are indicated. The ~6.5 Kb and ~6 Kb *SphI* fragments which were subcloned into pUC18 are indicated by the arrows. Markers were not included in the blot, so the approximate positions of the λ *HindIII* fragments were drawn.



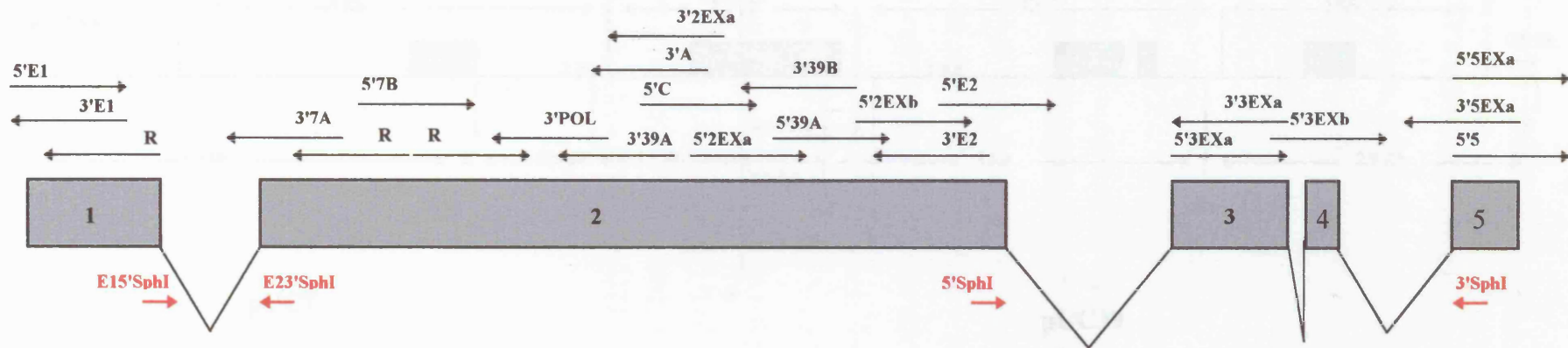


Figure 3.7.

Scheme illustrating the overlapping DNA fragments sequenced in order to obtain the complete sequence of the *D. virilis nonA* gene. Exons are indicated by the filled boxes and appropriately numbered. Introns (which are not to scale) are indicated by the thin lines. Each long arrow represents the average 200 bp sequenced using each primer. The primer name is written on top of each arrow. Small red arrows under the exons represent the position of the primers (whose names are indicated) used to amplify and sequence *D. virilis nonA* cDNA fragments in order to establish the exact exon-intron boundaries.

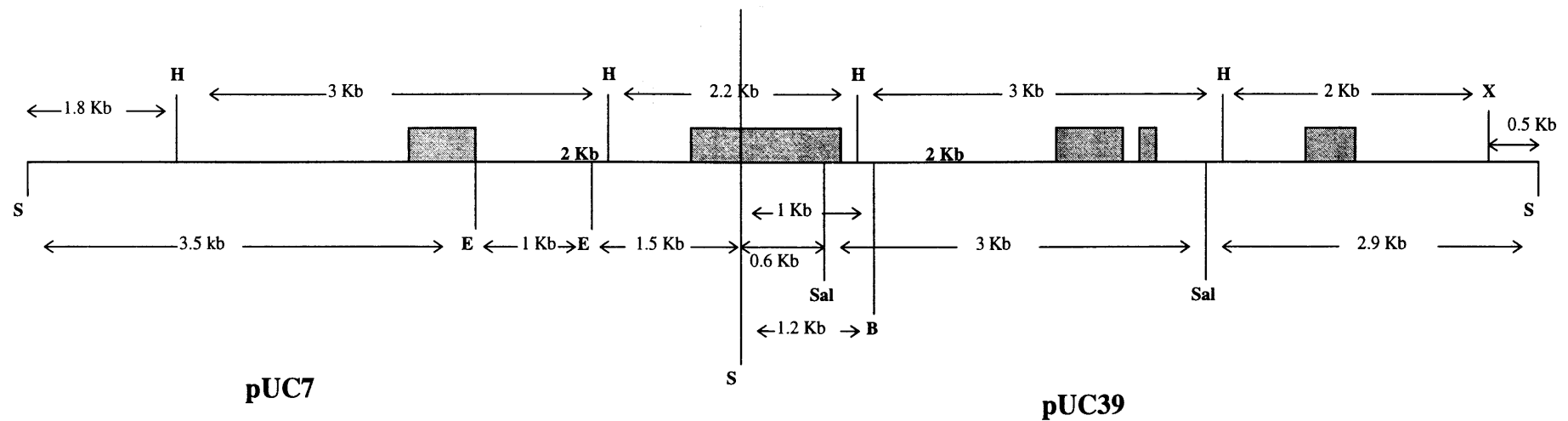


Figure 3.8.

Restriction map of the putative *D. virilis nonA* gene. Filled boxes represent the exons. Horizontal lines indicate the introns and 5' and 3' untranslated sequences. The long vertical line in the middle of the figure represents the *SphI* site which divides the gene into the two 6 Kb and 6.5 Kb fragments cloned into pUC (pUC 7 and pUC 39, see text). The sizes of the first two introns is indicated in bold.

S = *SphI*

E = *EcoRI*

H = *HindIII*

B = *BamHI*

X = *XbaI*

Sal = *SalI*

showed homology in a region towards the end of the *nonA* gene first exon. New primers extending the sequence in 5' and 3' were designed in order to complete the sequence of the putative *D.virilis nonA* first exon.

Sequence from the *D.virilis nonA* putative fifth exon was obtained simply by using a primer based on the sequence from the beginning of exon five of *D.melanogaster nonA* at position 7106-7125 (Jones and Rubin, 1990) (**5'5:** 5'-AGGACTCCTTCGCTTTTGA-3'). The sequence of this exon was subsequently confirmed several times as described below.

To obtain sequence from the third and fourth exons without having to sequence also the introns, I opted for an approach which employed RT-PCR (see chapter 2). RNA was isolated from *D.virilis* flies using the RNAzol[™]B method (see chapter 2) and reverse transcribed using a 3' primer binding to to the middle of the fifth exon (**3'SphI:** 3' GCGGCATGCTGCCGCCTTGGTTCATAT 5', see figure 3.7). A 510 bp cDNA fragment was then amplified using the **3'SphI** primer and a 5' primer annealing to the end of the second exon(**5'SphI:** 5' TATGCGATGCAGATGGAGTATGCTCGCT 3'). Both primers had an *SphI* site (underlined) allowing cloning of the 510 bp PCR product into PUC18. The Universal and Reverse Primers, in conjunction with the primers used for the cDNA amplification, were used to sequence the third, the fourth, and part of the fifth exons. These sequences were then reconfirmed after designing new primers annealing to the beginning and the end of the third exon (**5'3EXa**, **3'3EXa**, and **5'3EXb**, see figure 3.7, and appendix 3.1) and to the beginning and the end of the fifth exon (**5'5EXa**, **3'5EXa**, figure 3.7, appendix 3.1) and using the genomic clone as a template DNA (**pUC39**). Comparison between the genomic and the cDNA sequences allowed the localization of the intron-exon boundaries from the third exon to the end of the gene.

The exact intron-exon boundaries between the first and the second exons were identified using again RT-PCR. *D.virilis* RNA was isolated, and cDNA produced using a 3' primer annealing to the the beginning of the second exon (**E23'SphI:** 5'-TATGCGATGCTGCCCTGCGCTTGAAGTTGTG-3',see figure 3.7). A 225 bp cDNA fragment was then amplified using the above 3' primer and a 5' primer annealing to the

end of the first exon (**E15'*Sph*I**: 5'-TATGCATGCGCGAACAAGGGTGGATTT-3', see figure 3.7). Thanks to the *Sph*I sites designed in each primer (underlined in the primer sequence), the PCR product could be cloned into the *Sph*I site of pUC18. The Universal, Reverse, **E15'*Sph*I** and **E23'*Sph*I** primers were used to sequence the PCR fragment. The sequence obtained from the cDNA was then compared to the one obtained from the genomic clone in order to identify the exact intron-exon junctions.

Sequence of the *D.virilis nonA* promoter region.

A ~ 2 Kb DNA sequence upstream of the putative first codon of the *D.virilis nonA* gene was obtained by automatic sequencing performed at the University of Padova, Italy. The **pUC7** clone was used as the genomic DNA template and the sequence was progressively extended starting from the beginning of the putative *D.virilis nonA* first exon.

Computer analyses.

All sequence analyses were performed with the programs of the GCG package for molecular biology (Version 7; University of Wisconsin Genetics Computer Group, Madison, WI, USA; Devereux *et al.*, 1984).

Sequence alignments were generated either with the Bestfit or the Gap GCG programs. The first was used for aligning protein sequences, while the second produced the DNA sequences alignments. Dot matrix analysis were carried out with the Compare and Dotplot programs. The secondary structure of proteins was predicted using the programs Peplot, Peptidestructure, and Plotstructure. Codon usage patterns were studied with the Codonfrequency program. Sequence simplicity analysis was performed by J. Hancock using SIMPLE34 (Hancock and Armstrong, 1994), a modified version of Tautz *et al.*'s (1986) SIMPLE program. The TF

SEARCH program (from the Internet, <http://pdap1.trc.rwcp.or.jp/htbin/nph-tf>; Akiyama, 1995) was used to search for functional motifs in the 5' regulatory region of *nonA*. Possible protein motifs were investigated with the PSITE program (Version 1, Solovyev and Kolchanov, 1994) from the Internet (<http://dot.imgen.bcm.tmc.edu:9331/pssprediction/Help/psite.html>).

RESULTS & DISCUSSION

Structure of the *D.virilis nonA* gene

The five-exon-four-intron organization found in the *D.melanogaster nonA* gene is conserved in the *D.virilis* homologue (figure 3.8). The intron-exon boundaries are also conserved, as revealed by the cDNA and genomic DNA comparison analysis (see methods). Although the sequence of the four introns of the *D.virilis nonA* gene was not obtained, the approximate lengths of the first and second introns were calculated. This was done by measuring the length of PCR products obtained using primers annealing to the exon boundaries. Both introns are considerably longer than their *D.melanogaster* counterparts. The first, being ~2 kb long, is about twice the size of the first *D.melanogaster nonA* intron, the second, also measuring ~2 Kb, is about four times the size of the *D.melanogaster*'s. The exact length of the *D.virilis nonA* intron three was not directly calculated, however this intron appears to be as short as that of *D.melanogaster*. In fact a sequencing reaction using a primer annealing to the end of the third exon (**5'3Exb**, see figure 3.7) produced a sequence of the last 40 bases of intron three. The upstream intron sequence could not be read being too close to the **5'3Exb** primer. Assuming that on average the first ~30 bases from a given sequencing primer cannot be read, I calculated the length of the third intron to be no more than 70 bases. The length of the fourth intron is unknown.

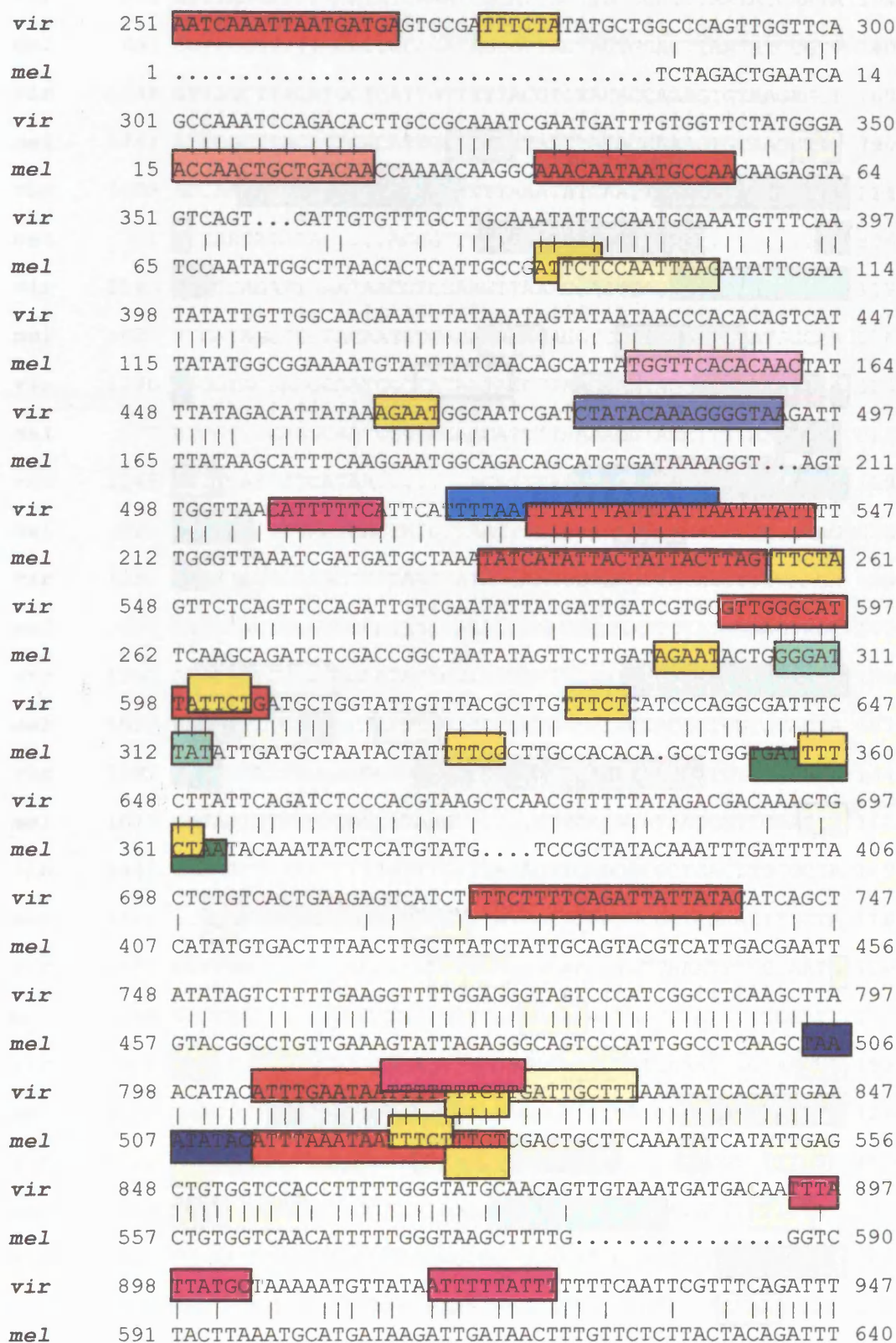
Jones and Rubin (1990) hypothesized the existence of a second splice form of NONA which would result from the translation of the beginning of the fourth intron and would add an extra 33 amino acids at the end of the fourth exon (see figure 3.1, FormII). Although the existence of this form II of NONA is doubtful (see introduction), I decided to check whether translation of the beginning of the fourth exon in *D.virilis nonA* could produce an extra stretch of amino acids as in *D.melanogaster*. The sequence of the first ~200 bases of the fourth *D.virilis nonA* intron was therefore obtained and translation in the three possible frames produced by the program Map of the GCG packet. The sequence is presented in appendix 3.2. 14 amino acids can be translated from the beginning of the fourth intron of *D.virilis* in the same reading frame of the end of the fourth exon. This stretch of amino acids is half the size of that predicted in *D.melanogaster* and the sequence, apart for the first two residues, is completely different. Although RNA studies, possibly via RT-PCR, could be done in order to check whether a second splice form with this extra 14 amino acid sequence at the end is actually produced from *nonA* in *D.virilis*, it seems unlikely.

The 5' regulatory region of *nonA*.

Figure 3.9 presents an alignment of the ~2 Kb 5' DNA sequence containing the putative promoter region of the *D.virilis nonA* gene aligned with the corresponding region of *D.melanogaster nonA*. The alignment was performed with the Gap program of the GCG packet (see methods), which finds the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. According to this alignment the two sequences share a 55.9 % identity.

A dot matrix analysis (window: 21; stringency:14) was performed between the *D.virilis* and the *D.melanogaster* sequences to graphically highlight the regions of homology (figure 3.10a). The areas of good conservation between the two sequences should reflect functional importance, given the long time of divergence (40 to 60 Myr) that separates *D.virilis* and *D.melanogaster* (Beverley and Wilson, 1984; Moriyama, 1987). The dotplot reveals seven regions of moderate-to-good conservation which are

Figure 3.9. Alignment of the 5' regulatory regions of *D. virilis* and *D. melanogaster nonA*. The putative transcription factor binding sites are highlighted in colour.



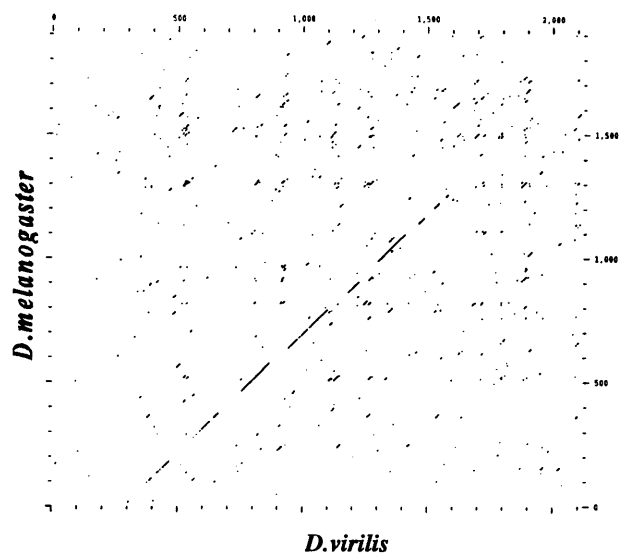
vir 948 GATAGAACCAATTGTACGGCAGGTTTTCTTGCCATCGGGGCATTGGAT 997
 mel 641 CATTGAACCTTTAGTACGACAAGTTTCTTGGCCATTGGAATGATTGGAT 690
 vir 998 GTTTAGGATTTACATATCAAATCGCATTGCTTGCCGATTTAATATCAATA 1047
 mel 691 GCTTGGGTTTTACATTCCAGATAGCCTTACTAGTGGACTTAATATCCGTT 740
 vir 1048 GTTGGCTTACATGCTCATTGTTTTTACGTCTACACCAAAGTGTAAGAGCT 1097
 mel 741 ATTGGCTTACATTCACATTGCTTCTATATTTATACCAAAGTGTAAGTTCT 790
 vir 1098 AGCAAGTATTAATATATGTTATTTAAATATCAATTTAACAATTTACTTT 1147
 mel 791 TSCAATATATA....ACCGCTTTCAAACATTAAATATTA.....TT 826
 vir 1148 TTTTCAGATTGAATAACGTCGAAGTTAAAGGACTGACGGTGTTATGGCAA 1197
 mel 827 TTCATAGCCTCTACAATGTGGAAAGAAAGGCTTTCTGTTCTATGGCAA 876
 vir 1198 GTGGTACGAGGCAATCGCTATAATACTTAAGAAGTAGGACATAAAAAACA 1247
 mel 877 GTCGTTTCGAGGCAATCGTTACAACATTTTAAAGGTATATTTTAAAGTTT 926
 vir 1248 TTCTATTTTTCATAA.....ACATTTATTTATGTTTTTGTGTTTAG 1290
 mel 927 TTGCCATCTAAGCACTATGTTAAGTTTACCAATTCTTATACTTTTCAG 976
 vir 1291 ATAGAATCGAAGCTCATAACTATATGAATCGCCAGCTATACCTTGCAACA 1340
 mel 977 GTCGCACTGAGTCCCACAACATATGAATCGGCAATTGTATCTAGCTACC 1026
 vir 1341 ATATTCTTTCTGCAATATTGTTCTTGTTCGACCACTCTGTCTACTA 1390
 mel 1027 ATATTCTTCTCGCTATTCTGTTTTGTTACCACTACCCTGTCTACTA 1076
 vir 1391 CGTTGTGTTTGAACCGTAAGCTGATTATTATTATTAGGTGTTAAAATT 1440
 mel 1077 TATAGTTTTTGCTGCTGTAAG.....TTAATATATAATCGTTGAATTT 1120
 vir 1441 GCAAGCTAATATCTTTGTTTGATTACAGTTGAAGGCGCTCACTTGCCTA 1490
 mel 1121 ..CGCAGTAATACAACCTTCTTTATAGCTGAAGGCGCTCACATTTGCTA 1167
 vir 1491 CGTTGGCAACTCTGGAATGTTTTCGCCGCAAGCTGTTAAATTTCCAATA 1540
 mel 1168 CTCTAAGTGCTTTTCATTTTGTGCGAAGAAACTGATGTATCTTCCGATT 1217
 vir 1541 GAAATGTTTCTTAAATACATTAAAAAGGGATTCTATGAAATTGGTAAGCG 1590
 mel 1218 GAAGTCTTATAAAACGGTTATTAAGAGGATGTCATGAATAGGTAATTG 1267
 vir 1591 AAATATTATTAAAGTATACACCTATTATTAT....CGCGTTTCTT 1636
 mel 1268 TGTAAAGAATGTGAATGCATAAAATATATTATATATGATTCTCTAT 1317
 vir 1637 TTAAATATAGATTGCCTAAGAGTGCTGGACATCCCACTTCAAAAACCAT 1686
 mel 1318 TGACTTGCAGATTGTATTCAAATCAAAGATGTTTCACATCAAGAAAGGGC 1367

vir 1687 ATTTTTTATTCA.....TCGAAATTC 1707
 mel 1368 TTTCCTGATGCACAACAACAGAAAATCAACGTGACAGTGATAAAATTA 1417
 vir 1708 AAAATTAATAATTTTCTATATATACTCCGAGTTTAAACTAAATAATGT 1757
 mel 1418 CAACTTTATAATCCTTGCCATATCTGAGATATTTAGAAGTAACAT 1467
 vir 1758 ACCATATTCAACAAGTTAAGCTGTAGTTTATTAAATTAAGTTTATTCT 1807
 mel 1468 AAGATGTTTTTAAAGTACATTTACATTTTATTATTGAACCTTCATTGT 1517
 vir 1808 CTTTGGGTTGATTTTTTTTGGCTTAGTCTTAATTCAATTGCGAAATTTCT 1857
 mel 1518 TCATTTGATCTTTTATTAATATCAATACAAAACAAATCGAATCTAATTA 1567
 vir 1858 GTTTTAGTGTTTAAAGATCTATTA.....TTCGTTTTTTTAAATTTT 1902
 mel 1568 TCGATAACAATTTTCAATCAGTGTACAAGGTCCATAATATTATAAACTT 1617
 vir 1903 TATCTAGCTATCGGTTTCAATTCGTAAAGGGAAAATATCTCGAATGCGTT 1952
 mel 1618 ATACCAATTATTTTAAACCCCTTCCAAATGTTTAATTTTTCTGTCTACTC 1667
 vir 1953 TCACATCACTATTTTATATCGATAGATCGATGCGACA.....AAATCAA 1996
 mel 1668 GTTTTTTAAACTTTTAAAAGCAAGCTATATTCGACATCTGTGTTTTTTT 1717
 vir 1997 CCCTGACAGATATAGAAGTGTTCGTTCCCAACAATCATTGTCATTTTATT 2046
 mel 1718 ATCTACCGATATTACGTTTGTTCATTCCATACATACCCAAGTAGGCATT 1767
 vir 2047 ACGCTTCAGTTTCTGTACGGGGCTGGACGGTTGTGTAATTGATTTTTTTT 2096
 mel 1768 GTGTCACCACCTATGCTGTGTACGTATAATACGTTTATCTGCTTTGTTCC 1817
 vir 2097 TAAACAATTTTACTTCAATCTTGTA..... 2122
 mel 1818 ATTACATTCAGTTTTCGATTTTGACGACGGTTGTCAACGAGTTGTAGCA 1867

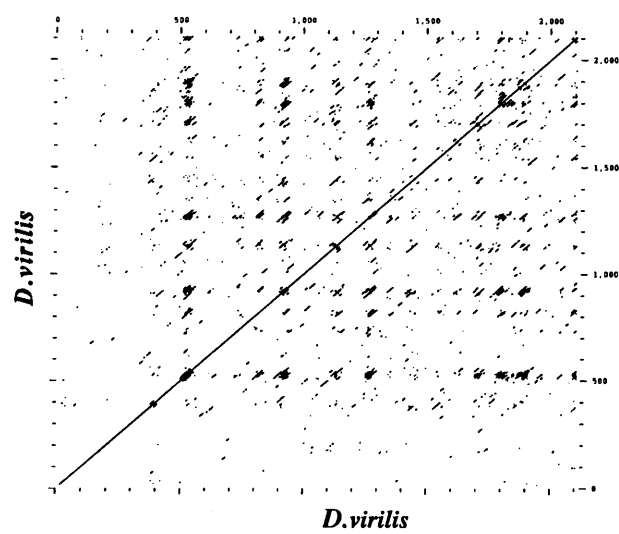
transcription start (Stanewsky, 1993)

	HSF		ABD-B
	DEFORMED		SU(H)
	HUNCHBACK		KRUPPEL
	DORSAL		SINGED
	BR-C		BICOID
	CF-II		TATA BOX (TF SEARCH)
	FTZ		TATA BOX (FINDPATTERNS)

A:



B:



C:

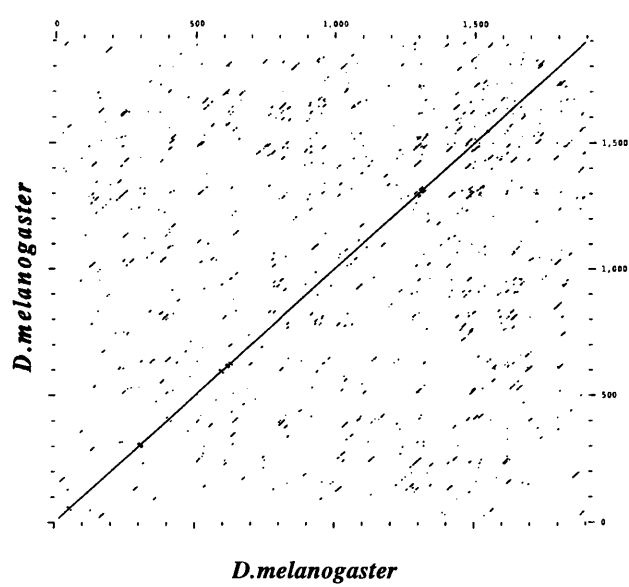


Figure 3.10. Dot plot comparisons of the *nonA* promoter region. **A:** *D.virilis* versus *D.melanogaster*. **B:** *D.virilis* versus itself. **C:** *D.melanogaster* versus itself.

interspaced by regions of complete divergence between the two sequences. In addition, the terminal third of the sequences appears to be very diverged.

The dot plot analysis was repeated by plotting each of the two sequences against itself (figure 3.10b and 3.10c), this time using a lower stringency (window: 35, stringency: 19). This gives an idea of the internal structure, in terms of repeatability, of each sequence. From the dotplot results it seems that repetitive DNA does not play a major role in the sequence composition of the *D.melanogaster* promoter region (figure 3.10c). However, the promoter of *D.virilis nonA* seems to have a considerable amount of repetitive DNA (figure 3.10b) which is concentrated in the regions which mostly differ from the *D.melanogaster* sequence (figure 3.10a).

When homologous genes from different species are compared, it is not rare to observe that areas of high divergence are often associated with regions of highly repetitive DNA. This phenomenon has been noted not only in coding sequences, for example the coding sequence of the *hunchback* (*hb*) gene (Treier *et al.*, 1989) and the Threonine-Glycine region of the *per* gene (Peixoto *et al.*, 1993), but also in regulatory regions such as the promoter and spacer regions of the ribosomal RNA genes (Tautz *et al.*, 1986, Tautz *et al.*, 1987; Hancock and Dover, 1988). The promoter region of *hb* has been recently characterized in four different species (*Musca domestica*, *D.melanogaster*, *D.virilis* and *Tribolium castaneum*) and high levels of short DNA motif repeats (simple DNA, Tautz *et al.*, 1986) were noticed in all the four species' promoter sequences (Hancock *et al.*, 1996). Events such as slippage and gene conversion seem to be the main molecular processes generating the rapid turnover of such redundant DNA sequences (Dover *et al.*, 1982; Dover, 1993; Tautz *et al.*, 1986).

The TF SEARCH program (Akiyama, 1995) was used to analyze the 5' regulatory region of the *D.melanogaster* and *D.virilis nonA* genes. This program highlights small nucleotide motifs homologous to known transcription factor binding sites from the TFMATRIX transcription factor binding site profile database of Wingender *et al.* (1997). Each motif is presented with a score of identity to known motifs in the database. Short motifs (< 10 nucleotides) were considered in the analysis only if scoring > 90, while longer motifs (> 10 nucleotides) were considered if they

scored > 85. The motifs found in each sequence and the number of times they occurred in the sequence are listed in table 3.3. Figure 3.9 highlights the position of the motifs found in each of the two aligned sequences.

Table 3.3. Transcription factor-binding motifs in the *D.virilis* and *D.melanogaster* promoter region.

	HSF	DFD	HB	DL	BRC-Z	CF2	ABD-B	SU(H)	FTZ	KR	SN	BCD
<i>vir</i>	23	14	10	1	9	1	1	1	1	0	0	0
<i>mel</i>	27	9	2	4	4	3	1	1	1	1	1	1

Both sequences have high numbers of motifs homologous to the heat shock factor binding motif (HSF). These are distributed throughout the whole sequence of *D.virilis* and *D.melanogaster* regulatory regions in both orientations. In *Drosophila* the HSF binding element is only five bases long and its consensus sequence is NGAAN, where the first nucleotide (N1) is usually an A and the three nucleotides GAA are highly conserved (Fernandes *et. al.*, 1994). A typical heat shock factor binding signal is constituted by several HSF units with alternate orientations. It has been shown *in vitro* that at least two adjacent HSF units are necessary to obtain a stable binding with the heat shock transcription factor. The distance between adjacent HSF units seems to vary in different genes (Fernandes *et. al.*, 1994). In *D.melanogaster* there are five pairs of HSF units (with a score > 90) in which the two HSF motifs are separated by less than 15 nucleotides. Similarly, in *D.virilis*, five pairs of nearby (< 15 nucleotides apart) HSF motifs are found. The presence of these signals could be sufficient to promote binding of heat shock transcription factors in the promoter regions of both *nonA* genes. This would assure production of NONA protein during stress conditions. It has to be noted, however, that, given the small size of the HSF motifs, homology between these motifs and regions in the examined sequences could be purely coincidental.

Nine motifs with homology (score > 85) to the homeotic protein Deformed binding motif were found in the *D.melanogaster* 5' sequence, and 14 were found in *D.virilis*. Within each sequence these motifs are found at regular intervals, and interestingly, on three occasions they are found in homologous positions within the two aligned sequences (see figure 3.9). The conservation of these motifs, whose sequence is 17 nucleotides long, may be more than coincidence, and could have functional implications. The same could be said about a *hb* product putative binding domain (score > 85) which was found at exactly the same position in the two *D.melanogaster* and *D.virilis* sequences (figure 3.9). However, it has to be noted that these homologies in the position of functional motifs all depend on the alignment of the two sequences. In fact, comparing the position of functional motifs acquires significance only if we assume that the alignment of the two examined sequences with the Gap program is indeed correct.

One other *hb* binding domain is found in the *D.melanogaster* sequence, and nine others in *D.virilis*. In the anterior part of the *Drosophila* embryo, levels of the *hb* gene product control the transcription of the gap gene *Krüppel* (by repressing it when at high concentrations, and activating it at lower concentrations) and simultaneously repress the transcription of the abdominal structures-forming gap genes *knirps* and *giant* (Rivera-Pomar and Jackle, 1996).

Four putative dorsal (*dl*) binding motifs with an identity score > 85 (the last two of them with score > 90) are present in the sequence of *D.melanogaster* (see figure 3.9) while only one (score > 85) is found in the *D.virilis* sequence in a region that is not alignable with the *D.melanogaster* sequence. In *Drosophila*, *dorsal* (*dl*) contributes to determining the destiny of the different embryonic regions along the dorso-ventral axis (Ray and Schüpbach, 1996). The product of *dl* is localized in the nucleus only on the ventral side of the embryo. Dorsal both activates and represses gene expression. It activates the genes *twist* (*twi*) and *snail* (*sna*), which are required for the development of ventral structures, and it represses the genes *decapentaplegic* (*dpp*) and *zerknüllt* (*zen*), which are required for the development of dorsal structures (Steward and Govind, 1993; Rusch and Levine, 1996).

Four motifs in the *D.melanogaster* sequence and nine in *D.virilis* are homologous (score > 85) to the sequence recognized by the transcription factors produced by the *Broad-Complex (BR-C)* genes. The transcription factors produced by the *BR-C* are known to bind and activate transcription of the glue protein gene *Sgs-4* in third instar larvae (Kalm *et al.*, 1994). The NONA protein also binds to the promoter of the *Sgs-4* gene when this is being actively transcribed (Saumweber, 1990; Besser *et al.*, 1990). It has been hypothesised that NONA could be processing the nascent RNA as it is being transcribed from the glue protein gene (Besser *et al.*, 1990). Could it be that the *BR-C* transcription factors activate the *Sgs-4* gene and, by promoting expression of the *nonA* gene as well, also ensure that the transcripts of the glue protein gene are properly processed?

Several other transcription factors binding motifs were found in both *D.melanogaster* and *D.virilis* sequences (such as the motifs for binding of the products of the genes: *Chorion factor 2 (Cf2)*, *fushi tarazu (ftz)*, *Abdominal B (Abd-B)* and *Suppressor of Hairless (Su(H))*, see table 3.3). However in all these cases the motifs were found in very different locations within the two sequences (see figure 3.9). It is not possible to speculate on the possible functional significance of these motifs. However it has to be mentioned that, in few instances, certain motifs scored very highly (> 90) for identity with known binding motifs. For example, the putative *ftz* binding motif found in the beginning of the *D.melanogaster* sequence (figure 3.9) had a score of 94.3. Similarly, the *Su(H)* binding domain identified in the sequence of *D.virilis* scored 90.4. Given that the *ftz* binding motif is 12 nucleotides long, and the *Su(H)* motif measures 15 nucleotides, their presence in these promoters could reflect some functional importance, however protein-DNA mobility assays should be used to shed light on these possibilities (e.g. Guille and Kneale, 1997).

Finally, several binding motifs were found exclusively in the *D.melanogaster* sequence (those for the gene products of *Kruppel (Kr)*, *singed (sn)* and *bicoid (bcd)*). In particular, one domain homologous to the *bcd* binding motif was assigned a very high score (96.2), as was a *Kr* putative binding motif (score = 93.5). Again, *in vivo*

and *in vitro* approaches could identify whether any of these binding sites are functional (Guille and Kneale, 1997; St Johnston and Nüsslein-Volhard, 1992).

Lastly, several sequences homologous to the TATA box consensus were localized by the TF SEARCH program. Only those with the appropriate orientation for the *nonA* gene are presented in figure 3.9. In the *D.virilis* promoter three putative TATA boxes are recognized, however the last one presents the highest homology (score = 88) to the consensus sequence (Brethnach and Chambon, 1981). Since the position of the transcriptional start site is not known in *D.virilis nonA*, it is not possible to speculate on the position of the TATA box element. In the sequence of *D.melanogaster* the only TATA box found by the program does not correspond to the one predicted by Besser *et al.* (1990) and does not fit with the position of the transcription start site localized by S1 mapping by Stanewsky *et al.* (1993) (nucleotide 1828 in the genomic sequence published by Jones and Rubin, 1990; indicated in figure 3.9). Approximately 30 bp upstream of the start of transcription (Stanewsky *et al.*, 1993), there is however a TATAATA motif (highlighted in figure 3.9), that the TF SEARCH program fails to detect, and which is likely to be the real TATA box in *D.melanogaster*. The program Findpattern of the GCG packet was therefore used as an additional means to localize sequences homologous to the TATA box consensus motif (Brethnach and Chambon, 1981) in the *D.virilis* promoter. All sequences homologous to the TATA A/T A/T A/T motif were searched with the aid of this program. Several homologous motifs were found, however only four, located in the terminal part of the sequence (after nucleotide 1000), were considered (respective positions: 1105, 1216, 1725, 1877, boxed in figure 6.9). Of these, the motif at position 1725 seems particularly interesting, since it was recognized also by the TF SEARCH program and was assigned the highest score among the possible TATA box motifs (see above).

The *nonA* region is genetically complex. An essential gene called *l(1)i19e* is associated with the *nonA* gene and partially overlaps with it so that the two genes cannot be molecularly separated (Stanewsky *et al.*, 1993). At present, only point mutations or small deletions 3' of a *SalI* site in the middle of the *nonA* second exon can

be assumed to affect the *nonA* gene only (Stanewsky *et al.*, 1993). This means that all the sequence upstream of the middle of the second exon of *nonA* must be shared between the two genes. Therefore the 5' upstream DNA region of *nonA* that has been studied here could also be a functional part of the *l(1)i19e* gene. If in *D.virilis* the same overlap exists between *nonA* and the *l(1)i19e* gene as in *D.melanogaster*, then the homology between the *D.melanogaster* and the *D.virilis* 5' regulatory regions could be reflecting functional importance not only for the *nonA* gene, but also for *l(1)i19e*.

The coding sequence of the *D.virilis nonA* gene

Figure 3.11 shows the nucleotide and amino acid sequence of the translated region of the *D.virilis nonA* gene. Arrows indicate the position of the introns. A dot matrix comparison (window: 21, stringency: 14) between the *D.virilis* and the *D.melanogaster nonA* coding sequences is presented in figure 3.12a and reveals an area of considerable divergence covering approximately the first half of the gene. This is followed by a region of very high homology between the two genes which extends for almost all the remaining length of the sequences. The Gap program was used to align the two sequences (appendix 3.3). After alignment, the two sequences shared 72.4 % identity. Not surprisingly, the area of high homology corresponds to the sequence for the RNA-binding domains (RRM1 and RRM2) and the charged domain (see Introduction). In addition, the sequence at the end of the *nonA* gene seems to be quite well conserved between *D.virilis* and *D.melanogaster*. However, a number of parallel shifts of the plot line indicate regions of gaps in the alignment of the two DNA sequences. Regions of very densely packed dots are also visible; these occur in several regions at the beginning and at the end of each of the two sequences, and indicate repetitive regions that cause numerous shifts in the plot line. Note that repetitive regions at the beginning of the *D.virilis nonA* gene align with sequences not only at the beginning, but also in the middle and at the end of the *D.melanogaster* sequence, and regions at the end of the *D.virilis* gene are similar to sequences found approximately in the middle of the *D.melanogaster* gene.

Figure 3.11. DNA and amino acid sequence of the *D. virilis nonA* gene. Vertical bars indicate the position of the introns. The sequence highlighted in yellow is homologous to the DBHS domain (see introduction). The two RNA-binding domains (RRM1 and RRM2) are underlined. The RNP1 motifs are shown in blue and the RNP2 in red.

```

1  ATGGAATTTCTGTTAAATGGACAATTCTGGCAATTCAACACCGCTGCCACAGCGCCAA  60
   M E N S V K M D N S G N S T P L P Q R Q

61  CGAAGAGCTAATAATCAACCGAATAAAAAATATCGGCAAACTCGGCCCACAGAAGCAAAAT  120
   R R A N N Q P N K N I G K L G P Q K Q N

121 GAAGGTGCATCCGACGGAGGCCAGCAGAAAAACGGCAACGTTTTGGCCCGAATAATCAG  180
   E G A S D G G P A E K R Q R F G P N N Q

181 AATGGTGGCGGTGGTAGCGTCGTCGGTGGTGGCGGCGGCGGCGGTGGAGGTGGCCAA  240
   N G G G G S V V G G G G G G G G G G G Q

241 AATCAAAATAAAATTTTCGGAACAAGGGTGGATTTGGCGGAGGTGGAATCGCAACCGC  300
   N Q N K N F A N K G G F G G G G N R N R

301 AATCGTGGCGGCAATCAAAATAGGTCTGAATTTTCAAAATCAGAATCAAAACCAAAAGTCC  360
   N R G G N Q N R S N F Q N Q N Q N Q K S

361 ACAACAGATGCACCAAAAGCGGACGGAGGAAATCTCAATGACAAATCTAACGAAGCTAAC  420
   T T D A P K A D G G N L N D K S N E A N

421 AATGCAAATCAATCCAATTCAAATTCAGCTGCCCCAAGCCCAGGCACAACCTCAAGCGCAG  480
   N A N Q S N S N S A A Q A Q A Q L Q A Q

481 GCCCAAGCTCATGCCCAGGCCCAAGCTCAGGCTCAGGCTCAGGCTCATGCTCAGGCCCCAG  540
   A Q A H A Q A Q A Q A Q A Q A H A Q A Q

541 GCTCAAGCTCATGCGCATGCTCAAAATCAAGCATTTAGGGCTCGCGGCGGTGGCGGCGGC  600
   A Q A H A H A Q N Q A F R A R G G G G G

601 GGCGGTGGTGGTGGTGGTGGAGGTGGTGGTGGCGGCGGCGGCGGCGGCGGAGGTGGAGGA  660
   G G G G G G G G G G G G G G G G G G G G

661 GGAGGAGGCGGACGTGATAGGAATCCGATCGGCGTGGTGGTGGTGGCGGCGGCGGACAA  720
   G G G G R D R N P D R R G G G G G G G Q

721 AACTCTGGCGGCGGTAACAATTCACAGCGTGGCGATGACTTCTTTTATAGCCAACGGTTA  780
   N S G G G N N S Q R G D D F F Y S Q R L

781 CGCAGCATTAGTGGTCCCACACATGAGCTACCACCGATCGAAGTCGCCCAGGAGACCAAG  840
   R S I S G P T H E L P P I E V A Q E T K

841 TTCTCTGGACGCAATCGTCTCTATGTTGGCAATTTGACCAATGACATTACCGATGAGGAG  900
   F S G R N R L Y V G N L T N D I T D E E

901 TTGCGCGAAATGTTTAAGCCGTACGGTGAAATAGGCGAGATATTTTCCAATCTTGAAAAA  960
   L R E M F K P Y G E I G E I F S N L E K

961 AACTTTACATTCTTAAGGTTGACTATCATATTAATGCCGAAAAGGCCAAGCGGCCCTTA  1020
   N F T F L K V D Y H I N A E K A K R P L

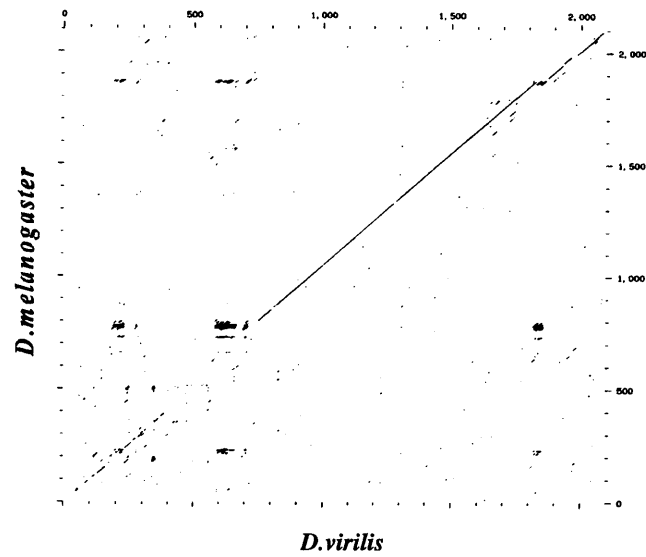
1021 GACGGCTCAATGCGTAAGGGTCGCCACGTTTCGCGTCCGTTTCGCCCCAACGCAACCATT  1080
   D G S M R K G R H V R V R F A P N A T I

1081 TTGCGTGTAGCAATCTGACACCGTTTCGTGTCCAACGAGCTGCTCTACAAGTCCTTTGAA  1140
   L R V S N L T P F V S N E L L Y K S F E

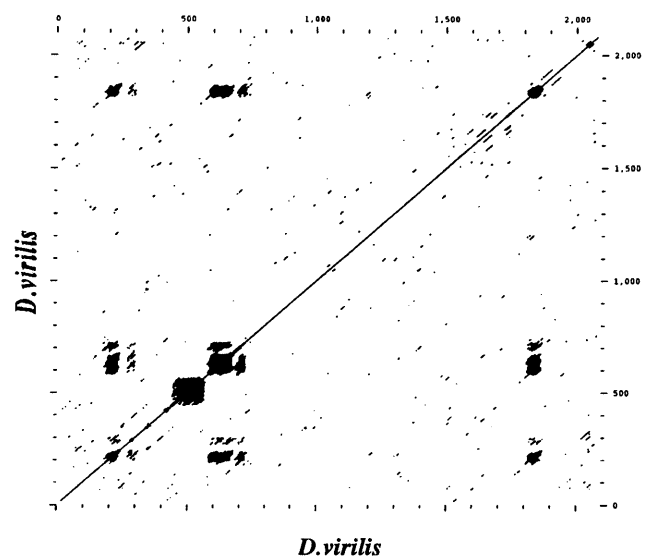
```

1141 ATATTTGGCCCCATTGAGCGGGCCAGCATTACAGTCGACGATCGTGGCAAGCATTGTTGGGC 1200
 I F G P I E R A S I T V D D R G K H L G
 1201 GAGGGCACCGTTGAGTTTGCCAAAAAATCATCGGCCAGCGCCTGTCTGCGCTTGTGCAAT 1260
 E G T V E F A K K S S A S A C L R L C N
 1261 GAGAAATGCTTCTTTTAAACGGCCTCGTTGCGTCCGTGCTTGGTGAACCGATGGAGGTT 1320
 E K C F F L T A S L R P C L V E P M E V
 1321 AACGATGACAACGACGGATTGCCCAGAGAAGGCCCTCAATAAGAAGCTGCAAGAGTTCAAT 1380
 N D D N D G L P E K A L N K K L Q E F N
 1381 CAGGAGCGCAGTGTGTTGGGCCCCGTTTCGCCGATCTAAATTCATTTGAGCACGAGTACGGC 1440
 Q E R S V G P R F A D L N S F E H E Y G
 1441 TCCAGGTGGAACAGTTGCATGATCTATTCAAAAGCAAGCAAGATGCTCTCAAGCGTGAG 1500
 S R W K Q L H D L F K S K Q D A L K R E
 1501 CTCAAAATGGAAGAAGAAAACTTGATGCTCAGATGGAGTATGCTCGCTATGAGCAAGAA 1560
 L K M E E E K L D A Q M E Y A R Y E Q E
 1561 ACAGAGCTTTTTCGTCAAGAAATTAAGAAAACGCGAATCTGATAATGAGCGTAAGAAATTG 1620
 T E L L R Q E L R K R E S D N E R K K L
 1621 GAATGGGAAATGCGCGAGAAACAGGCCGAGGAAATGCGCAAACGCGAAGAGGAGACCATG 1680
 E W E M R E K Q A E E M R K R E E E T M
 1681 CGTCGTCATCAAACCGAAATGCAGAGCCGTATGGTGGCGCCAGGAGGAGGATATGCGTCGT 1740
 R R H Q T E M Q S R M V R Q E E D M R R
 1741 CGTCAGCAGGAGAACACATTATTCATGCAGGCCCAGCAGCTAAACTCGCTGCTTGATCAA 1800
 R Q Q E N T L F M Q A Q Q L N S L L D Q
 1801 CAGGAAGGCTTCGGCGGTGGCAACGGTGGTGGCGGCGGCGGCGGTGGTGGCGGCGGCGGC 1860
 Q E G F G G G N G G G G G G G G G G G G G
 1861 GTGGGCAATTCAAACCTTTGATAACTTCGGGGGGCAACAGCAACTCCCCGTTCGAAGTGTTTC 1920
 V G N S N F D N F G G N S N S P F E V F
 1921 AGAGGTAATAACAACCTCGTCCATGGCTGGCAACAATGCCGGACCCGGTGCAAATAACCAG 1980
 R G N N N S S M A G N N A G P G A N N Q
 1981 CAGCAGGACTCCTTCGCTGCTTTTGAATTTGGGGTTAACAATATGAACCAAGGCGGCAAT 2040
 Q Q D S F A A F E F G V N N M N Q G G N
 2041 CAACGTGGAAATAATGGCGGAAATAATGTTCCATGGGGACGACGTCGTTTTTAGAC 2096
 Q R G N N G G N N V P W G R R R F *

A:



B :



C:

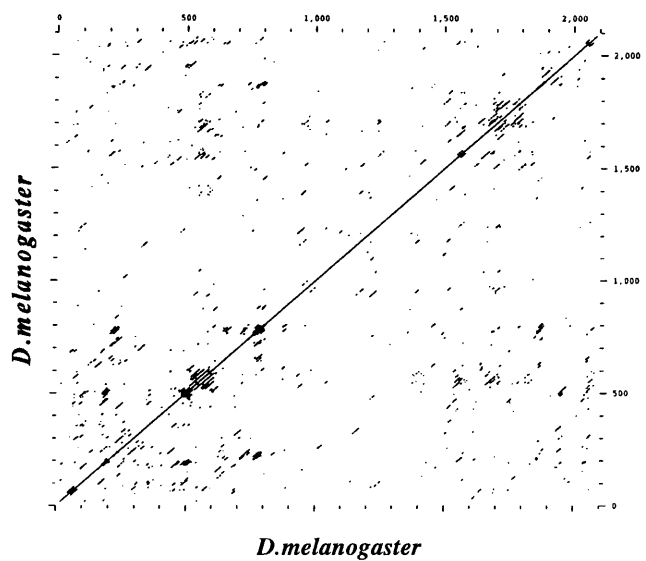


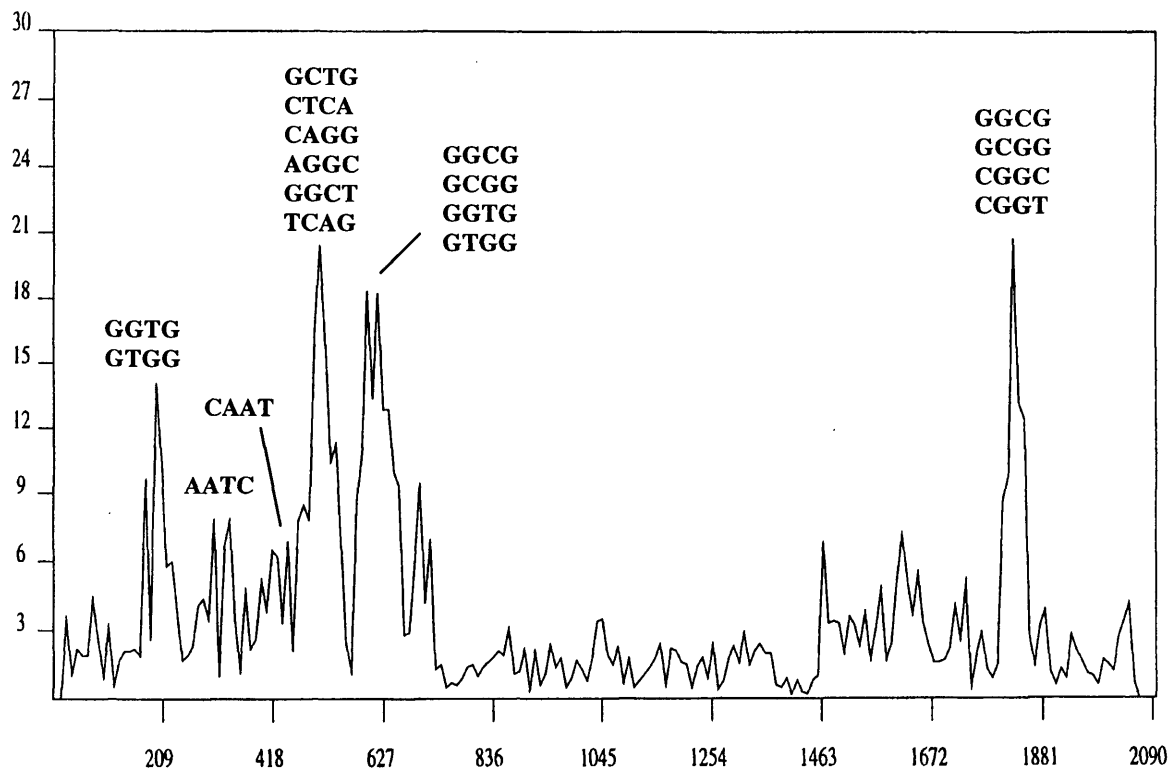
Figure 3.12. Dot plot comparisons of the *nonA* gene coding sequence. **A:** *D.virilis* versus *D.melanogaster*. **B:** *D.virilis* versus itself. **C:** *D.melanogaster* versus itself.

The dot matrix analysis was repeated after lowering the stringency (window: 35, stringency: 19) and plotting each of the two *nonA* sequences against itself. The output for the *D.virilis* gene (figure 3.12b) clearly shows numerous and large regions of repeated DNA, clustered especially at the beginning and the end of the gene. The *D.melanogaster nonA* gene (figure 3.12c) seems much less repetitive than its *virilis* homologue, however regions of repeats are still evident especially within the first third of the gene, and towards the end of it.

The frequent occurrence of several repeated short motifs in close proximity (cryptic simplicity) seems to be a widespread and important component of eukaryotic genomes (Tautz *et. al.*, 1986). To have a more accurate idea of the extent of the redundancy in each *nonA* gene (either of *D.melanogaster* or *D.virilis*) and to discover which sequence motifs are involved, a cryptic simplicity analysis was performed on the *D.melanogaster* and the *D.virilis nonA* gene coding sequences. The cryptic simplicity test was carried out by John Hancock, using the SIMPLE34 program (Hancock and Armstrong, 1994). The SIMPLE34 program, which is a modified version of Tautz *et. al.*'s (1986) SIMPLE program, analyzes a given sequence by moving a 64 bp window along it. For each window, it finds three or four base pair motifs at its centre, and searches for duplicates of those motifs within the window. It awards one point for each duplicate of the three base motif, and an extra three points if the four base motif is duplicated. It then calculates an average score for the entire sequence based on all the window scores, which is the "Simplicity Factor" (SF). It then repeats the same analysis for ten random sequences with the same base doublet frequency as the input sequence and the same length. This generates a "Randomized Simplicity Factor" (RSF). The "Relative Simplicity Factor" (RelSF) is the ratio of SF over RSF, and indicates how much more simplicity the input sequence contains than the random sequences.

The SF of the *D.melanogaster nonA* coding sequence was 3.034, and as the RSF was 1.954, the RelSF of *D.melanogaster nonA* was calculated as 1.552. The SF of the *D.virilis nonA* gene was 3.687, the RSF 1.943, and the RelSF 1.898. From these values it therefore appears that the *D.virilis nonA* sequence is more repetitive than that of *D.melanogaster*. Figure 3.13a and 3.13b are the graphic representation

A:



B:

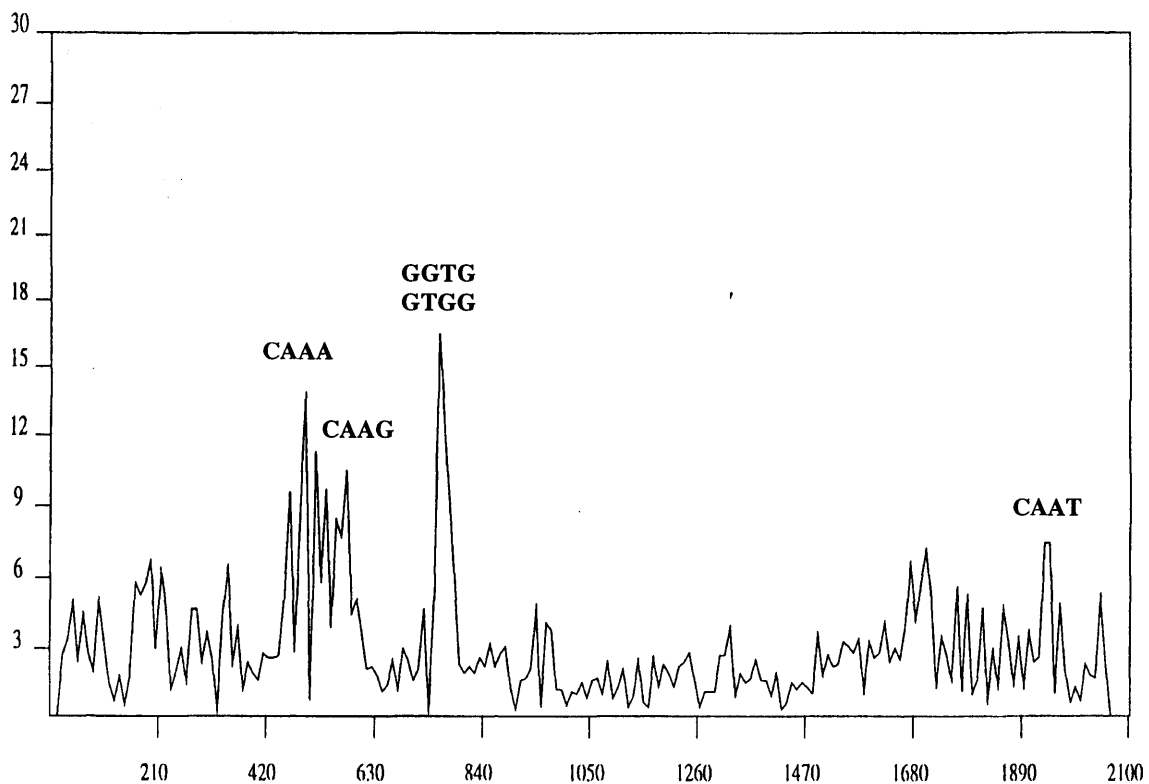


Figure 3.13. Results of the cryptic simplicity test for the coding sequence of *nonA*. **A:** *D. virilis*. **B:** *D. melanogaster*. The highest peaks represent motifs (whose sequence is indicated) found to be concentrated in the sequence to a degree unlikely to occur by chance.

for the results of the cryptic simplicity test. The high peaks found in both displays, indicate that both *nonA* genes have regions which are cryptically simple through the repetition of one short motif or another. The highest peaks represent the most statistically significant simple motifs (termed SSMs) in the sequence. These correspond to the motifs found to be concentrated in the sequence to a degree unlikely to occur by chance. The base compositions of these short motifs is indicated in the graphic display for each of the two sequences analyzed. Both *D.melanogaster* and *D.virilis nonA* genes present areas of simplicity especially towards the first third of the gene, even though *D.virilis* does so to a much larger extent. In both *nonA* genes there is also an additional repetitive region towards the end of the sequence, however, again, this is more pronounced in the *D.virilis* gene. Moreover, the very high peak towards the end of the *D.virilis* sequence indicates that a single motif constitutes a significantly large part of the repeats in this part of sequence. In the terminal part of the *D.melanogaster* sequence instead, although many repeats are found, no motif in particular is significantly over-represented.

Recent comparison of the promoter region of the *hb* gene (Hancock J., pers. comm.), has revealed not only that a large number of SSM motifs exists in this region of *D.melanogaster*, *D.virilis*, and *Musca domestica*, but also that many of these highly represented motifs are shared between the sequences of these different species. Similarly, in the coding region of *D.melanogaster* and *D.virilis nonA* some SSMs were found to be shared by the two sequences (GGTG and GTGG, see figure 3.13). However, several other SSMs were found in the *D.virilis* sequence but not in the *D.melanogaster*. A CAAT motif which is highly repeated at the beginning of *nonA* in *D.virilis*, was also found in the *nonA* sequence of *D.melanogaster*. In the *D.melanogaster* gene this motif is localized towards the end of the sequence and it is represented to a much lower extent. Most of the repeated motifs in *nonA* are constituted by cyclic permutations of the short sequence GGT or GGC. Note that GGT and GGC both code for Gly which constitutes a very large repeated region in the protein of the *D.virilis nonA* gene and a slighter smaller one in the protein of *D.melanogaster*.

Conceptual translational start of the *D.virilis nonA* gene

Figure 3.14 shows a partial sequence from the *D.virilis nonA* gene indicating the ATG motif which is believed to initiate translation. Conceptual translation of the sequence starting from this ATG produces an open reading frame with homology to that of the *D.melanogaster nonA* gene (see figure 3.15). Few other ATG motifs can be found upstream of the one highlighted in figure 3.14. However conceptual translation from these other ATGs contains multiple termination codons in all three reading frames. Moreover, the predicted ATG starting codon is in agreement with the most stringent requirements dictated by Kozak's consensus translation initiation sequence (GCCRCCATGG) (Kozak, 1987; 1991). In fact, in this case, a purine (G) is present three bases before the ATG motif, and a G is immediately following it.

Protein comparison

Figure 3.15 shows an alignment of the amino acid sequence of *D.virilis* and *D.melanogaster* NONA proteins, created with the Bestfit program of the GCG packet. This program finds optimal alignments by inserting gaps to maximize the number of matches using the algorithm of Smith and Waterman (1981). The percent of overall identity between the two is just over 75%, while the similarity is about 83%. Previous comparisons between several *D.virilis* and *D.melanogaster* homologous gene products have shown protein identities ranging between 50 and 83 %. Table 3.4 shows the percent similarity and identity shared by several *D.virilis* and *D.melanogaster* proteins, calculated with the algorithm of Smith and Waterman (1981). Note that, in general, developmental gene products seems to be most conserved in these two species.

Table 3.4. Percent similarity and identity shared by 10 *D.melanogaster* and *D.virilis* homologous proteins.

	NONA	EN	HB	KNIRPS	NOS	OSK	PER	RUNT	SEV	TRA
% sim	82.8	87.8	88.5	83	77.7	74.3	78.4	86.9	77.9	65.9
% id	75.1	82.9	81.2	74.8	64.2	59.4	67.6	80.9	64.5	50

Figure 3.14.

Conceptual translational start of the *D. virilis nonA* gene. **a**, **b**, and **c** are the three reading frames. The ATG motif believed to initiate translation is boxed in red. The first three amino acids of the *D. virilis* NONA protein are underlined in red. Other ATGs (boxed in white) are found in the ~ 400 nucleotides upstream, however, translation from these produces stop codons in all three reading frames.

```

AGTTTAAACTAAATAAATGTTACCATATTCAACAAGTTAAGCTGTAGTTTATTTAAATTAG
a S L N * I N V P Y S T S * A V V Y L N * -
b V * T K * M Y H I Q Q V K L * F I * I S -
c F K L N K C T I F N K L S C S L F K L V -

TTTTATTCTCTTTGGGTTGATTTTTTTGGCTTAGTCTTAATTCAATTGCGAAACTTTCT
a F Y F S L G * F F W L S L N S I A K L S -
b F I S L W V D F F G L V L I Q L R N F L -
c L F L F G L I F L A * S * F N C E T F C -

GTTTTAGTGTTTAAAGATCTATATATTTAACGTTTTTTAAATTTTATCTAGCTATCGGT
a V L V F K D L Y I * R F L N F Y L A I G -
b F * C L K I Y I F N V F * I F I * L S V -
c F S V * R S I Y L T F F K F L S S Y R F -

TTCATTTCGTAAAGGGAAAATATCTCGAATGCGTTTCACATCACTATTTTATATCGATAG
a F I S * R E N I S N A F H I T I L Y R * -
b S F R K G K I S R M R F T S L F Y I D R -
c H F V K G K Y L E C V S H H Y F I S I D -

ATCGATGCGACAAAATCAACCCTGACAGATATAGAACTGTTTCGTTCCCACAATCATTGT
a I D A T K S T L T D I E L F R S H N H C -
b S M R Q N Q P * Q I * N C F V P T I I V -
c R C D K I N P D R Y R T V S F P Q S L S -

CATTTTCATTACGCTTCAGTTTCTGTACGGGGCTGGACGGTTGTGTAATTGATTTTTTTTT
a H F I T L Q F L Y G A G R L C N * F F F -
b I S L R F S F C T G L D G C V I D F F L -
c F H Y A S V S V R G W T V V * L I F F * -

AAAACAATTTTACTTCAATCTTGTATGAAATTT
a K T I L L Q S C N G K -
b K Q F Y F N L V M E N -
c N N F T S I L * W K I -

```

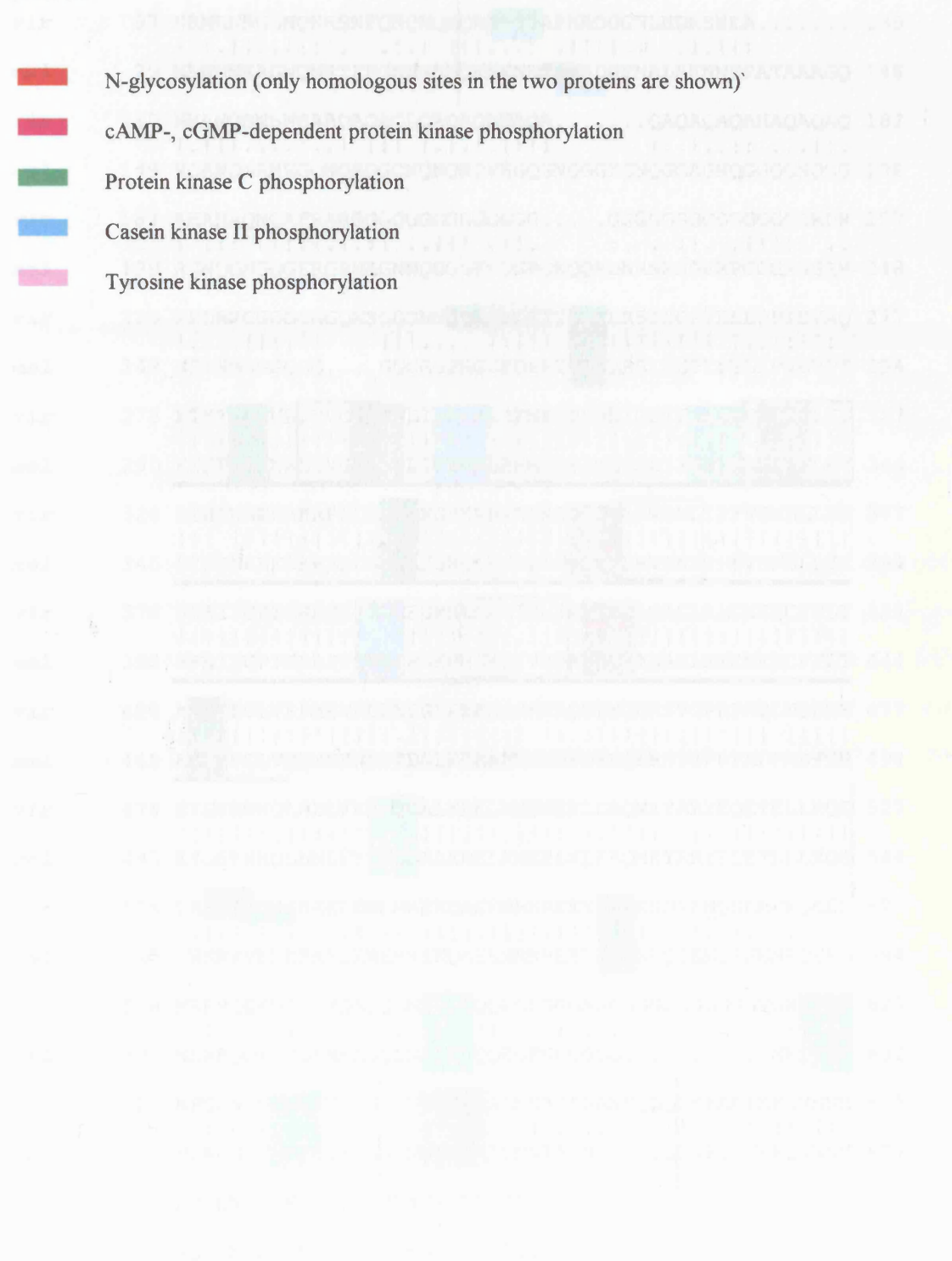
The first half of the protein is clearly the most diverged (figure 3.15), with the first exon of *nonA* showing only 60% identity, as is the case for the first half of the second exon. Note that these regions of divergence are constituted in large part by stretches of repeats. In the first exon stretches of poly-Gly are found especially in the *D.virilis* protein. In the beginning of the second exon, a QN and a degenerate GNQGGX repeat found in the *D.melanogaster* gene have been replaced by a QA and a very long poly-Gly repeat. A long stretch of 29 Gly residues is the most striking feature of the *nonA virilis* gene. 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 are believed to function primarily in protein-protein interactions, since all of these proteins are components of multiprotein ribonucleoprotein complexes. However, the auxiliary domains may also influence the RNA-binding activity of a protein. A comparison between the binding of A1 to single-stranded (ss) DNA and RNA with that of UP1, the truncated auxiliary domain-less derivative of A1, suggested that the auxiliary domain may modify the polynucleotide binding properties of RNA-binding domains (Bandziulis *et al.*, 1989).

Given the functional importance of the RNA-binding domain (RRM1+RRM2) and the adjacent charged domain of NONA (Rendahl *et al.*, 1996, Stanewsky *et al.*, 1996), it is not surprising that these regions of the protein are very well conserved between *D.virilis* and *D.melanogaster*. Only 5 non-conservative changes are in fact found in this protein region that spans a total of 311 amino acids. The amino acid sequence encoded by the second part of exon three is also very well conserved, except for a region in which a small poly-G repeat found in *D.melanogaster* seems to have expanded into a much larger one in *D.virilis*. The small exon four encodes for a very diverged part of the protein where alignment between the *D.virilis* and *D.melanogaster* amino acid sequences is possible only in regions containing N repeats (see figure 3.15).

Perhaps more interesting is the high degree of conservation found in the amino acid sequence encoded by the fifth exon. This part of the protein in fact does not

Figure 3.15. *D.virilis* and *D.melanogaster* NONA alignment. Breaks (vertical lines) in the sequence represent intron positions. The amino acid sequence in a yellow background represents the DBHS domain (see Introduction). The two RNA-binding domains (RRM1 and RRM2) are underlined, and the RNP1 and RNP2 motifs are boxed. Putative post-translational modification sites are highlighted in colour as follows:

- N-glycosylation (only homologous sites in the two proteins are shown)
- cAMP-, cGMP-dependent protein kinase phosphorylation
- Protein kinase C phosphorylation
- Casein kinase II phosphorylation
- Tyrosine kinase phosphorylation



[illegible]

appear to have any recognizable functional motif. Translation of the fifth exon appears however to be indispensable for the functioning of the NONA protein: *nonA* mutant flies (either *nonA*⁽⁻⁾ or *nonA*^{dis}) carrying a *nonA* transgene, in which an *in vitro* generated stop codon prevents translation of the fifth exon, do not show any rescue of the mutant phenotypes (Jones and Rubin, 1990; Rendahl *et al.*, 1992; Stanewsky *et al.*, 1993).

Protein motifs in NONA.

The PSITE program (Version 1, Solovyev and Kolchanov, 1994) searches for functional motifs in a protein using the PROSITE database (Bairoch and Bucher, 1994). The results of the program for the *D.virilis* and *D.melanogaster* NONA proteins are graphically illustrated in figure 3.15.

Eight possible N-glycosylation sites are found in the *D.virilis* NONA sequence, and fourteen in the *D.melanogaster*. Five of these are found at homologous positions in the two aligned proteins (see figure 3.15). Interestingly, a site found at position 645 of the *D.virilis* amino acid sequence (figure 3.15), is also present in the *D.melanogaster* sequence in spite of being located in a poorly conserved region of the protein (the region encoded by exon four). It would seem however improbable that NONA is extensively glycosylated, since from Western Blot studies with both monoclonal and polyclonal antibodies (Stanewsky *et al.*, 1993; Rendahl *et al.*, 1996) the protein does not appear to be highly modified and the band it produces in the blots is of the predicted size, based on the primary sequence.

Two putative cAMP- and cGMP-dependent protein kinase phosphorylation sites are present in the *D.virilis* protein, one of which is also found in the *D.melanogaster* protein. Interestingly, the site found only in the *D.virilis* amino acid sequence, owes its presence to a S residue which is found at amino acid position 534 of the *D.virilis* protein instead of a V residue (as in the *D.melanogaster* NONA). The same S residue in the *D.virilis* protein creates also a possible casein kinase II

phosphorylation site which is not found in the NONA protein of *D.melanogaster*. It is intriguing that the V-S substitution occurs in the *D.virilis* amino acid sequence within the highly conserved RRM + acidic domains and is one of the very few non-conservative changes in this region of the protein. This allows speculation that the V-S substitution could have some functional importance for the *D.virilis* protein. Another possibly interesting site is found at position 250 of the *D.virilis* amino acid sequence. It is a site for tyrosine kinase phosphorylation and it is not found in the *D.melanogaster* amino acid sequence. Finally, five protein kinase C phosphorylation sites are found at homologous positions within the two aligned NONA proteins, while three more are found only in the *D.virilis* protein in a region of high divergence (see figure 3.15).

NONA secondary structure

The secondary structure of the NONA protein of *D.virilis* and *D.melanogaster* was predicted using the GCG programs Pepplot, Peptide structure and Plotstructure. The output of Pepplot is shown in figure 3.16 and 3.17. The first panel of the Pepplot picture shows a schematic representation of the amino acid sequence. Each residue is represented by a line at the position where it occurs in the sequence. The lengths and colours of the lines are used to show chemically similar groups of amino acids as follows: RED= hydrophilic, charged (down=acidic, up=basic); YELLOW= hydrophilic, uncharged (short=amides, long=alcohols); GREEN= hydrophobic (short=aliphatic, long=aromatic); BLACK= Proline; UNMARKED= Alanine, Glycine, Cysteine. The second panel displays the residues which are beta-sheet forming and breaking, as defined by Chou and Fasman (1978). To produce beta-structures, there should be at least three beta-forming residues and not more than one breaking residue within a window of five. The third panel of the plot shows the Chou and Fasman (1978) propensity measures for alpha-helix and beta-sheet. As each curve rises above the threshold for its colour, it satisfies one criterion for propagation of an alpha-helix or beta-sheet structure. If the curves for alpha and beta propagation drop below the black threshold and if there is at least one breaking residue in four, then the structure

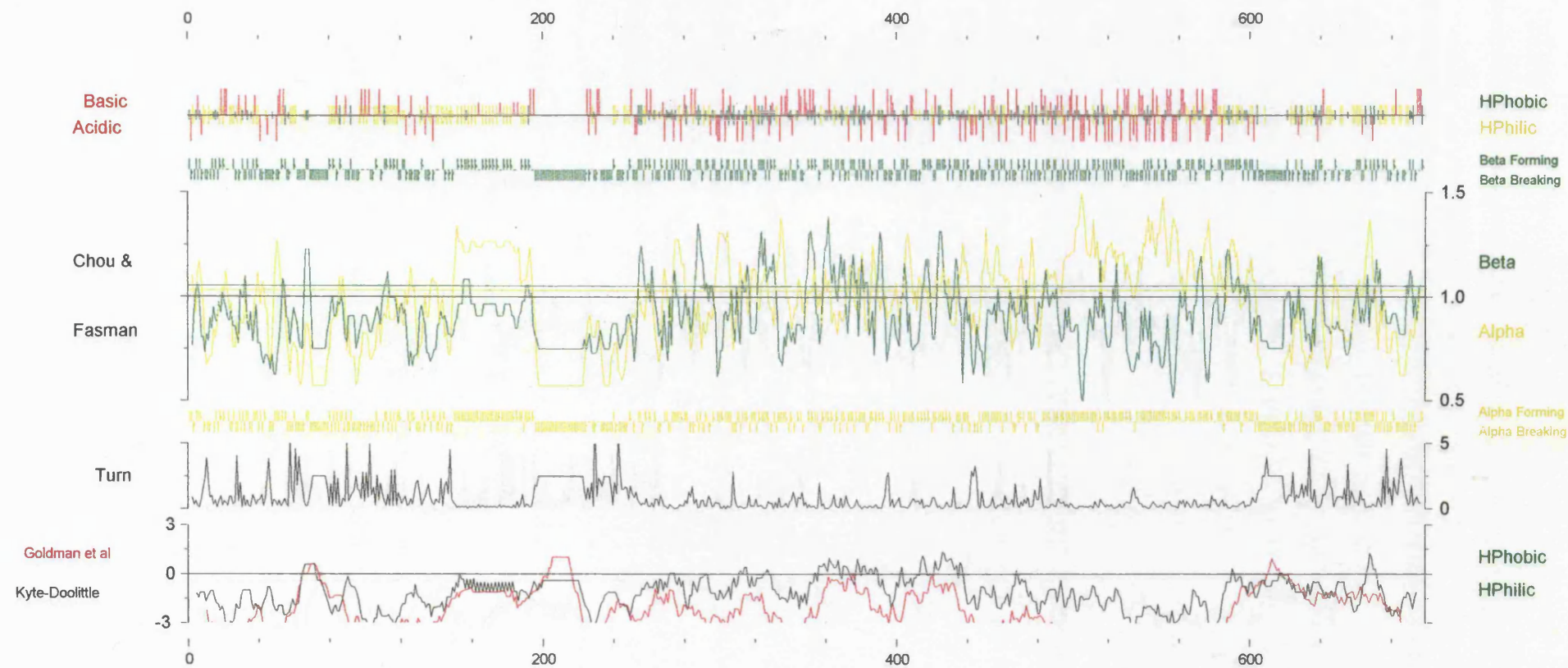


Figure 3.16. Peplot of the *D. virilis* NONA protein.

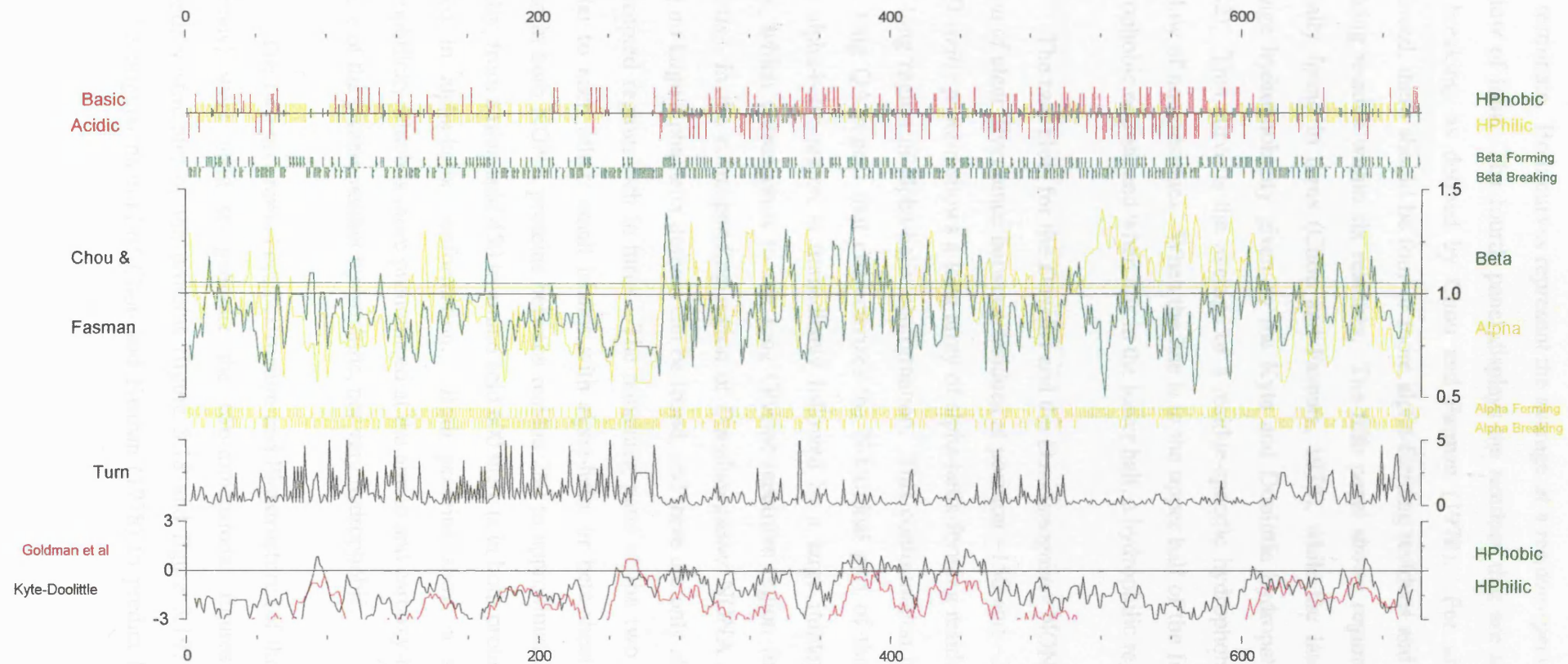


Figure 3.17. Peplot of the *D.melanogaster* NONA protein.

may terminate. Both curves represent the average of a residue-specific attribute over a window of four. The fourth panel displays the residues that are alpha-helix forming and breaking, as defined by Chou and Fasman (1978). For alpha helices to be produced, there should be four or more alpha-forming residues and no more than one breaking residue within six residues. The fifth panel shows regions of the sequence typically found in turns (Chou and Fasman, 1978), while the last panel shows the average hydrophobicity given by the Kyte and Doolittle hydrophathy measure curve (1982). This curve is the average of a residue-specific hydrophobicity index over a window of nine residues. When the line is in the upper half of the frame, it indicates a hydrophobic region, and when it is in the lower half, a hydrophilic region.

The two plots for the *D.virilis* and the *D.melanogaster* NONA proteins show a region of clear divergence between residues at position ~150 and ~250. In this region the *D.virilis* protein shows a long array of alpha-helix-forming residues which give rise to a long region of alpha-helix conformation. This conformation is generated by the very long QA repeat that characterizes the N-terminal part of the *D.virilis* protein. This alpha-helix region is immediately followed by a large uninterrupted domain of turns, which corresponds to the long Glycine-repetitive region found in NONA of *D.virilis*. In the corresponding region of *D.melanogaster* NONA (amino acids 150-250) no large alpha helix domain can be found, and there are only short and frequently interrupted regions rich in turns. The remaining parts of the two proteins are more similar to each other: small islands with alpha-helix or beta-sheet conformation are found in both NONA proteins between residue 280 to approximately amino acid 450. Finally, from amino acid 450 to amino acid 600 there is in both proteins a large domain folded in alpha-helix conformation. Both proteins show a similar pattern of hydrophilicity which is more pronounced at the amino and carboxy-termini, leaving the middle of the protein neither hydrophilic, nor very hydrophobic.

The two programs Peptidestructure and Plotstructure (of the GCG packet, see methods) were used to generate the two-dimensional figures representing the secondary structure of the proteins (figure 3.18 and figure 3.19). Peptidestructure uses the original method of Chou and Fasman (1978) to predict helices, sheets and

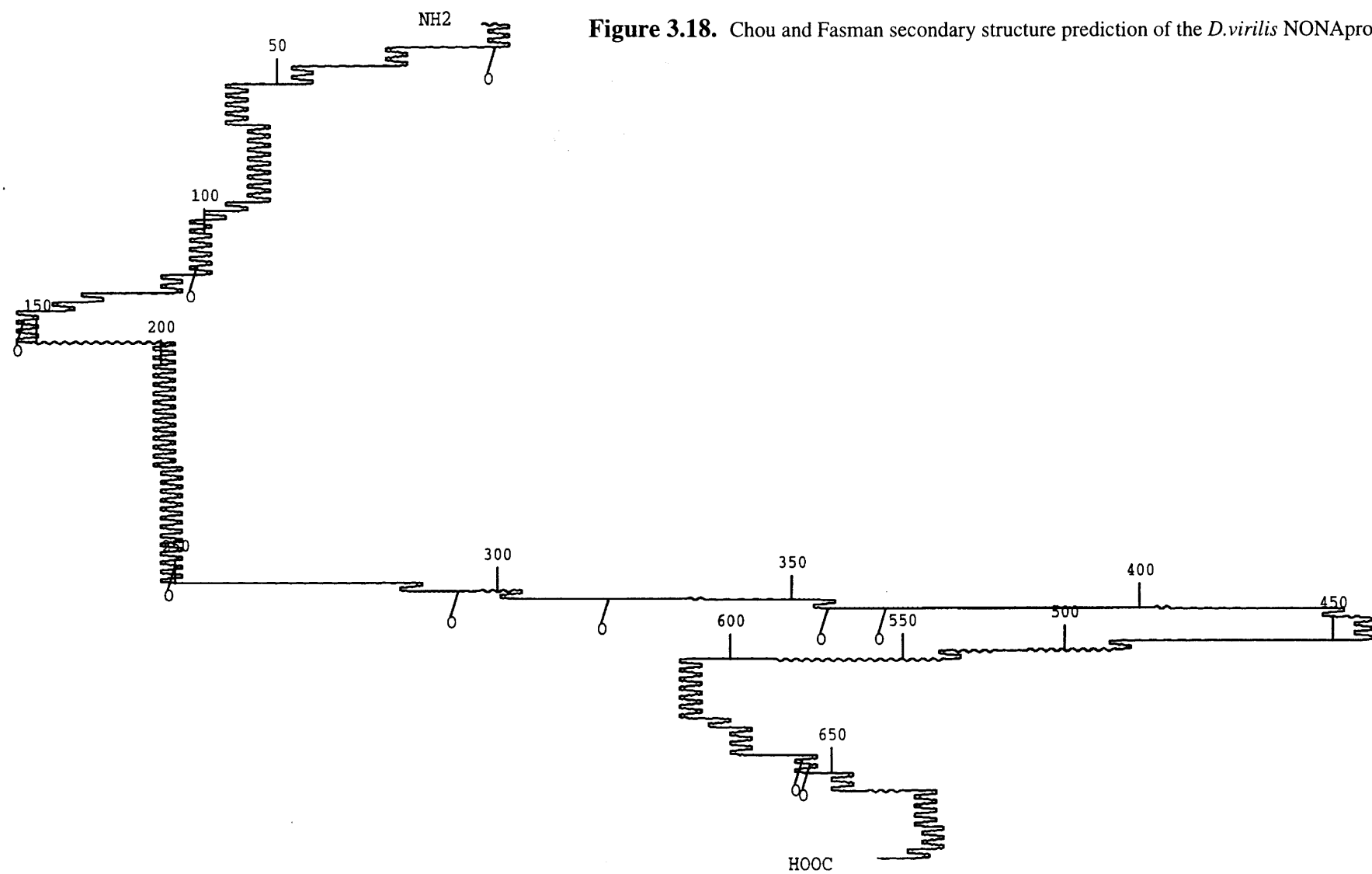
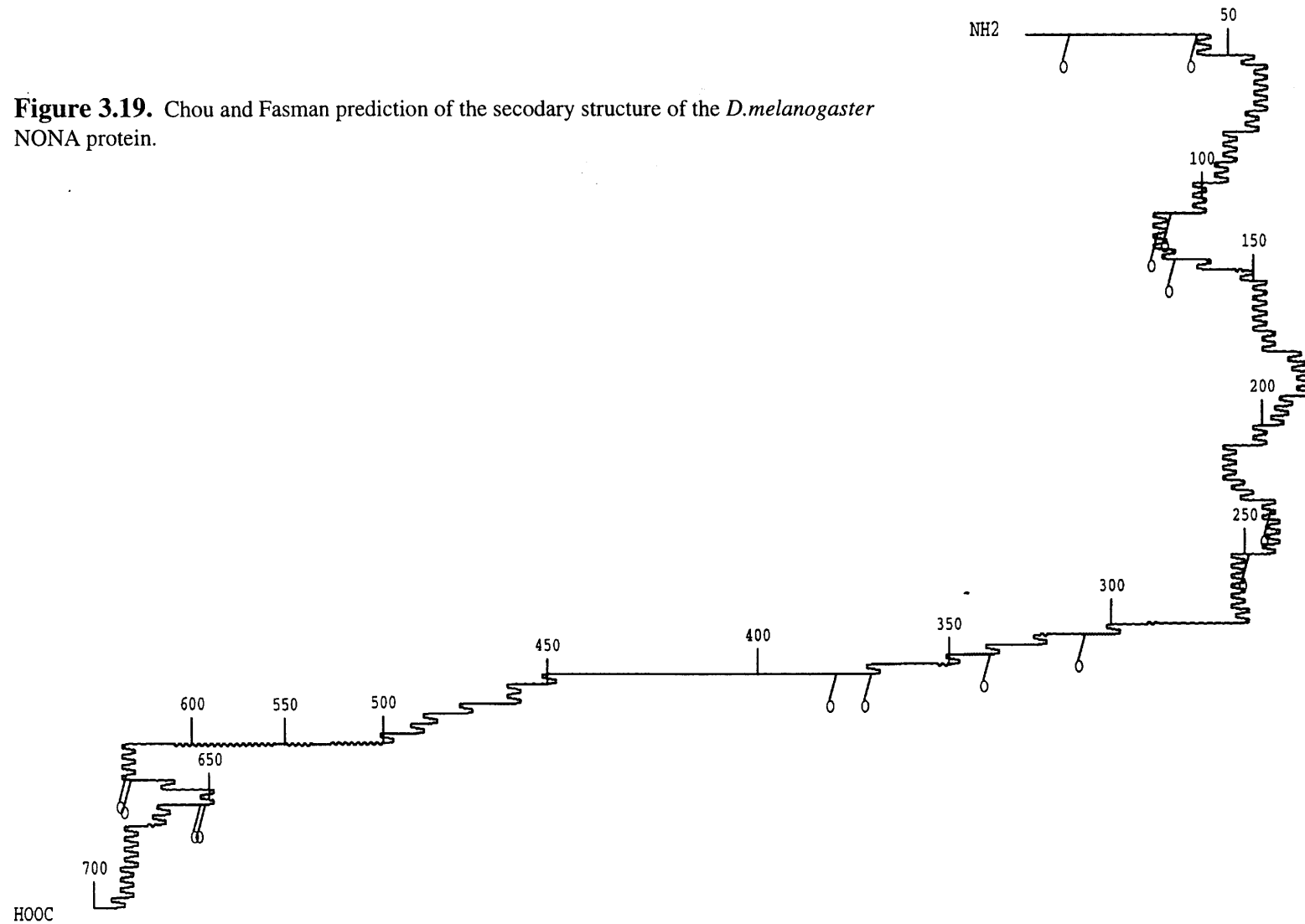


Figure 3.18. Chou and Fasman secondary structure prediction of the *D. virilis* NONAprotein.

Figure 3.19. Chou and Fasman prediction of the secondary structure of the *D.melanogaster* NONA protein.



turns. The minimum length of an alpha-helix is six residues and of a beta-sheet is four. Regions without adequate predictions are replaced by the conformational state of the next best probability. Glycosylation sites are also predicted for sites where the residues have the composition NXT or NXS. When X is D, W, or P, the site is taken to be a weak glycosylation site, otherwise it is a strong glycosylation site. The Plotstructure program plots the measures of secondary structure from the output file of Peptidestructure into a two dimensional representation. The two-dimensional pictures produced for the *D.virilis* and *D.melanogaster* NONAs (figure 3.18 and 3.19) nicely shows the striking difference between the two proteins from residues 150 and 250, where in *D.virilis* NONA the large region of alpha-helix conformation is represented by a sine wave, and the following uninterrupted region full of turns is indicated by 180 degree turns.

The Peptidestructure and Plotstructure programs were also used to obtain graphic representation of the secondary structure of the RNA-binding domain of the two NONA proteins, which are located approximately from amino acid 300 to 600 in each amino acid sequence (appendix 3.4 and 3.5). All RNA-binding structures determined to date reveal a structure composed by a large antiparallel beta-sheet which is sandwiched between two alpha-helices, and follow a general pattern of β - α - β - β - α - β (Siomi and Dreyfuss, 1997). The functionally important structure seems to be the antiparallel beta-sheet which forms a non-sequence specific RNA-binding platform. The secondary structure of the RNA-binding domains in the *D.melanogaster* and *D.virilis* NONA proteins seems to fit with the general conformational rules for RNA-binding proteins just described: an antiparallel beta-sheet is found in both NONAs and it is sandwiched between two alpha-helices domains (appendix 3.4 and 3.5).

Molecular evolution of the *nonA* gene

We have seen that the first half of the NONA protein is quite diverged between *D.virilis* and *D.melanogaster*, while the remaining half maintains a high level of conservation. Although the high conservation of the RNA-binding domains and the

adjacent charged region is not surprising, given that the same domains are functionally important and show conservation in proteins found in such distantly related organisms such as *Drosophila* and humans (Rendahl *et al.*, 1996), the divergence found in the first part of the protein could be explained in two opposite ways. Either the diverged regions of NONA are those regions of the protein which lack functional importance, and are thus free of selective constraints, or, alternatively, they serve species-specific functions and therefore their divergence reflects functional importance which could be the result of adaptive evolution.

Coding DNA sequences can undergo two types of substitutions: non-synonymous and, given the degeneracy of the genetic code, synonymous. Synonymous changes can be considered to be largely neutral, and therefore should reflect the neutral mutation rate, i.e. the ticking of the molecular clock (see Li, 1997). On the contrary, non-synonymous substitutions are likely to be mostly deleterious and therefore subjected to purifying selection. However, if a non-synonymous is advantageous, it will be fixed at a rate faster than the neutral mutation rate. When this happens, the number of non-synonymous substitutions can be greater than the number of the synonymous ones. In other words, if we calculate the number of non-synonymous substitutions (indicated as K_A) and the number of synonymous substitutions (K_S), a ratio K_A/K_S greater than one will indicate that positive selection is acting in maintaining the non-synonymous changes (Li and Graur, 1991)

The number of synonymous and non-synonymous substitutions per site between the *D.virilis* and the *D.melanogaster nonA* genes was calculated with the GCG NewDiverge program. The program is based on a variant of the method of Li *et al.* (1985), as modified by Li (1993) and Pamilo and Bianchi (1993). It uses a translation table to determine codon degeneracies, and applies Kimura's two-parameter method (1980) to correct for multiple hits and to account for the difference in substitution rates for transitions and transversions. The Assemble program of the GCG packet was used to remove DNA regions from each *nonA* sequence which could not be properly aligned, i.e. those regions which generated gaps in the proteins alignment.

The K_A/K_S test was performed first on the whole coding sequence of *nonA*, and then in two separate domains of the sequence: the first including the most diverged part of the genes (from nucleotide 1 to 813 of each newly assembled DNA sequence), and the second constituted by the more conserved part (from nucleotide 814 to the end of each gene).

The K_A/K_S test was performed also on the *D.virilis* and *D.melanogaster* *per*, *tra* and *engrailed* (*en*) homologous genes. Table 3.5 shows the results of the test for these genes, including the results obtained for the *timeless* (*tim*) gene by Ousley *et al.* (1998). It is clear that in all these genes the rate of non-synonymous substitutions is occurring at a much lower rate than the synonymous level. It is interesting that *tra*, known for the high divergence it displays across different species (O'Neil and Belote, 1992), shows the highest K_A/K_S of all the genes here studied. The *nonA* gene displays a lower ratio than *tra* even when the most diverged part of the gene only is taken into consideration.

Table 3.5. K_S , K_A , and K_A/K_S values calculated for 4 *D.melanogaster* and *D.virilis* homologous genes. The columns "*nonA* c" and "*nonA* d" show these values calculated only for the conserved (c) and the diverged (d) part of the *nonA* gene respectively. The column "*nonA*" gives the values for the entire *nonA* gene.

	<i>nonA</i> c	<i>nonA</i> d	<i>nonA</i>	<i>en</i>	<i>per</i>	<i>tim</i>	<i>tra</i>
K_S	1.045	2.155	1.242	1.035	1.234	1.264	1.515
K_A	0.056	0.458	0.190	0.130	0.243	0.102	0.493
K_A/K_S	0.054	0.213	0.153	0.126	0.197	0.081	0.325

It has to be pointed out that the K_S measure is likely to be an underestimate of the neutral mutation rate. It has been shown, in fact, that preferences in the usage of codons might exist, especially for highly expressed genes (Shiels *et al.*, 1988, Sharp *et al.*, 1988), these preferences being driven probably by the availability of different types of tRNAs (Kurland, 1987; Sharp *et al.*, 1988). Biases in codon usage are inversely

proportional to the silent mutation rate, and have been shown to slow down the rate of silent substitutions by as much as one half in some genes (Sharp and Li, 1989).

Rapid divergence driven by positive selection has rarely been demonstrated at the molecular level. One of the few examples is presented in a recent study on the evolution of the lysozyme gene in primates (Messier and Stewart, 1997). Using the K_A/K_S ratio as a tool to identify positive selection, Messier and Stewart have been able to show that adaptive changes occurred in the lysozyme gene. They identified two separate episodes of positive selection: one on the lineage leading to the common ancestor of the colobine mokeys, and a second, unexpectedly, on the lineage leading to the common ancestor of hominoids.

Genes encoding functional proteins are always expected to be subjected to some degree of purifying selection, eliminating most of the amino acids changes because of their deleterious effects. Whole-gene comparisons of K_A/K_S ratios reflect both positive and negative selective forces, consequently the value of the ratio may often be less than one even if some adaptive substitutions have occurred. For example, when 363 homologous genes were compared between the mouse and rat, only one had a ratio higher than one (Wolfe and Sharp, 1993). It would be surprising however if only one of all those genes had undergone adaptive changes during the divergence of the two species (Sharp, 1997). One way to overcome this problem is to know where to look for adaptive evolution within a gene sequence. Knowing in advance which sites within a protein are good candidates for adaptive changes, would allow focusing only on the corresponding part of the sequence, without including in the analysis many other positions subjected to purifying selection. An example of this is the study of the class I major histocompatibility complex (MHC) genes (Hughes and Nei, 1988). Based on the knowledge of which residues of the protein are part of the antigen recognition site (ARS), Hughes and Nei (1988) were able to show that these sites specifically have a K_A/K_S ratio greater than one, while in the remaining parts of these genes the ratio is always considerably lower. Similarly, analysis of the K_A/K_S ratio in the homeobox gene, *Pem*, has shown that the divergence of this specific domain was driven by adaptive selection (Sutton and Wilkinson, 1997).

It is now recognized that most nuclear genes in *Drosophila* evolve at rates more than two times faster than those in rodents and other mammals (Moriyama, 1987, Moriyama and Gojobori, 1992, Sharp and Li, 1989). Moreover, the rate of synonymous substitutions can vary between different genes of *Drosophila* (Sharp and Li, 1989). The rate of nucleotide substitutions depends on the mutation rate (Kimura, 1983), the generation time (or the number of replications), and the degree of selective constraints on the nucleotide sequence when the changes are not selectively neutral (Moriyama and Gojobori, 1992). One explanation for the higher rate of synonymous substitutions in *Drosophila* compared to mammals could be the shorter generation time of *Drosophila*, which implies more DNA replication per year (Britten, 1986; Li and Tanimura, 1987; Kimura, 1983; 1991). Differences in the accuracy of DNA replication and efficiencies in DNA repair could also account for the different substitution rates between genes of *Drosophila* and mammals (Moriyama and Gojobori, 1992).

A consistent explanation for the high and variable synonymous substitution rates in *Drosophila* has not been provided so far. However several pieces of evidence suggest that the rates of synonymous substitutions are correlated with the base composition of the nuclear genes. It seems that the G+C and C content at the third codon position of nuclear genes is negatively correlated with synonymous substitution rates, while the A content is positively correlated (Moriyama and Gojobori, 1992). Moriyama and Gojobori propose two possible explanations for the observed relationships. The first is that spontaneous mutations directed from C to A or T may frequently occur in the *Drosophila* nuclear genome. As a result, A and T have accumulated and C has been reduced at those sites under low functional constraints. This hypothesis was supported by data obtained from pseudogenes which indicated an accumulation of A and T in third positions (Moriyama and Gojobori, 1992). The second possible explanation is that mutations that cause accumulation of A or T in the coding regions of *Drosophila* nuclear genes may be subjected to some kind of selection. Consequently, genes under higher selective constraints may have a more biased codon usage (presumably towards G+C), as described by Sharp and Li (1989).

Using an estimate of the divergence time between *D.melanogaster* and *D.virilis* of 40 million years (Moriyama, 1987), the rate of synonymous substitutions was calculated to be 15.52×10^{-9} /site x year for *nonA*, 15.43×10^{-9} /site x year for *per*, 18.94×10^{-9} /site x year for *tra*, 15.8×10^{-9} /site x year for *tim* and 12.94×10^{-9} /site x year for *en*. Moriyama and Gojobori (1992) had already calculated the rate of synonymous substitutions in *per* to be about 16×10^{-9} /site x year after comparing the sequences of *D.virilis* and *D.melanogaster*. The small difference with my calculations for *per* are likely to be due to different methods used to align the two sequences and to calculate the synonymous substitutions. Moriyama and Gojobori (1992) divided the *Drosophila* genes they studied into three groups according to their average synonymous substitution rate. These three groups have an average synonymous substitution rate of 11×10^{-9} (group I), 17.5×10^{-9} (group II) and 27.1×10^{-9} (group III) respectively. All the genes included in my study, except maybe *en*, can be included in group II, and seem to be evolving at a rate determined by a similar molecular clock.

Codon usage

If synonymous mutations are selectively neutral, we should expect all the synonymous codons for one amino acid to be used with more or less equal frequencies. However since the first studies of Grantham *et. al.* (1980, 1981), it has become apparent that the usage of codons is not random. Moreover related species show similar patterns of choice between synonymous codons, i.e. codon biases seem to be species specific (Grantham *et. al.* 1980). In most species, however, there is also a considerable difference in codon usage between different genes (Sharp *et. al.*, 1988).

Biases in codon usage could constrain the rate of silent substitutions, and in fact it has been shown that the rate of synonymous substitutions is negatively correlated with the degree of codon usage bias (Sharp and Li, 1989). Sharp and Li (1989) have concluded that the variation in silent substitution rates among different nuclear genes in *Drosophila* is mainly due to differences in the strength of selective constraints on synonymous codons. In addition to this, Moriyama and Gojobori

(1992) found that the rates of synonymous substitutions are also correlated negatively with the C content and positively with the A content at the third codon positions (see above).

Here, the codon usage patterns were studied in ten homologous genes of *D.melanogaster* and *D.virilis*: *nonA*, *en*, *hb*, *knirps*, *nanos* (*nos*), *oskar* (*osk*), *per*, *runt*, *sevenless* (*sev*), and *tra*. The Codonfrequency program of the GCG packet was used to count codons and write their frequencies into codon frequency tables. Table 3.6 shows, for each amino acid in each *D.virilis* gene, which codon was mostly used. The biased codons for each gene of *D.melanogaster* are shown between brackets only when different from those used by *D.virilis*. For amino acids encoded by only two different codons, a gap means that no preference was shown in the use of one or the other.

The Relative Synonymous Codon Usage (RSCU) has been defined by Sharp *et al.* (1988) as the observed number of codons divided by the number expected if all codons for that amino acid were used equally. In table 3.6, codons of the *D.virilis* genes studied with a RSCU greater than 1.5 are marked by one asterisk, those with a RSCU greater than 2 are indicated with two asterisks.

It is immediately clear that the codon usage for *nonA* differs from that of other genes. Firstly, with the only exception of *tra*, *nonA* is the gene which displays the larger difference in codon usage between *D.virilis* and *D.melanogaster*. In *nonA*, in 10 amino acids out of 18, a different codon is used in *D.virilis* compared to *D.melanogaster*. In all the other genes always fewer than 8 amino acids show a different codon usage in the two species, the only exception being *tra* in which 11 amino acids present different codon bias between *D.virilis* and *D.melanogaster*.

Another striking feature of the *nonA* sequence, is the high A/T content found in the third codon position. Table 3.7 shows the percentage, within the number of biased codons, of codons that end with A or T in all the 10 *D.virilis* and *D.melanogaster* genes analyzed. Overall in both *D.virilis* and *D.melanogaster*, it appears that A and T ending codons form a substantial proportion of biased codons only in *nonA* and *tra*. If

Table 3.6. Codon bias table for 10 *D.virilis* (in black) and *D.melanogaster* (in red and brackets) genes. The biased codons of *D.melanogaster* are shown only when different from those of *D.virilis*. The number below each amino acid indicates the corresponding number of degenerate codons. For amino acids coded by two different codons, a gap indicates no bias for one or the other. For *D.virilis* only (in black) biased codons with a RSCU >1.5 are marked (*), codons with a RSCU >2 are also marked (**).

	<i>nonA</i>	<i>en</i>	<i>hb</i>	<i>knirps</i>	<i>nos</i>	<i>osk</i>	<i>per</i>	<i>runt</i>	<i>sev</i>	<i>tra</i>
Gly 4	GGC* (GGT)	GGC** (GGA)	GGC**	GGC**	GGC*	GGC**	GGC**	GGC**	GGC** (GGA)	GGC* (GGG/G GA)
Glu 2	GAG	GAG*	GAG*	GAG	GAA (GAG)	GAG (--)	GAG*	GAG	GAG*	GAG (GAA)
Asp 2	GAT (GAC)	GAC (GAT)	GAC (GAT)	GAT	GAC (GAC)	GAC (GAT)	GAT (GAC)	GAC	GAT	-- (GAT)
Val 4	GTT**	GTC** (GTG)	GTG**	GTG**	GTG*	GTG**	GTC*	GTG**	GTG**	GTG** (GTG/G TT)
Ala 4	GCC*	GCC**	GCC*	GCC*	GCA (GCG/G CC)	GCA* (GCC)	GCC*	GCC**	GCC*	GCG/GC C*
Arg 6	CGT** (CGC)	CGC**	CGC**	CGC**	CGA** (CGC)	CGC**	CGC**	CGC**	CGC**	CGC* (AGA)
Ser 6	TCC* (AGC)	TCC*	AGC**	AGC**	TCG	AGC/TC C (AGC)	AGC** (TCC)	AGC**	AGC**	AGC*
Lys 2	AAA (AAG)	AAG*	AAG*	AAG	AAG	AAG	AAG	AAG	AAG	AAG
Asn 2	AAT*	AAC	AAC (--)	AAC	AAC	AAC	-- (AAC)	AAC	AAT	AAT** (--)
Ile 3	ATT (ATT/AT C)	ATC**	ATC*	ATT (ATC)	ATT* (ATC)	ATC	ATC*	ATC**	ATC*	ATA/AT T* (ATC)
Thr 4	ACA** (ACC)	ACC*	ACG** (ACC)	ACC*	ACG/AC T (ACC)	ACG (ACC)	ACG	ACG* (ACC)	ACG* (ACC)	ACA** (ACC)
Cys 2	TGC* (--)	TGC*	TGC*	TGC*	TGT (TGC)	TGC	TGC*	TGC**	TGC	TGC**
Tyr 2	TAT	TAC	TAC	TAT (TAC)	TAC	TAT*	TAT (TAC)	TAT	TAC	TAT (TAC)
Leu 6	TTG* (CTG)	CTG**	CTG**	CTG*	CTG**	CTG**	CTG**	CTG**	CTG**	CTG** (TTG)
Phe 2	TTT (TTC)	TTC	TTC	TTC	TTT	TTC (--)	TTC	TTC*	TTT (TTC)	TTT
Gln 2	--	CAG*	CAG	CAG	CAG	CAG	CAG*	CAG	CAG*	CAG* (--)
His 2	CAT*	CAC	CAC	CAC	CAT (CAC)	CAC	-- (CAC)	CAT (CAC)	CAT (CAC)	CAT*
Pro 4	CCG*	CCG* (CCC)	CCG* (CCC)	CCG* (CCC)	CCC	CCA*	CCG/CC C* (CCG)	CCG* (CCC)	CCG**	CCG**

Table 3.7. For each *D.virilis* and *D.melanogaster* gene, the number and the percentage of biased codons ending in A or T is indicated.

		BIASED CODONS	A/T-ending	% A/T-ending
nonA	<i>vir</i>	17	10	58.8
	<i>mel</i>	16	6	37.5
en	<i>vir</i>	18	0	0
	<i>mel</i>	18	2	11.1
hb	<i>vir</i>	18	0	0
	<i>mel</i>	17	1	5.9
knirps	<i>vir</i>	18	3	16.7
	<i>mel</i>	18	1	5.6
nos	<i>vir</i>	18	8	44.4
	<i>mel</i>	18	1	5.6
osk	<i>vir</i>	18	3	16.7
	<i>mel</i>	16	3	18.8
per	<i>vir</i>	16	2	12.5
	<i>mel</i>	18	0	0
runt	<i>vir</i>	18	2	11.1
	<i>mel</i>	18	1	5.6
sev	<i>vir</i>	18	4	22.2
	<i>mel</i>	18	3	16.7
tra	<i>vir</i>	17	7	41.2
	<i>mel</i>	16	7	43.8

Table 3.8. Number of codons with RSCU >1 and >2 for each *D.virilis* gene.

	<i>nonA</i>	<i>en</i>	<i>hb</i>	<i>knirps</i>	<i>nos</i>	<i>osk</i>	<i>per</i>	<i>runt</i>	<i>sev</i>	<i>tra</i>
> 1.5	8	7	6	5	3	3	7	3	5	7
> 2	3	6	6	4	2	4	4	8	6	6

we focus on *D.virilis* only, we can add *nos* to this list. This unusual result could be interpreted to mean a lack of selective constraint on *nonA*, leading to the expression of the A/T mutation bias, as suggested by Moriyama and Gojobori (1992; see above). This should be reflected in a higher synonymous mutation rate for *nonA*. However, as described above, *nonA* does not seem to have elevated mutation rate compared to the other genes examined (*per*, *tra*, *tim* and *en*). Moreover, *tra* has a slightly lower number of A/T-ending biased codons compared to *nonA*, yet, instead of having a lower (or at least similar) synonymous mutation rate compared to *nonA*, it has an higher one. These observations therefore do not fit very well with those presented by Moriyama and Gojobori (1992).

Finally, to further complicate the picture, the *D.virilis nonA* pattern of codon usage seems to be very different from that of the other *D.virilis* and *D.melanogaster* genes (see table 3.6). In *D.virilis nonA* the codons GTT for Valine, CGT for Arginine, and ACA for Threonine have a RSCU greater than 2, meaning a very strong bias. Interestingly, *D.melanogaster* always prefers other codons for Arginine and Threonine in *nonA*. Similarly, in all the other genes studied, either in *D.virilis* or in *D.melanogaster*, the codons used for Val, Arg, and Thr are almost always different from those used in *nonA* of *D.virilis*. There could be an explanation for this if Val, Arg, and Thr were highly repeated amino acids, since repetition of a single amino acid could largely inflate the proportion of the codon encoding for it. However the *D.virilis nonA* amino acid sequence data does not support this hypothesis, since V, R, and T are not involved in large repeats. Similarly, the results from the cryptic symplicity analysis do not indicate the presence of any short repeat containing the motifs: GTT, CGT, or ACA. In contrast to these observations, those amino acids which in *D.virilis nonA* are actually involved in repeats, such as Gln, Ala, His and Gly, show a codon bias similar to that all the other *D.virilis* and *D.melanogaster* genes examined.

Biases in codon usage can result from a combination of several factors, for example mutation trends, translational selection among synonymous codons, or selection for particular structures in DNA. In general, highly expressed genes show

very strong biases towards a particular subset of codons, while lowly expressed genes show a more even codon usage (Sharp *et al.*, 1988). In *E.coli*, the favoured codons in highly expressed genes are those recognized by the most abundant tRNA species (Gouy and Gautier, 1982; Ikemura, 1985). In contrast, in lowly expressed genes, the deviation from equal usage of synonymous codons is most likely to be a result of mutation trends (Sharp and Li, 1986; Bulmer, 1988). To have an idea of the strength of the codon bias in the 10 *D.virilis* genes studied here, the number of codons with a RSCU >1.5 and >2 was counted (see table 3.8). From table 3.8 it is clear that *nos* and *osk* have the lowest bias strength of all. Interestingly, *nonA*, with the only exception of *nos*, has the lowest number of codons with a RSCU >2, showing therefore a rather poor bias strength. If we assume that strength in codon bias reflects the level of expression of a gene, then, from the data presented here, the results (table 3.8) predict that *nonA* is not as highly expressed as most of the other genes included in the study. Any differences in *nonA*'s codon usage would therefore simply be the result of mutational forces, and not selection based on tRNA availability. If this is the case, then these forces should give different codon usage patterns also to *nos* and *osk* (which also have weak bias - see table 3.8), compared to the other genes. However this is clearly not the case (see table 3.7), and *nonA*'s unusual bias for codons GTT (Val), GCT (Arg), ACA (Thr), TTG (Leu) and AAA (Lys) (the first four having a RSCU >1.5), remains something of an enigma.

In conclusion, we have seen that the *D.virilis nonA* gene has diverged considerably from that of *D.melanogaster*, particularly in the N-terminal half. The next step is to find out whether this divergence has any repercussions for the biological functions of the *nonA* gene. In other words, will the *D.virilis nonA* gene rescue the behavioural defects of the *D.melanogaster nonA* mutant? O'Neil and Belote (1992) showed, for example, that the *D.virilis tra* gene, in spite of being highly diverged from its *D.melanogaster* orthologue, was able to restore "femaleness" to diplo-X *D.melanogaster tra* pseudomales. If the *D.virilis nonA* gene can restore normal song and visual behaviour to *nonA* mutants, then it will be possible to assess whether the

transformants show any evidence for *D.virilis*-like song behaviour. In the case of the *per* gene, which involved transforming the *D.simulans per* gene into arrhythmic *per*⁰ *melanogaster* hosts, robust interspecific rescue was obtained for the arrhythmic song phenotype, and the *simulans* characteristics were transferred to the host (Wheeler *et al.*, 1991). However, *D.simulans* diverged from *D.melanogaster* about two million years ago (Cariou, 1987), whereas the *D.virilis/D.melanogaster* relationship is far more remote (Moriyama, 1987; Beverly and Wilson, 1984). The following chapter describes how the *D.virilis nonA* gene was cloned into a vector and transformed into *D.melanogaster* hosts and includes the genetic and molecular analysis of the various lines that were generated.

CHAPTER 4

Transformation of $nonA^{vir}$

INTRODUCTION

In chapter 3 I described how the *D.virilis* genomic clone containing the putative *nonA^{vir}* gene was subcloned in two fragments into the pUC18 vector. In order to transform *D.melanogaster* flies with the *nonA^{vir}* gene it was necessary to rejoin the two DNA fragments by cloning them into a suitable transformation vector. The *P* element vector pW8 (Klemenzen *et al.*, 1987) was preferred among others (Ashburner, 1989) for its large polylinker containing 12 unique restriction sites. This vector, derived from the *P* element end containing construct Carnegie4 (Rubin and Spradling, 1983), carries, as a marker, the *white⁺* gene fused to the *heat shock protein 70 (hsp70)* gene promoter. Expression of the *w⁺* gene completely or partially rescues the lack of red pigment in the eyes of the recipient transformant *w* flies.

METHODS

Figure 4.1 schematically illustrates the procedure used to join the two *nonA^{vir}* DNA fragments into the transformation vector pW8. As a first step, the ~6.5 Kb *nonA^{vir}* DNA fragment in pUC39 was shortened of ~0.5 Kb by cutting with *XbaI* which recognizes a site in the pUC polylinker and one at about 0.5 Kb from the 3' end of the *nonA^{vir}* DNA insert, and *SphI*. The released ~6 kb *nonA^{vir}* band was gel purified and subcloned into the *SphI-XbaI* sites of pUC18 (see chapter 2). This new construct was named pUC23. The 5' and 3' ends of the insert in pUC23 were sequenced using the pUC Universal and Reverse primers, in order to make sure that the 0.5 Kb fragment had been actually removed from the correct end of the insert. The next step was to clone the *SphI-XbaI* DNA fragment from pUC23 into the unique *SphI* and *XbaI* sites of the pW8 polylinker. To do so, the 6 Kb *SphI-XbaI* fragment was released from pUC23 by digesting with *SphI* and *XbaI*, gel purified, and ligated to pW8 digested with the same enzymes. Transformation of the construct in *E.coli* cells was

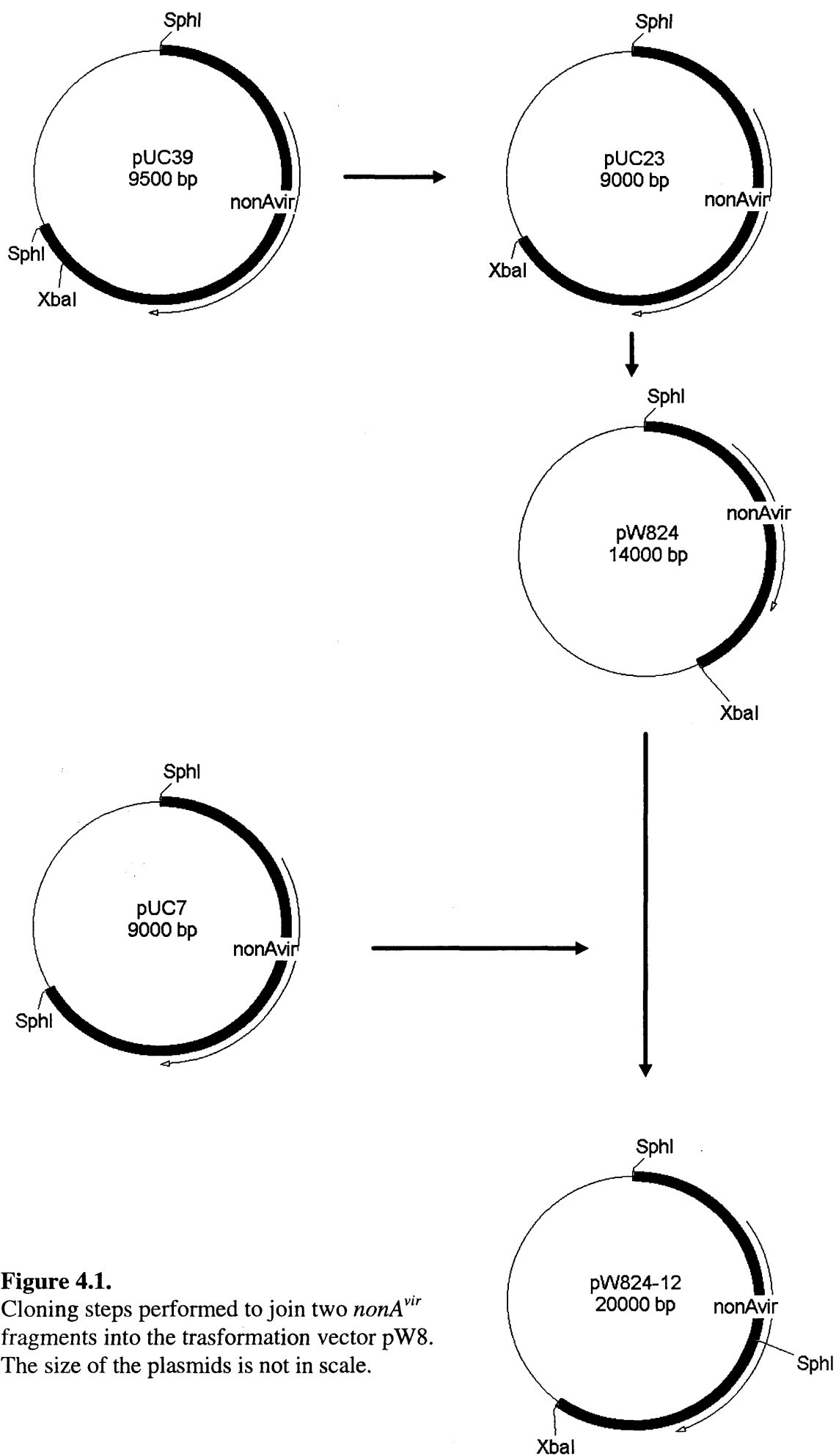


Figure 4.1.
Cloning steps performed to join two *nonA^{vir}* fragments into the transformation vector pW8. The size of the plasmids is not in scale.

performed by electroporation as described in chapter 2. This plasmid was called pW824.

In the second step the ~6 Kb *nonA*^{vir} DNA fragment contained in pUC7 was excised by digesting with *SphI* and cloned into the *SphI* site of construct pW824. Two insert-containing positive clones were obtained (pW824-12; pW824-25), and a restriction digestion confirmed that in both constructs the *SphI-SphI* fragment had inserted in the appropriate orientation reconstituting the intact *nonA*^{vir} gene. To further check whether no frame-shift mutations had occurred at the two inserts' junction site, the DNA from both pW824-12 and pW824-25 constructs was manually sequenced (as in chapter 2) using primers annealing upstream and downstream of the *SphI* junction. Large scale DNA preparation was made of construct pW824-12 (the procedure was described in chapter 2) and this DNA was used to transform flies as reported in chapter 2. From now on this construct will be referred to as the *nonA*^{vir} construct.

Transforming *w; e Sb Δ2-3/TM6* flies

When flies from the *w; e Δ2-3 Sb/TM6* stock were used as the injectees of the *nonA*^{vir} construct the source of transposase was provided by a stable P-element inserted on the third chromosome. This element, constructed *in vitro* by Laski *et al.* (1986) is missing the third (2-3) intron of the *P* element and although it is very stable, it causes mobilization of other elements at unusually high frequency (Robertson *et al.*, 1988). The injection procedures are described in chapter 2. The *nonA*^{vir} DNA was precipitated and resuspended in injection buffer (5 mM KCl and 0.1 mM NaH₂PO₄, pH 6.8) at a concentration of 600 ng/μl.

Transforming *w*¹¹¹⁸ flies

The helper plasmid pUCHSΠΔ2-3 (a gift from J-M. Dura) was used as a transposase source when the injected flies were *w*¹¹¹⁸. The DNA concentration of

nonA^{vir} construct: helper plasmid was 300:150 ng/μl in injection buffer. G₀ flies were backcrossed to the *w¹¹¹⁸* stock and their progeny (G₁) screened for transformants (marked by the red-eye colour).

Southern blotting of transformed lines

The DNA from 50 flies of each transformed line was extracted using the method described in chapter 2. The DNA was digested with the appropriate restriction enzyme, run overnight on a 0.8% agarose gel and blotted according to the methods in chapter 2. The probe to be used in the standard hybridization step was the 5' *P* element end of pW8. This 600 bp DNA fragment was released from pW8 after digesting with *Hind*III-*Xba*I, and, after gel purification, 20ng of DNA were used for the hybridization as described in chapter 2. A probe annealing to the 3' *P* element end of pW8 was used in some cases. This probe was generated by PCR amplification of a 600 bp fragment from a *D.melanogaster* 2.9 Kb *P* element. However only ~180 bp of this probe actually annealed to the pW8 3' *P* element end making the hybridization signal rather poor.

Mapping the inserts

The crosses performed in order to simultaneously map and balance the transgenes are presented in figure 4.2. Flies of the double-balanced stock *w; T2:3 ap^{Xa}/SM5CyTM3Sb* were crossed to flies from each transformant line. From the F₁, red-eyed individuals carrying the transgene and both balancers (marked with *Cy* and *Sb*) were crossed to double-balanced white-eyed flies. If the insert maps on the second chromosome, white-eyed, curly wing (*w; Cy*) F₂ flies are produced among the progeny. The proportion of the F₂ carrying these informative markers should be about 1/6 of the total progeny. If the insert is on the third chromosome, among the F₂ progeny should be flies carrying both *w* and *Sb*.

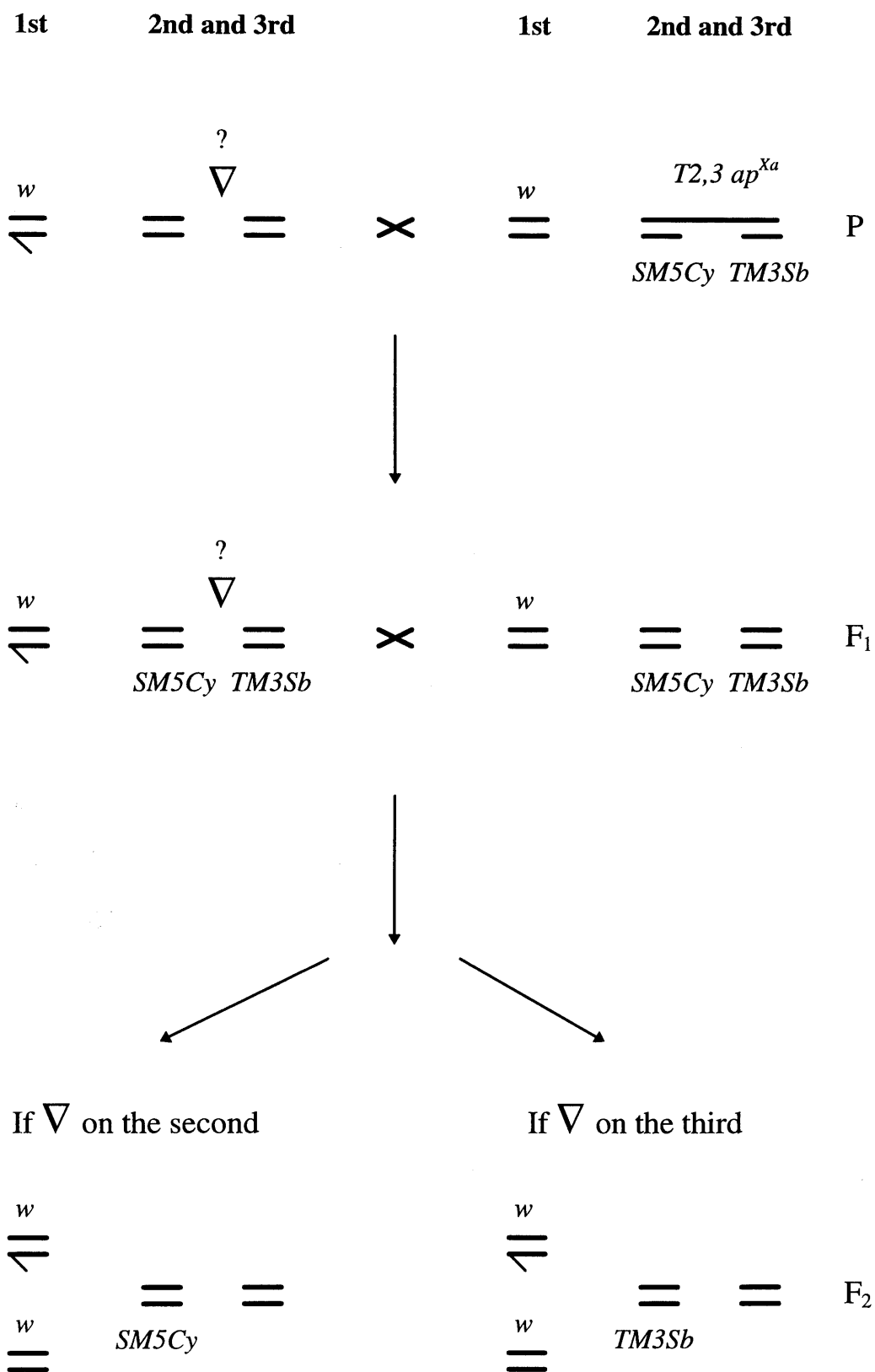


Figure 4.2.

Crosses for mapping transgenic inserts (▽).

Mobilization of X-linked inserts

Transgenes that map to the first chromosome can be mobilized to other chromosome locations if a source of transposase is added in their genome. In order to do this, transformant females were crossed to males of the *w; Sb e Δ2-3/TM6* stock (figure 4.3). Red-eyed males of the progeny marked with *Sb* carry the transposase-producing Δ2-3 element on their third chromosome. Transposase activity promotes mobilization of the insert in the germ cells of these males as well as in their somatic tissues. As a result some of these males displayed a red and white mosaic-eyed phenotype indicating that in some cells the insert had been lost. All the *w⁺; Sb e Δ2-3* males were singularly crossed to *w¹¹¹⁸* females and the male progeny screened for the presence of *w⁺* which would indicate a jump to an autosomal location.

RESULTS

Lines 112 and 113

Transformant lines 112 and 113 were obtained from ~180 injected G₀ *w; Sb e Δ2-3/TM6* flies. The insert in 112 was found to be homozygous viable on the X chromosome. Several extensive attempts to jump this insert into an autosomal location, using crosses to *w; Sb e Δ2-3/TM6* flies as described above, failed. The position of insert 112 on the X chromosome was roughly estimated by crossing 112 flies to a *yellow, white* (*y w*) strain and backcrossing the F₁ females to *y w* males. The number of recombinant flies (333 *y w⁺*, 400 *y⁺ w*) over the total number of F₂ flies (1572) gave an approximate map position of 1-47, very close to the estimated position of the endogenous *nonA* locus (1-52, Kulkarni *et al.*, 1988). An *in situ* hybridization to polytene chromosomes of line 112 was performed by Federica Sandrelli at the University of Padova, using a ~3 Kb *nonA^{vir}* specific probe generated by digesting the pUC7 construct (see figure 4.1 and chapter 3) with *HindIII*. This probe included the

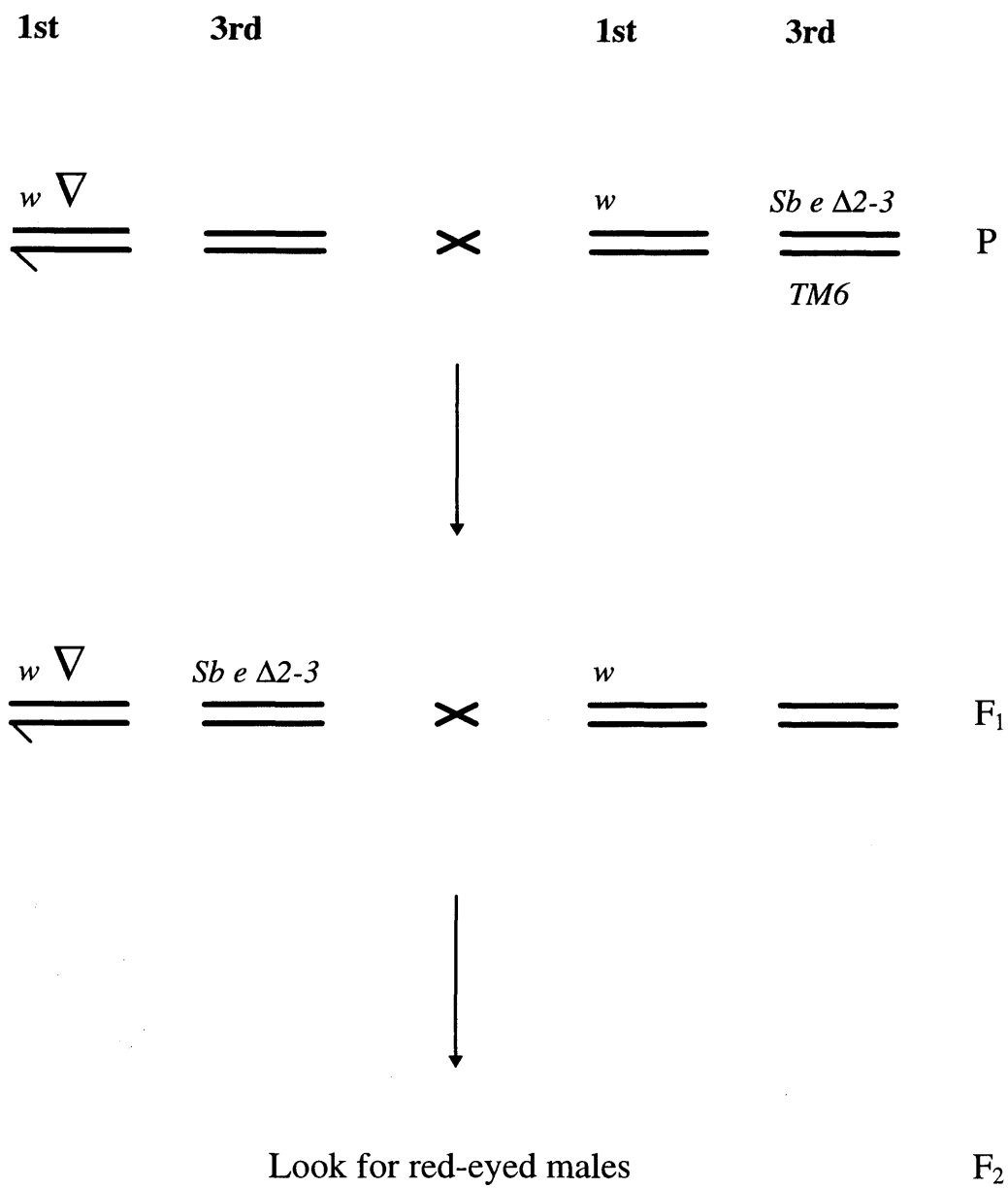


Figure 4.3.

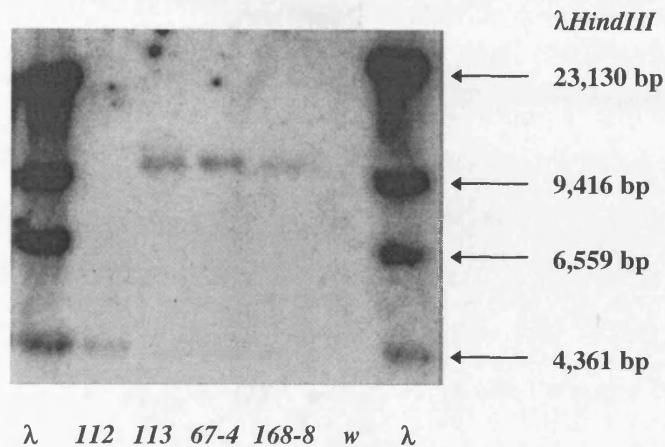
Crosses for mobilizing an X-linked transgene (∇).
 Only the first and the third chromosomes are indicated.

first exon of *nonA^{vir}*, some upstream sequence, and part of the first intron. From the results of the *in situ* hybridization, the *112* insert was mapped to 18F-19-A, again very close to the endogenous *nonA* gene which lies at 14C1-2 by *in situ* hybridization using a cDNA probe (Besser *et al.*, 1990). Given the proximity of *nonA^{vir}* to the endogenous *nonA* in line *112*, attempting to isolate a double cross-over of the *112* insert to an X chromosome carrying a deletion of the *nonA* locus (Stanewsky *et al.*, 1993) or the *nonA^{diss}* mutation seemed impractical. Consequently any behavioural phenotype produced by the insert in line *112* could not be tested on a *nonA* mutant background (see next chapters).

Southern blots of line *112* hybridized with a probe for the 5' *P* element end always revealed the presence of two bands (figure 4.7, 4.10, 4.11) indicating that either the 5' *P* element end had been somehow duplicated during integration of the transgene, or that the X chromosome contained two inserts. However the former seemed to be the most plausible explanation since only one band was found when using hybridization probes annealing to either the 3' *P* element end (figure 4.4) or the 6 Kb 3' half of the *nonA^{vir}* gene (figure 4.5).

The transgene of line *113* integrated on the *Sb e Δ2-3* chromosome. To avoid further mobilization of the insert, the $\Delta 2-3$ element was recombined off the chromosome by crossing red-eyed females to *w* males. To test for the loss of $\Delta 2-3$ the DNA of ten transformant males not marked with *Sb* was individually amplified using primers for the $\Delta 2-3$ element, and run on a gel alongside a positive control (amplified DNA of *w; Sb e Δ2-3/TM6* flies). None of the ten samples displayed a band corresponding to the amplification product of the positive control, indicating the absence of $\Delta 2-3$ (data not shown). Line *113* was initiated from the progeny of only one of these tested males after crossing to *w* females. The insert in line *113* is homozygous lethal, which means that after balancing the transgene-carrying third chromosome with *MKRS* (marked with *Sb*), the only viable *w⁺* flies carry *Sb*.

A



B

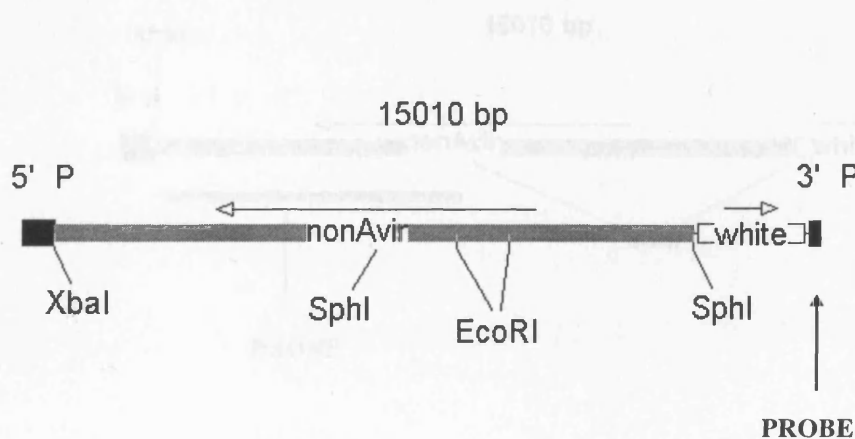
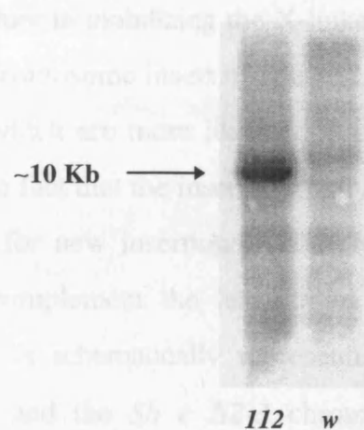


Figure 4.4.

A: Southern blot with *EcoRI* digested DNA from lines 112, 113, 67-4, and 168-8. *EcoRI* digested DNA from *w*¹¹¹⁸ flies (lane *w*) was included as a *P* element-free negative control sample. *HindIII* digested *λ* DNA was used as a marker. The probe used annealed only partially to the 3' *P* element end of pW8 (see methods) making the hybridization signal rather poor. **B:** schematic illustration showing the position and orientation of the *nonA*^{vir} construct in relation to the 5' and 3' *P* element ends of pW8 (named "5' P" and "3' P" in picture). The *white* gene in pW8 is also shown. The arrows above each gene indicate the 5' to 3' orientation. The annealing position of the 3' *P* element probe is also indicated. The construct is ~15 Kb long.

A



B

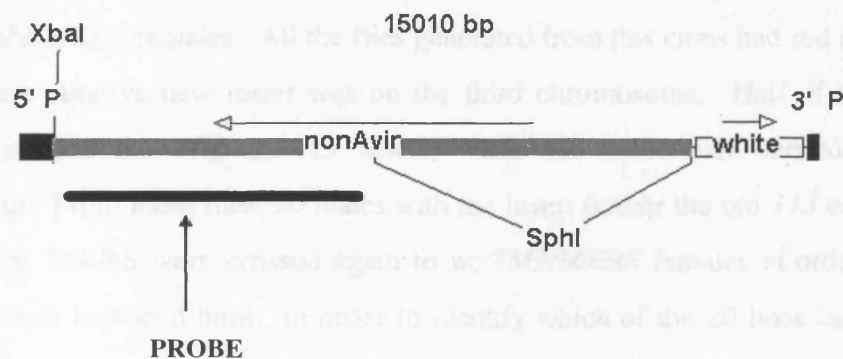


Figure 4.5.

A: Southern blot of *SphI* digested DNA from transformant line *112*. The hybridization probe used was the ~6 Kb *SphI*-*XbaI* fragment of *nonA^{vir}* (picture B). DNA from *w¹¹⁸* flies served as a the negative control in the hybridization. The presence of only one band in the *112* lane suggests that a single transgenic insert is present in this line. **B:** Schematic illustration showing the position and orientation of the *nonA^{vir}* construct in relation to the 5' and 3' P element end of pW8, as in figure 4.4. The annealing position of the probe is indicated by the vertical arrow.

Lines 168-8 and 67-4

After the failure in mobilizing the X-linked *112* insert, an attempt was made to mobilize the third chromosome insert of line *113*. In this case, even local jumps on the same chromosome, which are more likely to occur (Tower *et al.*, 1993), would have been welcomed. The fact that the insert in line *113* is homozygous lethal was used as a means of screening for new insertions. If the transgene moves to another genomic location it should complement the lethality of insert *113*. The cross designed to mobilize insert *113* is schematically represented in figure 4.6. About 200 males carrying insert *113* and the *Sb e Δ2-3* chromosome were individually crossed to females of line *113* (i.e. carrying insert *113* balanced by *MKRS*). The progeny were scored for any red-eyed flies which were not carrying the *Sb* marker. In theory these flies should carry two copies of the transgene: one at the original position of *113*, and one in a novel genomic location. Only one out of the 200 individual crosses (number 168) produced such flies among its F₂. One male with red eyes and no *Sb* marker was crossed to *w; TM2/MKRS* females. All the flies generated from this cross had red eyes, indicating that the putative new insert was on the third chromosome. Half of these flies therefore carried the original *113* insert, while the other half carried the hypothetical jump. From these flies, 20 males with the insert (either the old *113* or the jump) balanced by *MKRS* were crossed again to *w; TM2/MKRS* females in order to generate 20 different balanced lines. In order to identify which of the 20 lines carried the newly mobilized insert, a red-eyed male carrying *MKRS* was taken from each of the lines and crossed back to females of the original *113/MKRS* stock. Those whose progeny gave red-eyed *Sb*⁺ flies must originate from a line containing the insert in a novel position. Only one of these lines (number 168-8) was kept while all the others were discarded. The "new" insert in line *168-8* is homozygous lethal since the progeny of a male and a female carrying *168-8/MKRS* consist only of *w*⁺ *Sb* flies.

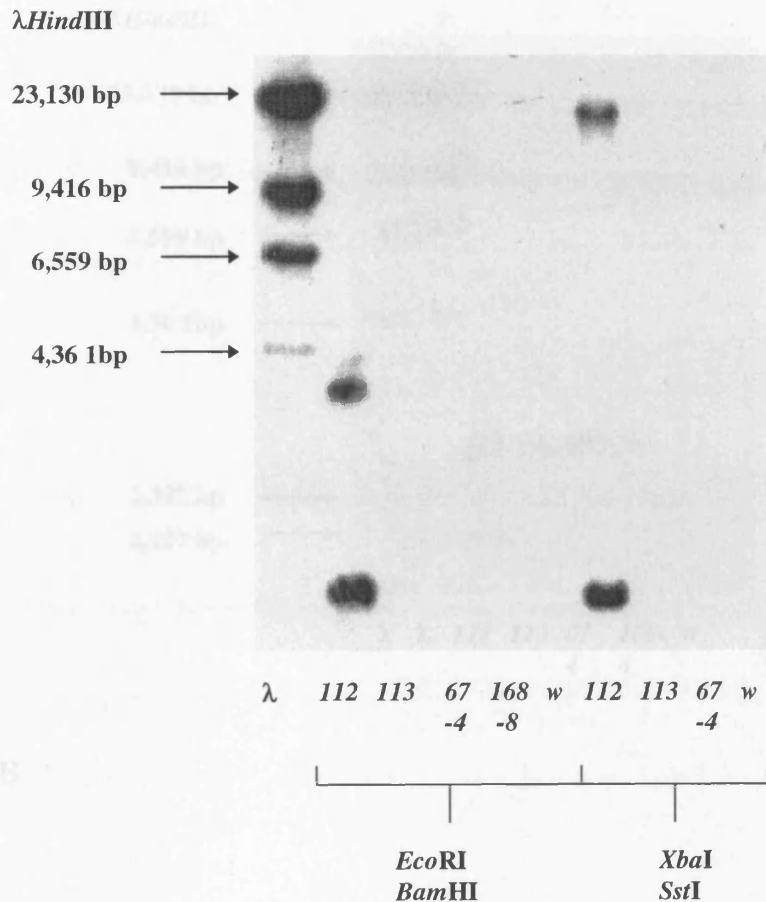
Line *67-4* was also generated from the cross described in figure 4.6. Some flies among the progeny generated by one male (numbered 67), carrying the *113* insert and *Δ2-3*, crossed to *113/MKRS* females, displayed a strikingly stronger eye colour than the parental one. However all these flies also carried *Sb* indicating that their insert was not complementing the lethality of insert *113*. One male (number 4) from the flies

bearing the new eye colour was crossed to *w*, *TM2/MKRS* females. Resulting males and females which had the insert balanced by *MKRS* were crossed to each other to generate line 67-4. Again, all the flies in this line are always *w*⁺ *Sb* indicating that insert 67-4 is homozygous lethal. My first hypothesis on this line was that it could have originated by a jump that had left a copy of the insert in its original position. The fact that the eye-colour of 67-4 flies was much darker than 113 individuals also supported the idea that in line 67-4 there could be two copies of the insert. However, when Southern blots were performed on lines 113, 168-8 and 67-4, the results suggested a much more complicated scenario (see below).

When a Southern blot with *EcoRI* digested DNA from lines 113, 168-8 and 67-4 was hybridized with a probe for the 3' *P* element end, the band detected was of the same size for each line, suggesting that the insert was in the same position in lines 168-8 and 67-4 as in line 113 (figure 4.4). Since only one band was detected in the DNA of line 67-4, it appeared that this line carried only one copy of the insert. Moreover, when a probe for the 5' *P* element end was used, no signal was detected from the DNA of any of these lines suggesting that the insert in each one of them had lost its 5' *P* element end (figure 4.7). Consequently, since the original 113 insert had somehow lost the 5' *P* element end, it was very unlikely that it could have jumped to generate lines 168-8 or 67-4. *In situ* hybridizations, performed by Federica Sandrelli at the University of Padova, also confirmed that the inserts in line 168-8 and 67-4 were localized exactly in the same position as in line 113 (80 A on the third chromosome).

A Southern blot was then performed to check whether the *XbaI* site adjacent to the 5' *P* element end had also been lost. This site marked the proximal extremity of the *nonA*^{vir} gene as it was cloned into pW8 (see figure 4.1), and its loss would indicate that also some of the 3' end DNA of the original *nonA*^{vir} had been lost. The Southern blots in figure 4.8 and 4.9 show that this was the case for line 113 and consequently for line 168-8 and 67-4. However a *HindIII* site 2Kb from the 3' end of *nonA*^{vir} was retained in lines 113, 168-8 and 67-4 (figure 4.8) suggesting that the *nonA*^{vir} sequence lost in these lines could have been anything between 0 and 2 Kb. This *HindIII* site falls within the fourth *nonA*^{vir} intron (see chapter 3 and figure 4.8). No additional experiments

A



B

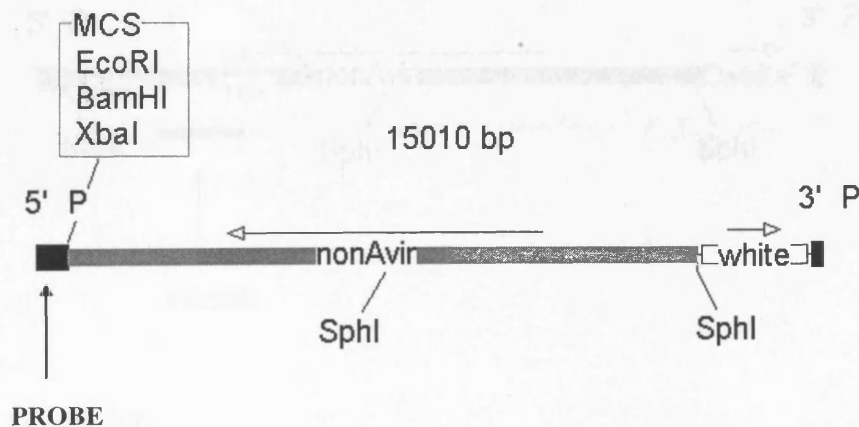
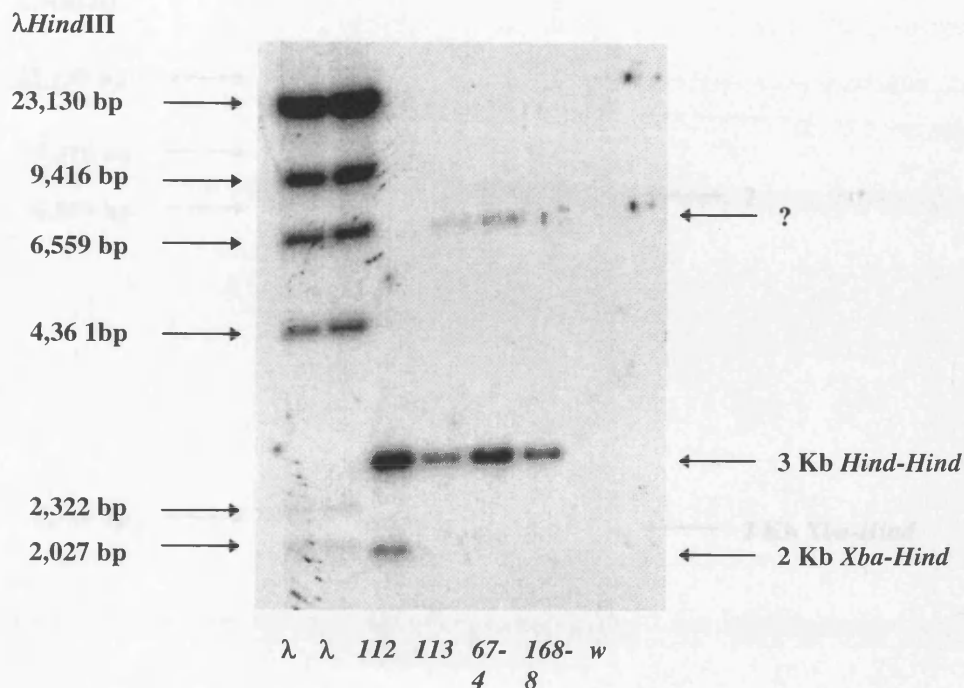


Figure 4.7.

A: Southern blot of lines 112, 113, 67-4, and 168-8, using a probe for the 5' *P* element end of pW8. The DNA from flies of each line was either double-digested with *EcoRI* and *BamHI* or with *XbaI* and *SstI* as indicated. DNA from a *P* element-free *w*¹¹⁸ (*w*) line was also included as a negative control. **B:** Schematic illustration showing the position and orientation of the *nonA*^{vir} construct in relation to the 5' and 3' *P* element ends of pW8 (named "5' P" and "3' P" in picture). The *white* gene in pW8 is also shown. The arrows above each gene indicate the 5' to 3' orientation. Some of the sites of pW8 multiple cloning site ("MCS") are indicated. The vertical arrow shows the annealing position of the 5' *P* element probe.

A



B

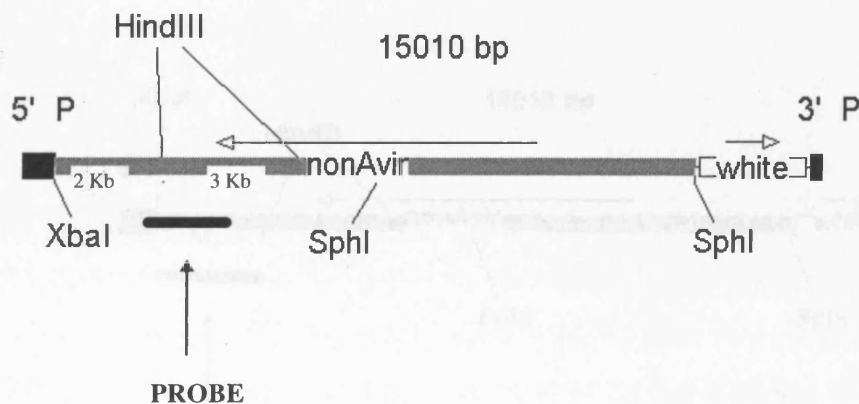


Figure 4.8.

A: Southern blot with *HindIII* and *XbaI* digested DNA from lines 112, 113, 67-4, and 168-8. DNA from w^{1118} was included as a negative control. As shown in picture B, an *HindIII*-*XbaI* cut releases two fragments of 2 and 3 Kb from the end of the *nonA^{vir}* construct. The probe, generated by PCR amplification of a ~950 bp fragment spanning from the end of exon three to the beginning of exon 5 of *nonA^{vir}*, anneals mostly to the 3 kb *HindIII*-*HindIII* fragment and in part also to the 2 kb *XbaI*-*HindIII* fragment. From the hybridization results in A it is clear that all the tested lines retained the 3 Kb *HindIII*-*HindIII* fragment. However line 113, 67-4, and 168-8 appear to have lost the *nonA^{vir}* terminal *XbaI* site since they do not display the expected 2Kb *XbaI*-*HindIII* band (as in line 112). Instead they show a much larger band (marked with "?" in A) resulting by a *XbaI* genomic cut distal to the insert (the signal of line 168-8 is not evident). The size of this band seems to be the same for 67-4 and 113 indicating that the extent of the 3' terminal *nonA^{vir}* deletion is the same in both inserts.

A

λ HindIII

23,130 bp

9,416 bp

6,559 bp

2,322 bp

← ?

← 2 Kb Xba-Hind

λ λ 112 113 67- 168- w
4 8

B

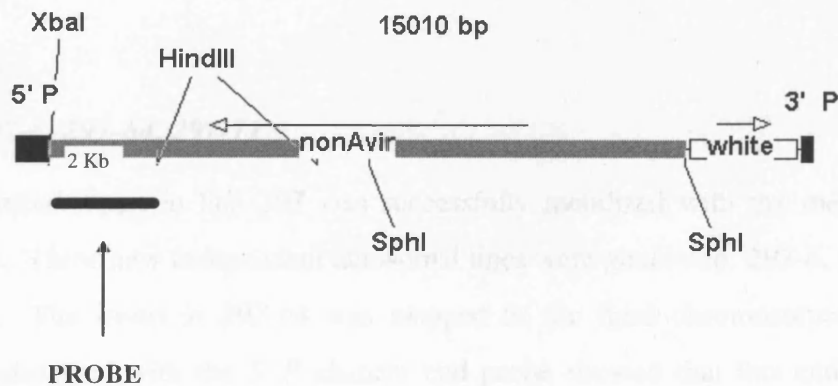


Figure 4.9.

A: Southern blot with *Hind*III-*Xba*I digested DNA from lines 112, 113, 67-4, and 168-8. As shown in picture **B** the hybridization probe used was the 2 Kb *Xba*I-*Hind*III terminal fragment of *nonA*^{vir}. For some reason, the use of this probe generated very poor hybridization results as seen by the quality of picture **A**. The results of this Southern confirm the data shown in figure 4.8. Again it is shown that only line 112 displays the expected 2 Kb *Xba*I-*Hind*III band while lines 113, 67-4, and 168-8 do not since in these lines the insert has lost the *nonA*^{vir} terminal *Xba*I site. The signal of line 113 here is not clear while the one of line 168-8 (which was not clear in the Southern of figure 4.8) is quite evident. The size of the band in 67-4 and 168-8 appears to be similar indicating that a similar deletion pattern occurred in the inserts of these two lines.

were performed in order to check whether the fifth *nonA^{vir}* exon was intact in the inserts of lines 113, 168-8 and 67-4. Further investigations aimed to evaluate the extent of 113, 168-8 and 67-4 *nonA^{vir}* sequence loss were unsuccessful due to the problems generated by the use of a probe for the 3' end of *nonA^{vir}* (see Southern in figure 4.9).

Lines 72, 135, and 297

w¹¹¹⁸ flies were injected as described in the methods above. From ~300 G₀ injected lines only three generated transformant progeny. Transformant lines 72 and 297 both had a homozygous viable X-linked insert. The only autosomal insert, in line 135, was mapped and balanced as described in the methods (figure 4.2): it is homozygous viable on the third chromosome. A Southern blot was performed on these lines using the 5' *P* element end of pW8 (figure 4.10). Again, it was found that one of the three lines (line 72) had lost the 5' *P* element end. The presence of a single hybridization signal from the blotted DNA of 135 and 297 indicated that in these lines only one copy of the transgene had inserted into the genome.

Lines 297-6, 297-64, 297-71

The X-linked insert in line 297 was successfully mobilized with the method illustrated above. Three new independent autosomal lines were generated: 297-6, 297-64 and 297-71. The insert in 297-64 was mapped to the third chromosome. A Southern blot hybridized with the 5' *P* element end probe showed that this end had been lost in the insert of lines 297-64 and 297-71 (figure 4.11). The insert in line 297-6 appeared to be in single copy.

Lines 75, 97, 191

Three new independent transformant lines were generated from ~200 *w¹¹¹⁸* injected flies. They all appeared to have autosomal inserts. A Southern blot checking

A

λ HindIII

23,130 bp

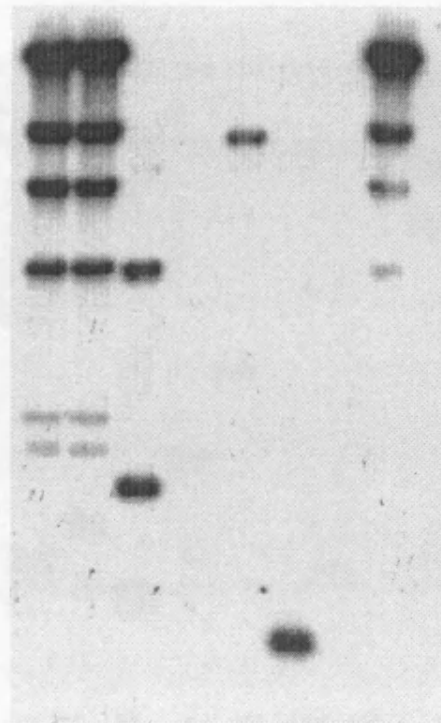
9,416 bp

6,559 bp

4,361 bp

2,322 bp

2,027 bp



λ λ 112 72 135 297 w λ

B

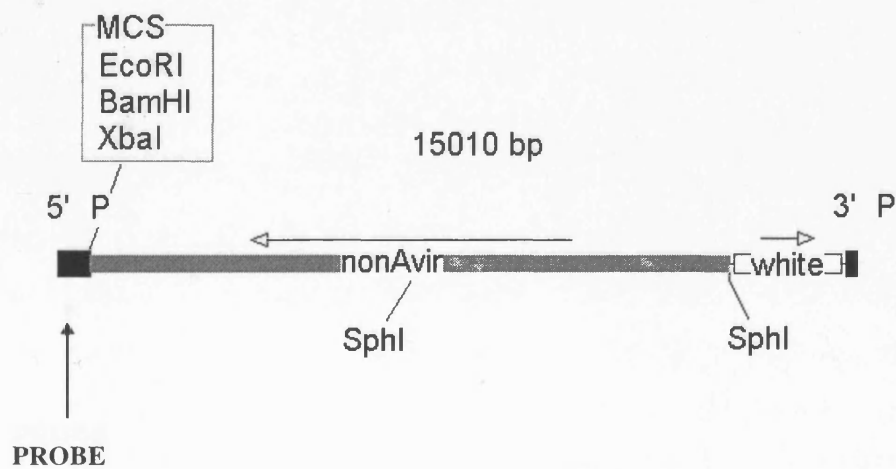


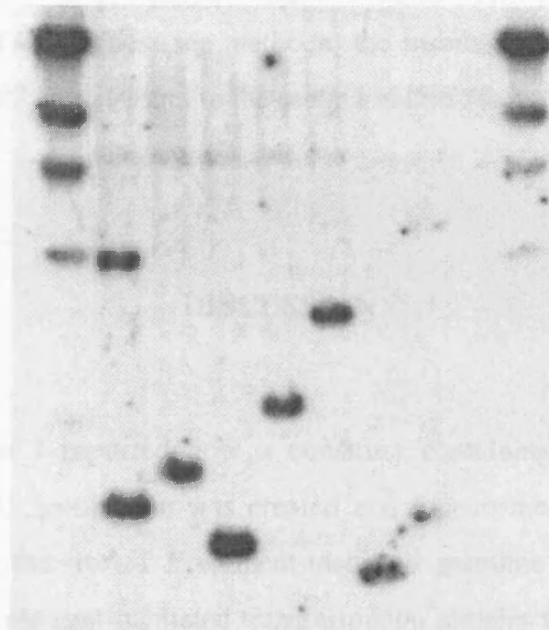
Figure 4.10.

A: Southern blot of *EcoRI* digested DNA from lines 112, 72, 135, and 297, using a probe for the 5' P element end of pW8. DNA from *w*¹¹⁸ (*w*) flies was included as a negative control. **B:** Same as figure 4.7B.

A

λ HindIII

23,130 bp →
9,416 bp →
6,559 bp →
4,361 bp →
2,322 bp →
2,027 bp →



λ 112 75 97 191 297 297 297 w λ
-6 -64 -71

B

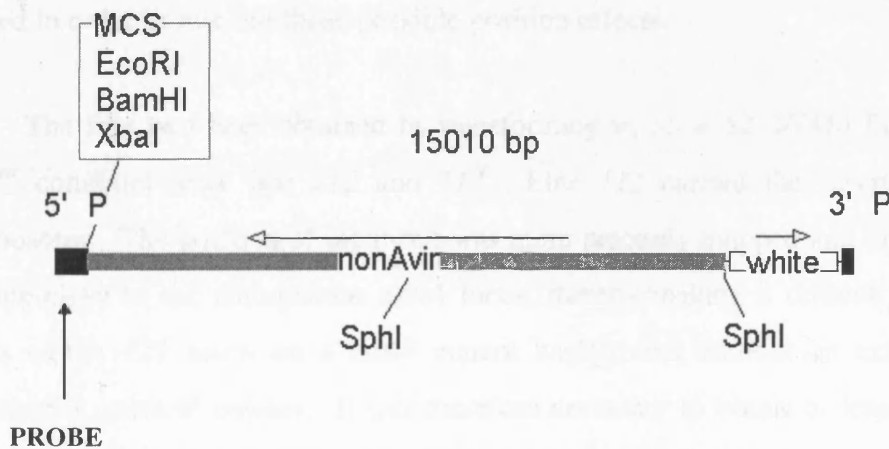


Figure 4.11.

A: Southern blot with *Eco*RI digested DNA from lines 112, 75, 97, 191, 297-6, 297-64, 297-71, and *w*¹¹⁸(*w*), hybridized with a probe for the 5' *P* element end of pW8. The signal in the lane corresponding to line 297-64 is an artifact due to a tear in the filter. No band was actually detected in 297-64 lane. **B:** Same as figure 4.7B.

the 5' *P* element end confirmed that all three independent inserts had retained it (figure 4.11). Only one copy of the transgene was present in each tested line. By crossing to *w*; *T2,3 ap^{Xa}/SM5CyTM3Sb* flies (see methods) the inserts were mapped to the third chromosome in lines 97 and 191 and to the second in line 75.

DISCUSSION

In this chapter I reported how a construct containing the entire predicted sequence of the *virilis nonA* gene was created and transformed into the genome of *D.melanogaster* with the use of *P* element-mediated germline transformation. The main problem with *P* element-mediated transformation remains the fact that no control can be imposed on where in the genome the transformed gene will insert. As a consequence, every independently derived transgene is localized in a unique chromosomal context that may affect its expression. When studying the phenotypic effects of a transgene, several independent insertions of the same transgene must be isolated in order to rule out these possible position effects.

The first two lines obtained by transforming *w*; *Sb e Δ2-3/TM6* flies with the *nonA^{vir}* construct were line 112 and 113. Line 112 carried the insert on the X chromosome. The position of the insert was more precisely mapped and turned out to be quite close to the endogenous *nonA* locus, thereby making it difficult to test the effects of the 112 insert on a *nonA* mutant background without an extensive and complicated series of crosses. It was therefore necessary to obtain at least one other autosomal line in addition to line 113. The easiest way to obtain autosomal lines given an X-linked transgene is to mobilize it and select for new insertions on the autosomes. The transposase necessary for mobilization of the insert was provided by the Δ2-3 element. Several extensive attempts to jump the 112 insert onto an autosomal location all failed. In order to transpose, a *P* element must have intact 5' and 3' ends. Approximately 150 bp of DNA at each end is required for transposition (O'Hare and Rubin, 1983). Contained within these sequences are a perfect terminal inverted repeat

(IR) of 31 bp and an 11 bp sequence found in inverted orientation ~130 bp from each end (O'Hare and Rubin, 1983; Engels, 1989). Characterization of the transposase protein has shown that this DNA-binding protein interacts with an internal 10 bp consensus sequence adjacent to the 31 terminal inverted repeat of each 5' and 3' end (Kaufman *et al.*, 1989). Another type of protein named IRBP (inverted repeat binding protein) interacts with the outer 16bp of both terminal inverted repeats (Staveley *et al.*, 1995) and it has been proposed to mediate transposition by increasing the affinity of the transposase for the internal 10 bp target sequence (Kaufman *et al.*, 1989). The presence of both 5' and 3' *P* element ends was checked in the *112* insert by Southern Blotting. However, no actual information was obtained about the integrity of the sequences required for transposition at each end. Therefore the stability displayed by insert *112* could be due to post-insertion damage to its terminal *P* element ends. Alternatively the genomic position of the *112* transgene could be responsible for its inability to transpose. Cases have been reported in which the genomic position of a certain *P* element was the only cause of abnormally low levels of transposition (Engels, 1989). A *P* element marked with the wild-type allele of the *white* gene ($P(w^+)$) inserted at cytological position 17B was found to have very low levels of transposition and excision. However, when a rare transposition to a new position was finally obtained, the new site displayed normal transposition levels (Engels, 1989).

Whenever the DNA from line *112* was digested with an enzyme that did not cut within the 5' *P* element sequence, and probed with the 5' *P* end HindIII-*Xba*I fragment from pW8, two bands were always displayed in the autoradiographs (figure 4.7, 4.10 and 4.11). This would suggest either the presence of two copies of the 5' *P* element end in the *112* insert or the presence of two copies of the *nonA*^{vir} construct. Since only one band was found when probing for the 3' *P* element end (figure 4.4) or for the second half of the *nonA*^{vir} construct (figure 4.5), it seems more likely that there is a single insert in line *112*. However its 5' *P* element end must have duplicated, possibly via some sort of slippage during the DNA replication required for the transgene's chromosomal insertion (Klecker, 1981; Shapiro, 1983).

An attempt was subsequently made to mobilize the autosomal insert *113*. The advantage of jumping this third chromosome transgene was that any new local insertions within the same chromosome should be easier to obtain compared to interchromosomal hops (Tower *et al.* 1992). The *113* insert is homozygous lethal. *P* elements can be mutagenic for a variety of reasons, depending on the position of the insertion in relation to the affected gene. *P* elements seem to preferentially insert at or near genes' transcription start sites (Engels, 1989) and consequently affect promoter function, usually reducing the level of wild type transcript. Alternatively, insertion into the 5' untranslated regions of genes could affect the stability of the transcript or result in its premature termination. *P* elements can also insert in the coding region of some genes, disrupting the production of a functional protein. One such example is the *white* locus in which four insertions in the coding sequence were recovered, three of which mapped to the same nucleotide (Rubin *et al.*, 1982; O'Hare and Rubin, 1983). Then there are more complicated cases of *P* element-gene interactions like the one described by Horowitz and Berg (1995). They showed that an *in vitro* generated *P*[*lacZ*, *rosy*⁺] element, caused a mutant phenotype by inducing aberrant splicing and termination of the *pipsqueak* gene transcript. It was found that the *P* element, which was inserted within an intron of *pipsqueak*, also carried the splice acceptor site and the transcription termination sequences of a gene called *l(3)S12*. Consequently an aberrant splicing event was joining the splice donor site upstream of the *pipsqueak* intron, in which the *P* element was inserted, to the acceptor site of the *l(3)S12* intron. The result was a fusion *pipsqueak-l(3)S12* transcript which yielded a protein with mutant effects.

Whatever causes the mutagenic effect of a *P* element insertion, it is possible to have reversion to wild-type simply by mobilizing the insert. However if the insert is located in the coding region of a gene, only a precise excision of the element can restore the proper reading frame. Bearing this in mind, the insert in line *113* was mobilized by providing a stable transposase source (the $\Delta 2-3$ element, see methods). It was believed that a new insertion could be detected because of its ability to complement the lethality of the original *113* insert. The third chromosome insert *168-8* displayed such a property in that flies heterozygous for *113* and *168-8* were fully viable. However when Southern analysis was performed, it was found that the insert in

168-8 was still at the same genomic position of 113. Further analysis revealed that both the original insert 113 and the new 168-8 had apparently lost their 5' *P* element end (figure 4.7). It was therefore unlikely that insert 113 could have transposed to a new genomic location in order to give the new insert 168-8.

However, even though insert 113 could not have transposed somewhere else, it is still possible that it might have been excised from the genomic DNA. The transposase could have mediated the excision after recognising the 10 bp target sequence internal to the terminal repeat of each *P* element end (Kaufman *et al.*, 1989). In fact, even though the Southern performed indicated that most of the sequence of the 5' *P* element end was gone from the 113 transgene, it cannot be excluded that sequences important for transposase binding, or for other factors interacting with the transposase (such as the IRBP, Staveley *et al.*, 1995; Kaufman *et al.*, 1989) had been retained in 113. Alternatively the transposase could have recognized the 8bp repeated sequences found at each side of *P* elements after genomic insertion (Engels, 1989; Paques and Wegnez, 1993). The excision of 113 would have left a double stranded gap which could have been repaired using the sister chromatid as a template (Engels *et al.*, 1990; Daniels and Chovnick, 1993; Staveley *et al.*, 1995).

In order to explain the ability of 168-8 to complement 113 lethal phenotype, it must be assumed that during the process of repair an error might have occurred that changed either sequence within the *nonA*^{vir} construct or some flanking genomic DNA sequences. It is common that after excision of a *P* element the gap produced is expanded by exonuclease action. Thus an error could possibly occur even during the repair of the sequences that actually lay outside each end of the insert (Engel *et al.*, 1990; Johnson-Schlitz and Engels, 1993). There is evidence that if the sequence within the *P* element end contains a fair amount of repetitive material, the repair process is especially error-prone due to frequent slippage-like events occurring during DNA synthesis (Paques and Wegnez, 1993; Paques *et al.*, 1996; Golic, 1994; Staveley *et al.*, 1995). How a sequence change occurring during gap repair, either within the *nonA*^{vir} construct or in some sequences outside it, could have reversed the effect of the lethal mutation caused by the 113 insert, remains a mystery. Moreover the "new" insert 168-

8 is also lethal when homozygous, suggesting a very complex mechanism underlying these events.

Line 67-4 was also isolated during the same screen that produced line 168-8. The insert in this case does not display complementation of the 113-caused lethality. However flies containing this insert have eyes which are of a much brighter and stronger shade of red than those of flies from line 113. Different eye colour displayed by transformant flies usually indicate different levels of expression of the *w*⁺ transgene suggesting different localization in the genome of recipient flies. The first hypothesis was that this line could have originated by a jump of 113 that had left a copy of the insert in the original position. This would have accounted for the lack of lethal complementation and the darker eye colour (since two copies of the *white*⁺ gene would have been present). However Southern blot analysis confirmed again that insert 113 had not moved in order to originate line 67-4 (figure 4.4). If we assume again that the 113 insert had undergone transposase-dependent excision, then 67-4 can be explained as the result of an event occurring during gap repair. Whatever happened to insert 67-4 must have affected the expression of the *white*⁺ gene encoded by the transgene. A possible duplication of the *white*⁺ gene cannot be excluded since it has not been tested.

As well as the 5' P element end, insert 113 and consequently inserts 168-8 and 67-4 have lost the terminal *Xba*I restriction site which marked the 3' end of the *nonA*^{vir} construct (see methods and figure 4.1, 4.8 and 4.9). All the three lines may therefore be missing some other 3' terminal sequences of the original *nonA*^{vir} construct. However little can be said about the exact extent of the deletion. A *Hind*III site found 2kb upstream of the 3' end of the *nonA*^{vir} construct was however retained in all the three lines (figure 4.8) indicating that the sequence loss from the end of the original construct after insertion must be between 0 to a maximum of 2kb. Comparing the Southern results of line 113 with those of lines 168-8 and 67-4 (figure 4.8 and 4.9), it seems that the extent of the 3' terminal *nonA*^{vir} deletion is the same in the three lines. Because of the way the Southern blots were designed (see legends in figure 4.8 and 4.9) an additional loss of 3' terminal *nonA*^{vir} sequence in the inserts of line 168-8 and 67-4 would have resulted in a smaller band displayed after hybridization of their DNA.

Instead, the bands resulting from lines 168-8 and 67-4 seem to be of the same size as that of line 113. However it must be noted that small differences in DNA sizes cannot be revealed by Southern blot.

Transformant lines 72, 135, 297, 75, 97 and 191 were all generated by injecting the *nonA*^{vir} construct into embryos of the *w*¹¹¹⁸ strain. This method proved to be more efficient than injecting *w*; *e* Δ2-3 *Sb/TM6* flies mainly because of the much greater survival rate of the *w*¹¹¹⁸ injected embryos. Lines 267-6, 267-64, and 267-71 were easily obtained by mobilizing the X-linked insert 297 to the autosomes. Southern blot analysis suggests that in lines 135, 297, 75, 97, 191 and 297-6 only one copy of the transgene has inserted into the genome (figure 4.10 and 4.11). When using a probe for the 5' P element end no hybridization signal was displayed by the DNA corresponding to lines 72, 297-64 and 297-71. Thus again the 5' P element end had been lost, possibly upon integration of the transgene into the genome of the recipient flies. Considering the fact that this end has been lost in a total of four independent integrations plus in another case it has been duplicated, it seems that replication slippage events at this end of the construct occur quite frequently. This suggests that the sequence at the 3' terminal end of the *nonA*^{vir} construct might be highly repetitive.

In spite of the enigmatic nature of lines 168-8 and 67-4 and the fact that they are not independent lines from 113, they were still used in all the behavioural tests described in the following chapters. No direct proof was obtained that these lines retained an intact fifth exon in their *nonA*^{vir} transgenes. However indirect proof of the existence of this exon in the three lines can be obtained by the behavioural results since the fifth exon has been shown to be indispensable for all the *nonA* biological functions (Jones and Rubin, 1990; Rendhal *et al.*, 1992). Among the other transformant lines, only lines 135, 97, 191 and 297-6 were used in the behavioural analyses.

CHAPTER 5

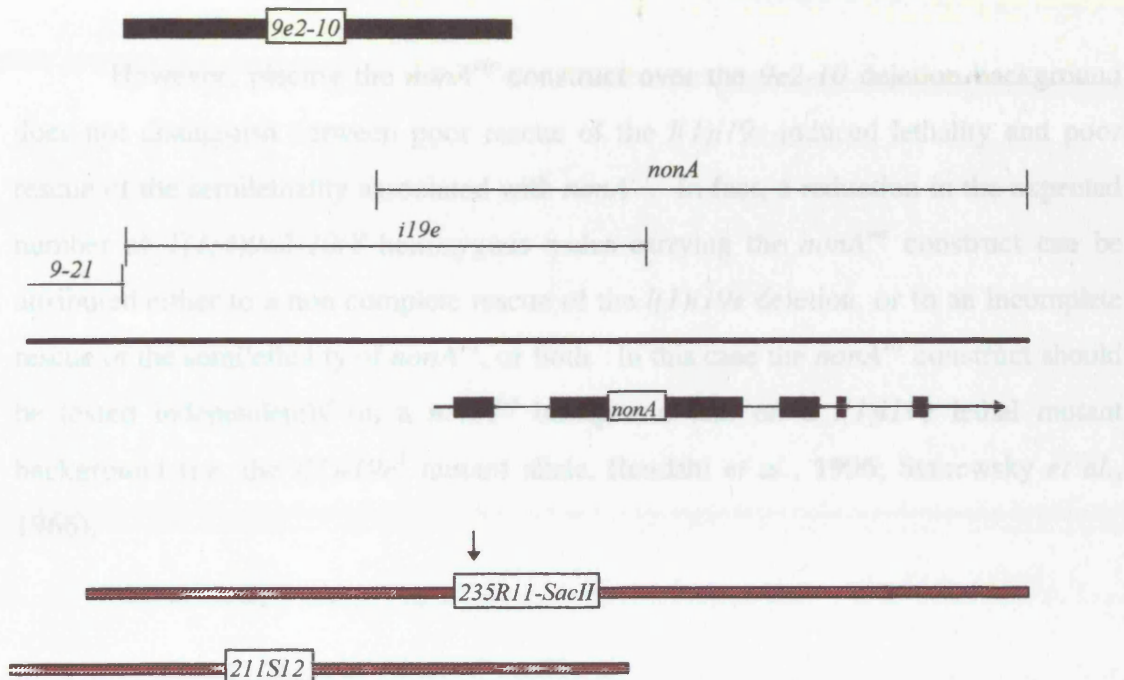
Viability tests

INTRODUCTION

This chapter reports the first tests performed on transformant flies carrying a *nonA^{vir}* gene. Stanewsky et al. (1993) have isolated a lethal reciprocal translocation between the first and the fourth chromosomes (*T(1;4)9e2-10*) which produces a small deletion uncovering the *nonA* gene and a partially overlapping, upstream locus (*l(1)i19e*) whose product is essential in *D.melanogaster*. By transforming flies carrying this deletion with transgenes encoding only for the essential gene *l(1)i19e*, Stanewsky et al.(1993) managed to create *nonA⁽⁻⁾* flies (see figure 5.1). Hemizygous *nonA⁽⁻⁾* males are subviable: only 10-30% of the expected number eclose from metamorphosis. The few *nonA⁽⁻⁾* males which reach the imaginal state are deficient in locomotor activity and flight, and less than one-third of them survive to become fertile males (Stanewsky et al., 1993). Not surprisingly *nonA⁽⁻⁾* flies are severely affected in all the *nonA*-related behaviours. Severe mutations affecting the first RRM domain (RRM1) of *nonA* have also proven to considerably reduce viability along with affecting both visual and song behaviours (Rendahl et al., 1996; Stanewsky et al., 1996). As a corollary, we can expect that mutations that affect viability will undoubtedly disrupt all the behavioural phenotypes controlled by *nonA*.

Since it was decided that the *nonA^{vir}* transgene should be tested on a *nonA⁽⁻⁾* background, it seemed sensible to first assess the efficiency of the *nonA^{vir}* gene in rescuing the viability deficits caused by *nonA⁽⁻⁾*. If the *nonA^{vir}* transgene does not ameliorate viability of *nonA⁽⁻⁾* males, there is little possibility that it will show any rescue of their behavioural defects. Bearing in mind that the *nonA^{vir}* construct used for transformation contains more than 3 Kb of DNA upstream of the first *nonA^{vir}* codon (see chapter 3, figure 3.8), I suspected that all the sequence of the *virilis* homologue of the *l(1)i19e* gene would also be contained in the transforming fragment. Therefore I decided to place the *nonA^{vir}* construct not in a *nonA⁽⁻⁾* background, but directly on a *T(1;4)9e2-10* genetic background. I knew that if any transformant males carrying the *nonA^{vir}* construct and the *l(1)i19e +nonA* double deletion could be produced, it would

A:



B:

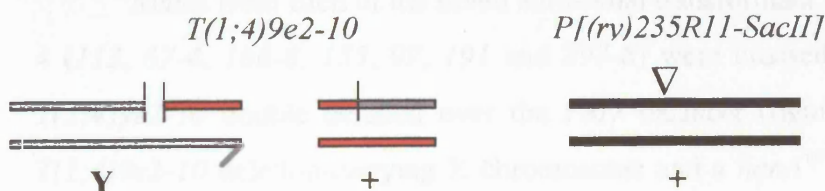


Figure 5.1. A: molecular and genetic map of the X chromosomal region 14B17-14C4. The physical position of the *9-21*, *l(1)i19e*, and *nonA* genes is indicated. The black bar on top of these genes shows the extent of the *9e2-10* deletion. The *nonA* primary transcript is also shown: black bars represent exons, thin lines stand for introns. Underneath, the two genomic DNA fragments used to generate *nonA*⁽⁻⁾ flies are indicated: the *211s12* fragment encodes only the first two exons of *nonA*, while the *235R11-SacII* has a frame shift mutation (indicated by the arrow) introduced at the *SacII* site in the first intron. Thus neither of these two DNA fragments can produce a full NONA protein.

B: chromosomal constitution of a *nonA*⁽⁻⁾ male. Grey bars: X and Y chromosomes; black bars: second or third chromosome; red bars: fourth chromosome. The X chromosome carries the *9e2-10* deletion uncovering both *l(1)i19e* and *nonA*. The reciprocal 1;4 translocation is schematically represented by the grey and red bars. On the autosomes (black bars) a transgene (V) encodes only the *l(1)i19e* gene product, this could be either the *235R11-SacII* or the *211s12* DNA fragment.

have meant that the *nonA^{vir}* transgene carries the *virilis* homologue of the *l(1)i19e* gene and that this gene is functioning in a *D.melanogaster* background.

However, placing the *nonA^{vir}* construct over the *9e2-10* deletion background does not distinguish between poor rescue of the *l(1)i19e*-induced lethality and poor rescue of the semilethality associated with *nonA⁽⁻⁾*. In fact, a reduction in the expected number of *T(1;4)9e2-10/Y* hemizygous males carrying the *nonA^{vir}* construct can be attributed either to a non complete rescue of the *l(1)i19e* deletion, or to an incomplete rescue of the semilethality of *nonA⁽⁻⁾*, or both. In this case the *nonA^{vir}* construct should be tested independently on a *nonA⁽⁻⁾* background and on a *l(1)i19e* lethal mutant background (i.e. the *l(1)i19e²* mutant allele, Rendahl *et al.*, 1996; Stanewsky *et al.*, 1966).

METHODS

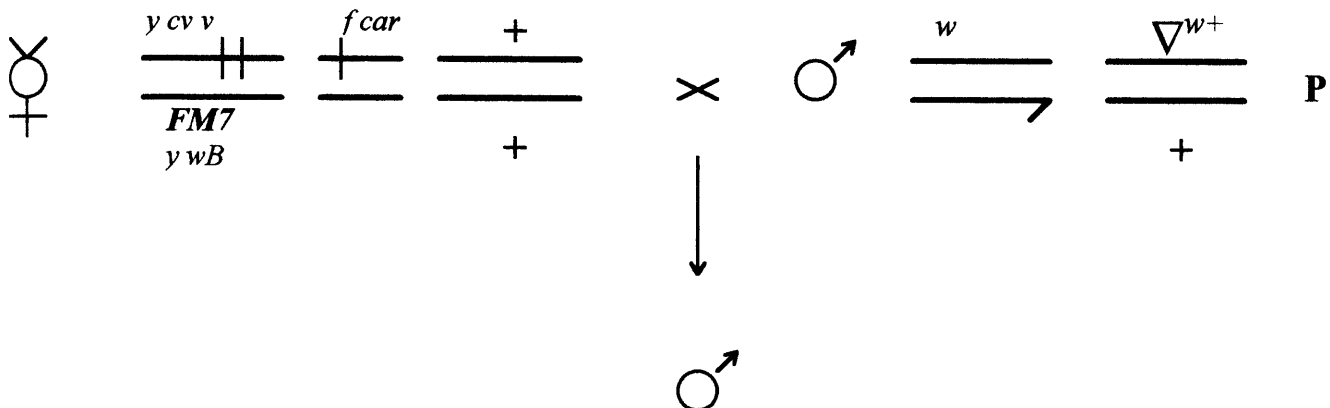
Placing the *nonA^{vir}* construct on a *T(1;4)9e2-10* background.

Males from each of the seven autosomal transformant lines described in chapter 4 (113, 67-4, 168-8, 135, 97, 191 and 297-6) were crossed to females carrying the *T(1;4)9e2-10* double deletion over the *FM7* balancer (figure 5.2). Males with the *T(1;4)9e2-10* deletion-carrying X chromosome and a *nonA^{vir}* autosomal insert should be distinguished by the presence of the morphological markers carried by the *T(1;4)9e2-10* X chromosome (*yellow^l* (*y^l*), *crossveinless* (*cv*), *vermillion* (*v*), *forked* (*f*), and *carnation* (*car*) (Stanewsky *et al.*, 1993).

Viability tests.

Large numbers of males from each transformant line were crossed to *T(1;4)9e2-10/FM7* females and all the males of the progeny were classified and

T(1;4)9e2-10



T(1;4)9e2-10

GENOTYPE

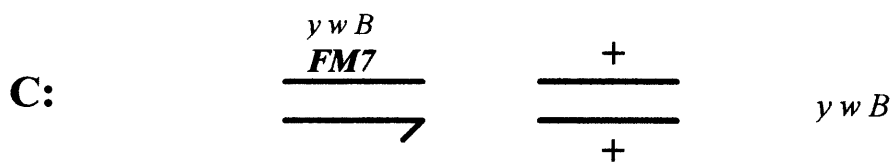
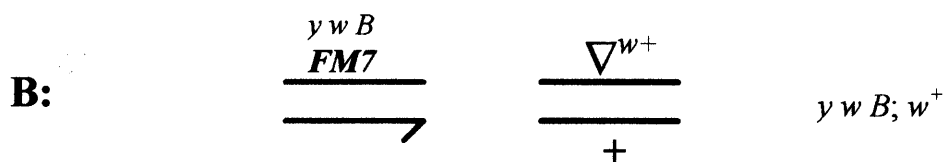
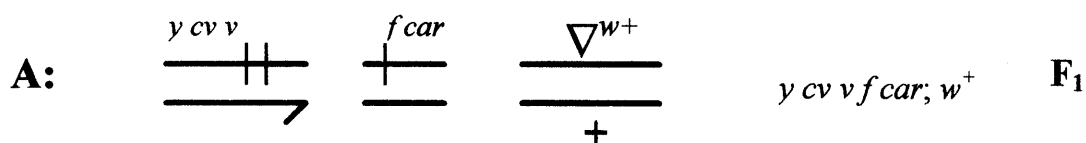


Figure 5.2. Cross for testing the viability of flies carrying the *T(1;4)9e2-10* translocation (marked with *y cv v f car*) and the *nonA^{vir}* transgene (represented by ∇ , marked with w^+). The three types of males generated by the cross are indicated. Only the X and one of the autosomes (either the second or the third) are shown for males B and C. In type A males the fourth chromosome involved in the *T(1;4)9e2-10* translocation is also illustrated. Grey bar = X and Y chromosome. Red bar = fourth chromosome. Black bar = second or third.

counted. The ratio of $y, cv, v, f, car; w^+$ males ($T(1;4)9e2-10 + nonA^{vir}$ insert) over the total number of males produced by the cross was calculated. The genotypes of the other two kinds of males obtained (figure 5.2) were $y, w, B; w^+$ and y, w, B ; since the *FM7* chromosome is marked with $y, w, Bar (B)$, (Lindsley and Zimm, 1992), and the transgene is marked with w^+ (Klementz *et al.*, 1987).

RESULTS

If males carrying the X-linked $T(1;4)9e2-10$ double deletion and a $nonA^{vir}$ autosomal insert were 100% viable, their number should equal 1/3 of the total number of males produced by the cross (i.e. 33.3 %). From table 5.1, it is clear that males carrying the $l(1)i19e + nonA$ deletion ($9e2-10$) and a $nonA^{vir}$ transgene are viable, confirming that the $nonA^{vir}$ construct used in the transformation (chapter 3 and chapter 4) must encode the complete product of the *virilis* homologue of the $l(1)i19e$ essential gene.

Most transformant lines show full rescue of the viability defects associated with $l(1)i19e$ and $nonA^{(-)}$. Unexpectedly, for most lines, far more than a third of the total males, produced by the cross, carried the $nonA^{vir}$ transgene on a $9e2-10$ deletion background. This effect is so marked that a statistical test is unnecessary. It is inevitable that the presence of the *FM7* chromosome in the two other types of males produced by the cross (B and C in table 5.1.), given the chromosomal rearrangements within the balancer (Lindsley and Zimm, 1992), will affect their viability.

Strangely, only 1 male of type C (see table 5.1) was produced in the cross involving the *l35* transgenic line. The reason for this bizarre result is unknown. Line *l91* generated significantly less than 33% $T(1;4)9e2-10/Y; \nabla/+$ males ($\chi^2 = 14.7$; $df = 2$; $p < 0.001$). A position effect is the most likely explanation for this result.

Table 5.1. Number of males of each genotype (A, B, or C) resulting from a cross *T(1;4)9e2-10/FM7* females x transgenic *nonA^{vir}* males (see figure 5.2). **A:** males carrying the X-linked deletion *T(1;4)9e2-10* and the *nonA^{vir}* insert on an autosome (indicated with the ∇ symbol). **B:** males with *FM7* and an autosomal *nonA^{vir}* insert. **C:** males with *FM7* only. In the last column the percentage of the A type of males is calculated over the total amount of males collected from the cross.

TRANSGENIC LINE	A <i>T(1;4)9e2-10/Y; ∇/+</i>	B <i>FM7/Y; ∇/+</i>	C <i>FM7/Y; +/+</i>	% A/(A+B+C)
113	92	64	38	47.4
67-4	64	42	22	71.1
168-8	85	58	43	45.7
135	109	121	1	47.2
97	78	44	20	54.9
191	41	72	37	27.3
297-6	54	63	31	36.5

DISCUSSION

In this chapter I have demonstrated that the *nonA^{vir}* construct (see chapter 3 and 4) fully rescues the lethality associated with a translocation *T(1;4)9e2-10* that causes a simultaneous deletion of the essential *l(1)i19e* gene and the *nonA* gene (Stanewsky *et al.*, 1993). This suggests that the *nonA^{vir}* transgene must contain the entire coding sequence for the *D.virilis* homologue of the *D.melanogaster l(1)i19e* gene, and that this gene product can function properly in a *D.melanogaster* background. Furthermore, the *nonA^{vir}* construct fully rescues the subviability associated with the absence of a functional *nonA* gene product in *nonA⁽⁻⁾* flies

(Stanewsky *et al.*, 1993). Therefore there is good indication that the *nonA^{vir}* transgene contains sequences encoding for the full product of the *D.virilis nonA* gene, and that the *virilis* NONA protein is working in a *D.melanogaster* background at least at levels sufficient to perform basic functions.

A useful consequence of these results is that, in order to perform the subsequent visual and courtship behavioural tests, there will be no need to produce *nonA⁽⁻⁾* flies (as illustrated in figure 5.1) to serve as a genetic background for testing the various *nonA^{vir}* inserts. These can in fact be studied directly on a *T(1;4)9e2-10* genetic background, saving a considerable amount of time consuming crosses that involve very unhealthy flies (i.e *nonA⁽⁻⁾* flies).

CHAPTER 6

Analysis of visual behaviour

INTRODUCTION

The *nonA* locus is one of the many genes which affect the normal functioning of the visual system of *D.melanogaster* (reviewed by O'Tousa, 1990; Pak, 1991; Smith *et al.*, 1991). *nonA* mutations were first isolated in screens for visual mutants which exhibited reduced phototaxis (Pak, 1970; Hotta and Benzer, 1970). Subsequently it was observed that the *nonA* visual phenotype involves multiple defects, as assayed electrophysiologically and behaviourally. *nonA* flies are deficient in optomotor responses (Heisenberg, 1971a) and exhibit an aberrant electroretinogram (ERG) (Hotta and Benzer, 1969; 1970; Heisenberg, 1971b; Pak *et al.*, 1970).

The *Drosophila* visual apparatus, like that of *Musca*, consists of a peripheral retina and three postretinal structures, the lamina, the medulla and the lobula, which together form the optic lobe (figure 6.1). The retina is composed of approximately 700 structural subunits called ommatidia, each optically isolated from the others by a layer of screening pigment cells containing pterins and ommachromes. These pigments give rise to the red colouration of the eye, and have a function similar to that of the pupil in the vertebrate eye. As well as the image-forming dioptric apparatus, each ommatidium bears a photoreceptive system, the retinula, which is composed of eight retinula cells (photoreceptors). Each retinula cell carries a rhabdomere, a differentiated membrane structure which contains the visual pigment. From each ommatidium, the rhabdomeres of cells R1-6 transfer their excitation to the monopolar neurons of the lamina, while those of cells R7-8 synapse directly with the medulla.

Recording the electroretinogram (ERG) of a fly can give an idea about the physiological and structural integrity of the peripheral retina and lamina in the compound eye. However interpretation of the results can be difficult, (Pak,1975). An ERG records the electrical potential change elicited by a light stimulus, between an electrode placed in the cornea and another electrode (the reference electrode) placed in some unresponsive portion of the animal (usually the abdomen or the dorsal surface of the head.) The electrical potential recorded arises from the summed electrical activities

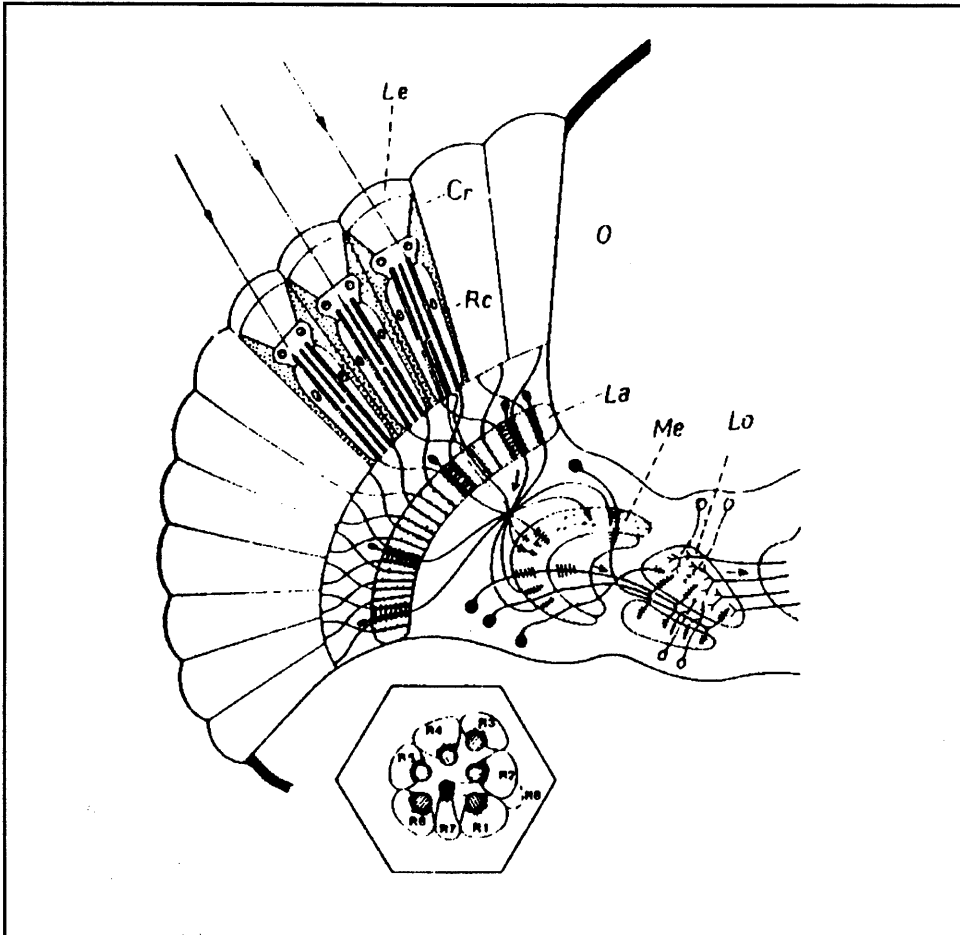


Figure 6.1. Schematic drawing of the dipteran eye and optic ganglia. *Cr*: crystalline cone; *La*: Lamina; *Le*: Lenslet; *Lo*: Lobula; *Me*: Medulla; *O*: Ommatidium; *Rc*: Receptor cell. Insert figure: crosssection of retinula cells in one ommatidium. From Heisenberg and Buchner (1977).

of the retinula cells as well as the neurons (and possibly the glial cells) of the optic lobe. A typical ERG displays a corneal-negative component which is maintained as long as the light is on. In addition, the ERG includes a rapid, corneal-positive transient (the “on-transient”) when the light is first turned on, and a corneal-negative transient (the “off-transient”) when the light is turned off. The corneal-negative, maintained component is believed to originate mainly in the retinula cells (Goldsmith, 1965), although minor contributions may come from cells in the optic lobe. A variety of evidence indicates that neither the on nor the off-transient originates from the retinula cells (Autrum and Hoffman, 1957, 1960; Burkhardt and Autrum, 1960; Wolbarsht *et al.*, 1966; Eichenbaum and Goldsmith, 1968; Alawi and Pak, 1971). Major components of these transients seem to originate in the lamina (Pak, 1975; Coombe, 1986; Coombe and Heisenberg, 1986).

The characteristic feature of an ERG recorded from *nonA* mutants is the lack of both on and off-transient spikes (Pak *et al.*, 1970; Hotta and Benzer, 1970; Heisenberg, 1971b). Thus the ERG of these mutants consists mainly of the corneal-negative, maintained component. An obvious interpretation of these ERGs is that in the mutants the receptor cells must be functioning normally, but either the transmission of the light signal to the neurons in the lamina is prevented, or the excitation of these laminar cells is somehow blocked (Pak *et al.*, 1969; Heisenberg, 1971b). Also, the possibility that the mutants’ visual defects are not confined to the lamina, but may extend to the medulla, is not ruled out by the electrophysiological evidence.

The hypothesis that *nonA* mutants have mainly a central defect in their visual system is supported by their poor optomotor response, which is the fly's tendency to follow movements of the visual environment (Götz, 1964). For an optomotor response to occur, the movement of the environment must be detected comparing the outputs from at least two photoreceptors. Since synaptic interconnections between different photoreceptor axons take place in the lamina, the first motion detectors are probably localized in this structure. Heisenberg (1972) characterized the flying and walking optomotor behaviour of the mutant *nonA*^{H2} which was further refined by Heisenberg and Buchner (1977). These investigators showed that, in optomotor tests,

nonA^{H2} seems to be defective in vision at high light levels in a system that specializes in the resolution of visual details (high acuity system). However *nonA*^{H2} defects did not seem to manifest themselves at the level of the retina, and were assumed to occur at a level proximal to it. Studies on the *optomotor-blind*^{H31} (*omb*^{H31}) mutant, in which the giant neurones of the lobula plate are missing or severely reduced, suggest that a set of three neurones (H-cells) at the anterior surface of the lobula plate may be part of the main pathway for the optomotor turning response (Heisenberg *et al.*, 1978). Electrophysiological experiments on larger flies have shown that the H-cells are sensitive to horizontal movement around the fly (Dvorak *et al.*, 1975; Hausen, 1976). Thus they are good candidates for mediating optomotor responses.

In this study the optomotor response of transformant flies carrying one, two or three copies of the *D.melanogaster* (*nonA*⁺) and *D.virilis non-A* gene (*nonA*^{vir}) were analyzed and compared to that of mutants *nonA*^{diss} and wild-type flies. An initial attempt was made to record ERGs from these flies, using equipment designed to measure ERGs in crustaceans. No reliable ERG recordings could be produced due to technical reasons. Considering the fact that ERGs artefacts are common even using better equipment, a decision was taken to use only the optomotor test in studying the visual behaviour of the flies of interest. The optomotor test, giving more than a single positive or negative result (in contrast to ERG measurements in which the only results are either presence or absence of transient spikes), was viewed as a more powerful test for uncovering subtle differences in the visual behaviour of the transformants.

MATERIALS & METHODS

The walking optomotor test was performed as described in chapter 2. Two sets of stripe widths were used: in one set each black and white pair subtended a small angle (26.6°). In the “large” stripes each black and white pair subtended a larger angle of 72°. 10 males for each genotype were tested in both patterns, with every male

subjected to twenty trials, in which the direction of the rotating stripes was alternatively changed.

GENOTYPES

A:Controls

Males from the *D.melanogaster* strain *Canton S* were used as a wild type positive control. The *nonA^{diss}* mutant male strain *diss/FM7a; ry⁵⁰⁶/MKRS* provided a negative control in the test. Unfortunately it proved almost impossible to generate *nonA⁽⁻⁾* flies to use as negative controls. As described in chapter 5, *nonA⁽⁻⁾* males can be produced by crossing females carrying the *T(1;4)9e2-10* translocation, which deletes both the *l(1)i19e* and the *nonA* genes, with transformant males carrying autosomal *235R11-SacII* or *211S12* inserts (which encode only the functional product of the *l(1)i19e* gene, see figure 5.1). For some reason, even after performing such a cross several times with large numbers of flies, only two *nonA⁽⁻⁾* males were ever produced. I realized therefore that it would have taken too much time to produce enough *nonA⁽⁻⁾* flies to serve as negative controls in any of behavioural tests, and decided to use *nonA^{diss}* mutants instead.

Hemizygous *nonA⁽⁻⁾*[*nonA⁺*]

As the experimental flies to be tested were transformants carrying the *D.virilis nonA* gene (*nonA^{vir}*), appropriate transformant controls carrying the *nonA⁺* gene from *D.melanogaster* were required. These were provided as follows. The translocation *T(1;4)9e2-10* created a simultaneous deletion of the essential gene *l(1)i19e* and the *nonA* locus (Stanewsky *et al.*, 1993, see chapter 5). Transformant flies carrying a 11kb DNA fragment encoding both the lethal gene *l(1)i19e* and the *nonA* gene of *D.melanogaster* (*235R11*, Jones and Rubin, 1990) were provided by the stock *P[(ry)235R11]D1* (see chapter 2) which carries the homozygous insert on the third

chromosome. Males from this stock were crossed to *T(1;4)9e2-10 /FM7* females. Since the X chromosome in *T(1;4)9e2-10* is marked with *y*, *cv*, *v*, *f*, and *car*, all the F1 red-eyed males carrying these alleles rescue simultaneously the loss of the lethal gene and *nonA*. These flies carried one copy of the *nonA*⁺ gene of *D.melanogaster* over a *nonA*⁽⁻⁾ background and were named *235R11/d*.

B: hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}]

Transformants carrying one copy of the *nonA* gene from *D.virilis* (*nonA*^{vir}) over a *nonA*⁽⁻⁾ background were generated by crossing *T(1;4)9e2-10/FM7* females to males for each of the seven transformed lines carrying autosomal *nonA*^{vir} inserts discussed in the previous chapter. The seven lines were provided by the transgenic strains *113*, *67-4*, *168-8*, *135*, *97*, *191*, *297-6*. From each cross, only F1 males carrying the *T(1;4)9e2-10* chromosome morphological markers were selected for behavioural testing. These males carry one copy of the *nonA*^{vir} gene over a *nonA*⁽⁻⁾ background. The fact that such viable males are produced by these crosses indicates that the *nonA*^{vir} construct inserted in each line must also carry the intact *virilis* homologue of the lethal gene *l(1)i19e* (see chapter 5). The males tested were respectively named: *113/d*, *67-4/d*, *168-8/d*, *135/d*, *97/d*, *191/d*, *297-6/d*, where *d* is used as a symbol for the double deletion.

C: heterozygous *nonA*^{vir}/*nonA*⁺

Transformants carrying one copy of the *nonA*^{vir} gene plus the endogenous copy of *melanogaster nonA*⁺ were generated as follows. The transformant males from line *112* (discussed in the previous chapter) were included in this analysis: these flies carry one copy of *nonA*^{vir} inserted not far from their endogenous copy of *nonA*⁺ on the X chromosome (carrying *w*, see chapter 4). These transformants were named *112/m*. In the transformant lines *113*, *67-4*, and *168-8* the insert on the third chromosome causes lethality if homozygous. As each strain carries the *w* mutation on the X chromosome,

every red eyed male in these lines has one copy of the *nonA^{vir}* gene plus the endogenous *nonA⁺*. The males analyzed from line 67-4 and 168-8 were called 67-4/*m*, and 168-8/*m* respectively. Transformant line 113 flies are distinguished by a very weak red eye colour. Bearing in mind that *white* flies are optomotor blind due to the lack of pigments in their eyes (see Hall, 1994) , it was necessary to test the effect of line 113 on a wild-type (red-eyed) background. In order to do this, males from line 113/*MKRS* (i.e. carrying the 113 transgene over the *MKRS* balancer) were crossed to females *Canton S*. F1 males carrying no balancer were then tested in the optomotor apparatus and are here referred to as 113/*m*. All the other *nonA^{vir}* inserts gave a strong red-eye colour when placed with a *w* X chromosome, so further crosses were not considered necessary. In lines 135 and 97 flies homozygous for the insert are fully viable. Males from lines 135 and 97 were therefore crossed with females from a *w¹¹⁸* stock. Red eyed male progeny carry a single copy of the insert. These males are called 135/*m* and 97/*m* respectively.

D: 3 *mel*

Males from the *P[(ry)235R11]D1* stock are homozygous for the *melanogaster nonA⁺* transgene on the third chromosome and carry the endogenous copy on their X chromosome. These flies were named 3*mel*.

E: 2 *mel*

Males from the *P[(ry)235R11]D1* stock were crossed to females of a *Canton S* strain. The resulting males (called 2*mel*) have one transformed copy of the *melanogaster nonA⁺* transgene on the third chromosome plus the endogenous copy on the X.

F: *mel+2vir*

Females from line *112* (*nonA^{vir}* insert on the X) which had their X chromosome balanced with *FM7*, were crossed to males from line *113* with their third chromosome balanced by *MKRS*. Resulting F1 males carrying no balancers must have one copy of *nonA^{vir}* on the third chromosome and one on the first, plus the endogenous *melanogaster nonA⁺* gene on the X. These flies were named *mel+2vir*

GENETIC BACKGROUND

When comparing the behaviour of independent transformants lines it is best to try to reduce possible differences due to the different genetic backgrounds. For this reason I tried to keep the background of the transformed flies analyzed here as similar as possible.

Lines *113* and *112* were obtained by transforming flies from the stock *w;Δ2-3/TM6* and successively crossing the transformants to flies from the laboratory stock *w Tully*. Lines *67-4* and *168-8* share the same background being derivatives of line *113*. Heterozygous *nonA^{vir}/mel nonA⁺* lines *112/m*, *67-4/m*, and *168-8/m* also have the same background, so does *mel+2vir*. *113/m* is slightly different since males of line *113* were crossed to *Canton S* females.

Lines *135*, *97*, *191* and *297-6* were all obtained by injecting another strain of *white* flies (stock *w¹¹¹⁸*) and crossing the transformants to flies of the same *w¹¹¹⁸* stock. Heterozygous lines *135/m* and *97/m* have the same background.

In line *235R11* (produced in Jeff Hall's lab) the *melanogaster nonA⁺* insert carries a *ry⁺* marker and is on a *ry⁵⁰⁶* background. This background is therefore shared with the two lines *2mel* and *3mel*.

RESULTS

Table 6.1 shows, for each genotype, the number of males out of 10 which gave a significantly correct optomotor response when tested in the optomotor apparatus, meaning 15 or more correct turns in 20 trials. A correct turn is achieved every time the fly turns in the same direction as the rotating stripes. A fly that performs 15 correct turns generates a significant χ^2 value of 5.0 (df=1; $p<0.05$) where the null hypothesis is random turning (i.e. 10 correct, 10 incorrect turns). A fly giving only 14 correct turns ($\chi^2=3.2$; $0.1>p>0.05$), or less, is therefore considered as visually compromised.

As shown in table 6.1, for small and large angle stripes, the genotype which stands out is *nonA^{diss}* with not a single male giving significantly correct responses. All the others generally produce predominantly correct responses.

Table 6.1. Number of males giving significantly correct optomotor responses. S and L stand for small and large stripes. *diss* refers to *nonA^{diss}*.

S T R I P E	<i>d</i> <i>i</i> <i>s</i> <i>s</i>	<i>C</i> <i>a</i> <i>n</i> <i>t</i> <i>i</i> <i>o</i> <i>n</i> <i>S</i>	2 3 5 R I I	1 1 3/ d	6 7- 4/ d	1 6 8- 8/ d	1 3 5/ d	9 7/ d	1 9 1/ d	2 9 7- 6/ d	1 1 2/ m	1 1 3/ m	6 7- 4/ m	1 6 8- 8/ m	1 3 5/ m	9 7/ m	2 m e l	3 m e l	<i>m</i> <i>e</i> <i>l</i> + 2 <i>v</i> <i>i</i> <i>r</i>	<i>v</i> <i>i</i> <i>r</i>
S	0	10	10	10	10	10	7	8	6	7	8	10	10	10	10	9	10	8	10	10
L	0	10	10	9	10	10	6	10	7	8	8	7	6	5	9	9	9	8	8	10

A χ^2 test of homogeneity, on the results obtained for small stripes, gave a significant value of 94.7 (df=19; $p<0.001$). The result was still significant even after excluding from the test the data of *nonA^{diss}* males ($\chi^2=39.01$; df=18; $p<0.01$). A

similar χ^2 performed on the data for all genotypes, with large stripes, gave a significant value of 68.1 (df=19; p<0.001). When the *nonA^{diss}* data were not included, the χ^2 value was still significant (χ^2 =32.6; df=18; p<0.02). A 2x2 contingency test was performed on the results obtained with the two different stripe widths (excluding the data from *nonA^{diss}*), the null hypothesis being that small and large stripes give the same results. The null hypothesis was rejected (χ^2 =4.03; df=1; p=0.045 with Yates' correction). From table 6.1 it is clear that genotypes *135/d*, *67-4/m*, *168-8/m*, perform comparatively poorly with large stripes.

This preliminary analysis can be further refined by taking into account, for each of the ten flies tested per genotype, the number of the correct turns performed out of twenty trials. Table 6.2 shows the results, obtained with these data, of a two way ANOVA with genotypes (18 in total) and stripe width (small or large) as the main factors. The data obtained from the *D.virilis* and the *nonA^{diss}* genotypes were not included since the *nonA^{diss}* data were clearly different, and the optomotor test of *D.virilis* flies had to be performed using a different method (see chapter 2). Thus only differences between transformant lines carrying one, two or three copies of *melanogaster nonA⁺* and *nonA^{vir}* genes are analyzed. The test scores of the positive control line *Canton S* were also included. Table 6.2 reveals a significant effect of Genotypes (p=0.0000) and Stripe width (p=0.0014). The interaction is also marginally significant (p=0.0389).

Table 6.2. Two way ANOVA of walking optomotor test

SOURCE	MS	df	F	p
GENOTYPE	21.22	17	5.96	0.0000*
STRIPE	36.74	1	10.33	0.0014*
G*S	6.10	17	1.71	0.0389*
ERROR	3.56	324		

Figure 6.2 illustrates the effect of stripe size collapsed over genotype. Clearly the larger set of stripes leads to significantly more turning errors than the narrower stripes. However, the presence of a marginally significant interaction indicates that this is not the case for all the genotypes analyzed. This is evident from figure 6.3 which illustrates the influence exerted by a change in stripe width on each genotype. In this picture the results obtained with *nonA^{dis}* and *D.virilis* were also included. Although for most genotypes a change of stripes into larger ones creates more incorrect responses, some genotypes behaved in the opposite way and actually performed better in response to large stripes (e.g. 97/d, 191/d, 297-6/d). In contrast, the response of other genotypes did not seem to be significantly influenced by the change of stripe size. This is the case of *Canton S*, 67-4/d, 135/m and 97/m.

Hemizygous *nonA⁽⁻⁾*[*nonA⁺*]

As expected, the visual behaviour of *nonA⁽⁻⁾* males carrying one copy of the *melanogaster nonA⁺* gene on their third chromosome (line 235R11/d on figure 6.3) showed no statistical difference from the results obtained using wild-type *Canton S* flies (see Newman-Keuls test results in appendix 6.1), indicating complete rescue of the visual defect of *nonA⁽⁻⁾* flies.

Hemizygous *nonA⁽⁻⁾*[*nonA^{vir}*].

Inspection of figure 6.3 reveals that the lines carrying one copy of *nonA^{vir}* on a *nonA⁽⁻⁾* background (from line 113/d to line 297-6/d) can be divided into two groups:

-lines 113/d, 67-4/d, and 168-8/d: these lines gave on average more correct turns than the other four lines, indicating a better rescue of the mutant visual defect. Their performance was not statistically different from that displayed by wild type flies or line 235R11/d (see appendix 6.1), revealing that the *nonA^{vir}* gene can function as

Figure 6.2. Main effect of stripe width (collapsed over genotype) on correct turning in walking optomotor test. Note that, in general, flies responded better to small stripes.

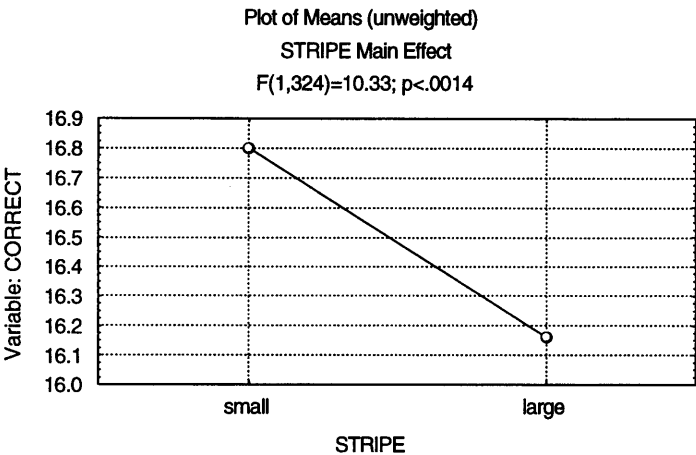
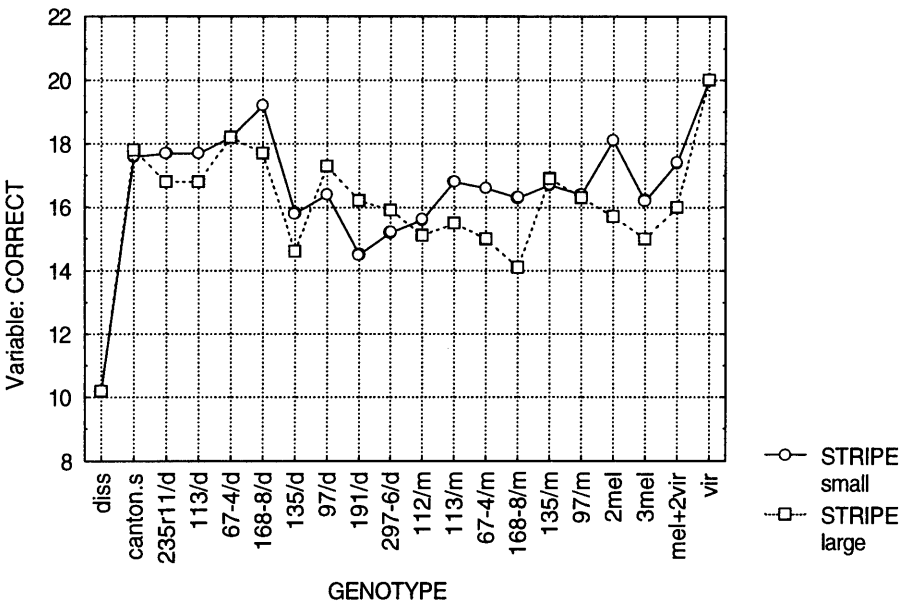


Figure 6.3. Average number of correct turns (out of 20 trials) given by each genotype on small and large stripes.



well as the *melanogaster nonA*⁺ gene in rescuing the mutant visual deficiencies. At this point it is useful to remark that, as explained in chapter 4, the P-element insert in line *113/d* has lost the 5'P-element end and may be missing part of the 3' end of the original *nonA*^{vir} construct. Moreover it appears that lines *67-4* and *168-8* are not independent lines from *113*, so even though the nature of any molecular alteration in lines *168-8* and *67-4* is enigmatic (see discussion in chapter 4), the insert is in the same chromosomal position in all the three lines.

-lines *135/d*, *97/d*, *191/d*, and *297-6/d*: those which achieved the poorest results were line *135/d* and line *191/d*. Line *135/d* behaved much worse when tested using the larger set of stripes. Its performance using large stripes is marginally poorer ($p=0.043$, see Newman-Keuls test's results in appendix 6.1) than that of *Canton S*, but not from *235R11/d* flies. However the transgene in this line is still rescuing the visual defect rather well, since its optomotor scores show no significant difference when compared to those obtained from *Canton S* on small stripes or *235R11/d* on both small and large stripes. Line *191/d* surprisingly performed more poorly when tested on small stripes than on large ones (even though the difference between stripes is not significant, see appendix 6.1). When tested on small stripes, line *191/d* has poorer scores than *235R11/d* ($p=0.038$) and *113/d* flies ($p=0.040$, see appendix 6.1). Therefore line *113* and its derivatives *67-4* and *168-8* rescue visual behaviour as well as the *mel nonA*⁺ transgene, and better than the intact *nonA*^{vir} transgenes *135*, *97*, *191*, and *297-6*. Lines *97/d* and *297-6/d* are not significantly different from *Canton S*, *235R11/d*, *113/d*, *67-4/d*, or *168-8/d* flies, for both sets of stripes.

Heterozygous *nonA*^{vir}/*nonA*⁺

Inspection of figure 6.3 reveals that the six lines analyzed (from line *112/m* to line *97/m*) showed similar scores to their hemizygous counterparts. Line *168-8/m* however showed relatively poor visual behaviour when tested on large stripes. Its performance was significantly poorer than the wild-type ($p=0.005$, Newman-Keuls test, appendix 6.1) and *235R11/d* when tested on large stripes ($p=0.006$). Interestingly

line 168-8/*m*, when using large stripes, produced significantly more errors compared to the insert on the *nonA*⁽⁻⁾ background (line 168-8/*d*, $p=0.007$). The same happened in the case of 67-4/*m*. When tested on large stripes, this line gave significantly less correct turns when compared to line 67-4/*d* ($p=0.048$). Line 112/*m*, which carries one copy of *nonA*^{vir} inserted on the X chromosome was not significantly different when compared to *Canton S* or to flies carrying only one copy of the *melanogaster nonA*⁺ gene or one copy of *nonA*^{vir}. In conclusion, for some inserts (113, 67-4, and 168-8) there is some evidence that adding a copy of the *mel nonA*⁺ gene increases the number of optomotor errors. However, for inserts 135, 97, 191, 297-6, there is little difference.

2 *mel*

Adding one extra copy of the gene *nonA* essentially had very little effect on the visual behaviour of the flies. As shown in figure 6.3, line 2*mel* showed a poorer performance when tested on large stripes, but in any case this performance was not significantly poorer than the one obtained with wild-type *Canton S* flies or with flies carrying only one copy of *nonA*⁺ (line 235R11/*d*).

3 *mel*

When tested on either stripes, flies carrying three copies of the *mel nonA*⁺ gene scored a little worse than flies carrying two copies of the gene but the difference is not statistically significant. The visual behaviour of flies carrying three doses of *nonA*⁺ was also not significantly different from the one displayed by *Canton S* flies or flies carrying only one copy of the *mel nonA*⁺ transgene.

mel+2 vir

The performance of flies carrying one dose of the *mel nonA*⁺ gene and two copies of the *nonA*^{vir} insert showed no significant difference to that of flies carrying only one copy of either gene.

DISCUSSION

In this chapter the visual behaviour of transformant flies carrying one, two or three copies of *nonA*⁺ and/or *nonA*^{vir} has been analyzed using the walking optomotor test. It was confirmed that a transformant line carrying the putative *D.melanogaster nonA*⁺ gene (235R11; Jones and Rubin, 1990) completely rescues the visual defect of *nonA*⁽⁻⁾ flies (Rendahl *et al.*, 1992).

In this study two different set of stripes were used in the optomotor test. The optomotor response is known to be dependent on the pattern wavelength of the visual stimulus (Götz, 1964; Heisenberg and Buchner, 1977; Blondeau and Heisenberg, 1982). The pattern wavelength λ is given by the angle subtended by a pair of black and white stripes. In the experiments reported in this chapter the pattern wavelength λ was 26.6° for the small stripes, and 72° in the case of the large stripes. An average response curve for walking wild-type flies, as a function of the pattern wavelength used, was first produced by Gotz (1964), and is presented by Heisenberg and Buchner (1977). The optomotor response tends to be at an optimum when pattern wavelengths between 55° and 19° are used, while it decreases precipitously at larger, and rather more gently at smaller values. Based on these observations, it would be expected that the large stripes used in the experiments described in this chapter ($\lambda=72^\circ$) represent a more challenging test when compared to the small ones. In contrast, the small ones should be eliciting the optimum optomotor response, under the same conditions. Indeed it was found that almost all genotypes performed better when tested on small stripes. However wild type flies were influenced very little by the change in wavelength pattern.

All the tranformants carrying one copy of the putative *nonA*^{vir} gene seem to rescue the visual defect of *nonA*⁽⁻⁾ rather well. I could not directly analyze *nonA*^{(-)/Y} males as they were almost impossible to obtain. Instead I used *nonA*^{diss} mutant males as the negative control. In my test, *nonA*^{diss} males cannot see the rotation of the stripes, as in Kulkarni *et al.* (1988) and Rendahl *et al.* (1992). Line 113/d and its

derivatives (line 67-4/d and 168-8/d) displayed the best optomotor responses indicating a complete rescue of the visual phenotype. Interestingly these lines carry a *nonA*^{vir} insert that appears to have lost some non-coding sequence at its 3' end (see chapter 4). The fact that these lines completely rescue the poor *nonA*⁽⁻⁾ optomotor response gives an indirect proof that their insert still carries the complete sequence for the fifth exon. This exon is in fact indispensable for the rescue of all the *nonA*-related biological functions (Jones and Rubin, 1990, Rendhal *et al.*, 1992). Whatever sequence has been lost from the transgene in lines 113, 67-4, and 168-8, it has to carry no biological function related to visual behaviour. This is particularly interesting when compared to the situation found in the case of the complex *optomotor-blind* (*omb*) gene. At the core of this locus lies a lethal complementation group (*l(1)omb*) whose mutations cause larval and pupal lethality (Plugfelder *et al.*, 1990; Plugfelder *et al.*, 1992). The recessive viable allele *In(1)omb*^{H31} which causes neuroanatomical and optomotor defects, is the result of an inversion affecting a proximal, non-transcribed part of the locus. This 40 Kb downstream regulatory region has been termed "OLR" for optic lobe regulatory region. Brunner *et al.* (1992) have shown that removing increasing amounts of DNA from this region results in a gradient of increasing severity of the neuroanatomical and visual defects.

On the basis of the results described above, it seems more likely that, in the case of *nonA*, all the neuroanatomical information related to vision is restricted to the coding region or to other sequences not very far from it. Recent studies involving *in-vitro* mutagenesis of *nonA* in order to study the behavioral effects of mutations affecting NONA's putative RNA-binding motives (Rendahl *et al.*, 1996; Stanewsky *et al.*, 1996), suggest that the N-terminal RNA-recognition motif (RRM1) is necessary for all the known biological functions of NONA. However, the C-terminal RRM2 domain appears to be primarily involved in vision. In agreement with this stands the fact that the chemically induced mutants with specific defects to the visual system (*nonA*^{H2}, *nonA*^{P14}) also map in or near the RRM2 domain (Rendahl *et al.*, 1996). Lines 135 and 191 seem to be slightly less effective than the other lines in rescuing the *nonA*⁽⁻⁾ defect, although their phenotype is far from being mutant. These discrepancies could simply be the result of position effects which imply that a particular genomic location

of the transduced DNA insert could lead to a lower than normal expression of the transformed gene (Hazelrigg *et al.*, 1984; Baylies *et al.*, 1987). Among all the *in vivo* and *in vitro* mutations isolated so far, not a single one has ever affected solely the singing behaviour. It seems as if, whenever a mutation in *nonA* occurs, the "first to go" is always the visual phenotype (Rendahl *et al.*, 1992; Rendahl *et al.*, 1996; Stanewsky *et al.*, 1996). Bearing this in mind, one could be tempted to think that a transgene rescuing the visual defect should consequently fully rescue the song phenotype as well.

However there have been cases in which particular *nonA*⁺ transgenes seemed to rescue better the visual defect than song phenotype. Rendahl *et al.* (1992) found that two transformed lines, carrying the 11Kb *EcoR1* genomic fragment from Jones and Rubin (1990), restored visual function better than the song phenotype in *nonA*^{diss} mutants. These investigators tried to explain this finding suggesting the possibility that the 11Kb fragment could be missing some particular "thoracic enhancers" causing poor restoration of the biological activity necessary for normal song. They argued that maybe larger genomic fragments might fully restore both *nonA* associated phenotypes. However this finding could be more parsimoniously explained as a result of a position effect, since other lines carrying the same fragment fully rescued the visual and song phenotypes (Rendahl *et al.*, 1996). With this in mind it is critical to analyze many independent lines carrying a particular DNA insert. If at least one of these lines can rescue both phenotypes, then the transduced DNA fragment has to contain all the essential enhancers for the relevant behaviours. In this study genomic DNA containing the putative *nonA*^{vir} gene seemed to rescue the optomotor deficiencies of *nonA*⁽⁻⁾ flies. Thus there is a strong possibility that the same fragment will also rescue the mutant song phenotype. Any deviation of the transformants' song pattern from wild-type could thus be due to species-specific song information encoded within *nonA*^{vir}.

The changes in behavioural phenotype caused by one or more copies of a particular transgene have been studied with the clock gene *per*. Males and females carrying one versus two copies of the same transgene produce alterations in the circadian period (Cooper *et al.*, 1994). In the case of *per*, the locomotor activity

rhythm of the fly is shortened by increasing the levels of *per* expression and lengthened by decreasing them (Cooper *et al.*, 1994; Baylies *et al.*, 1987). The possibility of a dosage effect of the *nonA*⁺ gene on visual behaviour was tested by Stanewsky *et al.* (1996). These investigators studied the optomotor response of flies carrying two or three doses of the *nonA*⁺ gene. They found that males of one line which carried a homozygous *nonA*⁺ insert on a *nonA*⁺ background (i.e. carrying three copies of *nonA*⁺) seemed to have a mutant optomotor response. However this turned out to be due to the site of insertion of that particular transgene, since the same insert, when tested with a different *nonA*⁺ insert, plus the endogenous *nonA*⁺ gene, gave normal visual behaviour.

In this chapter the optomotor response of flies carrying two or three copies of *nonA*⁺ has been tested again, but using a different *nonA*⁺ insert than the one described by Stanewsky *et al.* (1996). The results with these transformants confirm the findings of Stanewsky *et al.* that increasing the doses of *nonA*⁺ has no influence on the visual behaviour. Here we also tested the effect of adding one or two doses of *nonA*^{vir} to flies with a *nonA*⁺ background. In two cases (*168-8/m* and *67-4/m*), when compared to hemizygous males carrying *nonA*⁺ or *nonA*^{vir}, a significant increase of the number of errors made in the optomotor test was observed. Furthermore a similar result is obtained with line *113*. However, *168-8* and *67-4* were derived from *113*, and these effects are not observed with the other lines where the insert appears to be intact. When males bearing one dose of *nonA*⁺ and two copies of *nonA*^{vir} were tested on the optomotor (*mel+2vir*), they performed similarly to hemizygous males carrying only one copy of either gene. Thus increasing the dose of the *nonA*^{vir} gene seems to have no effect on optomotor behaviour.

Given the successful rescue of optomotor behaviour, song analysis in these transformants should provide a reasonably unambiguous interpretation of whether the *nonA*^{vir} gene rescues the mutant phenotype. These song studies are reported in the next chapter.

CHAPTER 7

Song analysis

INTRODUCTION

This chapter presents a detailed analysis of the courtship song produced by the same transformant lines whose visual behaviour has been described in the previous chapter. From the results of chapter 6 it is clear that the *nonA^{vir}* transgene is able to rescue the visual defect of *nonA⁽⁻⁾* flies. Here the lovesongs produced by *nonA⁽⁻⁾* males carrying the *nonA^{vir}* transgene will be compared to the songs generated by *nonA^{diss}* mutants, *nonA⁽⁻⁾[nonA⁺]* transformants, and *D.virilis* males. The aim is to investigate whether the *nonA^{vir}* gene rescues the *nonA⁽⁻⁾* song phenotype, and whether it carries any species-specific song characteristic. The songs of heterozygous *nonA⁺/nonA^{vir}*, and *2mel*, *3mel* and *mel+2vir* (see chapter 6) males will also be analysed in order to test the possibilities of *nonA* dosage effects, and if the *nonA^{vir}* gene carries any semidominant *virilis*-like song characteristics.

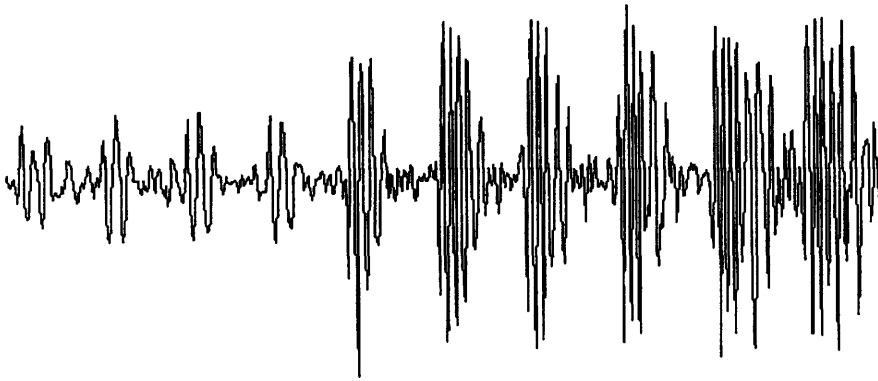
The *D.virilis* song differs from that of *D.melanogaster* in several quantitative aspects. Firstly, the sine song component, which, from my own observations, in *D.melanogaster* can often account for as much as half of the total song produced, is completely absent from the songs produced by *D.virilis* (Hoikkala, 1985). Figure 7.1 shows clearly the differences between the pulse song of *D.melanogaster* and *D.virilis*. *D.virilis* pulses are more polycyclic than those of *D.melanogaster* (range of 4.5-6.5 cycles compared to the 1-3 of *D.melanogaster*). Furthermore, the average *D.virilis* intrapulse frequency (PF) is higher, usually > 400 Hz in *D.virilis* compared to < 350 Hz in *D.melanogaster* (see results). In addition, pulse trains are shorter in *D.virilis* (rarely containing more than 10-11 pulses), and many of the pulses tend to fuse together so that there are short gaps between the end of one pulse and the beginning of the next. Consequently, *D.virilis* IPIs are shorter than those of *D.melanogaster* (Hoikkala, 1982).

The *nonA^{diss}* mutant lovesong has already been described in chapter 1. Its salient feature is the tendency of pulses towards the end of long trains (containing > 6 pulses) to become high both in cycles number (CPP) and amplitude (Kulkarni *et al.*,

D.melanogaster



nonA^{diss}



D.virilis

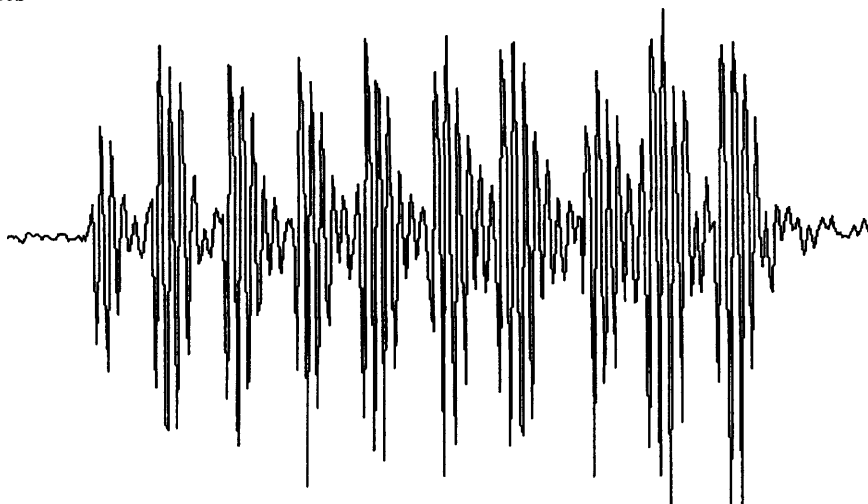
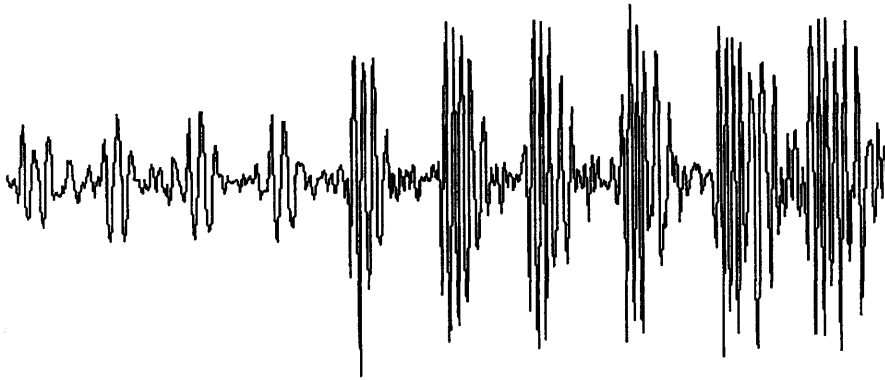


Figure 7.1. Pulse song of *D.melanogaster*, *nonA^{diss}* and *D.virilis*.

D.melanogaster



nonA^{diss}



D.virilis

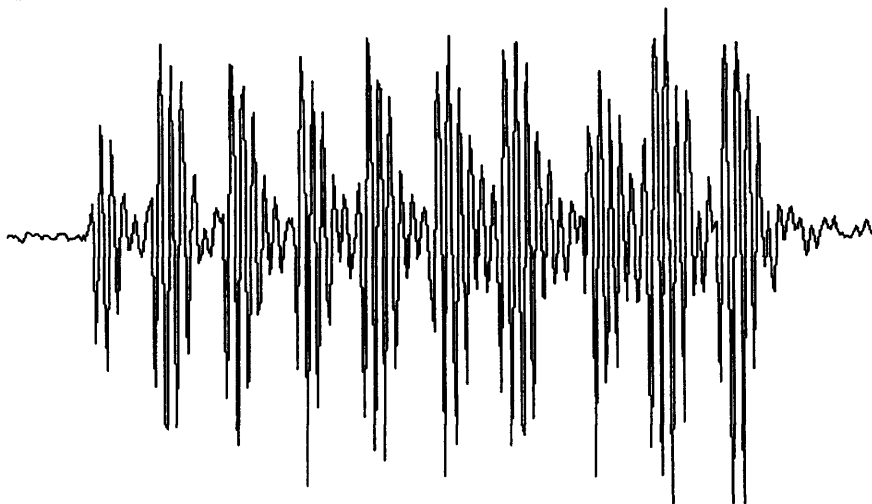


Figure 7.1. Pulse song of *D.melanogaster*, *nonA^{diss}* and *D.virilis*.

1989, see figure 7.1). Wheeler *et al.* (1989) also extended the study of these mutant songs and found previously undetected qualitative aberrations in the pulse and sine waveform shapes. In this chapter only quantitative characteristics of the songs will be taken into consideration. The parameters studied for each song were described fully in chapter 2 and are:

CPP: cycles per pulse.

bCPP: value of the CPP regression slope.

SSP: sine song proportion.

PF: pulse frequency.

bPF: value of the PF regression slope.

SSF: sine song frequency.

IPI: mean interpulse interval.

MATERIALS & METHODS

GENOTYPES

The genotypes analyzed were the same as those tested for optomotor behaviour (see chapter 6). *nonA* mutant songs were recorded from males of the strain *diss/Y; ry⁵⁰⁶/MKRS*. Other *diss* strains tested failed to produce an easily recognizable mutant song. These stocks had been kept in J.C. Hall's lab for several years and had probably accumulated modifiers. No songs were recorded from the *nonA⁽⁻⁾* genotype (see chapter 5) since no flies of this genotype could ever be generated (see chapter 6). Line *235R11/d* (see chapter 6) provided wild-type control songs. The same line had already been used as the positive control (under a different name, J. Hall, personal communication) in the song analysis performed by Rendahl *et al.* (1996). Therefore I expected to obtain reliable *melanogaster*-like wild-type songs from it.

SONG ANALYSIS

10 males from each genotype were recorded, as described in chapter 2. For most genotypes 10 min of song from each fly were processed and analyzed using the Spike2 program and a Microsoft Excel 5.0 macro written by M. Couchman, as described in chapter 2. Only 6 songs were analysed for the genotype 67-4/m. 5 min of song only were analyzed from flies of the genotypes *diss*, 97/d, 191/d, 296-6/d, and 135/m and *vir*.

RESULTS

It became clear during song recording that some flies sang more than others. In some cases the difference could be over an order of magnitude. For example, within ten minutes of recording, one fly of the *3mel* genotype produced a total of 171 seconds of song, while one fly of the *113/m* genotype sang for a total of only 15 seconds. If any of the song characters, such as CPP or bCPP, changed systematically in relation to the amount of song, then this would have to be covaried out of the song measure. For example, imagine that one fly sang twice as much as another, and CPP increased linearly with the amount of song, perhaps due to fatigue. An adjustment would have to be made between the two flies' CPP values, via regression, so that they could be compared in the absence of any differences in global song output.

As an index of song output I decided to use the sum of the total amount of pulse song plus the total amount of sine song produced by each fly (PS+SS). PS and SS were calculated as described in chapter 2. For each genotype, Pearson product-moment correlations were then performed between the PS+SS parameter and each of the song characters of interest. The results are presented in table 7.1.

Table 7.1. Correlation matrix between song output parameters (PS+SS) and other song characters in each genotype. CPP= cycles per pulse. bCPP= cycles per pulse gradient. SSP= proportion of sine song over the total amount of song (pulse song + sine song). PF= pulse frequency. bPF= pulse frequency gradient. SSF= sine song frequency. IPI= mean interpulse interval. * indicates significant r values at $p < 0.05$. ** indicates significant r at $p < 0.01$.

PS+SS	CPP	bCPP	SSP	PF	bPF	SSF	IPI
<i>diss</i>	-0.56	0.54	0.82**	-0.19	0.31	0.68*	0.03
<i>235R11/d</i>	0.81**	-0.50	0.41	-0.40	0.46	0.18	-0.15
<i>113/d</i>	-0.30	-0.12	0.71*	-0.29	-0.04	0.25	0.18
<i>67-4/d</i>	-0.03	-0.18	-0.01	-0.23	0.42	0.52	0.04
<i>168-8/d</i>	0.20	0.34	0.72*	-0.60	0.44	0.40	-0.06
<i>135/d</i>	0.65*	0.05	-0.03	0.22	0.15	0.48	-0.13
<i>97/d</i>	0.27	-0.04	0.46	-0.43	0.27	0.51	-0.54
<i>191/d</i>	0.53	0.11	0.32	-0.41	0.24	-0.04	-0.11
<i>297-6/d</i>	0.16	0.44	0.73*	-0.37	0.22	0.05	0.24
<i>112/m</i>	0.60	-0.16	0.45	-0.42	-0.06	-0.71*	0.70*
<i>113/m</i>	0.59	-0.24	0.56	-0.62	0.58	0.00	-0.03
<i>67-4/m</i>	-0.63	-0.34	0.92**	-0.64	0.75	-0.18	-0.21
<i>168-8/m</i>	-0.59	0.26	0.74*	-0.15	-0.31	0.11	0.17
<i>135/m</i>	0.59	-0.33	0.87**	-0.61	0.57	-0.07	0.29
<i>97/m</i>	-0.32	-0.04	0.22	-0.60	0.58	-0.17	0.49
<i>2mel</i>	0.20	0.03	0.34	-0.53	0.60	0.28	0.17
<i>3mel</i>	0.54	-0.65*	0.28	-0.25	0.63	-0.32	0.05
<i>mel+2vir</i>	0.13	-0.49	0.33	-0.69*	0.81**	0.60	-0.20
<i>vir</i>	0.76*	-0.64*	--	-0.51	-0.24	--	0.41

From table 7.1 it is clear that no consistent relationship between PS+SS and characters such as CPP, bCPP, bPF, SSF or IPI can be found. In some genotypes the correlation between song output index and each one of these parameters is positive, while in some others it is negative. Furthermore in almost all the cases these correlations are not significant. A different result is found when correlating PS+SS with SSP. In this case almost all the genotypes display a positive correlation between these two characters, and in several genotypes the relationship is highly significant. PF also seems to be related to the song output parameter: consistent negative (but largely non-significant) correlations between PS+SS and PF can be found for all the genotypes except one.

In conclusion, there is no compelling evidence that characters such as CPP, bCPP, or bPF, which could potentially be used to distinguish between species-specific, *virilis*-like song rescue, or incomplete, *diss*-mutant like rescue, are related to quantity of song output. However, SSP and PF do seem to genuinely related to the amount of song generated, so caution must be exercised in using these characters, and any effect of differential song production between genotypes must be removed before any comparison.

bCPP.

The salient feature of a mutant *nonA^{diss}* song is that the number of cycles per pulse (CPP) is not constant during each single pulse train. In long trains (containing more than 5-7 pulses) the pulses toward the end of the train become increasingly more polycyclic, sometimes bearing up to 30 cycles per pulse. Wild type songs never show such a feature and in general their pulse trains contain a constant number of cycles per pulse (usually between 1 and 3) (Kulkarni *et.al.*, 1987).

The first song feature to be taken into consideration in the analysis of the songs recorded was therefore the CPP regression gradient (bCPP). A CPP least squares regression was performed for each song as described in chapter 2. The bCPP gives an indication, for each song, of how the pulses change during pulse trains: a significant positive bCPP indicates a tendency of the number of CPPs to increase as the train of pulses progresses. A non significant, very small bCPP will suggest a tendency of the pulses to remain constant during pulse trains.

For each genotype, the number of songs out of ten which had a significant CPP regression gradient (bCPP) was calculated (table 7.2). For genotype 67-4/m this number was calculated out of only 6 songs analyzed. The significant gradients were divided into positive or negative. A negative gradient would suggest that the pulses at the beginning of the train were actually more polycyclic than those towards the end. The average value of bCPP was also calculated. Notice that the genotypes which had

very few songs with significant regressions also had a very small average bCPP value (see *113/m* or *3mel*). In these genotypes it seems that the CPP remains constant during the progression of a pulse train.

Table 7.2. bCPP (cycle per pulse regression gradient) for different genotypes. N= number of significant bCPPs out of 10 songs (* out of 6 songs). + = significant positive bCPPs. - = significant negative bCPPs. The mean bCPP value per genotype is also indicated.

GENOTYPE	N	+	-	Mean bCPP
<i>diss</i>	10	10	0	0.257
<i>235R11/d</i>	4	1	3	-0.002
<i>113/d</i>	7	7	0	0.004
<i>67-4/d</i>	8	8	0	0.019
<i>168-8/d</i>	8	8	0	0.016
<i>135/d</i>	8	8	0	0.022
<i>97/d</i>	7	6	1	0.012
<i>191/d</i>	7	6	1	0.024
<i>297-6/d</i>	7	6	1	0.014
<i>112/m</i>	5	3	2	0.001
<i>113/m</i>	2	0	2	-0.003
<i>67-4/m*</i>	4	4	0	0.014
<i>168-8/m</i>	5	5	0	0.010
<i>135/m</i>	5	4	1	0.005
<i>97/m</i>	3	3	0	0.004
<i>2mel</i>	8	7	1	0.009
<i>3mel</i>	2	0	2	0.000
<i>mel+2vir</i>	3	2	1	0.001
<i>vir</i>	10	10	0	0.186

As expected, all the *diss* song recorded had a strongly significant positive gradient and the average value of the positive slope was high. Unexpectedly, the results obtained with *vir* were strikingly similar to those given by *diss*: all the songs recorded had a significantly positive regression slope and the average value of bCPP was also very similar to that found in *diss*. Only four songs from the *235R11/d* genotype had a significant CPP regression and in three this regression had a negative

slope. The average bCPP for this genotype had a slightly negative value. Therefore in line *235R11/d* the amount of CPP did not increase with the progressing of the pulse train as opposed to what happens in *diss*, thus confirming that this line, which carries one copy of *D.melanogaster nonA*⁺ over a *nonA*⁽⁻⁾ background, gives effective wild-type rescue of this song character (Rendahl *et al.*, 1992; 1996).

All the transformant lines carrying one copy of *nonA*^{vir} over *nonA*⁽⁻⁾ background (from genotype *113/d* to *297-6/d* in table 7.2) produced songs with mostly significant positive CPP regressions. Also the average value of bCPP was positive in all these lines even though it was generally at least 10 times smaller than the average slope value for *diss* regressions (see also figure 7.2a). Heterozygous lines carrying one copy of *nonA*^{vir} plus the endogenous *nonA*⁺ gene (genotypes *112/m* to *97/m* in table 7.2) had a much lower number of significant regression slopes when compared to their hemizygous counterparts. Also some of the significant slopes were negative (in *113/m* 2 out of 2 significant slopes were negative). The average bCPP had also much lower values in these genotypes if compared with those of the *nonA*^{vir} hemizygous lines. In general these values resembled closely those found for line *235R11/d*. The only two exceptions were perhaps lines *67-4/m* and *168-8/m* for which all the significant regression gradients were positive and the average bCPP values were about ten times greater than those from the other heterozygous lines. A t-test comparing the proportions of positive significant slopes in hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] lines with those in heterozygous *nonA*^{vir}/*nonA*⁺ revealed a significant difference between these two groups (*t*=3.701; *p*=0.003).

Line *2mel* which carries two copies of the *nonA*⁺ gene had twice as many significant CPP regressions than the line carrying one copy of *nonA*⁺ (*235R11*). Also almost all the gradients were positive. Consequently the average bCPP value was also slightly positive. Line *3mel* carrying three copies of *nonA*⁺ had only two significant regressions both of them with a negative slope. The average value of bCPP for this genotype was 0 indicating that the song pulses maintain the same number of cycles throughout the pulse trains. The results obtained from the line carrying two copies of *nonA*^{vir} and one of *nonA*⁺ (*mel+2vir*) were very similar to those of line *3mel*. Both

lines behave very similarly to line 235R11/d, indicating that no consistent dosage effect was exerted on this song character, in spite of differences between line 2mel and line 235R11/d.

One way ANOVA was performed using the bCPP values of the songs analyzed for each genotype. When the data for all the genotypes was included, the result was highly significant ($F= 11.155$; $df = 18, 167$; $p < 0.001$). When the bCPP values of *diss* and *vir* were excluded from the analysis (note in figure 7.2a the huge variance in bCPP values for these two genotypes), one-way ANOVA still gave a significant main effect ($F= 4.263$; $df= 16, 149$; $p < 0.001$). Figure 7.2a shows the mean, the standard error and standard deviation of the bCPP values in each genotype, and figure 7.2b illustrates the same values for the transformant lines only. Newman-Keuls *a posteriori* test revealed that all the transformant lines were significantly different from both *diss* and *vir*.

From figure 7.2b it is evident that all the hemizygous $nonA^{(-)}[nonA^{vir}]$ lines have, in general, higher bCPP values than the other transformant lines. A planned comparison between the control line 235R11/d and all the hemizygous $nonA^{(-)}[nonA^{vir}]$ lines (from line 113/d to line 297-6/d, in figure 7.2) revealed that these two groups differ significantly ($F= 17.321$, $df= 1, 149$; $p < 0.001$). In particular, a Newman-Keuls test, performed excluding the data from *vir* and *diss* (see appendix 7.1), revealed that lines 67-4/d, 135/d and 191/d differ significantly from the positive control line 235R11/d. Heterozygous $nonA^{vir}/nonA^{+}$ lines showed, in general, similar or slightly higher bCPP values than those of the control 235R11/d. A planned comparison between all the $nonA^{vir}/nonA^{+}$ lines (from line 112/m to line 97/m in figure 7.2) and line 235R11/d showed that these two groups do not differ significantly ($F= 2.824$; $df= 1, 149$; $P= 0.095$) from each other.

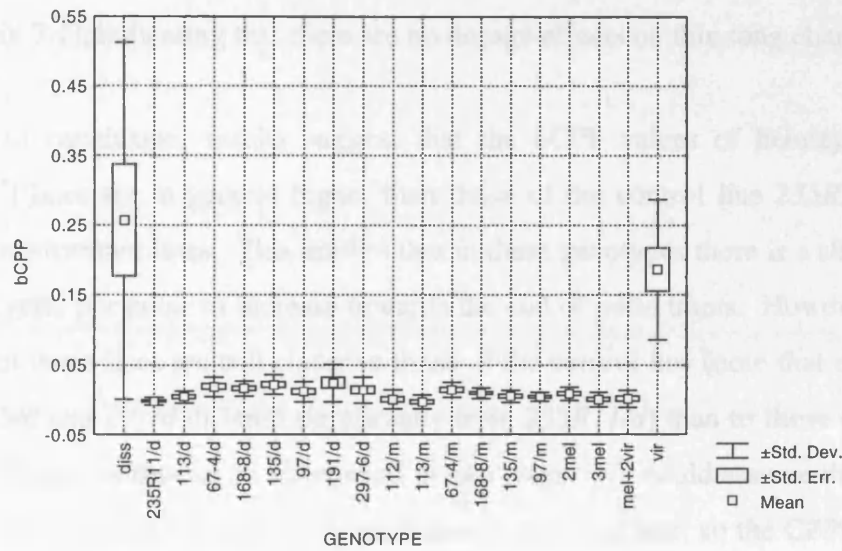
From figure 7.2b it is clear that all the hemizygous $nonA^{(-)}[nonA^{vir}]$ lines have higher bCPP values than the heterozygous $nonA^{vir}/nonA^{+}$ group. A planned comparison between these two groups showed in fact that they significantly differ ($F= 20.921$; $df= 1, 149$; $p < 0.001$) from each other. When comparing each hemizygous $nonA^{(-)}[nonA^{vir}]$ line with its heterozygous $nonA^{vir}/nonA^{+}$ counterpart (for example

Figure 7.2.

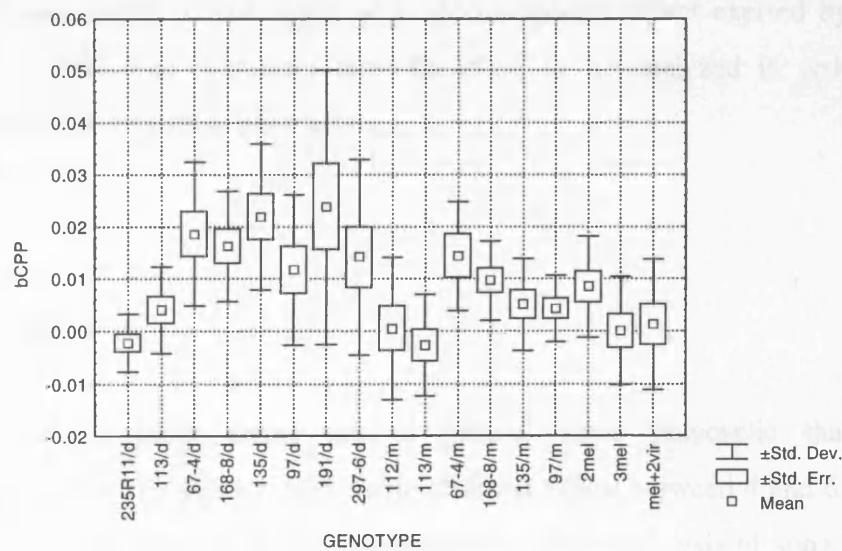
A: bCPP values of all genotypes.

B: bCPP values of transformant lines only.

A



B



113/d with *113/m*, in figure 7.2b) it is evident that the hemizygous lines have always higher bCPPs than their heterozygous counterparts. (although not significantly, see Newman-Keuls results in appendix 7.1)

bCPP values of line *2mel* are not significantly different from the positive control *235R11/d*, neither are those of lines *3mel* and *mel+2vir* (see Newman-Keuls in appendix 7.1), indicating that there are no dosage effects on this song character.

In conclusion, results suggest that the bCPP values of hemizygous *nonA⁽⁻⁾* [*nonA^{vir}*] lines are in general higher than those of the control line *235R11/d* and the other transformant lines. This implies that in these genotypes there is a slight tendency of the cycles per pulse to increase towards the end of pulse trains. However the bCPP values of these lines are still closer to those of the control line (note that only lines *67-4/d*, *135/d* and *191/d* differed significantly from *235R11/d*) than to those of either *diss* or *vir*. These results can be interpreted in two ways. We could assume that the rescue of this song character in *nonA⁽⁻⁾/nonA^{vir}* lines is not complete, so the CPPs still tend to increase slightly toward the end of each song. However *vir* songs also show a tendency of CPP to increase during pulse trains (the bCPP average values of *vir* are in fact very similar to those of *diss*, see table 7.1) so we could also argue that the higher bCPP values might be the result of a species-specific effect exerted by the *nonA^{vir}* gene. Other songs characters have therefore to be analyzed in order to try to distinguish between these alternatives.

CPP.

D.virilis pulse songs are in general more polycyclic than those of *D.melanogaster*, the average CPP value of *virilis* falling between 4 and 6 (Hoikkala *et al.*, 1982), as opposed to 1-2 in *melanogaster*. In the analysis of song produced by transformant lines it was essential to check whether transformants carrying the *nonA^{vir}* gene produced more polycyclic songs than transformants carrying the *nonA⁺* gene.

For each song, CPP values can be simply extrapolated from the CPP regression curve $y=bx+c$ at different values of x , where x represents pulse position in a train. An alternative analysis could consist of using the raw CPP values found in each song. Both approaches were used. In the first kind of analysis CPP values were calculated using the CPP regression of each song when $x=1$ (CPP1) and $x=10$ (CPP10). The second analysis consisted in calculating in each song how many pulses could be found that had a CPP value of 1, 1.5, 2, 2.5, 3, 3.5, 4, and so on. The results are presented below.

CPP1.

The CPP values based on the regression function ($CPP=bx+c$) when $x=1$ were calculated for each song. An one way ANOVA was employed to analyze these data. When the CPP1 values from all genotypes were used, the test gave a highly significant F ratio ($F= 22.696$; $df= 18, 167$; $p<< 0.001$). From figure 7.3a it is clear that in *vir* the first pulse in a train contains a relatively high number of cycles, and that this number tends to be quite variable. Large variation is found in the value of CPP1 of *diss* as well, even though in this genotype the average CPP1 value is lower than in *vir*, and more similar to that of the transformant lines. Newman-Keuls comparisons confirm that no significant difference exist between the average CPP1 value of *diss* and that of the transformant lines.

The ANOVA was repeated including only CPP1 values from the transformant lines. The F ratio was highly significant ($F= 15.501$; $df= 16, 149$; $p<< 0.001$). Figure 7.3b shows the mean, standard error and standard deviations of CPP1 values of the transformant lines only. A planned comparison between the control line *235R11/d* and the group composed of hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] lines showed that they differ significantly ($F= 16.413$; $df= 1, 149$; $p<< 0.001$). In particular, Newman-Keuls *a posteriori* comparisons (appendix 7.2) revealed that CPP1 values in *67-4/d*, *168-8/d*, *135/d* and *297-6/d* are significantly higher than in the control line *235R11/d*.

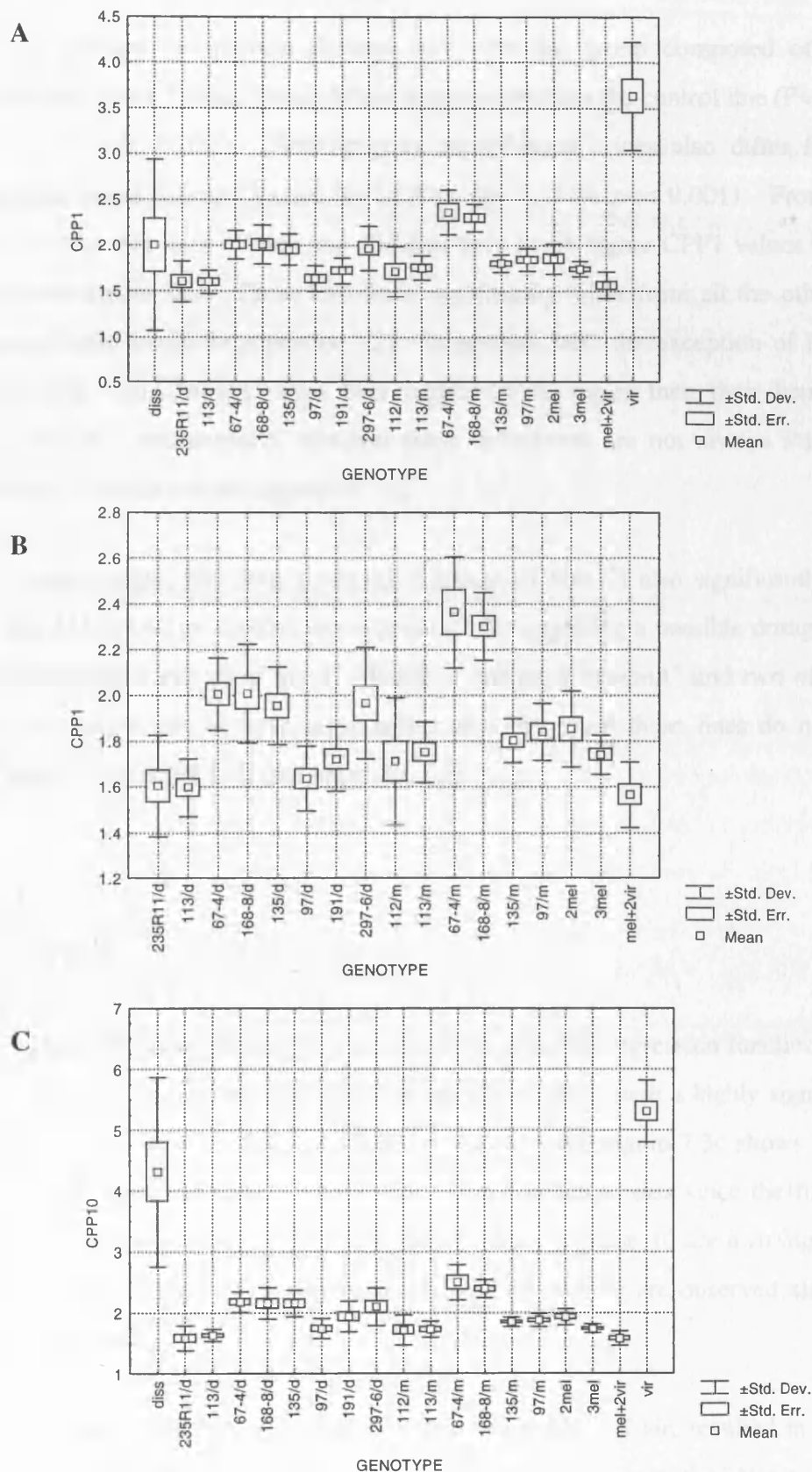


Figure 7.3.

A: CPP values at $x=1$ for all genotypes.

B: same as in A excluding *diss* and *vir*.

C: CPP values at $x=10$ for all genotypes.

A planned comparison showed that even the group composed of all the heterozygous *nonA^{vir}/nonA⁺* lines differs significantly from the control line ($F=36.084$; $df=1, 149$; $p < 0.001$). Heterozygous *nonA^{vir}/nonA⁺* lines also differ from the hemizygous *nonA⁽⁻⁾[nonA^{vir}]* ones ($F=14.870$; $df=1, 149$; $p < 0.001$). From figure 7.3b it is clear that lines 67-4/*m* and 168-8/*m* have much higher CPP1 values than any other transformant line. These two lines significantly differ from all the others (see Newman-Keuls results in appendix 7.2). In general, with the exception of line 135, heterozygous *nonA^{vir}/nonA⁺* lines bear higher CPP1 values than their hemizygous *nonA⁽⁻⁾[nonA^{vir}]* counterparts, however these differences are not always statistically significant (Newman-Keuls, appendix 7.2).

Interestingly, line 2*mel* (carrying 2 copies of *nonA⁺*) also significantly differs from line 235*R11/d* ($p=0.0436$, see appendix 7.2) suggesting a possible dosage effect. However having 3 copies of *nonA⁺* (3*mel*) or one copy of *nonA⁺* and two of *nonA^{vir}* (*mel*+2*vir*) seems not to have much effect on CPPs since these lines do not differ significantly from 235*R11/d* (see appendix 7.2).

CPP10.

The CPP value when $x=10$ was calculated using the regression function of each song. The data was tested with one-way ANOVA which gave a highly significant F ratio ($F=52.501$; $df=18, 167$; $p < 0.001$). As expected, figure 7.3c shows how the average CPP values at pulse 10 have more than doubled in *diss* since the first pulse (compare with figure 7.3a). The average CPP values at pulse 10 are also higher than in pulse 1 in *vir* songs. Slight increases of the CPP number are observed also in the transformant lines.

One-way ANOVA, excluding the data from *diss* and *vir*, resulted in a highly significant F ratio ($F=19.555$; $df=16, 149$; $p < 0.001$). *A posteriori* Newman-Keuls comparisons (see appendix 7.3) showed that at pulse 10 all the transformant hemizygous *nonA⁽⁻⁾[nonA^{vir}]* lines, except 113/*d* and 97/*d*, are significantly more

polycyclic than the control line *235R11/d*. Also all the heterozygous *nonA^{vir}/non-A⁺* lines, except *112/m* and *113/m*, differ significantly from line *235R11/d*. *67-4/m* and *168-8/m* are again significantly more polycyclic than any other line. All heterozygous *nonA^{vir}/nonA⁺* lines are still in general more polycyclic than their hemizygous *nonA⁽⁻⁾* [*nonA^{vir}*] counterparts, with the only exception of *135/m*. At pulse 10 the CPP values of *3mel* and *mel+2vir* are still not significantly different from the control line *235R11/d*, while *2mel* is significantly more polycyclic (see appendix 7.3).

To summarize, hemizygous *nonA⁽⁻⁾* [*nonA^{vir}*] lines are in general more polycyclic than the control line *235R11/d* both at pulse 1 and 10. However at pulse 10 they also significantly differ from either *diss* or *vir* (which are much more polycyclic). Heterozygous *nonA^{vir}/nonA⁺* lines are also more polycyclic than the control line at both pulse 1 or 10. In general, they are also more polycyclic than their hemizygous counterparts, even though pairwise comparisons are not always significantly different. Again, these results could be interpreted in two ways. We could assume that the increased polycyclicality seen in *nonA⁽⁻⁾* [*nonA^{vir}*] lines when compared to the control *nonA⁽⁻⁾* [*nonA⁺*] is actually due to a species-specific effect connected to the presence of the *nonA^{vir}* gene. The fact that heterozygous *nonA^{vir}/nonA⁺* lines are actually more polycyclic than their hemizygous *nonA⁽⁻⁾* [*nonA^{vir}*] counterparts could be similarly explained invoking some sort of dominant effect linked to the *nonA^{vir}* gene. However, the *mel+2vir* genotype is normal, contradicting the notion of *virilis* dominance.

Alternatively, the polycyclicality of *nonA⁽⁻⁾* [*nonA^{vir}*] lines could be explained by assuming that the *nonA^{vir}* gene is not giving complete rescue, and the results obtained with *nonA^{vir}/nonA⁺* lines could be due to a dosage effect. The higher polycyclicality of *2mel* could also be interpreted by invoking a dosage effect. However no dosage effect can be seen in the case of *3mel* and *mel+2vir*.

CPP distributions.

For each song recorded the proportion of pulses containing 1, 1.5, 2, 2.5, 3,

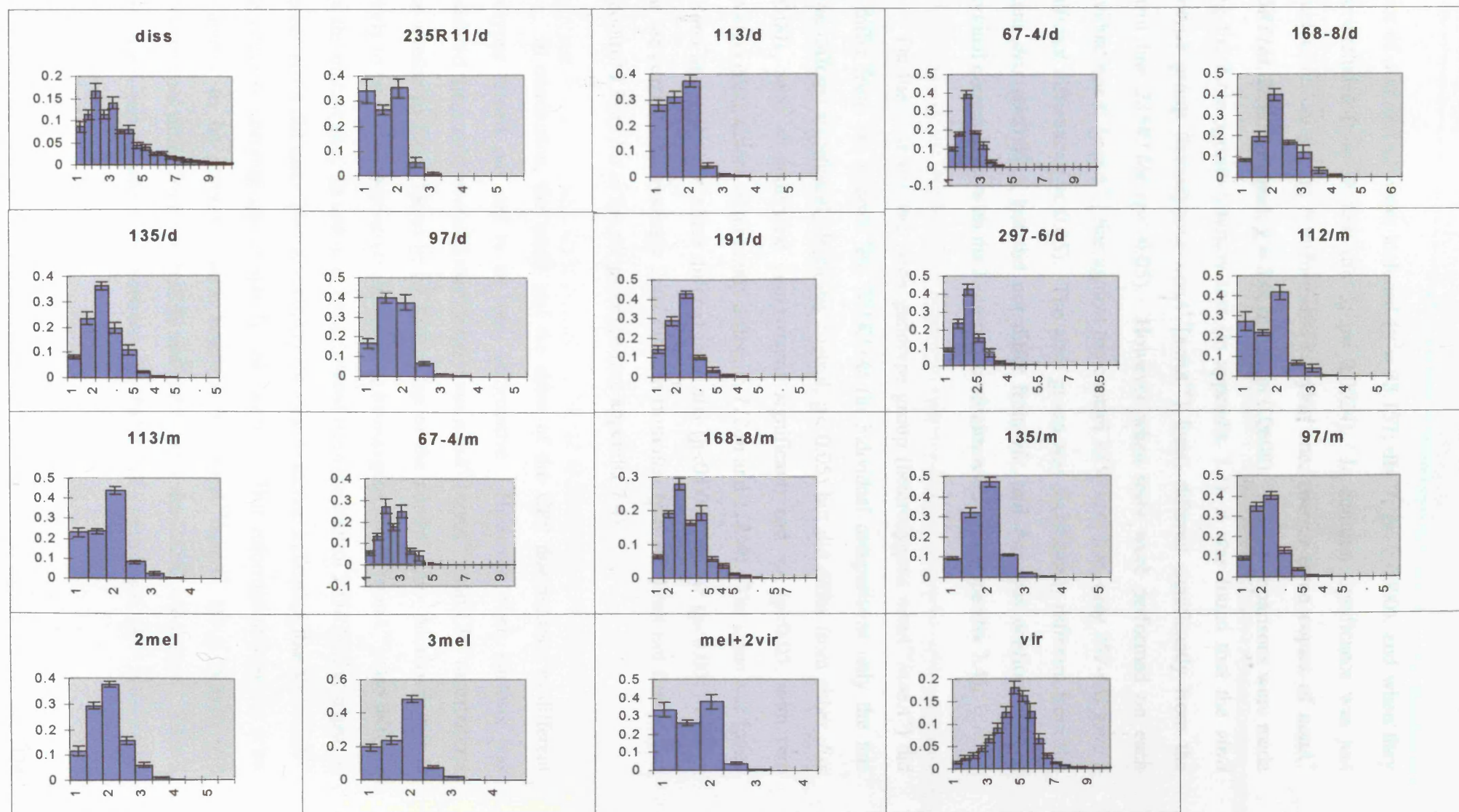
3.5, etc.. cycles was estimated. The average proportions per genotype were then calculated. Frequency histograms showing the average proportion of pulses with 1, 1.5, 2, 2.5, and so on, cycles in each genotype are presented in figure 7.4. On the x-axis the maximum number of cycles considered was 10, however it has to be noted that a small proportion of pulses containing more than 10 cycles could be found in both *diss* and *vir* (in some *diss* songs pulses with more than 30 cycles were found). To see how these CPP distributions differed between genotypes, three parameters were taken into consideration: the mode of the distribution, the skewness and the kurtosis.

The mode of each song was computed and a Kruskal-Wallis non parametric ANOVA was performed on these data. In order to reduce the number of independent (grouping) variables (which in this case are the different genotypes), all the hemizygous $nonA^{(-)}$ [$nonA^{vir}$] genotypes were grouped together as *vir/d*, and all the heterozygous $nonA^{vir}/nonA^{+}$ as *vir/m*. The result of the test was highly significant ($\chi^2 = 88.504$; $df = 7$; $p < 0.001$). However, when *diss* and *vir* were excluded from the analysis, the test was non-significant ($\chi^2 = 8.663$; $df = 5$; $p = 0.123$). A Kolmogorov-Smirnov two sample test performed on each pair of genotypes, confirmed that the only line that differs from all the others is *vir*. From figure 7.4 it is clear, in fact, that the mode of most genotypes, including *diss*, is 2. The fact that *diss* has a mode of 2 should not be surprising considering that in this mutant most of the pulses (all the pulses of shorter trains and all the pulses at the beginning of long trains) are still wild-type.

The skew values of the CPP distributions in each song were also analyzed using a Kruskal-Wallis ANOVA. Again, the number of independent variables was reduced by grouping all the hemizygous $nonA^{(-)}$ [$nonA^{vir}$] as *vir/d* and all the heterozygous $nonA^{vir}/nonA^{+}$ as *vir/m*. The test gave a significant result only when data from *diss* and *vir* were included in the analysis ($\chi^2 = 22.914$; $df = 7$; $p = 0.0018$). A Kolmogorov-Smirnov test conducted on pairs of genotypes, showed that *diss* was the genotype that differed from all the others. No significant difference was found between *vir* and any other transformant line.

The same kind of analysis was performed on the kurtosis values of each song CPP distribution. This time, the Kruskal-Wallis ANOVA was significant when the

Figure 7.4. CPP distributions in each genotype.



values of *diss* and *vir* were included ($\chi^2 = 33.157$; $df = 7$; $p = 0.0000$), and when they were excluded ($\chi^2 = 18.529$; $df = 5$; $p = 0.0024$). In addition significance was just missed when comparing transformants carrying one, two or three copies of *nonA*⁺ (*235R11/d*, *2mel* and *3mel*; $\chi^2 = 5.6$; $df = 2$; $p = 0.0608$). When comparisons were made using the Kolmogorov-Smirnov test (in appendix 7.4), it was found that the *vir/d* genotype group (hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] lines) differed significantly from the control line *235R11/d* ($p < 0.05$). However when tests were performed on each individual *nonA*⁽⁻⁾[*nonA*^{vir}] line against the control *235R11/d* line, only *297-6/d* gave a significant difference ($p < 0.05$). The *vir/d* group was significantly different from the mutant *diss* ($p < 0.001$), but did not differ from *vir*, and this was confirmed in all individual comparisons with the Kolmogorov-Smirnov test (see appendix 7.4).

On the contrary, the *vir/m* genotype group (heterozygous *nonA*^{vir}/*nonA*⁺) did not differ from the control line *235R11/d* (in individual comparisons only the line *135/m* differed significantly from the control, $p < 0.05$) but did differ from either *diss* ($p < 0.001$, with all individual comparisons significant) and *vir* ($p < 0.05$, with two individual comparisons significantly different, *112/m* and *113/m*). The same was found for *3mel* and *mel+2vir* which differed from *diss* ($p < 0.001$) and *vir* ($p < 0.001$) but not from the control. Interestingly *2mel* differed from *diss* ($p < 0.001$) but not from either the control *235R11/d* or the *vir* genotype (see appendix 7.4)

In conclusion, the mode and the skew of the CPP distributions in different genotypes turned out not to be very informative. However when kurtosis was considered the results revealed that hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] had CPP distributions more similar to those found in *vir* than those of the control line. Moreover this is unlikely to be due to incomplete rescue since hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] also differed from the mutant *diss*. In addition, *vir/d* differed from the control *235R11/d* as a group, whereas *vir/m* did not. This is probably the first evidence indicating that the *nonA*^{vir} gene could be carrying species-specific information. This information however does not appear to be dominant since heterozygous *nonA*^{vir}/*nonA*⁺ (as a group) and *mel+2vir* lines differ from *vir*, but do not differ from the *235R11/d* control. Finally *2mel* shows results which fall between the *235R11/d* control and *vir*. This could

possibly be due to some dosage effect linked to the *mel nonA* gene. Note however that lines carrying three copies of *mel nonA* (*3mel*) do not show this effect.

bPF.

For each song analyzed, a PF (pulse frequency) least square regression was performed as described in chapter 2 (see figure 2.2). Table 7.3 shows the number of songs, out of ten per genotype (out of six for *674/m*), which had a significant PF regression gradient (bPF).

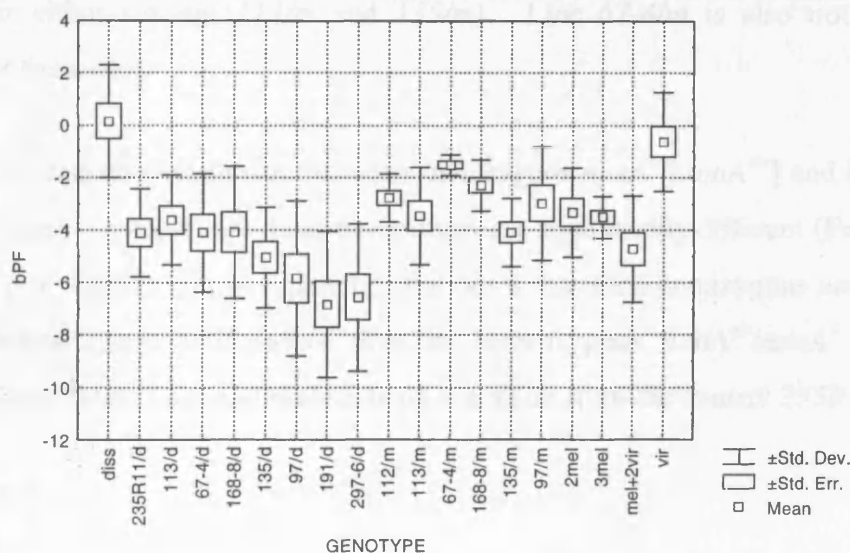
Table 7.3. bPF (pulse frequency gradient) for different genotypes. N= number of significant bPFs out of 10 songs (* out of 6 songs). + = significant positive bPFs. - = significant negative bPFs. The mean bPF value per genotype is also presented.

GENOTYPE	N	+	-	Mean bPF
<i>diss</i>	5	2	3	0.155
<i>235R11/d</i>	9	0	9	-4.107
<i>113/d</i>	10	0	10	-3.613
<i>67-4/d</i>	10	0	10	-4.104
<i>168-8/d</i>	8	0	8	-4.083
<i>135/d</i>	10	0	10	-5.041
<i>97/d</i>	8	0	8	-5.851
<i>191/d</i>	9	0	9	-6.841
<i>297-6/d</i>	10	0	10	-6.572
<i>112/m</i>	9	0	9	-2.781
<i>113/m</i>	10	0	10	-3.464
<i>67-4/m*</i>	6	0	6	-1.530
<i>168-8/m</i>	9	0	9	-2.293
<i>135/m</i>	10	0	10	-4.102
<i>97/m</i>	8	0	8	-2.988
<i>2mel</i>	9	0	9	-3.336
<i>3mel</i>	9	0	9	-3.513
<i>mel+2vir</i>	10	0	10	-4.725
<i>vir</i>	1	0	1	-0.630

From table 7.3 it is clear that in all the transformant lines (from line *235R11/d* to line *mel+vir*) almost all the songs analyzed had a significant negative bPF. In other

words, in all the transformant lines analyzed there was a trend for the pulses at the beginning of each train to have a higher frequency than those towards the end. *diss* and *vir*, on the contrary, showed a very different PF regression pattern: only 5 *diss* songs had a significant bPF, 2 of which were positive. Only one *vir* song had a significant negative PF gradient. This means that in these two genotypes the frequencies of pulses at the beginning of a pulse train are not very different from the frequencies of pulses at the end. The average bPF values per each genotype were also calculated and are shown in table 7.3. These values plus standard error and standard deviation are given in figure 7.5.

Figure 7.5. bPF values.



One-way ANOVA was employed on the bPF values of each song analyzed. The result was highly significant ($F = 8.066$; $df = 18, 167$; $p < 0.001$). When bPF values of *diss* and *vir* were removed from the analysis, the ANOVA still gave a highly significant F ratio ($F = 4.773$; $df = 16, 149$; $p < 0.001$). From figure 7.5 it appears that *diss* and *vir* have the highest bPF values of all the genotypes analyzed. A *posteriori* Newman-Keuls test (see appendix 7.5) confirms that these two lines significantly differ from the control line 235R11/d and from most of the other transformant lines, while they do not differ from each other.

The mean bPF values in hemizygous $nonA^{(-)}[nonA^{vir}]$ lines were found to be either similar or more negative than those of the control line 235R11/d (see figure 7.5). A planned comparison showed that the group including all the $nonA^{(-)}[nonA^{vir}]$ lines does not significantly differ from the control line ($F= 2.465$; $df= 1, 167$; $p= 0.118$). Newman-Keuls *a posteriori* comparisons revealed that only lines 191/d and 297-6/d have significantly lower bPF values than line 235R11/d.

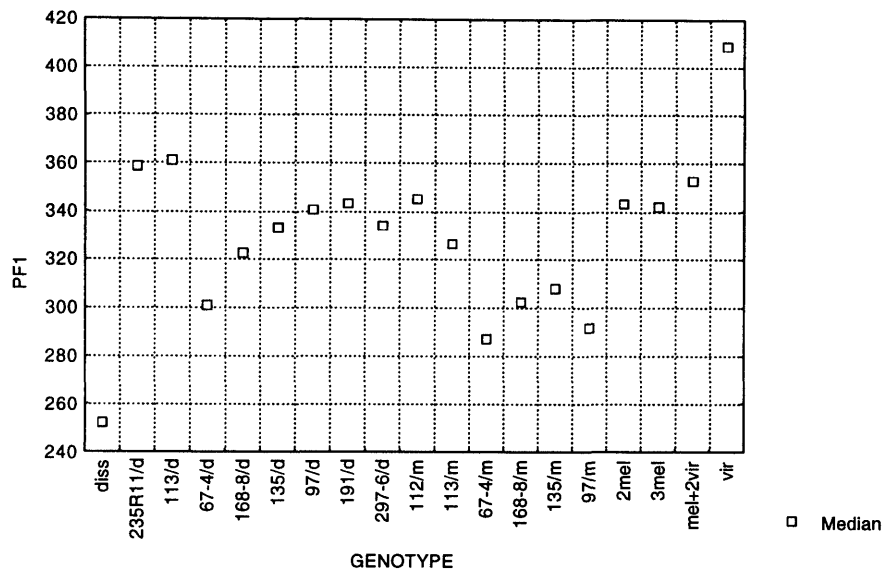
Heterozygous $nonA^{vir}/nonA^{+}$ lines have bPF values close, or slightly higher than those of the control 235R11/d (figure 7.5). A planned comparison between these lines and line 235R11/d gave a non-significant F ratio ($F= 3.35$; $df= 1, 167$; $p= 0.07$). According to the Newman-Keuls test (see appendix 7.5), no heterozygous $nonA^{vir}/nonA^{+}$ line differs from the control. However most of these lines do not differ from *vir* either (except 113/m and 135/m). Line 67-4/m is also not significantly different from *diss*.

A planned comparison between hemizygous $nonA^{(-)}[nonA^{vir}]$ and heterozygous $nonA^{vir}/nonA^{+}$ showed that these two groups are significantly different ($F= 41.083$; $df= 1, 167$; $p<< 0.001$). From figure 7.5 it appears that each hemizygous $nonA^{(-)}[nonA^{vir}]$ line displays lower bPF values than its heterozygous $nonA^{vir}/nonA^{+}$ counterpart. Finally lines 2*mel*, 3*mel* and *mel*+2*vir* do not differ from the control 235R11/d.

PF.

As discussed above, pulse frequency is one of the two song characters that seem to generally correlate (in this case negatively) with the total amount of song produced by a fly (see table 7.1). Therefore for analysis, it is necessary to take into consideration the amount of real song recorded for each fly. Consequently, for each song, the (PS+SS) value was used as the covariate for PF, which itself was calculated at $x= 1$ from the regression function $y=bx+c$ (PF1). Analysis of covariance revealed a highly significant F ratio ($F= 18.793$; $df = 18, 166$; $p= 0.00$). Corrected PF1 averages were produced for each genotype and their values are graphically plotted in figure 7.6.

Figure 7.6. Adjusted PF1 average values per genotype.



Again, it is obvious that the two genotypes that stand out are *diss* with much lower PF1 values than the rest, and *vir* which has much higher PF1s than the other lines. All the hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] lines, except *113/d*, show lower PF1 values than the control line *235R11/d*. However Newman-Keuls comparisons reveal that only lines *67-4/d* and *168-8/d* significantly differ from *235R11/d*. These two lines are also significantly different from *diss*, as are the rest of the *nonA*⁽⁻⁾[*nonA*^{vir}] lines.

In figure 7.6 it is clear how the heterozygous *nonA*^{vir}/*non-A*⁺ lines have in general even lower PF1 values than their hemizygous counterparts, although the only significant difference is between *97/d* and *97/m*. According to Newman-Keuls tests, four out of the six heterozygous *nonA*^{vir}/*nonA*⁺ lines (*67-4/m*, *168-8/m*, *135/m* and *97/m*) significantly differ ($p < 0.05$) from the control *235R11/d*, even though they also differ from *diss*. Lines *2mel*, *3mel* and *mel+2vir* had average PF1 values similar to the control line *235R11/d* and in fact they do not significantly differ from it.

To conclude, the *nonA*^{vir} gene, either on a *nonA*⁽⁻⁾ or in a *non-A*⁺ background, does not seem to have any species-specific effect on the pulse frequency of transformant flies. Furthermore the mutant *diss* pulse frequency appear to be rescued in *nonA*⁽⁻⁾[*nonA*^{vir}] lines. In fact, while *vir* PF values are on average quite high, all the

transformants carrying one or two copies of the *nonA^{vir}* gene display either similar or lower PFs than the control line transformed with *non-A⁺* (235R11). Two hemizygous *nonA⁽⁻⁾[nonA^{vir}]* lines have PF values that seem to be intermediate between the control and the mutant *diss*. This could be explained as a result of non-perfect rescue of this song character. However it is more difficult to explain why four of the heterozygous *nonA^{vir}/nonA⁺* also display low PF values.

SSP.

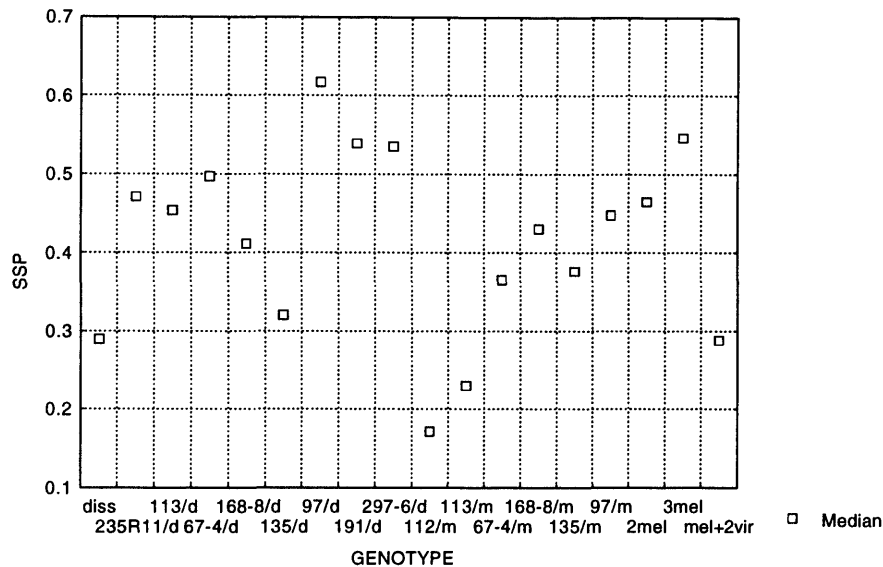
As previously described, the courtship song of *D.melanogaster* is not exclusively composed of trains of pulses. A second song component consisting of ~160 HZ hums (sine or hum song) is generally present interspersed within the pulse song. *D.virilis* males, on the contrary, produce courtship songs containing only trains of polycyclic pulses. It seemed therefore interesting to study the amount of sine song produced by transformant males carrying one or two copies of the *nonA^{vir}* gene in order to establish whether this gene had any possible effect in repressing sine song production.

The proportion of sine song produced during courtship by each male was calculated as the amount of sine song over the total amount of song (PS+SS) (see chapter 2). All the males of each genotype, including males of the *diss* line, produced at least one or two bursts of sine song. The only exception were two males of the *mel+2vir* line whose song was completely devoid of sine song.

SSP, as discussed above, seems to be almost always positively correlated with the total amount of song recorded for each fly (PS+SS). This was the case in all the genotypes analyzed except in line 67-4/d and 135/d for which the correlation was negative although not significant (see table 7.1). It seemed opportune therefore to take into account the PS+SS value of each song when analyzing SSP values. Analysis of covariance (where SSP was the dependent variable and PS+SS the covariate) gave a highly significant F ratio (F= 12.258; df genotype= 17, 157; p<< 0.001). The adjusted

SSP means per each genotype were calculated and are graphically represented in figure 7.7.

Figure 7.7. Adjusted SSP average values per genotype.



From figure 7.7, large variation of SSP averages can be noticed within each of the two groups constituted by hemizygous $nonA^{(-)}[nonA^{vir}]$ and heterozygous $nonA^{vir}/nonA^{+}$ lines. In general $nonA^{(-)}[nonA^{vir}]$ lines seemed to have higher SSPs than $nonA^{vir}/nonA^{+}$. Only line 135/d had significantly lower SSP values than the control 235R11/d, according to a Newman-Keuls test, while line 97/d had actually significantly higher SSPs. All the other $nonA^{(-)}[nonA^{vir}]$ lines did not differ from the control.

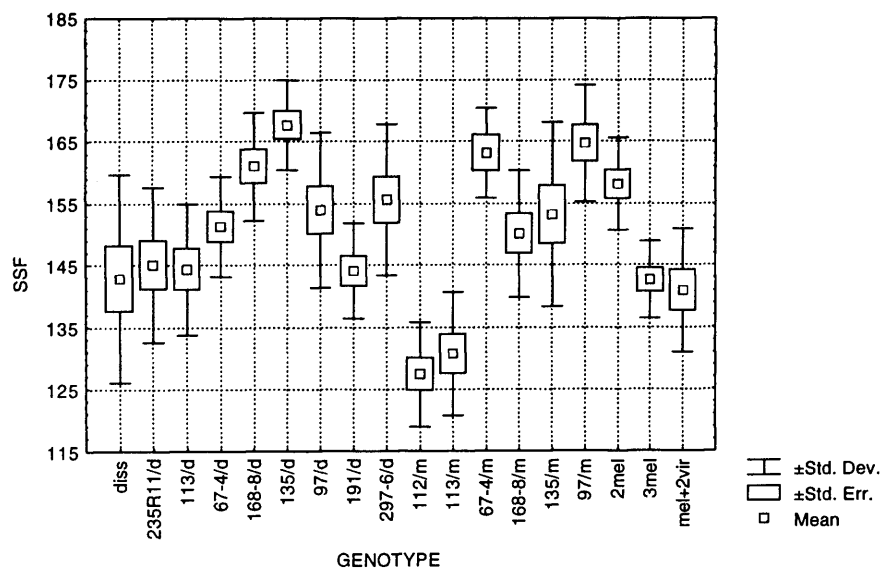
Lines 112/m and 113/m had the lowest SSP values among all the genotypes analyzed. However the four other $nonA^{vir}/nonA^{+}$ lines had much higher values which did not significantly differ from the control (235R11/d). When comparing each hemizygous $nonA^{(-)}[nonA^{vir}]$ line with its heterozygous $nonA^{vir}/nonA^{+}$ counterpart, no coherent pattern can be noticed. Finally, 2mel and 3mel had similar SSP values to those of the control 235R11/d, while diss and mel+2vir had significantly lower SSPs.

In conclusion, there is no argument to support the idea that the *nonA^{vir}* gene causes suppression of sine song. Neither *nonA⁽⁻⁾[nonA^{vir}]* or *nonA^{vir}/nonA⁺* lines show a consistent pattern of SSP values, and adding multiple doses of the *mel* or *vir nonA* gene does not help to clarify the picture.

SSF.

Wild type *D.melanogaster* sine song frequencies usually range between 100 and 200 Hz, with median values varying between 120 and 160 Hz (Wheeler *et al.*, 1989). SSF values were calculated for each song as the average frequency of ten sine song bouts, as described in chapter 2. The average SSF values found in each genotype are shown in figure 7.8.

Figure 7.8. Average SSF values for each genotype.



One-way ANOVA performed on the SSF values of each song analyzed gave a highly significant F ratio ($F=10.882$; $df = 17, 157$; $p < 0.001$). From figure 7.8 it is clear how *diss* seems to be normal in respect to its SSF values when compared to the

control 235R11/d. The Newman-Keuls *a posteriori* test confirmed that no significant difference exists between the SSFs of *diss* and those of 235R11/d (see appendix 7.6).

Hemizygous $nonA^{(-)}[nonA^{vir}]$ lines had SSF values either similar or higher than those of the control 235R11/d. A planned comparison between the group including all the $nonA^{(-)}[nonA^{vir}]$ lines and the control gave a significant F ratio ($F= 6.320$; $df= 1, 157$; $p= 0.013$). In particular Newman-Keuls tests showed that 168-8/d and 135/d display significantly higher SSFs than 235R11/d (see appendix 7.6).

Heterozygous $nonA^{vir}/nonA^{+}$ lines show a complicated pattern since four of them have higher SSF values than the control (67-4/m and 97/m have significantly higher SSFs than 235R11/d, see Newman-Keuls in appendix 7.6) while other two (112/m and 113/m) have significantly lower values (see appendix 7.6). Not surprisingly a planned comparison between line 235R11/d and the $nonA^{vir}/nonA^{+}$ lines was non-significant ($F= 0.745$; $df= 1, 157$; $p= 0.389$).

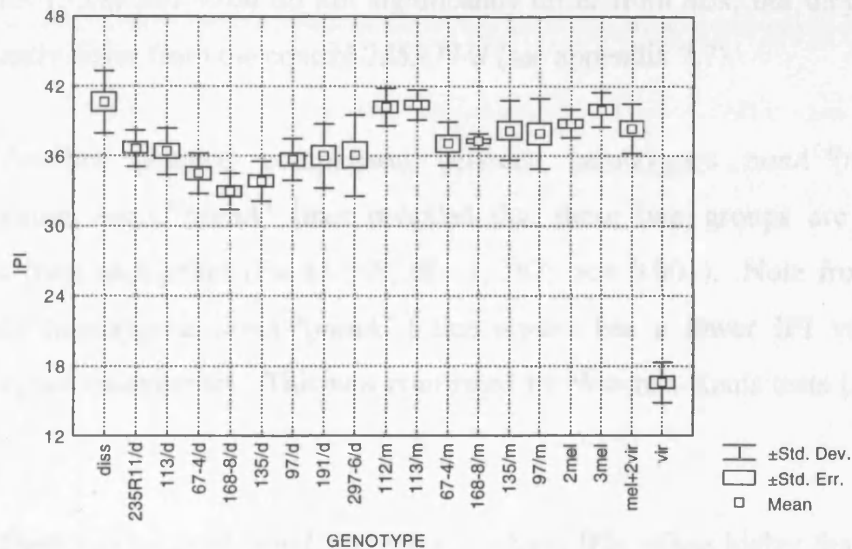
A planned comparison between hemizygous $nonA^{(-)}[nonA^{vir}]$ lines and heterozygous $nonA^{vir}/nonA^{+}$ was significant ($F= 9.316$; $df= 1, 157$; $p= 0.003$). However, comparing each hemizygous $nonA^{(-)}[nonA^{vir}]$ line with its heterozygous $nonA^{vir}/nonA^{+}$ counterpart (for example 113/d with 113/m) does not reveal any coherent pattern: in some cases the hemizygous line has higher value than the heterozygous (this is the case of lines 113, 168-8 and 135), in others the situation is reversed (compare 67-4/d with 67-4/m and 97/d with 97/m in figure 7.8). Finally, line 2mel has higher SSFs than the control, even though the difference is not significant, while 3mel and mel+2vir have values similar to 235R11/d.

In summary, it seems that the $nonA^{vir}$ gene plays no role in affecting the sine song frequency. This is not altogether surprising considered that it has been shown above that the $nonA^{vir}$ transgene does not affect the frequencies of single pulses either. *D.virilis* flies produce very high frequency sounds but this characteristic is obviously not determined by the *nonA* gene.

IPI.

The mean IPI values were calculated for each song as described in chapter 2. One way ANOVA performed on these data gave a highly significant F ratio ($F=64.492$; $df=18, 167$; $p<<0.001$). The result of the ANOVA was still significant even after excluding the data from *vir* and *diss* ($F=11.295$; $df=16, 149$; $p<<0.001$). Figure 7.9 graphically illustrates the mean, standard error and standard deviation of the IPI values of each genotype.

Figure 7.9. Mean IPI values of each genotype.



From figure 7.9 it is clear that *diss* displays the highest IPI values of all the genotypes analyzed. These values differ significantly from the IPI values shown by the control line *235R11/d*, as confirmed by the *a posteriori* test (see appendix 7.7). As expected from previous analysis of *virilis* songs (Hoikkala *et al.*, 1982), the mean IPIs of this species are very short.

The mean IPI values of hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] lines are either similar or lower than those of the control *235R11/d*. A planned comparison between line *235R11/d* and all the hemizygous lines *nonA*⁽⁻⁾[*nonA*^{vir}] (from line *113/d* to line *297-*

6/d, see figure 7.9) showed that these two groups are indeed significantly different ($F=5.298$; $df=1, 167$; $p=0.023$). In particular, a Newman-Keuls test revealed that the IPI values of lines 168-8/d and 135/d are significantly lower than those of the control 235R11/d (see appendix 7.7).

From figure 7.9 the mean IPIs of heterozygous $nonA^{vir}/nonA^+$ lines appear to be either similar or higher than those of line 235R11/d. A planned comparison showed that this group of genotypes significantly differs from the control line 235R11/d ($F=6.763$, $df=1, 167$; $p=0.010$). A Newman-Keuls test pointed out that the mean IPI values of lines 113/m and 112/m are significantly higher than those of the control 235R11/d. These two genotypes have IPI values that do not differ from those of *diss*. Also lines 135/m and 97/m do not significantly differ from *diss*, but they also do not significantly differ from the control 235R11/d (see appendix 7.7).

Another planned comparison between hemizygous $nonA^{(-)}[nonA^{vir}]$ and heterozygous $nonA^{vir}/nonA^+$ lines revealed that these two groups are significantly different from each other ($F=83.368$; $df=1, 167$; $p<<0.001$). Note from figure 7.9 that each hemizygous $nonA^{(-)}[nonA^{vir}]$ line always has a lower IPI value than its heterozygous counterpart. This was confirmed by Newman-Keuls tests (see appendix 7.7).

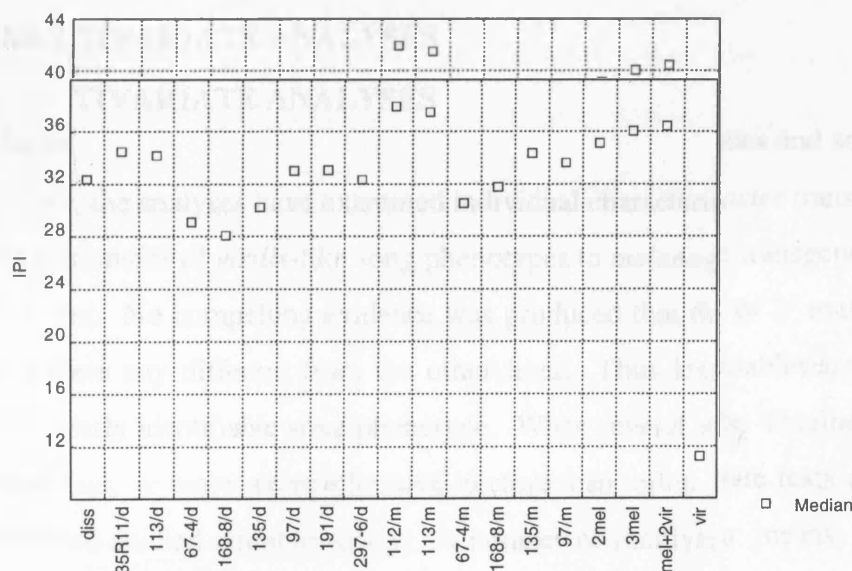
Finally, lines 2*mel*, 3*mel*, and *mel+2vir* have IPIs values higher than the control 235R11/d. All three lines do not significantly differ from *diss*, however, only line 3*mel* has significantly higher IPI values than the control line (see Newman-Keuls in appendix 7.7).

Considering that *vir* songs have very short mean IPIs, the fact the songs of flies carrying the $nonA^{vir}$ gene on a $nonA^{(-)}$ background tend to have lower IPI values than those of the control line is interesting and unexpected. However this effect on the IPIs is quite marginal, since, as seen above, only two $nonA^{(-)}[nonA^{vir}]$ lines differ significantly from the control. Dosage effects could explain why some $nonA^{vir}/nonA^+$ heterozygous lines, and lines with more than one copy of $nonA^+$ (for example 3*mel*) have higher IPIs than the control. In fact adding copies of $nonA^+$ (either *virilis* or *mel*)

increases IPIs, yet heterozygotes ($nonA^{vir}/nonA^{+}$, or $mel+2vir$) have in general lower IPIs than their hemizygous 2-dose or 3-dose *mel* counterparts.

The fact that *diss* has such long IPIs could however be the result of an artifact due to the high CPP and low frequencies of the pulses in this line. IPIs were here measured as the average distance between the maximum peak of the intrapulse oscillation of one pulse to the maximum peak of the next pulse. Thus low frequency, polycyclic pulses will have longer IPIs because the distance from the maximum peak of one such pulses to the peak of the next is actually greater than the same distance measured between high frequency, monocyclic pulses. In order to overcome this possible problem, I decided to correct each line's IPI values by taking into account their respective CPP (at $x=0$), bCPP, PF (at $x=0$) and bPF values. An analysis of covariance was performed in which CPP, bCPP, PF and bPF were the covariates of the IPI values. The resulting F ratio was highly significant ($F= 22.083$; $df= 18, 163$; $p < 0.001$). Figure 7.10 shows the adjusted IPI values for each genotype.

Figure 7.10. Adjusted IPI values.



Note that after correction the IPI values of *diss* become very similar to those of the control line 235R11/d. Hemizygous $nonA^{(-)}[nonA^{vir}]$ lines still have similar or

lower values than those of the control. Newman-Keuls *a posteriori* comparisons showed that the values of 67-4/d, 168-8/d, and 135/d are significantly lower than those of line 235R11/d. After correction the IPIs of lines 112/m and 113/m are still significantly higher than those of the control, while those of lines 67-4/m and 168-8/m are actually significantly lower. The adjusted IPI values of 2mel, 3mel and mel+2vir are not significantly different from those of the control line 235R11/d.

In summary, the most interesting result obtained after correcting the IPIs values for their respective CPP, bCPP, PF and bPF values, is that *diss* becomes indistinguishable from the control line. On the contrary, hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] lines retain generally lower values than line 235R11/d. After correction, the situation of the heterozygous *nonA*^{vir}/*nonA*⁺ lines changes slightly, since two of these lines now show significantly lower values than the control (before correction all *nonA*^{vir}/*nonA*⁺ lines had similar or higher values than 235R11/d, see figure 7.9). Also the results obtained with 2mel, 3mel and mel+2vir, were changed after correction, since these lines no longer differ from the control strain.

MULTIVARIATE ANALYSES

So far, the analyses have examined individual characteristics and some evidence for a modest transfer of *virilis*-like song phenotypes to *melanogaster* transformants has been presented. No compelling evidence was produced that the transgenes 113, 67-4, and 168-8 were any different from the other lines. Thus loss of 3' material did not generate a clearly identifiable song phenotype. When several variables are investigated as in this case, a more comprehensive picture can often be obtained if all the phenotypes are studied simultaneously. A number of multivariate tests are available, and discriminant analysis (Anderson, 1963) was chosen as a means of detecting differences between the experimental groups. SYSTAT 7.0 software was used and I am grateful to Yuri Dubrova for running the analysis on his computer and explaining the method and results to me.

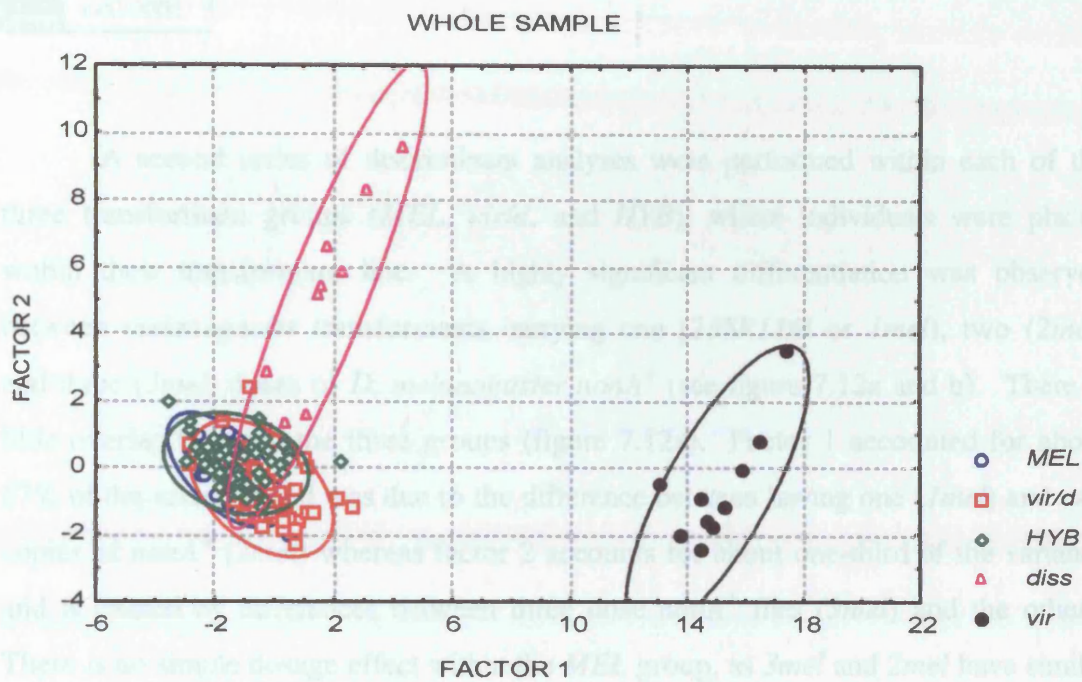
All the six song variables, CPP, bCPP, PF, bPF, SSF, and IPI, were used. SSF of course could not be used in tests which included *D. virilis* songs. One *diss* fly, which showed a negative intercept on its CPP and a correspondingly high bCPP was removed to generate a normal distribution for these characters. In practice, this makes little difference to the analysis, as *diss* mutants form a clearly separate group (see below).

Figure 7.11a and b show the results of the discriminant analysis on the whole sample. This first analysis placed all individuals into 5 groups: *vir* (*D. virilis*), *diss*, *vir/d* (all the *nonA*⁽⁻⁾[*nonA*^{vir}] transformants), the group named *MEL* which includes the control line 235R11/d (or 1*mel*), 2*mel* and 3*mel*, and the "hybrids" group (*HYB*) composed by *vir/m* (all the *nonA*^{vir}/*nonA*⁺ transformants) and *mel*+2*vir*. In figure 7.11a individual scores are plotted plus the ellipses defining the 95% confidence limits, thereby giving an idea of the overlap between the 5 groups, and in figure 7.11b the mean for each group is shown for clarity. The *vir* species and *diss* mutants form two outlying, nonoverlapping groups. Table 7.4 presents the summary of the analysis and clearly indicates that each behavioural characteristic contributes equally highly to the differentiation of these 5 groups. An extremely low probability of misclassification of groups (the F-statistic estimates the probability of Wilk's lambda value occurring by chance) is observed. This analysis reveals that about 87% of the total variance of all traits corresponds to the differences between *vir* and the rest of the groups (factor 1), whereas approximately 9% of the total variance appears due to the differences between *diss* and the others. Within this initial analysis, the transformant lines form one, largely overlapping group.

Table 7.4. Results of the discriminant analysis for the whole sample.

ALL GROUPS		
Traits	F (4, 180)	P
IPI	217.04	<<0.0001
PF	27.15	<<0.0001
CPP	68.41	<<0.0001
bCCP	122.84	<<0.0001
bPF	20.74	<<0.0001
SSF	--	--

A:



B:

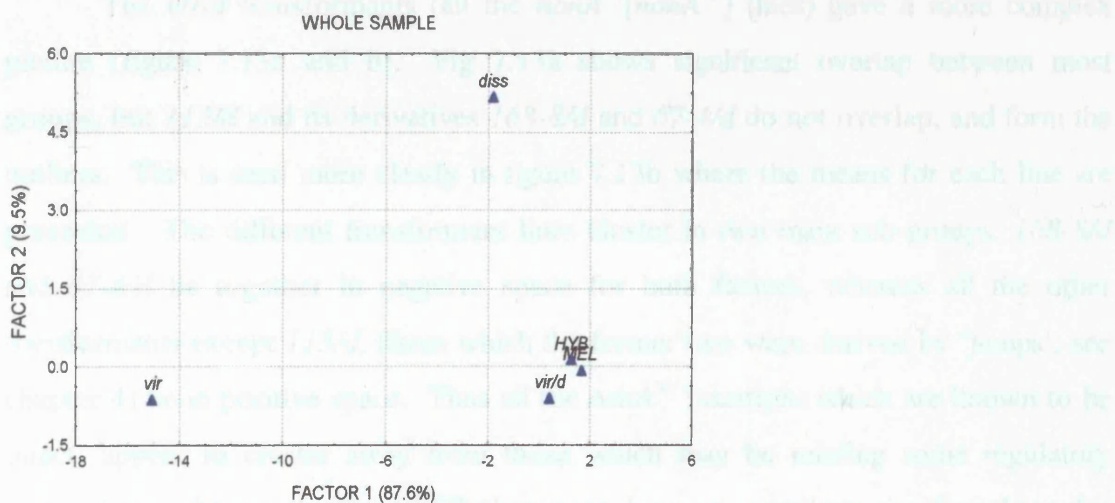


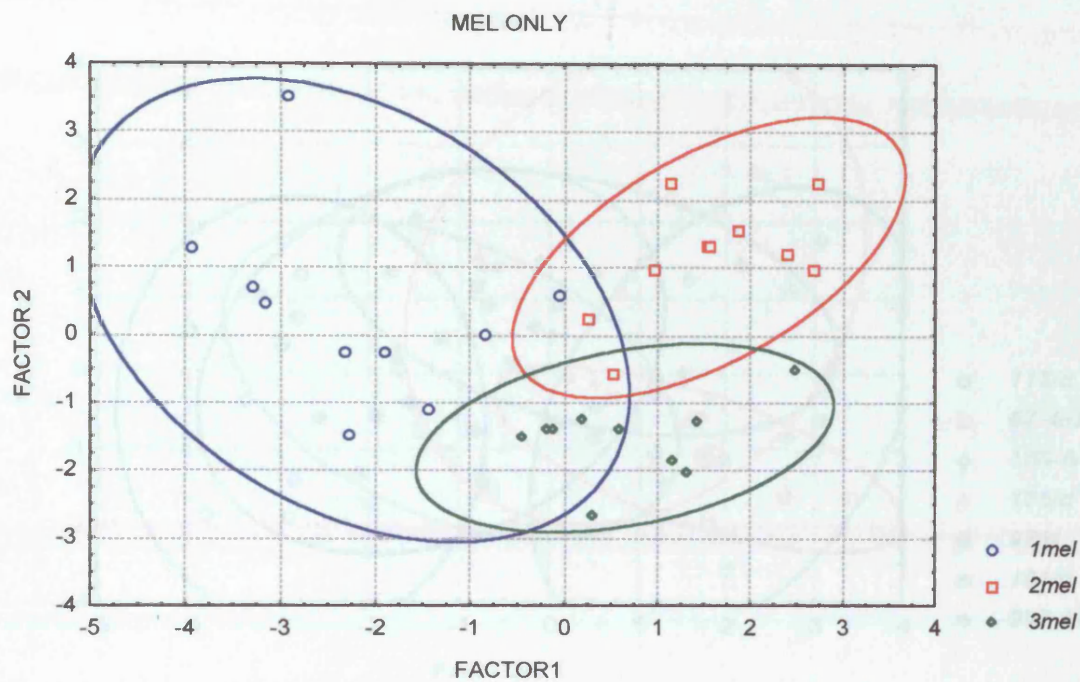
Figure 7.11. Discriminant analysis of all genotypes grouped as follows: MEL= 235R11/d, 2mel, 3mel. vir/d= 113/d, 67-4/d, 168-8/d, 135/d, 97/d, 191/d, 297-6/d. HYB= 112/m, 113/m, 67-4/m, 168-8/m, 135/m, 97/m, mel+2vir. diss= nonA^{diss}. vir= *D.virilis*. **A:** individual scores plus the ellipses defining the 95% confidence limits. **B:** mean score for each group. Factor 1 accounts for 87.6% of the total variability, while factor 2 accounts for 9.5%.

λ	0.0182
F	68.6
d.f.	20, 584
P	<<0.0001

A second series of discriminant analyses were performed within each of the three transformant groups (*MEL*, *vir/d*, and *HYB*), where individuals were placed within their transformant line. A highly significant differentiation was observed between *melanogaster* transformants carrying one (*235R11/d* or *1mel*), two (*2mel*) and three (*3mel*) doses of *D. melanogaster nonA*⁺ (see figure 7.12a and b). There is little overlap between the three groups (figure 7.12a). Factor 1 accounted for about 67% of the variance and was due to the difference between having one (*1mel*) and two copies of *nonA*⁺ (*2mel*) whereas factor 2 accounts for about one-third of the variance and is caused by differences between three dose *nonA*⁺ flies (*3mel*) and the others. There is no simple dosage effect within the *MEL* group, as *3mel* and *2mel* have similar scores on factor 1, as do *1mel* (*235R11/d*) and *2mel* on factor 2. Appendix 7.8 shows the summary statistics for this analysis, with bPF failing to contribute significantly to the discrimination.

The *vir/d* transformants (all the *nonA*⁽⁻⁾[*nonA*^{vir}] lines) gave a more complex picture (figure 7.13a and b). Fig 7.13a shows significant overlap between most groups, but *113/d* and its derivatives *168-8/d* and *67-4/d* do not overlap, and form the outliers. This is seen more clearly in figure 7.13b where the means for each line are presented. The different transformant lines cluster in two main sub-groups. *168-8/d* and *67-4/d* lie together in negative space for both factors, whereas all the other transformants except *113/d*, (from which the former two were derived by ‘jumps’, see chapter 4) lie in positive space. Thus all the *nonA*^{vir} insertions which are known to be intact, appear to cluster away from those which may be missing some regulatory material (see chapter 4). The bCCP phenotype does not contribute significantly to this discrimination (see Appendix 7.9). This pattern is repeated with the *vir/m* hybrids, again the *67-4/m* and *168-8/m* transformants clustering away from the other groups (see figure 7.14a and b and appendix 7.10).

A:



B:

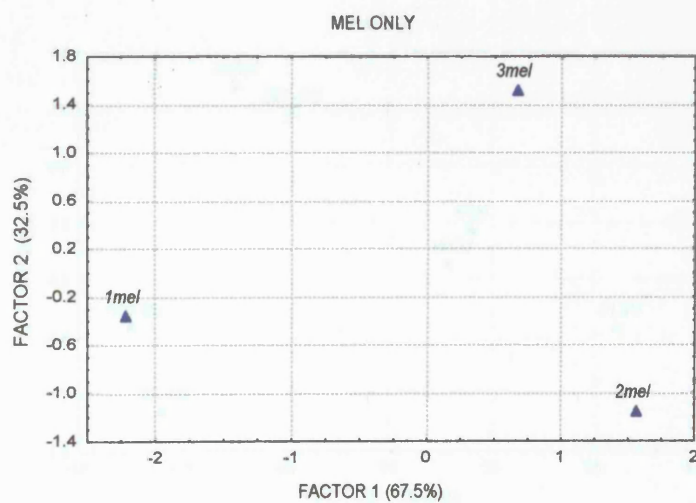
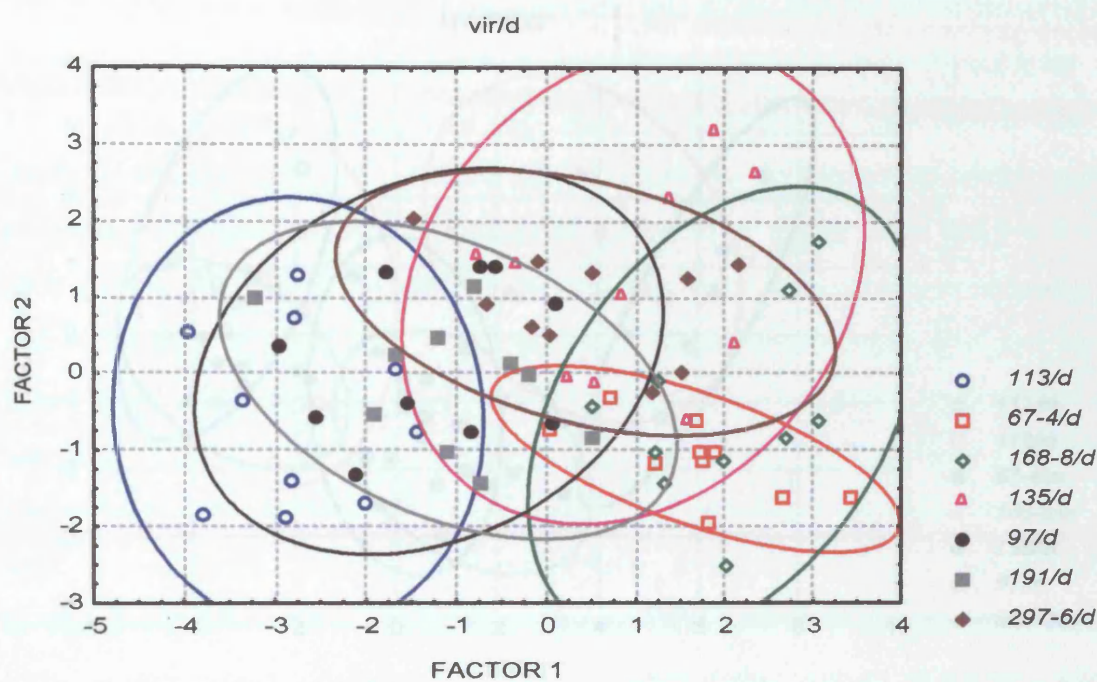


Figure 7.12. Discriminant analysis of the *MEL* group (*1mel*, *2mel*, *3mel*). **A:** the individual scores for each genotype are indicated plus the ellipses defining the 95% confidence limits. **B:** mean score for each genotype. Factor 1 accounts for 67.5% of the total variability, while factor 2 accounts for 32.5%.

A:



B:

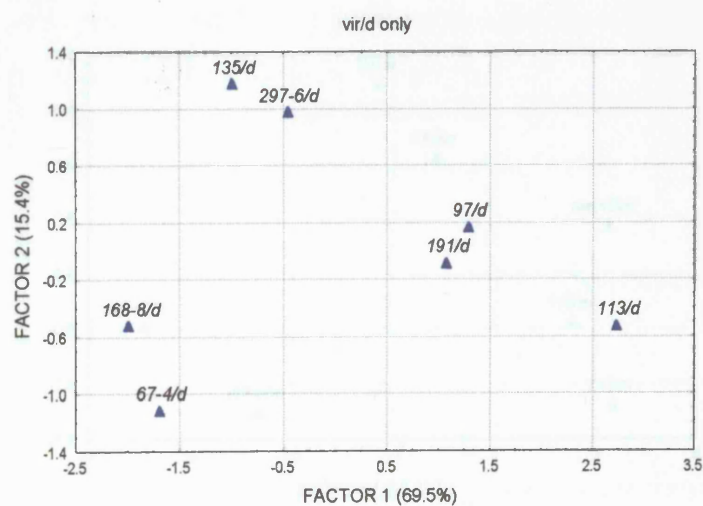
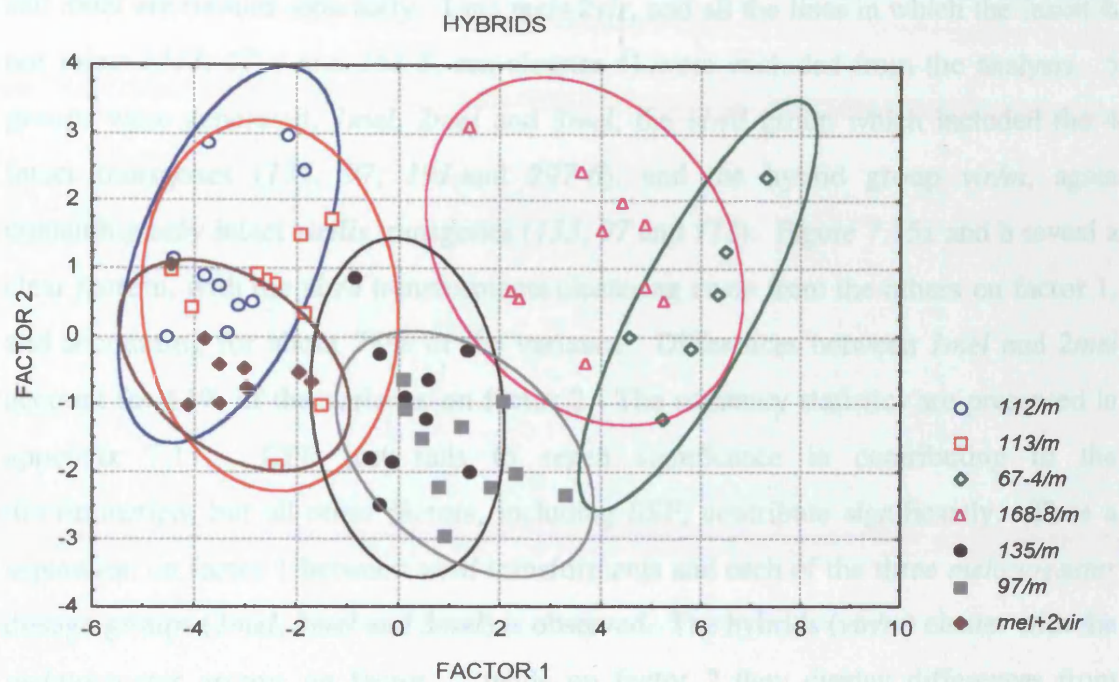


Figure 7.13. Discriminant analysis of the *vir/d* group (all the $nonA^{(-)}$ [$nonA^{vir}$] transformants). **A:** scores of the individuals of each genotype plus the ellipses defining the 95% confidence limits. **B:** mean for each transformant line. Factor 1 accounts for 69.5% of the total variability, while factor 2 accounts for 15.4%.

A:



B:

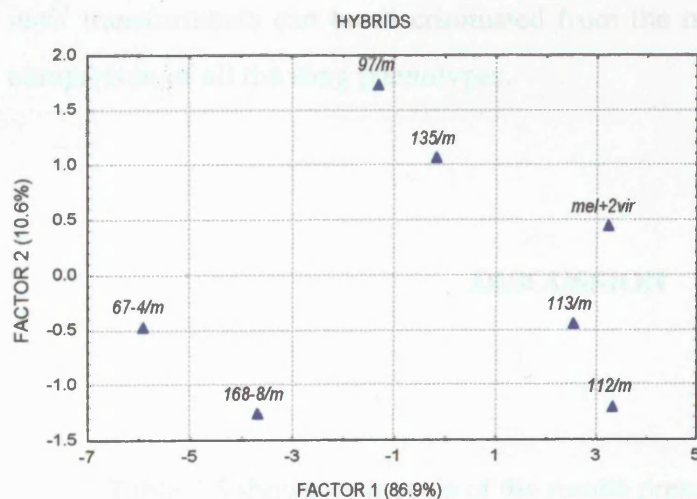


Figure 7.14 Discriminant analysis of the hybrids (HYB) group (including all the $nonA^{vir}/nonA^{+}$ transformants and $mel+2vir$). **A:** scores of the individuals of each genotype plus the ellipses defining the 95% confidence limits. **B:** mean score for each line. Factor 1 accounts for 86.9% of the total variability, while factor 2 accounts for 10.6%.

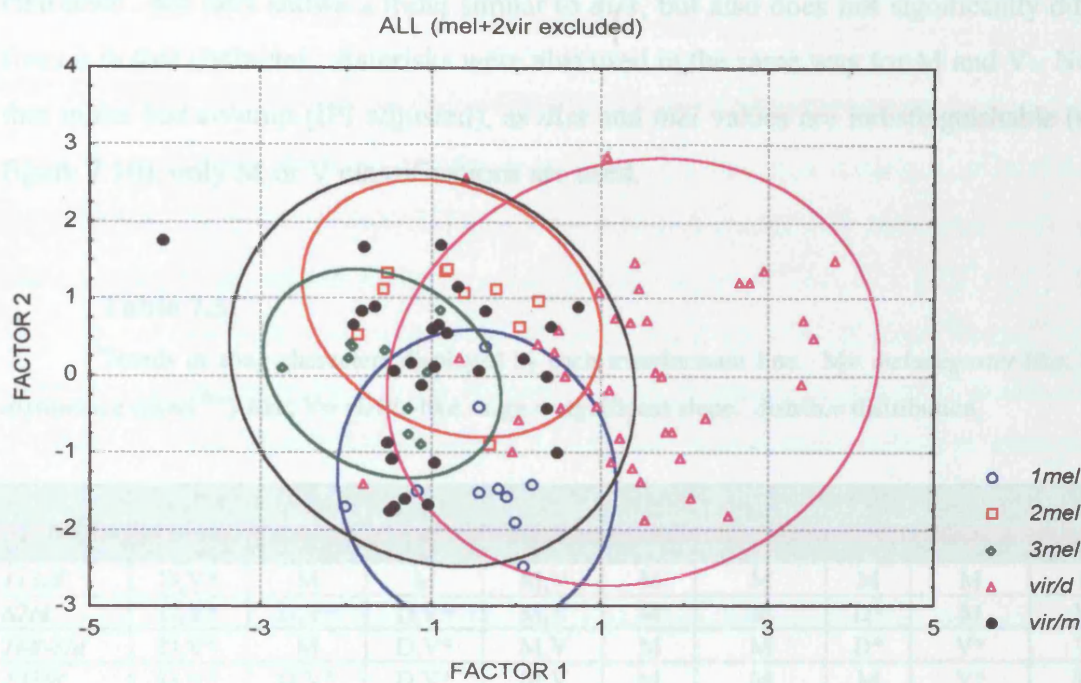
The final discriminant analysis puts all the individual lines into their respective groups, and maintains the different groups in terms of gene dosage. Thus *1mel*, *2mel* and *3mel* are treated separately. Line *mel+2vir*, and all the lines in which the insert is not intact (*113*, *67-4* and *168-8*, see chapter 4) were excluded from the analysis. 5 groups were generated, *1mel*, *2mel* and *3mel*, the *vir/d* group which included the 4 intact transgenes (*135*, *97*, *191*, and *297-6*), and the hybrid group *vir/m*, again containing only intact *virilis* transgenes (*135*, *97* and *112*). Figure 7.15a and b reveal a clear pattern, with the *vir/d* transformants clustering away from the others on factor 1, and accounting for about 78% of the variance. Differences between *1mel* and *2mel* account for 14% of the variance on factor 2. The summary statistics are presented in appendix 7.11. CPP just fails to reach significance in contributing to the discrimination, but all other factors, including SSF, contribute significantly. Thus a separation on factor 1 between *vir/d* transformants and each of the three *melanogaster* dosage groups (*1mel*, *2mel* and *3mel*) is observed. The hybrids (*vir/m*) cluster with the *melanogaster* groups on factor 1, while on factor 2 they display differences from groups *1mel* and *2mel*, but not from *3mel*, or *vir/d*, thereby revealing a recessive phenotype with respect to *nonA^{vir}* on factor 1, but a dominant one for factor 2.

In conclusion, the multivariate analyses have shown reasonably clearly that the *vir/d* transformants can be discriminated from the other groups by the simultaneous comparison of all the song phenotypes.

DISCUSSION

Table 7.5 shows a synopsis of the results presented. For each transformant line the trend shown by each song character analyzed (only informative characters are included in the table) has been classified as *dissonance*-like (D), *virilis*-like (V), or *melanogaster*-like (M). The classification was based on the very conservative *a posteriori* tests described in the results section. Any character which showed a trend

A:



B:

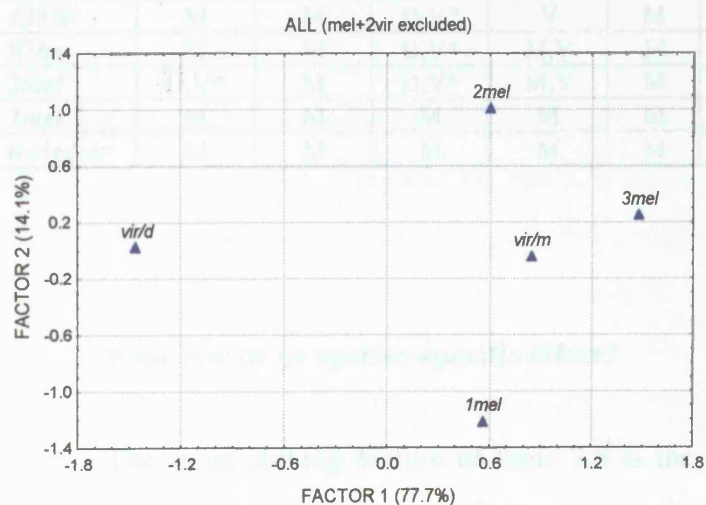


Figure 7.15. Discriminant analysis of all the genotypes, excluding the transformant lines in which the insert is not intact (*113/d*, *67-4/d*, *168-8/d*, *113/m*, *67-4/m*, *168-8/m* and *mel+2vir*, see text and chapter 4). **A:** scores of the individuals belonging to each group plus the ellipses defining the 95% confidence limit. **B:** mean of each group. Factor 1 accounts for 77.7% of the total variability, while factor 2 accounts for 14.1%.

similar, but still significantly different from, let us say, *diss*, was marked as D followed by an asterisk (D*). Therefore, a line which is marked as D (and no asterisk) for one character, not only shows a trend similar to *diss*, but also does not significantly differ from it in that character. Asterisks were also used in the same way for M and V. Note that in the last column (IPI adjusted), as *diss* and *mel* values are indistinguishable (see figure 7.10), only M or V classifications are used.

Table 7.5.

Trends in song characters displayed by each transformant line. M= *melanogaster*-like; D= *dissonance* (*nonA^{diss}*)-like; V= *virilis*-like. sign.= significant slope. distrib.= distribution.

	bCPP sign.	bCPP	CPP10	CPP distrib.	bPF sign.	bPF	PF1	IPI	IPI adjusted
113/d	D,V*	M	M	M,V	M	M	M	M	M
67/d	D,V*	D,V*	D,V*	M,V	M	M	D*	M	V*
168-8/d	D,V*	M	D,V*	M,V	M	M	D*	V*	V*
135/d	D,V*	D,V*	D,V*	M,V	M	M	M	V*	V*
97/d	D,V*	M	M	M,V	M	M	M	M	M
191/d	D,V*	D,V*	D,V*	M,V	M	M	M	M	M
297-6/d	D,V*	M	D,V*	V	M	M	M	M	M
112/m	M	M	M	M	M	M,V	M	D	M*
113/m	M	M	M	M	M	M	M	D	M*
67-4/m	M	M	D,V*	M,V	M	M,V,D	D*	M	V*
168-8/m	M	M	D,V*	M,V	M	M,V	D*	M	V*
135/m	M	M	D,V*	V	M	M	D*	M,D	M
97/m	M	M	D,V*	M,V	M	M,V	D*	M,D	M
2mel	D,V*	M	D,V*	M,V	M	M	M	M,D	M
3mel	M	M	M	M	M	M	M	D	M
mel+2vir	M	M	M	M	M	M	M	M,D	M

Poor rescue or species-specific effect?

The most striking feature of table 7.5 is the observation that in most of the cases it was impossible to distinguish between a *diss*-like phenotype (D) or a *vir*-like (V) one. In fact the unexpected positive regression found for CPP in *D.virilis* complicated interpretation of the results, since the mutant, *nonA^{diss}* carries a similar

feature. Thus a positive gradient could be interpreted as either *virilis*-like or mutant *diss* behaviour.

It was clear that all the hemizygous $nonA^{(-)}[nonA^{vir}]$ transformants had significantly more positive regressions than the control *235R11/d* and the heterozygous $nonA^{vir}/nonA^{+}$ lines (table 7.2 and figure 7.2b). If this was simply a reflection of poorer rescue in $nonA^{(-)}[nonA^{vir}]$ lines, then we should expect CPP measures to show a similar pattern, i.e. higher CPP in $nonA^{(-)}[nonA^{vir}] > nonA^{vir}/nonA^{+} > 235R11/d$. Yet this was not the case: CPP values (at both $x=1$ and $x=10$) for $nonA^{vir}/nonA^{+}$ lines were constantly higher than those of $nonA^{(-)}[nonA^{vir}]$, which in turn were higher than *235R11/d*. Since the CPP gradients of lines $nonA^{vir}/nonA^{+}$ were definitely wild-type-like, it is difficult to explain this result as a mutant-like phenotype (possibly caused by a dosage effect). Perhaps this result reflects species-specific information conveyed by the $nonA^{vir}$ gene? However, if this was true, then why does line *mel+2vir* have completely normal bCPP and CPP values?

The CPP distributions were informative when the fourth moment (kurtosis) was examined. As a group, the $nonA^{(-)}[nonA^{vir}]$ transformants had distributions more similar to those of *virilis*, and differed significantly from the *mel* control (*235R11/d*) and from the mutant *diss*. The group composed by $nonA^{vir}/nonA^{+}$ lines instead did not differ from the control, but significantly differed from either *diss* or *vir*. These results were also partially supported when individual comparisons were made between lines. Most importantly, each transformant line differed significantly from *diss*, suggesting that robust rescue was achieved for this song character. Furthermore, most of the individual transformant lines carrying the $nonA^{vir}$ gene had distributions falling in between *mel* and *vir* (see table 7.5). Consequently, there is again some evidence for species-specific effects being transferred, and the results are difficult to place within an "incomplete rescue" scenario.

When the PF regressions were examined, all the $nonA^{(-)}[nonA^{vir}]$ transformants behaved in a manner indistinguishable from the control line, showing good rescue for this character. When the bPF values were analyzed, some heterozygous $nonA^{vir}/nonA^{+}$ lines gave values that fell between *mel* and *vir*, and in one case (*67-4/m*) it was not

possible to distinguish from a *diss*-like phenotype. Therefore, again, this could reflect some evidence for a transfer of species-specific information, but, the fact that *diss* behaved similarly to *vir* also in this song character, again created problems (at least for line 67-4/m) in distinguishing between a mutant or a *virilis*-like phenotype.

When PF was analyzed, after correction for song output (figure 7.6), both the $nonA^{(-)}[nonA^{vir}]$ and $nonA^{vir}/nonA^{+}$ group lines gave values generally lower than the control 235R11/d and *vir*. In particular four out of the six heterozygous $nonA^{vir}/nonA^{+}$ lines analyzed, had PFs rather similar to (even though still statistically different from) *diss*. If this result was to be considered as evidence of poor rescue or mutant-like phenotype, then it is again difficult to understand why heterozygous $nonA^{vir}/nonA^{+}$ lines that show perfectly normal CPP gradients have actually lower PF values than the hemizygous $nonA^{(-)}[nonA^{vir}]$ lines whose CPP gradient might be considered as reflecting incomplete rescue. An alternative hypothesis to the results obtained with PF is that perhaps, in *D.melanogaster*, simply having more CPP predispose to lower PFs. This would explain why $nonA^{vir}/nonA^{+}$ lines, which have in general more CPP than $nonA^{(-)}[nonA^{vir}]$ lines, also have lower PF.

SSP and SSF were not informative. My idea that $nonA^{vir}$ might act as a factor repressing sine song in *D.virilis* was not supported. The only gene known so far to suppress SS in *D.melanogaster* is *doublesex* (*dsx*) (Villella and Hall, 1996).

Finally, the IPI results were quite interesting in that the hemizygous $nonA^{(-)}[nonA^{vir}]$ flies had IPIs in the *D.virilis* direction, and generally lower than those of heterozygous $nonA^{vir}/nonA^{+}$ lines. Some lines of the heterozygous $nonA^{vir}/nonA^{+}$ group had IPIs falling in between the control and *diss*, and two lines had IPI values actually more similar to *diss* than to the control. However, the higher *diss* IPI may be an artifact due to the increased CPP and lower PF of these mutants. Because of the way IPIs were calculated (i.e. as the distance from the highest peak of the intrapulse oscillation of one pulse to the highest peak of the next pulse), it may be inevitable that pulses with higher CPPs and low PFs show longer IPIs. In fact, after performing a covariance analysis in which the IPI values were corrected by taking into account the

corresponding CPP, bCPP, PF and bPF values, the IPIs of *diss* were found to be very similar to those of the control *235R11/d*.

However, even with *diss* it is difficult to know at what level IPI control occurs: in other words, does the actual space between two pulses stay the same even between polycyclic pulses, so that the IPIs measured as the peak to peak distance between such pulses is really longer than for monocyclic pulses? Or is it that when pulses are polycyclic the actual space between them is reduced so that the distance between the maximum peak of one pulse and the maximum peak of the next pulse remains unaltered if compared to monocyclic pulses? If the first case is true, then making a compensation for CPP and PF in the measure of IPIs is the correct thing to do. If the second case is correct, then the covariance analysis may not be meaningful.

To summarize, because of the positive bCPP values of hemizygous *nonA*⁽⁻⁾ [*nonA*^{vir}] lines, it proves very difficult to conclude whether they are completely rescued in their song phenotype, or whether they actually show any *virilis* song features. It must be noted however that in the PF characters they behave decisively like the control, and in a few characters resemble the *vir* phenotype (see their CPP distributions, and IPIs).

The existence of some species-specific information being conveyed by the *nonA*^{vir} gene can however be reasonably supported when taking into consideration the results obtained with the heterozygous *nonA*^{vir}/*nonA*⁺ lines. The fact that these lines have unusually high CPP in their songs, cannot be simply explained as a mutant phenotype caused by the males carrying one copy of *nonA*^{vir} and one endogenous copy of *nonA*⁺. In most of the song characters analyzed, these lines are indistinguishable from the control, and even for those characters which present some inconsistencies (i.e. PF1 and IPI), alternative explanations, other than mutant effects, can be argued.

In addition, the multivariate analyses also suggest that, in the *vir/d* transformants, species-specific transfer of some *virilis*-like information may have passed into the *melanogaster* hosts. The multivariate analyses clearly show a differentiation between the *vir/d* males and the others transformants. In these analyses

all the song characters contribute significantly to the differentiation. This includes measures of CPP (CPP or bCPP) PF (PF or bPF) IPI and SSF. Again, the differences seen between the *diss* mutants and *D.virilis* are so large (Fig 7.11a and b) that it is not possible to say whether the subtle differences observed between transformant groups reflect mutant or *virilis*-like behaviour. However, if we imagine that the *nonA^{vir}* transgene is robust enough to rescue visual defects, but that its "dosage" is not enough to ensure proper song rescue, (perhaps via regulatory region divergence), then we might expect the partial hybrids *vir/m*, to have a dosage that reflects more than one copy of the *melanogaster* gene but less than two copies. Thus the *vir/m* hybrids should fall in multivariate space between one and two dose *melanogaster* flies (*1mel* and *2mel*). Yet, on the main factor of the discriminant analysis the hybrids lie at the same point as *1mel* and *2mel* (figure 7.15b). On factor 2, which accounts for only 14% of the variance, they lie between *1mel* and *2mel* but are also at the same level as *3mel*. It is therefore difficult to argue that the *vir/d* phenotype is mutant due to less than optimal dosage, particularly as no clear picture on dosage emerges even from comparing *1mel*, *2mel*, and *3mel* transformants (see figure 7.12b).

Of additional interest was that the discriminant analysis also differentiated between the lines containing insert *113* (*113/d* and *113/m*), and those containing its two 'jumped' derivatives *67-4* and *168-8* (*67-4/d* and *67-4/m*, *168-8/d* and *168-8/m*, see figures 7.13a,b and 7.14a,b). While the *113* insert may be missing some of the flanking DNA for *nonA^{vir}*, the multivariate analysis argues that perhaps further regulatory material has 'gone' in inserts *67-4* and *168-8*, as the transformants carrying them are clearly outliers.

Line *mel+2vir* is something of an enigma, since it is consistently indistinguishable from the control line *235R11/d* in all the individual song characters analyzed. When the multivariate analysis is performed, this line clusters with other heterozygous lines carrying only one dose of *nonA^{vir}* and *mel nonA⁺* (figure 7.14b), as if the extra copy of the *nonA^{vir}* gene made no difference.

Dosage effect?

As discussed above, having one dose of *nonA*^{vir} and one of *mel nonA*⁺ does not seem to cause a song phenotype that could be reasonably classified as mutant. However, the results obtained from flies carrying two copies of *nonA*⁺ (*2mel*) are somehow puzzling. Even though their songs' bCPP values are not statistically different from those of the control *235R11/d*, most of the *2mel* songs analyzed had significantly positive CPP regressions. Moreover *2mel* had significantly higher CPP values than the control, and the CPP distribution in this line was intermediate between that of the control line and *D.virilis*. *2mel* IPI values were also quite high (between those of *235R11/d* and those of *diss*), however after correction with covariance they actually become very similar to those of the control. These results could only be explained by assuming that *2mel* flies are slightly mutant in their singing behaviour. Maybe having two doses of the *nonA* gene in males somehow affects its proper functioning. However if this is true, then why is it that flies carrying one copy of *nonA*^{vir} and one of *mel nonA*⁺, 3 copies of *nonA*⁺ (*3mel*), or one copy of *mel nonA*⁺ and two of *nonA*^{vir} do not seem to be similarly affected? As already mentioned, the multivariate analyses also showed that it is not possible to classify the results obtained with *1mel*, *2mel*, *3mel*, and the hybrids with a simple dosage effect explanation.

To conclude, this exhaustive analysis of the song data does bring a final message: clearly, no striking species-specific effect is carried over with the *nonA*^{vir} gene, as was the case with the *per* gene of *D.simulans* (Wheeler *et al.*, 1991) and *D.pseudoobscura* (Petersen *et al.*, 1988). However it may be that *nonA* does convey some relevant species information, but only a small proportion of the interspecific variation is transferred.

CHAPTER 8

DISCUSSION

The aim of this thesis was to investigate whether the *nonA* gene carries species-specific information concerning the male's lovesong. The success of similar studies using *D.simulans* and *D.pseudoobscura per* transgenes to examine lovesong and circadian locomotor patterns was a major motivation for my experiments (Petersen *et al.*, 1988; Wheeler *et al.*, 1991). The courtship song of *Drosophila* is undoubtedly important for recognition of conspecific mates (Kyriacou and Hall, 1980; 1982; Cowling and Burnet, 1981), and therefore is a character which could be directly involved in speciation (Coyne, 1992). In addition, locomotor activity patterns could also be viewed as an important factor in sexual isolation: if one species is active only in the morning while the other is active only at night, individuals from the two different species are temporally isolated and may actually never come into contact with each other even if they cohabit the same niche. This would ensure that no interspecific matings occur, thereby reducing the risk of producing wasteful inviable or sterile hybrid offspring. By controlling directly species-specific characteristics of courtship song and locomotor activity in flies, the *per* gene can contribute directly to the speciation process, and has therefore been named a "speciation gene" by Coyne (1992).

Superficially, *nonA* resembles the *per* gene, in that it also controls characteristic elements of the lovesong. Consequently, could *nonA* be in any way responsible for the differences seen in the lovesong of *D.melanogaster* and that of another species? To address this question it was essential to find a species that differed from *D.melanogaster* in song characters that might be controlled by *nonA*. The best candidate was *D.virilis*, whose song is very different from that of *D.melanogaster*, and whose polycyclic pulses actually resembled those found in the mutant *nonA^{diss}* (see figure 7.1). Other *Drosophila* species songs that have been well characterized seemed too similar to the *D.melanogaster* song to be considered useful for this study (Cowling and Burnet, 1981). Furthermore, in the case of *D.melanogaster* and *D.simulans* it had been shown that characters such as the number of cycles per pulse or the interpulse interval are actually controlled by autosomal genes (Wheeler *et al.*, 1988). Studies conducted on the species-specific differences between *D.virilis* and other species from the *montana* phylad instead suggested that lovesong characters such as IPI, CPP and

pulse train length might be controlled by X-linked genes (Hoikkala and Lumme, 1987). In the case of the differences between the song of *D.virilis* and *D.littoralis*, the involvement of sex-linked genes was actually demonstrated (Hoikkala and Lumme, 1987). Could *nonA* (which in *D.virilis* maps on the X as in *D.melanogaster*, Hoikkala, personal communication) be one of these genes? The fact that apparently no gene on the X chromosome controls lovesong characters such as CPP or IPIs between the *D.melanogaster* and *D.simulans* species is not contradictory, as it does not mean that *nonA* cannot be controlling these song differences between *D.virilis* and *D.melanogaster*. It is possible that in the lineage leading to *D.virilis*, changes which determine lovesong traits accumulated in the *nonA* gene. In the 1-2 million years since the divergence of *D.melanogaster* from *D.simulans*, clearly no X-chromosome gene (including *nonA*) which could affect these song characters, have had time to diverge. In fact, initial work I performed on the *D.simulans nonA* gene showed that its sequence is almost completely identical to that of *D.melanogaster*.

The fact that the *nonA* gene is sex-linked actually has implications for the evolution of new song patterns. As Ewing (1969) discussed, sex-linkage could provide the conditions for rapid fixation of a new song characteristic in a population, since a gene controlling song traits, if located on the X chromosome, would have immediate expression in the males. The process of fixation of a new song pattern could be further accelerated if the same gene controlling song production in the males, also controlled song receptivity in heterozygous females. If heterozygous females were able to respond to the new song pattern (implying some form of codominance), they would produce progeny in which the males produce the new pattern and the females are capable of responding to it. Under certain conditions the selective advantage for the new song characteristics could be high, and by virtue of being sex-linked, it could rapidly become fixed in the population. Such behavioural and genetic coupling between song production and song reception has sparked a controversial debate as to how this might evolve (Alexander, 1962; Kyriacou and Hall, 1986; Butlin and Ritchie, 1989).

The major problem in choosing to transfer the *D.virilis nonA* gene into *D.melanogaster* mutant hosts became evident only at the end of my work during the song analyses (chapter 7). As described in that chapter, it was found that the songs of *D.virilis* have positive CPP gradients, a result that was completely unexpected, since none of the many publications from Hoikkala's group had identified this characteristic (Hoikkala *et al.*, 1982; Hoikkala, 1985; Hoikkala and Lumme, 1987). This complicated the interpretations of my results, since one of the principle characters studied in order to determine whether a song is mutant or not, is the CPP gradient (bCPP). Wild-type *D.melanogaster* flies produce songs with very small or negative gradient, while mutant songs (i.e. *nonA^{diss}*) have very pronounced positive gradients (Kulkarni *et al.*, 1988; Rendahl *et al.*, 1992; 1996). When the transformant *nonA⁽⁻⁾* [*nonA^{vir}*] flies exhibited a mild positive CPP gradient in their lovesong, it became difficult to determine whether this was due to an incomplete rescue of the song phenotype, or a transfer of the *D.virilis* characteristic.

The work described in this thesis consisted in cloning, and characterizing the *D.virilis nonA* gene and studying its functional properties by *P* element-mediated transformation into *D.melanogaster* hosts. This kind of interspecific analysis of genes, including their molecular and functional characterization is not new. Several *D.melanogaster* and *D.virilis* orthologues have been studied (see for example Colot *et al.*, 1988; Treier *et al.*, 1989; Michael *et al.*, 1990; Heberlein and Rubin, 1990; Kassiss *et al.*, 1986; O'Neil and Belote, 1992; Hart *et al.*, 1993, Wallrath and Friedman, 1991; Yao and White, 1991; Ousley *et al.*, 1998) and several *D.virilis* genes have been tested, via *P* element-mediated transformation, for their ability to function properly in a *D.melanogaster* background (O'Neil and Belote, 1992; Wallrath and Friedman, 1991; Hart *et al.*, 1993; Heberlein and Rubin, 1990). In all the cases cited, it was found that the *D.virilis* gene was able to partially or completely rescue the corresponding mutant phenotype in *D.melanogaster*. Particularly interesting was the case of the *D.virilis tra* gene which, in spite of displaying a very high level of divergence from its *D.melanogaster* orthologue (the identity of the two species' TRA proteins is only 50%, see chapter 3), was able to rescue, at least partially, the mutant diplo-X *tra* pseudomale phenotype (O'Neil and Belote, 1992). These data were encouraging when initially

planning my project, as the chances of robust rescue of *nonA*⁽⁻⁾ phenotype was considered high.

The overall identity of the *D.virilis* and *D.melanogaster* NONA proteins was calculated to be 75%, a figure that did not significantly differ from the levels of identity found between several other *D.virilis* and *D.melanogaster* proteins (see chapter 3, table 3.4). The NONA protein could be divided into two distinct parts: the N-terminal which showed high level of divergence between the two species, and the C-terminal which was almost identical in the two proteins. Not surprisingly the C-terminal contains the two RNA-binding domains and the acidic domain of *nonA*, which are conserved between NONA and other mammalian and *Drosophila* putative RNA-binding proteins (Rendahl *et al.*, 1996). If the major role of NONA is indeed splicing of transcripts that are involved in the determination of visual and lovesong phenotypes, then we could expect that the *D.virilis nonA*, given that its RNA-binding domain is almost identical to that of *D.melanogaster*, might indeed provide these basic functions in *D.melanogaster* flies.

This was the case in the study of the *elav* gene, which also encodes a RNA-binding protein, and whose vital function appears to be that of splicing nervous-system-specific transcripts (Robinow and White, 1988; Robinow and White, 1991). Comparison of the *D.melanogaster* and *D.virilis* ELAV proteins revealed, as in NONA, a highly diverged N-terminal which was also rich in homopolymeric repeats (in this case alanine and glutamine). Three RNA-binding domains encoded within the C-terminal showed an astonishing 100% identity between the two species. Transformation experiments proved that the *D.virilis elav* ORF functioned in a manner indistinguishable from that of the *D.melanogaster*. Moreover, in trying to address the question of whether the diverged region of the protein carried functional information, or whether its presence was simply the result of lowered functional constraints in that part of the protein, Yao and White (1991) deleted the alanine/glutamine-rich domain by 40 amino acids. This deletion had no influence, but the consequences of a complete removal of the alanine/glutamine repeat were not tested.

One of the first decisions which had to be taken in the designing of the experiments, was "how much" of the *nonA*^{vir} clone was to be used to transform *D.melanogaster* flies. Most interspecific transformation experiments resort to using only the coding region of the donor species gene and placing it under the control of the corresponding *D.melanogaster* promoter. This ensures that the gene is properly expressed temporally and spatially in the *D.melanogaster* hosts (see for example Petersen *et al.*, 1988; Yao and White, 1991). Therefore, transforming a *nonA*^{vir} DNA fragment, which contained approximately 3 Kb upstream of the beginning of translation and probably more than 1 Kb after the stop codon (although this cannot be precisely determined, see chapter 3, figure 3.8), was slightly risky. There was the possibility that the 3 Kb upstream the ORF did not contain the intact promoter for *D.virilis nonA*. Similarly, I had no proof that the 3' untranslated material was enough to ensure proper termination and adenylation of the *nonA* transcript. Also, there was the possibility that, given the differences in sequence between the 5' regulatory region of *D.virilis* and *D.melanogaster* (see chapter 3), the promoter of *nonA*^{vir} would not function properly in the *D.melanogaster* background.

The transformation experiments, as described in chapter 4, did not go smoothly. First of all, the construct had the tendency to insert into the X chromosome, which made it difficult to cross into a *nonA*⁽⁻⁾ background. The mobilization of insert 113 caused major problems since I was initially confident in the results of the genetic crosses, which indicated that the insert had jumped to a new chromosomal position, at least in line 168-8, because it complemented the lethal 113 homozygous phenotype. As discussed in chapter 4, I do not know what really happened in lines 168-8 and 67-4 and I did not have the time to investigate further. I am curious to find out why the 5' P element end of the *nonA*^{vir} construct was lost, possibly at the moment of insertion in the genome, on so many occasions (inserts 113, 72, 297-64 and 297-71) and duplicated in one case (112). My suspicion is that the 3' end of the *nonA*^{vir} construct (which, after cloning into the *pW8* vector, is immediately adjacent to the 5' P element end, see chapter 4) might contain numerous repeated sequences. The presence of such repeats could create mispairing of DNA strands, and slippage events during the processes of chromosomal insertion of the construct, which might be likely to result in

the loss or the duplication of the *P* element end. It would therefore be of some interest to sequence the 3' untranslated region of the *nonA*^{vir} construct. I also would have liked to find out more about lines 168-8 and 67-4. I showed that the insert in both lines still contained a restriction site which indicated that the sequence lost at the 3' end of each insert could be from a minimum of few bp to a maximum of 2 Kb. Digesting the DNA of 168-8 and 67-4 flies with an enzyme that cuts frequently, would cut this 2 Kb terminal region into several small pieces. After Southern blotting, a probe binding to this 2 kb sequence could be used (as in chapter 4, figure 4.9). This would hybridize to a series of small fragments from the 2 Kb terminal region. If a control in which the terminal 2 Kb region is intact was included in the analysis, it would be possible to determine, by comparison, which of the small fragments are missing in the blots of 168-8 and 67-4, thus discovering more precisely the extent of the deletion (if any) affecting these two inserts.

Chapter 5 presented evidence that the *nonA*^{vir} gene completely rescues the semilethality caused by a loss of the *nonA* locus (*nonA*^(-/-)). Furthermore, it was shown that the *nonA*^{vir} construct included the sequence of the complete *l(1)i19e* essential gene since it also rescued the deletion affecting both *nonA* and *l(1)i19e* (*T(1;4)9e2-10*, see chapter 5, figure 5.1 and 5.2). These results indicated that *nonA*^{vir} was functioning in *D.melanogaster* at least at a basic level.

The visual behaviour experiments (chapter 6) revealed that the *nonA*^{vir} gene rescues the mutant (*nonA*^(-/-)) phenotype as well as the *mel nonA*⁺ transgene (control line 235R11/d). Only one transformed line (191/d) performed significantly worse than the control when tested on small stripes. This result is very likely to be due to position effect.

Two main results need to be underlined from the song analysis of the transformants described in chapter 7. First is the finding on CPP values. The fact that *nonA*^{vir}/*nonA*⁺ transformants had CPP values in general higher than the corresponding *nonA*^(-/-)[*nonA*^{vir}] transformants, but wild-type CPP gradients, could only be explained as a transfer of species-specific information (for higher CPP number) from the *D.virilis* gene. This result was reminiscent of the *D.pseudoobscura per* gene which manifested

its effects on the locomotor activity patterns when placed on a *D.melanogaster* wild-type background (Petersen *et al.*, 1988). When placed on the mutant *per*⁰ background, its effects were masked by incomplete rescue. Similarly, the effects of the *nonA*^{vir} gene on the mutant background (*nonA*⁽⁻⁾) are difficult to interpret because in these transformants the presence of a (moderately) positive CPP gradient does not allow us to rule out the hypothesis of incomplete rescue versus species-specific transfer. Since the *nonA*^{vir}/*nonA*⁺ transformants do not show a positive CPP gradient, they cannot be classified as "mutants" and therefore their increased CPP values are likely to be the result of the presence of *nonA*^{vir}. The second encouraging result was obtained when studying the fourth moment (kurtosis) of the distribution of CPP values in the *nonA*⁽⁻⁾[*nonA*^{vir}] transformants. The results indicated that these lines were more similar in this aspect to *D.virilis*.

The other interesting result was obtained after applying multivariate analysis to the song data. As seen from figure 7.15, when all the song characters are considered simultaneously, the *nonA*⁽⁻⁾[*nonA*^{vir}] (*vir/d*) transformant lines are clearly different from the other transformants carrying the *mel nonA*⁺ or the combination of *nonA*^{vir} and *nonA*⁺ (hybrids *nonA*^{vir}/*nonA*⁺). These results also suggest that subtle effects from the *nonA*^{vir} transgene percolate through to bCPP, PF, bPF, IPI and SSF, as all these variables contributed to the significant discrimination between groups. Thus *nonA*^{vir} has a small but detectable effect on several song features. While the results are not as striking or as clear-cut as those obtained with the *per* gene (Wheeler *et al.*, 1991), it must be remembered that *per* only transferred the IPI rhythm from *D.simulans* to *D.melanogaster*, but no other song character was affected (Wheeler *et al.*, 1991).

It would be worthwhile to assess whether the song phenotypes associated with the *nonA*^{vir} transgene can be attributed to the coding sequence, its regulatory regions or to both. The only definitive way to answer this question would be to construct and transform chimaeric genes composed of reciprocal combinations of the *D.virilis* and *D.melanogaster* coding and regulatory regions. Examination of Northern, Western blots, or immunohistochemistry (assuming the *D.melanogaster* NONA antibody recognizes the *D.virilis* epitopes) would reveal whether the *nonA*^{vir} transgene is

expressed normally in the transformant lines. Is it expressed in the right regions, for example in the thoracic ganglia responsible for song production, (Schilcher and Hall, 1979; Hall, 1994)? Are there any indications that the differences seen between the songs of *nonA*⁽⁻⁾[*nonA*⁺] and *nonA*⁽⁻⁾[*nonA*^{vir}] transformants could be due to incomplete rescue, perhaps by a quantitative or spatial difference in *nonA* expression?

If the promoter of *nonA*^{vir} is not functioning optimally in a *D.melanogaster* background, we might expect to see differences in the amount or even localization of the *nonA* product compared to that of the control line. This approach would also reveal whether some of discrepancies in the results found for different transformant lines (see for example the optomotor results in chapter 6) could be ascribed to changes in expression. However, if the song characteristics carried by the *nonA*^{vir} transgene were to be attributed exclusively to its coding region, this would further support the assumption that the *nonA*⁽⁻⁾[*nonA*^{vir}] transformants are not mutant in their song behaviour, because these flies are normal in their visual behaviour. No coding region mutation of *nonA* has been isolated that affects solely the song behaviour without affecting vision, so the visual phenotype appears to be the most sensitive to mutation (Rendahl *et al.*, 1992; 1996, Stanewsky *et al.*, 1996). Logically, then, if the *nonA*⁽⁻⁾[*nonA*^{vir}] transformants are rescued in their visual behaviour, they should be rescued in their song, indicating species-specific transfer of song traits.

As seen in chapter 3, the N-terminal of the NONA protein has diverged considerably between *D.virilis* and *D.melanogaster*. This makes it the most likely candidate for conveying any species-specific characteristics. Interestingly, the most striking difference in the N-terminal of *nonA*^{vir} is the presence of a very long uninterrupted poly-Glycine repeat. It would be valuable to see whether deleting (or shortening) this repeat has any effect on the visual or song behaviour. It was mentioned in chapter 3 that poly-Glycine regions in RNA-binding proteins could serve as protein-protein interaction domains (Bandziulis *et al.*, 1989). In the case of the *per* gene, the repetitive Threonine-Glycine tract, which lies in one of the most diverged regions of the protein (Peixoto *et al.*, 1993), is responsible for the transfer of species-

specific song information between *D.simulans* and *D.melanogaster* (Wheeler *et al.*, 1991).

Many more experiments are needed to fully understand the functions of the *nonA* gene. For instance, the *nonA* gene appears to be ubiquitously expressed throughout all the stages of the *Drosophila* life cycle (Rendahl *et al.*, 1992). The limited and specific arrays of phenotypes displayed by *nonA* mutants is therefore rather surprising. One of the questions that could be addressed is where (i.e. in which organs), and when, is the product of *nonA* needed for the fly to display normal behaviour? The question "when" is the *nonA* product necessary has already been answered by Rendahl and Hall (1996), (as mentioned in chapter 1) by placing *nonA* under the control of a heat shock promoter and switching its expression at determined developmental stages. Federica Sandrelli, a doctoral student from Padova, Italy, and I have started to investigate "where" the product of *nonA* must be expressed to generate normal behaviour. We have performed a functional dissection of the *D.melanogaster nonA* promoter, by making a series of nested deletions. Promoter sequences containing deletions ranging from a minimum of ~200 bp to a maximum of ~1700 bp were placed 5' of the yeast transcriptional activator gene GAL4 (Brand and Perrimon, 1993), while the *D.melanogaster nonA* cDNA (Besser *et al.*, 1990) was cloned downstream of the GAL4-responsive UAS (Brand and Perrimon, 1993). A transformant line carrying the fusion of UAS with the reporter gene *lacZ* was also used in order to determine the expression pattern of the different *nonA*-GAL4 constructs. At the same time, crosses between the *nonA*-GAL4 lines with the UAS-*nonA* line permitted us to study the behavioural effects (on a *nonA^{diss}* mutant background) of the promoter deletions. The ultimate goal of this project is to correlate changes in the temporal and spatial expression of UAS-*nonA* with the behavioural phenotypes of the transformants, and with the relevant promoter sequences. At the moment we are waiting for the behavioural results of this project.

The recent cloning of the *Dmca1A*-encoding locus (Smith *et al.*, 1996) whose *cac* mutation produces song and visual aberrations strikingly similar to those observed for *nonA^{diss}* (Kulkarni and Hall, 1987) permits a test of whether the *Dmca1A* transcript

could possibly be the target of NONA. The transcript of the *Dmca1A*-encoding locus displays a complicated pattern of alternative splicing which could be regulated by NONA, as already discussed in chapter 1. *nonA* mutants (*nonA*⁽⁻⁾ or *nonA*^{dis}) flies could be transformed with any of the *Dmca1A*-encoding cDNAs under the control of a heat shock promoter. If such a transgene rescued the visual and song phenotypes of *nonA* mutants it would be an indication that the *Dmca1A*-encoding gene is in fact regulated by the *nonA* gene, presumably by splicing. Under this scenario, *nonA*^{vir} might splice a different isoform of *Dmca1A* in the *D.melanogaster* transformants, suggesting a different profile of *Dmca1A* cDNAs in *D.virilis*.

To conclude, if my interpretation is correct, *nonA*^{vir} may carry some species-specific information for the song characteristics that I have studied, because subtle but detectable changes were observed in the transformants. However, the larger differences in the song patterns between *D.melanogaster* and *D.virilis* are undoubtedly encoded by many genes. *nonA* may contribute a small component of this variance, thereby conforming to the traditional view that species-specific characteristics accumulate by small changes at many loci (Mayr, 1977; Coyne, 1992).

APPENDICES

FOR CHAPTER 3

Appendix 3.1.

Sequence and position (relative to the *D.virilis nonA* gene sequence, figure 3..) of the primers used in the sequencing of the *D.virilis nonA* gene. *= the position of this primer is relative to the *D.melanogaster nonA* gene sequence by Jones and Rubin (1990).

PRIMER	SEQUENCE	POSITION
5'E1	5'-ATGGAAAATTCTGTAAA-3'	1-19
3'E1	5'-AAATCCACCCTTGTTTCGC-3'	276-261
3'7A	5'-ATTTCTCCGTCGCTTT-3'	393-376
5'7B	5'-CACAACCTCAAGCGCAGGCC-3'	464-483
3'POL	5'-CTAATGCTGCGTAACCGTTG-3'	791-772
5'C	5'-CGCGAAATGTTTAAGCCG-3'	904-921
3'39A	5'-GCGACCCCTTACGCATTGA-3'	1044-1027
5'2EXa	5'-CCAACGAGCTGCTCTACAAG-3'	1112-1131
3'A	5'-GCCCCGCTCAATGGGGCCAAA-3'	1163-1144
3'2EXa	5'-CTTGCCACGATCGTCGACTG-3'	1191-1172
5'39A	5'-AAATGCTTCTTTTAACG-3'	1264-1281
3'39B	5'-CTCCTGATTGAACTCTTG-3'	1386-1369
5'2EXb	5'-CAATCAGGAGCGCAGTGTTG-3'	1377-1396
5'E2	5'-CAGATGGAGTATGCTCGC-3'	1531-1548
3'E2	5'-GCGAGCATACTCCATCTG-3'	1548-1531
5'3EXa	5'-CGCGAATCTGATAATGAGCG-3'	1591-1610
5'3EXb	5'-CCGTTCGAAGTGTTTCAGAGG-3'	1906-1925
3'3EXa	5'-CCTCTGAACACTTCGAACGG-3'	1925-1906
5'5	5'-TAGGACTCCTTCGCTTTTGA-3'	7106-7125*
5'5EXa	5'-GGACTCCTTCGCTGCTTTTG-3'	1986-2005
3'5EXa	5'-CCTTGGTTCATATTGTAAAC-3'	2033-2014

Appendix 3.2. Translation in the three reading frames (a, b, c) of the beginning of the *D.virilis* *nonA* intron four.

```

1 GTATGTAAACGTACAGGACATAGTATTGTGAGCTGCTATATATAGCATATGGCATATGGT 60
a V C K R T G H S I V S C Y I * H M A Y G
b Y V N V Q D I V L * A A I Y S I W H M V
c M * T Y R T * Y C E L L Y I A Y G I W S

61 CGACTCCCATATGAAACCAGATGCCGCCGGCCTGTGCCCTTATTAATATAACATAAAAGT 120
a R L P Y E T R C R R P V P L L I * H K S
b D S H M K P D A A G L C P Y * Y N I K V
c T P I * N Q M P P A C A L I N I T * K L

121 TACCATTGTTGCCGTCGCTCATTTATCGGTTATATAATTTCAATGACTGTTCTCACTCAT 180
a Y H C C R R S F I G Y I I S M T V L T H
b T I V A V A H L S V I * F Q * L F S L M
c P L L P S L I Y R L Y N F N D C S H S C

181 GTAGATAGCAAATCAAA 197
a V D S K S
b * I A N Q
c R * Q I K

```

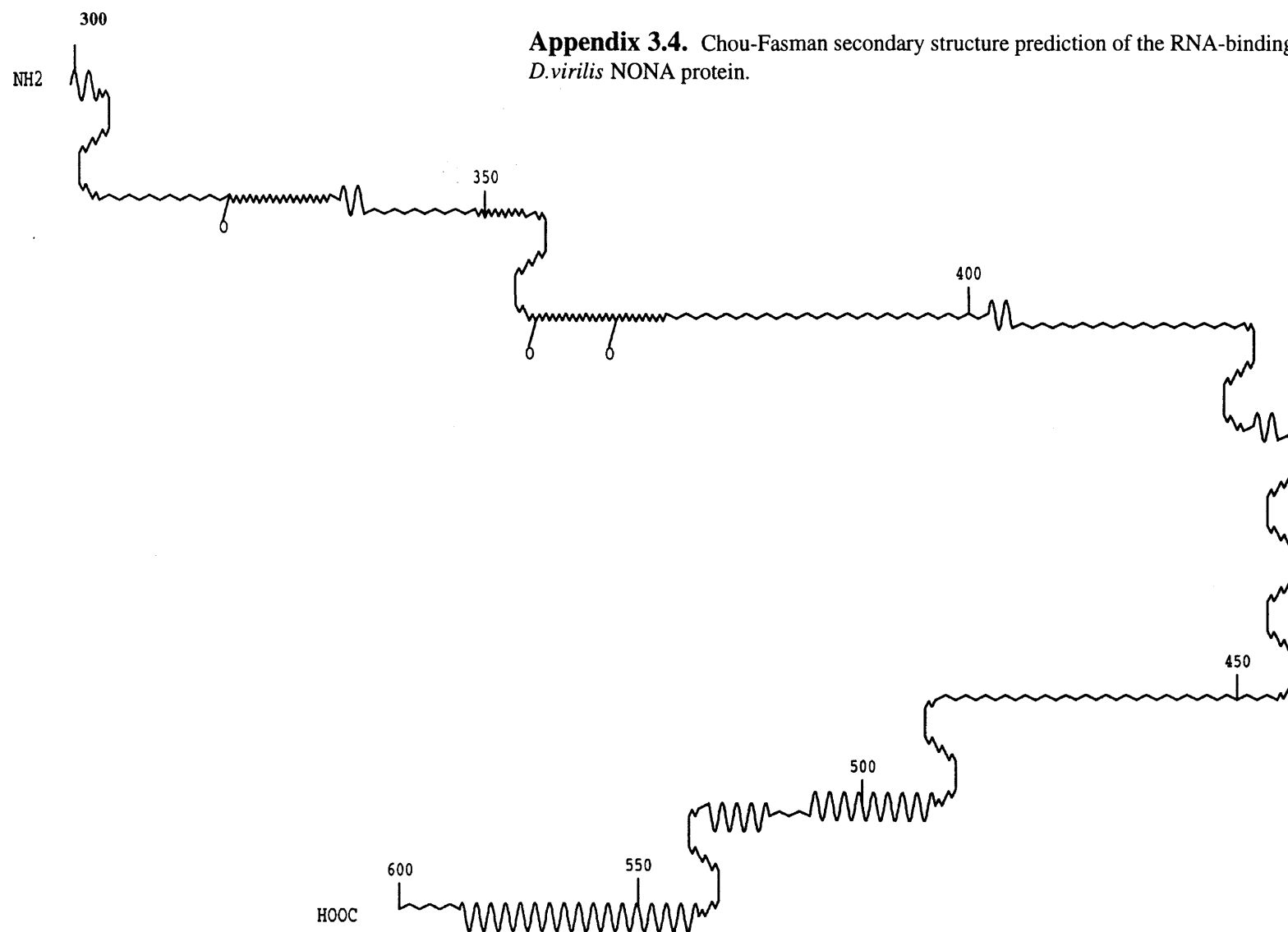
Appendix 3.3. Alignment of the *D.virilis* and *D.melanogaster nonA* DNA sequence. The sequence homologous to the DBHS domain (see Introduction) is highlighted in yellow. The two RNA-binding domains (RRM1 and RRM2) are underlined. The RNP1 motifs are shown in blue and the RNP2 in red.

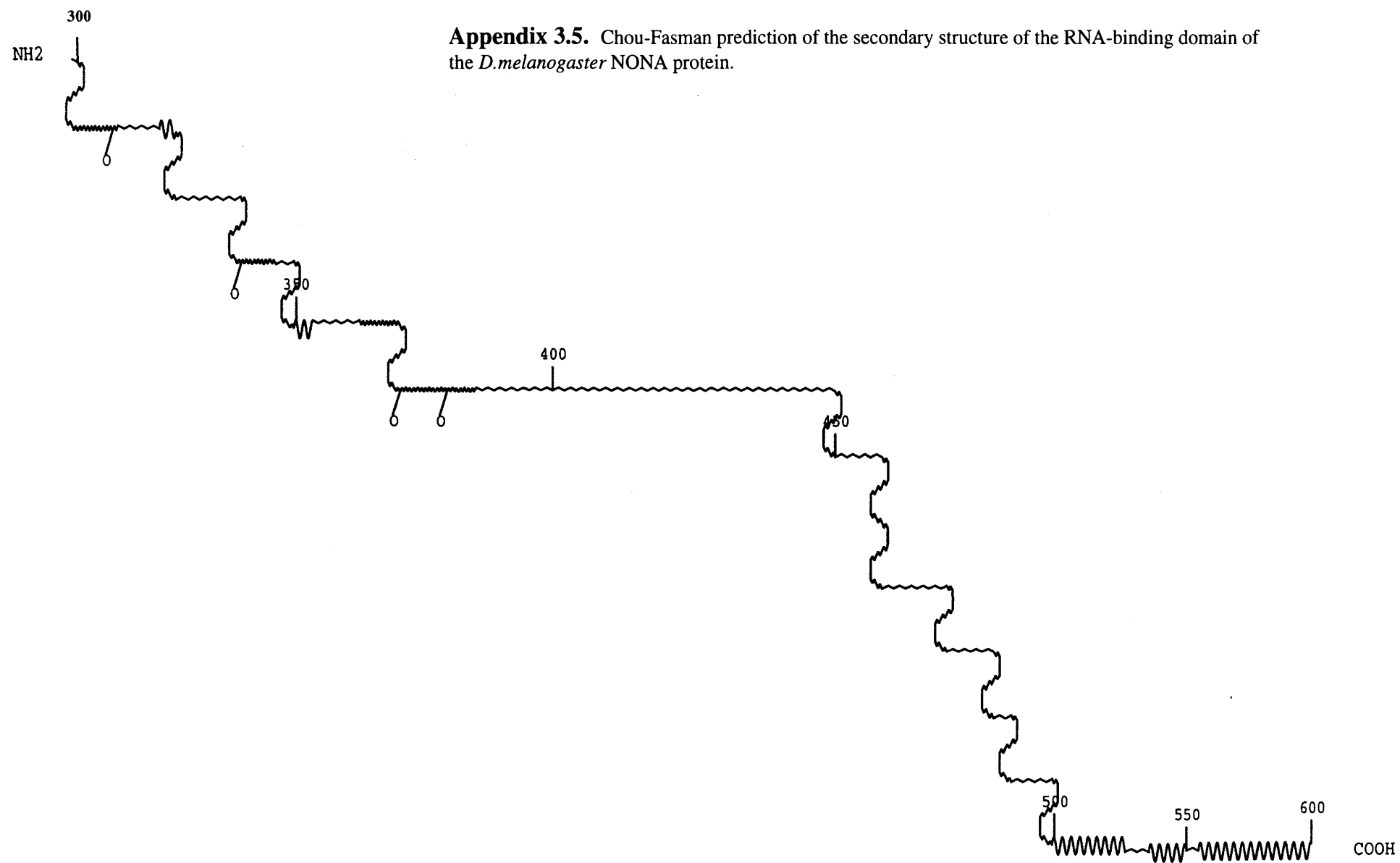
vir	1	ATGGA	AAAATTCTGT	TAAATGGACA	ATTCTGGCA	ATTCAACAC	CGCTGCC	50
mel	1	ATGGA	AAGTGCGG	GGAAACA	AAGATAATA	ACGCCA...	CGCAACAGTTGCC	47
vir	51	ACAGC	GCCAACGA	AGAGCTA	ATAATCA	ACCGAATA	AAAAATAT	CGGCAAAC 100
mel	48	GCAGC	GACAGC	AGCGGG	AAACCAG	CAGGCCA	ACAAAAAT	CTCGGCAAGC 97
vir	101	TCGGC	CCACAGA	AGCAAA	ATGAAG	TGCATCC	GACGGAG	GGCCAGCAGAA 150
mel	98	ATAAT	GCACAAA	AAACAAA	ATGA...	TTCCG	CGGACGG	AGGCCCGGCTGAA 144
vir	151	AAACG	GCAACG	TTTTTG		CCCGA	ATAATCA	GAAATGGTGGCGG 191
mel	145	AAGA	ACAACG	ATTTCG	TGGTCC	AAATGCC	CCAGAAT	CAAAACCAGAACCA 194
vir	192	TGGTA	GCCTCG	TGGTGG	TGGTGG	CGGCGG	CGGCGG	TGGAGGTGGCCAAA 241
mel	195	GAATC	AAAATG	TGGTGT	AACGGG	CGGTGG	TGGTGC	AGTTGGTGGCCCCA 244
vir	242	ATCAA	AATAAAA	ATTTTC	G...CG	AACAAG	GGTGGAT	TTGGCGGAGGTGGA 288
mel	245	ATCAA	AATAAG	AATTTT	TGGAAC	AACAAG	GGTGGG	TTCTGTTGAAACCGC 294
vir	289	AATCG	CAACCG	CAATCG	TGGCGG	CAATCAA	AAATAGG	TCGAATTTTCAAAA 338
mel	295	AATCG	CAACA	AATAAT	CGCGCT	TGGAAC	CAAAAC	CGGACTTTTCAGGCAA 344
vir	339	TCAGA	ATCAAA	ACCAAA	AGTCC	ACAAC	AGATGC	ACCAAAAGCGGACGGAG 388
mel	345	CAACA	ATTCAA	ATCAAA	AGCCCA	ATAAC	GAAAC	ATCAAAGGCGGACGGCC 394
vir	389	GAAAT	CTCAAT	GACAA	ATCTAA	CGAAG	CTAACA	ATGC.....A 426
mel	395	CTAAT	GCAC	TGCTA	AGAACA	ATGAAC	CGGCA	ACTGCAGCTGCTGGCCAA 444
vir	427	AATCA	ATCCA	ATTC	CAATTC	AGCTGC	CCCAAG	CCCAGGCACAACTTCAAGC 476
mel	445	AATCA	AGCG	AATC	AGAA...	CGCCA	ACAAGG	GCCAAAATCAACGGCAAGG 491
vir	477	GCAGG	CCCCA	AGCTCA	TGCCAG	GCCCA	AGCTCA	GGGCTCA..... 515
mel	492	ACAAA	ACCAAA	ATCAAA	ATCAAG	TCCATG	GTCAAG	GAAATCAAGGAGGAC 541
vir	516		GGCTC	AGGCTC	ATGCTC	AGGCC.....	CAGGCT	543
mel	542	CAGGA	AATCA	AGGGG	GAGCAG	GCAATC	AAGGAG	GCCAAGGAAATCAAGGA 591
vir	544	CAAGC	TGCG	CATGCT	CAAAAT	CAAGC	ATTTAG	GGCTCGCGGCGGTGG 593
mel	592	GGAGC	CGGAA	ATCAGG	GAAATG	GTCAAG	GATTCA	GAGGTCGCAATGCTGG 641
vir	594	CGGCG	GCGG	CGGTGG	TGGTGG	TGGTGG	TGGTGG	AGGTGGTGGTGGC.....GGCG 637
mel	642	TAACA	ATCA	AGGTG	GCGGTT	TCTCCG	GAGGTC	CGCAGAACAGCAGCGTG 691

vir 638 GCGGCGGCGGCGGAGGTGGAGGAGGAGGAGGCGGACGTGATAGGAATCCG 687
|
mel 692 ACAACCGAAACCGGAGTGGTCCACGCCCTGGCGGAGGTGCTGGTGGCGCG 741
vir 688 GATCGGCGTGGTGGTGGTGGCGGCGGCGGACAAACTCTGGCGGCGGTAA 737
|
mel 742 ATGAACAGCACCAATATGGGTGGTGGTGGTGGTGGCGGTGGAGGCGGT.. 789
vir 738 CAATTCACAGCGTGGCGATGACTTCTTTTATAGCCAACGGTTACGCAGCA 787
|
mel 790 .GGTCCGCGCGGCGGTGAAGACTTCTTCATTACCCAGCGGTACGCAGCA 838
vir 788 TTAGTGGTCCCACACATGAGCTACCACCGATCGAAGTCGCCCAGGAGACC 837
|
mel 839 TTTCTGGGCCACCTTTGAACTGGAACCGGTGGAAGTGCCACGGAGACG 888
vir 838 AAGTTCTCTGGACGCAATCGTCTCTATGTTGGCAATTTGACCAATGACAT 887
|
mel 889 AAGTTCTCTGGACGAAATCGTCTCTATGTGGGCAACCTGACCAATGACAT 938
vir 888 TACCGATGAGGAGTTGCGCGAAATGTTAAGCCGTACGGTGAAATAGGCG 937
|
mel 939 CACCGACGATGAGCTGCGCGAGATGTTCAAGCCATATGGCGAGATTAGCG 988
vir 938 AGATATTTTCCAATCTTGAAAAAACTTTACATTCTTAAGGTTGACTAT 987
|
mel 989 AGATCTTCTCGAACCTGGATAAAGAACTTTACATTCTTAAAGGTCGACTAT 1038
vir 988 CATATTAATGCCGAAAAGGCCAAGCGGCCCTTAGACGGCTCAATGCGTAA 1037
|
mel 1039 CATCCCAATCGCGAGAAGGCTAAACGCGCCTTGGACGGATCGATGCGAAA 1088
vir 1038 GGGTCGCCACGTTTCGCGTCCGTTTCGCCCCCAACGCAACCATTTTGCGGTG 1087
|
mel 1089 GGGTCGCCAGTTGCGTGTACGATTTCGCACCGAATGCCACCATAATTGCGGG 1138
vir 1088 TTAGCAATCTGACACCGTTCGTGTCCAACGAGCTGCTCTACAAGTCCTTT 1137
|
mel 1139 TTAGCAATCTCACACCGTTCGTTTCCAACGAGCTGCTGTACAAGTCCTTC 1188
vir 1138 GAAATATTTGGCCCCATTGAGCGGGCCAGCATTACAGTCGACGATCGTGG 1187
|
mel 1189 GAAATCTTTGGTCCCATCGAGAGGGCCAGCATCACCGTCGACGATCGTGG 1238
vir 1188 CAAGCATTGCGGCGAGGGCACCGTTGAGTTTGCCAAAAAATCATCGGCCA 1237
|
mel 1239 CAAGCATATGGGCGAGGGCATAGTTGAGTTTGCCAAAAGTCATCGGCCA 1288
vir 1238 GCGCCTGTCTGCGCTTGTGCAATGAGAAATGCTTCTTTTAAACGGCCTCG 1287
|
mel 1289 GCGCCTGTCTGCGGATGTGCAATGAGAAGTGCTTCTTCTGACTGCTTCC 1338
vir 1288 TTGCGTCCGTGCTTGGTGGAAACCGATGGAGGTTAACGATGACAACGACGG 1337
|
mel 1339 CTTGACCGTGTCTGGTTGATCCGATGGAGGTGAATGACGATACCGACGG 1388
vir 1338 ATTGCCCCGAGAAGGCCCTCAATAAGAAGCTGCAAGAGTTCAATCAGGAGC 1387
|
mel 1389 ACTGCCGAGAAGGCGTTCAACAAAAAGATGCCGACTTCAACCAGGAGC 1438
vir 1388 GCAGTGTTGGGCCCCGTTTCGCCGATCTAAATTCATTTGAGCACGAGTAC 1437
|
mel 1439 GTAGCATTGGTCCGCGCTTCGCCGATCCCAATTCGTTTGAACACGAGTAC 1488

vir	1438	GGCTCCAGGTGGAACAGTTGCATGATCTATTCAAAAGCAAGCAAGATGC	1487
mel	1489	GGTTCGCGCTGGAAGCAGCTGCACAATTTGTTCAAACCAAGCAGGACGC	1538
vir	1488	TCTCAAGCGTGAGCTCAAAATGGAAGAAGAAAAACTTGATGCTCAGATGG	1537
mel	1539	ACTCAAACGCGAAGTAAAATGGAGGAGGACAAGTTGGAGGCTCAGATGG	1588
vir	1538	AGTATGCTCGCTATGAGCAAGAACAGAGCTTTTTCGTCAAGAATTAAGA	1587
mel	1589	AGTATGCACGCTATGAGCAAGAACTGAACTGTTGCGTCAAGAGCTGCGA	1638
vir	1588	AAACGCGAATCTGATAATGAGCGTAAGAAATTGGAATGGGAAATGCGCGA	1637
mel	1639	AAACGCGAAGTTGACAACGAGCGCAAGAAATTGGAGTGGGAGATGCGCGA	1688
vir	1638	GAAACAGGCCCGAGGAAATGCGCAAACGCGAAGAGGAGACCATGCGTCGTC	1687
mel	1689	GAAGCAGGCCCGAGGAGATGCGCAAGCGCGAGGAGGAGACCATGCGACGTC	1738
vir	1688	ATCAAACCGAAATGCAGAGCCGTATGGTGC GCCAGGAGGAGATATGCGT	1737
mel	1739	ATCAGACCGAGATGCAGAGCCACATGAATCGCCAGGAGGAGACATGCTC	1788
vir	1738	CGTCGTCAGCAGGAGAACACATTATTCATGCAGGCC CAGCAGCTAAACTC	1787
mel	1789	CGTCGGCAGCAGGAG...ACACTGTT CATGAAGGCACAGCAGCTCAATTC	1835
vir	1788	GCTGCTTGATCAACAGGAAGGCTTCGGCGGTGGCAACGGTGGTGGCGGCG	1837
mel	1836	GTTGCTGGATCAGCAGGAGGGATT.....	1859
vir	1838	GCGGCGGTGGTGGCGGCGGCGGCGTGGGCAATTCAAAC TTTGATAACTTC	1887
mel	1860	TGGTGGCGGCGGCGGCGGTAACAACTCCACTTTTGACA ACTTC	1902
vir	1888	GGGGGCAACAGCAACTCCCCGTTTCAAGTGTT CAGAGG.....TAATAA	1931
mel	1903	GCTGGCAATAGCAATTCGCCATTCGAGTTTTTTAGAGGCAATAACAATAA	1952
vir	1932	CAACTCGTCCATGGCTGGCAACAATGCCG GACCCGGTGCAAATAACCAGC	1981
mel	1953	CAATTCCACAATGATTGGAAACAATGCTGCC CCCC AACACA.....	1992
vir	1982	AGCAGGACTCCTTCGCTGCTTTTGAATTTGGGGTTAACA ATATGAACCAA	2031
mel	1993	..CAGGACTCCTTC...GCTTTTGAATTTGGGGTTAACA ATATGAACCAA	2037
vir	2032	GGCGGCAATCAACGTGGAAATAATGGC...GGAAATAAT GTTCCATGGGG	2078
mel	2038	GGCGGCAATCAACGTGGAAATAATGGCGGAGGAAATAAT GTCCCATGGGG	2087
vir	2079	ACGACGTCGTTTTTTAGAC	2096
mel	2088	ACGTCGACGTTTTTTAG..	2103

Appendix 3.4. Chou-Fasman secondary structure prediction of the RNA-binding domain of the *D.virilis* NONA protein.





APPENDICES

FOR CHAPTER 6

Appendix 6.1. Newman-Keuls probability matrix based on the two way ANOVA for the optomotor test results. Any significant probability ($p < 0.05$) is shaded in grey. Bold numbers at the head of each column indicate each genotype's average correct turns.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)
	17.60	17.80	17.70	16.80	17.70	16.80	18.20	18.20	19.20	17.70	15.80	14.60	16.40	17.30	14.50	16.20	15.20	15.90	15.60	15.10	16.80	15.50	16.60	15.00	16.30	14.10	16.70	16.90	16.40	16.30	18.10	15.70	16.20	15.00	17.40	16
canton.s small (1)		0.999	0.906	0.965	0.992	0.934	0.992	0.997	0.616	0.999	0.797	0.074	0.943	0.933	0.054	0.945	0.364	0.841	0.683	0.301	0.878	0.617	0.960	0.257	0.947	0.010	0.964	0.840	0.921	0.929	0.992	0.744	0.930	0.244	0.813	0.878
canton.s large (2)	0.999		0.999	0.984	0.992	0.975	0.883	0.965	0.459	0.906	0.730	0.043	0.945	0.997	0.130	0.927	0.271	0.789	0.591	0.215	0.960	0.515	0.971	0.175	0.938	0.005	0.979	0.964	0.930	0.924	0.722	0.663	0.914	0.166	0.997	0.839
235r11/d small (3)	0.906	0.999		0.964	1.000	0.938	0.992	0.997	0.635	1.000	0.744	0.054	0.929	0.965	0.038	0.924	0.301	0.797	0.617	0.244	0.895	0.546	0.953	0.204	0.930	0.007	0.960	0.878	0.906	0.911	0.990	0.683	0.907	0.193	0.933	0.841
235r11/d large (4)	0.965	0.984	0.964		0.979	1.000	0.911	0.930	0.223	0.988	0.990	0.499	0.990	0.976	0.431	0.999	0.878	0.993	0.979	0.841	1.000	0.970	0.969	0.819	0.997	0.161	0.906	0.999	0.965	0.992	0.929	0.986	0.997	0.797	0.981	0.995
113/d small (5)	0.992	0.992	1.000	0.979		0.964	0.976	0.992	0.563	1.000	0.767	0.057	0.947	0.990	0.040	0.938	0.318	0.819	0.641	0.257	0.938	0.569	0.968	0.215	0.945	0.007	0.975	0.934	0.929	0.930	0.965	0.707	0.924	0.204	0.985	0.862
113/d large (6)	0.934	0.975	0.938	1.000	0.964		0.886	0.911	0.202	0.979	0.994	0.523	0.997	0.934	0.453	0.999	0.897	0.996	0.905	0.862	1.000	0.978	0.995	0.839	0.999	0.171	0.992	0.992	0.990	0.997	0.906	0.991	0.999	0.819	0.954	0.997
67-4/d small (7)	0.992	0.883	0.992	0.911	0.976	0.886		1.000	0.462	0.934	0.402	0.008	0.771	0.979	0.005	0.707	0.084	0.474	0.271	0.060	0.856	0.215	0.855	0.045	0.744	0.001	0.886	0.876	0.743	0.718	0.906	0.334	0.683	0.043	0.981	0.546
67-4/d large (8)	0.997	0.965	0.997	0.930	0.992	0.911	1.000		0.236	0.976	0.421	0.008	0.797	0.988	0.005	0.730	0.088	0.495	0.284	0.064	0.886	0.226	0.878	0.048	0.767	0.001	0.907	0.906	0.771	0.744	0.992	0.350	0.707	0.045	0.990	0.589
168-8/d small (9)	0.616	0.459	0.635	0.223	0.563	0.202	0.462	0.236		0.480	0.014	0.000	0.092	0.467	0.000	0.061	0.001	0.020	0.006	0.001	0.182	0.004	0.150	0.000	0.076	0.000	0.184	0.213	0.084	0.070	0.560	0.009	0.056	0.000	0.503	0.029
168-8/d large (10)	0.999	0.906	1.000	0.988	1.000	0.979	0.934	0.976	0.480		0.789	0.060	0.960	0.997	0.043	0.950	0.334	0.839	0.663	0.271	0.964	0.591	0.979	0.226	0.957	0.007	0.984	0.965	0.947	0.945	0.883	0.730	0.938	0.215	0.997	0.881
135/d small (11)	0.797	0.730	0.744	0.990	0.767	0.994	0.402	0.421	0.014	0.789		0.889	0.997	0.924	0.876	0.965	0.954	0.906	0.969	0.962	0.996	0.985	0.995	0.981	0.992	0.638	0.993	0.994	0.999	0.997	0.474	0.906	0.990	0.965	0.897	0.969
135/d large (12)	0.074	0.043	0.054	0.499	0.057	0.523	0.008	0.008	0.000	0.060	0.889		0.743	0.182	0.906	0.762	0.954	0.876	0.900	0.934	0.546	0.895	0.630	0.635	0.758	0.824	0.566	0.474	0.771	0.790	0.012	0.898	0.798	0.883	0.141	0.856
97/d small (13)	0.943	0.945	0.929	0.990	0.947	0.997	0.771	0.797	0.092	0.960	0.997	0.743		0.979	0.690	0.999	0.959	0.997	0.995	0.947	0.999	0.993	0.969	0.945	0.992	0.362	0.985	0.999	1.000	0.906	0.817	0.996	0.995	0.930	0.975	0.997
97/d large (14)	0.933	0.997	0.965	0.976	0.990	0.934	0.979	0.988	0.467	0.997	0.924	0.182	0.979		0.141	0.986	0.592	0.945	0.862	0.523	0.824	0.819	0.982	0.474	0.984	0.034	0.981	0.635	0.964	0.975	0.981	0.897	0.979	0.453	0.906	0.960
191/d small (15)	0.054	0.030	0.038	0.431	0.040	0.453	0.005	0.005	0.000	0.043	0.876	0.906	0.690	0.141		0.723	0.962	0.856	0.898	0.954	0.474	0.900	0.566	0.824	0.712	0.635	0.499	0.402	0.718	0.743	0.008	0.889	0.758	0.934	0.107	0.830
191/d large (16)	0.945	0.927	0.924	0.999	0.938	0.999	0.707	0.730	0.061	0.950	0.965	0.762	0.999	0.986	0.723		0.936	0.933	0.981	0.931	1.000	0.982	0.999	0.943	0.992	0.415	0.999	1.000	1.000	0.999	0.767	0.976	1.000	0.921	0.979	0.813
297-6/d small (17)	0.364	0.271	0.301	0.878	0.318	0.897	0.084	0.088	0.001	0.334	0.954	0.954	0.959	0.592	0.962	0.936		0.962	0.883	0.906	0.914	0.722	0.930	0.995	0.953	0.850	0.907	0.881	0.971	0.968	0.113	0.934	0.960	0.969	0.523	0.965
297-6/d large (18)	0.841	0.789	0.797	0.993	0.819	0.996	0.474	0.495	0.020	0.839	0.906	0.876	0.997	0.945	0.856	0.933	0.962		0.985	0.965	0.998	0.990	0.996	0.979	0.990	0.598	0.995	0.996	0.999	0.997	0.546	0.969	0.985	0.964	0.924	0.906
112/m small (19)	0.683	0.591	0.617	0.979	0.641	0.985	0.271	0.284	0.006	0.663	0.969	0.900	0.995	0.862	0.898	0.981	0.883	0.985		0.934	0.990	0.906	0.990	0.981	0.991	0.697	0.986	0.983	0.997	0.996	0.334	0.906	0.992	0.954	0.819	0.990
112/m large (20)	0.301	0.215	0.244	0.841	0.257	0.862	0.060	0.064	0.001	0.271	0.962	0.934	0.947	0.523	0.954	0.931	0.906	0.965	0.934		0.881	0.883	0.907	0.992	0.943	0.844	0.878	0.839	0.960	0.959	0.084	0.954	0.953	0.906	0.453	0.964
113/m small (21)	0.878	0.960	0.895	1.000	0.938	1.000	0.856	0.886	0.182	0.964	0.996	0.546	0.999	0.824	0.474	1.000	0.914	0.998	0.990	0.881		0.983	0.999	0.858	1.000	0.182	0.999	0.906	0.997	0.999	0.876	0.994	0.999	0.839	0.893	0.999
113/m large (22)	0.617	0.515	0.546	0.970	0.569	0.978	0.215	0.226	0.004	0.591	0.985	0.895	0.993	0.819	0.900	0.982	0.722	0.990	0.906	0.883	0.983		0.986	0.976	0.990	0.713	0.979	0.974	0.996	0.995	0.271	0.969	0.991	0.934	0.767	0.992
67-4/m small (23)	0.960	0.971	0.953	0.969	0.968	0.995	0.855	0.878	0.150	0.979	0.995	0.630	0.969	0.982	0.566	0.999	0.930	0.996	0.990	0.907	0.999	0.986		0.897	0.947	0.251	0.906	0.999	0.813	0.985	0.886	0.993	0.997	0.878	0.981	0.997
67-4/m large (24)	0.257	0.175	0.204	0.819	0.215	0.839	0.045	0.048	0.000	0.226	0.981	0.635	0.945	0.474	0.824	0.943	0.995	0.979	0.981	0.992	0.858	0.976	0.897		0.947	0.710	0.862	0.809	0.957	0.960	0.064	0.982	0.959	1.000	0.402	0.975
168-8/m small (25)	0.947	0.938	0.930	0.997	0.945	0.999	0.744	0.767	0.076	0.957	0.992	0.758	0.992	0.984	0.712	0.992	0.953	0.990	0.991	0.943	1.000	0.990	0.997	0.947		0.393	0.997	0.999	0.999	1.000	0.797	0.992	0.906	0.929	0.979	0.985
168-8/m large (26)	0.010	0.005	0.007	0.161	0.007	0.171	0.001	0.001	0.000	0.007	0.638	0.824	0.362	0.034	0.635	0.415	0.850	0.598	0.697	0.844	0.182	0.713	0.251	0.710	0.393		0.203	0.141	0.385	0.420	0.001	0.671	0.448	0.824	0.023	0.552
135/m small (27)	0.964	0.979	0.960	0.906	0.975	0.992	0.886	0.907	0.184	0.984	0.993	0.566	0.985	0.981	0.499	0.999	0.907	0.995	0.986	0.878	0.999	0.979	0.906	0.862	0.997	0.203		0.999	0.933	0.990	0.911	0.990	0.997	0.841	0.982	0.996
135/m large (28)	0.840	0.964	0.878	0.999	0.934	0.992	0.876	0.906	0.213	0.965	0.994	0.474	0.999	0.635	0.402	1.000	0.881	0.996	0.983	0.839	0.906	0.974	0.999	0.809	0.999	0.141	0.999		0.997	0.999	0.889	0.990	0.999	0.789	0.824	0.998
97/m small (29)	0.921	0.930	0.906	0.965	0.929	0.990	0.743	0.771	0.084	0.947	0.999	0.771	1.000	0.964	0.718	1.000	0.971	0.999	0.997	0.960	0.997	0.996	0.813	0.957	0.999	0.385	0.933	0.997		0.992	0.790	0.998	0.999	0.945	0.960	0.999
97/m large (30)	0.929	0.924	0.911	0.992	0.930	0.997	0.718	0.744	0.070	0.945	0.997	0.790	0.906	0.975	0.743	0.999	0.968	0.997	0.996	0.959	0.999	0.995	0.985	0.960	1.000	0.420	0.990	0.999	0.992		0.771	0.997	0.992	0.947	0.968	0.997
2mel small (31)	0.992	0.722	0.990	0.929	0.965	0.906	0.906	0.992	0.560	0.883	0.474	0.012	0.817	0.981	0.008	0.767	0.113	0.546	0.334	0.084	0.876	0.271	0.886	0.064	0.797	0.001	0.911	0.889	0.790	0.771		0.402	0.744	0.060	0.982	0.617
2mel large (32)	0.744	0.663	0.683	0.986	0.707																															

APPENDICES

FOR CHAPTER 7

Appendix 7.1.

Newman-Keuls probability matrix based on the ANOVA for bCPP. Only transformant lines are included. Any significant probability ($p < 0.05$) is shaded in grey.

Bold numbers at the head of each column indicate each genotype's average bCPP value.

	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}	{12}	{13}	{14}	{15}	{16}	{17}
	-0.002200	.0040800	.0186000	.0163000	.0218700	.0117600	.0238600	.0141900	.0005000	-.002690	.0143500	.0097000	.0050800	.0042900	.0085100	.0000500	.0013300
235R11/d {1}		0.818879	0.025003	0.076736	0.003412	0.330261	0.000887	0.154488	0.888717	0.933087	0.16482	0.515613	0.875398	0.876572	0.595733	0.699811	0.930554
113/d {2}	0.818879		0.274597	0.477282	0.082272	0.776245	0.034013	0.594013	0.812742	0.855729	0.647391	0.871695	0.983948	0.971296	0.87276	0.900728	0.637456
67-4/d {3}	0.025003	0.274597		0.693475	0.575228	0.767363	0.639421	0.874242	0.081814	0.021138	0.74668	0.647948	0.284139	0.255726	0.596342	0.074849	0.105025
168-8/d {4}	0.076736	0.477282	0.693475		0.605712	0.864454	0.565754	0.930473	0.196187	0.067792	0.738264	0.790123	0.465035	0.442455	0.765546	0.18634	0.2342
135/d {5}	0.003412	0.082272	0.575228	0.605712		0.510067	0.733098	0.681178	0.015486	0.002709	0.570144	0.361201	0.094178	0.077591	0.299255	0.013431	0.02202
97/d {6}	0.330261	0.776245	0.767363	0.864454	0.510067		0.368488	0.677101	0.530298	0.318471	0.897122	0.724086	0.661736	0.703523	0.842904	0.538561	0.55663
191/d {7}	0.000887	0.034013	0.639421	0.565754	0.733098	0.368488		0.560381	0.004886	0.000687	0.478236	0.228309	0.042325	0.03271	0.173855	0.004091	0.007398
297-6/d {8}	0.154488	0.594013	0.874242	0.930473	0.681178	0.677101	0.560381		0.314398	0.143294	0.978143	0.721805	0.522512	0.534034	0.764645	0.311751	0.349264
112/m {9}	0.888717	0.812742	0.081814	0.196187	0.015486	0.530298	0.004886	0.314398		0.94748	0.341846	0.697457	0.935089	0.915776	0.743542	0.938538	0.886905
113/m {10}	0.933087	0.855729	0.021138	0.067792	0.002709	0.318471	0.000687	0.143294	0.94748		0.151111	0.51081	0.887027	0.895899	0.600218	0.885594	0.958955
67-4/m {11}	0.16482	0.647391	0.74668	0.738264	0.570144	0.897122	0.478236	0.978143	0.341846	0.151111		0.855893	0.606139	0.599834	0.85512	0.334242	0.385148
168-8/m {12}	0.515613	0.871695	0.647948	0.790123	0.361201	0.724086	0.228309	0.721805	0.697457	0.51081	0.855893		0.708132	0.790323	0.838425	0.717243	0.705985
135/m {13}	0.875398	0.983948	0.284139	0.465035	0.094178	0.661736	0.042325	0.522512	0.935089	0.887027	0.606139	0.708132		0.892322	0.556675	0.955432	0.918153
97/m {14}	0.876572	0.971296	0.255726	0.442455	0.077591	0.703523	0.03271	0.534034	0.915776	0.895899	0.599834	0.790323	0.892322		0.749752	0.950408	0.867813
2mel {15}	0.595733	0.87276	0.596342	0.765546	0.299255	0.842904	0.173855	0.764645	0.743542	0.600218	0.85512	0.838425	0.556675	0.749752		0.774485	0.733573
3mel {16}	0.699811	0.900728	0.074849	0.18634	0.013431	0.538561	0.004091	0.311751	0.938538	0.885594	0.334242	0.717243	0.955432	0.950408	0.774485		0.973835
mel+2vir {17}	0.930554	0.637456	0.105025	0.2342	0.02202	0.55663	0.007398	0.349264	0.886905	0.958955	0.385148	0.705985	0.918153	0.867813	0.733573	0.973835	

Appendix 7.2.

Newman-Keuls probability matrix based on the ANOVA for CPP1. Only transformant lines are included. Any significant probability ($p < 0.05$) is shaded in grey.

Bold numbers at the head of each column indicate each genotype's average CPP1 value.

	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}	{12}	{13}	{14}	{15}	{16}	{17}
	1.604700	1.597880	2.005700	2.008700	1.955070	1.635890	1.722760	1.966710	1.712700	1.752780	2.363983	2.302790	1.803220	1.840920	1.853100	1.742160	1.565850
235R11/d {1}		0.931151	0.000040	0.000041	0.000414	0.692724	0.440017	0.000244	0.357638	0.416944	0.000026	0.000023	0.153695	0.055857	0.043625	0.408344	0.875026
113/d {2}	0.931151		0.000036	0.000038	0.000321	0.880035	0.508858	0.000198	0.465156	0.438811	0.000029	0.000026	0.155162	0.053457	0.040373	0.447571	0.684887
67-4/d {3}	0.000040	0.000036		0.969684	0.797195	0.000157	0.010206	0.621319	0.007848	0.022951	0.000040	0.000504	0.105900	0.225310	0.214023	0.019070	0.000025
168-8/d {4}	0.000041	0.000038	0.969684		0.904935	0.000158	0.010894	0.855610	0.008181	0.026131	0.000040	0.000201	0.124830	0.273865	0.280047	0.020975	0.000028
135/d {5}	0.000414	0.000321	0.797195	0.904935		0.001740	0.050741	0.882770	0.044445	0.077360	0.000023	0.000117	0.217936	0.317221	0.196391	0.075691	0.000066
97/d {6}	0.692724	0.880035	0.000157	0.000158	0.001740		0.513681	0.001168	0.330481	0.574928	0.000023	0.000020	0.276779	0.126561	0.107442	0.533258	0.811439
191/d {7}	0.440017	0.508858	0.010206	0.010894	0.050741	0.513681		0.041857	0.898590	0.923367	0.000017	0.000015	0.738079	0.564398	0.564248	0.805863	0.349141
297-6/d {8}	0.000244	0.000198	0.621319	0.855610	0.882770	0.001168	0.041857		0.035040	0.073158	0.000021	0.000127	0.232551	0.382173	0.320663	0.066874	0.000046
112/m {9}	0.357638	0.465156	0.007848	0.008181	0.044445	0.330481	0.898590	0.035040		0.957269	0.000020	0.000017	0.781484	0.582205	0.562476	0.926091	0.338888
113/m {10}	0.416944	0.438811	0.022951	0.026131	0.077360	0.574928	0.923367	0.073158	0.957269		0.000012	0.000010	0.522793	0.503726	0.581467	0.892978	0.257022
67-4/m {11}	0.000026	0.000029	0.000040	0.000040	0.000023	0.000023	0.000017	0.000021	0.000020	0.000012		0.438174	0.000010	0.000032	0.000026	0.000015	0.000033
168-8/m {12}	0.000023	0.000026	0.000504	0.000201	0.000117	0.000020	0.000015	0.000127	0.000017	0.000010	0.438174		0.000032	0.000026	0.000020	0.000012	0.000029
135/m {13}	0.153695	0.155162	0.105900	0.124830	0.217936	0.276779	0.738079	0.232551	0.781484	0.522793	0.000010	0.000032		0.632907	0.802522	0.719262	0.065811
97/m {14}	0.055857	0.053457	0.225310	0.273865	0.317221	0.126561	0.564398	0.382173	0.582205	0.503726	0.000032	0.000026	0.632907		0.877373	0.594122	0.017716
2mel {15}	0.043625	0.040373	0.214023	0.280047	0.196391	0.107442	0.564248	0.320663	0.562476	0.581467	0.000026	0.000020	0.802522	0.877373		0.624065	0.012297
3mel {16}	0.408344	0.447571	0.019070	0.020975	0.075691	0.533258	0.805863	0.066874	0.926091	0.892978	0.000015	0.000012	0.719262	0.594122	0.624065		0.277270
mel+2vir {17}	0.875026	0.684887	0.000025	0.000028	0.000066	0.811439	0.349141	0.000046	0.338888	0.257022	0.000033	0.000029	0.065811	0.017716	0.012297	0.277270	

Appendix 7.3.

Newman-Keuls probability matrix based on the ANOVA for CPP10. Only transformant lines are included. Any significant probability ($p < 0.05$) is shaded in grey. Bold numbers indicate each genotype's average CPP10 value.

	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}	{12}	{13}	{14}	{15}	{16}	{17}
	1.584900	1.634600	2.173100	2.155400	2.151900	1.741730	1.937500	2.094420	1.717200	1.728570	2.493133	2.390090	1.848940	1.879530	1.929690	1.742610	1.577820
235R11/d {1}		0.557404	0.000023	0.000020	0.000017	0.344013	0.001318	0.000015	0.262349	0.325720	0.000029	0.000026	0.030215	0.011871	0.001562	0.425991	0.933398
113/d {2}	0.557404		0.000020	0.000017	0.000015	0.585529	0.010541	0.000015	0.329517	0.508286	0.000026	0.000023	0.115326	0.058789	0.011648	0.706631	0.780777
67-4/d {3}	0.000023	0.000020		0.834505	0.966076	0.000027	0.043108	0.789392	0.000021	0.000022	0.000478	0.010432	0.003268	0.009554	0.046756	0.000022	0.000026
168-8/d {4}	0.000020	0.000017	0.834505		0.967047	0.000045	0.049571	0.751722	0.000026	0.000032	0.000413	0.015468	0.005518	0.014358	0.059355	0.000059	0.000023
135/d {5}	0.000017	0.000015	0.966076	0.967047		0.000064	0.030574	0.497428	0.000024	0.000029	0.000549	0.025403	0.004713	0.011412	0.043239	0.000051	0.000020
97/d {6}	0.344013	0.585529	0.000027	0.000045	0.000064		0.189434	0.000635	0.954830	0.876555	0.000017	0.000015	0.414715	0.363467	0.172618	0.991718	0.380627
191/d {7}	0.001318	0.010541	0.043108	0.049571	0.030574	0.189434		0.063966	0.155500	0.171618	0.000026	0.000021	0.722641	0.772644	0.926549	0.144569	0.001124
297-6/d {8}	0.000015	0.000015	0.789392	0.751722	0.497428	0.000635	0.063966		0.000316	0.000435	0.000055	0.004414	0.030821	0.054442	0.126315	0.000481	0.000017
112/m {9}	0.262349	0.329517	0.000021	0.000026	0.000024	0.954830	0.155500	0.000316		0.893237	0.000023	0.000020	0.526403	0.391991	0.156081	0.990629	0.353101
113/m {10}	0.325720	0.508286	0.000022	0.000032	0.000029	0.876555	0.171618	0.000435	0.893237		0.000020	0.000017	0.486142	0.384033	0.165289	0.984977	0.385502
67-4/m {11}	0.000029	0.000026	0.000478	0.000413	0.000549	0.000017	0.000026	0.000055	0.000023	0.000020		0.223827	0.000012	0.000010	0.000032	0.000015	0.000033
168-8/m {12}	0.000026	0.000023	0.010432	0.015468	0.025403	0.000015	0.000021	0.004414	0.000020	0.000017	0.223827		0.000010	0.000032	0.000026	0.000012	0.000029
135/m {13}	0.030215	0.115326	0.003268	0.005518	0.004713	0.414715	0.722641	0.030821	0.526403	0.486142	0.000012	0.000010		0.718026	0.606517	0.209402	0.029845
97/m {14}	0.011871	0.058789	0.009554	0.014358	0.011412	0.363467	0.772644	0.054442	0.391991	0.384033	0.000010	0.000032	0.718026		0.553762	0.238671	0.011097
2mel {15}	0.001562	0.011648	0.046756	0.059355	0.043239	0.172618	0.926549	0.126315	0.156081	0.165289	0.000032	0.000026	0.606517	0.553762		0.120883	0.001366
3mel {16}	0.425991	0.706631	0.000022	0.000059	0.000051	0.991718	0.144569	0.000481	0.990629	0.984977	0.000015	0.000012	0.209402	0.238671	0.120883		0.450081
mel+2vir {17}	0.933398	0.780777	0.000026	0.000023	0.000020	0.380627	0.001124	0.000017	0.353101	0.385502	0.000033	0.000029	0.029845	0.011097	0.001366	0.450081	

Appendix 7.4.

Kolmogorov-Smirnov tests. **A:** hemizygous *nonA*⁽⁻⁾[*nonA*^{vir}] and heterozygous *nonA*^{vir}/*nonA*⁺ lines are respectively grouped as *vir/d* and *vir/m*. **B:** each transformant line is compared to 235R11/*d*, *diss* and *vir*. Maximum differences between pair of genotypes are indicated. * = significant differences at $p < 0.05$. ** = significant differences at $p < 0.01$.

A:

	235R11/ <i>d</i>	<i>vir/d</i>	<i>vir/m</i>	<i>vir</i>
<i>diss</i>	-0.900**	-0.843**	-0.893**	-0.800**
235R11/ <i>d</i>		-0.486*	-0.429	-0.700*
<i>vir/d</i>			-0.214	-0.400
<i>vir/m</i>				-0.518*

B:

	235R11/ <i>d</i>	<i>diss</i>	<i>vir</i>
113/ <i>d</i>	-0.40	0.90**	0.50
67-4/ <i>d</i>	-0.60	0.80**	0.50
168-8/ <i>d</i>	-0.50	0.90**	0.60
135/ <i>d</i>	-0.50	1.00**	0.60
97/ <i>d</i>	-0.50	0.90**	0.40
191/ <i>d</i>	-0.60	0.80**	0.40
297-6/ <i>d</i>	-0.70*	0.70*	0.20
112/ <i>m</i>	-0.30	1.00**	0.90**
113/ <i>m</i>	-0.30	1.00**	0.80**
67-4/ <i>m</i>	-0.50	0.83*	0.43
168-8/ <i>m</i>	-0.50	0.90**	0.50
135/ <i>m</i>	-0.70*	0.90**	0.50
97/ <i>m</i>	-0.60	0.70*	0.40
2 <i>mel</i>	0.50	-1.00**	-0.60
3 <i>mel</i>	0.40	-1.00**	-0.80**
<i>mel</i> +2 <i>vir</i>	-0.40	-1.00**	-1.00**

Appendix 7.5.

Newman-Keuls probability matrix based on the ANOVA for bPF. Any significant probability ($p < 0.05$) is shaded in grey. Bold numbers at the head of each column indicate each genotype's average bPF value.

	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}	{12}	{13}	{14}	{15}	{16}	{17}	{18}	{19}
	.1551200	4.10710	3.61327	4.10402	4.08334	5.04059	5.85084	6.84139	6.57197	2.78128	3.46384	1.52972	2.29338	4.10209	2.98760	3.33644	3.51292	4.72546	6.30480
diss {1}		0.000207	0.001207	0.000182	0.000144	0.000030	0.000033	0.000039	0.000036	0.009836	0.001526	0.147308	0.033234	0.000156	0.006461	0.002051	0.001550	0.000031	0.383134
235R11/d {2}	0.000207		0.982186	0.997274	0.999993	0.553886	0.213007	0.028989	0.048832	0.903400	0.991800	0.154945	0.639455	0.999983	0.946986	0.989742	0.986218	0.492410	0.007434
113/d {3}	0.001207	0.982186		0.947972	0.601770	0.692340	0.202092	0.012529	0.028375	0.940572	0.984950	0.286199	0.765530	0.850218	0.957728	0.989938	0.911303	0.819852	0.026024
67-4/d {4}	0.000182	0.997274	0.947972		0.999710	0.726035	0.296451	0.038314	0.067640	0.870276	0.980729	0.136733	0.591797	0.998292	0.920203	0.979281	0.965528	0.769394	0.006463
168-8/d {5}	0.000144	0.999993	0.601770	0.999710		0.896138	0.438991	0.056235	0.104497	0.776912	0.901816	0.105273	0.490549	0.983406	0.829018	0.921666	0.801790	0.953643	0.004959
135/d {6}	0.000030	0.553886	0.692340	0.726035	0.896138		0.368387	0.188211	0.205125	0.334485	0.714908	0.007393	0.107628	0.835856	0.448277	0.674701	0.689989	0.726463	0.000118
97/d {7}	0.000033	0.213007	0.202092	0.296451	0.438991	0.368387		0.514251	0.423382	0.036968	0.194550	0.000179	0.006043	0.376720	0.065474	0.161124	0.188261	0.424098	0.000030
191/d {8}	0.000039	0.028989	0.012529	0.038314	0.056235	0.188211	0.514251		0.764870	0.000648	0.009676	0.000033	0.000076	0.048531	0.001551	0.006585	0.010033	0.129657	0.000036
297-6/d {9}	0.000036	0.048832	0.028375	0.067640	0.104497	0.205125	0.423382	0.764870		0.002092	0.023759	0.000031	0.000219	0.088086	0.004659	0.017145	0.023929	0.169858	0.000033
112/m {10}	0.009836	0.903400	0.940572	0.870276	0.776912	0.334485	0.036968	0.000648	0.002092		0.873365	0.346410	0.588059	0.825472	0.818846	0.811187	0.926980	0.534856	0.079403
113/m {11}	0.001526	0.991800	0.984950	0.980729	0.901816	0.714908	0.194550	0.009676	0.023759	0.873365		0.263034	0.691596	0.954626	0.857253	0.887536	0.956552	0.857369	0.027648
67-4/m {12}	0.147308	0.154945	0.286199	0.136733	0.105273	0.007393	0.000179	0.000076	0.000219	0.588059	0.691596		0.396558	0.118293	0.368081	0.263086	0.294403	0.023151	0.318130
168-8/m {13}	0.033234	0.639455	0.765530	0.591797	0.490549	0.107628	0.006043	0.000076	0.000219	0.588059	0.691596	0.396558		0.537667	0.721033	0.653420	0.754585	0.225960	0.154739
135/m {14}	0.000156	0.999983	0.850218	0.998292	0.983406	0.835856	0.376720	0.048531	0.088086	0.825472	0.954626	0.118293	0.537667		0.879673	0.957991	0.914161	0.900178	0.005509
97/m {15}	0.006461	0.946986	0.957728	0.920203	0.829018	0.448277	0.065474	0.001551	0.004659	0.818846	0.857253	0.368081	0.721033	0.879673		0.698555	0.937185	0.648881	0.067320
2mel {16}	0.002051	0.989742	0.989938	0.979281	0.921666	0.674701	0.161124	0.006585	0.017145	0.811187	0.887536	0.263086	0.653420	0.957991	0.698555		0.979070	0.835652	0.031885
3mel {17}	0.001550	0.986218	0.911303	0.965528	0.801790	0.689989	0.188261	0.010033	0.023929	0.926980	0.956552	0.294403	0.754585	0.914161	0.937185	0.979070		0.830059	0.029896
mel+2vir {18}	0.000031	0.492410	0.819852	0.769394	0.953643	0.726463	0.424098	0.129657	0.169858	0.534856	0.857369	0.023151	0.225960	0.900178	0.648881	0.835652	0.830059		0.000476
vir {19}	0.383134	0.007434	0.026024	0.006463	0.004959	0.000118	0.000030	0.000036	0.000033	0.079403	0.027648	0.318130	0.154739	0.005509	0.067320	0.031885	0.029896	0.000476	

Appendix 7.6.

Newman-Keuls probability matrix based on the ANOVA for SSF. Any significant probability ($p < 0.05$) is shaded in grey. Bold numbers at the head of each column indicate each genotype's average SSF value.

	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}	{12}	{13}	{14}	{15}	{16}	{17}	{18}
	142.9000	145.1000	144.4000	151.3000	161.0000	167.7000	153.9000	144.1000	155.6000	127.4000	130.7000	163.1667	150.1000	153.2000	164.7000	158.1000	142.6000	140.8889
diss {1}		0.967704	0.947238	0.494456	0.007203	0.000040	0.293584	0.801931	0.164222	0.010498	0.052509	0.001386	0.559029	0.321359	0.000390	0.047922	0.949996	0.907173
235R11/d {2}	0.967704		0.883663	0.397308	0.020032	0.000132	0.350587	0.976204	0.239874	0.005298	0.041705	0.005004	0.295884	0.327112	0.001724	0.093791	0.985101	0.951297
113/d {3}	0.947238	0.883663		0.472710	0.015277	0.000084	0.350248	0.949996	0.224209	0.006977	0.048010	0.003486	0.458138	0.350587	0.001134	0.080007	0.981849	0.948598
67-4/d {4}	0.494456	0.397308	0.472710		0.326314	0.017627	0.849770	0.559029	0.805374	0.000036	0.000583	0.166158	0.801931	0.691210	0.094548	0.613648	0.535168	0.365976
168-8/d {5}	0.007203	0.020032	0.015277	0.326314		0.498851	0.446974	0.014995	0.496193	0.000026	0.000023	0.650576	0.254249	0.477619	0.719309	0.544335	0.006704	0.001859
135/d {6}	0.000040	0.000132	0.000084	0.017627	0.498851		0.059844	0.000077	0.115505	0.000036	0.000033	0.610016	0.008858	0.049936	0.530544	0.262505	0.000040	0.000031
97/d {7}	0.293584	0.350587	0.350248	0.849770	0.446974	0.059844		0.383734	0.722299	0.000019	0.000077	0.297480	0.857011	0.883663	0.211595	0.654182	0.304906	0.165418
191/d {8}	0.801931	0.976204	0.949996	0.559029	0.014995	0.000077	0.383734		0.239080	0.006407	0.040725	0.003276	0.592125	0.400519	0.001011	0.082384	0.947238	0.907950
297-6/d {9}	0.164222	0.239874	0.224209	0.805374	0.496193	0.115505	0.722299	0.239080		0.000020	0.000029	0.388966	0.779837	0.870471	0.315911	0.601219	0.166329	0.076256
112/m {10}	0.010498	0.005298	0.006977	0.000036	0.000026	0.000036	0.000019	0.006407	0.000020		0.490262	0.000029	0.000106	0.000018	0.000033	0.000023	0.000094	0.013327
113/m {11}	0.052509	0.041705	0.048010	0.000583	0.000023	0.000033	0.000077	0.040725	0.000029	0.490262		0.000026	0.001306	0.000119	0.000029	0.000021	0.034355	0.033169
67-4/m {12}	0.001386	0.005004	0.003486	0.166158	0.650576	0.610016	0.297480	0.003276	0.388966	0.000029	0.000026		0.113097		0.000029	0.000021	0.000288	
168-8/m {13}	0.559029	0.295884	0.458138	0.801931	0.254249	0.008858	0.857011	0.592125	0.779837	0.000106	0.001306	0.113097		0.295783	0.748554	0.539487	0.001222	0.000288
135/m {14}	0.321359	0.327112	0.350587	0.691210	0.477619	0.049936	0.883663	0.400519	0.870471	0.000018	0.000119	0.295783	0.793453		0.196599	0.735134	0.341870	0.197616
97/m {15}	0.000390	0.001724	0.001134	0.094548	0.719309	0.530544	0.211595	0.001011	0.315911	0.000033	0.000029	0.748554	0.057836	0.196599		0.512044	0.000341	0.000086
2mel {16}	0.047922	0.093791	0.080007	0.613648	0.544335	0.262505	0.654182	0.082384	0.601219	0.000023	0.000021	0.539487	0.550109	0.735134	0.512044		0.046770	0.016777
3mel {17}	0.949996	0.985101	0.981849	0.535168	0.006704	0.000040	0.304906	0.947238	0.166329	0.008094	0.034355	0.001222	0.619819	0.341870	0.000341	0.046770		0.720560
mel+2vir {18}	0.907173	0.951297	0.948598	0.365976	0.001859	0.000031	0.165418	0.907950	0.076256	0.013327	0.033169	0.000288	0.463086	0.197616	0.000086	0.016777	0.720560	

Appendix 7.7.

Newman-Keuls probability matrix based on the ANOVA for IPI. Any significant probability ($p < 0.05$) is shaded in grey. Bold numbers at the head of each column indicate each genotype's average IPI value.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
	40.64680	36.66730	36.39270	34.50150	32.96240	33.83140	35.67480	36.01960	36.06460	40.23460	40.40280	37.07833	37.26490	38.18800	37.90400	38.81080	39.98780	38.36440	16.57010
diss {1}		0.001130	0.000363	0.000029	0.000036	0.000033	0.000036	0.000088	0.000093	0.898951	0.794686	0.005503	0.009412	0.119116	0.067591	0.286659	0.895978	0.144411	0.000039
235R11/d {2}	0.001130		0.769612	0.189826	0.001993	0.039966	0.827680	0.900638	0.796412	0.004496	0.002734	0.661087	0.799470	0.483171	0.550692	0.250515	0.009487	0.459075	0.000010
113/d {3}	0.000363	0.769612		0.257605	0.004738	0.069091	0.869897	0.916419	0.726382	0.001722	0.000986	0.744887	0.788582	0.392852	0.489610	0.163427	0.004009	0.350694	0.000032
67-4/d {4}	0.000029	0.189826	0.257605		0.228181	0.474775	0.210770	0.237454	0.341146	0.000023	0.000026	0.086605	0.063552	0.003368	0.008689	0.000279	0.000021	0.001898	0.000008
168-8/d {5}	0.000036	0.001993	0.004738	0.228181		0.353996	0.019946	0.009605	0.012062	0.000029	0.000033	0.000410	0.000199	0.000019	0.000021	0.000023	0.000026	0.000021	0.000009
135/d {6}	0.000033	0.039966	0.069091	0.474775	0.353996		0.120660	0.090426	0.120190	0.000026	0.000029	0.012501	0.007698	0.000186	0.000616	0.000027	0.000023	0.000098	0.000022
97/d {7}	0.000036	0.827680	0.869897	0.210770	0.019946	0.120660		0.713058	0.909130	0.000102	0.000060	0.666215	0.618741	0.154624	0.252251	0.033648	0.000274	0.114391	0.000017
191/d {8}	0.000088	0.900638	0.916419	0.237454	0.009605	0.090426	0.713058		0.961722	0.000438	0.000230	0.791079	0.769372	0.286340	0.408193	0.035631	0.001189	0.231241	0.000020
297-6/d {9}	0.000093	0.796412	0.726382	0.341146	0.012062	0.120190	0.909130	0.961722		0.000455	0.000243	0.701012	0.703437	0.261236	0.364447	0.081869	0.001202	0.215997	0.000026
112/m {10}	0.898951	0.004496	0.001722	0.000023	0.000029	0.000026	0.000102	0.000438	0.000455		0.857633	0.017391	0.025801	0.186092	0.128189	0.282097	0.792384	0.189872	0.000033
113/m {11}	0.794686	0.002734	0.000986	0.000026	0.000033	0.000029	0.000060	0.000230	0.000243	0.857633		0.011734	0.018568	0.169650	0.107178	0.324644	0.897647	0.189481	0.000036
67-4/m {12}	0.005503	0.661087	0.744887	0.086605	0.000410	0.012501	0.666215	0.791079	0.701012	0.017391	0.011734		0.842282	0.637247	0.652531	0.434812	0.031518	0.645870	0.000012
168-8/m {13}	0.009412	0.799470	0.788582	0.063552	0.000199	0.007698	0.618741	0.769372	0.703437	0.025801	0.018568	0.842282		0.586622	0.495450	0.466013	0.042800	0.644109	0.000015
135/m {14}	0.119116	0.483171	0.392852	0.003368	0.000019	0.000186	0.154624	0.286340	0.261236	0.186092	0.169650	0.637247	0.586622		0.761960	0.784223	0.219670	0.850772	0.000020
97/m {15}	0.067591	0.550692	0.489610	0.008689	0.000021	0.000616	0.252251	0.408193	0.364447	0.128189	0.107178	0.652531	0.495450	0.761960		0.768127	0.171251	0.875567	0.000017
2mel {16}	0.286659	0.250515	0.163427	0.000279	0.000023	0.000027	0.033648	0.085631	0.081869	0.282097	0.324644	0.434812	0.466013	0.784223	0.768127		0.209334	0.633979	0.000026
3mel {17}	0.895978	0.009487	0.004009	0.000021	0.000026	0.000023	0.000274	0.001189	0.001202	0.792384	0.897647	0.031518	0.042800	0.219670	0.171251	0.209334		0.193413	0.000029
mel+2vir {18}	0.144411	0.459075	0.350694	0.001898	0.000021	0.000098	0.114391	0.231241	0.215997	0.189872	0.189481	0.645870	0.644109	0.850772	0.875567	0.633979	0.193413		0.000023
vir {19}	0.000039	0.000010	0.000032	0.000008	0.000009	0.000022	0.000017	0.000020	0.000026	0.000033	0.000036	0.000012	0.000015	0.000020	0.000017	0.000026	0.000029	0.000023	

Appendix 7.8. Results of the discriminant analysis for the *MEL* group.

MEL		
Traits	F (2, 27)	P
IPI	13.7	<<0.0001
PF	9.58	0.0007
CPP	4.82	0.016
bCCP	4.18	0.026
bPF	0.78	0.47
SSF	8.32	0.0015

λ	0.107
F	7.53
d.f.	12, 44
P	<<0.001

Appendix 7.9. Results of the discriminant analysis for the *vir/d* group.

<i>vir/d</i>		
Traits	F (6, 63)	P
IPI	3.4	0.0057
PF	6.55	<<0.0001
CPP	10.11	<<0.0001
bCCP	1.71	0.133
bPF	2.75	0.0195
SSF	7.56	<<0.0001

λ	0.089
F	5.25
d.f.	36, 257
P	<<0.001

Appendix 7.10. Results of the discriminant analysis for the hybrids (*vir/m* and *mel+2vir*).

HYB		
Traits	F (6, 58)	P
IPI	4.21	0.0014
PF	23.29	<<0.0001
CPP	23.20	<<0.0001
bCCP	2.57	0.028
bPF	3.89	0.003
SSF	18.22	<<0.0001

λ	0.027
F	8.34
d.f.	36, 235
P	<<0.0001
P	<<0.001

Appendix 7.11. Results of the discriminant analysis for all the transformants (excluding *mel+2vir*).

ALL (except <i>mel+2vir</i>)		
Traits	F (4, 94)	P
IPI	15.08	<<0.0001
PF	7.46	<0.0001
CPP	2.26	0.0680
bCCP	7.67	<0.0001
bPF	9.64	<0.0001
SSF	3.29	0.014

λ	0.274
F	5.82
d.f.	24, 311
P	<<0.0001

REFERENCES

- ACKERMAN, S.L., and SIEGEL, R.W. (1986) Chemically reinforced conditioned courtship in *Drosophila*: responses of wild-type and the *dunce*, *amnesiac* and *don giovanni* mutants. *J. Neurogenet.* 3: 111-123.
- ADAM, S.A., NAKAGAWA, T., SWANSON, M.S., WOODRUFF, T.K., and DREYFUSS, G. (1986) m-RNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell. Biol.* 6: 2932-294.
- ALAWI, A.A., and PAK, W.L. (1971) On-transient of insect electroretinogram: Its cellular origin. *Science* 172: 1055-1057.
- ALBAGLI, O., DHORDAIN, P., DEWEINDT, C., LECOCQ, G., and LEPRINCE, D. (1995) The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth Differ.* 6: 1193-1198.
- ALEXANDER, R.D. (1962) Evolutionary change in cricket acoustical communication. *Evolution* 16: 443-467.
- ALT, S., RINGO, J., TALYN, B., BRAY, W., and DOWSE, H. (1998) The *period* gene controls courtship song cycles in *D.melanogaster*. *Anim. Behav.* In press.
- AMREIN, H., GORMAN, M., and NOTHIGER, R. (1988) The sex-determining gene *tra-2* of drosophila encodes a putative RNA binding protein. *Cell* 55: 1025-1035.
- AMREIN, H., MANIATIS, T., and NOTHIGER, R. (1990) Alternatively spliced transcripts of the sex-determining gene *tra-2* of *Drosophila* encode functional proteins of different size. *EMBO J.* 9: 3619-3629.
- ANDERSON, T.W. (1963) An introduction to multivariate statistical analysis. N.Y. Wiley & Sons.
- ASHBURNER, M. (1989) *Drosophila*. A laboratory handbook. Cold Spring Harbor Laboratory Press.
- AUTRUM, H., and HOFFMAN, C. (1957) Die Wirkung von Pikrotoxin und Nikotin auf das Retinogramm von Insekten. *Z. Naturforsch.* 12B: 752-757.
- AUTRUM, H., and HOFFMAN, C. (1960) Diphasic and monophasic responses in the compound eye of *Calliphora*. *J. Insect Physiol.* 4: 122-127.
- BACHMANN, B., LUKE, W., and HUNSMANN, G. (1990) Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Res.* 18: 1309.
- BAIROCH, A., and BUCHER, P. (1994) Prosite: recent developments. *Nucleic Acids Res.* 22; 3583-3589.
- BAKER, B.S., and RIDGE, K.A. (1980) Sex and the single cell. On the action of major foci affecting sex-determination in *Drosophila melanogaster*. *Genetics* 94: 383-423.

- BAKER, B.S., and WOLFNER, M.F. (1988) A molecular analysis of *doublesex*, a bifunctional gene that controls both male and female sexual-differentiation in *Drosophila melanogaster*. *Genes Dev.* 2: 477-489.
- BAKER, B.S. (1989) Sex in flies: the splice of life. *Nature* 340: 521-524.
- BANDZIULIS, R.J., SWANSON, M.S., and DREYFUSS, G. (1989) RNA-binding proteins as developmental regulators. *Genes Dev.* 3: 431-437.
- BARGIELLO, T.A., and YOUNG, M.W., (1984) Molecular genetics of a biological clock in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 81: 2142-2146.
- BARGIELLO, T.A., JACKSON, F.R., and YOUNG, M.W. (1984) Restoration of circadian behavioral rhythms by gene-transfer in *Drosophila*. *Nature* 312: 752-754.
- BAYLIES, M.K., BARGIELLO, T.A., JACKSON, F.R., YOUNG, M.W. (1987) Changes in abundance or structure of the *per* gene product can alter periodicity of the *Drosophila* clock. *Nature* 326: 390-392.
- BELL, L.R., MAINE, E.M., SCHEDL, P., and CLINE, T.W. (1988) *Sex-lethal*, a *Drosophila* sex-determination gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* 5: 1037-1046.
- BELL, L.R., HORABIN, J.I., SCHEDL, P., and CLINE, T.W. (1991) Positive autoregulation of *Sex-lethal* by alternative splicing maintains the female determined state in *Drosophila*. *Cell* 65: 229-240.
- BELLEN, H.J., KOOYER, S., D'EVELYN, D., and PEARLMAN, J. (1992) The *Drosophila* couch potato protein is expressed in nuclei of peripheral neuronal precursors and shows homology to RNA-binding proteins. *Genes Dev.* 6: 2125-2136.
- BELOTE, J.M., and BAKER, B.S. (1982) Sex determination in *Drosophila melanogaster*: analysis of *transformer-2*, a sex-transforming locus. *Proc. Natl. Acad. Sci. USA* 79: 1568-1572.
- BELOTE, J.M., and BAKER, B.S. (1987) Sexual-behavior: its genetic control during development and adulthood in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 84: 8026-8030.
- BELOTE, J.M., McKEOWN, M., BOGGS, R.T., OHKAWA, R., SOSNOWSKI, B.A. (1989) Molecular genetics of *transformer*, a genetic switch controlling sexual differentiation in *Drosophila*. *Dev. Genet.* 10: 143-154.
- BELOTE, J.M. (1992) Sex determination in *Drosophila melanogaster*: from the X:A ratio to *doublesex*. *Dev. Biol.* 3: 319-330.
- BENNET-CLARK, H.C., and EWING, A.W. (1969) Pulse interval as a critical parameter in the courtship sound of *Drosophila melanogaster*. *Anim. Behav.* 17: 755-759.
- BENZER, S. (1967) Behavioral mutants of *Drosophila* isolated by countercurrent distribution. *Proc. Natl. Acad. Sci. USA* 58: 1112-1119.

- BENZER, S. (1971) From the gene to behavior. *J. Amer. Med. Assoc.* 218: 1015-1022.
- BENZER, S. (1973) Genetic dissection of behavior. *Sci. Amer.* 229: 24-37.
- BERNSTEIN, A.S., NEUMANN, E.K., and HALL, J.C. (1992) Temporal analysis of tone pulses within the courtship songs of 2 sibling *Drosophila* species, their interspecific hybrid, and behavioral mutants of *D.melanogaster* (Diptera: Drosophilidae). *J. Ins. Behav.* 5: 15-36.
- BESSER, H.V., SCHNABEL, P., WIELAND, C., FRITZ, E., STANEWSKY, R., and SAUMWEBER, H. (1990) The puff-specific *Drosophila* protein Bj6, encoded by the gene *no-on-transientA*, shows homology to RNA-binding proteins. *Chromosoma* 100: 37-47.
- BEVERLEY, S.M., and WILSON, A.C. (1984) Molecular evolution in *Drosophila* and the higher diptera. A time scale for fly evolution. *J. Mol. Evol.* 21: 1-13.
- BEYER, A.L., and OSHEIN, Y.N (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2: 754-765.
- BLONDEAU, J., and HEISENBERG, M. (1982) The three-dimensional optomotor torque system of *Drosophila melanogaster*. *J. Comp. Physiol.* 145: 321-329.
- BOGGS, R.T., GREGOR, P., IDRIS, S., BELOTE, J.M., and MCKEOWN, M. (1987) Regulation of sexual-differentiation in *Drosophila melanogaster* via alternative splicing of RNA from the *transformer* gene. *Cell* 50: 739-747.
- BRAND, A.H., and PERRIMON, N. (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118: 401-415.
- BRETHNACH, R., and CHAMBON, P. (1981) Organization and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50: 349-383.
- BRITTEN, R.J. (1986) Rates of dna-sequence evolution differ between taxonomic groups. *Science* 231: 1393-1398.
- BRUNNER, A., WOLF, R., PFLUGFELDER, G.O., POECK, B., and HEISENBERG, M. (1992) Mutations in the proximal region of the *optomotor-blind* locus of *Drosophila melanogaster* reveal a gradient of neuroanatomical and behavioral phenotypes. *J. Neurogenetics* 8: 43-55.
- BULMER, M. (1988) Are codon usage patterns in unicellular organisms determined by selection-mutation balance? *J. Evol. Biol.* 1: 15-26.
- BURG, M.G., SARTHY, P.V., KOLIANITZ, G., and PAK, W.L. (1993) Genetic and molecular identification of a *Drosophila* histidine decarboxylase gene required in photoreceptor transmitter synthesis. *EMBO J.* 12: 911-919.

- BURKHARDT, D., and AUTRUM, H. (1960) Die Belichtungspotentiale Einzelner Sehzellen von *Calliphora erythrocephala* Meig. *Z. Naturforsch.* 15B 612-616.
- BURNET, B., and BECK, J. (1968) Phenogenetic studies on visual acuity in *Drosophila melanogaster*. *J. Insect Physiol.* 14: 855-860.
- BURTIS, K.C., and BAKER, B.S. (1989) *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced messenger-RNAs encoding related sex-specific polypeptides. *Cell* 56: 997-1010.
- BURTIS, K.C. (1993) The regulation of sex determination and sexually dimorphic differentiation in *Drosophila*. *Curr. Opin. Cell. Biol.* 5: 1006-1014.
- BUTLIN, R.K. and RITCHIE, M.G. (1989) Genetic coupling in mate recognition systems: what is the evidence? *Biological Journal of the Linnean Society* 37: 237-246.
- BYERS, D., DAVIS, R.L., and KIGER, J.A. (1981) Defect in cAMP phosphodiesterase due to the *dunce* mutation of learning in *Drosophila melanogaster*. *Nature* 289: 79-81.
- CARIOU, M.-L. (1987) Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D.sechellia* and *D.orena*. *Genet. Res. Camb.* 50: 181-185.
- CARLSON, J. (1991) Olfaction in *Drosophila*: genetic and molecular analysis. *Trends Neurosci.* 14: 520-524.
- CASTRILLON, D.H., GÖNCZY, S., ALEXANDER, R., RAWSON, C.G., EBERHART, C.G., VISWANATHAN, S., DINARDO, S., and WASSERMAN, S.A. (1993) Toward a molecular-genetic analysis of spermatogenesis in *Drosophila melanogaster*: characterization of male-sterile mutants generated by single *P* element mutagenesis. *Genetics* 135: 489-505.
- CHAMPLIN, D.T., FRASCH, M., SAUMWEBER, H., and LIS, J.T. (1991) Characterization of a *Drosophila* protein associated with boundaries of transcriptionally active chromatin. *Genes Dev.* 5: 1611-1621.
- CHANG, H.H., and MILLER, D.D. (1978) Courtship and mating sounds in species of the *Drosophila affinis* subgroup. *Evolution* 32: 540-550.
- CHOU, P.Y., and FASMAN, U.D. (1974) Prediction of protein conformation. *Biochemistry* 13: 211-215.
- CLINE, T.W. (1984) Autoregulatory functioning of a *Drosophila* gene-product that establishes and maintains the sexually determined state. *Genetics* 107: 231-277.
- COLOT, H.V., HALL, J.C., and ROSBASH, M. (1988) Interspecific comparison of the *period* gene reveals large blocks of non-conserved coding DNA. *EMBO J.* 7: 3929-3937.

- CONNOLLY, K., and COOK, R. (1973) Rejection responses by female *Drosophila melanogaster*: their ontogeny, causality and effects upon the behaviour of the courting male. *Behaviour* 52:142-166.
- COOMBE, P.E. (1986) The large monopolar cells L1 and L2 are responsible for ERG transients in *Drosophila*. *J. Comp. Physiol.* 159: 655-665.
- COOMBE, P.E., and HEISENBERG, M. (1986) The structural brain mutant *Vacuolar medulla* of *Drosophila melanogaster* with specific behavioral defects and cell degeneration in the adult. *J. Neurogenet.* 3: 135-158.
- COOPER, M.K., HAMBLIN-COYLE, M.J., LIU, X., RUTILA, J.E., and HALL, J.C. (1994) Dosage compensation of the *period* gene in *Drosophila melanogaster*. *Genetics* 138: 721-732.
- COWLING, D.E., and BURNET, B. (1981) Courtship songs and genetic-control of their acoustic characteristics in sibling species of the *Drosophila melanogaster* subgroup. *Anim. Behav.* 29: 924-935.
- COYNE, J.A. (1992) Genetics and speciation. *Nature* 355: 511-515.
- CURRIE, D.A., and BATE, M. (1995) Innervation is essential for the development and differentiation of a sex-specific adult muscle in *Drosophila melanogaster*. *Development* 121: 2549-2557.
- DANIELS, S.B., and CHOVNICK, A. (1993) *P* element transposition in *Drosophila melanogaster*: an analysis of sister-chromatid pairs and the formation of intragenic secondary insertions during meiosis. *Genetics* 133: 623-636.
- DAVIS, R.L. (1996) Physiology and biochemistry of *Drosophila* learning mutants. *Physiological Reviews* 76: 299-317.
- DAVIS, R.L., and KIGER, J.A. (1981) *dunce* mutants of *Drosophila melanogaster*: mutants defective in the cAMP phosphodiesterase enzyme-system. *J. Cell Biol.* 90: 101-107.
- deBELLE, J.S., and HEISENBERG, M. (1994) Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies. *Science* 263: 692-695.
- DEMETRIADES, M.K. (1997) Behavioural genetic analysis of biorhythms in the *melanogaster* subgroup of *Drosophila*. Ph.D. Thesis.
- DEVEREUX, J., HAEBERLI, P., SMITHIES, O. (1984) A comprehensive set of sequence-analysis programs for the VAX. *Nucleic Acids Res.* 12: 387-395.
- DOBZHANSKY, T. (1940) Speciation as a stage in evolutionary divergence. *Amer. Natur.* 74: 312-321.
- DONG, B.H., HOROWITZ, D.S., KOBAYASHI, R., and KRAINER, A.R. (1993) Purification and cDNA cloning of hela-cell p54^{nrb}, a nuclear-protein with two RNA recognition motifs and extensive homology to human splicing factor PSF and *Drosophila* NONA/Bj6. *Nucleic Acids Res.* 21: 4085-4092.

- DOVER, G.A., STRACHAN, T., and COEN, E.S. (1982) Molecular drive. *Science* 218: 1069.
- DOVER, G.A. (1983) Molecular drive. *New Scientist* 98: 808.
- DREYFUSS, G., MATUNIS, M.J., PINOL-ROMA, S., and BURD, C.G. (1993) hnRNP proteins and the biogenesis of m-RNA. *Annu. Rev. Biochem.* 62: 289-321.
- DUDAI, Y. (1985) Genes, enzymes and learning in *Drosophila*. *Trends in Neurosci.* 8:18-21.
- DVORAK, D.R., BISHOP, L.G, and ECKERT, H.E. (1975) On the identification of movement detectors in the fly optic lobe. *J. Comp. Physiol.* 100: 5-23.
- EICHENBAUM, D.M., and GOLDSMITH, T.H. (1968) Properties of insect photoreceptor cells lacking synapses. *J. Exp. Zool.* 169: 15-32.
- ENGELS, W. (1989) *P* element in *Drosophila melanogaster*. pp437-484 in *Mobile DNA*, edited by D. BERG and M. HOWE. American Society for Microbiology, Washington, D.C.
- ENGELS, W.R., JOHNSON-SCHLITZ, D.M., EGGLESTON, W.B., and SVED, J. (1990) High-frequency *P* element loss in *Drosophila* is homolog dependent. *Cell* 62: 515-525.
- EWER, J., ROSBASH, M., and HALL, J.C. (1988) An inducible promoter fused to the *period* gene in *Drosophila* conditionally rescues adult *per* mutant arrhythmicity. *Nature* 333: 82-84.
- EWER, J., HAMBLENCOLE, M., ROSBASH, M., and HALL, J.C. (1990) Requirement for *period* gene expression in the adult and not during development for locomotor activity rhythms of imaginal *Drosophila melanogaster*. *J. Neurogen.* 7: 31.
- EWING, A.W., and BENNET-CARK, H.C. (1968) The courtship songs of *Drosophila*. *Behaviour* 31: 288-301.
- EWING, A.W. (1969) The genetic basis of sound production in *Drosophila pseudoobscura* and *D.persimilis*. *Anim. Behav.* 17: 555-560.
- EWING, A.W. (1970) The evolution of courtship song in *Drosophila*. *Rev. Comp. Animal.-T.* 4:3-8.
- EWING, A.W. (1979) Complex courtship songs in the *Drosophila funebris* species group: escape from an evolutionary bottleneck. *Anim. Behav.* 27: 343-349.
- EWING, A.W., and MIYAN, J.A. (1986) Sexual selection, sexual isolation and the evolution of song in the *Drosophila repleta* group of species. *Anim. Behav.* 34: 421-429.
- EWING, L.S., and EWING, A.W. (1987) Courtship of *Drosophila melanogaster* in large observation chambers: the influence of female reproductive state. *Behaviour* 101: 243-252.
- FEANY, M.B., and QUINN, W.G. (1995) A neuropeptide gene defined by the *Drosophila* memory mutant *amnesiac*. *Science* 268: 869-873.

- FERNANDES, M., XIAO, H., LIS, J.T. (1993) Fine-structure analyses of the *Drosophila* and *Saccharomyces* heat-shock factor: heat-shock element interactions. *Nucleic Acids Res.* 22: 167-173.
- FERVEUR, J.F., and JALLON, J.M. (1993) *Nerd*, a locus on chromosome III, affects male reproductive-behavior in *Drosophila melanogaster*. *Naturwissenschaften* 80: 474-475.
- FERVEUR, J.F., STORTKUHL, K.F., STOCKER, R.F., and GREENSPAN, R.J. (1995) Genetic feminization of brain structures and changed sexual orientation in male *Drosophila*. *Science* 267: 902-905.
- FERVEUR, J.F., and JALLON, J.M. (1996) Genetic control of male cuticular hydrocarbons in *Drosophila melanogaster*. *Genetical Research* 67: 211-218.
- FERVEUR, J.F., and SUREAU, G. (1996) Simultaneous influence on male courtship of stimulatory and inhibitory pheromones produced by live sex-mosaic *Drosophila melanogaster*. *Pr. R. Soc. London series B-Biol Sci.* 263: 967-973.
- FERVEUR, J.F., COBB, M., BOUKELLA, H., and JALLON, J.M. (1996) World-wide variation in *Drosophila melanogaster* sex-pheromone: behavioral-effects, genetic bases and potential evolutionary consequences. *Genetica* 97: 73-80.
- FINLEY, K.D., TAYLOR, B.J., MILSTEIN, M., and McKEOWN, M. (1997) *dissatisfaction*, a gene involved in sex-specific behavior and neural development of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 94: 913-918.
- FISHER, R.A. (1930) Genetical theory of natural selection. *Clarendon Press*, Oxford.
- FOLKERS, E. (1982) Visual learning and memory of *Drosophila melanogaster* wild type and the mutants *dunce*, *amnesiac*, *turnip* and *rutabaga*. *J. Ins. Physiol.* 28: 535-539.
- FRASCH, M., and SAUMWEBER, H. (1989) Two proteins from *Drosophila* nuclei are bound to chromatin and are detected in a series of puffs on polytene chromosomes. *Chromosoma* 97: 272-281.
- FRISCHAUF, A.M., LEHRACH, H., POUSTKA, A., and MURRAY, N. (1983) Lambda-replacement vectors carrying polylinker sequences. *J. Mol. Biol.* 170: 827-842.
- FU, X.D., MAYEDA, A., MANIATIS, T., and KRAINER, A.R. (1992) General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5' and 3' splice site selection. *Proc. Natl. Acad. Sci. USA* 89: 11224-11228.
- GAILEY, D.A., JACKSON, F.R., and SIEGEL, R.W. (1982) Male courtship in *Drosophila*: the conditioned response to immature males and its genetic control. *Genetics* 102: 771-782.

- GAILEY, D.A., JACKSON, F.R., and SIEGEL, R.W. (1984) Conditioning mutations in *Drosophila melanogaster* affect experience-dependent behavioural modification in courting males. *Genetics* 106: 613-623.
- GAILEY, D.A., LACAILLADE, R.C., and HALL, J.C. (1986) Chemosensory elements of courtship in normal and mutant, olfaction-deficient *Drosophila melanogaster*. *Behav. Genet.* 16: 375-405.
- GAILEY, D.A., and HALL, J.C. (1989) Behavior and cytogenetics of *fruitless* in *Drosophila melanogaster*: different courtship defects caused by separate, closely linked lesions. *Genetics* 121: 773-785.
- GAILEY, D.A., TAYLOR, B.J., and HALL, J.C. (1991) Elements of the *fruitless* locus regulate development of the Muscle of Lawrence, a male-specific structure in the abdomen of *Drosophila melanogaster* adults. *Development* 113: 879-890.
- GALTON, F. (1869) *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan.
- GILL, K.S. (1963) A mutation causing abnormal courtship and mating behaviour in males of *Drosophila melanogaster*. (Abstract) *Am. Zool.* 3: 507.
- GLOOR, G, and ENGELS, W. (1990) Single-fly DNA preps for PCR. *Drosophila Information Newsletter* 1.
- GOLDSMITH, T.H. (1965) Do flies have a red receptor? *J. Gen. Physiol.* 49:265-287.
- GOLIC, K.G. (1994) Local transposition of *P* elements in *Drosophila melanogaster* and recombination between duplicated elements using a site-specific recombinase. *Genetics* 137: 551-563.
- GOODE, B.L., and FEINSTEIN, S.C. (1992) Speedprep purification of template for double-stranded DNA sequencing. *Biotechniques* 3: 374-375.
- GORALSKI, T.J., EDSTROM, J.E., and BAKER, B.S. (1989) The sex determination locus *transformer-2* of *Drosophila* encodes a polypeptide with similarity to RNA-binding proteins. *Cell* 56: 1011-1018.
- GOTZ, K.G. (1964) Optomotorische Untersuchung des visuellen Systems einiger Augenmutanten der Fruchtfliege. *Drosophila. Kybernetik* 2: 77-92.
- GOUY, M, and GAUTIER, C. (1982) Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10: 7055-7074.
- GRANTHAM, R., GAUTIER, C., GOUY, M., MERCIER, R, and PAVE, A. (1980) Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 8: r49-r62.
- GRANTHAM, R., GAUTIER, C., GOUY, M., JACOBZONE, M., and MERCIER, R. (1981) Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* 9: r43-r74.
- GRAUMANN, P. and MARAHIEL, A. (1996) A case of convergent evolution of nucleic acid binding modules. *BioEssays* 18: 309-315.

- GREENACRE, M.L., RITCHIE, M.G., BYRNE, B.C., and KYRIACOU, C.P. (1993) Female song preference and the *period* gene in *Drosophila*. *Behav. Genet.* 23: 85-90.
- GREENSPAN, R.J. (1995) Understanding the genetic construction of behaviour. *Scientific American* 272: 74-79.
- GRIFFITH, L.C., VERSELIS, L.M., AITKEN, K.M., KYRIACOU, C.P., DANHO, W., GREENSPAN, R.J. (1993) Inhibition of calcium calmodulin-dependent protein-kinase in *Drosophila* disrupts behavioral plasticity. *Neuron* 10: 501-509.
- GUILLE, M.J., and KNEALE, G.G. (1997) Methods for the analysis of DNA-protein interactions. *Molecular Biotechnology* 8: 35-52.
- HALL, J.C. (1977) Portions of the central nervous system controlling reproductive behaviour in *Drosophila melanogaster*. *Behav. Genet.* 7: 291-312.
- HALL, J.C. (1978) Courtship among males due to a male-sterile mutation in *Drosophila melanogaster*. *Behav. Genet.* 8: 125-141.
- HALL, J.C. (1979) Control of male reproductive behavior by the central nervous system of *Drosophila*: Dissection of a courtship pathway by genetic mosaics. *Genetics* 92: 437-457.
- HALL, J.C. (1986) Learning and rhythms in courting, mutant *Drosophila*. *Trends in Neurosci.* 9: 414-418.
- HALL, J.C. (1994) The mating of a fly. *Science* 264: 1702-1714.
- HANCOCK, J.M., and ARMSTRONG, J.S. (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. *Computer Applications in the Biosciences* 10: 67-70.
- HANCOCK, J.M., and DOVER, G.A. (1988) Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. *Mol. Biol. Evol.* 5: 377-391.
- HART, A.C., HARRISON, S.D., VAN VACTOR, D.L. JR., RUBIN, G.M., and ZIPURSKY, S.L. (1993) The interaction of bride of sevenless with sevenless is conserved between *Drosophila virilis* and *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 90: 5047-5051.
- HAUSEN, K. (1976) Functional characterization and anatomical identification of motion sensitive neurons in the lobula plate of the blowfly *Calliphora erythrocephala*. *Z. Naturforsch* 31C: 629-633.
- HAZELRIGG, T., LEVIS, R., RUBIN, G.M. (1984) Transformation of *white* locus DNA in *Drosophila*: dosage compensation, *zeste* interaction and position effects. *Cell* 36: 469-481.
- HEBERLEIN, U., and RUBIN, G.M. (1990) Structural and functional comparisons of the *Drosophila virilis* and *Drosophila melanogaster* rough genes. *Proc. Natl. Acad. Sci. USA* 87: 5916-5920.

- HEDLEY, M.L., and MANIATIS, T. (1991) Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to *tra-2* protein *in vitro*. *Cell* 65: 579-586.
- HEISENBERG, M. (1971a) Isolation of mutants lacking the optomotor response. *Drosophila Inf. Serv.* 46: 68.
- HEISENBERG, M. (1971b) Separation of receptor and lamina potentials in the electroretinogram of normal and mutant *Drosophila*. *J. Exp. Biol.* 55: 85-100.
- HEISENBERG, M. (1972) Comparative behavioral studies of two visual mutants of *Drosophila*. *J. Comp. Physiol.* 80: 119-136.
- HEISENBERG, M., and GOTZ, K.G. (1975) The use of mutations for the partial degradation of vision in *Drosophila melanogaster*. *J. Comp. Physiol.* 98: 217-241.
- HEISENBERG, M., and BUCHNER, E. (1977) The role of retinula cell types in visual behavior of *Drosophila melanogaster*. *J. Comp. Physiol.* 117: 127-162.
- HEISENBERG, M., WONNEBERGER, R., and WOLF, R. (1978) *Optomotor-blind*^{H31} - a *Drosophila* mutant of the lobula plate giant neurons. *J. Comp. Physiol.* 124: 287-296.
- HEISENBERG, M., BORST, A., WAGNER, S., and BYERS, D. (1985) *Drosophila* mushroom body mutants are deficient in olfactory learning. *J. Neurogenet.* 2: 1-30.
- HODGKIN, J. (1989) *Drosophila* sex determination: a cascade of regulated splicing. *Cell* 56: 905-906.
- HODGKIN, J. (1990) Sex determination compared in *Drosophila* and *Caenorhabditis*. *Nature* 344: 721-728.
- HOIKKALA, A., LAKOVAARA, S. and ROMPPAINEN, E. (1982) Mating behaviour and male courtship sounds in the *Drosophila virilis* group. *Advances in Genetics, Development, and Evolution of Drosophila* Edited by Seppo Lakovaara (Plenum Publishing Corporation, 1982).
- HOIKKALA, A., and LUMME, J. (1984) Genetic control of the difference in male courtship sound between *Drosophila virilis* and *Drosophila lummei*. *Behav. Genet.* 14: 257-268.
- HOIKKALA, A. (1985) Evolution of the male courtship sound in the species of the *Drosophila virilis* group. *Acta Universitatis Ouluensis, Series A, Scientiae Rerum Naturalium* No. 175. Biologica No. 25. Distributed by Oulu University Library, oulu, Finland.
- HOIKKALA, A., and LUMME, J. (1987) The genetic-basis of evolution of the male courtship sounds in the *Drosophila virilis* group. *Evolution* 41: 827-845.
- HOROWIZ, H., and BERG, C.A. (1995) Aberrant splicing and transcription termination caused by *P* element insertion into the intron of a *Drosophila* gene. *Genetics* 139: 327-335.

- HOSHIIJIMA, K., INOUE, K., HIGUCHI, I., SAKAMOTO, H., and SHIMURA, Y. (1991) Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. *Science* 252: 833-836.
- HOTTA, Y., and BENZER, S. (1969) Abnormal electroretinograms in visual mutants of *Drosophila*. *Nature* 222:354-356.
- HOTTA, Y., and BENZER, S. (1970) Genetic dissection of the *Drosophila* nervous system by means of mosaics. *Proc. Natl. Acad. Sci. USA* 67: 1156-1163.
- HOTTA, Y., and BENZER, S. (1976) Courtship in *Drosophila* mosaics: sex specific foci for sequential action patterns. *Proc. Natl. Acad. Sci. USA* 73: 4154-4158.
- HUGHES, A.L., and NEI, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167-170.
- IKEDA, H., TAKABATAKE, I., and SAWADA, N. (1980) Variation in courtship sounds among three geographical strains of *Drosophila mercatorum*. *Behav. Genet.* 10: 361-375.
- IKEMURA, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2: 13-34.
- INOUE, K., HOSHIIJIMA, K., SAKAMOTO, H., and SHIMURA, Y. (1990) Binding of the *Drosophila Sex-lethal* gene-product to the alternative splice site of *transformer* primary transcript. *Nature* 344: 461-463.
- ITO, H., FUJITANI, K., USUI, K., SHIMIZUNISHIKAWA, K., TANAKA, S., and YAMAMOTO, D. (1996) Sexual orientation in *Drosophila* is altered by the *satori* mutation in the sex-determination gene *fruitless* that encodes a zinc-finger protein with a BTB domain. *Proc. Natl. Acad. Sci. USA* 93: 9687-9692.
- JAFFE, E., and LAIRD, C. (1986) Dosage compensation in *Drosophila*. *Trends Genet.* 2: 316-321.
- JALLON, J-M (1984) A few chemical words exchanged by *Drosophila* during courtship and mating. *Behav. Genet.* 14: 441-478.
- JALLON, J.M., and DAVID, J.R. (1987) Variations in cuticular hydrocarbons among the 8 species of the *drosophila-melanogaster* subgroup. *Evolution* 41:294-302.
- JEFFREYS, A.J., WILSON, V., NEUMANN, R., and KEYTE, J. (1988) Amplification of human minisatellites by the polymerase chain-reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.* 23: 10953-10971.
- JOHNSON-SCHLITZ, D.M., and ENGELS, W.R. (1993) *P*-element-induced interallelic gene conversion of insertions and deletions in *Drosophila melanogaster*. *Molecular and Cellular Biology* 13: 7006-7018.
- JONES, K.R., and RUBIN, G.M. (1990) Molecular analysis of *no-on-transient A*, a gene required for normal vision in *Drosophila*. *Neuron* 4: 711-723.

- KAISER, K. (1993) A second generation of enhancer-traps. *Curr. Biol.* 3: 560-562.
- KALM, L.V., CROSSGROVE, K., VONSEGGER, D., GUILD, G.M., and BECKENDORF, S.K. (1994) The BROAD-COMPLEX directly controls a tissue-specific response to the steroid-hormone ecdysone at the onset of *Drosophila* metamorphosis. *Embo J.* 13: 3505-3516.
- KASSIS, J.A. POOLE, S.J., WRIGHT, D.K. and O'FARRELL, P.H. (1986) Sequence conservation in the coding and intron regions of the *engrailed* transcription unit. *EMBO J.* 5: 3583-3589.
- KAUFMAN, P.D., DOLL, R.F., and RIO, D.C. (1989) *Drosophila* P element transposase recognizes internal P element DNA sequences. *Cell* 59: 359-371.
- KAWANISHI, M. and WATANABE, T.K. (1980) Genetic variations of courtship song of *Drosophila melanogaster* and *D.simulans*. *Jpn. J. Genet.* 55: 235-240.
- KELLER, G.H., HUANG, D.P., and MANAK, M.M. (1989) Labeling of DNA probes with a photoactivatable hapten. *Analytical Biochemistry* 177: 392-395.
- KENAN, D.J., QUERY, C.C., and KEENE, J.D (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.* 16: 214-220.
- KILEDJIAN, M. and DREYFUSS, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* 11: 2655-2664.
- KIM, Y-J., ZUO, P., MANLEY, J.L., and BAKER, B.S. (1992) The *Drosophila* RNA-binding protein RBP1 is localized to transcriptionally active sites of chromosomes and shows a functional similarity to human splicing factor ASF/SF2. *Genes Dev.* 6: 2569-2579.
- KIMURA, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.
- KIMURA, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- KIMURA, M. (1991) The neutral theory of molecular evolution: a review of recent-evidence. *Jpn. J. Genet.* 66: 367-386.
- KLECKNER, N. (1981) Transposable elements in prokaryotes. *Ann. Rev. Genet.* 15: 341-404.
- KLEMENZ, R., WEBER, U., and GEHRING, W.J. (1987) The *white* gene as a marker in a new P-element vector for gene transfer in *Drosophila*. *Nucleic Acids Res.* 15: 3947-3959.
- KONOPKA, R.J., and BENZER, S. (1971) Clock mutants of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 68: 2112-2116.
- KOZAK, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate mRNAs. *Nucleic Acids Res.* 15: 8125-8148.

- KOZAK, M. (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* 115: 887-903.
- KULKARNI, S.J., and HALL, J.C. (1987) Behavioral and cytogenetic analysis of the *cacophony* courtship song mutant and interacting genetic-variants in *Drosophila melanogaster*. *Genetics* 115: 461-475.
- KULKARNI, S.J., STEINLAUF, A.F., and HALL, J.C. (1988) The *dissonance* mutant of courtship song in *Drosophila melanogaster*: isolation, behavior and cytogenetics. *Genetics* 118: 267-285.
- KURLAND, C.G. (1987) Strategies for efficiency and accuracy in gene expression. *Trends Biochem. Sci.* 12: 126-128.
- KYRIACOU, C.P., and HALL, J.C. (1980) Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proc. Natl. Acad. Sci. USA* 77: 6729-6733.
- KYRIACOU, C.P., and HALL, J.C. (1982) The function of courtship song rhythms in *Drosophila*. *Anim. Behav.* 30: 794-801.
- KYRIACOU, C.P., and HALL, J.C. (1984) Learning and memory mutations impair acoustic priming of mating behavior in *Drosophila*. *Nature* 308: 62-65.
- KYRIACOU, C.P., and HALL, J.C. (1986) Interspecific genetic-control of courtship song production and reception in *Drosophila*. *Science* 232: 494-497.
- KYRIACOU, C.P., and HALL, J.C. (1989) Spectral analysis of *Drosophila* courtship song rhythms. *Anim. Behav.* 37: 850-859.
- KYRIACOU, C.P., and HALL, J.C. (1990) Song rhythms in *Drosophila*. *Trends Ecol. Evol.* 5: 125-126.
- KYTE, J., and DOOLITTLE, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157: 105-132.
- LAKOVAARA, S. and HOIKKALA, A. (1979) Analysis of the male courtship sounds of certain species of the *Drosophila virilis* group. *Aquilo. Ser. Zool.* 20: 94-99.
- LASKI, F.A., RIO, D.C., and RUBIN, G.M. (1986) Tissue specificity of *Drosophila P* element transposition is regulated at the level of mRNA splicing. *Cell* 44: 7-19.
- LAWRENCE, P.A., and JOHNSTON, P. (1986) The muscle pattern of a segment of *Drosophila* may be determined by neurons and not by contributing myoblasts. *Cell* 45: 505-513.
- LEVIN, L.R., HAN, P.L., HWANG, P.M., FEINSTEIN, P.G., DAVIS, R.L., REED, R.R. (1992) The *Drosophila* learning and memory gene *rutabaga* encodes a Ca^{2+} /Calmodulin-responsive adenylyl cyclase. *Cell* 68: 479-489.
- LI, W-H., WU, C-I., and LUO, C-C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* 2: 150-174.

- LI, W-H., and TANIMURA, M (1987) The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326: 93-96.
- LI, W-H., and GRAUR, D. (1991) *Fundamentals of Molecular Evolution* (Sinauer, Sunderland, MA).
- LI, W-H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* 36: 96-99.
- LI, W-H. (1997) *Molecular Evolution*. Chapter 8: 215-235 (Sinauer, Sunderland, MA).
- LINDSLEY, D.L., and ZIMM, G.G. (1992) The genome of *Drosophila melanogaster*. Academic Press, Inc. London.
- LIVINGSTONE, M.S. (1985) Genetic dissection of *Drosophila* adenylate-cyclase. *Proc. Natl. Acad. Sci. USA* 82: 5992-5996.
- MANNING, A. (1975) Behaviour genetics and the study of behavioural evolution. From Function and Evolution in Behaviour. Essays in honour of professor Niko Tinbergen, F.R.S. G. Baerends, C. Beer and A. Manning Ed. Clarendon Press. Oxford.
- MATTOX, W., and BAKER, B.S. (1991) Autoregulation of the splicing of transcripts from the *transformer-2* gene of *Drosophila*. *Genes Dev.* 5: 786-796.
- MAYEDA, A., ZAHLER, A.M., KRAINER, A.R., and ROTH, M.B. (1992) Two members of a conserved family of nuclear phosphoproteins are involved in pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA* 89: 1301-1304.
- MAYR, E. (1977) Populations, Species, and Evolution. An abridgment of *Animal Species and Evolution*. The Belknap Press of Harvard University Press. Cambridge, MA and London, England.
- McKEOWN, M. (1994) Sex determination and differentiation. *Dev. Genet.* 15: 201-204.
- McROBERT, S.P., and TOMPKINS, L. (1985) The effect of *transformer*, *doublesex* and *intersex* mutations on the sexual-behavior of *Drosophila melanogaster*. *Genetics* 111: 89-96.
- MESSIER, W. and STEWART, C-B (1997) Episodic adaptive evolution of primate lysozymes. *Nature* 385: 151-154.
- MICHAEL, W.M., BOWTELL, D.D.L., and RUBIN, G.M. (1990) Comparison of the sevenless genes of *Drosophila virilis* and *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 87: 5351-5353.
- MILLER, D.D., GOLDSTEIN, R.B., and PATTY, R.A. (1975) Semispecies of *Drosophila athabasca* distinguishable by male courtship sounds. *Evolution* 29: 531-544.
- MORIYAMA, E.N. (1987) Higher rates of nucleotide substitution in *Drosophila* than in mammals. *Jpn. J. Genet.* 62: 139-147.
- MORIYAMA, E.N., and GOJOBORI, T. (1992) Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. *Genetics* 130: 855-864.

- MOSES, K. (1989) The glass gene encodes a zinc-finger protein required by *Drosophila* photoreceptor cells. *Nature* 340: 531-536.
- MOTTES, J., SABLE, E., TAKAGAWA, K., and IVERSON, L. (1993) Regulation of splicing of *Shaker* potassium channel transcripts: An *in vitro* assay (abstract). Neurobiology of *Drosophila*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- NAKAMURA, M., OKANO, H., BLENDY, J.A., and MONTELL, C. (1994) Musashi, a neural RNA-binding protein required for *Drosophila* adult external sensory organ development. *Neuron* 13: 67-81.
- NOOR, M.A. (1995) Speciation driven by natural selection in *Drosophila*. *Nature* 375: 674-675.
- O'DELL, K.M.C., ARMSTRONG, J.D., YANG, M.Y., and KAISER, K. (1995) Functional dissection of the *Drosophila* mushroom bodies by selective feminization of genetically defined subcompartments. *Neuron* 15: 55-61.
- O'DELL, K.M.C, and KAISER, K. (1997) Sexual behavior: Secrets and flies. *Curr. Biol.* 7: R345-R347.
- O'HARE, K., and RUBIN, G.M. (1983) Structures of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* 34: 25-35.
- O'NEIL, M.T., and BELOTE, J.M. (1992) Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics* 131: 113-128.
- O'TOUSA, J.E. (1990) Genetic analysis of phototransduction. In *Drosophila. Sem. Neurosci.* 2: 207-215.
- OUSLEY, A., ZAFARULLAH, K., CHEN, Y., EMERSON, M., HICKMAN, L., and SEHGAL, A. (1998) Phylogenetic analysis of the *timeless* clock gene reveals a novel functional domain. *Genetics* In Press.
- PAK, W.L., GROSSFIELD, J., and WHITE, N.V. (1969) Nonphototactict mutants in a study of vision of *Drosophila*. *Nature* 222: 351-354.
- PAK, W.L., GROSSFIELD, J., and ARNOLD, K.S. (1970) Mutants of the visual pathway of *Drosophila melanogaster*. *Nature* 227: 518-520.
- PAK, W.L. (1975) Mutations affecting the vision of *Drosophila melanogaster*. In King, R.C., (Ed), *Handbook of genetics*, New York, Plenum Press, 3: 703-732.
- PAK, W.L. (1991) Molecular genetic studies of photoreceptor function using *Drosophila* mutants. *Prog. Clin. Biol. Res.* 362: 1-32.
- PAMILO, P., and BIANCHI, N.O. (1993) Evolution of the *Zfx* and *Zfy* genes: Rates and interdependence between the genes. *Mol. Biol. Evol.* 10: 271-281.
- PAQUES, F., and WEGNEZ, M. (1993) Deletions and amplifications of tandemly arranged ribosomal 5S genes internal to a *P* element occur at a high rate in a dysgenic context. *Genetics* 135: 469-476.

- PAQUES, F., BUCHETON, B., and WEGNEZ, M. (1996) Rearrangements involving repeated sequences within a *P* element preferentially occur between units close to the transposon extremities. *Genetics* 142: 459-470.
- PASTINK, A., SCHALET, A.P., VREEKEN, C., PARADI, E., and EEKEN, J.C.J. (1987) The nature of radiation-induced mutations at the *white* locus of *Drosophila melanogaster*. *Mutation Research* 177: 101-115.
- PATTERSON, J.T., and STONE, W.S. (1952) Evolution in the genus *Drosophila*. The Macmillan Company, New York, 610 pp.
- PATTON, J.G., PORRO, E.B., GÁLGERAN, J., TEMPST, P., and NADAL-GINARD, B. (1993) Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.* 7: 393-406.
- PEIXOTO, A.A., CAMPESAN, S., COSTA, R., and KYRIACOU, C.P. (1993) Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. *Mol. Biol. Evol.* 10: 127-139.
- PETERSEN, G., HALL, J.C., and ROSBASH, M. (1988) The *period* gene of *Drosophila* carries species-specific behavioral instructions. *EMBO J.* 12: 3939-3947.
- PFLUGFELDER, G.O., SCHWARZ, H., ROTH, H., POECK, B., SIGL, A., KERSCHER, S., JONSKER, B., PAK, W.L., and HEISENBERG, M. (1990) Genetic and molecular characterization of the *optomotor-blind* gene locus in *Drosophila melanogaster*. *Genetics* 126: 91-104.
- PFLUGFELDER, G.O. (1992) *optomotor-blind*: a key gene in the development of the optic lobes in *Drosophila*. *European Journal of Neuroscience* No.55: 236.
- PFLUGFELDER, G.O., ROTH, H., POECK, B., KERSCHER, S., SCHWARZ, H., JONSKER, B., and HEISENBERG, M. (1992). The *lethal (1) optomotor-blind* gene of *Drosophila melanogaster* is a major organizer of optic lobe development: isolation and characterization of the gene.
- QUINN, W.G., and GREENSPAN, R.J. (1984) Learning and courtship in *Drosophila*: 2 stories with mutants. *Ann. Rev. Neurosci.* 7: 67-93.
- RAY, R.P., and SCHÜPBACH (1996) Intercellular signaling and the polarization of body axes during *Drosophila* oogenesis. *Genes Dev.* 10: 1711-1723.
- REDDY, P., ZEHRING, W.A., WHEELER, D.A., PIRROTTA, V., HADFIELD, C., HALL, J.C., and ROSBASH, M. (1984) Molecular analysis of the *period* locus in *Drosophila melanogaster* and identification of a transcript involved in biological rhythms. *Cell* 38: 701-710.
- REND AHL, K.G., JONES, K.R., KULKARNI, S.J., BAGULLY, S.H., and HALL, J.C. (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., and HALL, J.C. (1996) Temporally manipulated rescue of visual and courtship abnormalities caused by a *nonA* mutation in *Drosophila*. *J. Neurogen.* 10: 247-256.
- RENDAHL, K.G., GAUKHSHTEYN, N., WHEELER, D.A., FRY, T.A., and HALL, J.C. (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*. *The Journal of Neuroscience* 15: 1511-1522.
- RIO, D.C. (1993) Splicing of pre-mRNA: mechanism, regulation and role in development. *Curr. Opin. Genet. Dev.* 3: 574-584.
- RIVERA-POMAR, R., JACKLE, H. (1996) From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trend Genet.* 12: 478-483.
- ROBERTSON, H.M., PRESTON, C.R., PHILLIS, R.W., JOHNSON-SCHLITZ, D.M., BENZ, W.K., and ENGELS, W.R. (1988) A stable genomic source of *P* element transposase in *Drosophila melanogaster*. *Genetics* 118: 461-470.
- ROBINOW, S., and WHITE, K. (1988) The locus *elav* of *Drosophila melanogaster* is expressed in neurons at all developmental stages. *Dev. Biol.* 126: 294-303.
- ROBINOW, S., CAMPOS, A.R., YAO, K.M., and WHITE, K. (1988) The *elav* gene-product of *Drosophila*, required in neurons, has 3 RNP consensus motifs. *Science* 242: 1570-1572.
- ROBINOW, S., and WHITE, K. (1991) Characterization and spatial distribution of the ELAV protein during *Drosophila melanogaster* development. *J. Neurobiol.* 22: 443-461.
- RUBIN, G.M., and SPRADLING, A.C. (1982) Genetic-transformation of *Drosophila* with transposable element vectors. *Science* 218: 348-353.
- RUBIN, G.M., KIDWELL, M.G., and BINGHAM, P.M. (1982) The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* 29: 987-994.
- RUBIN, G.M., and SPRADLING, A.C. (1983) Vectors for *P* element-mediated gene transfer in *Drosophila*. *Nucleic Acids Res.* 11: 6341-6351.
- RUBIN, G.M. (1988) *sevenless*, a gene encoding a putative receptor for positional information. *Cold Spring Harbor Symposia on Quantitative Biology* 53: 417-424.
- RUSCH, J. and LEVINE, M. (1996) Threshold responses to the dorsal regulatory gradient and the subdivision of primary tissue territories in the *Drosophila* embryo. *Curr. Opin. Genet. Dev.* 6: 416-423.
- RYNER, L.C., and BAKER, B.S. (1991) Regulation of *doublesex* pre-mRNA processing occurs by 3' splice site activation. *Genes Dev.* 5: 2071-2085.
- RYNER, L.C., GOODWIN, S.F., CASTRILLON, D.H., ANAND, A., VILLELLA, A., BAKER, B.S., HALL, J.C., TAYLOR, B.J., and WASSERMAN,

- S.A. (1996) Control of male sexual-behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* 87: 1079-1089.
- SACHS, A.B., BOND, M.W., and KORNBERG, R.D. (1986) A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain-structure and expression. *Cell* 45: 827-835.
- SAMBROOK, J., MANIATIS, T. and FRITSCH (1989) Molecular cloning, a laboratory manual. Cold Spring Harbour Laboratory Press, USA.
- SANGER, F., NICKLEN, S., and COULSON, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.
- SAUMWEBER, H., FRASCH, M., and KORGE, G. (1990) Two puff-specific proteins bind within the 2.5 kb upstream region of the *Drosophila melanogaster* *Sgs-4* gene. *Chromosoma* 99: 52-60.
- SCHILCHER, F. von. (1976a) The role of auditory stimuli in the courtship of *Drosophila melanogaster*. *Anim. Behav.* 24: 18-26.
- SCHILCHER, F. von. (1976b) The function of pulse song and sine song in the courtship of *Drosophila melanogaster*. *Anim. Behav.* 24: 622-625.
- SCHILCHER, F. von. (1977) A mutation which changes courtship song in *Drosophila melanogaster*. *Behav. Genet.* 7: 251-259.
- SCHILCHER, F., and HALL, J.C. (1979) Neural topography of courtship song in sex mosaics of *Drosophila melanogaster*. *J. Comp. Physiol.* 129: 85-95.
- SCOTT, D. (1986) Sexual mimicry regulates the attractiveness of mated *Drosophila melanogaster* females. *Proc. Nat. Sci. USA* 83: 8429-8433.
- SCOTT, D., and JACKSON, L.L. (1988) Interstrain comparison of male-predominant antiaphrodisiacs in *Drosophila melanogaster*. *J. Ins. Physiol.* 34: 863-871.
- SCOTT, D., RICHMOND, R.C., and CARLSON, D.A. (1988) Pheromones exchanged during mating: a mechanism for mate assessment in *Drosophila*. *Anim. Behav.* 36: 1164-1173.
- SENTRY, J.W., GOODWIN, S.F., MILLIGAN, C.D., DUNCANSON, A., YANG, M. and KAISER, K. (1994) Reverse genetics of *Drosophila* brain structure and function. In "Progress in Neurobiology" Vol. 42 (eds. G.A. Kerkurt and J.W. Phillis), pp. 299-308, Pergamon Press.
- SHAPIRO, J.A. (1983) Mobile Genetic Elements. (New York: Academic Press).
- SHARP, P.M., and LI, W.H. (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for rare codons. *Nucleic Acids Res.* 14: 7737-7749
- SHARP, P.M., COWE, E., HIGGINS, D.G., SHIELDS, D.C., WOLFE, K.H., and WRIGHT, F. (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.
- SHARP, P.M., LI, W-H. (1989) On the rate of DNA sequence evolution in *Drosophila*. *J. Mol. Evol.* 28: 398-402.
- SHARP, P.M. (1997) In search of molecular darwinism. *Nature* 385: 111-112.
- SHIELDS, D.C., SHARP, P.M., HIGGINS, D.G., and WRIGHT, F. (1988) Silent sites in *Drosophila* genes are not neutral; evidence of selection among synonymous codons. *Mol. Biol. Evol.* 5: 704-716.
- SIEGEL, R.W., and HALL, J.C. (1979) Conditioned responses in courtship behavior of normal and mutant *Drosophila*. *Proc. Natl. Acad. Sci. USA* 76: 3430-3434.
- SIMON, M.A., BOWTELL, D.D.L., and RUBIN, G.M. (1989) Structure and activity of the sevenless protein: a protein tyrosine kinase receptor required for photoreceptor development in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 86: 8333-8337.
- SIOMI, H., and DREYFUSS, G. (1997) RNA-binding proteins as regulators of gene expression. *Curr. Opin. Genet. Dev.* 7: 345-353.
- SMITH, D.P., STAMNES, M.A., and ZUCKER, C.S. (1991) Signal transduction in the visual system of *Drosophila*. *Annu. Rev. Cell. Biol.* 7: 161-190.
- SMITH, L.A., WANG, X.J., PEIXOTO, A.A., NEUMANN, E.K., HALL, L.M., and HALL, J.C. (1996) A *Drosophila* Calcium channel $\alpha 1$ subunit gene maps to a genetic locus associated with behavioral and visual defects. *J. Neurosci.* 16: 7868-7879.
- SMITH, L.A., PEIXOTO, A.A., WRIGHT, E., and HALL, J.C. (1997) Courtship and visual defects of *cacophony* mutants reveal functional complexity of a Calcium channel $\alpha 1$ subunit in *Drosophila*. (In Press).
- SMITH, T.F., and WATERMAN, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.* 147: 195-197.
- SOLOVYEV, V.V., and KOLCHANOV, N.A. (1994) Search for functional sites using consensus. In *Computer Analysis of Genetic Macromolecules*. (eds. Kolchanov, N.A., Lim, H.A.), World Scientific, pp. 16-21.
- SOSNOWSKI, B.A., BELOTE, J.M., McKEOWN, M. (1989) Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence dependent splice site blockage. *Cell* 58: 449-459.
- SPRADLING, A.C. and RUBIN, G.M. (1982) Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Science* 218: 341-347.
- ST. JOHNSTON, D., and NÜSSLEIN-VOLHARD, C. (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68: 201-219.
- STANEWSKY, R., RENDAHL, K.G., DILL, M., and SAUMWEBER, H. (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., FRY, T.A., REIM, I., SAUMWEBER, H., and HALL, J.C. (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.
- STAVELEY, B.E., HESLIP, T.R., HODGETTS, R.B., and BELL, J.B. (1995) Protected *P*-element termini suggest a role for inverted-repeat-binding protein in transposase-induced gap repair in *Drosophila melanogaster*. *Genetics* 139: 1321-1329.
- STEWART, R. and GOVIND, S. (1993) Dorsal-ventral polarity in the *Drosophila* embryo. *Curr. Opin. Genet. Dev.* 3: 556-561.
- STOCKER, R.F., and GENDRE, N. (1989) Courtship behavior of *Drosophila* genetically or surgically deprived of basiconic sensilla. *Behav. Genet.* 19: 371-385.
- STOCKER, R.F. (1994) The organization of the chemosensory system in *Drosophila melanogaster*: a review. *Cell and Tissue Research* 275: 3-26.
- STONE, W.S., GUEST, W.C. and WILSON, F.D. (1960) The evolutionary implications of the cytological polymorphism and phylogeny of the *virilis* group of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 46: 350.
- SUTTON, K.A., and WILKINSON, M.F. (1997) Rapid evolution of a homeodomain: Evidence for positive selection. *J. Mol. Evol.* 45: 579-588.
- TAUTZ, D., TRICK, M., and DOVER, G.A. (1986) Cryptic simplicity in DNA is a major source of genetic-variation. *Nature* 322: 652-656.
- TAUTZ, D., TAUTZ, C., WEBB, D., and DOVER, G.A. (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.
- TAYLOR, B.J. (1992) Differentiation of a male-specific muscle in *Drosophila melanogaster* does not require the sex-determining genes *doublesex* or *intersex*. *Genetics* 132: 179-191.
- TAYLOR, B.J., and TRUMAN, J.W. (1992) Commitment of abdominal neuroblasts in *Drosophila* to a male or female fate is dependent on genes of the sex-determining hierarchy. *Development* 114: 625-642.
- TAYLOR, B.J., VILLELLA, A., RYNER, L.C., BAKER, B.S., and HALL, J.C. (1994) Behavioral and neurobiological implications of sex-determining factors in *Drosophila*. *Dev. Genet.* 15: 275-296.
- TAYLOR, B.J., and KNITTEL, L.M. (1995) Sex-specific differentiation of a male-specific abdominal muscle, the Muscle of Lawrence, is abnormal in hydroxyurea-treated and in *fruitless* male flies. *Development* 121: 3079-3088.

- THACKERAY, J.R. , and GANETZKY, B. (1994) Developmentally regulated alternative splicing generates a complex array of *Drosophila para* sodium-channel isoforms. *J. Neurosci.* 14: 2569-2578.
- THROCKMORTON, L.H. (1982) The *virilis* species group. In: Ashburner, M., Carson, H.L., and Thompson, J.N., Jr. (eds.) *The Genetics and Biology of Drosophila*, Vol 3b, Academic Press, London, pp. 227-296.
- TIAN, M., and MANIATIS, T. (1993) A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. *Cell* 74: 105-114.
- TOMARU, M., and OGUMA, Y. (1994a) Genetic-basis and evolution of species-specific courtship song in the *Drosophila auraria* complex. *Genet. Res.* 63: 11-17.
- TOMARU, M., and OGUMA, Y. (1994b) Differences in courtship song in the species of the *Drosophila auraria* complex. *Anim. Behav.* 47: 133-140.
- TOMPKINS, L., HALL, J.C., and HALL, L.M. (1980) Courtship-stimulating volatile compounds from normal and mutant *Drosophila*. *J. Ins. Physiol.* 26: 689-697.
- TOMPKINS, L., and HALL, J.C. (1981) The different effects on courtship of volatile compounds from mated and virgin *Drosophila* females. *J. Ins. Behav.* 27: 17-21.
- TOMPKINS, L., and HALL, J.C. (1983) Identification of brain sites controlling female receptivity in mosaics of *Drosophila melanogaster*. *Genetics* 103: 179-195.
- TOMPKINS, L. (1984a) Genetic analysis of sex appeal in *Drosophila*. *Behav. Genet.* 14: 411-440.
- TOMPKINS, L. (1984b) The effect of *Sex-lethal* mutations on the sexual behaviour of *D.melanogaster*. *Genetics* 107: 107-115.
- TOMPKINS, L. (1986) Genetic control of sexual-behavior in *Drosophila melanogaster*. *Trends Genet.* 2:14-17.
- TOMPKINS, L., and McROBERT, S.P. (1989) Regulation of behavioral and pheromonal aspects of sex determination in *Drosophila melanogaster* by the *sex-lethal* gene. *Genetics* 123: 535-541.
- TOWER, J., KARPEN, G.H., CRAIG, N., and SPRADLING , A.C. (1993) Preferential transposition of *Drosophila P* elements to nearby chromosomal sites. *Genetics* 133: 347-359.
- TREIER, M., PFEIFLE, C., and TAUTZ, D. (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.
- TRYON, R.C. (1940) Genetic differences in maze learning ability in rats. *National Society for the Study of Education 39th Yearbook*, 39: 111-119.
- TULLY, T., GERGEN, J.P. (1986) Deletion mapping of the *Drosophila* memory mutant *amnesiac*. *Journal of Neurogenetics* 3: 33-47.

- TULLY, T. (1987) *Drosophila* learning and memory revisited. *Trends Neurosci.* 10: 330-335.
- VAIAS, L.J., NAPOLITANO, L.M., and TOMPKINS, L. (1993) Identification of stimuli that mediate experience-dependent modification of homosexual courtship in *Drosophila-melanogaster*. *Behav. Genet.* 23: 91-97.
- VIEIRA, J., and MESSING, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19: 259-268.
- VILLELLA, A., and HALL, J.C. (1996) Courtship anomalies caused by *doublesex* mutations in *Drosophila melanogaster*. *Genetics* 143: 331-344.
- VILLELLA, A., GAILEY, D.A., BERWALD, B., OHSHIMA, S., BARNES, P.T., and HALL, J.C. (1997) Extended reproductive roles of the *fruitless* gene in *Drosophila melanogaster* revealed by behavioral analysis of new *fru* mutants. *Genetics* 147: 1107-1130.
- WALDRON, I. (1964) Courtship and sound production in two sympatric sibling *Drosophila* species. *Science* 144: 191-192.
- WALLRATH, L.L., and FRIEDMAN, T.B. (1991) Species differences in the temporal pattern of *Drosophila* urate oxidase gene expression are attributed to trans-acting regulatory changes. *Proc. Natl. Acad. Sci. USA* 88: 5489-5493.
- WELBERGEN, P., VANDIJKEN, F.R., and SCHARLOO, W. (1987) Collation of the courtship behavior of the sympatric species *Drosophila melanogaster* and *Drosophila simulans*. *Behaviour* 101: 253-274.
- WHEELER, D.A., FIELDS, W.L., and HALL, J.C. (1988) Spectral analysis of *Drosophila* courtship songs: *D.melanogaster*, *D.simulans*, and their interspecific hybrid. *Behav. Genet.* 18: 675-703.
- WHEELER, D.A., KULKARNI, S.J., GAILEY, D.A., and HALL, J.C. (1989) Spectral-analysis of courtship songs in behavioral mutants of *Drosophila melanogaster*. *Behav. Genet.* 19: 503-528.
- WHEELER, D.A., KYRIACOU, C.P., GREENACRE, M.L., YU, Q., RUTILA, J.E., ROSBASH, M., and HALL, J.C. (1991) Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. *Science* 251: 1082-1085.
- WILSON, R., BURNET, B., EASTWOOD, L., and CONNOLLY, K. (1976) Behavioural pleiotropy of the *yellow* gene in *Drosophila melanogaster*. *Genet. Res.* 28: 75-88.
- WINER, B.J. (1971) Statistical principles in experimental design. (2nd ed.) McGraw-Hill Kogakusha, Ltd. London.
- WINGENDER, E., KEL, A.E., KEL, O.V., KARAS, H., HEINEMEYER, T., DIETZE, P., KNUPPEL, R., ROMASCHENKO, A.G., AND KOLCHANOV, N.A. (1997) TRANSFAC, TRRD and COMPEL: Towards a federated database system on transcriptional regulation. *Nucleic Acids Res.* 25: 265-268.

- WOLBARSH, M.L., WAGNER, H., and BODENSTEIN, D. (1966) Origin of electrical responses in the eye of *Periplaneta americana*. In *The Functional Organization of the Compound eye*, edited by C. G. Bernhard, pp. 207-217, Pergamon Press, London.
- WOLFE, K.H., and SHARP, P.M. (1993) Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *J. Mol. Evol.* 37: 441-456.
- WU, C.F., and GANETZKY, B. (1992). In *Ion Channels*. T. Narahashi, Ed. (Plenum, New York, 1992), vol.3, pp 261-314.
- YANG, Y.S., HANKE, J.H., CARAYANNOPOULOS, L., CRAFT, C.M., CAPRA, J.D., and TUCKER, P.W. (1993) NonO, a non-POU-domain-containing, octamer-binding protein, is the mammalian homolog of *Drosophila nonA^{diss}*. *Mol. Cell. Biol.* 13: 5593-5603.
- YAO, K-M., and WHITE, K. (1991) Organizational analysis of *elav* gene and functional analysis of ELAV protein of *D.melanogaster* and *D.virilis*. *Mol. Cell. Biol.* 11: 2994-3000.
- YAO, K-M., SAMSON, M.L., REEVES, R., and WHITE, K. (1993) Gene *elav* of *Drosophila melanogaster*: a prototype for neuronal-specific RNA binding protein gene family that is conserved in flies and humans. *J. Neurobiol.* 24: 723-739.
- YOKOKURA, T., UEDA, R., and YAMAMOTO, D. (1995) Phenotypic and molecular characterization of *croaker*, a new mating behavior mutant of *Drosophila melanogaster*. *Jpn. J. Genet.* 70: 103-117.
- YU, Q., COLOT, H.V., KYRIACOU, C.P., HALL, J.C., and ROSBASH, M. (1987) Behaviour modification by *in vitro* mutagenesis of a variable region within the *period* gene of *Drosophila*. *Nature* 326: 765-769.
- ZAHLER, A.M., LANE, W.S., STOLK, J.A., and ROTH, M.B. (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* 6: 837-847.
- ZAMORE, P.D., PATTON, J.G. and GREEN, M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355: 609-614.
- ZAWISTOWSKI, S. (1988) A replication demonstrating reduced courtship of *Drosophila melanogaster* by associative learning. *J. Comp. Psychol.* 102: 174-176.
- ZEHRING, W.A., WHEELER, D.A., REDDY, P., KONOPKA, R.J., KYRIACOU C.P., ROSBASH, M., and HALL, J.C. (1984) P-element transformation with *period* locus DNA restores rhythmicity to mutant, arrhythmic *Drosophila*. *Cell* 2: 369-376.
- ZHANG, S.D., and ODENWALD, W.F. (1995) Misexpression of the *white (w)* gene triggers male-male courtship in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 92: 5525-5529.