Towards developing a male-only Queensland fruit fly strain

Dr. Deborah Shearman
Fruit Fly Research Centre

Project Number: HG02037

HG02037

This report is published by Horticulture Australia Ltd to pass on information concerning horticultural research and development undertaken for the Australian horticulture industry.

The research contained in this report was funded by Horticulture Australia Ltd with the financial support of Fruit Fly Research Centre (on behalf of contributors).

All expressions of opinion are not to be regarded as expressing the opinion of Horticulture Australia Ltd or any authority of the Australian Government.

The Company and the Australian Government accept no responsibility for any of the opinions or the accuracy of the information contained in this report and readers should rely upon their own enquiries in making decisions concerning their own interests.

ISBN 0 7341 1082 0

Published and distributed by:
Horticultural Australia Ltd
Level 1
50 Carrington Street
Sydney NSW 2000
Telephone: (02) 8295 2300
Fax: (02) 8295 2399
E-Mail: horticulture@horticulture.com.au

© Copyright 2005

Know-how for Horticulture™

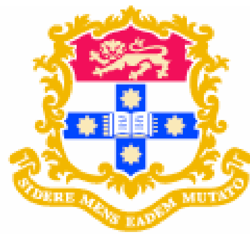
HORTICULTURE AUSTRALIA LTD

FINAL REPORT

PROJECT HG02037 (01/09/2002-30/06/2004)

Towards developing a male-only Queensland fruit fly strain

Dr. Deborah C. A. Shearman
Fruit Fly Research Centre,
School of Biological Sciences,
The University of Sydney



FRUIT FLY RESEARCH CENTRE
The University of Sydney

Horticulture Australia Project number: HG02037

Project/Program Leader
Dr Deborah C. A. Shearman

Principal Investigator
Dr Deborah C. A. Shearman*, Fruit Fly Research Centre (FFRC), University of Sydney

***Address: Fruit Fly Research Centre, A12, University of Sydney, NSW 2006, Australia**
E-mail: shearman@bio.usyd.edu.au
Telephone: + 61 2 9351 2386
FAX: + 61 2 9351 4771

Collaborative Investigators: Ms Merryl Robson (Administration), FFRC.

Purpose of report

This report is an account of the research into the development of a male-only SIT strain for the Queensland fruit fly, Bactrocera tryoni. *The research was an investigation into suitable sex-specific genes that might be used in female-killing constructs in* B. tryon*i and related pest species, such as* Bactrocera dorsalis *(Oriental fruit fly). Elements of the sex-determination gene* doublesex *were identified in a number of* Bactrocera *species. A set of constructs has been made using control regions of the* doublesex *gene from* B. tryoni, *with a cell-death gene* grim, *which when stably genetically transformed into* B. tryoni *should function to kill female flies early in their development in the SIT facility.*

Funding:	**Horticulture Australia Ltd.** **Baxter Perpetual Trust** **FFRC (on behalf of voluntary contributors)**
Collaborating Institutions:	**None**

Disclaimer

Any recommendations contained in this publication do not necessarily represent current Horticulture Australia Ltd policy. No person should act on the basis of the contents of this publication, whether as to matters of fact or opinion or other content, without first obtaining specific, independent professional advice in respect of the matters set out in this publication..

Date of report: January 2005

CONTENTS

MEDIA SUMMARY 2
TECHNICAL SUMMARY 3
INTRODUCTION 4
1. Distribution and pest status 4
2. Mixed-sex versus single-sex sterile release strains 4
3. Genetic-sexing and genetically-modified SIT strains 4
4. Transformation systems and candidate genes 4
MATERIALS AND METHODS 6
1. Fly raising conditions/stocks 6
2. Expression of the Yp genes 6
3. Genomic DNA extraction 6
4. Total RNA extraction 6
5. Amplification conditions and cycles 6
6. Amplification with degenerate primers 7
7. Amplification of lethal gene constructs 7
8. RT-PCR 7
9. Inverse PCR 7
10. Fragment isolation, cloning and sequencing 8
11. Southern blot analysis 8
12. Sequence alignment 8
RESULTS 9
(1) Isolation of the *B. tryoni* (Q-fly) *Yolk protein* (*Yp*) intergenic region 9
(a) The *B. tryoni Yp1* and *Yp2* gene homologues 9
(b) The amino acid sequences of the *B. tryoni Yp* genes 9
(c) The gene structure of the *B. tryoni Yp* genes 9
(d) Putative regulatory sequences in the intergenic regions of *Yp1* and *Yp2* genes of *B. tryoni* 13
(2) Isolation of the *Yp* genes from representative Bactrocera species 15
(3) Isolation of the *dsx* genes from representative Bactrocera species 15
(4) Construction of *dsx-grim* lethal gene cassettes 19
DISCUSSION 22
(1) The *Yp* genes of *B. tryoni* 22
(2) The *Yp* genes of other Bactrocera species 23
(3) The *dsx* genes of other Bactrocera species 23
(4) Lethal gene constructs of *dsx* gene and the *D. melanogaster grim* gene 23
(5) Conclusions 24
TECHNOLOGY TRANSFER 24
RECOMMENDATIONS 24
ACKNOWLEDGMENTS 25
BIBLIOGRAPHY OF LITERATURE CITED 25

MEDIA SUMMARY

The Queensland fruit fly or Q-fly, *Bactrocera tryoni,* has been regarded as the most serious pest of horticulture in NSW and Queensland for many years. In spite of interstate quarantine measures (bait-spraying, trapping, exclusion zones, monitoring grids and quarantine policing), isolated outbreaks have occurred in Victoria, South Australia and Western Australia, a resident population has become established in Alice Springs and resident populations also exist in a number of inland towns around the Fruit Fly Exclusion Zone (FFEZ). The current pre-harvest control methods for Q-fly rely heavily on cover sprays, containing one of the systemic broad-spectrum insecticides which also destroy important beneficial insects of the orchard.

The Sterile Insect Technique (SIT) is a biological control method which has been widely adopted by industry to minimise the amount of chemical spraying. In this method, huge numbers of insects are reared, sterilised and released to mate with the wild insect pests, so that wild females produce no offspring and pest numbers are dramatically reduced. Overseas work on the related pest species, Med-fly, has shown that SIT is both cheaper and more reliable when sterile males only are released, rather than sterile flies of both sexes. Thus there is a demand from the horticultural industries for research to develop a male-only strain for Q-fly.

Because Australia is surrounded by countries which contain endemic fruit fly species with invasive pest status even greater than that of Q-fly, there is a need to develop male-only SIT systems that can be readily transferred to other species. The use of gene constructs that kill females in a fly-rearing factory will provide the best outcomes for transfer between species in the future.

This project has provided an investigation of sex-specific genes that might be used in female-killing constructs in Q-fly and related pest species, such as Oriental fruit fly. It has shown that the female-specific *Yolk protein* genes cannot be as readily adopted as the sex-determination gene *doublesex*. Accordingly, a set of constructs has been made using control regions of *doublesex*, with a cell-death gene *grim*, which should function to kill female flies early in their development in the SIT facility.

Future R&D requires the further development of stable, reliable genetic transformation systems for Q-fly, and a testing of the *doublesex* constructs for their efficacy in killing female embryos. A comparative investigation of other sex-determination genes in Q-fly, for their utility in male-only SIT would also be productive at this stage of research.

TECHNICAL SUMMARY

The need for improved protocols in the Sterile Insect Technique has long been recognised, and the development of male-only strains of the Mediterranean fruit fly, *Ceratitis capitata*, has been underway for the past 20 years. The initial *C. capitata* male-only strains utilised a serendipitously-isolated temperature-sensitive lethal (*tsl)* and Y-autosome translocations, to yield a strain in which females in the release broods are killed by a short, high-temperature treatment at the egg and early embryo stages. Tests of this strain from factories in South and Central America and from the FAO/IAEA laboratories in Austria have shown that the male-only strain is highly effective, both in species knock-down and in relative costs. Thus, there is an identified need to develop male-only strains for other fruit fly pests. In Australia, the particular need is related to the Queensland fruit fly, *Bactrocera tryoni*, but also other pests of the dorsalis or oriental fruit fly complex, such as *B. dorsalis*, *B. papayae*, *B. carambolae* and *B. phillipinnensis*. Realistically, there are not enough resources and time to develop a *tsl* system for even one of these species, so the development of female-killing constructs with conditional expression is the only solution.

This project was set up to provide data towards female-killing constructs. The aims were:
(1) To investigate the expression of the female-specific *Yolk protein* genes in *B. tryoni*, and to identify regulatory sequences in the bidirectional intergenic promoter, by comparison with the promoter region of the Drosophila *Yolk protein* genes.
(2) To isolate *Yolk protein* genes from pest species of the genus Bactrocera, for comparative purposes and for future utility.
(3) To isolate female-specific splicing-regulatory regions in intron 3 and exon 4 of the sex-determination gene, *doublesex*, from species of the genus Bactrocera. To compare the isolated sequences with previously analysed sequences from *B. tryoni* and *D. melanogaster*, in order to determine the conservation of sequences responsible for female-specific splicing
(4) To make female-specific lethal gene constructs, based on the *doublesex* female-specific splicing region or the yolk-protein promoter, plus a suitable lethal gene from Drosophila.

Analysis of expression of the *B tryoni Yolk protein* genes (*Yp1* and *Yp2*) showed that expression of *Yp2* is leaky in males and that the sequences are relatively poorly conserved, making identification of regulatory sequences difficult. Thus the *Yolk protein* regulatory regions could not be used for gene constructs without detailed functional analysis in the future. On the other hand, the *doublesex* gene proved to be highly conserved in the Bactrocera species tested (*B dorsalis*, *B. papayae* and *B. jarvisi*). Therefore a series of constructs were made using the intron and splicing signals of *doublesex* from *B. melanogaster* and *B. tryoni*, coupled with the cell-death gene, *grim*, from *D melanogaster*. These constructs await cloning for bulk preparation and testing by transformation into *D. melanogaster* and *B. tryoni.* The *D. melanogaster* constructs were made as a control, to test the system in a Dipteran for which reliable and stable gene transformation systems have been developed.

It appears that endogenous genes and promoters function better in non-drosophilid insect transformation systems than the *D. melanogaster* counterparts. Therefore future work to isolate and incorporate the *grim* homologue from Bactrocera is recommended before final decisions are made about the efficiency of the female-killing construct. Several types of female-specific lethal constructs will need to be tested before the most reliable systems could be used in the field and, indeed, more than one system may be required for fail-safe operation. Therefore, at this stage of research, we recommend that work be carried out on the utility of other genes of the sex-determination pathway in male-only SIT. It is known that the *transformer* gene can be used to produce male-only broods in *C. capitata* by RNA interference (Pane *et al*., 2002), but the dominant *Male determiner* has not been isolated or tested.

A further important recommendation to Horticulture Australia and the horticultural industries is that information about the use of genetic modification in SIT strains be widely disseminated, since the problems and advantages are very different to those of the well-known genetic modification of food crops. In SIT:
(1) the modified strain would be maintained in a fly factory, under conditions of stringent containment, more analogous to the widespread use of genetic modification in medicine.
(2) the only insects released are sterile males, so that transfer of genetic material to subsequent generations would not be a problem.
(3) only genes endogenous to the pest insects would be used so that, for example, predation of sterile males by birds would not result in their ingestion of any DNA different to that of their normal diet.

We propose that the production of male-only strains, by the introduction of female-lethal constructs into factory flies, should be approved for release at an earlier stage than current protocols on genetic modification of crops.

INTRODUCTION

1. Distribution and pest status

The Queensland fruit fly or Q-fly, *Bactrocera tryoni* (Froggatt), is a member of the family Tephritidae (true fruit flies) and is native to north-eastern Australia. This species utilises a wide host range of native plant families, in particular many rainforest species of Queensland, but has also easily adapted to nearly all the non-native fruits that are grown commercially. *B. tryoni* has a wide bioclimatic potential which has ensured its survival and aided its incursion into more temperate regions (Meats, 1981). The expansion of its range south is thought to have been coincident with the spread of horticulture in Australia such that it is now found along the entire eastern seaboard of Australia (Meats, 1981). Q-fly is a serious horticultural pest and the management of fruit fly infestations are of major concern, both in terms of associated costs and access to current and future export fruit markets. The current pre-harvest control methods for Q-fly rely heavily on cover sprays, although "greener" alternatives to cover sprays and more effective strategies of pest management are always being sought.

One method of non-chemically based control is the biological control method called the Sterile Insect Technique (SIT). This method is based on the concept that the release of sterile insects in sufficient numbers to overflood and out-compete the wild population, over a significant geographical area, will bring about a decrease in the size of the wild population (Knipling, 1955). SIT is the only environmentally-sustainable and species-specific biological method for area-wide control of fruit fly pests. SIT is an effective eradication, suppression and control method in areas that are currently fly-free and it has been used to successfully eradicate a number of introduced pest species such as the New World screwworm, *Cochliomyia hominovorrax*, from North America by 1982, and the Mediterranean fruit fly, *Ceratitis capitata* (Med-fly), from Chile and Mexico (reviewed in Robinson, 2002). In 1994, a successful SIT trial was carried out against Q-fly in the Murrumbidgee Irrigation Area (MIA), south-western New South Wales and one important finding of this exercise was that the cost was half that of current control methods (Horwood and Keenan, 1996).

2. Mixed-sex versus single-sex sterile release strains

The standard SIT strains (first generation sterile-release strains) are generally laboratory-reared insects that do not carry any genetic markings and produce both male and female progeny, which are sterilised by irradiation before release. The current belief is that SIT strains which can produce male-only progeny for release are more effective than mixed-sex strains. Studies on *C. capitata* found that when sterile flies of both sexes were released, the sterile males accounted for only 1/4 of the matings involving wild females, whereas, in single-sex releases, sterile males obtained >3/4 of the matings with wild females (Robinson *et al*., 1986). McInnis *et al*. (1994) found that the released males moved further from the site of release, as measured by their presence in bait traps, suggesting a heightened searching for mates rather than a greater attraction to the bait traps. A further advantage of male-only release is the predicted cost savings, both in rearing and transportation. Another concern in mixed-sex release is that the sterile females may still attempt to oviposit into fruit (sterile stings). This behaviour may lead to the introduction of bacteria and fruit rot fungi into the fruit (Cayol *et al*., 1994) which is of particular concern in pome fruit, tomato and stone fruit industries in Australia (Horwood and Keenan, 1996).

3. Genetic-sexing and genetically-modified SIT strains

Genetic-sexing strains are those strains that facilitate the large-scale separation of males from females. The strains are made by classical genetic methods involving the isolation of Y:autosome translocations, where the translocation carries a dominant wild-type allele for a selectable gene (Franz *et al*., 1994). The current genetic-sexing (male-only) strains in Med-fly have been constructed by conventional genetic mutagenesis and breeding programs using a temperature-sensitive lethal (*tsl*) mutation. Females can be eliminated at an early (embryonal) stage by subjecting the brood to high temperatures (Hendrichs *et al*., 1995). However, this protocol for generating male-only strains relies on serendipity (despite much subsequent effort, only one egg-stage *tsl* mutation has ever been isolated, by chance, only in Med-fly) and has taken about 20 years to develop. These systems based on chromosome aberrations also tend to be unstable and reduce the fitness of the insects, making them less effective agents for SIT (Robinson *et al*., 1999). A more rapid and targeted approach to the development of genetic-sexing or male-only strains involves the use of genetic transformation technology to modify strains with known gene constructs. When induced, these gene constructs will function to eliminate females from the brood. It is to be noted that in male-only SIT, modifications introduced into the strains include only genes specific to fruit flies. All genetically-modified fertile flies are maintained in fly factories under stringent containment conditions. Only sterile males are released so that, unlike the situation with some GM crops, transmission of any genetic changes to the next generation of wild flies is prevented.

4. Transformation systems and candidate genes

The means to integrate these gene constructs into the genome of the species of interest has been the major stumbling block to date as, until recently, transposon-mediated germline transformation of a dipteran species other than the

laboratory fly, *Drosophila melanogaster*, had not been reported. In recent years, transformation of Med-fly (Loukeris *et al*., 1995; Handler *et al*., 1998), has been achieved using the *piggyBac* (IFP2) transposable element from the cabbage looper, *Trichoplusia ni*, as part of a bipartite vector-helper system.

Candidate genes for the production of genetic-sexing strains include those genes which are expressed in a sex-specific manner, such as genes involved in sex determination (for example, the *doublesex* [*dsx*] gene), or in reproduction (for example, the *Yolk protein* [*Yp*] genes). The isolation of these genes was originally carried out in *D. melanogaster*. The *dsx* gene encodes sex-specific proteins (Ryner and Baker, 1991) and, in females, the production of the female-specific *dsx* protein is the result of sex-specific splicing of the *dsx* mRNA transcripts (Figure 1). The *Yp* genes (*Yp1* and *Yp2*) are expressed in a female-specific manner in the ovaries and fat body of adult *D. melanogaster* flies and this sex-specific expression has been shown to be regulated by a number of different proteins (Garabedian *et al*., 1985; Burtis *et al*., 1991; Lossky and Wensink, 1995; Søndergaard *et al*., 1995; Bownes *et al*., 1996).

One of the first isolations of a sex-determination gene in a non-drosophilid fly was *dsx* in Q-fly (Shearman and Frommer, 1998) and isolation of similar genes in other species is now proceeding. The Q-fly *dsx* homologue is expressed in a sex-specific manner as it is in *D. melanogaster* and splicing regulatory sequences have also been identified in the 3' untranslated region of exon 4 of this gene (Shearman and Frommer, 1998) similar to those depicted in Figure 1. Portions of the *Yp* genes have also been isolated from Q-fly (Shearman, 1999). The binding sites for the regulatory elements (as identified in *D. melanogaster*) should also be present and identifiable in the control regions of the Q-fly *Yp* genes. It should be possible to make a lethal-gene construct using the sex-specific regulatory elements of the *dsx* gene and a lethal gene, like the cell-death gene *grim*. This construct, when its expression is induced, should direct the splicing of the construct into a functional and lethal product only in females. Similarly, it should also be possible to incorporate the regulatory elements of the *Yp* genes into a construct that is only expressed in females.

In those pest tephritid species that have not been kept in culture for any length of time but are sufficiently close genetically to other species that have been studied in more detail (such as Med-fly and Q-fly), the rapid production of sterile-release strains may be possible with the knowledge gained from our research. The aims of the project are to isolate the *Yp* genes and regulatory regions for female-specific splicing of the *dsx* gene from pest fruit flies, relatives of Q-fly in the genus Bactrocera, and to make a series of constructs based on the *dsx*, the *Yolk protein* and lethal genes that should show female-specific expression.

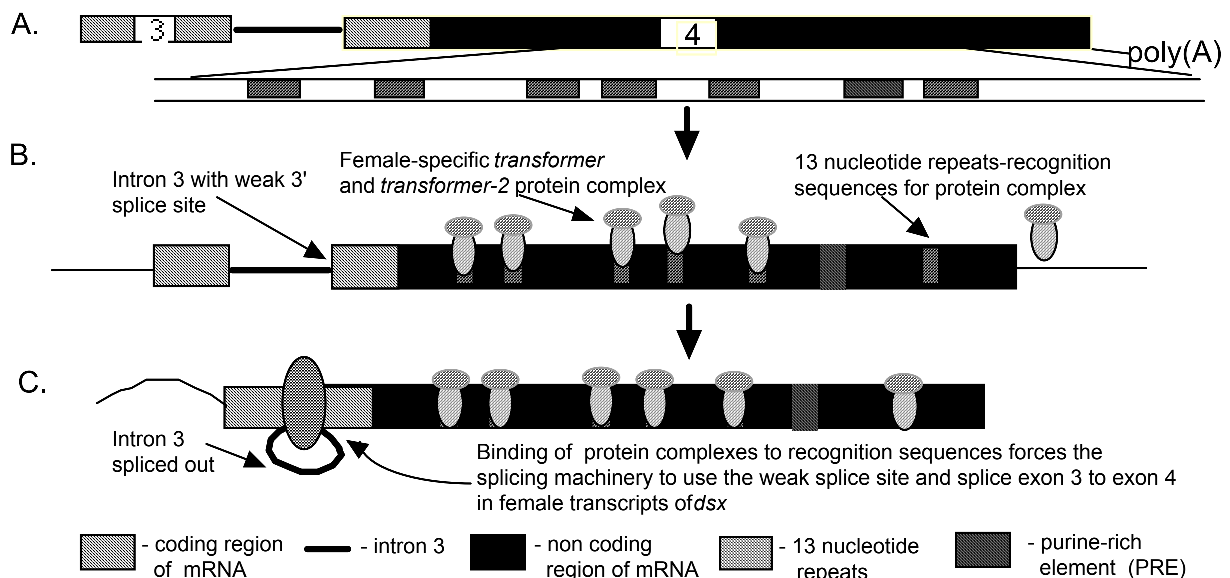


Figure 1. The mechanism of female-specific splicing of the *D. melanogaster dsx* gene
A. The *dsx* repeat element (*dsx*RE) of the *D. melanogaster dsx* gene lies within the non-coding regions of exon 4 and contains the splicing regulatory sequences: the 13 nucleotide repeats and the purine-rich element.
B. The protein complexes of the *transformer*/*transformer2* gene products (TRA/TRA-2) bind to the regulatory elements in the *dsx*RE.
C. The binding of the TRA/TRA-2 protein complexes forces the splicing machinery to use the weak 3' splice site of intron 3 resulting in the splicing of exon 3 to exon 4 (the female-specific exon).

MATERIALS AND METHODS

1. Fly raising conditions/stocks

A laboratory stock of *B. tryoni* flies has been maintained in cages over a period of up to 10 years under conditions of constant temperature (25°C) and natural light cycles (14L/10D). Flies were collected 2 weeks post emergence and snap frozen in liquid nitrogen if not used immediately. Stock flies have never exhibited any form of sex ratio or sexual phenotype anomalies.

2. Expression of the Yp genes

The laboratory stock of *B. tryoni* were sexed upon eclosion and separated into different treatments. The treatments were as follows:

Age (hours)		NE	24	48	98	192	336	504	672
	Sex	No. of flies per treatment							
No protein - Not mated	M	12	12	12	12	12	12	12	12
	F	12	12	12	12	12	12	12	12
Protein - Not mated	M	12	12	12	12	12	12	12	12
	F	12	12	12	12	12	12	12	12
No protein - Mated	M	ND	ND	ND	12	12	12	12	12
	F	ND	ND	ND	12	12	12	12	12
Protein - Mated	M	ND	ND	ND	12	12	12	12	12
	F	ND	ND	ND	12	12	12	12	12

NE – Newly emerged (0 hrs); ND – Not done

In the "No protein" treatment flies were fed water and sugar only from eclosion. In the "Protein" treatment flies were fed protein *ad libitum* as well as sugar and water from eclosion. Protein was in the form of 1 part yeast hydrolysate to 1 part sugar. "Not mated" male and female flies were kept in separate cages from eclosion, whereas males and females in the "Mated" treatments were caged together from eclosion. At the end of the treatment period flies were anaesthetised with carbon dioxide and then dissected into four parts – head, ovaries, abdomen (fat body) and thorax.

3. Genomic DNA extraction

Genomic DNA was extracted using a modified method of Bender *et al*. (1983). Briefly, ~1g of flies was ground to a powder under liquid N_2, mixed with the extraction buffer (100mM Tris-Cl, pH8.0; 50mM NaCl; 50mM EDTA, pH8.0; 200mM sucrose), Sarkosyl (0.5%) and Proteinase K (50µg/ml), and incubated at 55°C for 36 hr. RNase from bovine pancreas (Boehringer Mannheim) was added to 100µg/ml, mixed and the incubation continued at 37°C for 2 hr. This was followed by one extraction with 1 vol. Tris-buffered phenol and one extraction with 1 vol. chloroform. 0.1 vol. 3M NaOAc was added and 2.5 vol. ethanol (100%) layered onto the mixture and inverted several times without disrupting the layers. The DNA that precipitated at the interface was pipetted off, washed with 70% ethanol, air dried and resuspended in deionised H_2O. Genomic DNA was also extracted using the Promega Wizard Genomic DNA Extraction Kit according to the manufacturer's instructions.

4. Total RNA extraction

Total RNA was extracted from approximately 100 flies using the TRIzol reagent according to the manufacturer's instructions. Portions of dissected frozen flies were homogenised in 1ml TRIzol and incubated for 5 min at RT. The mixture was extracted with 1/5 vol choloroform then incubated at RT for 2-3 min, and centrifuged for 15 min at 4°C. The top layer was removed and 0.5ml isopropanol added and incubated for 10 min at RT. The mixture was centrifuged at 4°C for 10 min, the supernatant removed and the pellet washed with 1ml 75% ethanol by vortexing. The mixture was centrifuged at ≤ 7,500xg at 4°C for 5 min, the supernatant removed, the pellet air-dried for 5-10 min and dissolved in RNase-free deionised H_2O.

5. Amplification conditions and cycles

All amplification reactions were carried out in a Hybaid Omni Gene thermal cycler.
Standard thermal cycling reactions used the buffer supplied by the manufacturer (Biotech International: 50mM KCl, 10mM Tris, 0.001% gelatin) and also contained 2.0mM $MgCl_2$, 250µM each dNTP, 12.5pmol each primer, ~200ng genomic DNA or ~ 100pg PCR product and 1U Biotech *Tth* plus DNA polymerase in a total volume of 25µL. Standard (hot start) cycling conditions were as follows: 94°C 4 min, hold at 72°C while the *Tth* plus is added, then one cycle of 55°C 2 min and 72°C

2 min; 94°C 1 min, 55°C 1 min, 72°C 2 min for 3 cycles; 91°C 40 sec, 55°C 30 sec, 72°C 1 min for 28 cycles; 72°C 7 min for 1 cycle.

For screening of clones using vector-specific primers cycling conditions were as follows:
94°C 4 min, one cycle; 94°C 30 sec, 55°C 30 sec, 72°C 50 sec, for 30 cycles; 72°C 7 min for 1 cycle.

6. Amplification with degenerate primers

A portion of the *dsx* DNA sequence was amplified from genomic DNA using degenerate primers as listed below:

dsx-DT – 5' RGAGCATTGYCARAARCTATTRGAG
dsx-GT – 5' TRCGCAGCTCACCCCCGTCATA

Portions of the *Yp* gene were amplified from genomic DNA using degenerate nested primers as listed below:
Yp-A – 5' GCCAMYMRGAARYTRGTNCARGC
Yp-H – 5' GAATGTAGTCACCYTCCAAATCG

where Y = pyrimidine; R = purine; M = A or C; N = G, A, T or C

Amplification reactions using degenerate primers used the amplification reagents as described above, except that primer concentrations were increased to 25pmol/25µL reaction.

For the *Yp* primers the cycle conditions were as follows: 94°C 4 min, hold at 72°C while the *Tth* plus is added, then one cycle of 40°C 2 min and 72°C 2 min; 94°C 1 min, 40°C 1 min, 72°C 2 min for 3 cycles; 94°C 1 min, 55°C 30 sec, 72°C 1 min for one cycle; 91°C 30 sec, 60°C 1 min, 72°C 40 sec for 29 cycles; 72°C 7 min for 1 cycle.

For the *dsx* primers the cycle conditions were as follows: 94°C 4 min, hold at 72°C while the *Tth* plus is added, then one cycle of 40°C 2 min and 72°C 2 min; 94°C 1 min, 40°C 1 min, 72°C 2 min for 3 cycles; 94°C 1 min, 55°C 30 sec, 72°C 1 min for 1 cycle; 91°C 30 sec, 60°C 1 min, 72°C 40 sec for 29 cycles; 72°C 7 min for 1 cycle.

7. Amplification of lethal gene constructs

Amplification conditions for joining the fragments of the lethal-gene constructs were as follows: reactions used the buffer supplied by the manufacturer (Promega: 50mM KCl, 10mM Tris, 0.001% gelatin) and also contained 2.0mM $MgCl_2$, 250µM each dNTP and a stoichiometric ratio of 1:1 for each fragment up to a total of at least 0.5ng in the reaction in a total volume of 25µL. The reactions were set up with all reagents except the primers (12.5pmol each primer) and the polymerase (1U Promega Platinum Taq) which were added to the reaction as described below. The cycle conditions were as follows: 94°C 4 min, hold at 72°C while the Platinum Taq is added, then 55°C 30 sec and 72°C 3 min followed by 94°C 1 min, 55°C 1 min, 72°C 5 min for 5 cycles; hold at 72°C while the primer mix is added then 55°C 1 min, 72°C 1 min for 1 cycle; 91°C 50 sec, 60°C 30 sec, 72°C 2 min for 25 cycles; 72°C 7 min for 1 cycle.

8. RT-PCR

Total RNA was extracted from dissected tissues of females and males as described above. As an internal control for both the extraction procedure and the amplification reaction 2ng kanamycin RNA was included with each tissue sample. The mRNA portion of the total RNA was reverse transcribed to first strand cDNA using Superscript II Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. The resultant cDNA was used as a template for amplification with the following primers:

yp1RTx – 5' CCACAGTCGTCTCCTTCTGTTGC
yp1RTz – 5' CATTCGTTAGTCAGTTCGACGAGAG
yp2RTu – 5' GGAAACGCCGGCTGAGGAGG
yp2RTv – 5' CCTGTTCAGAGTCAGCGTCTTGAC
kan1F – 5' GCCATTCTCACCGGATTCAGTCG
kan2R – 5' ATCAGGTGCGACAATCTA

Cycles were as given above for standard hot start amplification reactions.

9. Inverse PCR

B. tryoni genomic DNA in 1.5µg lots in a total volume of 30µL were digested overnight with one of a number of different restriction enzymes in the appropriate buffer (eg. *Bam*HI, *Eco*RI or *Hin*dIII). The enzymes were heat inactivated at 85°C

for 20 min or according to the manufacturer's specification. For each of the ligation reactions the following reagents were combined on ice: 30µl of digested sample; 40µl of 10x ligation buffer (Roche) and 9.6 Weiss units of T4 DNA ligase (Roche) in a total volume of 400µl. The sample was mixed thoroughly by inverting, before an overnight incubation at 16°C. The resulting ligation mix was then ethanol precipitated ($1/10^{th}$ vol 3M NaOAc, mix, then add 2.5vols 100% EtOH and leave at –20°C for 30 min). Following this procedure, each sample ('IPCR ligate') was resuspended in 20µl of H_2O. As amplification products may be large, the DNA polymerase Elongase (Invitrogen) was used. Reactions used the buffers A and B supplied by the manufacturer (5X Buffer A [300mM Tris-SO_4, (pH9.1 at 25°C), 90mM $(NH_4)_2SO_4$ and 5mM $MgSO_4$] and 5X Buffer B [300mM Tris-SO_4, (pH9.1 at 25°C), 90mM $(NH_4)_2SO_4$ and 10mM $MgSO_4$] in a ratio that would give a final concentration of 1.5mM $MgCl_2$. Reactions also contained 1250µM each dNTP, 12.5pmol each primer, 1U ELongase DNA polymerase and ~35ng IPCR ligate (0.5µL) in a total volume of 25µL. Amplification conditions were as follows: 94°C 3 min , 72°C 30 sec, hold at 72°C and add Elongase, then 55°C 90 sec, 68°C 8 min for 1 cycle then 94°C 30 sec, 55°C 30 sec, 68°C 7 min for 24 cycles then 68°C 8 min for 1 cycle. Products were visualised on a 1% agarose gel and any distinct bands were excised from the gel and cloned as described in the next section.

Primers for these amplification reactions were as follows:
ypk 5' GACGCTGTCCATTGTAGCGCTGC
ypm 5' GTGCAACTCGTTACTTCGCCG
ypk2 5' CACTCGTCCGTTGTAGGCTTGC
ypl 5' GGAGCCCAAGCTGATGACCTAA

10. Fragment isolation, cloning and sequencing

Amplification products were excised from 1% agarose gels (Progen) and purified using the Wizard gel extraction kit (Promega) according to the manufacturer's instructions. pBluescript vector DNA (Stratagene) was T-tailed using the method of Marchuk *et al.* (1991), ligated to the amplification products using T4 DNA ligase (Invitrogen) according to the manufacturer's instructions in the buffer supplied. Recombinant plasmids were transformed into DH5α competent cells which were prepared by the method of Inoue *et al.* (1990). To screen transformants for plasmid inserts colonies were picked from agar plates, eluted in 75µL deionised water, boiled for 5 min then used as template in PCR using vector-specific primers under amplification conditions listed above. Plasmids from positive clones were isolated by using a plasmid isolation kit (Promega) according to the manufacturer's instructions. Automated sequencing was carried out on an ABI systems 373A automated DNA sequencer at the SUPAMAC sequencing facility or by Macrogen (Korea).

11. Southern blot analysis

Southern blotting was carried out essentially as described in Sambrook *et al.* (1988) and according to the standard protocol recommended in the Zeta-Probe instruction manual (Biorad). Genomic DNA (10µg), digested with *Bam*HI, *Sal*I or *Hin*dIII, was electrophoresed on a 0.8% agarose gel and blotted onto Hybond N+ nylon membrane. The membrane was prehybridised in 1mM EDTA, 0.5M NaH_2PO_4, pH7.2, 7% SDS for 5 min at 65°C. Probes were prepared using an endonuclease restriction fragment of both the *Yp1* and the *Yp2* genes as template. The fragments were purified from agarose (as described above), eluted in 10µL deionised H_2O and the final concentration of template determined by gel electrophoresis on a 1% agarose gel against known DNA standards. Probes were radioactively-labelled with α-^{32}P-dCTP, using a random primed DNA labelling kit (Roche) according to the manufacturer's instructions. The probe quality was checked by electrophoresis of 0.1 vol. and 0.05 vol. of the labelled probe on a 1% agarose gel, then the gel photographed. Probe was added and the membrane hybridised for 20 hr at 65°C. The membrane was washed twice in 1mM EDTA, 40mM $NaHPO_4$, pH7.2, 5% SDS for 30 min at 65°C. The membrane was then washed twice in 1mM EDTA, 40mM $NaHPO_4$, pH7.2, 1% SDS for 30 min at 65°C. The membranes were exposed to a phospor screen for 2 hours and analysed using a Biorad phosphoimager.

12. Sequence alignment

Programs Pileup and Bestfit in the GCG suite of programs and the BLASTX program were used for DNA and protein sequence alignments.

RESULTS

The choice of species of fruit flies to be investigated was determined by the availability within the collection of the FFRC of suitable specimens from which reasonable quality DNA could be extracted. *B. dorsalis* (Oriental fruit fly) and *B. papayae* were chosen as representatives of the dorsalis complex within the subgenus Bactrocera (which also includes the tryoni complex). Both these species are serious pests in SE Asia as well as other parts of the world where they have invaded. *B. jarvisi* was chosen as representative of the subgenus *Afrodacus*. Although not considered as serious a pest as the former two species, *B. jarvisi* will infest commercial hosts such as mangoes and bananas.

(1) Isolation of the *B. tryoni* (Q-fly) *Yolk protein* (*Yp*) intergenic region

In *D. melanogaster,* the major protein components of the yolk storage granules found in eggs are encoded by the *Yolk protein* (*Yp*) genes. *Yp* gene expression is first detected around 24 hr after eclosion but maximum levels of *Yp* gene expression are only seen after the ingestion of protein. The sex-specific expression of the *Yolk protein* (*Yp*) genes is controlled at the level of transcription and, in females, the *Yp* genes (*Yp1*, *Yp2*, and *Yp3*) are expressed not only in the ovary but also in the fat body, a tissue common to both sexes (Belote *et al*., 1985). This sex-specific expression is controlled through a number of regulatory sequences which have been identified in the regions upstream of the coding regions (Garabedian *et al*., 1985; Burtis *et al*., 1991; Lossky and Wensink, 1995; Søndergaard *et al*., 1995; Bownes *et al*., 1996). The isolation of these gene sequences and the identification of potential regulatory sequences from Bactrocera species will be of use in the future construction of sex-specifically expressed transgene constructs in male-only strains.

(a) The *B. tryoni Yp1* and *Yp2* gene homologues

In *B. tryoni*, fragments of the *Yp1* and *Yp2* gene had been isolated previously (Shearman, 1999). These sequences were used to design further primers for inverse PCR which was carried out to isolate the 5' regions of the gene as well as the intergenic region as shown in Figure 2. The 3' regions of the genes have not been isolated. Sequence analysis of the clones of the 5' regions of the *Yp* genes revealed there were two alleles for each of the *Yp* gene homologues, *Yp1* and *Yp2,* and these have been designated *Yp1β, Yp2α, Yp1γ* and *Yp2δ*, following convention of differentiating these alleles by using Greek symbols in the Med-fly, *Ceratitis capitata* (as given in Appendix 1). At the DNA sequence level and within the protein coding region of sequence, the alleles of each of the *Yp1 and Yp2* genes are almost identical except for a few base differences (as shown in Appendix 1). Of these base changes only two differences change the amino acid sequence in the *Yp2* genes. The coding regions of the *B. tryoni Yp* genes show the same bias for an A or T in the third base position of the codon as previously found in the *dsx* gene (Shearman and Frommer, 1998) and the *white* gene (Bennett and Frommer, 1998) suggesting that this bias may be characteristic of a diverse range of *B. tryoni* coding sequences.

(b) The amino acid sequences of the *B. tryoni Yp* genes

The conceptual translations of the fragments of the *B. tryoni Yp1* and *Yp2* gene alleles were aligned. There were five nucleotide changes with no amino acid changes between the *Yp1* alleles and seven nucleotide changes with three amino acids changes between the two alleles. The amino acid sequences of the *B. tryoni Yp* genes were aligned with the amino acid sequences of the *C. capitata* yolk protein sequences, VG1 and VG2, the *D. melanogaster* yolk protein sequences, YP1 and YP2, using the PileUp program in the GCG suite of programs (Figure 3). The amino acid matches across the *Yp1* genes are indicated by shading, across the *Yp2* genes by shading and across all gene sequences by an asterisk (*) between the sequences which also includes conservative substitutions. Pairwise comparisons were made between the protein sequences of the two *B. tryoni* genes, the three *D. melanogaster* genes and the two *C. capitata* genes and the percent identity between pairs of sequences across the carboxy terminal region is given in Table 2. As shown, each of the *Yp* protein homologues has a higher degree of identity with its interspecies counterpart than with its intraspecies counterpart, especially across the carboxyl region of the protein. For example, the *B. tryoni Yp1* (Btyp1) protein sequence is 84% identical to the *C. capitata Vg1-γ* (CcVg1) protein sequence whereas it is only 61% identical to the *B. tryoni Yp2* (Btyp2) protein sequence. The *B. tryoni Yp1* homologues show a higher degree of identity to the *D. melanogaster Yp1* protein sequence and also a significant degree of identity with the *D. melanogaster Yp3* gene at both the DNA and protein level.

(c) The gene structure of the *B. tryoni Yp* genes

The gene structure and the distances (in DNA base pairs) between each of the *Yp1* and *Yp2* genes in *B. tryoni* were determined by Southern blotting, as shown in Figure 4, although the exact orientation and the distance between each gene

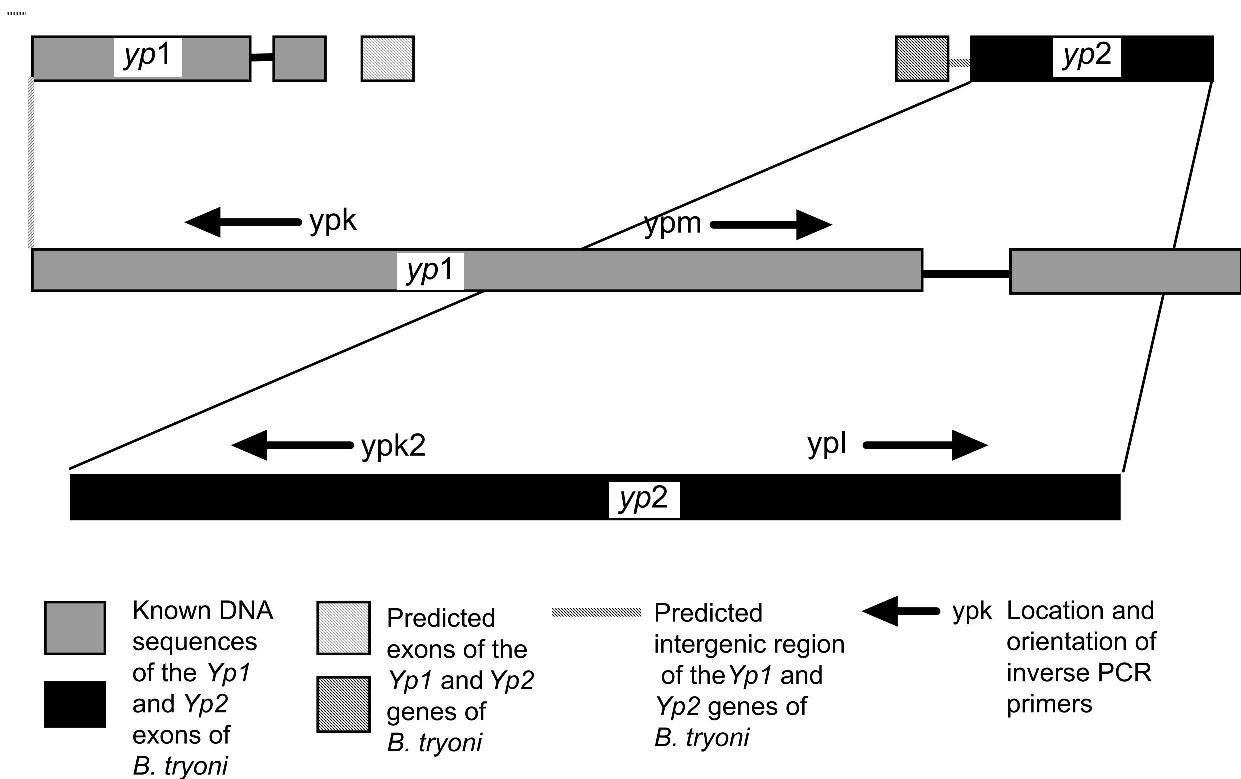


Figure 2. The *Yp* gene structure of *B. tryoni* and location of inverse PCR primers

	Btyp1	Btyp2	Bcyp1	Bcyp2	Ccyp1	Ccyp2	Dmyp1	Dmyp2	Dmyp3
Btyp1	100%								
Btyp2	57%	100%							
Bcyp1	75%	57%	100%						
Bcyp2	57%	89%	57%	100%					
Ccyp1	84%	61%	76%	58%	100%				
Ccyp2	55%	89%	54%	87%	59%	100%			
Dmyp1	61%	55%	55%	52%	62%	52%	100%		
Dmyp2	57%	58%	54%	53%	48%	57%	57%	100%	
Dmyp3	60%	58%	57%	54%	60%	51%	60%	57%	100%

Table 2 – Percent identity between carboxyl terminal regions of the protein sequences of each of the Yp genes of *B. tryoni* (Bt); *C. capitata* (Cc) and *D. melanogaster* (Dm)

```
Bt1 MNPLKIFCFMAFVLAIA...NASPKHGKNKDNV..NSLKPVDWLSASELESMPSVEDITLQQLENMSMEDAQRKIEKLYHLSQINHALEPSFVPSPSNVPVILMKPNGQPERTNL
Cc1 MNPLKIFCFLALVIAVA...SAN.KHGKNKDNAGPNSLKPTDWLSVEELQSMTAIDDITLQQLENMSVEDAERKIEKIYHLSQINHALEPSYVPSPSNVPVMLMKPNGQSQQTNH
Dm1 MNPMRVLSLLA.CLAVA....ALAKPNGRMDNSVNQALKPSQWLSGSQLEAIPALDDFTIERLENMNLERGAELLQQVYHLSQIHHNVEPNYV..PSGIQVYVPKPNGDKTVAPL
Dm3 MMSLRI.CLLATCLLVA...AHASKDASN......DRLKPTKWLTATELENVPSLNDITWERLENQPLEQGAKVIEKIYHVGQIKHDLTPSFVPSPSNVPVWIIKSNGQKVECKL
    *         *                    *       *  *               *   *     *         **  *      *   ***   *        *
Bt2 MSPLSIFCLVALLA........TATPVCARGNSIRDNLKPTEWISPRELENAPSVDEITFEKLQETPAEEAAELVNQIYHLSQLSRKIEPSYAPSPSDIPVYTYTPTGQRVNTKL
Cc2 MNPLTIFCLVAVLL........SAATAHRGSNAIRNNLQPSGXLSPRELEDMPAINEITFEKLQEMPAEEAADLVNKIYHLSQMSRNIEPSYAPSPNQIPAYTYTPTGQRVNFNL
Dm2 MNPLRTLCVMACLLAVAMGNPQSGNRSGRRSNSLDNVEQPSNWVNPREVEELPNLKEVTLKKLQEMSMEEGATLLDKLYHLSQFNHVFKPDYTPEPSQIRGYIVGERGQKIEFNL
    1                                                                              #                             125

Bt1 NELVQTAKQQPNFGDEEVTIFITGLPQS.SPSVAKANKKLIQAYMQRYNGQRQ...PISNNQDY.DYGNNKDNQGATSSEEDYSESWK....NPKPTKGNLVVISLGSTLTNMKR
Cc1 NELVEAAKQQPNFGDEEVTIFITGMPQT.SSAVLKANKKLVQAYMQRYNGQQQ...PINGNKDY.DYGSSQGNQGATSSEEDYSESWK....NQKSTKGNLVIINLGSTLTNMKR
Dm1 NEMIQRLKQKQNFGEDEVTIIVTGLPQT.SETVKKATRKLVQAYMQRYNLQQQRQHGKNGNQDYQDQSNEQRKNQRTSSEEDYSEEVK....NAKTQSGDIIVIDLGSKLNTYE
Dm3 NNYVETAKAQPGFGEDEVTIVLTGLPKT.SPAQQKAMRRLIQAYVQKYNLQQLQKNAQEQQQQLKS.....SDYDYTSSEEA.ADQWK....SAKAASGDLIIIDLGSTLTNFKR
    *  *     *  **  ****   * *         *   *****  * **                              ****                 *  ***  **
Bt2 NQLVSTVQQQPHFGQQEVTIFITGLPQQKLRDAARANKKLIQAYLQAYNGRVQVKGGQDAD.........SEQDTSSSEE....TS...NSKQTQPSGNLVVIDLGAVIRNFEE
Cc2 NQLVATAQQQPNFGKQEVTVFITGLP.NKSSAMLTANQKLVQAYLQAYNGRVQVQGEQGDD.........SNQDTSSSEE....SSNRPNGQQPKPNGNLVVIDLGAVIRNFED
Dm2 NTLVEKVKRQQKFGDDEVTIFIQGLPETN.TQVQKATRKLVQAYQQRYNLQPYETTDYSNE.........EQSQRSSSEEQQTQRRKQNGEQDDTKTGDLIVIQLGNAIEDFEQ
    126                                                                                                 #         250

Bt1 LALIDVEQTGNMIGKALVELTNECDVPQEIIHIVGQGVGAQVAGAAGRQYKRLTGHQLRRITALDPAKLFAKDKDMLTGLARGDADFVDAIHTSTCGMGTRQRVGDVDFYVNGPA
Cc1 FALLDVEQTGNMIGKTLVQLTNEVDVPQEIIHIVAQCIGAQVAGAAGRQYKRLTGHQLRRITALDPSKIFAKNRNALTGLARGDADFVDAIHTSTCGMGTRQRVGDVDFYVNGPA
Dm1 YAMLDIEKTGAKIGKWIVQMVNELDMPFDTIHLIGQNVGAHVAGAAAQEFTRLTGHKLRRVTGLDPSKIVAKSKNTLTGLARGDAEFVDAIHTSVYGMGTPIRSGDVDFYPNGPA
Dm3 YAMLDVLNTGAMIGQTLIDLTNK.GVPQEIIHLIGQGISAHVAGAAGNKYTAQTGHKLRRITGLDPAKVLSKRPQILGGLSRGDADFVDAIHTSTFAMGTPIRCGDVDFYPNGPS
         *   *  **   *        *   *        * ***  *       **   *****  **  *        * **  **  * ********   ***  *   ***  ***
Bt2 LVLLDINRVGAAIGNSLVQLTSQTDVPQEVIYIVAQGIGAHVAGAAARQYTRQTGNKLRRITAMDPTKIFARKPNTLVGLARGNADFVDAIHTSAYGLGSAARAGDVDFYPNGPS
Cc2 LVLLDINRVGAAIGNSLVQLTAQADVPQEVINIVAQGIAAHVAGAAARQYTRQTGNTLRRITAMDPSKIYARKPNTLVGLARGNADFVDAIHTSAYGLGTTTRAGDVDFYPNGPS
Dm2 YATLNIERLGEIIGNRLVELTNTVNVPQEIIHLIGSGPAAHVAGVAGRQFTRQTGHKLRRITALDPTKIYGKPEERLTGLARGDADFVDAIHTSAYGMGTSQRLANVDFFPNGPS
    251                                                                                                            375

Bt1  SAAPGATNVIEATMRATRYFAESVRPGNERNFPAVAANSMDQYENNDGAGKRVYMG..................................................
Cc1  STAPGTNNVVEASMRATRYFAESLRPGNERNFPAVAANSLNQYENNEGNGKRAYMGIATDFDLEGDYILKVNPKSPFGKSAPAQKQRRYHGLHQSWKSGKNQNQE
Dm1  AGVPGASNVVEAAMRATRYFAESVRPGNERSFPAVPANSLQQYKQNDGFGKRAYMGIDTAHDLEGDYILQVNPKSPFGRNAPAQKQSSYHGVHQAWNTNQDSKDYQ
Dm3  TGVPGSENVIEAVARATRYFAESVRPGSERNFPAVPANSLKQYKEQDGFGKRAYMGLQIDYDLRGDYILEVNAKSPFGQRSPAHKQAAYHGMHHAQN........
        **  ****  *****  **   **  * **  * * *   *    *  ***  ***
Bt2  VAMPGTDNIIEASLRATRYFAETVRPGNDRNFPAVAAESLQQYKNNNGNGRRAYMGIAADYDLEGDYILQVNAKTPF..............................
Cc2  VNMPGTDDIIEASLRATRYLAETVLPGNDRNFPAVAAESLQQYKNNNGNGRRAYMGIAADYDLEGDYILQVNAKSPFGKSAPAQKQNSYHGIHQGAGRPN.....
Dm2  TGVPGADNVVEATMRATRYFAESVRPGNERNFPSVAASSYQEYKQNKGYGKRGYMGIATDFDLQGDYILQVNSKSPFGRSTPAQKQTGYHQVHQPWRQSSSNQGSRRQ
     376                                                                                                  451
```

Identical amino acid residues in *Yp1/3* across three species

Identical amino acid residues in *Yp2* across three species

* Identical amino acid residues in *Yp1*, *Yp2*. and *Yp3*

\# Intron/exon boundary

Bt – *B. tryoni*; Dm – *D. melanogaster* Cc – *C. capitata* **Bt1 – *B. tryoni* yolk protein 1 (yp1) sequence**

Figure 3. Alignment and comparison of the protein sequences of the yolk proteins of *D. melanogaster*, *C. capitiata* and *B. tryoni*

The figure shows the alignment of the protein sequences of the yolk proteins from three species using the computer program Pileup. Boxed segments indicate regions of identity between sequences.

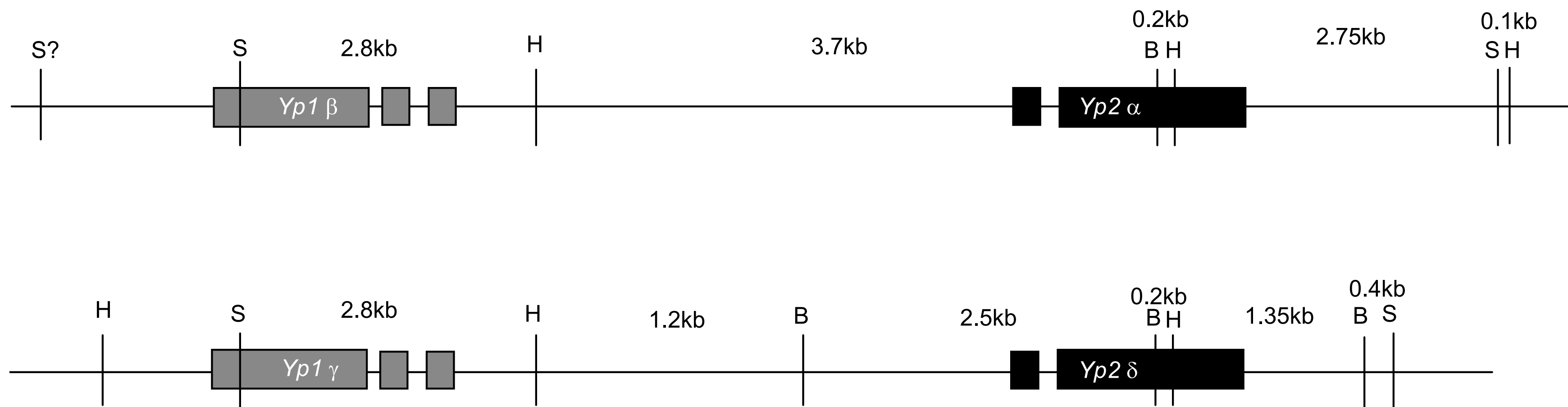


Figure 4. Restriction map of *yp1* and *yp2* gene regions of *B. tryoni*

Distances between restriction sites is indicated in kilobase pairs (kb). B - *Bam*HI; H - *Hin*dIII; S - *Sal*I

pair have not been determined. The chromosomal organisation of the *B. tryoni Yp1* and *Yp2* genes was found to be the same as that of the *D. melanogaster* and the *C. capitata* gene pairs. The *B. tryoni Yp1* gene homologues were found to have two introns while the *Yp2* homologues have only one intron. The *B. tryoni Yp1* gene structure is the same as that found in the *D. melanogaster Yp3* gene and the *C. capitata Vg-1* genes, but differs to the *D. melanogaster Yp1* gene, which contains only one intron. The *B. tryoni Yp2* genes have the same structure as the *D. melanogaster Yp2* and *C. capitata Vg-2* genes (Figure 4). Based of the alignment of the amino acid sequences of the *Yp* genes (Figure 3) the introns in the *Yp1* homologues of *B. tryoni* lie in the same location as the introns of the *C. capitata Vg1* and the *D. melanogaster Yp3* genes. The single intron of the *B. tryoni Yp2* genes lies in the same location as the single intron of the *D. melanogaster Yp2* and the *C. capitata Vg-2* genes as indicated in Figure 3.

The region between the *Yp* genes, the intergenic region (IGR), was isolated by inverse PCR, sequenced (Appendix 1) and the regions from both pairs of genes were compared. In *B. tryoni* this region was found to be 4,600 base pairs (bp) long whereas in *D. melanogaster* the corresponding region is around 1,300bp long. The IGR of the *B. tryoni Yp* genes is similar in size to the same region of the *C. capitata Yp* genes, although the region in *C. capitata* has not been sequenced. The non-coding regions of both these pairs of genes in *B. tryoni* also show a high degree of DNA sequence identity with only 70 base changes and 10 insertion/deletions (indels) within this 4.6kb region of sequence (Appendix 1).

(d) Putative regulatory sequences in the intergenic regions of *Yp1* and *Yp2* genes of *B. tryoni*

A number of both sex-specific (GATAβ, OE1, OE2 and *dsx*) and non sex-specific control elements (Lossky and Wensink, 1995; Logan and Wensink, 1990; Chung *et al*., 1996; Burtis *et al*., 1991) have been identified within the IGR region of the *D. melanogaster Yp1* and *Yp2* genes (Figure 5). Given that the *Yp1* and *Yp2* genes of *B. tryoni* lie in the same orientation as those of *D. melanogaster*, it is expected that regulatory sequences, which are located within the intergenic region and control the co-ordinated expression of these genes in *D. melanogaster*, are likely to be conserved in *B. tryoni.* However the arrangement and orientation of these regulatory sequences may differ between the two species, as has been found for the regulatory sequences in other Drosophila species such as *Drosophila grimshawi* (Piano *et al*., 1999).

The IGR of the *B. tryoni Yp* genes was searched for sequences with homology to the fat body- and ovary-specific regulatory elements of *D. melanogaster* and of other dipteran species. The location of these putative regulatory sequences is shown in Figure 5. Not all the regulatory regions identified in *D. melano*gaster can be easily identified in *B. tryoni*, as the sequence is quite divergent, although the *B. tryoni* sequences do appear to represent a number of the regulatory element binding sites. The location of the AEF-1/*dsx*/C/EBPbinding site does look like a significant match whereas the *ref1* protein binding site and ecdysone response elements do not look like particularly good matches, but this has yet to be proven.

(e) Regulation of expression of the *Yp* genes in *B. tryoni* by protein feeding

In *D. melanogaster*, the level of the *Yp* gene expression relies on the ingestion of protein, with only minimum levels of gene expression being detected in females that are not given access to protein. Maximum levels of *Yp* gene expression are detected in females that have been fed protein from eclosion and these levels reach a peak at around day 4 after eclosion.

The effect of protein feeding and starvation on the expression of the *Yp* genes and the maturation of the ovaries in *B. tryoni* females was investigated in the wild type lab stock Postharvest. The relative levels of expression of the *Yp* genes in females that had been fed protein and those that had been denied protein were determined by a technique called RT-PCR (reverse transcription – PCR) and the results are summarised in Table 3. As can be seen in the table a low level of *Yp* expression in both protein-fed and protein-starved females can be detected after 24 hours (post emergence). In the protein fed females this level increases over the next three days until it reaches a maximum level around eight days post emergence. There appeared to be a small amount of *Yp2* mRNA present in males 24 hr after emergence, although this amount did not increase and disappeared after a few days (data not shown).

The effect of protein feeding on ovarian maturation and egg development was investigated in *B. tryoni* in females that were 21 days old. The dramatic differences between the abdomens and ovaries of females of the two treatments are shown in Figure 6. The abdomens of protein-fed females appear more distended (due to the presence of many mature eggs) and are darker in colour compared to those of the protein-starved females (as shown in Figure 6a). The presence of many mature eggs is obvious in the ovaries of protein-fed females (Figure 6b). The ovaries of females that were protein starved did not develop when compared to ovaries of protein-fed females and no enlarging ovarioles can be seen (Figure 6b). The ovaries of protein-starved females were identical to those of newly-emerged females (not shown).

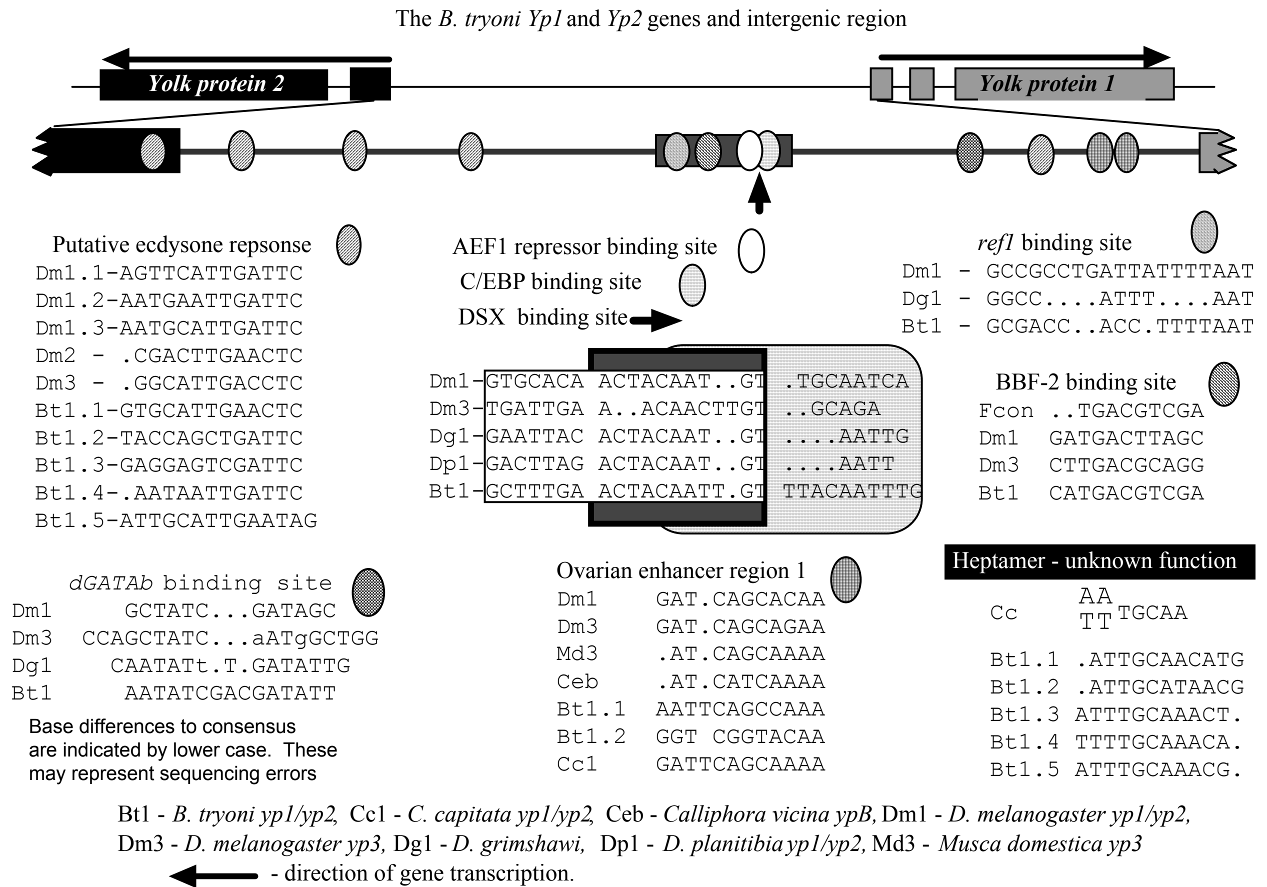


Figure 5. Putative regulatory sequences and their locations within the intergenic region of the *Yp* genes of *B. tryoni*
Regions that show some similarity in either sequence or location with the *D. melanogaster* regulatory elements were identified in *B. tryoni.*

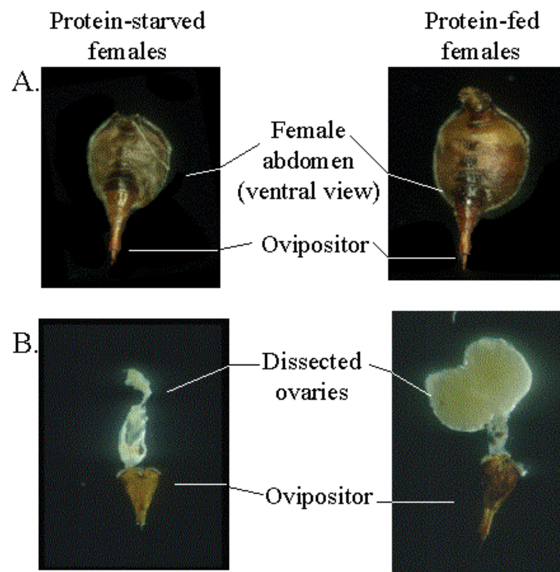


Figure 6. A domens (a) and ovaries (b) of protein-starved and protein-fed female *B. tryoni.*

Age (hours)		NE	24	48	98	192	336	504	672
	Gene	Level of expression							
No protein - Not mated	*Yp1*	-	-/+	+	+	+	+	+	+
	Yp2	-	+	+	+	+	+	+	+
Protein - Not mated	*Yp1*	-	-/+	+	++	+++	+++	+++	+++
	Yp2	-	+	++	+++	+++	+++	+++	+++
No protein - Mated	*Yp1*	ND	ND	ND	+	+	+	+	+
	Yp2	ND	ND	ND	+	+	+	+	+
Protein - Mated	*Yp1*	ND	ND	ND	+++	+++	+++	+++	+++
	Yp2	ND	ND	ND	+++	+++	+++	+++	+++

NE – newly emerged; 192 = 8 days; 336 = 14 days; 504 = 21 days; 672 = 28 days
ND – not done; **-** Gene expression not detected; + - low level of *yp* gene expression;
++ - medium level of *yp* gene expression; +++ - high (maximum) level of *yp* gene expression
Table 3 – Relative levels of *yp* gene expression in protein-fed and protein-starved *B. tryoni* females

(2) Isolation of the *Yp* genes from representative Bactrocera species

The DNA sequences of the *Yp* genes of *B. tryoni* and *D. melanogaster* were aligned and degenerate primers were designed to regions around the most conserved portion of the sequence, that is the carboxyl two-thirds of the protein. Good quality genomic DNA was available in the lab for *B. dorsalis* and this species was tried first. Portions of the *Yp1* and *Yp2* genes were amplified from *B. dorsalis* using the degenerate primers yp-A and yp-H and the fragments were cloned. Three clones of each gene were sequenced and the consensus DNA sequences aligned with those of the *B. tryoni Yp1* and *Yp2* genes, as shown in Figures 7 and 8. It is not possible to tell if there are two alleles of each of the *Yp1* and *Yp2* genes in *B. dorsalis* as not enough clones of each gene were obtained. The *B. dorsalis Yp2* gene shows a very high degree of identity to the *B. tryoni Yp2* genes at the DNA level with only 14 base changes out of 681 bases between them. The level of identity for the *Yp1* genes is quite low in comparison with the *Yp2* genes, with 193 base changes out of 775 bases across the exon as well as 1 indel (insertion/deletion) of 3 bases and 1 indel of 9 bases differing between the species.

The DNA sequence of the *B. dorsalis Yp* genes was translated into protein sequence and this was compared to the sequences of the *Yp* proteins of *B. tryoni* and *C. capitata* as shown in Figures 9 and 10. As expected, based on the similarity of the DNA sequence, the *B. dorsalis Yp2* amino acid sequence is almost identical to that of the *B. tryoni Yp2* sequence except for five amino acid differences. The *B. dorsalis Yp1* sequence shows 50 amino acid differences compared to the *B. tryoni Yp1* sequence across 256 amino acids.

DNA was extracted from *B. papayae* and *B. jarvisi* using the Promega Wizard genomic DNA extraction kit. Fragments of the *Yp1* gene of *B. jarvisi* and the *Yp2* gene of *B. papayae* were amplified from this genomic DNA. Although several attempts were made to amplify the *Yp1* gene from *B. papayae*, this has not been successful to date. Similarly, the *Yp1* gene was difficult to isolate from *B. tryoni* and did not always amplify to the same level as ~~that of~~ the *Yp2* fragments. Whether this inability to amplify the *Yp1* gene from *B. papayae* points towards a difference in the genome of *B. papayae* compared to that of the closely-related species *B. dorsalis* or whether this merely indicates a difference in the quality of the DNA that was obtained for each species is not clear. Again the cloning of these fragments proved difficult, with only three clones of the *B. jarvisi Yp1* gene and two clones of the *B. papayae Yp2* gene being isolated. These clones were sequenced and the sequences are shown in Figures 7 and 8. The *B. jarvisi Yp1* sequence differs from the *B. tryoni Yp1* sequence by 190 base changes out of 775 bases across the exon as well as 1 indel of 3 bases and 1 indel of 9 bases. The *B. jarvisi* sequence differs from the *B. dorsalis* sequence by only 16 bases out of 775 bases. The *B. papayae Yp2* sequence differs from *B. tryoni* by 14 base changes out of 681 bases. The *B. papayae* sequence differs from the *B. dorsalis* sequence by 8 bases out of 681. The sequences from *B. jarvisi* and *B. papayae* were translated into protein and these were compared to the *B. tryoni Yp1* and *Yp2* sequences (as shown in Figures 9 and 10). The *B. jarvisi Yp1* sequence differs from the *B. tryoni Yp1* sequence by 59 amino acids and from the *B. dorsalis* sequence by seven amino acids. The *B. papayae Yp2* sequence differs from the *B. tryoni Yp2* sequence by 6 amino acids and from the *B. dorsalis* sequence by one amino acid.

(3) Isolation of the *dsx* genes from representative Bactrocera species

Whereas the *Yp* genes are only expressed in one sex, the *dsx* gene is expressed in both sexes and encodes sex-specific gene products. The sex-specific splicing of the *dsx* pre-mRNA is controlled through the binding of the female- specific

```
Btyp1β ........GA AATTGGTTCA AGCTTATATG CAGCGCTACA ATGGACAGCG TCAACCAATC A...GCAATA ACCAGGACTA TGACTATGGC AACAATAAGG
Btyp1γ GCAAACAAAA AATTGATCCA GGCTTATATG CAGCGCTACA ATGGACAGCG TCAACCAATC A...GCAATA ACCAGGACTA TGACTATGGC AACAATAAGG
Bdyp1  GCCAACAGGA AATTGGTACA GGCTTACATG CAACGCTACT ACGGACAACA GCAGCCAGTG AATGCCAACA ACCAGGATTA TGATT..... ....CTAATG
Bjyp1  GCCACCAGGA AATTGGTGCA GGCTTACATG CAACGTTACT ACGGACAACA GCAGCCAGTG AATGCCAACA ACCAGGATTA TGATT..... ....CTAATG

       101                                          150 151                                          200
Btyp1β ACAATCAAGG TGCCACTTCA AGTGAAGAAG ACTACAGTGA ATCGTGGAAG AATCCCAAAC CCACAAAGGG CAACCTTGTG GTAAGTTTTA ACACCAAATG
Btyp1γ ACAATCAAGG TGCCACTTCA AGTGAAGAAG ACTACAGTGA ATCGTGGAAG AACCCCAAAC CCACAAAGGG CAACCTTGTG GTAAGTTTTA ACACCAAATG
Bdyp1  AGCGTCAGCA AGCCACTTCC AGCGAAGAGG ACTACAGCGA ATCGTGGAAG CAGCAAAAAT CGAACAGAGG CAATCTTGTC GTAGGTTT.. ..........
Bjyp1  AGCGTCAGCC AGCCACTTCC AGCGAAGAGG ACTACAGCGA ATCGTGGAAG CAGCAAAAAT CGAACAGAGG CAATCTTGTC GTAGGTTT.. ..........

       201                                          250 251                                          300
Btyp1β TGTATTTTAT GATTTTTCTA TCATTATCTT TAACTATTTT CTGTTTATTC ATTTAGGTCA TCAGCTTGGG CTCCACCCTC ACGAACATGA AACGTTTAGC
Btyp1γ TGTATTTTAT GATTTTTCTA TCATTATCAT TAACTATTTT CTGTTTATTC ATTTAGGTCA TCAGCTTGGG CTCCACCCTC ACCAACATGA AACGTTTAGC
Bdyp1  TCTACTCG.A AAAAATTTTA CCTTCGTCCA AAAAGTTTTC CTTTT..... ...TAGATCA TTAAATTGGG CGCTGTCTTG ACGAACTTTA AACGCTATGC
Bjyp1  TCTACTCGTA CGAAACCCGA GCTTCTTCTA AAAAGGTTTC TTCTTTCTCC A..TAGATAA TCAAATTAGG CGCTGTCTTG ACGAACTTGA AACGCTATGC

       301                                          350 351                                          400
Btyp1β TCTCATCGAT GTAGAACAAA CCGGTAACAT GATCGGTAAA GCTCTCGTCG AACTGACCAA CGAATGTGAT GTACCACAAG AGATCATTCA TATTGTTGGA
Btyp1γ TCTCATCGAT GTAGAACAAA CCGGTAACAT GATCGGTAAA GCTCTCGTCG AACTGACTAA CGAATGTGAT GTACCACAAG AGATCATTCA TATTGTTGGA
Bdyp1  TTTGCTCGAT GTCGAACAAA CCGGCCAAAT GATCGGCAAG ACGCTCGTCC AACTTACCGA CGAGGCTGAC GTGCCACAGG AGATCATCCA TTTGATTGGT
Bjyp1  TTTCCTCGAT GTCGAACAAA CCGGTCAAAT GATCGGCAAG ACCCTCGTCC AACTCACCGA TGAGGCTGAT GTGCCACAGG AGATCATCCA TTTGATTGGT

       401                                          450 451                                          500
Btyp1β CAAGGTGTTG GTGCCCAAGT TGCAGGAGCT GCTGGACGTC AATACAAACG TTTGACAGGT CATCAATTGC GTCGTATCAC AGCTTTGGAC CCTGCAAAAT
Btyp1γ CAAGGTGTTG GTGCCCAAGT TGCAGGAGCT GCTGGACGTC AATACAAACG TTTGACAGGT CATCAATTGC GTCGTATCAC AGCTTTGGAC CCTGCAAAAT
Bdyp1  CAAGGTATTG GCGCCCAGGT CGCCGGTGCT GCTGGTCGTC AATACAAGCG CTGGACTGGT CATCAATTGC GTCGTATTAC CGCTTTGGAC CCCGCCAAGA
Bjyp1  CAAGGTATTG GCGCACAGGT CGCCGGTGCT GCTGGTCGTC AATACAAGCG CTGGACTGGT CATCAATTGC GTCGTATTAC CGCTTTGGAC CCCGCTAAGA

       501                                          550 551                                          600
Btyp1β TGTTCGCTAA GGACAAGGAT ATGTTAACTG GTTTGGCTCG TGGTGATGCT GATTTCGTTG ATGCCATTCA CACTTCGACT TGTGGAATGG GAACACGCCA
Btyp1γ TGTTCGCTAA GGACAAGGAT ATGTTAACTG GTTTGGCTCG TGGTGATGCT GATTTCGTTG ATGCCATTCA CACTTCGACT TGTGGAATGG GAACACGCCA
Bdyp1  TCTTCGCCAG CAACAAGAAT GTTTTGAGCG GTTTGGCGCG TGGTGATGCC GATTTCGTCG ATGCCATACA CAGCAGCACC TGCGGCATGG GCACACGCCA
Bjyp1  TCTTCGCCAG CAACAAGAAT GTTTTGACCG GTTTGGCGCG TGGTGATGCC GATTTCGTCG ATGCCATACA CAGCAGCACC TGCGGCATGG GCACACGCCA

       601                                          650 651                                          700
Btyp1β ACGAGTTGGT GATGTCGACT TCTACGTCAA CGGTCCAGCT TCCGCTGCTC CAGGCGCTAC CAATGTAATT GAAGCAACTA TGCGTGCAAC TCGTTACTTC
Btyp1γ ACGAGTTGGT GATGTCGACT TCTACGTCAA CGGTCCAGCT TCCGCTGCTC CAGGCGCT.. .......... .......... .......... ..........
Bdyp1  ACGTGTCGGN GATGTTGATT TCTACGTCAA TGGTCCAGCC TCTGCTGCGC CAgGTGCTGA CAATGTCGTC GAgGCAgcTA TGCGCGCTAC ACGCTaCTtC
Bjyp1  ACGTGTAGGT GATGTTGACT TCTACGTCAA TGGTCCAGCC TCTGTTGCGC CAGGTGCTGA CAATGTCGTC GAGGCAGCAA TGCGCGCTAC ACGCTACTAC

       701                                                                                                  810
Btyp1β GCCGAATCTGTACGCCCAGGTAATGAACGTAACTTCCCTGCTGTCGCCGCCAACTCCATGGACCAATACGAAAATAACGATGGCGCTGGCAAACGCGTTACATGGGTATT
Btyp1γ ..............................................................................................................
Bdyp1  GCCGAGTCGGtGCGTCCCGGAAATGAgCGCAaCTTTCcTGcTGTACCAgCcAGTTCgAAgCAGCAGTATGAAAACAATGAGGGGTATGGcAAACgGgcCACATGGGTATT
Bjyp1  GCCGAGTCGGTGCGCCCCGGAAATGAGCGCAACTTCGCTGCTGTGCCAGCCAGTTCGAAGCAGCAGTATGAGAGCAATGAGGGTTATGGCAAACGCGCTACATGGGTATT
```

Figure 7. The DNA sequences of the *Yp1* genes of *B. tryoni* (Bt) (β and γ alleles), *B. dorsalis* (Bd) and *B. jarvisi* (Bj).

Primer binding sites are indicated by boxed text; changes in DNA sequence are indicated by underlining; intron sequences are indicated by zigzag underlining (GATTTTT). Note - region around primer sequence yp-H not shown.

```
        1                                                 50  51                                                100
Btyp2α  GCCAACAAGA AACTCATTCA GGCTTATTTG CAAGCCTATA ACGGACGAGT GCAAGTACAA GGTGGTCAAG ACGCTGACTC TGAACAGGAT ACTTCATCGA
Btyp2δ  GCCAACAAGA AACTCATTCA GGCTTATTTG CAAGCCTACA ACGGACGAGT GCAAGTAAAA GGTGGTCAAG ACGCTGACTC TGAACAGGAT ACTTCATCGA
Bdyp2   GCCACCCGGA AATTGGTTCA GGCTTATTTG CAGGCCTACA ACGGACAAGT GCAAGTACAA GGCGGCCAAG ACGCTGACTC CGAACAGGAT ACTTCATCGA
Bpyp2   GCCAATAGGA AACTGGTTCA GGCTTATTTG CAGGCCTACA ACGGACAAGT GCAAGTACAA GGTGGTCAAG ACGCTGACTC TGAACAGGAT ACTTCATCGA

        101                                              150 151                                                200
Btyp2α  GCGAGGAATC TTCTGACAGC AAACAGACCC AACCTAGTGG TAATTTGGTG GTTATTGATT TGGGCGCCGT CATACGCAAT TTCGAAGAAC TTGTTCTGCT
Btyp2δ  GCGAGGAAAC TTCTAACAGC AAACAGACCC AACCTAGTGG TAATTTGGTG GTTATTGATT TGGGCGCCGT CATACGCAAT TTCGAAGAAC TTGTTCTGCT
Bdyp2   GCGAGGAATC TTCTAACAGC AAACAGACCA AACCTAGTGG TAATTTGGTG GTTATTGATT TGGGCGCCGT CATACGCAAT TTCGAAGAAC TTGTGCTGCT
Bpyp2   GCGAGGAATC TTCTAACAAC AAACAGACCA AACCTAGTGG TAATTTGGTG GTTATTGATT TGGGCGCCGT CATACGCAAT TTCGAAGAAC TTGTGCTGCT

        201                                              250  251                                               300
Btyp2α  CGATATCAAC CGCGTCGGTG CTGCGATCGG TAACAGTTTG GTCCAGCTTA CGTCGCAAAC TGATGTACCC CAGGAAGTGA TCTATATTGT CGCACAAGGT
Btyp2δ  CGATATCAAC CGCGTCGGTG CTGCGATCGG TAACAGTTTG GTCCAACTTA CGTCGCAAAC TGATGTACCC CAGGAAGTGA TCTATATTGT CGCACAAGGT
Bdyp2   CGATATCAAC CGCGTCGGTG CTGCGATCGG TAACACTTTG GTTCAACTTA CGTCGCAAAC TGATGTACCC CAGGAAGTGA TCTATATTGT CGCACAAGGT
Bpyp2   CGATATCAAC CGCGTCGGTG CTGCGATCGG TAACACTTTG GTTCAACTTA CGTCGCAAAC TGATGTACCC CAGGAAGTGA TCTATATTGT CGCACAAGGT

         301                                             350 351                                                400
Btyp2α  ATTGGCGCTC ATGTTGCCGG TGCTGCTGCT CGCCAATACA CACGTCAAGC AGGCAACAAG TTGCGTCGTA TCACTGCCAT GGATCCCACA AAAATCTTTG
Btyp2δ  ATTGGCGCTC ATGTTGCCGG TGCTGCTGCT CGCCAATACA CACGTCAAAC AGGCAACAAG TTGCGTCGTA TCACTGCCAT GGATCCCACA AAAATCTTTG
Bdyp2   ATTGGCGCTC ATGTTGCCGG TGCTGCTGCT CGCCAATACA CACGTCAAAC AGGAAACAAG TTGCGTCGTA TCACTGCTAT GGATCCCTCA AAAATCTTTG
Bpyp2   ATTGGTGCTC ATGTTGCCGG TGCTGCTGCT CGCCAATACA CACGTCAAAC AGGAAACAAG TTGCGTCGTA TCACTGCTAT GGATCCCTCA AAAATCTTTG

         401                                              450 451                                              500
Btyp2α  CACGCAAACC CAACACTTTG GTCGGCTTGG CTCGCGGCAA TGCTGATTTC GTTGATGCCA TCCACACTTC TGCTTATGGC TTGGGTAGCG CTGCTCGTGC
Btyp2δ  CACGCAAACC CAACACTTTG GTCGGCTTGG CTCGCGGCAA TGCTGATTTC GTTGATGCCA TCCACACTTC TGCTTATGGC TTGGGTAGCG CTGCTCGTGC
Bdyp2   CACGCAAACC CAACACTTTG GTCGGCTTGG CTCGCGGCAA TGCTGATTTC GTTGATGCCA TCCACACTTC TGCTTATGGC TTGGGTAGCG CTGCTCGTGC
Bpyp2   CACGCAAACC CAACACTTTG GTCGGCTTGG CTCGCGGCAA TGCTGATTTC GTTGATGCTA TCCACACTTC TGCTTATGGC TTGGGTAGCG CTGCTCGTGC

        501                                              550 551                                                600
Btyp2α  CGGTGATGTT GACTTCTATC CCAACGGTCC TTCTGTTGCC ATGCCCGGAA CTGATAACAT AATCGAAGCT TCCTTACGTG CAACCCGTTA CTTCGCCGAG
Btyp2δ  CGGTGATGTT GACTTCTATC CCAACGGTCC TTCTGTTGCC ATGCCCGGAA CTGATAACAT AATCGAAGCT TCCTTACGTG CAACCCGTTA CTTCGCCGAG
Bdyp2   CGGTGATGTT GACTTCTATC CCAACGGTCC TTCTGTTGCC ATGCCCGGCA CTGATAACAT TATCGAAGCT TCCTTACGTG CAACACGTTA CTTCGCCGAG
Bpyp2   CGGTGATGTT GATTTCTATC CCAACGGTCC TTCTGTTGCC ATGCCCGGCA CTGATAACAT TATCGAAGCT TCCTTACGTG CAACACGTTA CTTCGCCGAG

         601                                              650 651                                             704
Btyp2α  ACAGTGCGCC CAGGTAATGA TCGCAACTTC CCAGCTGTCG CAGCCGAATC TCTACAACAG TATAAAAACAACAATGGCAACGGCAGACGCGCTTATATGGGTAT
Btyp2δ  ACAGTGCGCC CAGGTAATGA TCGCAACTTC CCAGCTGTCG CAGCCGAATC CCTACAACAG TACAAAAACAACAATGGCAACGGCAGACGCGCTTATATGGGTAT
Bdyp2   ACGGTGCGTC CAGGTAATGA TCGCAACTTC CCAGCTGTCG CAGCCGAATC CCTACAACAG TATAAAAACAACAATGGCAATGGCAAACGCGCCTACATGGGTAT
Bpyp2   ACGGTGCGTC CAGGTAATGA TCGCAACTTC CCAGCTGTCG CAGCCGAATC GCTACAACAG TATAAAAACAACAATGGCAATGGCAGACGCGCTTATATGGGTAT
```

Figure 8. Alignment of the *Yp2* sequences of *B. tryoni* (Bt) (α and δ alleles), *B. dorsalis* and *B. papayae.*

Primer binding sites are indicated by boxed text; changes in DNA sequence are indicated by underlining. Note - region around primer sequence yp-H not shown.

```
       1                                                   50 51                                                 100
Bdyp1  ANRKLVQAYM QRYYGQQQPV NANNQDYD.. .SNERQQATS SEEDYSESWK QQKSNRGNLV IIKLGAVLTN FKRYALLDVE QTGQMIGKTL VQLTDEADVP
Bjyp1  ATRKLVQAYM QRYYGQQQPV NANNQDYD.. .SNERQPATS SEEDYSESWK QQKSNRGNLV IIKLGAVLTN LKRYAFLDVE QTGQMIGKTL VQLTDEADVP
Btyp1  ANKKLIQAYM QRYNGQRQPI .SNNQDYDYG NNKDNQGATS SEEDYSESWK NPKPTKGNLV VISLGSTLTN MKRLALIDVE QTGNMIGKAL VELTNECDVP
                     *  *  * **      ** ***** *               ** ***     * *  **    *  * **       *    *   *  * *

       101                                              150 151                                              200
Bdyp1  QEIIHLIGQG IGAQVAGAAG RQYKRWTGHQ LRRITALDPA KIFASNKNVL SGLARGDADF VDAIHSSTCG MGTRQRVGDV DFYVNGPASA APGADNVVEA
Bjyp1  QEIIHLIGQG IGAQVAGAAG RQYKRWTGHQ LRRITALDPA KIFASNKNVL TGLARGDADF VDAIHSSTCG MGTRQRVGDV DFYVNGPASV APGADNVVEA
Btyp1  QEIIHIVGQG VGAQVAGAAG RQYKRLTGHQ LRRITALDPA KLFAKDKDML TGLARGDADF VDAIHTSTCG MGTRQRVGDV DFYVNGPASA APGATNVIEA
            **    *               *                    ** **  *               *                         *     *  *

       201                                               250   256
Bdyp1  AMRATRYFAE SVRPGNERNF PAVPASSKQQ YENNEGYGKR AYMGIAAQYD LEGDYI
Bjyp1  AMRATRYYAE SVRPGNERNF AAVPASSKQQ YESNEGYGKR AYMGIAAQYD LEGDYI
Btyp1  TMRATRYFAE SVRPGNERNF PAVAANSMDQ YENNDGAGKR VYMGIAAQYD LEGDYI
       *      *              *  * * **    * * *    *
```

Figure 9. The amino acid sequences of the *Yp* genes of *B. tryoni* (Bt), *B. dorsalis* (Bd) and *B. jarvisi* (Bj).
The amino acid changes between the species is indicated by an asterisk.

```
       1                                                  50 51                                                 100
Bdyp2  ATRKLVQAYL QAYNGQVQVQ GGQDADSEQD TSSSEESSNS KQTKPSGNLV VIDLGAVIRN FEELVLLDIN RVGAAIGNTL VQLTSQTDVP QEVIYIVAQG
Bpyp2  ANRKLVQAYL QAYNGQVQVQ GGQDADSEQD TSSSEESSNN KQTKPSGNLV VIDLGAVIRN FEELVLLDIN RVGAAIGNTL VQLTSQTDVP QEVIYIVAQG
Btyp2  ANKKLIQAYL QAYNGRVQVK GGQDADSEQD TSSSEETSNS KQTQPSGNLV VIDLGAVIRN FEELVLLDIN RVGAAIGNSL VQLTSQTDVP QEVIYIVAQG
                       *   *                     *    *                                     *

       101                                              150 151                                              200
Bdyp2  IGAHVAGAAA RQYTRQTGNK LRRITAMDPS KIFARKPNTL VGLARGNADF VDAIHTSAYG LGSAARAGDV DFYPNGPSVA MPGTDNIIEA SLRATRYFAE
Bpyp2  IGAHVAGAAA RQYTRQTGNK LRRITAMDPS KIFARKPNTL VGLARGNADF VDAIHTSAYG LGSAARAGDV DFYPNGPSVA MPGTDNIIEA SLRATRYFAE
Btyp2  IGAHVAGAAA RQYTRQTGNK LRRITAMDPT KIFARKPNTL VGLARGNADF VDAIHTSAYG LGSAARAGDV DFYPNGPSVA MPGTDNIIEA SLRATRYFAE
                                      *
       201                                          246
Bdyp2  TVRPGNDRNF PAVAAESLQQ YKNNNGNGKR AYMGIAADYD LEGDYI
Bpyp2  TVRPGNDRNF PAVAAESLQQ YKNNNGNGRR AYMGIAADYD LEGDYI
Btyp2  TVRPGNDRNF PAVAAESLQQ YKNNNGNGRR AYMGIAADYD LEGDYI
```

Figure 10. The amino acid sequences of the *Yp* genes of *B. tryoni*, *B. dorsalis* and *B. papayae*.
The amino acid changes between the species is indicated by an asterisk.

TRA/TRA-2 proteins to the 13 nucleotide repeat elements located in the non-coding region of exon 4. This forces the splicing machinery to splice exon 3 to exon 4, thereby creating the female-specific transcript which encodes the female-specific *dsx* protein. In theory, the presence of intron 3 (with the weak splice site) and the regulatory sequences (from the non-coding region of exon 4) in any other gene sequence (such as a lethal gene) should direct the female-specific splicing of the lethal gene sequence into a functional product. If these regulatory sequences are similar across a range of species from the Bactrocera genus then it should be possible to engineer a generic *dsx*-lethal gene construct that will function in any Bactrocera species when it is genetically transformed into that species.

The DNA sequences of the *dsx* genes of *B. tryoni* and *D. melanogaster* were aligned and degenerate primers were designed to regions around intron 3 in exons 3 and 4. Primers were biased towards the *dsx* gene of *B. tryoni* to improve the amplification of this gene from the other Bactrocera species. A fragment of the *B. dorsalis dsx* gene was isolated first and this DNA sequence is shown together with the *B. tryoni dsx* sequence from the same region in Figure 11.

The *dsx* intron 3 sequences were also amplified from *B. papayae* and *B. jarvisi*. Again the cloning of these fragments proved difficult and only two clones from each species were obtained. These clones were sequenced and the resultant sequences were aligned with the *B. tryoni* sequence of the same region as shown in Figure 11. There is only one change in coding region of the portion of exon 3 that was amplified and this does not change the amino acid sequence. The *B. papayae* and *B. dorsalis* sequences were found to be identical across the entire fragment. There are 15 changes, including four indels between *B. papayae*, *B. dorsalis* and *B. tryoni*, and nine changes including one indel between *B. dorsalis*, *B. papayae* and *B. jarvisi*. Between *B. jarvisi* and *B. tryoni* there are also are nine changes, including three indels. The species are all identical across the intron/exon junction regions, in particular across the 3' splice junction which is critical in the control of the sex-specific splicing. There is also another region that is almost identical between all sequences, the region between bases 190 and 249, which may also contain an important regulatory element for correct splicing of this intron.

(4) Construction of *dsx-grim* lethal gene cassettes

The presence of some yolk protein (*Yp2)* mRNA in newly emerged males indicated that any attempt to make a female-specific lethal using the*Yp1*/*Yp2* regulatory region would not be feasible. Therefore, lethal-gene constructs using the regulatory elements of the *dsx* gene and the cell-death gene *grim* were engineered as shown in Figure 12. These constructs consist of the female-specific intron, intron 3 of the *dsx* gene joined into the coding region of the *grim* gene of *D. melanogaster*. The 5' end of the construct consists of the 5' untranslated region of the g*rim* gene and the 3' end of the construct consists of the 3' untranslated region of exon 4 of the *dsx* gene (as shown in Figure 12). Two different constructs were designed for both *D. melanogaster* and *B. tryoni*. This was to try to ensure that at least one of the constructs for each species did not inadvertently place the intron sequence next to any unidentified enhancer sequences in the *grim* gene sequence, which may alter the expression of the gene construct. Primers which overlapped the ends of consecutive fragments were designed as shown in Figure 12. As some of these primers contained sequences that were common recognition sequences for intron splicing or were basically poor primer sequences (eg. all the -c primers) the region encompassing these sequences within the *dsx* gene was amplified with the primers Dm-y and Dm-h2 for *D. melanogaster* and Bt-y and Bt-h2 for *B. tryoni* as shown in Figure 12. The fragments bounded by primers –c and –d as well as –g and –h were amplified using the h-y dsx gene fragment as template. The *grim* gene fragments bounded by the –a and –b and also the –e and –f primers, were amplified from *D. melanogaster* genomic DNA. Pairs of adjacent fragments were combined in the same amplification reaction and joined by allowing the overlapping ends to anneal and then extending the sequences and shown in Figure 11. Primers were then added to the reaction and the amplification continued. This process was continued using the –a2 and –h2 primers until the full length constructs were obtained. These primers did not contain any restriction sites and were located within the sequence of the constructs so that amplification was not inhibited by the self annealing of primers, which can occur when primers contain restriction sites. The full length constructs were then amplified again using the –a and –h primers which contain unique restriction sites such as *Not*I. The constructs are waiting for the problems with cloning to be resolved before they are cloned. These constructs, when transferred into either *D. melanogaster* or *B. tryoni*, and expression is induced, should direct the splicing of the construct into a functional product of the cell death gene *grim* only in females

```
       1                                                 50
Bdint3 AGAGCATTGT CAAAAACTAT TGGAGAAATT TCGATATCCA TGGGAGATGA
Bpint3 GGAGCATTGT CAAAAGCTAT TGGAGAAATT TCGATATCCA TGGGAGATGA
Bjint3 GGAGCATTGC CAGAAGCTAT TGGAGAAATT TCGATATCCA TGGGAGATGA
Btint3 GGAGCATTGT CAAAAACTAT TGGAGAAATT TCGATATCCA TGGGAGATGA

       51                                               100
Bdint3 TGCCATTCAT GTATGTGATA TTAAAAGATG CTGGGGCAGA TATTGAAGAG
Bpint3 TGCCATTAAT GTATGTGATA TTAAAAGATG CTGGGGCAGA TATTGAAGAG
Bjint3 TGCCATTAAT GTATGTGATA TTAAAAGATG CTGGGGCAGA TATTGAAGAG
Btint3 TGCCATTAAT GTATGTGATA TTAAAAGATG CTGGGGCAGA TATTGAGGAA

       101                                              150
Bdint3 GCTTCAAGAC GCATTGAGGA AGGTAAGTGT GTATTTCAGC ATTTCCGTTA
Bpint3 GCTTCAAGAC GCATTGAGGA AGGTAAGTGT GTATTTCAGC ATTTCCGTTA
Bjint3 GCTTCAAGAC GCATTGAGGA AGGTAAGTTT GCATTTCAGC ATTTCCGTTA
Btint3 GCTTCAAGAC GCATTGAGGA AGGTAAGT.T GCATTTCAGC ATTCCCGTTA

       151                                              200
Bdint3 C.AAAAGGTG TTTATATCAT ATTCTC.AAA ATGACTTAAT TTACCATGAT
Bpint3 C.AAAAGGTG TTTATATCAT ATTCTC.AAA ATGACTTAAT TTACCATGAT
Bjint3 CAAAAACGTG TTTACATCAT ATACTC.AAA TTGACATAAT TAACCATGAT
Btint3 CAAAATCGTG TTTACATCGT ATACTCGAAA TTGACTTAAT TTACCATGAT

       201                                              250
Bdint3 AAGTACGCAT TTATTATTAT GCAATCAAAT TACAATAAT. TAAAATATTT
Bpint3 AAGTACGCAT TTATTATTAT GCAATCAAAT TACAATAAT. TAAAATATTT
Bjint3 AAGTACGCAC TTATTATTAT GCAATCAAAT TACAATCAT. TAAAATATTT
Btint3 AAGTACGCAC TTATTATTAT GCAATCAAAG TACAATAATG CAAAATATTT

       251                                              300
Bdint3 TAGGCCAACA TGTCGTAAAT GAATACTCCC GTCAACACAA TCTGAATATA
Bpint3 TAGGCCAACA TGTCGTAAAC GAATACTCCC GTCAACACAA TCTGAATATA
Bjint3 TAGGCCAACA TGTCGTAAAC GAATACTCCC GTCAACACAA TCTGAATATA
Btint3 TAGGCCAACA TGTCGTAAAC GAATACTCCC GTCAACACAA TCTGAATATA

       301                   322
Bdint3 TATGACGGGG GTGAGCTGCG ..
Bpint3 TATGACGGGG GTGAGCTGCG GA
Bjint3 TATGACGGGG GTGAGCTGCG GA
Btint3 TATGACGGGG GTGAGCTGCG ..
```

Figure 11. The sequence of intron 3 of the *dsx* gene of *B. tryoni*, *B. dorsalis*, *B. papayae* and *B. jarvisi*.

Primer binding sites are indicated by boxed text; changes in DNA sequence are indicated by underlining; intron splice site recognition seqeunces are indicated by zigzag underlining (GATTTTT).

Figure 12. Construction of the *B. tryoni dsx-grim* lethal gene constructs

A. Location of primers within *dsx* and *grim* genes for *dsx-grim* gene constructs. The generation of the h-y template for amplification of fragments containing intron/exon junction sequences. Note only *B. tryoni* sequence is shown here

B. Double-stranded - PCR (ds-PCR) products showing overlapping ends generated by primer sequences

C. Final dsx-grim gene constructs. Two different types of constructs were generated for each species.

DISCUSSION

(1) The *Yp* genes of *B. tryoni*

As *D. melanogaster* is the most well studied dipteran insect species, the study of any genes from other dipteran insects will necessarily involve some form of comparison with the vast body of knowledge that exists for the vinegar fly, *D. melanogaster*. Although *D. melanogaster* is taken as a model dipteran species, in many respects it must be considered to be a highly evolved species and, as such, not all relevant findings can be extended to other dipteran species.

In Q-fly, *B. tryoni*, there are two alleles of the each of the *Yp1* and *Yp2* genes, that is, a total of four *Yp* genes compared to only three such genes in *D. melanogaster*. In Med-fly, *C. capitata*, another tephritid species, there are also two alleles of each of the *Yp1* and *Yp2* homologues, designated *Vg-1* and *Vg-2*. The *Yp* genes of *D. melanogaster* are thought to represent duplicated copies of a single ancestral gene based on the high degree of sequence identity that exists between all the *Yp* genes. In *C. capitata* the suggestion has been that the gene pairs represent a recent duplication event due to the high degree of sequence identity between the alleles (Rina and Mintzas, 1987) although this suspected duplication event may not be as recent as proposed, discussed below. Whether all the copies of the *Yp* genes are expressed and functional in *B. tryoni* is not known, although it is likely that they are because there are only minor changes between the each pair of genes (including the control regions) and the protein sequences encoded by each allele are also almost identical.

The protein sequences of the yolk proteins of *B. tryoni* and *D. melanogaster* (and also *C. capitata*) show a degree of conservation, in particular across those regions that are thought to be functionally important. In *D. melanogaster*, it has been found that the *Yp* proteins can readily substitute for one another, as indicated in mutant flies which lacked one of the *Yp* genes yet still produced viable eggs (Bownes *et al*., 1991). In *D. melanogaster*, the *Yp1* and *Yp2* sequences show a higher level of identity to one another than found in the other species examined. In the pest fruit fly species, there is greater conservation of the protein sequence in the *Yp1* (or *Yp2*) homologues between species than between the *Yp1* and *Yp2* genes within species. This suggests that the duplication and sequence divergence of the *Yp* genes took place long before speciation

The high degree of similarity in the *Yp2* protein sequences in all the species examined, compared to the lower level of similarity in the *Yp1* protein sequences, may also point towards some sort of specialised function for this *Yp2*. Alternately, these differences may indicate that the yolk proteins have similar, but subtly different functions and, as such, the YP1 and YP2 proteins may represent two different subunits of a larger yolk protein molecule.

The regulation of the *Yp* genes in both the ovaries and the fat body in *D. melanogaster* is complex and factors such as nutrition, hormones (20-hydroxyecdysone and juvenile hormone), tissue-specific factors (such as GATA , OE1 and OE2), as well as the sex-determination gene *dsx*, all play a role in the expression of the *Yp* genes (Garabedian *et al*., 1985; Burtis *et al*., 1991; Lossky and Wensink, 1995; Søndergaard *et al*., 1995; Bownes *et al*., 1996). Nutrition seems to be important in both *D. melanogaster* and *B. tryoni* for the production of yolk proteins, as in both species protein must be ingested in order to produce mature eggs. Regions which confer a nutritional response have been identified in the intergenic region of the *Yp1* and *Yp2* genes of *D. melanogaster* (Søndergaard *et al*., 1995). There is a redundancy in these sequences which confer a nutritional response, similar to that seen for the elements which determine female specificity of the *Yp* genes. It has been suggested that nutrition acts to modify the level of a *trans*-acting factor, and that several DNA binding proteins are likely to interact and regulate transcription of the *yp* genes. The expression of the *Yp* genes of *B. tryoni* is likely to be controlled by the same regulatory elements as those identified in *D. melanogaster*. However, the actual binding sites of these regulatory proteins are not always obvious in *B. tryoni*, as the DNA sequence of the IGR differs significantly to the same region in *D. melanogaster*. Major differences in the IGR of *Yp1* and *Yp2* is also observed for other Drosophila species such as *D. grimshawi*, making the identification of the regulatory sequences, responsible for female-specific expression, quite difficult. Given the evolutionary distance between *B. tryoni* and *D. melanogaster* it is not surprising that these sequence are difficult to identify based on sequence homology alone. Functional studies will have to be performed to determine the exact binding sites for the regulatory proteins.

The examination of the expression of the *Yp* genes in *B. tryoni* indicates when the regulatory sequences function and the difference in level of activity between males and females. There appeared to be a small amount of *Yp2* mRNA present in males 24 hr after emergence, although this amount did not increase and disappeared after a few days

(data not shown). It is possible that there are cryptic control regions that limited the expression of *Yp1* to females but the same controls are not active on the *Yp2* gene.
Although the poor conservation of *Yp* regulatory regions means that we could not define regulatory sites as we could for *dsx*, the results suggest that only *Yp1*-specific regulatory regions should be used in any female-killing construct and might be useful as an adult genetic sexing system to back up a system active in early development.

(2) The *Yp* genes of other Bactrocera species

Insufficient numbers of clones were isolated from *B. dorsalis*, *B. papayae* and *B. jarvisi* to determine whether there were more than one allele of each of the *Yp1* and *Yp2* genes in these species. Given that there are two alleles of each of the *Yp* genes in both *B. tryoni* and *C. capitata* it is likely that other Bactrocera species will also possess two alleles of each *Yp* gene. The sequence similarity that was noted between the *Yp* sequences of *B. tryoni* and *B. dorsalis* suggests that the regulatory sequences will be very similar to one another. The protein sequences that have been obtained will allow isolation of the regulatory sequences in other Bactrocera species, as well as in Med-fly, and comparisons between these should lead to a better identification of regulatory regions.

(3) The *dsx* genes of other Bactrocera species

The sequence of the *dsx* gene that was isolated from other Bactrocera species included some of exon 3 and exon 4 and the complete intron 3. An alignment of the sequences from *B. tryoni*, *B. jarvisi*, *B. dorsalis* and *B. papayae* show that all contain the characteristic weak 3' splice site that is necessary for the splicing of the *dsx* pre-mRNA into the female-specific form in female flies. Given the similarity of the intron sequences and the known similarity of the exon sequences, the designing of female-lethal constructs that function in all species should be straightforward.

(4) Lethal gene constructs of *dsx* gene and the *D. melanogaster grim* gene

Processing of the *D. melanogaster dsx* pre-mRNA into the female-specific form occurs by 3'-splice site activation and is the result of the utilisation of an upstream 3'-terminal exon, exon 4 (Ryner and Baker, 1991). This splicing is activated by the binding of both TRA and TRA-2 to the cis-acting 13 nucleotide repeats within the *dsx*RE, which lies in the non-coding region of exon 4 (Ryner and Baker, 1991; Inoue *et al*., 1992; Tian and Maniatis, 1993). The *dsx*RE is required, not only for female-specific splicing but also for female-specific polyadenylation (Hedley and Maniatis, 1991). Alteration of any of these cis-acting sequences either by deletion or inversion of the elements (Nagoshi and Baker, 1990; Hoshijima *et al*., 1991) or base substitutions within the elements (Inoue *et al*., 1992) results in the failure of the female-specific splicing of exon 3 to exon 4. Within the putative non-coding region of exon 4 of the *B. tryoni dsx* transcript, four sequences that are 10-13 bases identical to particular cis-elements of *D. melanogaster* have been identified (Shearman and Frommer, 1998). The *D. melanogaster dsx*RE contains six cis-elements whereas the *B. tryoni* homologue contains only four elements. Similarly, *D. virilis* has only four elements within the female-specific fourth exon (Hertel *et al*., 1996). The purine-rich enhancer (PRE) sequence which lies within the *dsx*RE of *D. melanogaster* has been shown to be necessary for the binding of TRA-2 to the *dsx*RE (Lynch and Maniatis, 1995). A purine-rich region, which does not show sequence identity to the *D. melanogaster* PRE but which appears to be functionally equivalent, has been identified within the *dsx*RE of *D. virilis*, immediately downstream of the fourth repeat element (Hertel *et al*., 1996). Similarly, a region that is purine-rich, but not identical to that of the *D. melanogaster* sequence, can be identified within the putative exon 4 sequence of *B. tryoni* downstream of the fourth repeat element.

The cell-death gene *grim* of *D. melanogaster* has been characterised and studies have suggested that it has an apoptotic function. The *grim* gene activity parallels that of *reaper* (*rpr*), however cell killing by *grim* does not require the functions of *rpr* or *hid* (Chen *et al.,* 1996). The *grim* gene has also been found to trigger apoptosis in mammalian cell lines, and therefore we hypothesise that it may be potentially more useful that *rpr* or *hid* as a killing system in divergent dipteran insect species. The constructs that we have produced, when transferred into either *D. melanogaster* or *B. tryoni*, and expression induced, should direct the splicing of the construct into a functional product of the cell death gene *grim* only in females. If expression of these constructs is induced at a critical stage of development to a level sufficient to induce cell death, it may be possible to eliminate females from the SIT strain leaving only males in the brood. The constructs made for Drosophila will allow testing of its efficacy in a laboratory system where gene transformation is routine. The constructs made for Q-fly will provide a means of testing sophisticated gene-transformation protocols.

(5) Conclusions

(1) The planned research outcomes were achieved. The currently unexplained low cloning efficiencies with highly-competent cells meant that there were fewer than expected clones of isolated fragments. However, all isolated fragments of *Yp* and *dsx* genes have been checked by sequencing and the original fragments are in storage, so that more clones can be made as required. In the light of this problem with cloning, it was decided not to attempt to clone the large female-lethal constructs until the problem was resolved.

(2) The bidirectional regulatory region of the *Yolk protein* (*Yp1* and *Yp2*) genes of *B tryoni* could not be used to produce a female-specific cell-death construct because low-level expression of the *Yp2* gene was observed in males. Furthermore, the IGR of the *Yolk protein* genes were relatively poorly conserved between species, so further analysis is required to identify the female-specific regulatory sequences. The results suggest that only *Yp1*-specific regulatory regions should be used in any female-killing construct and might be useful as an adult genetic sexing system to back up a system active in early development.

(3) By contrast, the *doublesex* gene sequences were extremely well conserved within the fruit fly species examined. Therefore, constructs were made which should be lethal to females when transformed into fruit fly embryos. Constructs were made which would produce female-specific splicing of the Drosophila cell-death gene, *grim*, when transformed into embryos of the laboratory fly, *D. melanogaster.* This constitutes a control to test the concept of female-specific lethality caused by *grim*. The constructs made using the female-specific splicing signals from the *B. tryoni doublesex* gene will be used to test both gene transformation efficiencies and female-specific lethality in Q-fly.

TECHNOLOGY TRANSFER

The work carried out in this project is preliminary in nature. Its future development to allow new biological control protocols will require much further laboratory testing, as well as the negotiation of field protocols to allow release of the
sterile genetically-modified males. We note that factory containment protocols in many sites in the world, where factories are located in areas free of fruit flies or where flies are currently being eradicated, are already developed. We also note that, because any released flies would be sterile, SIT with genetically-modified male-only strains will probably be one of the earlier protocols to be certified as genuinely environmentally benign.

In the meantime, a continuation of laboratory research to develop stable constructs and transformation protocols is required. The Fruit Fly Research Centre is active in discussions with Industry groups and Government departments of Primary Industries, to ensure that an understanding of issues, problems and advantages is widespread throughout the community.

RECOMMENDATIONS

(1) Future R&D on a male-only strain based on the *doublesex* sex-determination gene and the *grim* cell death gene requires:
(a) The further development of stable, reliable genetic transformation systems for Q-fly,
(b) A testing of the *doublesex*/*grim* constructs for their efficacy in killing female embryos.
(c) It appears that endogenous genes and regulatory sequences function better in transformation systems for pest insects than the frequently-used *D. melanogaster* counterparts. Therefore future work to isolate and incorporate the *grim* gene from Bactrocera, in place of the currently-available Drosophila gene, is recommended before final decisions are made about the efficacy of the female-killing construct.

(2) Several types of female-specific lethal constructs will need to be tested before the most reliable systems could be used in the field and, indeed, more than one system may be required for fail-safe operation. Therefore, at this stage of research, we recommend that work be carried out to investigate other genes of the sex-determination pathway of fruit flies, in particular, the male-determining factor and the *transformer* genes, to establish which is the most reliable for male-only SIT

(3) An important further recommendation to Horticulture Australia and the horticultural industries is that information about the use of genetic modification in SIT strains be widely disseminated, since the problems and advantages are very

different to those of the well-known genetic modification of food crops. We propose that the production of male-only strains, by the introduction of female-lethal constructs into factory flies, should be approved for release at an earlier stage than current protocols on genetic modification of crops.

(4) We note that this work, as strategic research with no expected outcomes that could be adopted by the industry for several years, presents a problem for funding. We recommend that a funding model be developed that provides separately for projects that can be immediately adopted and for projects that will have important implications for Australia's competitiveness in the future.

ACKNOWLEDGMENTS

We are grateful to Daniele Crisafulli and Marianne Frommer for their assistance in this study. The work was supported by a grant from Horticulture Australia Limited. We are grateful to Merryl Robson for negotiating funding towards this project.

BIBLIOGRAPHY OF LITERATURE CITED

Belote, J. M, Handler, A. M., Wolfner, M. F., Livak, K. J and Baker, B. S. 1985. Sex-specific regulation of *Yolk protein* gene expression in Drosophila. *Cell* **40**(2):339-348.

Bender, W., Spierer, P. and Hogness, D. S. 1983. Chromosomal walking and jumping to isolate DNA from the *Ace* and *rosy* loci and the *bithorax* complex in *Drosophila melanogaster*. *J Mol Biol* **168**:17-33.

Bennett, C. L. and Frommer, M. 1997. The *white* gene of the tephritid fruit fly *Bactrocera tryoni* is characterized by a long untranslated 5' leader and a 12kb first intron. *Insect Mol Biol.* 6(4):343-356.

Bownes, M. Shirras, A. d. and Saunders, R. D. C. 1987. A critical evealutation of the factors essential for correct expression of the genes coding for Drosophila yolk proteins. In Proceedings of a Monsanto UCLA Symposium, 49 Steamboat Springs, Colarado, USA.

Bownes, M., Lineruth, K. and Mauchline, D. 1991. Egg production and fertility in Drosophila depend upon the number of *Yolk protein* gene copies. *Mol. Gen. Genet.* **228**(1-2):324-327.

Bownes M, Ronaldson E, Mauchline D. 1996. 20-Hydroxyecdysone, but not juvenile hormone, regulation of yolk protein gene expression can be mapped to cis-acting DNA sequences. *Dev Biol*. **173**(2):475-489

Burtis, K. C., & Baker, B. S. 1989. Drosophila *doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56**:997-1010.

Burtis, K. C., Coschigano, K. T., Baker, B. S., & Wensink, P. C. 1991. The *doublesex* proteins of *Drosophila melanogaster* bind directly to a sex-specific *yolk protein* gene enhancer. *EMBO J* **10**:2577-2582.

Cayol, J. P. Causse, R. Louis, C. Barthes, J. 1994. Medfly *Ceratitis capitata* Wiedemann (Dipt., Trypetidae) as a rot vector in laboratory conditions. *J. App Entomol.* **117**(4):338-343.

Chen P, Nordstrom W, Gish B, Aβrams JM. 1996. *grim*, a novel cell death gene in Drosophila. *Genes Dev.* **10**(14):1773-1782.

Chung YD, Kwon HC, Chung KW, Kim SJ, Kim K, Lee CC. 1996. Identification of ovarian enhancer-binding factors which bind to *ovarian enhancer 1* of the Drosophila genes *Yp1* and *Yp2*. *Mol Gen Genet.* **251**(3):347-351.

Franz, G., Gencheva, E. and Kerremans, P. 1994.. Improved stability of genetic sex-separation strains for the Mediterranean fruit fly, Ceratitis capitata [Article] *Genome* **37**(1): 72-82.

Garabedian, M. J., Hung, M. C., and Wensink, P. C. 1985. Independent control elements that determine yolk protein gene expression in alternative Drosophila tissues. *Proc Natl Acad Sci U S A*. **82**(5):1396-1400.

Handler, A. M., McCombs, S. D., Fraser, M. J and Saul, S. H. 1998. The lepidopteran transposon vector, piggyBac, mediates germ-line transformation in the Mediterranean fruit fly. *Proc Natl Acad Sci U S A*. **95**(13):7520-7525.

Hedley, M. L., & Maniatis, T. 1991. Sex-specific splicing and polyadenylation of *dsx* pre-messenger RNA requires a sequence that binds specifically to *tra-2* protein *in vitro*. *Cell* **65**: 579-586.

Hendrichs, J., Franz, G. and Rendon, P. 1995. Increased effectiveness and applicability of sterile insect technique through male-only releases for control of Mediterranean fruits flies during fruiting seasons [Article] Journal of Applied Entomology. **119**(5): 371-377.

Hendrichs, J., Robinson, A. S, Cayol, J. P. and Enkerlin, W. 2002. Medfly areawide sterile insect technique programmes for prevention, suppression or eradication: The importance of mating behavior studies. *Florida Entomologist* **85**(1):1-13.

Hertel KJ, Lynch KW, Hsiao EC, Liu EH, Maniatis T. 1996. Structural and functional conservation of the Drosophila *doublesex* splicing enhancer repeat elements. *RNA.* **2**(10):969-981.

Horwood, M. A. and Keenan, P. A. 1996. Queensland fruit fly eradication. *Horticultural Research and Development Corporation* – Final Report, CT336.

Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., & Shimura, Y. (1991). Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in Drosophila. *Science* **252**: 833-836.

Inoue (1990). High efficiency transformation of *E. coli* with plasmids. *Gene* **96**: 23-28.

Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H., & Shimura, Y. (1992). Binding of the Drosophila *transformer* and *transformer-2* proteins to the regulatory elements of *doublesex* primary transcript for sex-specific RNA processing. *Proc Natl Acad Sci USA* **89**: 8092-8096.

Knipling, E. F. 1955. Possibilities of insect control or eradication throuhg the use of sterile males. *J. Econ Entomol* **8**

Logan, S. K. and Wensink, P. C. 1990. Ovarian follicle cell enhancers from the Drosophila yolk protein genes: different segments of one enhancer have different cell-type specificities that interact to give normal expression. *Genes Dev.* **4**(4):613-23.

Lossky, M. and Wensink, P. C. 1995. Regulation of Drosophila *Yolk protein* genes by an ovary-specific GATA factor. *Mol Cell Biol*. **15**(12):6943-52.

Loukeris, T. G, Livadaras, I., Arca, B., Zabalou S. and Savakis, C. 1995. Gene transfer into the medfly, Ceratitis capitata, with a Drosophila hydei transposable element. *Scienc*. **270**(5244):2002-2005.

Lynch, K. W., & Maniatis, T. (1995). Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev* **9**(3): 284-293.

Marchuk, D., Drumm, M., Saulino, A., & Collins, F. C. (1991). Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res* **19**: 1154.

McInnis, D. O., Tam, S., Grace, C. and Miyashita, D. 1994. Population suppression and sterility rates induced by variable sex ratio sterile insect releases of *Ceratitis capitata* (Diptera: Tephritidae) in Hawaii. *Ann. Entomol. Soc. Amer.* **87**(2): 231-240.

Meats, A. 1981. The bioclimatic potential of the Queensland fruit fly, *Dacus* (*Bactrocera) tryoni*, in Australia. *Proceedings of the Ecological Society of Australia* **11**:151-161.

Nagoshi, R. N. and Baker, B. S. 1990. Regulation of sex-specific RNA splicing at the Drosophila *doublesex* gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes Dev.* **4**(1):89-97.

O'Brochta, D. A. & Atkinson, P. W. 1998. Building the better bug. *Sci. Am.* **279**(6), 60-66

O'Brochta, D. A and Atkinson, P. W. 1996. Transposable elements and gene transformation in non-drosophilid insects. *Insect Biochem Mol Biol*. **26**(8-9):739-53. Review.

Pane, A., Salvemini, M., Delli Bovi, P., Polito, C and Saccone, G. 2002. The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate. *Development* **129**(15):3715-3725.

Piano, F., Parisi, M. J., Karess, R. and Kambysellis, M. P. 1999. Evidence for redundancy but not trans factor-cis element coevolution in the regulation of Drosophila *Yp* genes. *Genetics* **152**(2):605-16.

Rendon P, McInnis D, Lance D, Stewart J. 2004. Medfly (Diptera: Tephritidae) genetic sexing: large-scale field comparison of males-only and bisexual sterile fly releases in Guatemala. *J. Econ. Entomol.* **97**(5):1547-53.

Rina, M. D. and Mintzas, A. C. 1987. Two vitellins-vitellogenins of the Mediterranean fruit fly, *Ceratitis capitata*: a comparative biochemical and immunological study. *Comp Biochem. Physiol*. **86B**: 801-808

Robinson, A. S., Riva, M. E. and Zapater, M. 1986. Genetic-sexing in the Mediterranean fruit fly, *Ceratitis capitata* using the alcohol dehydrogenase locus. *Theor. Appl. Genet.* **72**(4): 455-457

Robinson, A.S., 2002. Mutations and their use in insect control. *Mut. Res.* **511**:113–132.

Robinson, A. S. Cayol, J. P. Hendrichs, J. 2002. Recent findings on medfly sexual behavior: Implications for SIT. *Florida Entomologist* **85**(1):171-181.

Robinson, A. S., Franz, G. and Fisher, K. 1999. Genetic sexing strains in the medfly, *Ceratitis capitata*: Development, mass rearing and field applications. *Trends Entomol* **2**:81–104.

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.

Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. (2nd ed.) New York: Cold Spring Harbor Laboratory Press, USA.

Shearman, D. C. & M. Frommer, 1998. The *Bactrocera tryoni* homologue of the *Drosophila melanogaster* sex-determination gene *doublesex*. *Insect Mol Biol* **7**(4):355-366.

Shearman, D. C. A. 1999. Sex-specific genes of the Queensland fruit fly, *Bactrocera tryoni.* Ph.D. Thesis. School of Biological Sciences, University of Sydney, NSW, Australia.

Sondergaard L, Mauchline D, Egetoft P, White N, Wulff P, Bownes M. 1995. Nutritional response in a Drosophila yolk protein gene promoter. *Mol Gen Genet*. **248**(1):25-32.

Tian, M. and Maniatis, T. 1993. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* **74**(1):105-114.

Appendix 1. The DNA sequences of the *Yp1β*, *Yp2α*, *Yp1γ* and *Yp2δ* genes and the intergenic regions.

```
      1                                                                    50                                                                    100
yp1β  ATACCCATGT AAACGCGTTT GCCAGCGCCA TCGTTATTTT CGTATTGGTC CATGGAGTTG GCGGCGACAG CAGGGAAGTT ACGTTCATTA CCTGGGCGTA
yp1γ  ATACCCATGT AAACGCGTTT GCCAGCGCCA TCGTTATTTT CGTATTGGTC CATGGAGTTG GCGGCGACAG CAGGGAAGTT ACGTTCATTA CCTGGGCGTA

      101                                                                  150                                                                   200
yp1β  CAGATTCGGC GAAGTAACGA GTTGCACGCA TAGTTGCTTC AATTACATTG GTAGCGCCTG GAGCAGCGGA AGCTGGACCG TTGACGTAGA AGTCGACATC
yp1γ  CAGATTCGGC GAAGTAACGA GTTGCACGCA TAGTTGCTTC AATTACATTG GTAGCGCCTG GAGCAGCGGA AGCTGGACCG TTGACGTAGA AGTCGACATC

      201                                                                  250                                                                   300
yp1β  ACCAACTCGT TGGCGTGTTC CCATTCCACA AGTCGAAGTG TGAATGGCAT CAACGAAATC AGCATCACCA CGAGCCAAAC CAGTTAACAT ATCCTTGTCC
yp1γ  ACCAACTCGT TGGCGTGTTC CCATTCCACA AGTCGAAGTG TGAATGGCAT CAACGAAATC AGCATCACCA CGAGCCAAAC CAGTTAACAT ATCCTTGTCC

      301                                                                  350                                                                   400
yp1β  TTAGCGAACA ATTTTGCAGG GTCCAAAGCT GTGATACGAC GCAATTGATG ACCTGTCAAA CGTTTGTATT GACGTCCAGC AGCTCCTGCA ACTTGGGCAC
yp1γ  TTAGCGAACA ATTTTGCAGG GTCCAAAGCT GTGATACGAC GCAATTGATG ACCTGTCAAA CGTTTGTATT GACGTCCAGC AGCTCCTGCA ACTTGTGCAC

      401                                                                  450                                                                   500
yp1β  CAACACCTTG TCCAACAATA TGAATGATCT CTTGTGGTAC ATCACATTCG TTAGTCAGTT CGACGAGAGC TTTACCGATC ATGTTACCGG TTTGTTCTAC
yp1γ  CAACACCTTG TCCAACAATA TGAATGATCT CTTGTGGTAC ATCACATTCG TTGGTCAGTT CGACGAGAGC TTTACCGATC ATGTTACCGG TTTGTTCTAC

      501                                                                  550                                                                   600
yp1β  ATCGATGAGA GCTAAACGTT TCATGTTGGT GAGGGTGGAG CCCAAGCTGA TGACCTAAAT GAATAAACAG AAAATAGTTA AAGATAATGA TAGAAAAATC
yp1γ  ATCGATGAGA GCTAAACGTT TCATGTTCGT GAGGGTGGAG CCCAAGCTGA TGACCTAAAT GAATAAACAG AAAATAGTTA ATGATAATGA TAGAAAAATC

      601                                                                  650                                                                   700
yp1β  ATAAAATACA CATTTGGTGT TAAAACTTAC CACAAGGTTG CCCTTTGTGG GTTTGGGGTT CTTCCACGAT TCACTGTAGT CTTCTTCACT TGAAGTGGCA
yp1γ  ATAAAATACA CATTTGGTGT TAAAACTTAC CACAAGGTTG CCCTTTGTGG GTTTGGGATT CTTCCACGAT TCACTGTAGT CTTCTTCACT TGAAGTGGCA

      701                                                                  750                                                                   800
yp1β  CCTTGATTGT CCTTATTGTT GCCATAGTCA TAGTCCTGGT TATTGCTGAT TGGTTGACGC TGTCCATTGT AGCGCTGCAT ATAAGCCTGG ATCAATTTTT
yp1γ  CCTTGATTGT CCTTATTGTT GCCATAGTCA TAGTCCTGGT CATTGCTGAT TGGTTGACGC TGTCCATTGT AGCGCTGCAT ATAAGCCTGG ATCAATTTCT

      801                                                                  850                                                                   900
yp1β  TGTTTGCCTT TGCAACAGAA GGAGACGACT GTGGCAAACC AGTGATGAAA ATGGTTACTT CCTCATCGCC AAAGTTGGGC TGTTGCTTGG CAGTTTGGAC
yp1γ  CGTTTGCCTT TGCAACAGAA GGAGACGACT GTGGCAAACC AGTGATGAAA ATGGTTACTT CCTCATCGCC AAAGTTGGGC TGTTGCTTGG CAGTTTGGAC

      901                                                                  950                                                                   1000
yp1β  TAGTTCATTG AGATTGGTAC GCTCGGGCTG TCCATTGGGT TTCATCAAGA TAACCGGCAC ATTACTGGGG CTCGGAACAA AGGACGGTTC TAGAGCATGA
yp1γ  TAGTTCATTG AGATTGGTAC ACTCGGGCTG TCCATTGGGT TTCATCAAGA TAACCGGCAC ATTACTGGGG CTCGGAACAA AGGACGGTTC TAGAGCGTGA

      1001                                                                 1050                                                                  1100
yp1β  TTAATTTGAG ACAAGTGATC TAAAGGAAGA AAAAATATTA TATTTTATAG TACCTAAATT AGAATTAATC GAACCATTTG TAACACTTAC ATAGCTTCTC
yp1γ  TTAATTTGCG ACAAGTGATC TAAAGGAAGA AAAAATATTA TATTTTATAG ....TAAATT AGAATTAATC GAACCATTTG TAACACTTAC ATAGCTTCTC
```

```
      1101                                                    1150                                                    1200
yp1β  AATCTTTCGC TGGGCATCTT CCATTGACAT GTTCTCCAAC TGTTGCAAAG TGATATCCTC GACTGATGGC ATCGATTCTA ATTCAGAGGC CGATAACCAA
yp1γ  AATCTTTCGC TGGGCATCTT CCATTGACAT GTTCTCCAAC TGTTGCAAAG TGATATCCTC GACTGATGGC ATCGATTCTA ATTCAGAGGC CGATAACCAA

      1201                                                    1250                                                    1300
yp1β  TCCACAGGTT TTAAGCTATT CACGTTGTCC TTGTTCTTGC CATGCTTGGG TGAAGCATTA GCAATGGCTA ATACGAAAGC CATAAAACAG AAAATCTTCA
yp1γ  TCCACAGGTT TTAAGCTATT CACGTTGTCC TTGTTCTTGC CATGCTTGGG TGAAGCATTA GCAATGGCTA ATACGAAAGC CATAAAACAG AAAATCTTCA

      1301                                                    1350                                                    1400
yp1β  GAGGATTCAT ACTGATCACG TTTGCAAATC GGTTTGAACT GTGCCTATAA CTTAATTGTC GATGATTTAT ATACGAAAAT GCAGTCTACC TTCGTAAGCC
yp1γ  GAGGATTCAT ACTGATCACG TTTGCAAATC GGTTTGAACT GTGCCTATAA CTTAATTGTC GATGATTTAT ATACGAAAAT GCAGTCTACC TTCGTAAGCC

      1401                                                    1450                                                    1500
igrαβ CTGCCTGCAT AGAAACTTTT GACAGACACC TTATACGTTT GTTTGCAAAA AGTTTCGGAG CATTTACA.. ......TATA AGTAAATGCA TTTAGTTTTT
igrγδ CTGCCTGCAA AGAAACTTTT GACAGACACC TTATACATTT ATTTGCAAAA AGTTTCGGAG CATTTACATA TGTATGTATA AGTAAATGCA TTTAGTTTTT

      1501                                                    1550                                                    1600
igrαβ AGATTATATT CCGAGTATAT GACTGCAGAA GGGCGTTGAC CGCACCTTGA ACGAGTGCTA GTGGACGGCA ATGTTGCGCG TGCAATTTTG GCTGAATTGC
igrγδ AGATTATATT CTGAGTATAT GACTGCAGAA GGGCGTTGAC CGCACCTTGA ACGAGTGCTA GTGGACGGCA ATGTTGCGCG TGCAATTTTG GCTGAATTGC

      1601                                                    1650                                                    1700
igrαβ AGTATTTGCC GGTCGGTACA ACAACAACAC GAATCTATAA TCAACTTTGT TGCCAAAATG TGATGCGCTG TCGCAAGTGA TAAGGGCTAA ACGAGCCAAC
igrγδ AGTATTTGCC GGTCGGTACA ACAACAACAC AAGTCTATAA TCAACTTTGT TGCCAAAATG TGATGCGCTG TCGCAAGTGA TAAGGGCTAA ACGAGCCAAC

      1701                                                    1750                                                    1800
igrαβ CTGTGCAGTG TGCAAACTCA CGTATATAGT TGCTATGTGT AAGTAGTTTA TTACTTTTTG TTGCTC.... ACATATGGTA TTCTAGACTC CAAACTACTC
igrγδ CTGTGCAGTG TGCAAACTCA CGTATATAGT TGCTATGTGT AAGTAGTTTA TTACTTTTTG TTGCTCACAT ACATATAGTA TTCTAGACTC CAAACTACTG

      1801                                                    1850                                                    1900
igrαβ ATTGCATTGA ATAGTAAAAA AATATTCTTT TAGTTAGTGG AAAATGCGCA ATGTTCCTAG CTATCTTTTC AATGCCTTTG TATTTGGTAC AAAAATCAAG
igrγδ ATTGCATTGA ATAGTAAAAA AATATTCTTT TAGTTAGTGG AAAATGCGCA ATGTTCCTAG CTATCTTTTC AATGCCTCTG TATTTGGTAC AAAAATCAAG

      1901                                                    1950                                                    2000
igrαβ AAATGTTTTT AAGTATCCGT ACTAAATAAA CTGCCTATTA CTATATTATC TAAAGCATAA TTTCCGGGTT CATAGAGCTT AGAACCTGGG TAGGGTAGCT
igrγδ AAATGTTTTT AAGTATCCGT ACTAAATAAA CTGTCTATTA CTATATTATC TAAAGCATAA TTTCCGGGTT CATAGAGCTT AGAACTTGGG TAGGGTAGCT

      2001                                                    2050                                                    2100
igrαβ CACCAACTGA TCATGGAAAT TTATTGTTTG CGTTTGAGTA GAAACATTAT TTTGTGGAGT TTGAAAGTGA TGTGGTGGTA CTTATGCAAT AAACTAAGAT
igrγδ AACCAACTGA TCATGGAAAT TTATTGTTTG CGTTTGAGTA TAAACATTAT TTTGTGGAGT TTGAAAGTGA TGTGGTGGTA CTTATGAAAT AAACTAAGAT

      2101                                                    2150                                                    2200
igrαβ AATCAATCAA ACGGCAGGTG AAAGAAACTT ATAAAAAGTT GTATATGATA AAAAAATCTT AATATAATAG TATCCCATTC A..GGCGTGG CTACAACTTC
igrγδ AATCAATCAA ACGGCAGATG AAAGAAACTT ATAAAAAGTT GTATATGATA AAAAAATCTT AATATAATAG TATCCCATTC AGTGGCGTGG CTACAACTTC
```

```
      2201                                              2250                                               2300
igrαβ  GGGCGCCCGG GGCCAAGGAT GTTCTGCCGC CCCTCTTTGT TTACCTACCT ACAAGTTTCA TGAGGTGGTT CAAACAAAAG TGAATTTTTA CAAAATATCG
igrγδ  GGGCGCCCGG GGCCAAGGAT GTTCTGCCGC CCC.CTTTGT TTACCTACCT ACAAGTTTCA TGGGGTGGTT CAAACAAAAG TGAATTTTTA CAAAATATCT

      2301                                              2350                                               2400
igrαβ  TCGATATTAA GTATATAAAA TTTAGGAACT AGAAGCCACA CAAATTTTGA AGTTCCAAAA TTCTCACAAT ACGGTTTTTG AGGAAATTTA TAGTAAATTT
igrγδ  TCGATATTAA GTATATAAAA TTTAGGAACT AGAAGCCACA CAAATTTTGA AGTTCCAAAA TTCTCACAAT ACGGTTTTTG AGGAAATTTA TAGTACATTT

      2401                                              2450                                               2500
igrαβ  ACAAGACATC TTATTAAAGG CCGTAGAAAA AAAACCATTT GCCCTATATC TTGTACACTT TGGACACAAT TTGCCGGAAT TTTTGCCCGA ATTGGAGTTT
igrγδ  ACAAGACATC TTATTAAAAG CCATAGAAAA AAAACCATTT GCCCTATATC TTGTACACTT TGGACACAAT TTGCCGGAAT TTTTGCCCGA ATTGGAGTTT

      2501                                              2550                                               2600
igrαβ  TAAAAAAGAG GGAAGATCGT TTTGCCGCTC CCAAGACGTT GCGCCCGGGG CGACGGCCCC CTTTGTTTGT ACAAAATAAT TTAAATAAAG ACTTTAAACC
igrγδ  TAAAAAAGAG GGAAGATCGT TTTGCCGCTC CCAAGACGTT GCGCCCGGGG CGACGGCGCC CTTTGTTTGT ACAAAATAAT TTAAATAAAG ACTTTAGACC

      2601                                              2650                                               2700
igrαβ  TTTTAGTGCC GGGAGCCGAT ATATTGGCTT CAGCTGTACT TGCGAAAAGT GCCAGAGCCG ATATACTCCT ACTTATGTAT CGACTCGCGC CTTGTTGCGG
igrγδ  TTTTAGTACC GGGAGCCGAT ATATTGGCTT CAGCTGTACT TGCGAAAAGT GCCAGAGCCG ATATACTCCT ACTTATGTAT CGACTCGCAC CTTGTTGCGG

      2701                                              2750                                               2800
igrαβ  TTTACTAATC GCATTGCGGG AAAACCATAT CTCAAAATAA ATTTGTAA.. TGCGTCATAA ATGATCTAAT TTCATTA.CA AGATAATACG CAGCTTTTTG
igrγδ  TTTACTAATC GCATTGCGGG AAAACTCTAT CTCAAAATAA ATTTGTAATT TGCGTCATAA ATGATCTGAT TTCATTAGCA AGATAATACG CAGCTTTTTG

      2801                                              2850                                               2900
igrαβ  AATCATGTAA ATATTGTTTC TCGTTTGGTT TTTACAAGCA TTTTTCCGAA CCCTTGTAAA ACTGTGATAT TCGGACGAGC CTGTCGTAAC CTAATTTACC
igrγδ  AATCATGTAA ATATCGTTTC CGGTATCGTT TTTACAAGCA TTTTTCCGAA CCCTTGTAAA TCTGTGAAAT TCGGACGAGC CTGTCGTAAC ATTATTTACC

      2901                                              2950                                               3000
igrαβ  ACTCAACTTT GAAATCCTGG ACTCTCGCAA CAAGTAAAAG AAAACCATTT GACACTATTT TTGCCTAACT AAAACTTTCT GGAAACATGT TGAAACAAAT
igrγδ  ACTCAACTTT GAAATCCTGG ACTCTCTGAA CAAGTAAAAG AAAACAATTT GGCACTATTT TTGCCTAACT AAAACTTTCT GGAAACACGT TGAAACAAAT

      3001                                              3050                                               3100
igrαβ  TGTAAACAAT TGTAGTTCAA AGCTTGAGCT ACGCTTACAT TTACTGCGAA AATTCGATGA CTCATGAGGG TTGTCAGCGG CTGCATGTGT GCGGTGACCA
igrγδ  TGTAAACAAT TGTAGTTCAA AGCTTGAGCT ACGCTTACAT TTACTGCGAA AATTCGATGA CTCATGGGGG TTGTCAGCGG CTGCATGTGT GCGGTGACCA

      3101                                              3150                                               3200
igrαβ  AGAGCGACCA CCTTTTAATC CAGGGTGTTA TGCGACTCCC GTGTCCGTTG GATGATTTGC AACAAGGAGG TTAATTCGGC TGTATTCGAA CGAAACCTTT
igrγδ  AGAGCGACCA CCTTTTAATC CAGGGTGTTA TGCGACTCCC GTGTCCATTG GATGATTTGC AACAAGGAGG TTAATTCGGC TGTATTCGAA CAAAACCTTT

      3201                                              3250                                               3300
igrαβ  CCGACACCGG GCTGAATTTT GAAATGACGG TGGTCTTACC GCGTCACGGG GCTCTGGCGC GGCGGATCGT CTCGTTCCAG CTACGGCGTT GAG..TAAGG
igrγδ  CCGACACCGG GCTGAATTTT GAAATGACGG TGGCCTTATC GCGTCACGGG GCTCTGGGGC GGCGGGTCGT CTCGTTCCAG CTACGGCGTT GAGCCCAAGG
```

```
     3301                                                                      3350                                                                      3400
igrαβ  GTAGTACGAA AAGGAAGACA AAACCCAACA ACAACAACAA AACCAAACCA AAAATTACTA CCACAAGAAA ACAACAGACT CGGGTGACGG TTAAGGACAA
igrγδ  TCCGTACGAA AAGGAAGACA AAATCCAGCA ACAACAACAA AACCAAACCA AAAACTACTA CCACAAAAAA ACAACAGACT CGGGTGACGG TTAAGGACAA

     3401                                                                      3450                                                                      3500
igrαβ  CGTCAGCATC GCTGCAGTTA AAGAGTGTGT CGTTTAGCTT CCTGGACGCA CTTTGGAAGA CTTCTGAGGT ACGAGAAGTA CTAGCGCGAA ACAGGGTTAA
igrγδ  CGTCAGCATC GCTGCAGTTA AAGAGTGTGT CGTTTAGCTT CCTGGACGCA CTTTGGAAGA CTTCTGAGGT ACGAGAAGTA CTAGCGCGAA ACAGGGTTAA

     3501                                                                      3550                                                                      3600
igrαβ  ACGCAGTATT GTCTCTCCTG TATCGAAGGC AAGCAAAGTT TGCAAATCCG CATCTGTTGT CACTTGAGGT AGTGGGGCTA AGAGCGCCAA TCCCAACCGT
igrγδ  ACGCAGTATT GTCTCTCCTG TATCGAAGGA AAGCAAAGTT TGCAAATCCG CATCCATTGC CACTTGAGGT AGTGGGGCTA AGAGCGCCAA TCCCAACCGT

     3601                                                                      3650                                                                      3700
igrαβ  CAGGCGAGGT AAAGAAAAGG CAGCGCCTTC GATCACCCAA ACCGCGAAAC AGAAAAACCG GCAAATAAAT CCACAGGAGC CTCCCAGCTC TAAAAGGCGA
igrγδ  CAGGCGAGGT AAAGAAAAGG CAGTGCCTTC GACCACCCAA ACCGCGAAAC AGAAAAACCG GCAAATAAAT CCACAGGAGC CTCCCAGCTC TAA.......

     3701                                                                      3750                                                                      3800
igrαβ  GAAGTAACAA CGCACAGATG GCCGCCTCAA ATCTATTAAA AACTAGACAA GTTCAAGCAA GTGATTGATC GCAGGAATGT ACCAAAGCAT TAGCCAAAGG
igrγδ  ........AA CGTACAGATG GCCGCCTCAA ATCTATTAAA AACTAGACAA GTTCAAGCAA GTGATTGATC GCAGGAATGT ACCAAAGCAT TAGCCAAAGG

     3801                                                                      3850                                                                      3900
igrαβ  GTTTGGGTGA CCTCGGAGTT TAGATATATA TTTAATTCTT TTCAGATCGG ACCACTATAA TATGTAGCTG CT......AA AAAAAACCAT CAGCGATCAA
igrγδ  GTTTGGGTGA CCTCGGAGTT TAGATATATA TTTAATTCTT TTCAGATCGG ACCACTATAA TATGTAGCTG CTAAAAAAAA AAAAAACCAT CAGCGATCAA

     3901                                                                      3950                                                                      4000
igrαβ  AATCAAGTTC TTGTACGGAA AACTTTTTTA TTTGACAAAA TATCTTCACC AAATTTGGCA GCGCTACAAT ATCCGAAGAA ATTTCGCAGA TCGCACCTTT
igrγδ  AATCAAGTTC TTGTACGGAA AACTTTTTTA TTTGACAAAA TATCTTCACC AAATTTGGCA ACGCTACAAT ATCTGAAGAA ATTTCGCAGA TCGGACCTTT

     4001                                                                      4050                                                                      4100
igrαβ  ATAGATCAAA TAACTAAATC TGATCGATCG ATCGATATAA AGATTTTGCA TGGAAACTTT CTTATTTGTG AAAAGTATTT TACCTTCGAT GGAACCGAAC
igrγδ  ATAGATCAAA TAACTAAATC TGATCGATCG ATCGATATAA AGATTTTGCA TGGAAACTTT CTTATTTATG AAAAGTATTT TACCTTCGAT GGCACCGAAC

     4101                                                                      4150                                                                      4200
igrαβ  TTAACTGTTT TTCTTGTTTC GAATCAATTA TTTGAAATTG GTCTATAATT TCTTTTGCCG ATTTTTATTT ACATTATTAC TTATAA.... ..........
igrγδ  TTAACTGTTT TCCTTGTTTC GAATCAATTA TTTGAAATTG GTCTATAATT TCTTTTGCCG ATTTTTATTT ACATTATTAC TTATAATACT AAAGGGGATC

     4201                                                                      4250                                                                      4300
igrαβ  ..TTTCGAGT TTCCCTAGTT TTTTAAAGAA AAGACTCAGA AATTTCAAAT TTATTGAGGA ATGTTTATTA TCATTTGAAA GCACATTTTC TGGCATTTAT
igrγδ  CATTTCGAGT TTCCCTAGTT TTTTAAAGAA AAAATCCAGA AATTTCAAAT TTATTGAGGA ATGTTTATCA TCATTTGAAA GCACATTTTC TGGCATTTAT

     4301                                                                      4350                                                                      4400
igrαβ  TTTTTTAAG ATTATAACTT TCCAACGTTG GCCACTCGAC CAACTCAAAA CGTGAAATTA TCTG...... GAAGTGTTCT CTCTATAAAT CCATCAATTG
igrγδ  TTTTTTAAG ATTATGTCTT TCCAACGTTG GCCACTCGAC CAACTTAAAA CGTGAAATTA TCTGCTTACC GAAGTGTTCT CTCTATAAAT CCATCAATTG
```

```
       4401                                               4450                                                    4500
igrαβ  ATGCGCTGAG TGGGAAGTGG CAACATCTTG TTGAAACCTA ATGTCGCCGA GATCACGAAC TTCAATTTTA GGCATCAAAT AGTCGGTTAG CCATTGACGG
igrγδ  ATGCGCTGAG TGGGAAGTGG CAACATCTTG TTGAAACCTA ATGTCGCCGA GATCACGAAC TTCAATTTTA GGCATCAAAT AGTCGGTTAG CCATTGACGG

       4501                                               4550                                                    4600
igrαβ  TTATGTTCTC ACCGGCATCA TTTTGAAAAT GAACCACACT AGCCCGTTGC TTTTATGGAT GGAATGGCAG CTCTTAAATC TCTTCAGGTT GCTTTTGGTC
igrγδ  TTATGTTCTC ACCGGCATCA TTTTGAAAAT GAACCACACT AGCCCGTTGC TTTTATGGAT GGAATGGCAG CTCTTAAATC TCTTCAGGTT GCTTTTGGTC

       4601                                               4650                                                    4700
igrαβ  CCTAATGCGG CAATTTTGCT TGTTTACATA TACATTGTGC CAGAAATGAG CCTCATTGCT GAACAAAATT TTCCTCGATC ATCGGATCTT CTTGGAACTT
igrγδ  CCTAATGCGG CAATTTTGCT TGTTTACATA TACATTGTGC CAGAAATGAG CCTCATTGCT GAACAAAATT TTCCTCGATC ATCGGATCTT CTTGGAACTT

       4701                                               4750                                                    4800
igrαβ  TTCAAGAGCC CATAGAGTGA AACGATGTCG CTTGGGAAGG TTGAGCGGCA CTCTTTAATT TAATATGTCA ATGTAAAATG CACCTAGTTG TTCCATATGT
igrγδ  TTCAAGAGCC CATAGAGTGA AACGATGTCG CTTGGGAAGG TTGAGCGGCA CTCTTTAATT TAATATGTCA ATGTAAAATG CACCTAGTTG TTCCATATGT

       4801                                               4850                                                    4900
igrαβ  TAGTCCAAGT TGCTGCGAAC GGCGCTGAAT CGACTCCTCG GTCTTCGTGT ACACTCTCAG CTACGGCTGA GATATTTTCT TCACTGCGTG CTGCACATGA
igrγδ  TAGTCCAAGT TGCTGCGAAC GGCGCTGAAT CGACTCCTCG GTCTTCGTGT ACACTCTCAG CTACGGCTGA GATATTTTCT TCACTGCGTG CTGCACATGA

       4901                                               4950                                                    5000
igrαβ  TCTATTCGGT CGAATATTAT CCAATAATTC ATGCTGGTTC TCAAGATAGA TGATGATGTA GCGAATAATT CGCTCAATAA GTTGAGCGAA GGCGTAAGTC
igrγδ  TCTATTCGGT CGAATATTAT CCAATAATTC ATGCTGGTTC TCAAGATAGA TGATGATGTA GCGAATAATT CGCTCAATAA GTTGAGCGAA GGCGTAAGTC

       5001                                               5050                                                    5100
igrαβ  TTTCTACGAT GAAATGCCAA ACAATGCTGA AAAAAATAAC ATGACAGCTT GACACGACTC ACGCGTGATC TGTGTAGAAA AGGATATTGA AAAAGTATCT
igrγδ  TTTCTACGAT GAAATGCCAA ACAATGCTGA AAAAAATAAC ATGACAGCTT GACACGACTC ACGCGTGATC TGTGTAGAAA AGGATATTGA AAAAGTATCT

       5101                                               5150                                                    5200
igrαβ  CTACTTGGAT CACTCGTTAT ATGTCAACAA AATGCTCTGT GTTGTGTACA ATTTTAATTT GACCAATTCG GATTGGGTGG GTTGCTCATA TTTCCAACTT
igrγδ  CTACTTGGAT CACTCGTTAT ATGTCAACAA AATGCTCTGT GTTGTGTACA ATTTTAATTT GACCAATTCG GATTGGGTGG GTTGCTCATA TTTCCAACTT

       5201                                               5250                                                    5300
igrαβ  CTTGGTTCCT TTTTTATTTA ATGTTCGGTA AATTAATGAA TTTCGCCACC ACCAAAAACT TCGTAATCTG CATGCTTTTC GTAAACAAAG TGTTTTATTT
igrγδ  CTTGGTTCCT TTTTTATTTA ATGTTCGGTA AATTAATGAA TTTCGCCACC ACCAAAAACT TCGTAATCTG CATGCTTTTC GTAAACAAAG TGTTTTATTT

       5301                                               5350                                                    5400
igrαβ  ACCAAAATAT AACGGAAAGG CGGTTTCATA TAGCCCGTTA TGCAATGGTT GTTTTGAAAT GAAAATAATA ATATATTCCC ATATGTTGTC GTTTTGAAGA
igrγδ  ACCAAAATAT AACGGAAAGG CGGTTTCATA TAGCCCGTTA TGCAATGGTT GTTTTGAAAT GAAAATAATA ATATATTCCC ATATGTTGTC GTTTTGAAGA

       5401                                               5450                                                    5500
igrαβ  GTATTTCACA TATGTTGGTA ATATAAAACG CCGAAATTTC GCGAGCTCAT GTGTGTGGCC AACAAAGTTA AAAAAATAAA ATGAAAAAAT AAAAAAATTA
igrγδ  GTATTTCACA TATGTTGGTA ATATAAAACG CCGAAATTTC GCGAGCTCAT GTGTGTGGCC AACAAAGTTA AAAAAATAAA ATGAAAAAAT AAAAAAATTA


       5501                                               5550                                                    5600
```

```
igrαβ  AAAATAAAGA AAAATTGTAA AATTACTAAA CAAAAACAAC AATAGTTTGA TGTCAGCATG TTGCAATGCT CGCAGCCATC GGTGGGCAAG CAATTGTGCT
igrγδ  AAAATAAAGA AAAATTGTAA AATTACTAAA CAAAAACAAC AATAGTTTGA TGTCAGCATG TTGCAATGCT CGCAGCCATC GGTGGGCAAG CAATTGTGCT

       5601                                                5650                                                5700
igrαβ  AGCAAGCGAG CTTATCACTT TCTTTCTTAG CTTTAATCGC AAAAAAAGAA AAAACGCACA GATAAGCGAA ACTTACGTCT AAGCTGCGAT TGCACGCGCA
igrγδ  AGCAAGCGAG CTTATCACTT TCTTTCTTAG CTTTAATCGC AAAAAAAGAA AAAACGCACA GATAAGCGAA ACTTACGTCT AAGCTGCGAT TGCACGCGCA

       5701                                                5750                                                5800
igrαβ  GCAGTAGTTG TAGATAGCGA TTCCTTAAAT GATTTACAAA TCAATAAATG TATTTCTATT TCCCTGCAAC AGCAGTTTTC AAAGCGAGTT CAATGCACCC
igrγδ  GCAGTAGTTG TAGATAGCGA TTCCTTAAAT GATTTACAAA TCAATAAATG TATTTCTATT TCCCTGCAAC AGCAGTTTTC AAAGCGAGTT CAATGCACCC

       5801                                                5850                                                5900
igrαβ  GACTGAATTA ACGTTTTCTC AGCATTCTGA GTTGTGCTAC GCAAATAACT TTAAGGTATA AAAGCTATAT TTTCCATTTT GTCAACACCA CAGTTCAATT
igrγδ  GACTGAATTA ACGTTTTCTC AGCATTCTGA GTTGTGCTAC GCAAATAACT TTAAGGTATA AAAGCTATAT TTTCCATTTT GTCAACACCA CAGTTCAATT

       5901                                                5950                                                6000
yp2α   CGACTGCTCC CGGCATACGA CATAACGGCA ACGAAGAGCC ATGAGTCCTT TAAGTATTTT TTGTTTGGTG GCCCTGTTGG CTACAGCAAC ACCAGTGTGC
yp2δ   CGACTGCTCC CGGCATACGA CATAACGGCA ACGAAGAGCC ATGAGTCCTT TAAGTATTTT TTGTTTGGTG GCCCTGTTGG CTACAGCAAC ACCAGTGTGC

       6001                                                6050                                                6100
yp2α   GCTAGAGGCA ACTCAATACG TGATAACCTG AAACCCACTG AGTGGATTTC GCCGCGTGAA TTGGAGAATG CGCCTTCAGT GGATGAGATC ACTTTCGAAA
yp2δ   GCTAGAGGCA ACTCAATACG TGATAACCTG AAACCCACTG AGTGGATTTC GCCGCGTGAA TTGGAGAATG CGCCTTCAGT GGATGAGATC ACTTTCGAAA

       6101                                                6150                                                6200
yp2α   AGTTGCAGGA AACGCCGGCT GAGGAGGCTG CCGAGTTGGT GAACCAGATT TGTAAGTATT GAAGCGTCTG GTTTGGAATG TAAGATTCAA AGTAATGCAG
yp2δ   AGTTGCAGGA AACGCCGGCT GAGGAGGCTG CCGAGTTGGT GAACCAGATT TGTAAGTATT GAAGCGTCTG GTTTGGAATG TAAGATTCAA AGTAATGCAG

       6201                                                6250                                                6300
yp2α   GAAAATTTTA CTTGTAGACC ACTTGTCGCA GTTGAGCCGG AAAATTGAGC CCAGTTATGC CCCCAGTCCC AGCGATATTC CTGTCTACAC CTACACACCC
yp2δ   GAAAATTTTA CTTGTAGACC ACTTGTCGCA GTTGAGCCGG AAAATTGAGC CCAGTTATGC CCCCAGTCCC AGCGATATTC CTGTCTACAC CTACACACCC

       6301                                                6350                                                6400
yp2α   ACTGGTCAGC GCGTGAACAC CAAATTGAAT CAGCTGGTAT CCACCGTCCA ACAACAACCT CATTTCGGCC ACAAGAAGTC ACCATTTTCA TCACTGGTCT
yp2δ   ACTGGTCAGC GCGTGAACAC CAAATTGAAT CAGCTGGTAT CCACCGTCCA ACAACAACCT CATTTCGGCC ACAAGAAGTC ACCATTTTCA TCACTGGTCT

       6401                                                6450                                                6500
yp2α   GCCACAACAG AAACTGCGTG ACGCCGCACG CGCCAACAAG AAACTCATTC AGGCTTATTT GCAAGCCTAT AACGGACGAG TGCAAGTACA AGGTGGTCAA
yp2δ   GCCACAACAG AAACTGCGTG ACGCCGCACG CGCCAACAAG AAACTCATTC AGGCTTATTT GCAAGCCTAC AACGGACGAG TGCAAGTAAA AGGTGGTCAA

       6501                                                6550                                                6600
yp2α   GACGCTGACT CTGAACAGGA TACTTCATCG AGCGAGGAAT CTTCTGACAG CAAACAGACC CAACCTAGTG GTAATTTGGT GGTTATTGAT TTGGGCGCCG
yp2δ   GACGCTGACT CTGAACAGGA TACTTCATCG AGCGAGGAAA CTTCTAACAG CAAACAGACC CAACCTAGTG GTAATTTGGT GGTTATTGAT TTGGGCGCCG


       6601                                                6650                                                6700
yp2α   TCATACGCAA TTTCGAAGAA CTTGTTCTGC TCGATATCAA CCGCGTCGGT GCTGCGATCG GTAACAGTTT GGTCCAGCTT ACGTCGCAAA CTGATGTACC
```

```
yp2δ   TCATACGCAA TTTCGAAGAA CTTGTTCTGC TCGATATCAA CCGCGTCGGT GCTGCGATCG GTAACAGTTT GGTCCAACTT ACGTCGCAAA CTGATGTACC

       6701                                        6750                                                    6800
yp2α   CCAGGAAGTG ATCTATATTG TCGCACAAGG TATTGGCGCT CATGTTGCCG GTGCTGCTGC TCGCCAATGC ACACGTCAAG CAGGCAACAA GTTGCGTCGT
yp2δ   CCAGGAAGTG ATCTATATTG TCGCACAAGG TATTGGCGCT CATGTTGCCG GTGCTGCTGC TCGCCAATAC ACACGTCAAA CAGGCAACAA GTTGCGTCGT

       6801                                        6850                                                    6900
yp2α   ATCACTGCCA TGGATCCCAC AAAAATCTTT GCACGCAAAC CCAACACTTT GGTCGGCTTG GCTCGCGGCA ATGCTGATTT CGTTGATGCC ATCCACACTT
yp2δ   ATCACTGCCA TGGATCCCAC AAAAATCTTT GCACGCAAAC CCAACACTTT GGTCGGCTTG GCTCGCGGCA ATGCTGATTT CGTTGATGCC ATCCACACTT

       6901                                        6950                                                    7000
yp2α   CTGCTTATGG CTTGGGTAGC GCTGCTCGTG CCGGTGATGT TGACTTCTAT CCCAACGGTC CTTCTGTTGC CATGCCCGGA ACTGATAACA TAATCGAAGC
yp2δ   CTGCTTATGG CTTGGGTAGC GCTGCTCGTG CCGGTGATGT TGACTTCTAT CCCAACGGTC CTTCTGTTGC CATGCCCGGA ACTGATAACA TAATCGAAGC

       7001                                        7050                                                    7100
yp2α   TTCCTTACGT GCAACCCGTT ACTTCGCCGA GACAGTGCGC CCAGGTAATG ATCGCAACTT CCCAGCTGTC GCAGCCGAAT CCCTACAACA GTATAAAAAC
yp2δ   TTCCTTACGT GCAACCCGTT ACTTCGCCGA GACAGTGCGC CCAGGTAATG ATCGCAACTT CCCAGCTGTC GCAGCCGAAT CCCTACAACA GTATAAAAAC

       7101                                        7150                  7170
yp2α   AACAATGGCA ACGGCAGACG CGCTTATATG GGTATTGCCG CTGACTACGA TTTGGAAGGT GACTACATTC
yp2δ   AACAATGGCA ACGGCAGACG CGCTTATATG GGTATTGCCG CTGACTACGA TTTGGAAGGT GACTACATTC
```

Appendix 1. The DNA sequences of the *Yp1β*, *Yp2α*, *Yp1γ* and *Yp2δ* genes and the intergenic regions.

Primer binding sites are indicated by boxed text; changes in DNA sequence are indicated by unerlining; intron seqeunces are indicated by zigzag underlining (GATTTTT). Protein start codons indicated by double underlining (ATG). Gaps in sequence are indicated by dots (…).