A cDNA Subtraction Approach to Isolate Male-Specific Genes From *Ceratitis capitata*

A thesis presented in partial fulfilment of the
requirements for the degree of
Master of Science in Genetics
At Massey University, Palmerston North
New Zealand

Simon Francis Kahu Hills
2003

Abstract

The Mediterranean fruit fly, *Ceratitis capitata* (medfly), is a significant world wide agricultural pest. Sterile Insect Technique (SIT) is a biological method that has been used to control medfly successfully in several parts of the world for around two decades. SIT involves the release of sterile insects into wild populations which, due to sterile matings, lead to a reduction in the size of the wild population. The effectiveness of this technique is significantly increased when only sterile males are released. This can be achieved by using sexing strains, but these strains are prone to breakdown as a result of recombination leading to the disassociation of the selection gene from the Y chromosome. An alternative system, that could be more robust, would involve the control of the expression of the male determining gene.

The aim of this thesis was to identify the male determining gene of medfly by creating subtracted cDNA libraries enriched for male-specific transcripts. Subtracted libraries were made by subtractive suppression PCR, using the ClonTech cDNA subtraction kit. The libraries were screened by Southern hybridisation analysis using male and female total cDNA probes. Only one clone appeared to display a bias toward male-specific hybridisation, but this was found to be a result of unequal transfer of DNA. A selection of clones were individually used to probe membrane bound genomic DNA. These hybridisation analyses indicated a general lack of male–specific enrichment. In addition to this, sequence analysis of a selection of clones revealed a number of mitochondrial gene fragments, showing that there had been insufficient subtraction. As results indicated that the creation of subtracted, male-specifically enriched libraries had been unsuccessful another approach to the identification of the male determining gene was attempted. Genomic DNA was screened with an *Sxl* probe, under low stringency hybridisation conditions, to identify Y-linked encoding RNA binding proteins distantly related *Sxl*, which could represent the male determining gene. This screen showed that there were no male-specific RNA binding proteins, of the SXL family in medfly.

Acknowledgements

Firstly I must thank my supervisor Dr Max Scott for his guidance, advice and seemingly unending patience.

I would like to thank the IAEA for the grant that made this project possible. Thanks also to Manaaki Whenua - Landcare Research for a scholarship that allowed me to keep my financial head above water.

Many thanks to the members of the Fly Spot lab, past and present. To Rebecca and Jörg for getting me started in the lab. To Asela and Wayne for ensuring that the work place was never boring. And to all the others Emma, Vicky, Xuelei it has been a pleasure working with you. Thanks also to members of other labs who offered valuable advice or the use of equipment.

Finally, special thanks to my family, who were unsure if this would ever be finished. Without your support (and occasional funding) this would never have been possible.

Table of Contents

List of Figures

List of Tables

Abbreviations

α	alpha
β	beta
λ	lambda
°C	degrees Celsius
ADP	adenine diphosphate
ATP	adenine triphosphate
BLAST	basic logic Alignment Search Tool
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
DEPC	diethyl pyrocarbonate
DNA	deoxyribose nucleic acid
DNase	deoxyribonuclease
dNTPs	dinucleotide triphosphates
DTT	dithioreitol
EDTA	ethylenediamide tetraacetic acid
g	gram
IPTG	isopropyl-ß-D-thiogalactopyranoside
kb	kilo base pairs
L	litres
M	molar
MOPS	4-morpholine-propanesulphonic acid
mg	milligrams
mL	millilitres
mM	millimolar
µCi	micro Curies
µg	micrograms
µL	microlitres
mRNA	messenger RNA
ng	nanograms
nm	nanometres

OD	optical density
Ω	ohms
PCR	polymerase chain reaction
Poly A RNA	polyadenylated RNA
RFLP	restriction fragment length ploymorphisim
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
RT-PCR	reverse transcriptase PCR
SIT	sterile insect technique
UTR	untranslated region
UV	Ultraviolet
V	volts
v/v	volume per volume
w/v	weight per volume
X-gal	5-Bromo-4-Chloro-3indolyl β-D-galactopyranoside

1 Introduction

1.1 The Mediterranean Fruit Fly

The Mediterranean fruit fly (medfly) *Ceratitis capitata* is one of the world's most significant agricultural pests. Medflies lay their eggs on ripe, tree bound fruit, where the eggs hatch, and the larvae feed. In addition to this the Medfly has a very broad host range of over 200 host plant species (Carry 1991, Sheppard *et al.* 1992, McPheron *et al.* 1994), around 100 of which are of economical importance. This coupled with a wide distribution of the pest (White and Elson-Harris 1992), leads to large annual income loss due to medfly infestation. World wide losses due to medfly have been estimated to be in the range of hundreds of millions of dollars annually (Ashburner 1995).

The medfly is thought to have originated in tropical Africa, and the Mediterranean region. From there it spread through southern Africa, Australia, Hawaii, Central and South America (Davies *et al.* 1999). It continues to threaten agriculture in the USA, Mexico, and Asia. Outbreaks and infestations occasionally occur in other regions such as New Zealand. These small scale outbreaks are usually a result of the importation of contaminated fruit.

There has been a large amount of study into the population genetics and bionomics of medfly, because of its economic impact. A reasonable proportion of this work has been devoted to developing genetic identification methods that would allow an infestation to be traced back to its source population. It is thought that this will aid in the understanding of infestations, and lead to a better ability to control outbreaks at an early stage, before major damage is caused by the pest.

1.2 Control of Medfly

Due to the large economic impact of medfly, there have been many different eradication and pest management strategies developed to deal with the pest. More traditional control methods consist of insecticide bait sprays. These usually have to be applied 2 to 8 times per year, and depend greatly on weather conditions. These poison drops also represent important ecological and health issues. Large scale use of pesticide sprays can have major detrimental ecological effects by eradicating potentially beneficial, non-target species (Oakeshott *et al.* 1993). Studies into the effects of these spraying programs on human health, have shown that the wide spread use of pesticides can cause illness among residents of the target area (Oakeshott *et al.* 1993). Furthermore the effectiveness of these sprays is questionable. Elimination of a large population in a single spray is virtually impossible, so continual use of the spray is required to control the pest. Continual use of the spray can lead to selection for resistance to the pesticide in the pest species. After resistance has developed the use of the pesticide will provide selection pressure that will allow the resistant population to thrive. Once resistance has developed the spraying processes must be carried out with a new pesticide. In addition to this the extremely wide distribution of medfly means that effective spraying operations are large and expensive.

A system which does not involve large scale use of potentially dangerous chemicals is needed to effectively control the medfly, and other pests. The most effective system that fits these parameters, that has been implemented for the control of several insect species, is known as Sterile Insect Technique.

1.3 Sterile Insect Technique

The sterile insect technique (SIT) was first developed by Dr E. Knipling to control insect populations on a large scale (Knipling 1960). The technique involves the mass rearing and release of large numbers of sterile insects. The sterile insects released compete with the wild population for mates. Every mating between a sterile individual and a wild individual will not produce progeny. In this way large numbers of sterile insects competing for mating will cause a marked decrease in the numbers, of the wild population, in the subsequent generation. When this type of release

program is repeated over several generations of the target species, the wild population can be effectively wiped out completely.

SIT was first used to control populations of the New World Screwworm in the South-Eastern USA, in 1957.

The technique has possible applications for many species of agricultural pests, and some species that carry disease. Indeed, sterile insect technique has been used successfully to control several insect species including; tsetse flies, codling moth, melon fly, Mexican fruit fly, and the Mediterranean fruit fly. Several species of mosquitoes have also been trialled for SIT programs. In addition to these, a number of species of insect are currently under development for use in SIT, including the Australian sheep blow fly *Lucilia cuprina*.

1.3.1 Medfly in SIT

C. capitata has been the target of several sterile insect control programs. The largest control effort was the MOSCAMED program in Central America (Linares and Valenzula 1993, Liedo *et al.* 1993, Orozco *et al.* 1994). This program was developed by the Mexican government, in conjunction with the USA and Guatemala, after the detection of medfly in Guatemala in 1976. This followed the rapid infestation of other parts of Central America by medfly. By 1980 a mass rearing facility had been established, and was producing its goal of 500 million sterile flies per week. Combined with small scale chemical suppression of the wild population, the sterile insect release program led to the eradication of wild medfly from Mexico by 1982. After this a 'sterile fly barrier' was set up at the Mexico-Guatemala border to prevent the northward spread of the fruit fly into Mexico, and the USA beyond. This 'sterile fly barrier is still maintained, and it protects the fruit crop industry in the lower USA such as in the California region.

1.3.2 Sexing Systems

SIT is much more efficient when only sterile males are released (Economopoulos 1996). It is advantageous to release males only, as females even though they are sterile and cannot lay fertile eggs, will still attempt to lay and cause damage to the fruit. The damage to the fruit from sterile females can then be invaded by other pathogens, such as fungi and bacteria (Ashburner 1995). This would undermine the

effectiveness of the system in areas where fruit crops are in immediate threat. Also the effectiveness of the technique can be increased with the release of males only, as this will eliminate the occurrence of matings between sterile flies (Franz and Kerremans 1994). Another advantage of releasing only sterile males is that a single male fly will mate with many females, unlike the female which will mate only once.

Sex separation systems have been developed to allow the separation of male and female flies. These systems are based on morphological, or biochemical differences that allow mass separation of the sexes. Where no adequate natural differences are found, genetically engineered sexing strains need to be developed. Genetic sexing strains provide a system whereby the different sexes can be easily separated so only males are released.

Sexing strains are based on translocations between a chromosome containing one allele, usually the wild-type, of a chosen selection gene, and the sex determining Y chromosome. Females are homozygous recessive for the mutant allele of the selection gene. All males will express the wild-type phenotype for the mutant allele of the selection gene as it is attached to the male determining chromosome. The selection system distinguishes between mutant and wild-type expression, and selects for only wild-type there by producing only males.

The downfall of this translocation based system is that it is prone to breakdown after several generations. This breakdown is the result of recombination in the males which uncouples the wild-type selection marker gene from the sex determining chromosome (Willhoeft *et al.* 1996, Franz *et al.* 1996). In the subsequent generations this produces wild-type females and mutant males, thus rendering the selection system ineffective.

Sexing strains can be divided into two general categories, sex sorting and sex killing (Robinson 1990, Rossler *et al.* 1994).

Sex sorting involves physical separation of males from females based on observable differences. These strains are based on genes where different morphologies can be seen in the sexes before adulthood. This makes males easily distinguishable from females. Females can then be physically removed. Sex sorting systems are hampered by a need for specialised sorting equipment, and that this equipment is not totally accurate and can damage pupae.

Sex killing strains are based on selection genes which have biological function. Mutant genes used for such strains cause an inability to complete a metabolic

function, or provide resistance to an environmental condition. Selection is achieved by applying the correct environmental condition to induce female death, during the rearing process. Genes that have been used for this type of sexing strain include; pesticide resistance (Wood 1990), alcohol dehydrogenase (Riva Francocs 1990), and temperature sensitive lethals (Franz *et al.* 1996). The problem with these sexing strains is that they require the addition of expensive chemical agents to induce female specific death.

1.3.3 Medfly Sexing Systems

Research has established that there are no natural sexually dimorphic traits or mutations that can be used for separation of male and female medflies. This has lead to the development of sexing strains produced by genetic modification. Two widely used sexing strains use pupal colour, or temperature sensitive lethals as selection genes.

Pupal colour was used in the first medfly sexing strain that was applied in the field. Males express the wild-type phenotype, and have brown pupae. Females express the mutant phenotype, which is black (Walder 1990, Zapter 1990), but white pupal mutants have been utilised as well (Economopoulos 1990). The sexes are separated using optical sorting machines. There are a number of genes which have been identified as having a role in pupal colour. Several of these have been used for the creation of sexing strains.

More recently sexing stains based on temperature sensitive lethal genes have been developed (Franz *et al.* 1996, Rendon *et al.* 1996). These sex killing strains result in female specific death after heat shock. Temperature sensitive lethal sexing strains are replacing the pupal colour based strains in medfly rearing plants. This is due to the cost reduction of not needing sorting equipment and, because selection is applied in late embryogenesis, savings are made in larval feed costs.

1.4 Sex Determination

1.4.1 Drosophila

The most well studied Dipteran sex determination system is that of *Drosophila melanogaster*. In *D. melanogaster* sex is determined by the ratio of X chromosomes to autosomes (A). Male development is determined by a ratio of 1X:2A, and female development is determined by 2X:2A (for review see Cline 1993). In females the chromosomal ratio switches on the key gene *Sex-lethal* (*Sxl*). SXL is an RNA binding protein that leads to the splicing of its targets. SXL binds both *transformer* (*tra*) pre-mRNA, and its own RNA. Splicing of its own RNA leads to the autoregulation of *Sxl*, continuing female specific expression. Unspliced *Sxl* mRNA does not produce functional protein due to the presence of multiple in-frame stop codons.

The splicing of *tra* is the next step in the sex determination pathway, and is mediated by SXL. Without splicing *tra* mRNA, like *Sxl*, will not produce functional protein. TRA protein resulting from correct splicing of *tra* mRNA promotes the splicing of *double-sex* (*dsx*). DSX is a transcription factor, which controls the global expression of further sex specific genes. The splicing of *dsx* mRNA is dependent on the presence of both TRA and the constitutively expressed TRA-2. In the presence of both TRA and TRA-2, *dsx* is spliced female-specifically. When functional TRA is not present *dsx* is spliced male-specifically. This leads to two sex-specific forms of the DSX protein, DSX^M and DSX^F. Male specific DSX^M represses female specific genes and activates male specific genes. DSX^F represses male specific genes and activates female specific genes. These two gene products differ in the C terminal ends. This difference provides the sex specific response, otherwise the proteins are identical. The conserved amino terminal domain is responsible for DNA binding (Schutt and Nothiger 2000). It has been demonstrated that this system is well conserved across the *Drosophila* genus (Erickson and Cline 1998). Figure 1.1 gives a graphic summary of the *Drosophila* sex determination pathway.

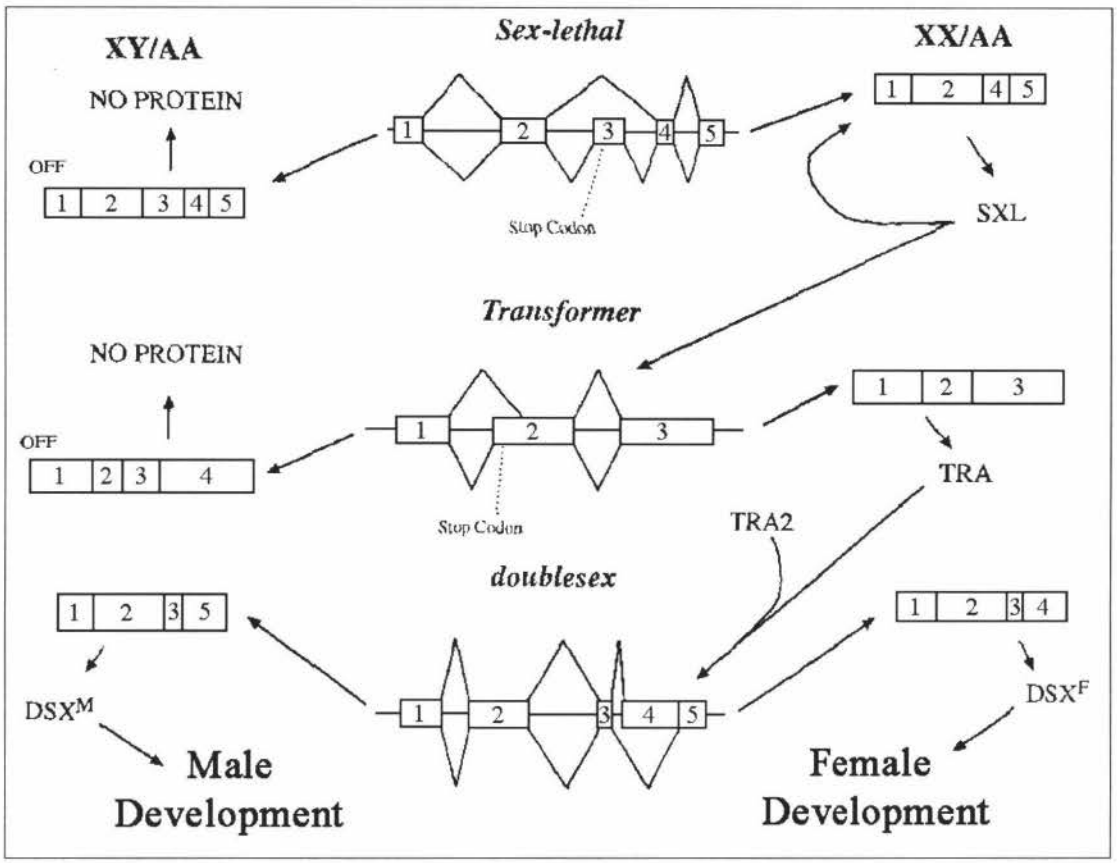

Figure 1.1: The pathway of sex determination in *D. melanogaster*.

In females the X:A ratio signals for female specific splicing of the mRNA of the *Sxl* gene. The SXL protein then promotes female specific splicing of *tra*. TRA then promotes female specific splicing *dsx*, which controls global female gene expression. In males *sxl* is not spliced. An in-frame stop codon leads to translation non-functional protein, which is degraded. Without SXL protein, *tra* is not spliced. In the same way as *Sxl*, an in-frame stop in *tra* codon leads to non-functional protein which is degraded. In the absence of TRA, *dsx* defaults to a male specific form. DSX^M then controls the global expression of male specific genes.

(Figure adapted from Saccone et al 2002)

1.4.2 Other Dipteran flies

Across the greater group of Diptera, comparisons have found that the *Drosophila* system is not widely conserved (reviewed in Schutt and Nothiger 2000).

As *sex-lethal* is the master switch gene it has been a focal point in the studies of other insect sex determination systems. In the housefly *Musca domestica* the *sex-lethal* gene has been found to be highly conserved. However the SXL protein is expressed in the same form in both sexes in the housefly (Meise *et al.* 1998). In addition to the differences in *sxl*, the primary sex determination signal is different in *M. domestica*. The determination of sex in housefly is dependent on the M and F genes. M determines maleness, and is the primary switch. F is required for female development, and is thought to represent a key gene, which is analogous in function to *sxl* (Hilfiker-Kleiner *et al.* 1993). The activity of the M element has been demonstrated to consist of two parts. One is located on the short arm of the Y chromosome, and the other on the long arm. The separate activities of these elements appear to be additive, where the presence of only one copy will result in an intersex phenotype (Hediger *et al.* 1998). Neither M nor F have been characterised at this time.

The system seen in *M. domestica* appears to be similar to *L. cuprina* and could be similar to *C. capitata*. It is probable that the *L. cuprina* sex determination system is more like that of *C. capitata* then the *M. domestica* system. This is because the M factor of *M. domestica* has been found in different linkage groups in different strains, meaning that the gene can be translocated and remain active. *L. cuprina* and *C. capitata* both appear to have M factors that are immobile on the Y chromosome (Shearman 2002).

In contrast to the interspecific sequence similarities and functional differences observed in *sxl*, there have been studies which have shown that there is a high degree of difference in the coding sequence of *tra*. Even within the *Drosophila* genus *tra* exhibits a high degree of divergence, and is considered to be one of the fastest evolving *Drosophila* genes (O'Neil and Belote 1992). However evidence suggests that *tra* has functional relatives in other species (Pane *et al.* 2002), and that *tra-2* has functional conservation in a species as distantly related as humans (Dauwalder *et al.* 1996).

Studies have indicated that *dsx* may be highly conserved across a broad range of species (Marin and Baker 1998, Graham 2003). A *dsx* homologue has been identified in *Bactrocera tryoni* (Queensland fruit fly) (Shearman and Frommer 1998), and evidence suggests that the distantly related *Megaselia scalaris* also exhibits sex-specific *dsx* RNAs (Kuhn *et al.* 2000). Evidence exists for the conservation of *dsx* homologues across greater evolutionary distances, given similarities seen in the *dsx* of *D. melanogaster* and the *Mab-3* gene of *C. elegans* (Marin and Baker 1998)

This evidence indicates that sex determination cascades evolve from the bottom up. The most ancestral and well conserved genes (e.g. *dsx*) occupy the bottom of the pathway, and the more recently evolved or co-opted genes occupy the top. This provides evidence that the master control genes and primary signals at the top of sex determination pathways can be variable between different species (Marin and Baker 1998, Graham *et al.* 2003)

1.4.3 Medfly

Male development is determined by the presence of the Y chromosome in *Ceratitis capitata*. Specifically there is a dominant male determining region located on the long arm of the *C. capitata* Y chromosome, near the centromere (Willhoeft and Franz 1996). This appears to be similar to the sex determination system found in both *M. domestica* and *L. cuprina*. It is currently unknown whether this region contains one or several male determining factors, although theoretically one would be sufficient.

Studies with *C. capitata* have demonstrated that the *sex lethal* protein sequence is highly conserved, but uniformly expressed in both sexes (Saccone *et al.* 1998). Further, no evidence has been found for sex-specific splicing. This indicates that *sxl* has no role in sex determination in medfly.

Recent studies with *C. capitata tra* have demonstrated that while the sequence is not conserved, its function is similar to *D. melanogaster*. At the sequence level *C. capitata tra* only has 32-40% identity to *D. melanogaster tra* (Graham 2003). Pane *et al.* (2002) have demonstrated that *C. capitata tra* is required for female specific splicing of *dsx*. In addition to this the presence of TRA/TRA2 binding elements indicate that a *C. capitata tra-2* homologue may exist (Pane *et al.* 2002). Pane *et al.* (2002) also demonstrated that *tra* exhibits a self-regulation loop much the same as *D. melanogaster sxl*. As a result of this *tra* work a model for *Ceratitis capitata* sex determination has been suggested (Figure 1.2).

Compared to *D. melanogaster*, the *C. capitata dsx* control function is quite well conserved. TRA/TRA-2 binding sites are observed in similar positions in the *dsx* sequences of *D. melanogaster and C. capitata* (Saccone t al 2002). In addition to this the structure of *dsx* mRNA is conserved between the two species (Saccone et al 1998).

1.5 Male Determination and SIT

The medfly male determining gene would allow females to be transformed into males by induction during early development. This could decrease the cost of SIT by eliminating the waste of resources, which results from the removal females from the release population. In addition it will alleviate the problem of breakdown of sexing strains as a result of Y linkage collapsing. The system may be adapted to work in related pest species such as *Ceratitis rosa*, which is a significant pest in South Africa. Utilisation of this in SIT would involve the creation of a construct containing the male determining gene under the control of a heat-shock promoter. Medfly successfully transformed with this construct would normally not express this copy of the male-determining gene. However after incubation at a temperature that activates the promoter, expression of the transformed male-determining gene would then be activated. This heat shock would have to be applied during early development for male development to be properly induced. Thus heat shock of a population of a strain containing such a construct would result in an entirely male population.

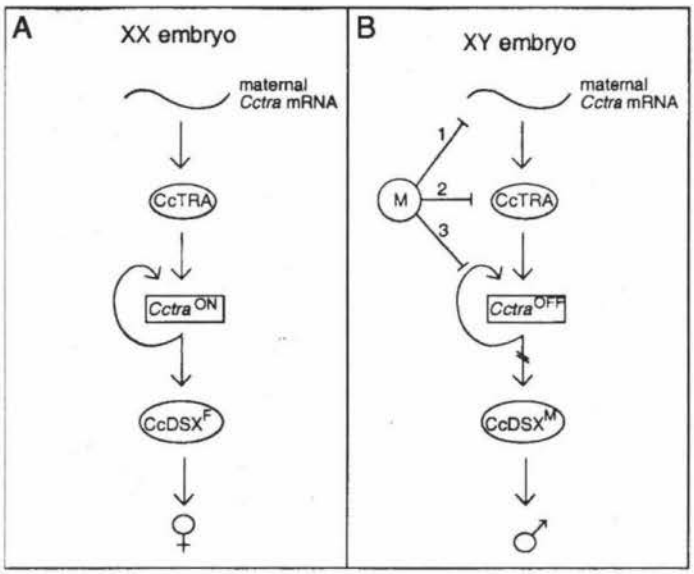

Figure 1.2: A model for medfly sex determination.

In the female medfly *tra* mRNA is spliced into an active from that is capable of autoregulating its own expression, there by maintaining female expression (A). In the male a male determining factor prevents splicing of *tra*, so an in-frame stop codon leads to an inactive protein (B).

The Male determining factor of medfly could act in one of three ways; prevention of translation of *tra* mRNA (1), direct inhibition of the TRA protein (2), or interference with TRA autoregulation to cause its function to be switched off (3). (Figure from Pane et al. 2002)

1.6 cDNA Subtraction

Subtractive cloning is an effective method for isolating differentially expressed genes in populations of mRNA or DNA. The technique is regularly used to identify novel genes expressed in cancer cells by comparing with gene expression of non-cancer cells. It has also been used for the identification of developmental genes by comparison of cells at different developmental stages. In insects the technique has been used successfully to identify a novel cold shock recovery gene (Goto 2001), and a novel early developmental patterning gene (Taylor 2000) in *D. melanogaster.*

The subtraction method requires two populations of mRNA or DNA, a 'tester' sample and a 'driver' sample. The tester sample contains the target fragments while the driver sample contains everything in the tester except the target sequences. The method works by denaturing the tester and driver populations, allowing them to hybridise together with an excess of driver, then all driver/driver and tester/driver hybrids are removed. This process is then repeated a number of times to ensure efficient subtraction. In general there are two approaches to subtraction, which differ in the method for removing driver hybrids. Traditional methods make use of biotinylated driver and remove hybrids with streptavidin and phenol extraction (Ausubel *et al.* 1997, Sive and St. John 1988). More recently PCR based subtraction methods have become available. PCR based subtraction techniques have several advantages over traditional methods. Firstly the PCR based method is quicker and simpler to use. Secondly the PCR based method is more sensitive, and can isolate rare fragments. Thirdly, commercial cDNA subtraction kits are available, which make the process significantly more efficient.

PCR based cDNA subtraction makes use of two techniques to produce a subtracted and enriched library, subtractive hybridisation and suppression PCR. Subtractive hybridisation allows removal of sequences that are common to two populations of cDNAs. Suppression PCR allows only specific target sequences to be amplified. The Clontech PCR-Select™ cDNA subtraction kit utilises amplification primers embedded in adapters, which are ligated onto samples of the tester and driver cDNAs. Through a combination of hybridisations and PCR subtraction is achieved. Figure 1.3 illustrates the events of subtraction using the Clontech kit.

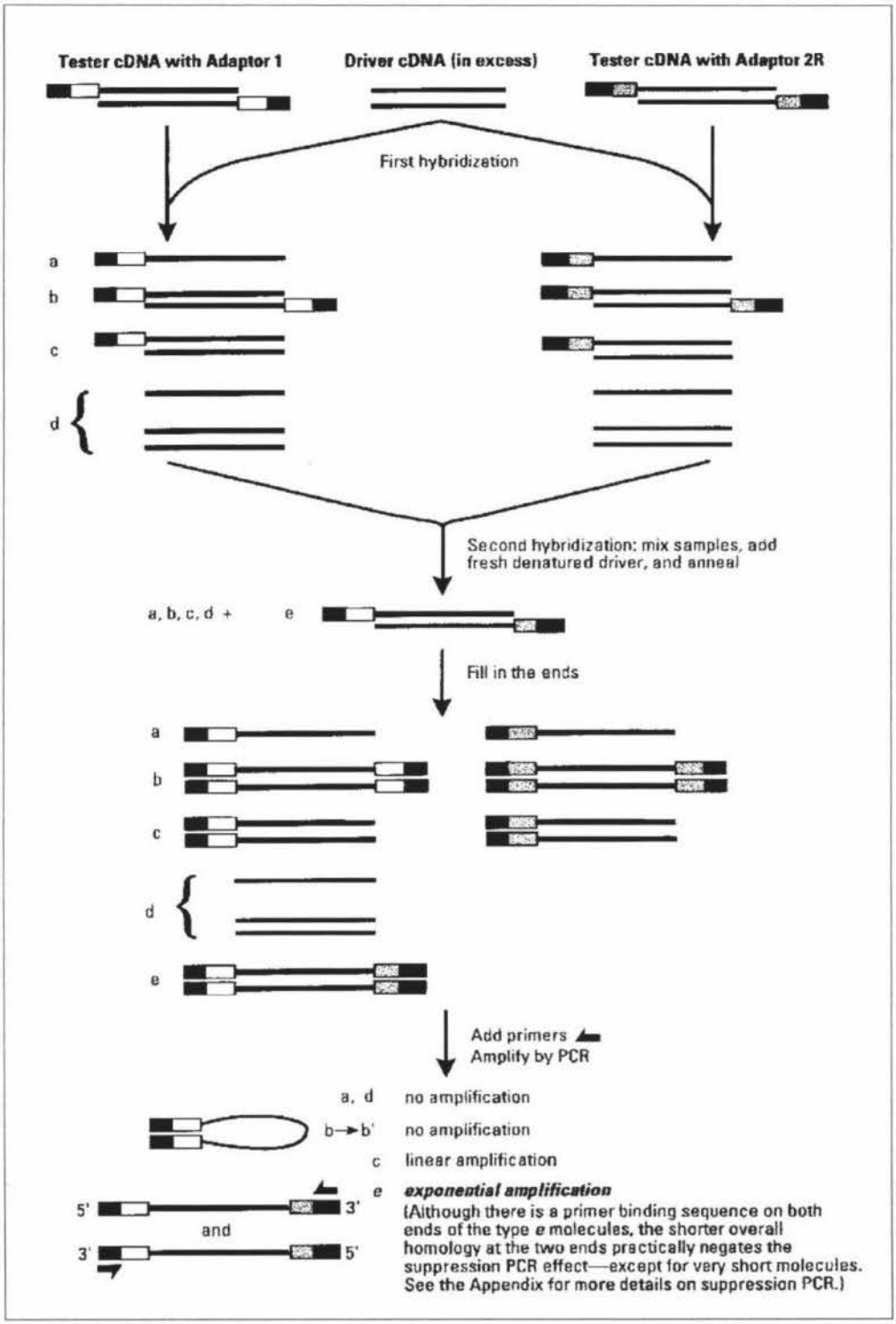

Figure 1.3: A schematic diagram of the events of PCR-Select cDNA subtraction. Bold lines represent *Rsa* I digested cDNA. Solid boxes represent the parts of two adapters that are the same, and contain a common PCR priming site. The clear and shaded boxes represent the parts of the two adapters that are different these contain specific nested PCR primers, that are unique to each adapter. (Figure taken from CLONTECH PCR-Select^{TM} cDNA Subtraction Kit Users Manual)

1.7 Research Objectives

The aim of the research detailed herein was to attempt to isolate the male determining gene of *Ceratitis capitata* by screening a subtracted cDNA library for likely candidates.

This aim was to be achieved by through the following objectives:

- Screening an existing subtracted cDNA library for male-specific candidates
- confirming candidates by Southern hybridisation with total genomic DNA targets
- sequence candidates, and identify known homologues

When the first subtracted library was found to be of limited quality the aim of this research was expanded to include the production of a subtracted cDNA library. The objectives of the project at this time were to:

- produce a subtracted, male-specifically enriched cDNA library
- screen subtracted library by colony hybridisation
- confirm candidates by Southern hybridisation
- sequence candidates, and identify known homologues

After the second subtracted library was found to lack any male-specific sequences a contingency experiment was conducted. The aim of this experiment was to identify male determining gene candidates through the identification of male-specific genes encoding RNA binding proteins. This would be achieved screening genomic DNA for Y-linked genes encoding RNA binding proteins by low stringency Southern hybridisation.