University of Otago

BIOMEDICAL SCIENCES

RESEARCH REPORT

Genome Architecture and Phenotypic Plasticity: Is the *Lethal (2) Essential for Life* cluster epigenetically regulated during ovary activation in the honeybee, *Apis mellifera*?

Mackenzie Lovegrove

A report submitted in partial fulfilment of the Degree of Bachelor of Biomedical Science with Honours.

Genetics, University of Otago, Dunedin. New Zealand

October 2013

ABSTRACT

Phenotypic plasticity is the ability of an organism to alter its phenotype, without altering its genome, in response to environmental cues. There is mounting evidence it is involved in human development, where it has been implicated in the risk of developing noncommunicable adult diseases. Studying the molecular basis of this in mammals can be difficult, particularly separating out single influences from complex environmental interactions. The honey bee, *Apis mellifera*, provides a useful model in which to study plasticity because of its well-controlled, easily triggered plastic responses. Queen bees are normally the only reproductively active females within a hive, but workers can activate their ovaries in response to the loss of the queen. During this process, over a third of the genome shows altered gene expression, implying that coordinated gene regulation within a chromatin domain may play a role. We have identified a candidate cluster for investigating this hypothesis, the *Lethal (2) Essential for Life* (*L(2)efl*) group. The genes of which are down-regulated as the workers undergo ovary activation. The findings of this study show that the original boundaries of the chromatin domain had been underestimated, and that the CTCF insulator element binding sites which flank the genes of the *Lethal(2)efl* cluster, *LOC100576174* and *Gmap,* appear to be the boundaries of the coordinated regulation. All of the genes within these sites show co-ordinated regulation, with expression occurring in the terminal filament cells of the ovary in queens, workers and active workers. As ovary activation is a phenotypically plastic response to an environmental cue, it was hypothesised that the mechanisms which underlie it are epigenetic in nature, with previous work identifying the repressive histone mark H3K27me3 as likely playing a role in ovary activation. Potential binding sites for the ecdysteroid-regulated transcription factors BR-C

Z1 and Z4 were found for all of the genes within the CTCF binding sites, and none directly outside it (*LOC411452* and *LOC412824*). The proposed model for the coordinated regulation of the genes within the chromatin domain containing the *L(2)efl* group is through an interaction of both histone modifications and ecdysteroid-regulated transcription factors. This work provides evidence for large scale, coordinated changes in gene expression leading to phenotypic plasticity in response to an environmental influence.

ABBREVIATIONS

~ = approximately

Amp = ampicillin

AP = anti-dioxygenin-alkaline phosphatase (AP) antibody

BCIP = 5-Bromo-4-chloro-3-indolyl phosphate

BLAST = Basic local alignment search tool

Bp = base pairs

BR-C = Broad complex

BSA = bovine serum albumin

cDNA = complementary DNA

CF2-II = chorion factor 2

ChIP = Chromatin immunoprecipitation

Croc = Crocodile

DAPI = 4',6-diamidino-2-phenyl-indole, dihydrochloride

dH_2O = distilled water

DIG = digoxygenin

dl = dorsal

dNTP = deoxyribonucleotide triphosphate

Elf-1 = Grainy Head

gDNA = genomic DNA

HSF = Heat shock factor

Kb = kilobase

Kr = Kruppel

LB = Luria Broth

MeOH = methanol

Min = minute

NBT = nitro blue tetrazolium chloride

PBS = Phosphate-buffered saline

PTw = 0.1% Tween-20 in 1x PBS

qPCR = Quantitative Polymerase Chain Reaction

-RT = minus Reverse transcriptase

RT-qPCR = Real-time Quantitative Polymerase Chain Reaction

SB = Sodium borate

TSS = Transcriptional start site

UV = ultraviolet

w/v = weight per volume

TABLE OF CONTENTS

1.1Phenotypic Plasticity

1.1.1 Phenotypic plasticity in nature

Phenotypic plasticity is defined as the ability of an organism to alter its phenotype in response to environmental cues, without altering its genome (Bateson et al., 2004; West-Eberhard, 1989). There is evidence of phenotypic plasticity occurring in almost all animals and plants (Doughty and Reznick, 2004; Schlichting, 1986) with one classical example being that of the Meadow vole (*Mictous pennsylvanicus*), which shows phenotypic plasticity in the form of a predictive adaptive response (PAR). Predictive adaptive responses are a type of plasticity that occurs in response to environmental cues in the early life stages. These are hypothesised to confer benefit to the organism by allowing it to adapt to the anticipated environment of later life (Gluckman et al., 2005). In this example, the coat thickness of the meadow vole offspring is determined prenatally by the maternal sensing of day length. This is presumed to be due to the transplacental movement of melatonin, a hormone which is secreted nocturnally from the pineal gland (Lee and Zucker, 1988). This allows the meadow vole to anticipate the seasonal conditions into which it will be born, and then adapt accordingly by altering the phenotype of its coat thickness. In this way the meadow vole does not spend time being maladapted, but instead increase their fitness by having an appropriate coat already developed.

1.1.2 Phenotypic plasticity in human development

There is mounting evidence that phenotypic plasticity is involved in human development (Kaati et al., 2002; Kaati et al., 2007; Pembrey et al., 2006). It has been implicated in the risk of developing non-communicable adult diseases due to a mismatch occurring between

anticipated and actual environments, leading to pathogenicity (Fig.1.1) (Gluckman and Hanson, 2004). For example, as mismatch may occur when a foetus is under nourished *in utero*, and so 'anticipates' an environment where resources are scarce after birth, thus altering gene expression to ensure that as much energy as possible is stored and conserved. If the child is born into a world of plenty that it is not adapted for, then this is a mismatch. Mismatch is thought to lead to disease phenotypes such as obesity through a the individual being 'primed' to store energy as fat, as well as a predisposition towards reduced activity in order to conserve energy (Forrester et al., 2012; Gluckman et al., 2007; Vickers et al., 2005). Many diseases implicated as being the result of phenotypic plasticity are common in the populations of the developed world. Examples include; osteoporosis (Dennison et al., 2001), polycystic ovarian syndrome (Ibáñez et al., 2001), mood disorders and psychoses (Thompson et al., 2001; Wahlbeck et al., 2001), hypertension (Stanner et al., 1997), cardiovascular disease (Kaati et al., 2002) and type 2 diabetes mellitus (Stanner et al., 1997). This gives an indication of the varied nature of the systems affected.

It has been suggested that organ systems are more susceptible to environmental damage at different stages of development, due to the varied windows of plasticity. This was highlighted in work which investigated individuals subjected to reduced nutrition *in utero*, as their mothers were malnourished during the Dutch famine. It was discovered that individuals affected in mid gestation had higher rates of obstructive airway disease, whereas those affected in later gestation showed a decreased glucose tolerance (Painter et al., 2005), highlighting that the timing of the environmental influence is critical to the disease phenotype.

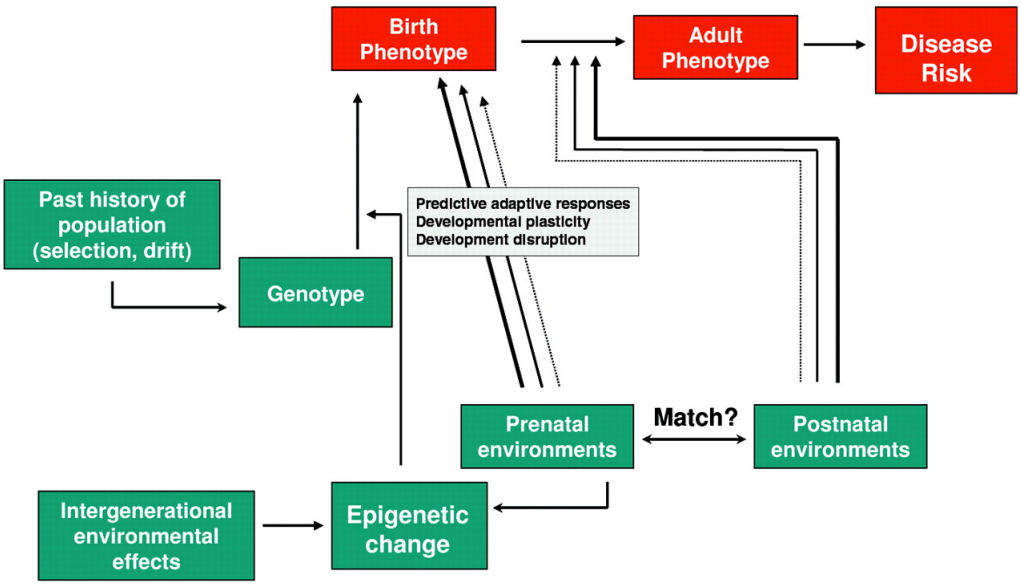


Figure 1.1: A proposed model for the interactions between environment, intergenerational effects and genetics that influence disease risk in later life. A mismatch between pre and post natal environments may be pathogenic. Adapted from Gluckman and Hanson, 2004.

Rodent models have provided a means of studying the role early environment has on phenotype, either prenatally or postnatally. The prenatal role was highlighted by maternal under-nutrition leading to the offspring being predisposed to select fat and carbohydrate rich foods, and so suggested a difference in appetite control and regulation (Bellinger et al., 2004). One example of a postnatal effect was the increased stress response of offspring that were shown poor maternal care, such as reduced licking (Francis et al., 1999). These studies imply that early environment plays a vital role in the disease risk of the adult.

The main method of gathering data in human studies is epidemiological, the major flaw of which is its retrospective nature. This is useful for understanding the correlation between an environmental influence and a disease phenotype, but not for discovering causation or mechanisms. Animal studies allow for some insight into phenotypic plasticity, however animal trials are often both expensive, and largely restricted in the scope of environmental

manipulation that is possible or permissible by ethical guidelines. Study at the molecular level therefore proves to be difficult in mammals, due to the wide range of complex environmental interactions that influence phenotypic plasticity being almost impossible to single out. This work is vital due to the gaps that exist within our knowledge on this topic, which centre upon three main questions; how do organisms sense this change in their environment? Once sensed, how is this signal transmitted to tissues? Lastly, how do large scale changes in gene expression occur? At this point in time, how the altered phenotypes that we are observing have occurred is a mystery, one that will only be unravelled by not only further research, but research using appropriate organisms and systems that will allow for elucidation of individual environmental effects.

1.2 Invertebrate model- the honeybee

1.2.1 Honeybees as a model organism

Invertebrate models, such as the honeybee, *Apis mellifera*, allow a unique opportunity to study the molecular basis of phenotypic plasticity. The usefulness of this organism has become even greater over the past decade, due to the work of the Honeybee Genome Sequence Consortium in 2006, which provided both great insight and a valuable resource for working with honeybees. Studies of phenotypic plasticity in honeybees have relevance to plasticity in other animals, including ourselves, as plasticity is a feature of all major branches of life, implying conserved mechanisms may underpin it (Nijhout, 2003). The honeybee exhibits two well characterised plastic events in response to limited environmental changes (Thompson et al., 2007). Female honeybees share the same genetic complement, yet develop into two distinct castes (queens or workers) depending on the environmental factor of larval diet (Kucharski et al., 2008; Townsend and Lucas, 1940).

The phenotypes of these castes differ, queen honeybees are larger with different pigmentation, a lack of pollen collecting scopae (Thorp, 2000) and mandibles with sharp cutting edges. Internally, only the queen bee has fully developed ovaries. The queen honeybee is normally the only reproductively active female in the hive, laying up to 2,000 eggs a day (Maleszka, 2008). She possesses large meroistic ovaries (ovaries possessing nurse cells connected to oocytes) with hundreds of ovarioles that consist (from anterior to posterior) of terminal filament cells containing germ cells and clusters of nurse cells alternating with oocytes that mature as they become more posterior (Fig. 1.2) (Dearden, 2006). Worker ovaries show arrested oocyte development and a reduced number of ovarioles (2-10). Both queen and worker ovarioles are contained within an intima.

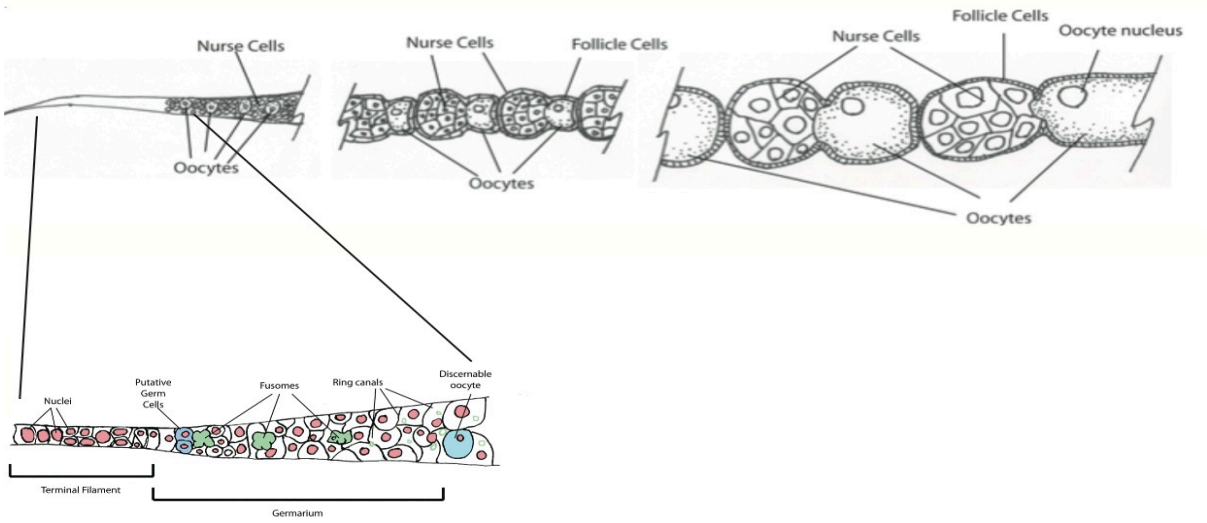


Figure. 1.2. Diagram showing the morphology of an ovariole from a mated queen honeybee ovary. (Modified from Dearden, 2006).

1.2.2 Ovary activation

In the absence of a queen, workers sense the environmental change of reduced levels of Queen Mandibular Pheromone (QMP), and undergo ovary activation (Butler et al., 1959). This can be broken down into three stages of activation, the indicators for which are

deposition of yolk, oocytes and mature elongated oocytes (Fig. 1.3). In this state ovaries become active tissues full of haploid eggs, that if nurtured will become drones (male bees) (Groot and Voogd, 1954). This progression occurs rapidly, within a week in some species (Page and Erickson, 1988). Over this time, over one third of the genome shows altered gene expression (Duncan and Dearden et al., unpublished data). An explanation for the rapidity and scale of this change in gene expression is that there are many genes responsible, and that they are located next to each other on the chromosome. This could allow coordinated gene expression through these genes being controlled as a chromatin domain, allowing a single environmental influence to alter the gene expression of all the genes within this domain.

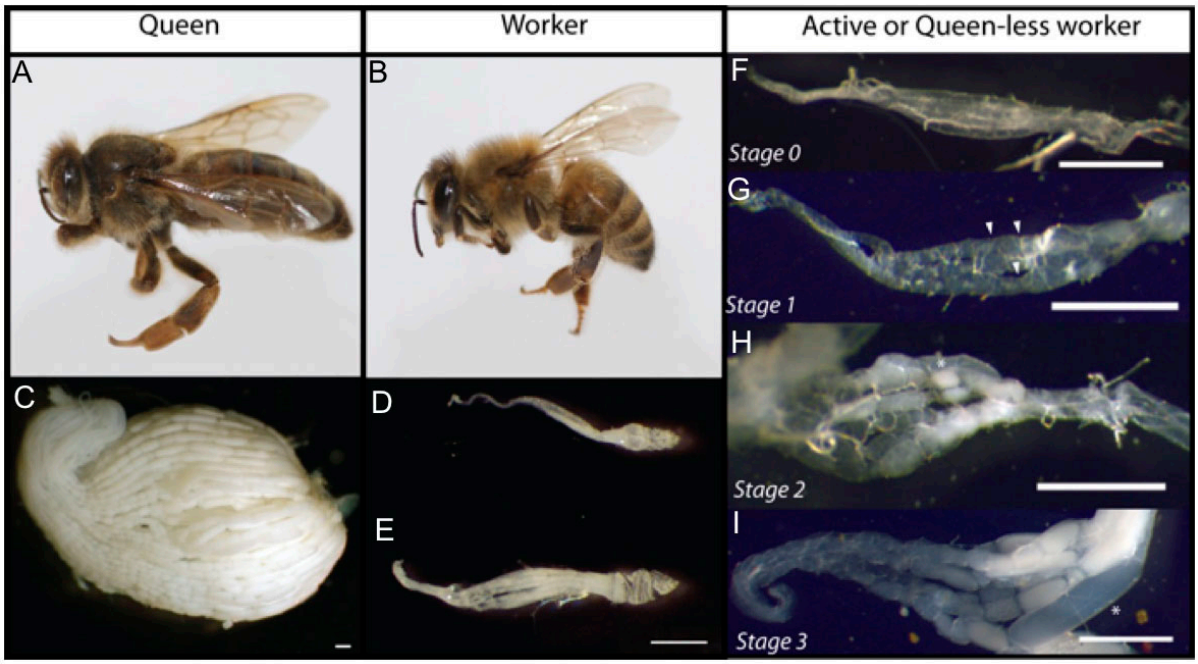


Figure 1.3. Images showing the differences in phenotype between queen (A) and worker (B) honeybees, including size and pigmentation. Images highlighting differences in size, number of ovarioles and stages of activation between queen (C) and worker (D, E) ovaries, note the differences in scale (scale bars = 500 μm). Images showing the three stages of ovary activation in the worker honeybee after the loss of the queen (F-I). Figure supplied by Dr Elizabeth Duncan.

RNA-seq analysis allows for the generation of a quantitative view of the RNA portion of the transcriptome. This has been performed in the honeybee, and has identified one such region, the Lethal (2) Essential for Life cluster (*L(2)efl*), located on chromosome 2, containing 7 genes (Romeril, 2013). This cluster was found to be the product of a gene duplication 200 million years ago in hymenoptera that has been maintained in the honeybee (Romeril, 2013). There are also seven paralogs of the *L(2)efl* gene in the Jewel Wasp *Nasonia vitripennis* and the bumblebee *Bombus impatiens*. However other members of the hymenoptera have only single orthologs of the *L(2)efl*, such as the fire ant species *Solenopsis invicta* and the parasitoid wasp species *Macrocentrus cingulum, Venturia canescens* and *Microctonus hyperodae* (Romeril, 2013). In honeybees these genes are down-regulated (in comparison to housekeeping genes) as the workers sense the loss of the queen through the absence of QMP (Figure 1.4).

1.2.3 Coordinated gene regulation

The possibility of these genes being located next to one another on the chromosome, as well as being co-ordinately regulated is interesting, as there are very few examples of co-ordinately regulated gene regions in invertebrate species. These have historically been found as part of gene complexes that are conserved across species, such as *Wnt, Fox, Runt, Hox* and *Enhancer of Split* (Duncan et al., 2008; Duncan and Dearden, 2010; Krumlauf, 1992; Mazet et al., 2006; Nusse, 2001). The fact that the *L(2)efl* genes have remained clustered together over 200 million years of evolution suggests a vital biochemical process, developmental role or some form of selective advantage, as evolutionary forces are

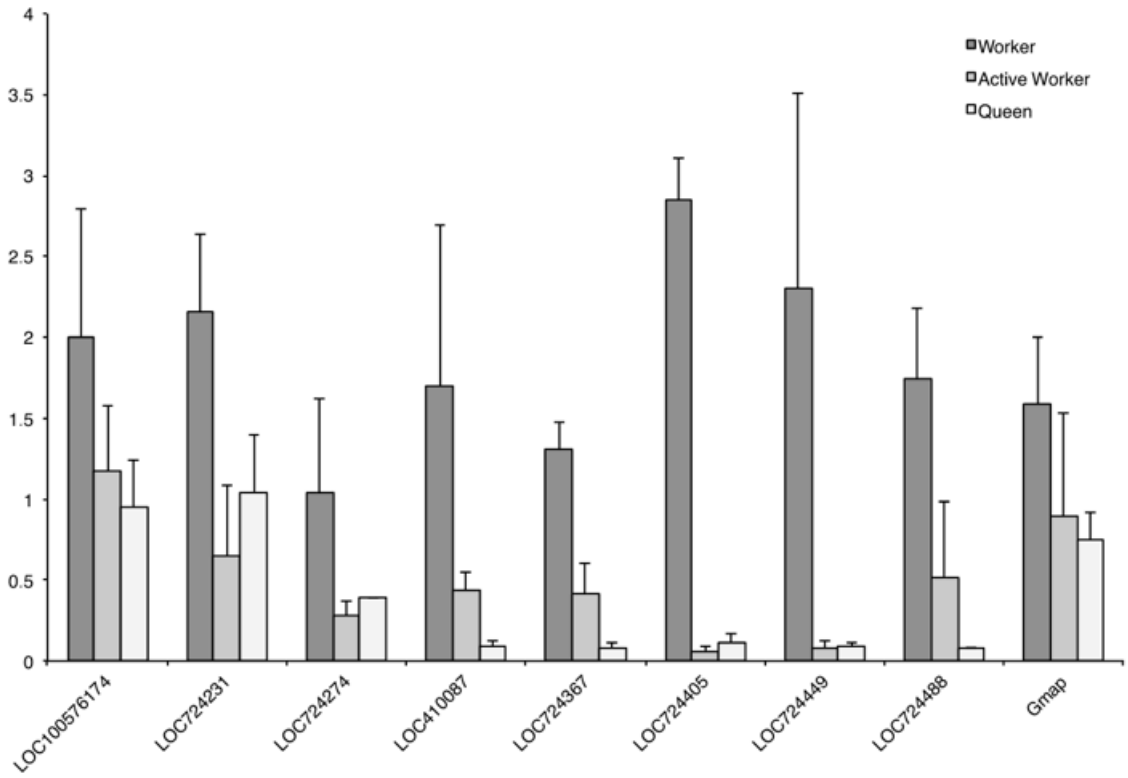


Figure 1.4. Quantitative RT-PCT data for the genes of the *L(2)efl* cluster (*LOC724231-LOC724488*), one gene upstream and one downstream (*LOC100576174* and *Gmap* respectively). *L(2)efl* genes are more highly expressed in workers, and this expression is reduced in active workers (Romeril, 2013).

preventing these genes from being split apart by recombination. Duplicated genes tend not to stay together over evolutionary time in invertebrates, in particular in the honeybee, which has a high recombination rate (Honeybee Genome Sequencing Consortium, 2006). This makes the *L(2)efl* cluster of particular interest, as the conservation suggests that the organisation of these genes has some significance.

The *L(2)efl* cluster can be used to aid our understanding of how the environment regulates large scale changes in gene expression. Particularly in regard to fact that studies in mammalian systems prove difficult in regard to separating out single influences from complex environmental interactions. The honeybee allows for an easily manipulable and

tractable system to study the molecular mechanisms of phenotypic plasticity in response to a single influence, in this case a lack of QMP.

1.3Molecular mechanisms underpinning environmental responsiveness

1.3.1 The agouti mouse

As previously discussed, the molecular mechanisms underlying plasticity are still poorly understood. One model that is well studied is the viable yellow agouti (A^{vy}) mouse. This mouse contains coat colour variation linked to epigenetic marks. Epigenetics allows for stable changes in gene expression without altering the underlying DNA sequence. In this case, the epigenetic modification is DNA methylation, which is the addition of a methyl group to a cytosine residue. This occurs at an IAP (a murine retrotransposon upstream of the *Agouti* gene), which is an epiallele. These are alleles that are identical, but show differential expression depending on epigenetic modifications that occur during early development (Rakyan et al., 2002). While not being an example of phenotypic plasticity, it provides a model system that can be perturbed by the environment. In the wild type mouse the Agouti gene is expressed in hair follicles, producing a brown phenotype (Duhl et al., 1994). The $A^{(vy)}$ IAP, however, is driven by a promoter that induces expression in all cells, leading to a phenotype of yellow fur, adult-onset obesity, diabetes and tumorigenesis (Miltenberger et al., 1997; Morgan et al., 1999). The phenotype does, however, vary greatly due to the level of methylation present in the 5' region of the Long Terminal Repeats (LTR) of the IAP epiallele. Colours produced range from yellow (unmethylated) to pseudoagouti/brown (methylated) (Dolinoy, 2008).

The agouti mouse was utilised in a study investigating the effects of maternal diet on the foetal epigenome. Genistein is a major isoflavone present in soy that has been shown to be

active in biological systems, including oestrogen receptor and non- oestrogen receptor-mediated signalling pathways (Lamartiniere et al., 2002; Tatiana et al., 2004). When the maternal diet was supplemented with genistein, the coat colour of the A^{vy}/a offspring was shifted towards pseudoagouti (brown). Analysis of the IAP element revealed the increased methylation of 6 CpG sites. As this occurred in many different tissue types, it suggested that this occurred early within development (Dolinoy et al., 2006). This change persisted into adulthood, and so protected the offspring from obesity later in life. This study helps to show the effects that early environment, such as maternal diet, can have on phenotypic plasticity, however, it does not answer the question of how diet is influencing an epiallele. On possible explanation is through Hsp90, a heat shock protein. When mutated or functionally impaired it has been proposed to 'promote evolutionary change in otherwise entrenched developmental processes' (Rutherford and Lindquist, 1998). Genistein has been shown to inhibit the function of Hsp90 (Basak et al., 2008), and provides a possible mechanism through which diet affects the function of systems which normally buffer variation.

Some may argue that studies on agouti mice are not ideal, as they are a model system that has a unique genetic feature that may not be directly applicable to other organisms. This is not necessarily the case, as they aid in proof of concept work, and may have greater applicability in the fact that 40% of the human genome is transposable elements, 9% of which are retrotransposons (Lander et al., 2001). A similar principle applies to the *L(2)efl* cluster in the honeybee, as research into the mechanisms underlying this can only aid in our understanding of the regulation of large scale gene expression changes regulated by the environment.

1.3.2 Rodent studies

One example of environmental effects acting postnatally is a study by Meaney, which examined the effect of mothers who did not exhibit correct maternal behaviour, such as licking their pups and normal nursing posture on the brain of the offspring (Francis et al., 1999). This study was the first to show nongenomic transmission of a behavioural state. This was later linked to a difference seen in the level of DNA methylation at the glucocorticoid receptor gene promoter, as well as other associated epigenetic modifications to histone proteins in this region in the pups (Weaver et al., 2004). This effect occurred within the first week of life, and was reversible by altering the early environment. However, if not reversed, the epigenetic effects were shown to persist into adulthood, thus affecting the stress response of the offspring through the hypothalamic-pituitary-adrenal axis (HPA) (Weaver et al., 2004). This exhibits how important the early life stages are in determining the adult phenotype of an organism, not only prenatally, but postnatally also.

In support of this is a study in rodents on restricted uterine blood flow to the offspring. This has been associated with intrauterine growth retardation and adult onset disease, such as hypertension, in both human and animal studies (Bassan et al., 2000; Zidar et al., 1998). Reduced uterine blood flow was shown to alter methylation of the renal *p53* gene, at the promoter and in exons 5-8. The proposed mechanism is altered DNA methylation of *p53* (involved in cell cycle regulation and genomic stability (Ashcroft et al., 1999)) affecting levels of mRNA transcripts of apoptosis-related genes. In turn, this increased renal apoptosis (cell death) and reduces glomeruli number in the kidney, associated with a hypertensive phenotype in adulthood (Pham et al., 2003).

Other examples of evidence for the mechanisms underlying plasticity include candidate gene studies in rodents looking at the effects of maternal dietary restriction during

gestation. This was shown to cause a decrease in methylation status of two genes in the liver; the glucocorticoid receptor (GR) and PPARα (peroxisomal proliferator-activated receptor). As these changes persisted after removal of the environmental influence (weaning), it suggested that these were stable alterations which occurred to the epigenetic regulation of these two transcription factors (Lillycrop et al., 2005). Interestingly, analysis of promoter methylation of another gene, PPARγ, suggested that the effects seen with maternal dietary restriction are gene specific. This raises the possibility that different environmental influences in early development are acting upon different parts of the genome to affect the phenotype of the individual. This implies that the adult phenotype may be the product of a myriad of interactions between genes and environmental cues.

These studies highlight the varied nature of environmental effects that can alter the phenotype of an organism. However, studies such as these focus on candidate genes, and so often only find a small number of genes showing an association. Candidate gene studies are often useful as an efficient way of using time and resources, however they often show only results that are expected. There is a need to look at the whole genome in order to extend our sphere of knowledge, as it is often the unexpected results that are the most significant. This is one of the major lessons that should have been learnt from the Human Genome Project, where the one gene one phenotype hypothesis lead many to believe sequencing alone would provide answers to disease risk. However, it is often not a single gene, or even several genes, but many interacting with varied levels of effect. Therefore by similar logic, the genes affected by a single environmental influence associated with phenotypic plasticity are unlikely to be singular either. Overall, despite the insights research such as this gives, it does show the importance of needing to understand how large scale gene expression is regulated.

1.4 Epigenetics

1.4.1 DNA methylation

Epigenetic modifications allow for long term changes in gene expression without altering the underlying gene sequence. This involves a variety of mechanisms, of which DNA methylation is one of the best understood. It occurs by the addition of a methyl group to the number 5 carbon of a cytosine at a CpG site, forming a 5-methyl-cytosine. *De novo* methylation is carried out in mammals by the DNA methyltransferases (Dnmts) 3a and 3b (Okano et al., 1999) and maintained across cell divisions by Dnmt1 (Bestor, 2000). The human genome can have up to 90% of its CpG dinucleotides methylated, typically spanning the promoter regions of genes and is associated with transcriptional repression (Jaenisch and Bird, 2003; Wolffe and Matzke, 1999). In humans, DNA methylation is involved in a variety of processes, including parental imprinting (Dittrich et al., 1996), cellular differentiation (Shiota, 2004), X-inactivation (Mohandas et al., 1981) and silencing repetitive elements. It has also been observed in the promoter regions of cancer cells in culture (Esteller and Herman, 2002), which implies a vital role in our development, both normal and abnormal.

Most genomic methylation patterns are stable across tissues and throughout life, with a few exceptions being the rapid demethylation of the paternal genome after fertilisation, followed by remethylation in development in mammalian systems (Esteller and Herman, 2002; Oswald et al., 2000), or the reprogramming of the maternal genome to match the paternal in zebrafish (Jiang et al., 2013; Potok et al., 2013). Demethylation is carried out by Ten-eleven translocation (TET) family of enzymes, which convert 5-methylcytosine into 5-hydroxymethylcytosine, shown to be vital in early development in mice (Gu et al., 2011). This process of regulating methylation patterns is vital, as many pluripotency genes are silenced by hypermethylation upon differentiation, but previous to this their promoters

must be hypomethylated to allow for gene expression (Farthing et al., 2008; Hattori et al., 2004; Imamura et al., 2006). Therefore proper regulation of DNA methylation is essential for early development.

DNA methylation is present in honeybees (Wang et al., 2006), with similar mechanisms to those in mammals (Holliday and Pugh, 1975; Riggs, 1975), the majority of which is localised to gene bodies (Moczek and Snell-Rood, 2008). Methylation of gene bodies is observed in a wide range of eukaryotes, and is present in insects (except for *Drosophila* and *Tribolium* which do not possess a functional DNA methylation system), plants and mammals (Feng et al., 2010; Sarda et al., 2012; Zemach et al., 2010). Promoter methylation appears to have evolved in the vertebrate lineage, whereas gene body and transposon methylation is present in the last common ancestor of plants and animals (Feng et al., 2010; Zemach et al., 2010). There is evidence for gene body methylation having various functions such as repressing intragenic promoter activity (Maunakea et al., 2010), alternative splicing (Falckenhayn et al., 2010; Foret et al., 2012; Li-Byarlay et al., 2013; Lyko and Maleszka, 2011; Sati et al., 2012; Shukla et al., 2011), controlling transcriptional elongation (Lorincz et al., 2004) and ensuring that the first and last exons are included in the transcript (Sati et al., 2012). As gene body methylation is a feature of ancient eukaryotic genomes it allows for the use of model organisms, such as honeybees, in investigating the role of DNA methylation.

DNA methylation can be investigated using bisulphite sequencing. It is a technique that uses bisulphite to convert unmethylated cytosines into uracil, but leaves methylated cytosines unchanged. Site specific differences can be seen by comparing treated and untreated sequences. The same principle applied to TAB-seq (Tet Assisted Bisulphite Sequencing), used to investigate demethylation through 5-hydroxymethylcytosine (5hmC), which utilises βglucosyltransferase to protect 5hmC, while Tet1 is used to oxidise 5mC,

which is converted into thymine by bisulphite (Yu et al., 2012). Another method that is widely used is Methylation-Specific High Resolution Melting (HRM), which detects differences in levels of methylation along a stretch of DNA (Wojdacz and Dobrovic, 2007). This technique also utilises bisulphite conversion. Differences in methylation status generate a difference in nucleotide composition in the PCR fragments. When these fragments are melted they have slightly different properties depending on how many cytosines are methylated (Wojdacz and Dobrovic, 2007). This technique has been shown to be capable of detecting one hemimethylated residue in a 150 base-pair fragment (Duncan and Dearden, unpublished data), and so is both sensitive and informative.

1.4.2 Histone modifications

Histone modifications are another vital epigenetic mechanism. Eukaryote genomes are packaged into chromatin consisting of nucleosomes, which are a combination of 146 base pairs of DNA wrapped around a histone protein octamer (Luger et al., 1997). Nucleosomes are essential to the organisation of genetic material, with the positioning of nucleosomes and modifications to the histones having a great effect on the expression of the associated DNA (Narlikar et al., 2002). These histones are subject to a wide range of post translational modifications, most often at the N-terminal tails, which induce active or repressed sequence depending on the modification (Nightingale et al., 2006). There are many examples of links between specific modifications to histones (or combinations thereof) and altered gene expression. Transcriptionally silent chromatin is often associated with overall histone under-acetylation and di- or tri- methylation of H3K9 (histone 3 lysine 9), whereas active promoter regions are often associated with hyperacetylation and increased tri-methylation of H3K4 (Litt et al., 2001; Mutskov and Felsenfeld, 2003; Pokholok et al., 2005; Schübeler et al., 2004). It is not only the marks themselves that are important, but the context of these also, as the locations at which these modifications

occur, and the combinations of modifications, give an indication of the regulation of expression at these sites.

Despite the findings to date, there is still much that remains unknown in regard to the role histone modifications play in regulating gene expression. The main tool for investigating this is chromatin immunoprecipitation. This technique for investigating protein-DNA binding. Using antibodies that are specific for particular proteins (such as specific histone modifications), it allows for enrichment and isolation of the DNA sequences that these proteins are bound to (O'Neill and Turner, 1996). Hence it is possible to ascertain whether a region of DNA is likely to be active or repressed. On a whole-genome scale, however, specific regions must be validated. This technique was successfully utilised in a study looking a whole genome enrichment of the repressive histone mark H3K27me3 in the ovaries of honeybees that were either queens, workers or active workers. It highlighted that after the loss of the environmental influence of QMP inducing the phenotypically plastic response of ovary activation, differences could be seen in the enrichment of H3K27me3 (Leask, 2014). The pattern of enrichment for active workers was shown to become an intermediate between workers and queens in some regions of the genome, and queen-like in others. This illustrates the role that histones play in phenotypic plasticity in response to an environmental influence.

1.4.3 Insulator elements

CTCF insulator elements play a crucial role in regulating gene expression. Insulators function as barriers to the influences of neighbouring *cis*-acting elements, for example, to prevent gene activation when located between an enhancer and promoter (Bell et al., 1999). Differentially expressed genes can be neighbours at distances over which enhancers may act, yet are independently regulated. There must, therefore, be mechanisms which

prevent the inappropriate activation of genes, one of which is insulation. CTCF often flanks regions of the genome which require precise regulation, and it has been shown to be methylation-sensitive, suggesting an interaction between epigenetic mechanisms (Bell and Felsenfeld, 2000). It is possible CTCF is playing a role in phenotypic plasticity during ovary activation, as CTCF is highly conserved between divergent species (Filippova et al., 1996).

1.4.4 Transcription factors

It is well established that transcription factors play a major role in regulating gene expression. They act by binding to specific sequences of DNA to cause activation, or to block transcription machinery. It is highly likely that these are acting during ovary activation to regulate gene expression. It is probable that when the *Lethal(2)efl* gene was duplicated in evolutionary history that the *cis*-regulatory elements were duplicated also. Assuming that evolution has not acted on these regions, there is the possibility that one, or several, transcription factors are acting to co-ordinately regulate the genes of the ancestral *L(2)efl* cluster. Potential binding sites for both CTCF and transcription factors can be found using consensus sequences and bioinformatics techniques. Any putative binding sites, however, must be validated with other techniques, such as chromatin immunoprecipitation.

1.4.5 Limitations of epigenetic studies

Regardless of the knowledge gained to date about the role of epigenetic mechanisms, such as DNA methylation and histone modifications, substantial gaps in our understanding still remain. One of the largest issues with epigenetics is that at this point it is almost impossible to separate out cause and effect. It is currently unknown whether the epigenetic modifications we observe to be associated with a phenotype are causative, or whether they

are simply the product of another mechanism, and therefore are correlative only. In order for this to be ascertained there needs to be a greater level of specificity in the studies which occur. At this stage investigations are limited to large scale manipulation of epigenetic profiles, such as using inhibitors of DNA methylation, which are nonspecific. This does not allow for interrogation of a specific site, such as a single CpG residue, or histone modification. There is also a great limitation due to cell type specificity issues, as the techniques used are often not capable of analysis on a single cell, meaning results are often the product of an 'average' from a pooled sample of different cell types. This greatly reduces the power of a study, and hinders the understanding of the role single modifications are playing in phenotypic plasticity. Ideally, interrogation of a single site utilising an environmental influence would allow for understanding of the role it plays on specific cell types in a phenotypically plastic response. At the root of the issue is epigenetics, and finding a way to answer the question of how the environment and DNA are interacting.

1.5 Aims, experimental design and significance

The aims of this research were to investigate how environmental influences result in large scale changes in gene expression associated with phenotypic plasticity, using the *L(2)efl* cluster in *A. mellifera* as a model. This was carried using *in situ* hybridisation determine which cell types of the *L(2)efl* genes are expressed in. Coordinated regulation would suggest that these genes are part of a chromatin domain, and therefore environmental cues alter the expression of all the genes within this domain. The environmental influence in this case is the absence of QMP following the removal of the queen bee, leading to the phenotypically plastic event of ovary activation.

Investigations were also carried out into which epigenetic mechanisms may underpin this regulation. If coordinated regulation was observed, then it was possible that the histone modifications were playing a role, and that different patterns of enrichment would be seen when comparing workers and active workers. The alternate hypothesis was that when these genes were duplicated in evolutionary history, their *cis*-regulatory regions were duplicated also, and therefore bioinformatics was used to look for conserved potential binding sites for transcription factors that could be co-ordinately regulating gene expression. Bioinformatics was also to be used to look for binding sites for CTCF insulator elements which may be regulating gene expression. These investigations of the *L(2)efl* cluster regulation aids to better our understanding of the epigenetic mechanisms which may regulate large scale changes in gene expression in response to environmental influence (in this case the loss of the single influence of QMP) making *A. mellifera* a valuable model system for understanding their role in phenotypic plasticity.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 Primers:

LOC411452 forward – GACACACCGGAAGTGTGATG	IDT
LOC411452 reverse – GGCGTCAAACCCAGAAAATA	IDT
LOC100576174 forward – TGACAACTGTTGCTGTAACTTCG	Invitrogen
LOC100576174 reverse – TGTTTTATATTGCAAATTGTATTCCA	Invitrogen
LOC274231 (Lethal 1) forward – ATTGAATGTTCGCGCTTCTT	Invitrogen
LOC724231 (Lethal 1) reverse – TCTCTTTTCTTGCTGCAGTGA	Invitrogen
LOC410087 (Lethal 3) forward – CAGAGATCGAGTGACGAGTGC	Invitrogen
LOC410087 (Lethal 3) reverse – CCCCGTTTGTTCGATTTTTA	Invitrogen
LOC724405 (Lethal 5) forward – TACAGCAGTCCGTGTGGAAC	Invitrogen
LOC724405 (Lethal 5) reverse – AGATTCGCGCAACGAATAAT	Invitrogen
LOC724488 (Lethal 7) forward – CACTCGAGAGAGAAGCTTA	Invitrogen
LOC724488 (Lethal 7) reverse – TTCCTTCAATGCCGGTTTAC	Invitrogen
Gmap forward – GAACTCGATCAGCAACACGA	Invitrogen
Gmap reverse – ACCTGTCCAACGACATCCTC	Invitrogen
LOC412824 forward – TATTTGACACGCGGACGTAA	IDT
LOC412824 reverse – CTCGTCGCATTGCTATTTCA	IDT
M13 forward – GTTTTCCCAGTCACGAC	Invitrogen
M13 reverse – CAGGAAACAGCTATGAC	Invitrogen

IDT = Integrated DNA Technologies.

2.1.2 ChIP qPCR primers:

LOC100576174 forward – CGTGACCAACTAGGCAACAA IDT

LOC100576174 reverse – CGAAGTTACAGCAACAGTTGTCA IDT

LOC410087 (Lethal 3) forward – CGTCTAAAGACCCCCTCATTG IDT

LOC410087 (Lethal 3) reverse – CGCACTCGTCACTCGATCT IDT

LOC724405 (Lethal 5) forward – ATTGCGCCTGCTTGTAAAAT IDT

LOC724405 (Lethal 5) reverse – GTTCCACACGGACTGCTGTA IDT

LOC724488 (Lethal 7) forward – CTATCGGTAGCCGGTGCTAT IDT

LOC724488 (Lethal 7) reverse – TCAAGCTTCTCTCTCGAGTGTTT IDT

Gmap forward – AAATGGGAAAAGTCCCATGTAATC IDT

Gmap reverse – TAACAGGAACCACGCACGTA IDT

2.1.3 qPCR primers

BR-C forward – CGGACTTGGCAGGAACTT IDT

BR-C reverse – TCGCAGGTATAGCACACAACC IDT

E75 forward – CGAGAAACGCCTAACTATACCG IDT

E75 reverse – CGAGAACCTCTTCGAGAAATCTT IDT

2.1.4 Table 2.1 Luria Broth (LB) and Luria Broth Agar

Luria Broth (LB)	1% Bacto tryptone (w/v), 0.5% Bacto yeast extract (w/v) and 1% NaCl (w/v) diluted in dH_2O (pH7) and autoclaved. Addition of ampicillin

	(50 ng/mL)
Luria Broth Agar	Luria Broth, 1.2% Bacto agar (w/v) autoclaved. Addition of ampicillin (50 µg/mL)

2.1.5 DNA amplification and cloning

10x conc. PCR buffer	Invitrogen
10x conc. $MgCl_2$	Invitrogen
5mM dNTPs	Roche
Taq DNA Polymerase	Invitrogen
XLI Blue cells	Propagated in the lab
pBluescript II KS (+) vector	Stratagene
2x conc. Rapid Ligation Buffer	Roche
T4 DNA Ligase	Roche
QIAquick Gel Extraction Kit	Qiagen
High Pure Plasmid Isolation Kit	Roche

2.1.6 Restriction enzymes (and buffers)

BamHI (buffer B)	Roche
HindIII (buffer B)	Roche
Kpn1 (Buffer L and BSA)	Roche

2.1.7 *Run-off in vitro transcription*

Transcription Buffer 10x conc.	Roche
DIG RNA labelling mix 10x conc.	Roche
RNaseOUT Ribonuclease Inhibitor	Invitrogen
T7 RNA Polymerase	Roche
T3 RNA Polymerase	Roche
DNase I Amplification Grade	Invitrogen

2.1.8 RNA in situ Reagents Table 2.2

Hybridisation Buffer	50% deionised formamide, 4x SSC, 1x denhardts, 250 µg/mL tRNA, 250 µg/mL ssDNA, 50 µg/mL heparin, 0.1% Tween-20, 5% dextran sulphate
Wash Buffer	50% formamide, 2x SSC, 0.1% Tween-20
PTw	PBS, 0.1% Tween-20
Carbonate Buffer	120 mM Na2CO3, 80 mM NaHCO3, pH 10.2
PBTw	PBS, 0.1% Tween-20, 0.1% Bovine Serum Albumin
Alkaline Phosphotase Buffer	100 mM Tris pH 9.5, 100 mM NaCl, 50 mM MgCl2, 0.1% Tween-20
Anti-dioxygenin-alkaline phosphatase (AP) antibody	Roche

NBT	75 mg/mL nitro blue tetrazolium chloride in dimethylformamide (Roche)
BCIP	50 mg/mL 5-Bromo-4-chromo-3-indolyl phosphate disodium salt in dimethylformamide (Roche)
Phosphate Buffered Saline (PBS) 10x conc.	21 mM NaH2PO4, 84 mM Na2HPO4, 1.7 M NaCl, pH 7.4

2.1.9 Miscellaneous

Bovine Serum Albumin (Invitrogen)

4', 6-diamidino-2-phenyl-indole, dihydrochloride (DAPI) (Invitrogen)

70% ultrapure Glycerol (Sigma)

Tween-20 (Sigma)

Formaldehyde (37%) (Sigma)

Proteinase K 20 µg/mL (Roche)

2.1.10. Chromatin immunopreciptitation

Glycine	Invitrogen
Lysis Buffer	Invitrogen
Protease Inhibitors	Invitrogen
Reverse Crosslinking Buffer	Invitrogen
Proteinase K	Invitrogen
Elution Buffer	Invitrogen

2.2 Equipment

2.2.1 RNA in situ hybridisation

In situ hybridisation microscopy was carried out using the Olympus BX61 compound microscope with an Optronics DP71 camera. *A. mellifera* dissection and monitoring of RNA *in situ* hybridisation was achieved using the Lecia MS5 dissection microscope. Nucleic acid levels were quantified on a Thermo Scientific NanoDrop 2000 spectrophotometer. Agarose gels were photographed using a Biorad GelDoc and the Quantity One program.

2.2.2 Chromatin immunoprecipitation

Tissue was homogenised using a 22 gauge hypodermic needle and syringe and a dounce homogeniser with B pestle. Chromatin was sheared using Covaris Adaptive Focused Acoustic Technology, with the sample contained in a Covaris MicroTube with AFA fibre.

2.2.3 Sonication Conditions Table 2.3

Duty Factor	5%
Cycles per Burst	70 (Intensity 2/CBP 200)
Cycle time	60 Seconds
Number of Cycles	12
Temperature of Water Bath	4 Degrees
Frequency of Sweep	Sweeping
Degassing Mode	Continuous

2.3 Polymerase Chain Reaction Protocol

PCR was carried out in a BioRad C1000 Thermal Cycler. Unless otherwise stated the PCR protocol was as I summarised in Table 2.4

Step	Temperature	Duration
1	94°C	3 minutes
2 (Denaturing)	94°C	30 seconds
3 (Annealing) 35 cycles	55°C	30 seconds
4 (Extension)	72°C	1 minute
5	72°C	5 minutes
6	4°C	Hold

2.4.Molecular Cloning of Lethal(2)efl genes 1, 3, 5 and 7, LOC100576174, Gmap, LOC411452 and LOC412824

2.4.1 PCR

PCR was performed on *A. mellifera* cDNA using *Lethal(2)efl 1, 3, 5* and *7, LOC100576174, Gmap, LOC411452* and *LOC412824* primers. The PCR reaction for each pair of primers contained 1x PCR buffer, 20 pmol of each forward and reverse primer, 50 ng of cDNA, 0.25 mM of dNTPs, 1 unit of *Taq* polymerase, and made up to the final volume of 20 µl with dH_20. A template free negative control was also set up for each reaction. The PCR products were combined with 1x electrophoresis loading dye (25% bromophenol blue, 25% xylene cyanol, 30% glycerol) and run on a 1% agaorse gel in 1x Sodium Borate (SB) buffer (20x SB buffer: 0.8% NaOH (w/v), 4.5% boric acid (w/v)) alongside an Invitrogen 1KB+ electrophoresis size standard, with imaging and photography taking place under a UV Biorad Geldoc.

2.4.2 PCR Purification

PCR products were purified using the Qiagen QIAquick Gel Extraction Kit according to the manufacturer's instructions.

2.4.3 Ligation Reaction

To ligate the genes in the *lethal(2)efl* cluster of interest, as well as the surrounding genes of interest into plasmids 40 µl of PCR product from the above PCR reactions was combined with 4 µl pBluescript II KS (+) vector, 1x ligation buffer and 1 unit T4 DNA ligase. This mix was then incubated overnight at 16°C.

2.4.4 *Transformation*

Previously prepared competent *E.coli* XLI- Blue cells were thawed on ice and then 10 µl of the ligation reaction was added. The bacteria mix was incubated on ice for 20 min, mixing every 2-3 min. This was then heat shocked at 37°C for 5 min, followed by a cold shock on ice for 1 min. Once cold shocked 0.9 mL of LB was added to the bacteria mix and this was incubated at 37°C with shaking for one hour. After the incubation period, cells were spun down, supernatant removed and 100 µl of the bacterial culture was plated onto the LB and ampicillin plate, and incubated overnight at 37°C.

2.4.5 Colony PCR

Single bacterial colonies (eight for each plate) were picked and each used to for patch plating onto a new LB and ampicillin plate (for use in later overnight cultures) and for use in colony PCR. These were added to 8 well strips that contained 1x PCR buffer, 0.25 mM dNTPs, M13 forward and reverse primers (20 pmol), 1 unit Taq polymerase and dH_2O to make up the final volume of 20 µl. The PCR products were combined with 1x electrophoresis loading dye and run on a 1% agarose gel, then photographed under UV

light. Colonies positive for the insert were used to inoculate overnight cultures consisting of 3 mL LB and amp (50 μg/mL) antibiotic in sterile universals. Plasmid DNA was isolated using Roche Applied Biosystems High Pure Plasmid Isolation Kit.

2.5 Sequencing

For each DNA sample two reactions were sent for sequencing. In one reaction 100-150 ng/μl of sample DNA was combined with M13 forward primer (3 pmol/μl) and made up to 5μl total volume with dH_2O. In the second reaction 100-150 ng/μl of sample DNA was combined with M13 reverse primer (3 pmol/μl) and made up to 5 μl total volume with dH_2O. The samples were sequenced using the ABI 3730xl DNA Analyser by Genetic Analysis Services in the Anatomy Department, Otago University.

2.6 Probe Preparation for RNA in situ Hybridisation

Table 2.5 Restriction enzymes, buffer and RNA polymerases used for each RNA in situ hybridisation probe.

Probe	Restriction Enzyme and Buffer	RNA Polymerase
LOC411452 sense	BamHI, buffer B	T3
LOC411452 antisense	HindIII, buffer B	T7
LOC100576174 sense	KpnI, buffer L	T7
LOC100576174 antisense	BamHI, buffer B	T3
LOC724231 (Lethal 1) sense	BamHI, buffer B	T3
LOC724231 (Lethal 1) antisense	KpnI, buffer L	T7
LOC410087 (Lethal 3) sense	HindIII, buffer B	T7

LOC410087 (Lethal 3) antisense	BamHI, buffer B	T3
LOC274405 (Lethal 5) sense	HindIII, buffer B	T7
LOC274405 (Lethal 5) antisense	BamHI, buffer B	T3
LOC724488 (Lethal 7) sense	BamHI, buffer B	T7
LOC724488 (Lethal 7) antisense	KpnI, buffer L	T3
Gmap sense	HindIII, buffer B	T7
Gmap antisense	BamHI, buffer B	T3
LOC412824 sense	KpnI, buffer L	T7
LOC412824 antisense	BamHI, buffer B	T3

2.6.1 Restriction enzyme digests

Approximately 1 µg of pBluescript II KS (+) containing the gene of interest was digested with the appropriate restriction enzyme and buffer (Table 2.5) in the following mix: 1 µg of plasmid, 1x restriction enzyme buffer, 20 units appropriate enzyme and made up to 50 µl with dH_2O then incubated at 37°C for 2 hours.

2.6.2 Ethanol Precipitation

The restriction digest was run on a 1% agarose gel, and the products visualised under UV light in a Biorad Geldoc. Fully digested samples were then made up to 200 µl with dH_2O. An equal volume of phenol chloroform was added and then centrifuged for 5 min at max speed on a bench top microcentrifuge. The top aqueous layer was removed and combined with an equal volume of chloroform. This mix was centrifuged at maximum speed for 5 min. The top aqueous layer was removed once again and 1/10 of the volume of sodium acetate (3 M, pH 5.2) was added, followed by 3x the volume of ice-cold ethanol. This was transferred to -20°C overnight. This mix was then centrifuged at max speed for 15 min in a bench top microcentrifuge. The supernatant was removed and the pellet washed in 70%

ethanol. The pellet was then air dried and then resuspended in 30 µl dH_2O. The amount of DNA present was quantified using a Nanodrop spectrophotometer.

2.6.3 Run-off in vitro transcription

The remaining 28 µl of DNA was combined with 1x Transcription Buffer, 1x digoxygenin (DIG) RNA labelling buffer, 4 units RNase OUT and 2 units of appropriate RNA polymerase (Table 2.5). This mix was incubated at 37°C for 4 hours. After incubation 2 units of DNase was added and the mix was returned to 37°C for a further 15 min. To this mix 1/10 volume of sodium acetate (3 M, pH 5.2) was added and mixed, followed by 100 µl ice-cold 100% ethanol. This was incubated at -20°C overnight. This mix was then centrifuged in a bench top microcentrifuge at max speed for 30 min in a 4°C cold room. The supernatant was removed and the pellet washed in 70% ethanol then centrifuged at max speed again for 5 min. This step was repeated once. The supernatant was removed and the pellet air dried and resuspended in 10 µl DEPC H_2O. The RNA sample was quantified using a Nanodrop spectrophotometer, and the quality of the probe was assessed via agarose gel electrophoresis. The remaining probe sample was combined with 40 µl of hybridisation buffer and was stored at -20°C.

2.7 Honeybee Maintenance

Honeybees used in this research were a New Zealand variety of *Apis mellifera,* originally from Italy, and these were maintained in Langstroth hives. Activation of ovaries was carried out by removing the queen from the hive, then waiting 2 weeks minimum before dissecting out the ovaries from ~50 active worker honeybees in order to determine the level of ovary activation.

2.8 Ovary collection and fixation

Honeybees from Langstroth hives were collected into containers, and then stored at -20°C until the honeybees became motionless. Honeybees were dissected in 1x PBS (18.6 mM NaH_2PO_4, 84.1 mM $NaHPO_4$ and 175 mM NaCl, pH 7.4). The honeybees were dissected from the anterior end of the abdomen using tweezers to remove the digestive system, and then remove the ovaries. The ovaries were transferred into PBS on ice. Dissecting time did not exceed 30 min. After dissecting ovaries were fixed for either 15 min (active worker ovaries) or 5 min (worker ovaries) by rocking in 400 µl PBS, 100 µl formaldehyde and 500 µl heptan. After fixing, the bottom layer of solution was removed and replaced with 1 mL of ice-cold methanol. The ovaries were then washed a further 4x in methanol and stored at -20°C in methanol.

Table 2.6 Probes used in ovary RNA in situ hybridisation

Gene	Probe	Treatment	Volume (µl)
LOC411452	Sense	Undigested	0.50
	Antisense	Undigested	0.50
LOC100576174	Sense	Undigested	0.50
	Antisense	Undigested	1.00
LOC724231	Sense	Undigested	0.50
	Antisense	Undigested	2.00
LOC410087	Sense	Undigested	0.25
	Antisense	Undigested	1.00
LOC724405	Sense	Undigested	1.00
	Antisense	Undigested	0.50
LOC724488	Sense	Digested	0.50
	Antisense	Digested	0.50
Gmap	Sense	Digested	0.25
	Antisense	Digested	0.25
LOC412824	Sense	Undigested	0.50
	Antisense	Undigested	0.50

2.9 Honeybee ovary RNA in situ hybridisation

Fixed ovaries in methanol were rehydrated through a methanol/PBS and 0.1% Tween-20 (PTw) series (3:1, 1:1, 1:3), and washed 3x with PTw. Ovaries were then transferred into a petri dish containing PTw and separated into individual ovarioles using tweezers, and transferred back into an eppendorf containing 1 mL PTw. Proteinase K (2 µl of 20mg/ml stock) was added to the PTw containing the ovarioles to increase probe permeability, and incubated at room temperature for 15 min (queen ovaries), 8 min (active worker ovaries) or 5 min (worker ovaries). The ovaries were washed in PTw, then re-fixed in 4% formaldehyde solution by rocking at room temperature for 15 min. The fixative was removed, and the tissue washed 6x with PTw, before a prehybridisation step for 2 hours in 1 mL of hybridisation solution incubated at 52°C. At the end of the incubation period the probes for genes *Lethal(2)efl 7* and *Gmap* were digested with an equal volume of carbonate buffer for 10 min at 60°C. The hybridisation solution was removed from around the tissue and replaced with probe solution, then incubated overnight at 52°C.

The ovaries were washed in wash buffer 6x at increasing time intervals from 5 min to one hour at 52°C. The last wash was incubated overnight at 52°C. The sample was washed of residual wash buffer with PTw. A blocking step was performed incubating the ovaries in PBTw for 30 min. This solution was replaced with PBTw containing a 1:1000 dilution of Anti-digoxygenin-alkaline phosphatase (AP) antibody and incubated at room temperature for 90 min. The tissue was washed in PTw and following this in AP buffer. The AP buffer was replaced with 1mL AP buffer containing 450 µg NBT and 175 µg BCIP. The ovaries were then transferred to a staining dish, and kept in the dark while staining was monitored every 5-10 min. After staining ovaries were de-stained and washed in MeOH, followed by

PTw. Tissue was stored at -20°C in 70% glycerol with 1 µl DAPI overnight. Ovaries were mounted on microscope slides for imaging.

2.10 Locating potential CTCF binding sites

CTCF binding sites were located using the Gene Palette program. The *Lethal(2)efl* cluster and surrounding genes were loaded into the library as a sequence, and the consensus sequence for CTCF (CNNNAGNNGGCGC) (Van Bortle et al., 2012) from *Drosophila* was added as a feature, not allowing for mismatches.

2.11. Locating Potential Transcription Factor binding sites

There are very few transcription factor binding sites known for the honeybee, therefore sequences used for binding sites were from known *Drosophila melanogaster* sequences. Potential binding sites for transcription factors were located using gene sequences from NCBI Gene (http://www.ncbi.nlm.nih.gov/gene) put into Match-1.0 Public (2005), which uses positional weight matrices from TRANSFAC Public 6.0. Settings were altered to minimise false positives, and to use invertebrate sequences. Potential transcription factor binding sites were located in the 500 bp upstream of the transcriptional start site.

2.12.Primer design for ChIP qPCR for genes LOC100576174, Lethal(2)efl 3, 5 and 7, and Gmap

Primers were designed for ChIP qPCR (Quantitative Polymerase Chain Reaction) by using NCBI to find the genes *LOC100576174, Lethal(2)efl 3, 5* and *7*. The transcriptional start site (TSS) of the gene was found, and 500 bp of sequence upstream and downstream used to get a 1kb fragment. This was pasted into Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi), and the target site adjusted to the 500bp point of the 1kb fragment, in order to design primers around the TSS. Optimum product size was set to 80-120bp, and primer GC% at 40-60%. Primers were designed, and then pasted into Free Beacon Designer (http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1), set to SYBR Green, and ΔG values for cross dimer and self dimer were checked to ensure they were between -3.0 and 0. The energy required for a secondary structure to form is ΔG, therefore the more negative the value, the more likely it is to form spontaneously as less energy is required. The arbitrary cut-off value is -3.0, but ideally values are closer to zero, as positive values require energy being put into the system for secondary structures to form. Lastly, they were put into primerBLAST on NCBI, with the species set to *Apis mellifera* to check specificity.

2.13.Primer design for qPCR primers for the transcription factors BR-C and E75

Spidey (http://www.ncbi.nlm.nih.gov/spidey/) was used to align gDNA and cDNA in order to find exon boundaries. The cDNA was then pasted into Primer3Plus and the parameters changed to design primers around exon boundaries, produce an optimum product size of

80-120bp, and CG% of 40-60%. Primer sequences were put into Free Beacon Designer, with SYBR green as the assay method. Checked that the self dimer and cross dimer ΔG values were between -3.0 and 0. Primer were put into primerBLAST on NCBI, with the species set to *A. mellifera* to test specificity.

2.14. PCR to test specificity of ChIP primers for LOC100576174, Lethal(2)efl 3, 5 and 7, Gmap, and qPCR primers for BR-C and E75

PCR was performed on *A.mellifera* gDNA using ChIP primers for LOC100576174, *Lethal (2)efl 3, 5* and *7*, *Gmap*, and cDNA for qPCR primers for BR-C and E75 (cDNA and gDNA provided by Elizabeth Duncan). The PCR reaction for each pair of primers contained 1xPCR buffer, 20 pmol of each forward and reverse primer, 50 ng of cDNA (for qPCR primers- one queen sample, one active worker sample) or 100 ng gDNA (for ChIP primers and qPCR primers), 0.25 mM dNTPs, 1 unit of *Taq* polymerase, and made up to the final volume of 20 µl with dH_20. A template free negative control was set up for each reaction. qPCR primers also had a –RT reaction set up for each type of cDNA. The PCR products were combined with 1x electrophoresis loading dye (25% bromophenol blue, 25% xylene cyanol, 30% glycerol) and run on a 1% agaorse gel in 1x Sodium Borate (SB) buffer (20x SB buffer: 0.8% NaOH (w/v), 4.5% boric acid (w/v)) alongside an Invitrogen 1kb+ electrophoresis size standard, with imaging and photography taking place under UV light.

2.15. Chromatin immunoprecipitation

2.15.1 Honeybee ovary collection

Ovary tissue was collected by dissection and transferred to PBS on ice, 250 µl of PBS was added for every 50 µg of tissue. Ovary collection time did not exceed 1.5 hours.

2.15.2 Chromatin extraction

Honeybee ovary tissue was homogenised in PBS using a 22 gauge hypodermic needle and syringe, then transferred to a dounce homogeniser and B pestle for 20 strokes. 1% formaldehyde was added and cells were fixed for 10 min. Glycine (0.25 mM) was added after the incubation to stop the fixing process. The cells were collected by centrifugation at 4°C at 4000 rpm, the supernatant removed and cells washed by resuspension in ice cold PBS. The centrifugation and wash step was repeated twice more. Centrifugation followed this, and the supernatant was replaced by 150 µl lysis buffer and protease inhibitors per 50 mg of tissue. The tissue was incubated on ice, then transferred to -80°C for storage.

2.15.3 Sonication

The chromatin was sheared using the Covaris AFA (Adaptive Focused Acoustic) technology. This utilises isothermal and controllable shearing, providing focused burst of ultrasonic acoustic energy at a frequency 15-30 times greater than conventional probe sonicators. The high frequency creates a wavelength of a few millimetres in length, which allows for the energy to be focused at the sample n a vessel immersed in a temperature controlled water bath.

The water bath was filled to level 15, the instrument degassed and cooled to 4°C. The Covaris MicroTube with AFA fibre was loaded with 100 µl cell lysate sample, this was inserted into a Covaris tube holder and transferred into the water bath. The sample was

sonicated using the conditions outlined in Table 2.3, and then transferred into a new tube and centrifuged at 20,000 g at 4°C for 5 min to remove cell debris. The supernatant containing the chromatin was transferred into new tubes and stored at -80°C.

2.15.4 DNA purification of Input Control

Input control (10 µl) was combined with 43 µl of Reverse Crosslinking Buffer and 1µl of Proteinase K and mixed. Samples were incubated at 55°C for 15 min, spun down and incubated again at 65°C for 15 min. Samples were cooled on ice, then magnetic beads added (50 µl purification buffer and 20 µl DNA Purification Magnetic Beads). Samples were mixed, and the DynaMag –PCR Magnet used to create a pellet, which was washed with 150 µl of Wash Buffer, and this step repeated. Supernatant was removed from the pellet and replaced with 150 µl of DNA Elution Buffer. This was incubated at 55°C for 20 min, then spun down to collect sample. A pellet was formed by the magnetic beads using the DynaMag-PCR Magnet once again, and the liquid removed to fresh tube and stored at -20°C.

2.15.5Fragment length determination

The treated cell lysates and purified DNA were visualised using gel electrophoresis with a 2% agarose gel. 10 µl of purified DNA and 20 µl of cell lysate were combined separately with 1x electrophoresis loading dye (25% bromophenol blue, 25% xylene cyanol, 30% glycerol) and run on a 2% agaorse gel in 1x Sodium Borate (SB) buffer (20x SB buffer: 0.8% NaOH (w/v), 4.5% boric acid (w/v)) alongside an Invitrogen 1kb+ electrophoresis size standard, with imaging and photography taking place under a UV light.

CHAPTER 3: RESULTS

3.1 Molecular cloning

PCR was performed on *A. mellifera* cDNA using gene specific primers for *lethal(2)efl* genes *1, 3, 5* and *7, LOC100576174, Gmap, LOC411452* and *LOC412824*. Agarose gel electrophoresis was used to visualise the PCR products (Figure 3.1). Products from the *Lethal(2)efl* genes *1, 3, 5* and *7* produced DNA fragments that ran a distance on the gel corresponding to ~900bp, ~600bp, ~800bp and ~700bp respectively (A-D). All reactions produced fragments of the expected size and had a template free negative control. The DNA fragments from *LOC100576174, Gmap, LOC411452* and *LOC412824* ran a distance on the gel corresponding to ~650bp, ~950bp, ~850bp and ~900bp respectively (E-H). These also produced fragments of the expected size, and each reaction had a negative template free control.

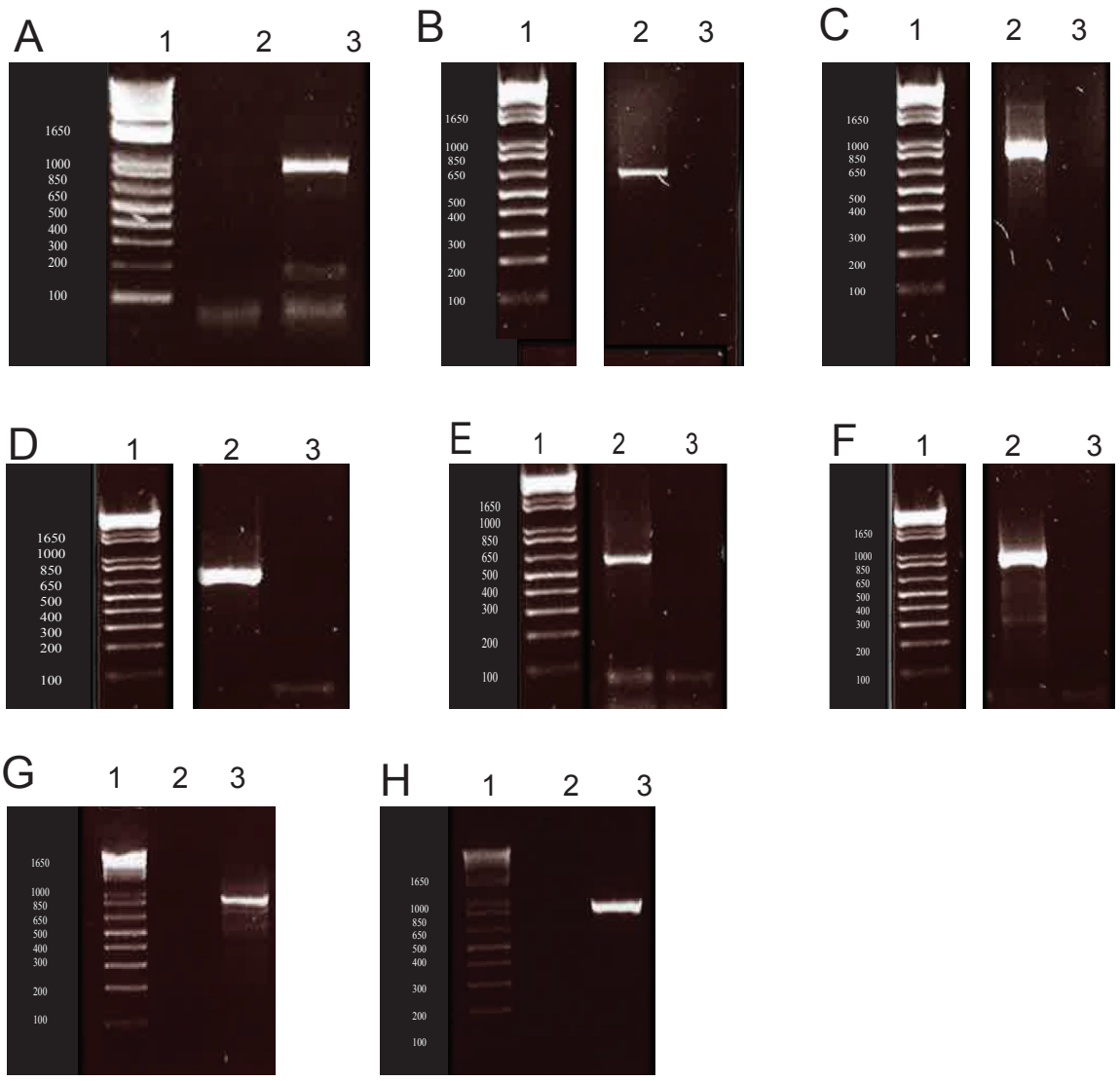


Figure 3.1 PCR products for Lethal genes 1, 3, 5, and 7, LOC1576174, Gmap, LOC411452 and LOC412824. (A) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lane 3: PCR product the expected size for Lethal 1. Lane 2: Template free negative control. (B) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lane 2: PCR product, the expected size of Lethal 3. Lane 3: Template free negative control (irrelevant lanes removed). (C) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lane 2: PCR product, the expected size of Lethal 5. Lane 3: Template free negative control (irrelevant lanes removed). (D) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lane 2: Template free negative control. Lane 3: PCR product, the expected size of Lethal 7 (irrelevant lanes removed). (E) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lane 2: PCR product, the expected size of LOC100576174. Lane 3: Template free negative control (irrelevant lanes removed from figure). (F) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lane 2: PCR product, the expected size of Gmap. Lane 3: Template free negative control (irrelevant lanes removed). (G) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lane 2: Template free negative control. Lane 3: PCR product, the expected size of LOC411452. (H) Lane 1: kb+ ladder, to the left the band sizes in base pairs. Lane 2: Template free negative control. Lane 3: PCR product, the expected size of LOC412824.

As detailed in the materials and methods, amplified DNA fragments that corresponded to the coding regions of *Lethal(2)efl 1, 3, 5* and *7, LOC100576174, Gmap, LOC411452* and *LOC412824* were cloned separately into pBluescript plasmid, then used to transform XLI-Blue *E.coli* cells. Colony PCR was used to identify positive colonies, which were used to inoculate overnight cultures. The plasmid DNA was extracted from these and sent for sequencing, the presence of all of the genes of interest was confirmed using BLAST, as well as the orientation of the inserts relative to the T3 and T7 promoters.

Plasmid DNA was then used to make DIG-labelled RNA antisense and sense probes for *Lethal(2)efl* genes *1, 3, 5* and *7, LOC100576174, Gmap, LOC411452* and *LOC412824*. This was carried out using the restriction enzyme digest method. The digests were visualised using gel electrophoresis (Figures 3.2-3.8). Fragments the size of fully digested DNA for both sense and antisense probes is shown for the *Lethal(2)efl* genes *1, 3, 5* and *7* in figures 3.2-3.5. Products ran on the gel the distance expected for fully digested sense and antisense probes for *LOC100576174, Gmap, LOC411452* and *LOC412824* also, shown in figures 3.6-3.8 respectively. Run-off in vitro transcription was used in the final step of making DIG (digoxygenin) -labelled RNA probes, with Figures 3.2-3.8 showing the presence of sense and antisense probes for each gene (excluding *LOC412824,* which has only as sense probe, as the antisense reaction failed).

3.2 A. mellifera RNA in situ hybridisation

3.2.1 Lethal(2)efl genes 1, 3, 5 and 7

In *Drosophila*, the single copy of *l(2)efl* is expressed in visceral and somatic muscle of stage 13-16 embryos (Tomancak et al., 2002). *A. mellifera* possess 7 copies of *L(2)efl* due to an ancient gene duplication (Romeril, 2013). Expression patterns for these genes had not

been previously determined. RNA *in situ* hybridisation was performed using a DIG-labelled antisense probe for *Lethal (2)efl* genes *1, 3, 5* and *7, LOC100576174, Gmap* and *LOC411452* on *A. mellifera* ovaries for queens, workers and active workers to determine the expression pattern for these genes. RNA-seq and RT-PCR analysis (Figure 1.4) implied the *Lethal(2)efl* genes are more highly expressed in workers, and that expression is reduced as workers activate their ovaries (Romeril, 2013). The *Lethal(2)efl* genes were hypothesised to be co-ordinately regulated, and therefore are expected to be expressed in the same cell types of the ovary. RNA *in situ* hybridisation showed the same pattern of expression for the *Lethal(2)efl* genes *1, 3, 5* and *7* in queen ovaries (Figures 3.2-3.3 (D), 3.4-3.45 (C)). In each case RNA for these genes was detected in cells of the terminal filament, the most anterior region of the ovary (For details of ovariole morphology see Figure 1.2). In some *in situs* staining was also detected in the germarium, however this study focused on the more anterior region, as it is thought to contain the germline stem cells. The exact function of the terminal filament is unknown, however several studies in *Drosophila* indicate these cells might be important in communicating to and maintaining the stem cell niche (González-Reyes, 2003). In all cases where worker and active worker tissue was examined expression of *Lethal(2)efl* was detected in the terminal filament. This was observed in active worker and worker ovaries for the *Lethal(2)efl* gene *3* (Figure 3.3 (F and H)), (arrows indicate the boundaries of the terminal filament as determined by DAPI staining). The worker terminal filament shown in Figure 3.3 (H) had folded within the intima layer, and so (from left to right) is showing posterior to anterior. RNA sense controls were carried out for each *in situ* (Figures 3.2-3.5), these showed no terminal filament staining. Images were taken at differing magnifications, scale bars are shown in figures. Sense controls are used to determine and exclude any pattern of artefact staining through a comparison of the sense and antisense *in situs.* Figure 3.4 (E) shows faint

terminal filament staining for active worker ovaries, as well as staining in the germarium for the *Lethal(2)efl* gene *5*, the region where oocytes differentiate from germline stem cells.

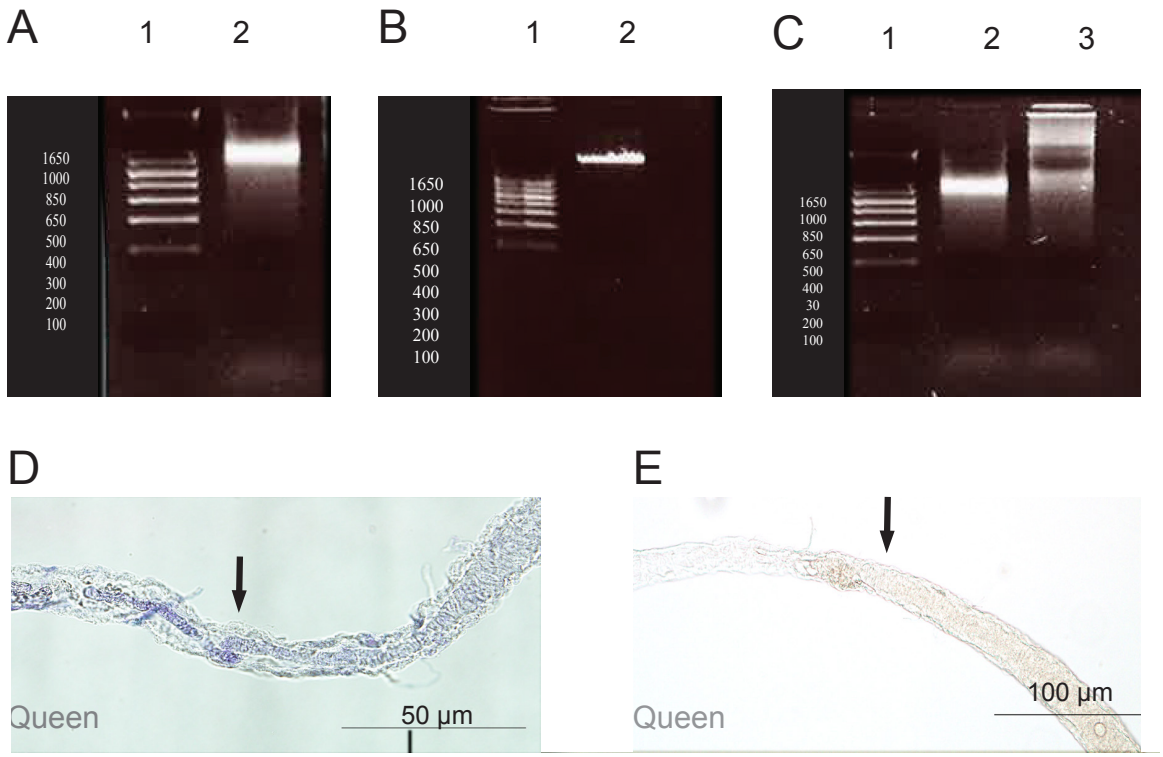


Figure 3.2 Agarose gel electrophoresis restriction digests and probes, and expression of Lethal 1 in honeybee ovaries. (A) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lane 2: Restriction enzyme digest for the sense probe. (B) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lane 2: Restriction enzyme digest for the antisense probe. (C) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Sense and antisense probes respectively.

Honeybee ovaries were hybridised with a DIG-labelled Lethal 1 antisense probe. The anterior region of the ovary is to the left, the posterior to the right. (D) Expression in queen ovaries is detected in the terminal filament (arrows indicate approximate boundaries of terminal filaments). (E) No staining in the terminal filament of sense controls.

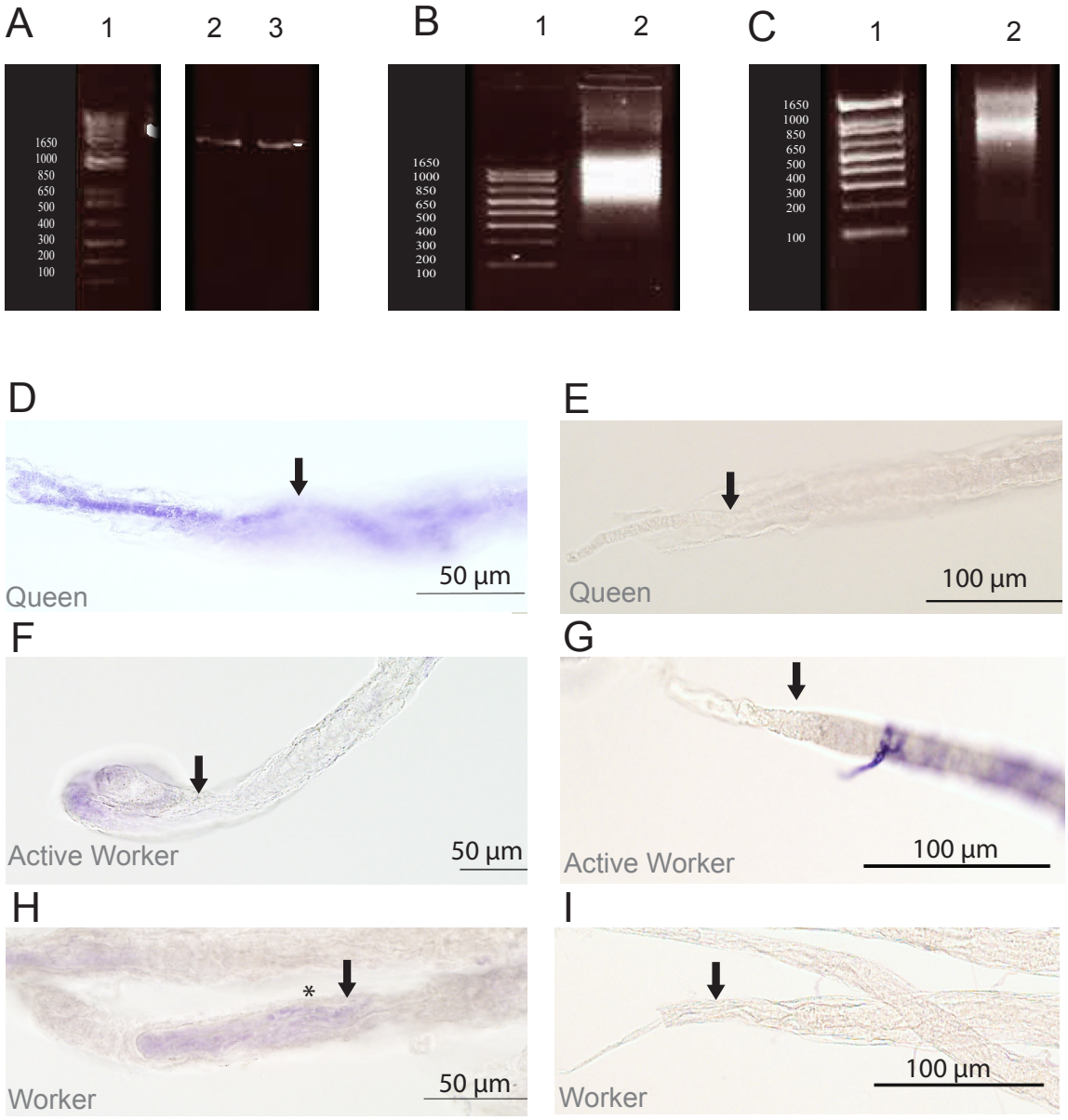


Figure 3.3 Agarose gel electrophoresis of restriction digests and probes, and expression of Lethal 3 in honeybee ovaries. (A) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Restriction enzyme digests for the sense and antisense probes respectively (irrelevant lanes removed). (B) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lane 2: Sense probe. (C) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lane 2: Antisense probe.

Honeybee ovaries were hybridised with a DIG-labelled Lethal 3 antisense probe. The anterior region of the ovary is to the left, the posterior to the right. Expression in queen, active worker and worker ovaries (D, F and H respectively) is detected in the terminal filament (arrows indicate approximate boundary of terminal filament). No terminal filament staining is observed in the sense contorols (E, G and I). * indicates that the terminal filamnet has been folded over, and so is shown from posterior to anterior inside the intima.

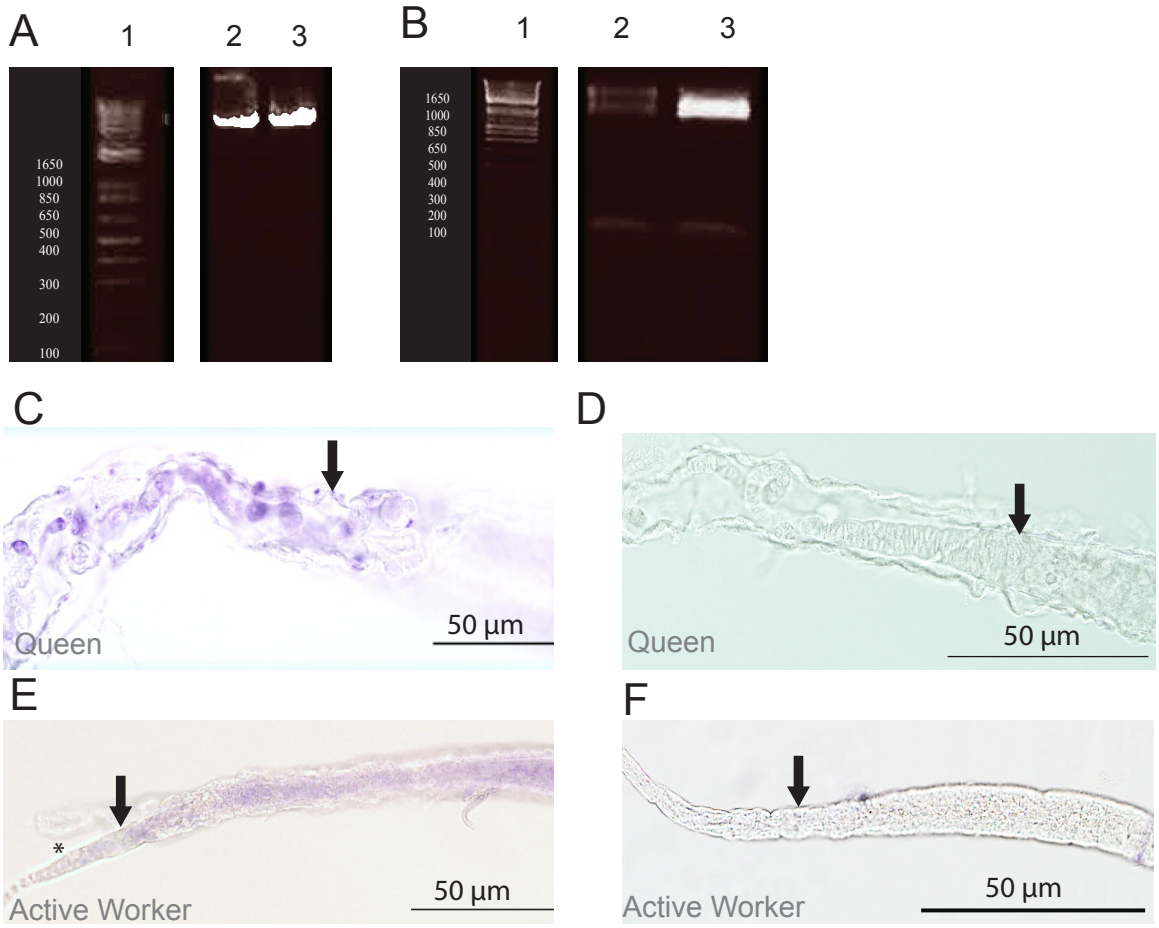


Figure 3.4 Agarose gel electrophoresis of restriction digests and probes, and expression of Lethal 5 in honeybee ovaries. (A) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Restriction enzyme digests for the sense and antisense probes respectively (irrelevant lanes removed). (B) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Sense and antisense probes respectively (irrelevant lanes removed).

Honeybee ovaries were hybridised with a DIG-labelled Lethal 5 antisense probe. The anterior region of the ovary is to the left, the posterior to the right. Expression in queen and active worker ovaries (C and E respectively) is detected in the terminal filament (arrows indicate approximate boundaries of terminal filament). Terminal filament staining is not detected in sense controls (D and F). * indicates faint terminal filament staining, region to the right of the arrow is germarium staining (E).

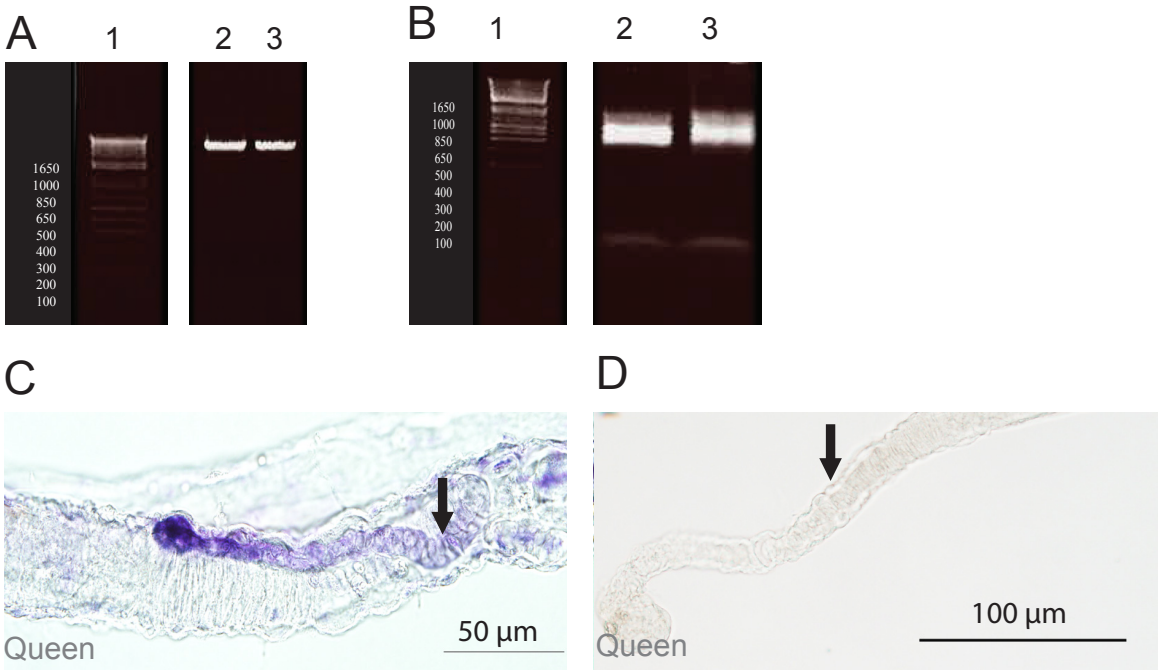


Figure 3.5 Agarose gel electrophoresis restriction enzyme digests and probes, and expression of Lethal 7 in honeybee ovaries. (A) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Restriction enzyme digests for the sense and antisense probes respectively (irrelevant lanes removed). (B) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Sense and antisense probes respectively (irrelevant lanes removed).
Honeybee ovaries were hybridised with a DIG-labelled Lethal 7 antisense probe. The anterior region of the ovary is to the left, the posterior to the right. (C) Expression in queen ovaries is detected in the terminal filament (arrows indicate approximate boundaries of terminal filaments). (D) No terminal filament staining detected in sense control.

3.2.2 LOC100576174 and Gmap

RNA-seq analysis indicated that *Gmap* and *LOC100576174* showed the same trend of higher expression in worker ovaries, and reduced expression in active workers as was seen for the *Lethal(2)efl* genes. This was not statistically significant for *Gmap*, and was thus not predicted to inside the chromatin domain, and so was not expected to show terminal filament staining. RNA *in situ* hybridisation was performed using DIG-labelled antisense probe for *LOC100576174* and *Gmap* in queen, worker and active worker ovaries. The pattern of terminal filament expression observed in *Lethal(2)efl* genes was also detected for *LOC100576174* and *Gmap* in queen ovaries (Figures 3.6-3.7 (C)), active worker ovaries (Figures 3.6-3.7 (E)) and worker ovaries (Figures 3.6- 3.7 (G)). RNA sense controls were carried out for each *in situ*, these showed no terminal filament staining (Figures 3.6-3.7).

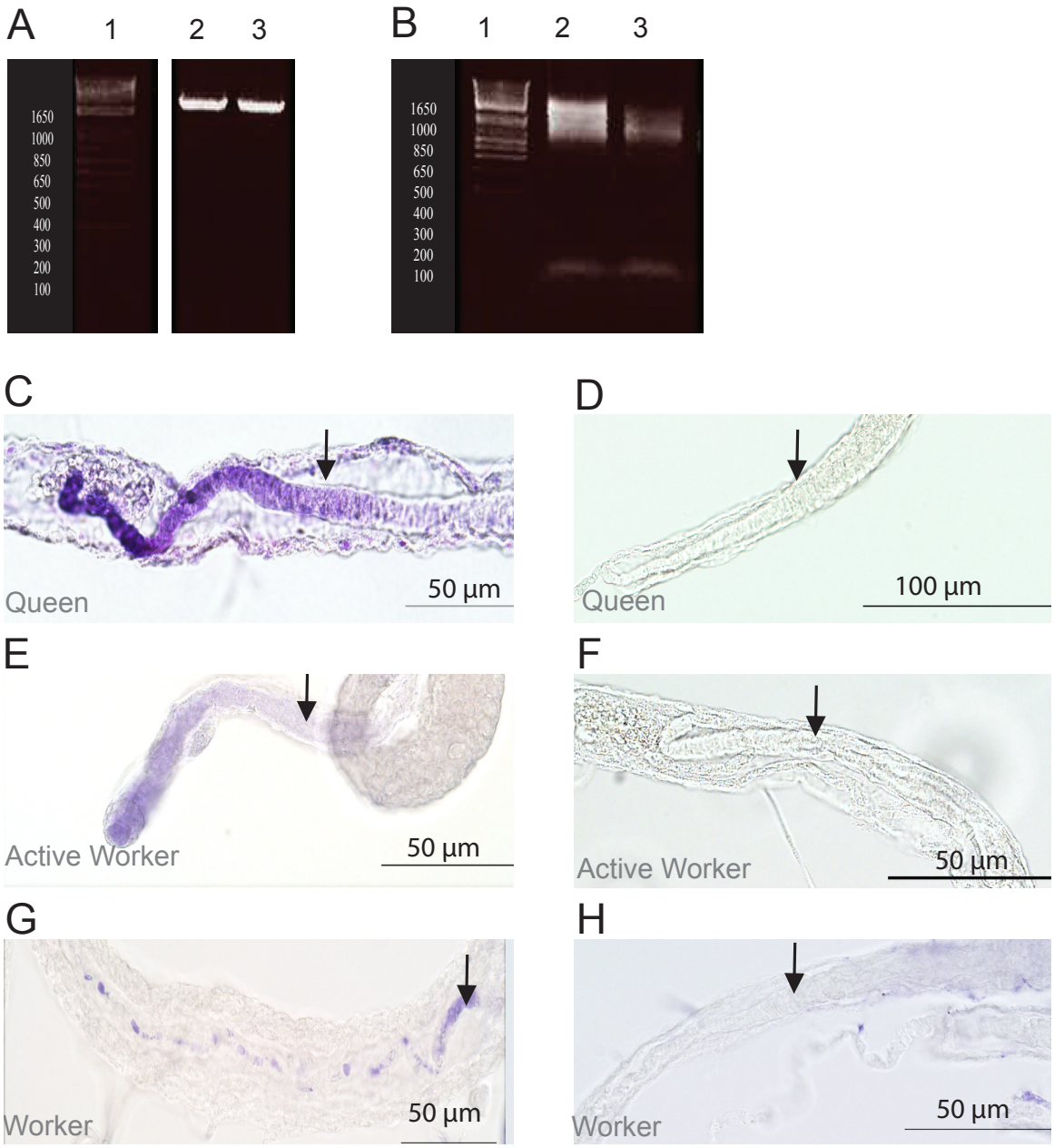


Figure 3.6 Agarose gel electrophoresis of restriction digests and probes, and expression of LOC100576174 in honeybee ovaries. (A) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Restriction enzyme digests for the sense and antisense probes respectively (irrelevant lanes removed from figure). (B) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Sense and Antisense probes respectively.
Honeybee ovaries were hybridised with a DIG-labelled LOC100576174 antisense probe. The anterior region of the ovary is to the left, the posterior to the right. In queen, active worker and worker ovaries (C, E and G respectively) expression is detected in the terminal filament region (arrows indicate approximate boundary of terminal filament). No terminal filament staining is detetcted in the sense controls (D, F and H).

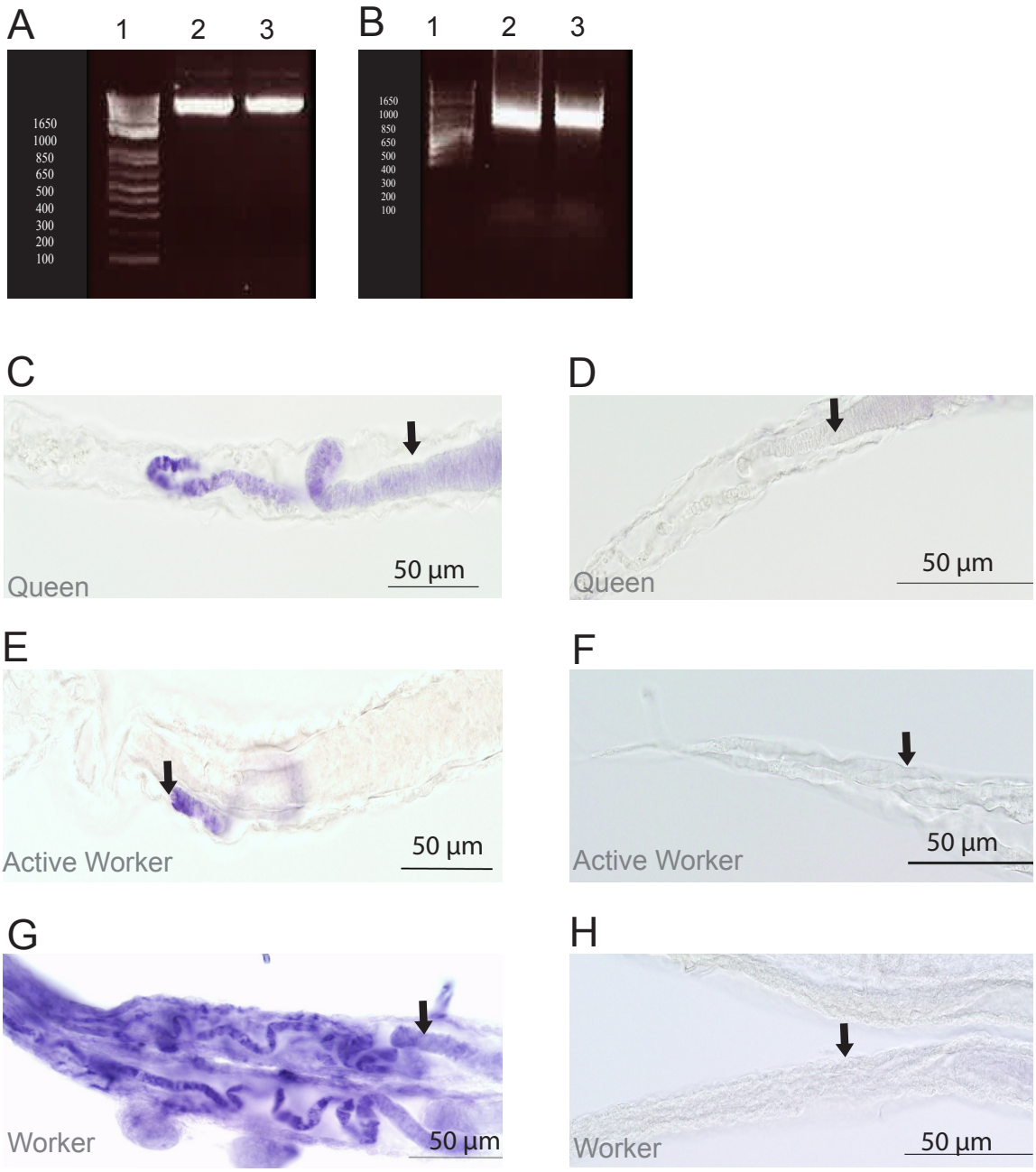


Figure 3.7 Agarose gel electrophoresis of restriction digests and probes, and expression of Gmap in honeybee ovaries. (A) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lanes 2 and 3: Restriction enzyme digests for sense and antisense probes respectively. (B) Lane 1: 1kb+ ladder, to the leftband sizes in base pairs. Lanes 2 and 3: Sense and antisense probes respectively.
Honeybee ovaries were hybridised with a DIG-labelled Gmap antisense probe. The anterior region of the ovary is to the left, the posterior to the right. Expression in queen, active worker and worker ovaries (C, E and G respectively) is detected in the terminal filament (arrows indicate approximate boundaries of terminal filaments). No terminal filament staining is detected in sense controls (D, F, and H).

3.2.3 LOC411452

As the same trend was observed in *LOC100576174, Gmap* and the *Lethal(2)efl* genes, genes were used that lay outside the CTCF sites flanking the lethal cluster (Figure 3.9), to target the boundaries of the chromatin domain. These were genes *LOC411452* and *LOC412824*. Due to time restrictions, RNA *in situ* hybridisation was only performed using a DIG-labelled antisense probe for one gene; *LOC411452*. This was carried out in active worker ovaries. Expression was detected in the terminal filament region also (Figure 3.8 (E). RNA sense controls were carried out for this *in situ*, some of these showed terminal filament staining also, whereas some were clear. Therefore it is possible that the terminal filament staining observed in the antisense was an artefact.

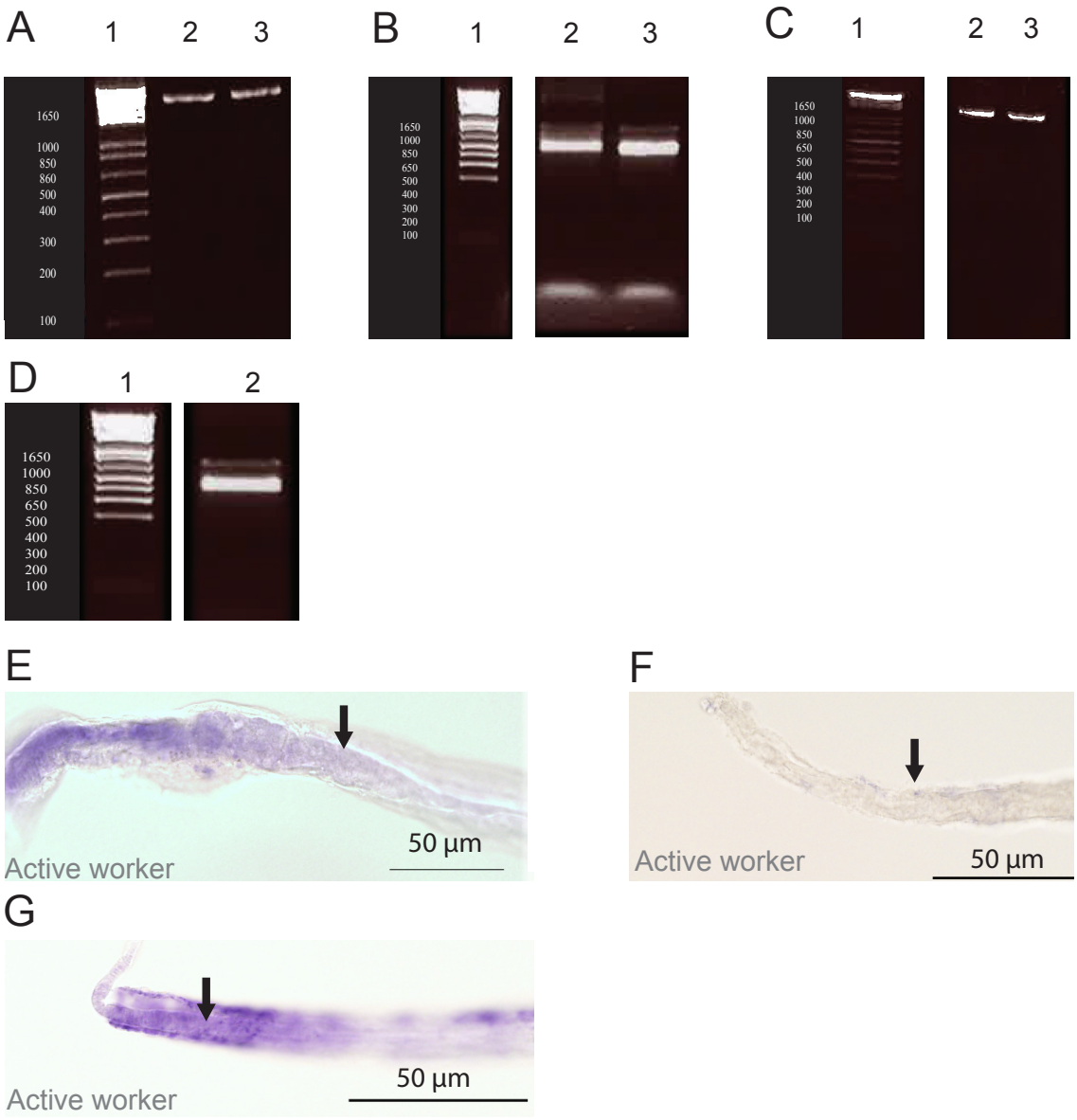


Figure 3.8 Agarose gel electrophoresis of restriction digests and probes for genes LOC411452 and LOC412824, and expression of LOC411452 in honeybee ovaries. (A) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lanes 2 and 3: Restriction enzyme digests for the sense and antisense probes respectively for LOC411452. (B) Lane 1: 1kb+ ladder, to the left band sizes in base paies. Lanes 2 and 3: Sense and antisense probes respectively (irrelevant lanes removed) for LOC411452. (C) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 3: Restriction enzyme digests for the sense and antisense probes respectively (irrelevant lanes removed) for LOC412824. (D) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lane 2: Sense probe (irrelevant lanes removed) for LOC412824.

Honeybee ovaries were hybridised with a DIG-labelled LOC411452 antisense probe. The anterior region of the ovary is to the left, the posterior to the right. (E) Expression in active worker ovaries is detected in the terminal filament (arrows indicate approximate boundaries of terminal filaments). (F) No terminal filamnet staining was detected in some of the sense controls, however some was detected in a proprotion of the sense controls (G).

3.3 Potential CTCF binding sites

As the *in situ* hybridisation investigation had revealed a region of the genome that was showing overlapping patterns of gene expression, and therefore it was likely being co-ordinately regulated, the next step was to determine the boundaries of this region. Potential binding sites for CTCF insulator elements were looked for as a possible boundary for the chromatin domain, as CTCF sites tend to flank regions of the genome that require precise regulation. Their presence could therefore indicate that in regards to gene expression, what is occurring to one gene may be occurring to all the genes within the chromatin domain. Consensus sequences used were from (*Drosophila* (Van Bortle et al., 2012)), for Motif 1 (which is similar to that found in all animals) without allowing for mismatches. CTCF sites were found at two regions with close proximity to the *L(2)efl* cluster, flanking *Gmap*, and upstream of *LOC100576174* (Figure 3.9). The first of these is located between *LOC411452* and *LOC100576174*, at position 4851-4863 (relative to contig 003377928.1), 84bp downstream from *LOC411452*, and 76.5 kb upstream from *LOC100576174* (Figure 3.9). The second is located between *Gmap* and LOC412824, at position 101090-101102 (relative to contig 003377965.1), 764bp downstream from *Gmap*, and ~101kb upstream LOC412824. These CTCF sites contain all of the genes of the *Lethal(2)efl* cluster, as well as *Gmap* and *LOC100576174*, no other genes are contained within this region.

3.4 Potential Transcription Factor Binding Sites

Very few transcription factor binding sites are known for *A. mellifera*, therefore binding sites used were from known *Drosophila* sequences. TRANSFAC Match was used to look for potential binding sites in the 500bp upstream of the transcriptional start site for all of the genes within the CTCF sites, as well as the closest genes upstream and downstream

(*LOC411452* and *LOC412824*). As the *Lethal(2)efl* genes were duplicated in evolutionary history, it is possible that the regulatory regions were duplicated also. This would mean that one (or several) transcription factors bind to all of the genes within the cluster, and that they most likely would not bind to the genes outside the CTCF sites. There were nine transcription factors that had binding sites for the genes of interest (none for all of the genes). These were BR-C Z1, BR-C Z4, CF2-II, Croc, dI, Elf-1, Hairy, HSF and Kr (Table 3.1 and Figure 3.9). Three of these, BR-C Z1, BR-C Z4 and CF2-II had binding sites for all of the genes within the CTCF sites (*Lethal(2)efl* genes as well as *Gmap* and *LOC100576174*). None of these had sites for the genes outside the insulator elements (*LOC411524* and *LOC412824*). Some had binding sites for only some of the genes within the CTCF sites, and none outside. These included *Crocodile* (Croc) which functions in head formation in *Drosophila,* but does not show expression in ovaries (Häcker et al., 1995). It does, however, show expression in honeybee ovaries, and had binding sites for *LOC100576174, Lethal(2)efl* genes *5* and *7*. Others included *Dorsal* (dl) which functions in dorso-ventral patterning in *Drosophila* and shows moderate expression in ovaries of both *Drosophila* and *A. mellifera*, with binding sites upstream of the *Lethal(2)efl 3* and *7*. *Hairy* (Hry), required for patterning early *Drosophila* embryos, which shows low ovary expression (Zhan et al., 2010), with binding sites by *LOC100576174* and *Lethal(2)efl* gene *3*. *Heat shock factor* (HSF), required for oogenesis and early larval development (Jedlicka et al., 1997) shows high expression in *Drosophila* ovaries was the last of this group, and had binding sites upstream of the *Lethal(2)efl* genes *1* and *3*. One gene, *Grainy Head* (Elf-1), required for *Drosophila* embryogenesis (Bray and Kafatos, 1991) does not show ovary expression in *Drosophila* ovaries does in *A. mellifera*, had binding sites upstream of genes both inside the CTCF sites, and outside it. These were *LOC411452, Lethal(2)efl 7, Gmap* and *LOC412824*. *Kruppel* (Kr) which functions in *Drosophila* to establish the anterior-

posterior boundary and is not expressed in the *Drosophila* ovary, yet is in *A. mellifera*, was the only transcription factor that had binding sites only upstream of a gene outside of the CTCF sites, *LOC412824*.

The transcription factors of interest were those that had binding sites upstream from all of the genes within the CTCF sites, and none outside of them. This was BR-C Z1, BR-C Z4 and CF2-II. Using BLAST (Altschul et al., 1990) it was determined that *A. mellifera* does not have an ortholog of *Chorionic factor 2* (CF2-II) involved in *Drosophila* oogenesis (Hsu et al., 1996) and shows moderate expression in *Drosophila* ovaries, therefore it was excluded from further investigation. BR-C Z1 and Z4 do have orthologs in honeybees, which have been shown to be expressed in ovaries and brains (Paul et al., 2006).

Transcription Factors	Binding Sites
BR-C Z1	21085 (*LOC100576174*), 21157 (*LOC100576174*), 21341 (*LOC100576174*), 21361 (*LOC100576174*), 24614 (*Lethal 1*), 24667 (*Lethal 1*), 24689 (*Lethal 1*), 24783 (*Lethal 1*), 24803 (*Lethal 1*), 24818 (*Lethal 1*), 24833 (*Lethal 1*), 24867 *(Lethal 1*), 25060 (*Lethal 1*), 28750 (*Lethal 3*), 28785 *(Lethal 3*), 28987 (*Lethal 3*), 29028 (*Lethal 3*), 29035 (*Lethal 3*), 33768 (*Lethal 5*), 33817 (*Lethal 5*), 39754 (*Lethal 7*), 39805 (*Lethal 7*), 39886 (*Lethal 7*), 100400 (*Gmap*), 100563 (*Gmap*), 100814 (*Gmap*)
BR-C Z4	21100 (*LOC100576174*), 21064 (*LOC100576174*), 21195 (*LOC100576174*), 21249 (*LOC100576174*), 24261 (*LOC100576174*), 21308 (*LOC100576174*), 21353 (*LOC100576174*), 24617 (*Lethal 1*), 24623 *(Lethal 1*), 24640 (*Lethal 1*), 24740 (*Lethal 1*), 24758 (*Lethal 1*), 24772 (*Lethal 1),* 24775 (*Lethal 1*), 24786 (*Lethal 1*), 24870 (*Lethal 1*), 24912 (*Lethal 1*), 24956 (*Lethal 1*), 25013 (*Lethal 1*), 25038 (*Lethal 1*), 28920-28923 (*lethal 3*), 28925 (*Lethal 3*), 28928 (*Lethal 3*), 28935 (*Lethal 3*), 28941 (*Lethal 3*), 28947 (*Lethal 3*), 28952 (*Lethal 3*), 28982 (*Lethal 3*), 29038 (*Letha*l *3*), 33587 (*Lethal 5*), 33593 (*Lethal 5*), 33653 (*Lethal 5*), 33715 (*Lethal 5*), 33729 (*Lethal 5*), 33921 (*Lethal 5*), 33947 *(Lethal 5*), 33950 (*Lethal 5*), 39717 (*Lethal 7*), 39802 (*Lethal 7*), 39912 (*Lethal 7*), 39932 (*Lethal 7*), 40041 (*Lethal 7*), 40138 (*Lethal 7*), 100423 (*Gmap*), 100427 (*Gmap*), 100525 (*Gmap*), 100541 (*Gmap*)
CF2-II	21100 (*LOC100576174*), 21226-21231 (*LOC100576174*), 21240-21243 (*LOC100576174*), 21245 (*LOC100576174*), 21286-21297 (*LOC100576174*), 21299 (*LOC 100576174*), 21332-21339 (*LOC100576174*), 24992 (*Lethal 1*), 24993 (*Lethal 1*), 25050 (*Lethal 1*), 25056-25058 (*Lethal 1*), 28932 (*Lethal 3*), 28935 *(Lethal 3*), 29199 (*Lethal 3*), 29200 (*Lethal 3*), 33959 (*Lethal 5*), 33965 (*Lethal 5*), 39793 (*Lethal 7*), 39874 (*Lethal 7*) , 39875 (*Lethal 7*), 40138 (*Lethal 7*), 100339 *(Gmap*), 100369 (*Gmap*), 100370 (*Gmap*), 100419 (*Gmap*), 100420 (*Gmap*), 100421 (*Gmap*)
Croc	28933 (*LOC100576174*), 33627 (*Lethal 5*), 39744 (*Lethal 7*), 40140 (*Lethal 7*)
dI	29141 (*Lethal 3*), 39700 (*Lethal 7*)
Elf-1	4807 (*LOC411452*), 39860 (*Lethal 7*), 100625 (*Gmap*), 100759 (*Gmap*), 375222 (*LOC412824*)
Hairy	21053 (*LOC100576174*), 29167 (*Lethal 3*)
HSF	25083 (*Lethal 1*), 29108 (*Lethal 3*), 29135 (*Lethal 3*), 29140 *(Lethal 3*)
Kr	375213 (*LOC412824)*

Table 3.1 Transcription factor binding sites shown in relation to the contigs 003377928.1 (for gene *LOC411452*), 003377965.1 (for *LOC100576174, Lethal(2)efl* genes *1, 3, 5* and *7* and *Gmap*), and 003377976.1 (for gene *LOC412824*).

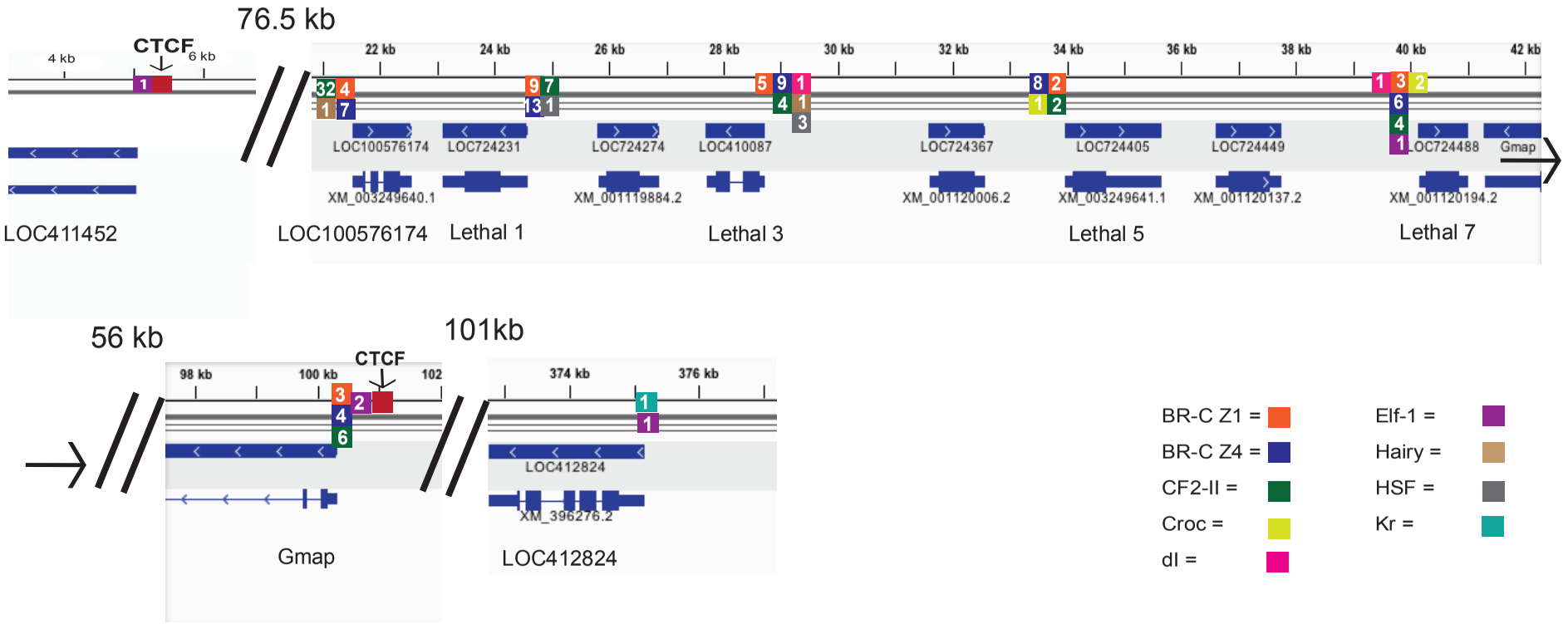


Figure 3.9 Diagram showing location and number of binding sites for transcription factors and CTCF. Figure showing the location of potential binding sites found for transcription factors in the 500bp upstream of the transcriptional start site for the genes LOC411452, LOC100576174, Lethal 1, 3 ,5 and 7, Gmap annd LOC 412824. Coloured boxes represent approximate binding sites (boxes not to scale), and numbers inside each box indicates the number of binding sites for each transcription factor for the gene in closest proximity. CTCF insulator element binding sites are labelled, and located upstream of LOC10576174 and downstream of Gmap.

3.5 Testing qPCR primers for B-RC and E75

In order to test if BR-C is expressed in *A. mellifera,* qPCR was to be utilised to quantify mRNA expression. In order to test the specificity of the primers PCR was performed on *A. mellifera* cDNA and gDNA using gene specific primers for BR-C and E75. Only one publication has investigated the expression of BR-C in honeybees, which looked at this and another ecdysteroid –regulated gene E75 (Paul et al., 2006). This was shown to be expressed in the same region of the queen ovary, and a different neuron type of the worker brain. As it shows similar expression patterns to BR-C and is ecdysteroid-regulated also, but does not have potential binding sites for any of the genes within the chromatin domain, it was used as a control. Agarose gel electrophoresis was used to visualise the PCR products (Figure 3.10). As can be seen in figure 3.10 (A), the DNA fragments in lanes 2 and 4 ran at a distance in the gel corresponding to ~100bp, the expected size of the product for BR-C. Bands were also seen in these lanes corresponding to sizes less than 100 bp. These are primer dimers which prevent these primers from being used, as they reduce the sensitivity of the system, rendering the quantitative capabilities useless. New primers must be designed before BR-C can be quantified. Figure 3.10 (B) shows the DNA fragments in lanes 2 and 4 that ran at a distance in the gel corresponding to ~100bp. The specificity of these primers is highlighted by the PCR product in lane 6 from gDNA which ran a distance on the gel corresponding to ~450bp, as the same genes with the introns intact produce different products. The primers for qPCR of E75 are therefore specific and can be used.

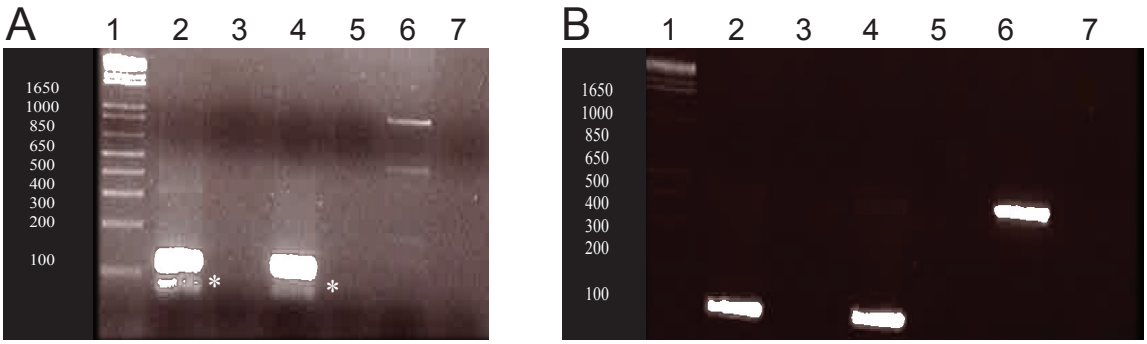


Figure 3.10. Agarose gel electrophoresis of PCR products from qPCR primers for BR-C and E75. (A) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 4: PCR product from cDNA, showing a band at the expected size, and also primer dimer (highlighted by an *) for queen and active worker cDNA respectively. Lanes 3 and 5: -RT controls for queens and active workers respectively, no bands seen. Lane 6: Products from gDNA, showing 3 distinct bands. Lane 7: Template free negative control. (B) Lane 1: 1kb+ ladder, to the left band sizes in base pairs. Lanes 2 and 4: PCR products from cDNA, showing bands at the expected size foe queens and active workers respectively. Lanes 3 and 5: -RT controls for queens and active workers respectively, no bands seen. Lane 6: PRC product from gDNA showing a single band. Lane 7: Template free negative control.

3.6 Chromatin immunoprecipitation

As ovary activation is a phenotypically plastic event in response to an environmental stimuli, it is likely that the mechanisms which underlie it are epigenetic in nature. A study investigating whole genome enrichment of the repressive histone mark H3K27me3 has shown that the enrichment profiles are altered in the *L(2)efl* genes as workers activate their ovaries. In order to investigate this further, chromatin immunoprecipitation qPCR was to be carried out to validate these findings. This works by quantifying the amount of DNA for each gene of interest that is bound to H3K27me3. An antibody for H3K27me3 bound to magnetic beads is used to bind the chromatin. It allows for isolation of the DNA that H3K27me3 binds. After DNA is unbound, qPCR is used to quantify the DNA. Regions that show high levels of enrichment require fewer cycles of qPCR amplify above the threshold level, as more DNA was isolated from the original sample.

3.6.1 Testing quality of chromatin

The quality of the purified DNA and chromatin extracted from worker bees, as well as the effectiveness of sonication was tested using gel electrophoresis. Figure 3.11 (C) Lane 2 shows that DNA was not present, and was lost at some stage of the purification process. Lane 3 however, shows that the sonication protocol has been properly optimised and that the chromatin has been sheared uniformly (as only one band is seen on the gel) to fragments of ~150 bp. Due to time restrictions and the availability of tissue, this experiment could not be repeated.

3.6.2 Testing specificity of chromatin immunoprecipiataion primers

The specificity of the chromatin immunoprecipiataion primers was tested by performing PCR on *A. mellifera* gDNA using specific primers for *LOC10576174, Lethal(2)efl* genes *3, 5* and *7* and *Gmap*. Agarose gel electrophoresis was used to visualise products. Figure

3.11 (B and C) Show 100bp products for *Lethal(2)efl* genes *3, 5* and *7*, as well as *Gmap. LOC100576174* failed to amplify twice, and therefore the primers will be redesigned. All genes had a template free negative control (Figure 3.11). All other primers are specific and can be used for ChIP qPCR.

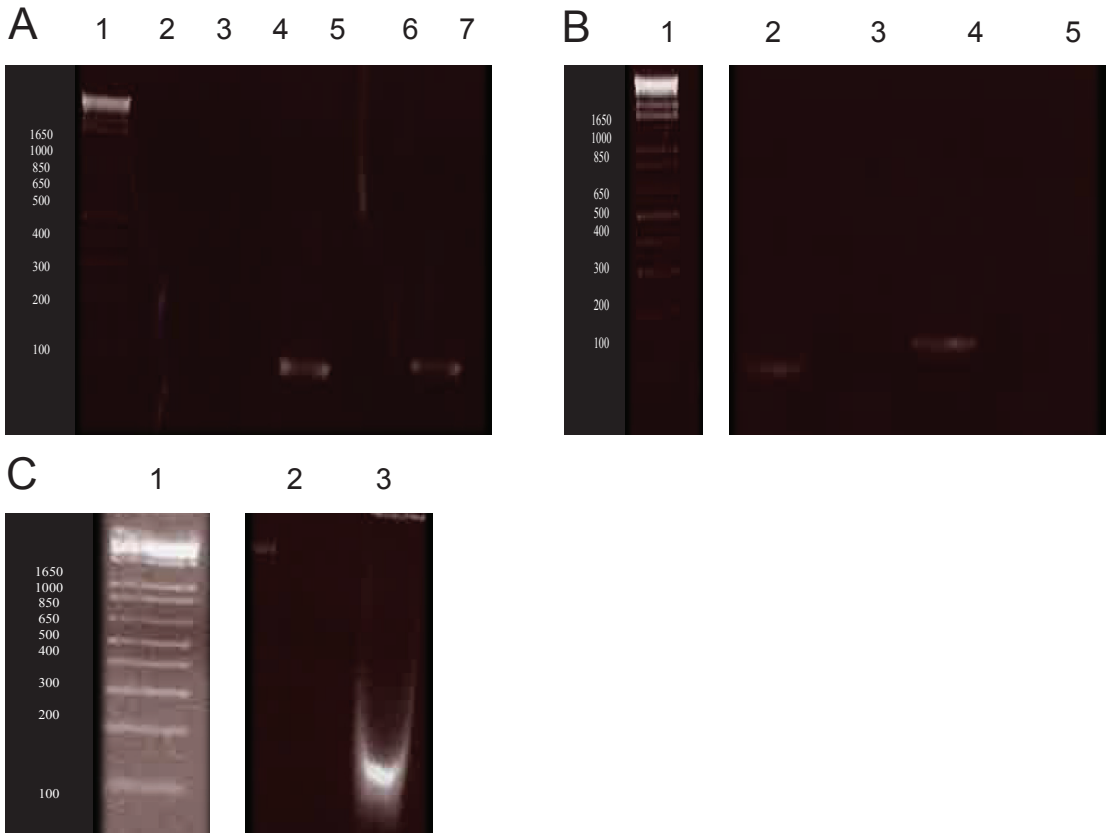


Figure 3.11. Agarose gel electrophoresis of PCR products from ChIP primers for LOC100576174, Lethal 3, 5, 7 and Gmap, and chromatin sample and prurified DNA. (A) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lanes 4 and 6: PCR products, the expected size for Lethal 3 and 5 respectively. Lane 2: No band seen for LOC100576174. Lanes 3, 5 and 7: Template free negative controls. (B) Lane 1: 1kb+ ladder, to the left the band sizes in base pairs. Lanes 2 and 4: PCR products, the expected sizes for Lethal 7 and Gmap respectively. Lanes 3 and 5: Template free negative controls. (C) Lane 1: 1kb+ ladder, to the band sizes in base pairs. Lane 2: Purified DNA sample, no band seen. Lane 3: Sheared chromatin fragment at ~150 bp, the expected size of the sonicated chromatin.

CHAPTER 4: DISCUSSION

4.1 Coordinated regulation of the genes within the Lethal (2) Essential for Life cluster

Investigations into previously uncharacterised gene complexes involved in ovary activation in the honeybee, *Apis mellifera*, has highlighted the Lethal (2) Essential for Life (*L(2)efl*) cluster as a possible candidate region of the genome involved in this phenotypically plastic process (Romeril, 2013). RNA-seq and RT-qPCR data showed that all of the seven genes within this cluster showed the same statistically significant trend of high levels of gene expression in workers, with reduced expression as workers activate their ovaries in response to the environmental event of loss of the queen. This suggested the possibility of large scale changes in gene expression in response to environmental influences leading to phenotypic plasticity.

Previous work in the field has aimed to identify regions of the honeybee genome associated with ovary activation, such as the QTL (quantitative trait loci) study in anarchistic honeybees (Oxley et al., 2008). It found four loci that are responsible for ~25% of the phenotypic variance observed in ovary activation, none of which are linked to the *L(2)efl* cluster investigated in this study. However, it is possible that as these workers are able to activate their ovaries in the presence of the queen, the mechanism used to do so is not representative of the mechanism used by wild type honey bees. Proteomic studies which identified a number of proteins that change in abundance during ovary activation have also not associated the *L(2)efl* group with ovary activation (Cardoen et al., 2012).

In order to test the theory of large scale changes in gene expression in response to environmental influences leading to phenotypic plasticity, evidence of coordinated regulation of gene expression within a chromatin domain was needed. This was investigated using genes of the *L(2)efl* cluster, looking for gene expression in the same cell types of the *A. mellifera* ovary, which would imply coordinated regulation. RNA *in situ* hybridisation was performed to visualise gene expression in queen, worker and active worker ovaries using antisense RNA probes for the *Lethal(2)efl* genes *1, 3, 5* and *7* (every second gene was investigated in this study). This ensured that both ends of the cluster were targeted, allowing for the boundaries of this putative chromatin domain to be interrogated. Expression of all of the genes was detected in the terminal filament region of the ovary for all *Lethal(2)efl* genes investigated for queens, workers and active workers (due to time restrictions and technical difficulties not all expression patterns were obtained for active worker and worker ovaries). This pattern of gene expression being detected in the same cell types for all of the *L(2)efl* genes investigated suggests that these genes are being co-ordinately regulated. However, weak terminal filament expression was detected in active workers for the *Lethal(2)efl* gene *5*, and also in the germarium region. This still supports coordinated regulation in the terminal filament, but it also implies that in the germarium the *Lethal(2)efl* gene *5* is being independently regulated. This implies that regulation of these genes differs depending on the cell type.

The terminal filament is the most anterior region of the ovary, next to where the germline stem cells are hypothesised to reside (Figure 1.2), however the exact function of these cells remains unknown. Despite there being only 5 cell types in the honeybee ovary (oocytes, nurse cells, terminal filament cells, germline stem cells and follicle cells) terminal filament staining is rare, and therefore a consistent expression pattern specific to these cells types is unlikely to be an artefact of staining (Dr Elizabeth Duncan, personal communication) and

is an interesting expression pattern. Further work needs to be undertaken to gather a complete set of *in situ* hybridisation patterns from worker and active worker ovaries for the *L(2)efl* genes in order to gain further evidence for coordinated regulation.

4.2 Interrogating the boundaries of the chromatin domain

In order to determine if the genes *Lethal(2)efl 1* and *7* were the boundaries of the chromatin domain RNA *in situ* hybridisation was performed on *A. mellifera* ovaries from queen, worker and active worker ovaries using antisense probes for the genes *LOC100576174* and *Gmap* (upstream and downstream respectively). Previous RNA-seq data had shown the same trend of high gene expression in workers, with expression being reduced as workers activate their ovaries for these genes, however this trend was not statistically significant for *Gmap*, suggesting these genes were not co-ordinately regulated with the *Lethal(2)efl* genes, and were not part of the chromatin domain. The expression patterns determined from *in situ* hybridisation showed that these genes are expressed in the terminal filament in ovaries from all three types of female honeybees. This would imply that these genes are being regulated in the same way, and are part of the putative chromatin domain that contains the *L(2)efl* genes. These findings imply that the original boundaries of the chromatin domain were underestimated, and that *LOC100576174* and *Gmap* are inside the domain. This also highlights that the RNA-seq analysis was unable to accurately detect the boundaries of this chromatin domain, and that specific regions need to be validated with other techniques.

Further supporting the hypothesis that *LOC100576174*, the *Lethal(2)efl* genes and *Gmap* are within a chromatin domain is the presence of CTCF insulator element binding sites flanking these genes (Figure 3.9). Bioinformatic analysis using CTCF consensus

sequences from *Drosophila* (which are consistent with binding sites in other animals) showed the presence of binding sites upstream of *LOC100576174* (~76.5 kb), and downstream of *Gmap* (~750 bp). There were no genes other than those already discussed present inside these insulator elements. The discovery of CTFC binding sites supports the hypothesis that *LOC100576174* and *Gmap* are within the boundaries of the chromatin domain, as CTCF sites tend to flank regions of the genome that require precise regulation. This implies that, in regard to gene expression, what is regulating one gene within the domain, is likely regulating all genes, a hypothesis supported by *in situ* hybridisation data. Regions flanked by CTCF often serve vital biological function, which makes sense in regards to the proposed role of the genes in ovary activation, a biological function that requires precise regulation in a eusocial insect. It is important that these CTCF sites have only been identified by their consensus sequences for binding sites, and it is unknown whether they are functional in the honeybee ovary. In order to test this chromatin immunoprecipitation, using an antibody specific to honeybee CTCF, could be used to determine if CTCF is binding in any of the three different types of ovary in *A. mellifera*.

4.3 Mechanisms involved in regulation

As ovary activation is a phenotypically plastic response to an environmental influence, it was hypothesised that the mechanisms which underlie this may be epigenetic in nature. This allows for long term changes in gene expression without altering the underlying genetic sequence. One such mechanism is histone modifications. Work has already been conducted in the Laboratory for Evolution and Development into whole genome enrichment of H3K27me3 (a repressive histone mark) in queen, worker and active worker ovaries (Leask, unpublished data). The results suggested that the pattern of enrichment in

active workers is altered to be an intermediate of queens and workers, or more queen-like depending on the region of the genome. This provided evidence that histone modifications are altered in response to environmental cues. H3K27me3 enrichment was also altered at sites within the *L(2)efl* chromatin domain. One example is seen around the transcriptional start site for *Lethal(2)efl 5*, where both queens and active workers show peaks of enrichment, whereas workers do not. This presence of the repressive mark correlates with RNA-seq data for this gene, which shows higher expression in workers, suggesting histones are playing a role in regulating gene expression in response to environmental cues. In order to investigate this further ChIP qPCR primers have been designed for a number of genes within the cluster, and chromatin collected and sheared into 150bp fragments. qPCR has not yet been performed due difficulty in obtaining bees in time restrictions.

Although epigenetic regulation of this genomic region may be important, it is also possible that when these genes were duplicated (to form the seven genes of the cluster) in evolutionary history, the *cis*-regulatory regions were duplicated and retained. Therefore it was possible that one, or a small number, of transcription factors had functional binding sites for all of the genes within the cluster, and were acting to co-ordinately regulate gene expression. To investigate this hypothesis potential binding sites were bioinformatically searched for in the 500bp upstream of the transcriptional start site for all the genes within the chromatin domain, as well as the genes immediately upstream and downstream of the CTCF sites (*LOC411452* and *LOC412824* respectively). As there are few known transcription factor binding sites for *A. mellifera,* sequences used were from *Drosophila* to search for binding sites using TRANSFAC Match. The results showed nine different types of transcription factors which had binding sites, however only three were found to have

binding sites for all of the genes within the chromatin domain, and not for the genes outside of the CTCF sites.

These transcription factors were BR-C Z1 and Z4 (members of the Broad family of transcription factors) and CF2-II. CF2-II was able to be excluded from further investigation as *A. mellifera* does not possess an ortholog of this gene (Dr Elizabeth Duncan, personal communication). The Broad transcription factors have been previously shown through *in situ* hybridisation to be expressed in the worker bee brain (large-type Kenyon cells of the mushroom bodies) and follicle cells of the ovary in queen bees (Paul et al., 2006). However, the staining was carried out on whole abdominal sections, and therefore it is possible there is no data for the terminal filament due to its small diameter and unknown function. It is possible therefore that these transcription factors are expressed in the terminal filament. BR-C is an ecdysteroid-regulated transcription factor, which suggests that hormonal regulation may be playing a role in regulating this system. The same study also investigated E75, an ecdysteroid-regulated transcription factor also, as it shows similar expression patterns to BR-C, but as does not have potential binding sites for any of the genes within the chromatin domain, it was used as a control. It is likely that the ecdysteroid-regulated transcription factor BR-C is acting in ovary activation, as it has been shown to be vital for vitellogenesis (the process of yolk formation as nutrients are deposited in the oocyte) in the mosquito *Aedes aegypti* (Chen et al., 2004). There is also evidence to suggest it a link between ecdysteroids and reproduction in honeybees, as studies have shown a link between this and ovary size in workers (Wang et al., 2012), that in worker ovaries cholesterol is converted into intermediate ecdysteroids in the ovary (Yamazaki et al., 2011), and that ecdysteroid receptors are expressed in the worker and queen brain, and nurse cells of the queen ovary (Takeuchi et al., 2007). The BR-C protein itself has been shown to be significantly upregulated in all regions of the active worker

body, showing that it is likely acting at a functional level (Cardoen et al., 2012). All of this evidence implies that BR-C may be acting to regulate ovary activation.

Primers have been designed in order to make probes for the BR-C transcription factors to perform *in situ* hybridisation, and this will be carried out in an ongoing investigation. In order to quantify the expression of these transcription factors, RT-qPCR will be carried out, with primers being designed and tested already. However due to primer dimer for the BR-C primers, they will need to be redesigned. It must be noted that the binding sites found are potential binding sites and further work, such as chromatin immunoprecipitation needs to be carried out to determine if these sites are functional. The evidence for both histone modifications and transcription factors acting as mechanisms to regulate coordinated gene expression of the *Lethal (2) efl* complex has led to the new hypothesis that these mechanisms are interacting, as opposed to being two alternative hypotheses.

4.4 Genes outside CTCF sites

This study has proposed that the CTCF binding sites flanking the chromatin domain containing the *Lethal(2)efl* cluster are the boundaries of the domain. This implies that the genes outside this (*LOC411452* upstream and *LOC412824* downstream) are regulated in a different manner, and therefore would show different transcription factor binding and patterns of gene expression in *A. mellifera* ovaries. RNA-seq data showed that unlike *LOC100576174* and *Gmap* which showed a similar trend of gene expression to that seen in the *Lethal(2)efl* genes as workers activated their ovaries, *LOC411452* and *LOC412824* did not show a similar trend. This suggested they were regulated in another way, as did their being outside of the CTCF sites, and BR-C Z1 or BR-C Z4 not having potential binding sites. Therefore they are not likely to be regulated co-ordinately with the *Lethal(2)efl*

genes. In order to test this hypothesis *in situ* hybridisation was performed on active worker ovaries using an antisense probe for *LOC411452* (one ovary type and one gene due to time restrictions). The expression pattern determined from this, unexpectedly, showed terminal filament staining, the same pattern observed from the genes within the CTCF binding sites. However, these results must interpreted with caution, as terminal filament staining was detected in some sense controls, therefore further work is required to determine the validity of this expression pattern. If this is the true expression pattern it could suggest that the CTCF binding sites are not the boundaries of the chromatin domain, and the genes being co-ordinately regulated are greater in number than previously anticipated. However, there are other explanations for the expression pattern observed. Firstly, there are only five cell types in *A. mellifera* ovaries; nurse cells, oocytes, follicle cells, germline stem cells and terminal filament cells. As the germline stem cells have not yet been conclusively identified (personal communication Dr Elizabeth Duncan), this leaves only 4 possible cell types in which staining can occur. It is possible therefore, that despite terminal filament staining being relatively rare, the staining observed matches the pattern inside the chromatin domain by chance. Another possibility is that as the CTCF sites have been located by bioinformatics only, it may be that these are non-functional, and therefore they are not the boundaries of the domain. This would allow for the terminal filament expression pattern that seems to be indicative of coordinated regulation to be acting on a greater region of the genome than previously thought. It is also a possibility that this investigation has located a dynamic region of the genome with unknown function, where many genes are detected in the relatively rare pattern of terminal filament expression, yet not necessarily functioning in the plastic response of ovary activation. Further investigations are needed to determine if this pattern can be observed in worker and queen

ovaries, and if it extends further upstream from the chromatin domain, or downstream as well.

4.5 Future Work

Understanding which genes appear to be co-ordinately regulated is difficult when full expression patterns for all of the genes within the CTCF sites for queen, worker and active worker ovaries is not complete. Therefore future work will focus on gathering a full set of expression patterns for these genes. *In situ* hybridisation will also be used to determine expression patterns *LOC411452* and *LOC412824*, upstream and downstream of the CTCF sites. The functionality of these CTCF sites also needs to be determined. Chromatin immunoprecipitation (ChIP) could be carried out to determine if CTCF is binding, and if so, in which ovary type. This technique could also be applied to investigating the transcription factor BR-C, in order to determine if these are playing a role in regulating gene expression. These transcription factors also need to be quantified with qPCR, and the expression pattern determined by *in situ* hybridisation. The data which shows altered H2K27me3 in the *Lethal(2)efl* cluster from previous work needs to be validated, and in order to do this ChIP qPCR will be performed.

Other than these experiments which are required to complete this study, there is further work required to help answer the main questions raised by this investigation. They include asking how does this system work? What is the function of these genes? And does it directly contribute to ovary activation? One mechanism which may be playing a role in this system is DNA methylation. Although gene body methylation, such as that seen in *A. mellifera* does not function to regulate gene expression, it is possible that this is interacting with histone modifications (Ooi et al., 2007).The technique of high resolution melting after

bisulphite treatment could be used to investigate this possibility. Primers for this technique have already been designed for two of the *Lethal(2)efl* genes. Another way in which DNA methylation could be acting to influence this system is through affecting CTCF binding. It has been shown in mammalian systems that DNA methylation at CTCF sites can inhibit the binding, and therefore insulative activity of CTCF (Hark et al., 2000). Preliminary data suggest that DNA methylation plays a role in in ovary activation, however the mechanism it is influencing is currently unknown (Duncan, unpublished data). This makes DNA methylation an interesting avenue of investigation.

In order to determine the role these genes are playing in ovary activation their function needs to be disrupted. One technique that would be particularly useful in answering the questions of how this system is working, and determining the role of the *L(2)efl* genes, is reducing the expression of these genes in adult bees. RNA interference, the standard method for testing the function of genes, does not work in adult bees as the RNA accumulates in the fat body (Jarosch and Moritz, 2011). Ideally, transgenic bees would allow for the role of the *L(2)efl* genes to be investigated by observing the proportion of workers that activate their ovaries with these some, or all, of these genes knocked out both in and out of the presence of the queen. However, this is not yet a technique that has been developed. However, genome editinh techniques, such as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) could be used to mutagenize these genes or knock them out. CRISPR uses guide RNAs to target Cas9 endonuclease to homologous DNA (Blackburn et al., 2013; Hwang et al., 2013). It induces a double stand break, which can be utilised to induce mutations, or targeting two sites can allow for deletion of a region, or the addition of a sequence of interest. This could be used to disrupt the CTCF binding sites, or mutate any of the genes within the *L(2)efl* chromatin domain to determine any functional role they may be playing in ovary activation. It could also be used to introduce new CTCF

sites within the domain to investigate if they are capable of disrupting coordinated regulation. The role of transcription factor binding could also be investigated, by introducing binding sites for BR-C outside of the CTCF sites, or by deleting some from within the domain, thereby determining if the transcription factor binding sites are capable of inducing coordinated regulation without other factors acting.

4.6 Limitations

There were two main limitations in this study. The first of these was the cell type specificity issue. Many analyses are carried out on mixed cell populations, which confounds the sensitivity of the study. In this case, this issue was raised in the intention to carry out analysis such as chromatin immunoprecipitation and high resolution melting. This was exacerbated by the fact the cell type of interest (terminal filament cells) is the smallest cell population in the ovary (excluding the germline stem cells which have not yet been located). Technical difficulties prevent the terminal filament cells exclusively from being dissected out and used solely for analysis. This means that any specific pattern of methylation or histone marks detected is not necessarily representative of what is occurring in the terminal filament cells, but is rather an average across all cell types in the ovary. Even having a single cell type for analysis still is confounded to a certain extent by inherent variation, therefore an ideal system would allow for methylation and histone analysis to be carried out on a single cell. However, until such a time, mixed cell population analysis is all that is possible within *A. mellifera* ovaries to investigate terminal filament cells. The second limitation to this study is seasonality. Despite no studies having been conducted into this area, as worker bees have a limited lifespan, throughout the year different worker bees will live in very dissimilar environments. As it is well established

they are sensitive to environmental perturbation, it is therefore possible there are different regions of the genome expressed during different seasons. It has been observed that the likelihood of a worker bee activating their ovaries differs depending on the time of the season (Hoover et al., 2006). It may be that issues with collecting *in situ* expression patterns arose from using some tissue from summer bees, and some from winter bees. This problem can be avoided in future by collecting sufficient tissue for the year in one season and storing it.

4.7 Conclusions

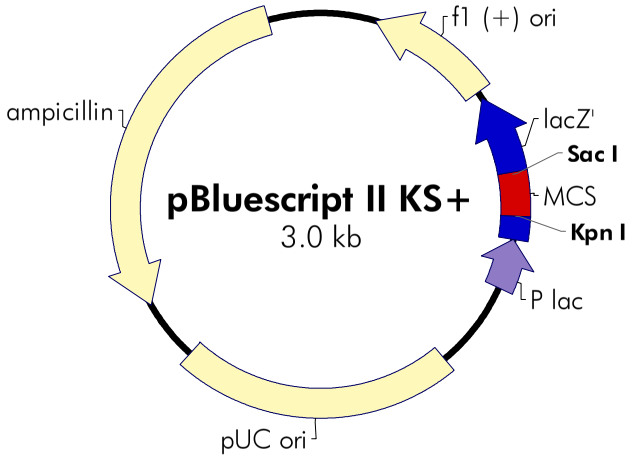
This study aimed to investigate large scale changes in gene expression associated with phenotypic plasticity in response to environmental stimuli. In particular, the phenotypically plastic event of ovary activation in worker *A. mellifera* in response to the environmental stimuli of loss of QMP from the queen. Evidence for the coordinated regulation of the *Lethal (2) Essential for Life* cluster was investigated, as well as the mechanisms which may underlie this. The findings of this study showed that the original boundaries of the chromatin domain had been underestimated, and that the CTCF insulator element binding sites which flank the genes of the *Lethal(2)efl* cluster, as well as *LOC100576174* and *Gmap* appear to be the boundaries of the coordinated regulation. All of the genes within these sites appear to be co-ordinately regulated in response to environmental cues, with expression of these genes occurring in the terminal filament cells of the ovary in queens, workers and active workers. The proposed model for the mechanisms which underlie this is an interaction of histone modifications and ecdysteroid-regulated transcription factors. From here, further work is required to determine the functionality of the findings associated with histones, transcription factors and CTCF sites,

as well as to establish a full set of expression patterns. Furthermore, to investigate if DNA methylation is acting as a mechanism in the process of ovary activation. This work provides evidence for large scale, coordinated changes in gene expression leading to phenotypic plasticity in response to an environmental influence.

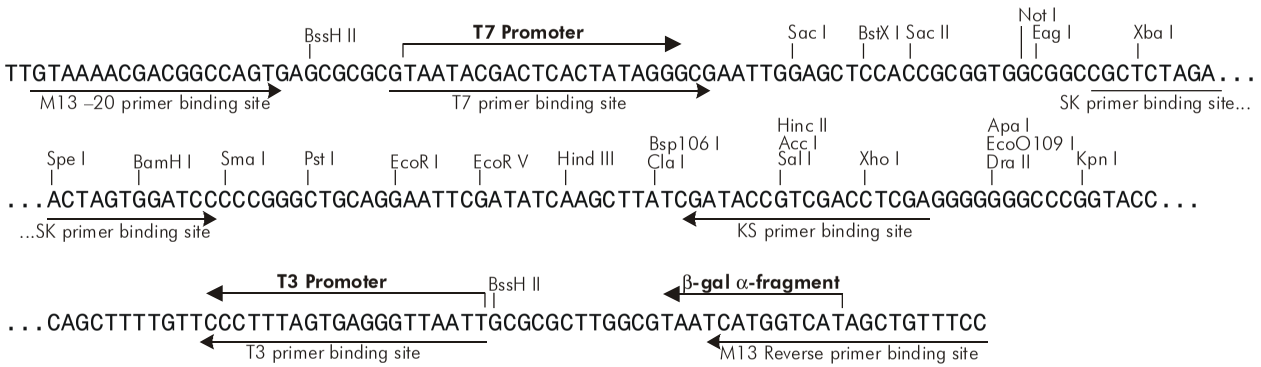
ACKNOWLEDGEMENTS

I would like to thank my fantastic supervisors Associate Professor Peter Dearden and Dr Elizabeth Duncan, for all of their help, support and vast amounts of patience. I would also like to thank the other members of the Laboratory for Evolution and Development who have been so welcoming and fun. I can't imagine a better group of people to work with. Last but not least, I would like to thank my family, who have supported and encouraged me throughout my years at university. This wouldn't have been possible without you all.

f1 (+) origin 135–441
β-galactosidase α-fragment 460–816
multiple cloning site 653–760
***lac* promoter** 817–938
pUC origin 1158–1825
ampicillin resistance (*bla*) ORF 1976–2833



pBluescript II KS (+/–) Multiple Cloning Site Region (sequence shown 598–826)



REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology 215, 403-410.

Ashcroft, M., Kubbutat, M.H., Vousden, K.H., 1999. Regulation of p53 function and stability by phosphorylation. Molecular and cellular biology 19, 1751-1758.

Basak, S., Pookot, D., Noonan, E.J., Dahiya, R., 2008. Genistein down-regulates androgen receptor by modulating HDAC6-Hsp90 chaperone function. Molecular cancer therapeutics 7, 3195-3202.

Bassan, H., Leider Trejo, L., Kariv, N., Bassan, M., Berger, E., Fattal, A., Gozes, I., Harel, S., 2000. Experimental intrauterine growth retardation alters renal development. Pediatr Nephrol 15, 192-195.

Bateson, P., Barker, D., Clutton-Brock, T., Debal, D., et al., 2004. Developmental plasticity and human health. Nature 430, 419-421.

Bell, A.C., Felsenfeld, G., 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482-485.

Bell, A.C., West, A.G., Felsenfeld, G., 1999. The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. Cell 98, 387-396.

Bellinger, L., Lilley, C., Langley-Evans, S.C., 2004. Prenatal exposure to a maternal low-protein diet programmes a preference for high-fat foods in the young adult rat. British Journal of Nutrition 92, 513-520.

Bestor, T.H., 2000. The DNA methyltransferases of mammals. Human Molecular Genetics 9, 2395-2402.

Blackburn, P.R., Campbell, J.M., Clark, K.J., Ekker, S.C., 2013. The CRISPR System—Keeping Zebrafish Gene Targeting Fresh. Zebrafish 10, 116-118.

Bray, S., Kafatos, F., 1991. Developmental function of Elf-1: an essential transcription factor during embryogenesis in Drosophila. Genes & development 5, 1672-1683.

Butler, C., Callow, R., Johnston, N.C., 1959. Extraction and purification of 'queen substance'from queen bees. Nature 184, 1871-1871.

Cardoen, D., Ernst, U.R., Boerjan, B., Bogaerts, A., Formesyn, E., de Graaf, D.C., Wenseleers, T., Schoofs, L., Verleyen, P., 2012. Worker Honeybee Sterility: A Proteomic Analysis of Suppressed Ovary Activation. Journal of Proteome Research 11, 2838-2850.

Chen, L., Zhu, J., Sun, G., Raikhel, A., 2004. The early gene Broad is involved in the ecdysteroid hierarchy governing vitellogenesis of the mosquito Aedes aegypti. Journal of molecular endocrinology 33, 743-761.

Dearden, P., 2006. Germ cell development in the Honeybee (Apis mellifera); Vasa and Nanos expression. BMC Developmental Biology 6, 6.

Dennison, E.M., Arden, N.K., Keen, R.W., Syddall, H., Day, I.N.M., Spector, T.D., Cooper, C., 2001. Birthweight, vitamin D receptor genotype and the programming of osteoporosis. Paediatric and Perinatal Epidemiology 15, 211-219.

Dittrich, B., Buiting, K., Korn, B., Rickard, S., Buxton, J., Saitoh, S., Nicholls, R.D., Poustka, A., Winterpacht, A., Zabel, B., 1996. Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene. Nature genetics 14, 163-170.

Dolinoy, D.C., 2008. The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. Nutrition Reviews 66, S7-S11.

Dolinoy, D.C., Weidman, J.R., Waterland, R.A., Jirtle, R.L., 2006. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by

modifying the fetal epigenome. Environmental health perspectives 114, 567.

Doughty, P., Reznick, D., 2004. Patterns and analysis of adaptive phenotypic plasticity in animals.

Duhl, D.M., Vrieling, H., Miller, K.A., Wolff, G.L., Barsh, G.S., 1994. Neomorphic agouti mutations in obese yellow mice. Nature genetics 8, 59-65.

Duncan, E., Wilson, M., Smith, J., Dearden, P., 2008. Evolutionary origin and genomic organisation of runt-domain containing genes in arthropods. BMC Genomics 9, 558.

Duncan, E.J., Dearden, P.K., 2010. Evolution of a genomic regulatory domain: The role of gene co-option and gene duplication in the Enhancer of split complex. Genome Research 20, 917-928.

Esteller, M., Herman, J.G., 2002. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. The Journal of pathology 196, 1-7.

Falckenhayn, C., Foret, S., Kucharski, R., Lyko, F., Maleszka, R., Wolf, S., 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biology 8.

Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.-F., Andrews, S., Dean, W., Hemberger, M., Reik, W., 2008. Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS genetics 4, e1000116.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., Ukomadu, C., Sadler, K.C., Pradhan, S., Pellegrini, M., Jacobsen, S.E., 2010. Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences 107, 8689-8694.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., Lobanenkov, V.V., 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Molecular and cellular biology 16, 2802-2813.

Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S.E., Robinson, G.E., Maleszka, R., 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proceedings of the National Academy of Sciences 109, 4968-4973.

Forrester, T.E., Badaloo, A.V., Boyne, M.S., Osmond, C., Thompson, D., Green, C., Taylor-Bryan, C., Barnett, A., Soares-Wynter, S., Hanson, M.A., 2012. Prenatal factors contribute to the emergence of kwashiorkor or marasmus in severe undernutrition: evidence for the predictive adaptation model. PloS one 7, e35907.

Francis, D., Diorio, J., Liu, D., Meaney, M.J., 1999. Nongenomic transmission across generations of maternal behavior and stress responses in the rat. Science 286, 1155-1158.

Gluckman, P.D., Hanson, M.A., 2004. Living with the Past: Evolution, Development, and Patterns of Disease. Science 305, 1733-1736.

Gluckman, P.D., Hanson, M.A., Spencer, H.G., 2005. Predictive adaptive responses and human evolution. Trends in Ecology & Evolution 20, 527-533.

Gluckman, P.D., Lillycrop, K.A., Vickers, M.H., Pleasants, A.B., Phillips, E.S., Beedle, A.S., Burdge, G.C., Hanson, M.A., 2007. Metabolic plasticity during mammalian development is directionally dependent on early nutritional status. Proceedings of the National Academy of Sciences 104, 12796-12800.

González-Reyes, A., 2003. Stem cells, niches and cadherins: a view from Drosophila. Journal of cell science 116, 949-954.

Groot, A.P., Voogd, S., 1954. On the ovary development in queenless worker bees (Apis mellifica L.). Experientia 10, 384-385.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S.-g., 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477, 606-610.

Häcker, U., Kaufmann, E., Hartmann, C., Jürgens, G., Knöchel, W., Jäckle, H., 1995. The Drosophila fork head domain protein crocodile is required for the establishment of head structures. The EMBO journal 14, 5306.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., Tilghman, S.M., 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405, 486-489.

Hattori, N., Nishino, K., Ko, Y.-g., Hattori, N., Ohgane, J., Tanaka, S., Shiota, K., 2004. Epigenetic Control of Mouse Oct-4 Gene Expression in Embryonic Stem Cells and Trophoblast Stem Cells. Journal of Biological Chemistry 279, 17063-17069.

Holliday, R., Pugh, J., 1975. DNA modification mechanisms and gene activity during development. Science 187, 226-232.

Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee Apis mellifera. Nature 443, 931+.

Hoover, S.E., Higo, H.A., Winston, M.L., 2006. Worker honey bee ovary development: seasonal variation and the influence of larval and adult nutrition. Journal of Comparative Physiology B 176, 55-63.

Hsu, T., Bagni, C., Sutherland, J.D., Kafatos, F.C., 1996. The transcriptional factor CF2 is a mediator of EGF-R-activated dorsoventral patterning in Drosophila oogenesis. Genes & development 10, 1411-1421.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.J., Joung, J.K., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229.

Ibáñez, L., Valls, C., Potau, N., Marcos, M.V., De Zegher, F., 2001. Polycystic ovary syndrome after precocious pubarche: ontogeny of the low-birthweight effect. Clinical Endocrinology 55, 667-672.

Imamura, M., Miura, K., Iwabuchi, K., Ichisaka, T., Nakagawa, M., Lee, J., Kanatsu-Shinohara, M., Shinohara, T., Yamanaka, S., 2006. Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. BMC developmental biology 6, 34.

Jaenisch, R., Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature Genetics 33 Suppl 1, 245-254.

Jarosch, A., Moritz, R., 2011. Systemic RNA-interference in the honeybee< i> Apis mellifera</i>: Tissue dependent uptake of fluorescent siRNA after intra-abdominal application observed by laser-scanning microscopy. Journal of insect physiology 57, 851-857.

Jedlicka, P., Mortin, M.A., Wu, C., 1997. Multiple functions of Drosophila heat shock transcription factor in vivo. The EMBO journal 16, 2452-2462.

Jiang, L., Zhang, J., Wang, J.-J., Wang, L., Zhang, L., Li, G., Yang, X., Ma, X., Sun, X., Cai, J., Zhang, J., Huang, X., Yu, M., Wang, X., Liu, F., Wu, C.-I., He, C., Zhang, B., Ci, W., Liu, J., 2013. Sperm, but Not Oocyte, DNA Methylome Is Inherited by Zebrafish Early Embryos. Cell 153, 773-784.

Kaati, G., Bygren, L.O., Edvinsson, S., 2002. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. European Journal of Human Genetics : EJHG 10, 682-688.

Kaati, G., Bygren, L.O., Pembrey, M., Sjöström, M., 2007. Transgenerational response to nutrition, early life circumstances and longevity. European Journal of Human Genetics : EJHG 15, 784-790.

Krumlauf, R., 1992. Evolution of the vertebrate Hox homeobox genes. BioEssays 14, 245-252.

Kucharski, R., Maleszka, J., Foret, S., Maleszka, R., 2008. Nutritional Control of Reproductive Status in Honeybees via DNA Methylation. Science 319, 1827-1830.

Lamartiniere, C.A., Cotroneo, M.S., Fritz, W.A., Wang, J., Mentor-Marcel, R., Elgavish, A., 2002. Genistein Chemoprevention: Timing and Mechanisms of Action in Murine Mammary and Prostate. The Journal of Nutrition 132, 552S-558S.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Leask, M.P., 2014. Epigenetics in the honeybee ovary. University of Otago.

Lee, T.M., Zucker, I., 1988. Vole infant development is influenced perinatally by maternal photoperiodic history. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology 255, R831-R838.

Li-Byarlay, H., Li, Y., Stroud, H., Feng, S., Newman, T.C., Kaneda, M., Hou, K.K., Worley, K.C., Elsik, C.G., Wickline, S.A., Jacobsen, S.E., Ma, J., Robinson, G.E., 2013. RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. Proceedings of the National Academy of Sciences 110, 12750-12755.

Lillycrop, K.A., Phillips, E.S., Jackson, A.A., Hanson, M.A., Burdge, G.C., 2005. Dietary Protein Restriction of Pregnant Rats Induces and Folic Acid Supplementation Prevents Epigenetic Modification of Hepatic Gene Expression in the Offspring. The Journal of Nutrition 135, 1382-1386.

Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D., Felsenfeld, G., 2001. Correlation Between Histone Lysine Methylation and Developmental Changes at the Chicken beta -Globin Locus. Science 293, 2453-2455.

Lorincz, M.C., Dickerson, D.R., Schmitt, M., Groudine, M., 2004. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nature structural & molecular biology 11, 1068-1075.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389, 251-260.

Lyko, F., Maleszka, R., 2011. Insects as innovative models for functional studies of DNA methylation. Trends in Genetics 27, 127-131.

Maleszka, R., 2008. Epigenetic integration of environmental and genomic signals in honey bees: the critical interplay of nutritional, brain and reproductive networks. Epigenetics 3, 188-192.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466, 253-257.

Mazet, F., Amemiya, C.T., Shimeld, S.M., 2006. An ancient Fox gene cluster in bilaterian animals. Current Biology 16, R314-R316.

Miltenberger, R.J., Mynatt, R.L., Wilkinson, J.E., Woychik, R.P., 1997. The Role of the agouti Gene in the Yellow Obese Syndrome. The Journal of Nutrition 127, 1902S-1907S.

Moczek, A.P., Snell-Rood, E.C., 2008. The basis of bee-ing different: the role of gene silencing in plasticity. Evolution & Development 10, 511-513.

Mohandas, T., Sparkes, R., Shapiro, L., 1981. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science 211, 393-396.

Morgan, H.D., Sutherland, H.G., Martin, D.I., Whitelaw, E., 1999. Epigenetic inheritance at the agouti locus in the mouse. Nature genetics 23, 314-318.

Mutskov, V., Felsenfeld, G., 2003. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. The EMBO journal 23, 138-149.

Narlikar, G.J., Fan, H.-Y., Kingston, R.E., 2002. Cooperation between Complexes that Regulate Chromatin Structure and Transcription. Cell 108, 475-487.

Nightingale, K.P., O'Neill, L.P., Turner, B.M., 2006. Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. Current Opinion in Genetics & Development 16, 125-136.

Nijhout, H.F., 2003. Development and evolution of adaptive polyphenisms. Evolution & Development 5, 9-18.

Nusse, R., 2001. An ancient cluster of Wnt paralogues. Trends in Genetics 17, 443.

O'Neill, L.P., Turner, B.M., 1996. Immunoprecipitation of chromatin, in: Sankar, A. (Ed.), Methods in Enzymology. Academic Press, pp. 189-197.

Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell 99, 247-257.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714-717.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., Walter, J., 2000. Active demethylation of the paternal genome in the mouse zygote. Current Biology 10, 475-478.

Oxley, P.R., Thompson, G.J., Oldroyd, B.P., 2008. Four quantitative trait loci that influence worker sterility in the honeybee (Apis mellifera). Genetics 179, 1337-1343.

Page, R.E., Jr., Erickson, E.H., Jr., 1988. Reproduction by Worker Honey Bees (Apis mellifera L.). Behavioral Ecology and Sociobiology 23, 117-126.

Painter, R.C., Roseboom, T.J., Bleker, O.P., 2005. Prenatal exposure to the Dutch famine and disease in later life: An overview. Reproductive Toxicology 20, 345-352.

Paul, R.K., Takeuchi, H., Kubo, T., 2006. Expression of Two Ecdysteroid-Regulated Genes, Broad-Complex and E75, in the Brain and Ovary of the Honeybee (Apis mellifera L.). Zoological Science 23, 1085-1092.

Pembrey, M.E., Bygren, L.O., Kaati, G., Edvinsson, S., Northstone, K., Sjöström, M., Golding, J., 2006. Sex-specific, male-line transgenerational responses in humans. European Journal of Human Genetics 14, 159-166.

Pham, T.D., MacLennan, N.K., Chiu, C.T., Laksana, G.S., Hsu, J.L., Lane, R.H., 2003. Uteroplacental insufficiency increases apoptosis and alters p53 gene methylation in the full-term IUGR rat kidney. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology 285, R962-R970.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K., Young, R.A., 2005. Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. Cell 122, 517-527.

Potok, Magdalena E., Nix, David A., Parnell, Timothy J., Cairns, Bradley R., 2013. Reprogramming the Maternal Zebrafish Genome after Fertilization to Match the Paternal Methylation Pattern. Cell 153, 759-772.

Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I., Whitelaw, E., 2002. Metastable epialleles in mammals. Trends in Genetics 18, 348-351.

Riggs, A.D., 1975. X inactivation, differentiation, and DNA methylation. Cytogenetic and Genome Research 14, 9-25.

Romeril, A.J., 2013. Gene Complexes and their role in Worker Honeybee (Apis mellifera) Ovary Activation. University of Otago.

Rutherford, S.L., Lindquist, S., 1998. Hsp90 as a capacitor for morphological evolution. Nature 396, 336-342.

Sarda, S., Zeng, J., Hunt, B.G., Yi, S.V., 2012. The Evolution of Invertebrate Gene Body Methylation. Molecular Biology and Evolution 29, 1907-1916.

Sati, S., Tanwar, V.S., Kumar, K.A., Patowary, A., Jain, V., Ghosh, S., Ahmad, S., Singh, M., Reddy, S.U., Chandak, G.R., 2012. High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region. PloS one 7, e31621.

Schlichting, C.D., 1986. The evolution of phenotypic plasticity in plants. Annual review of ecology and systematics 17, 667-693.

Schübeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J., 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes & development 18, 1263-1271.

Shiota, K., 2004. DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. Cytogenetic and genome research 105, 325-334.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., Oberdoerffer, S., 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479, 74-79.

Stanner, S.A., Bulmer, K., Andres, C., Lantseva, O.E., et al., 1997. Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross sectional study. British Medical Journal 315, 1342-1348.

Takeuchi, H., Paul, R.K., Matsuzaka, E., Kubo, T., 2007. EcR-A expression in the brain and ovary of the honeybee (Apis mellifera L.). Zoological science 24, 596-603.

Tatiana, V., Veronika, S., Daniel, S., 2004. Cellular and Physiological Effects of Soy Flavonoids. Mini Reviews in Medicinal Chemistry 4, 881-887.

Thompson, C., Syddall, H., Rodin, I., Osmond, C., Barker, D.J.P., 2001. Birth weight and the risk of depressive disorder in late life. The British Journal of Psychiatry 179, 450-455.

Thompson, G.J., Yockey, H., Lim, J., Oldroyd, B.P., 2007. Experimental manipulation of ovary activation and gene expression in honey bee (Apis mellifera) queens and workers: testing hypotheses of reproductive regulation. Journal of Experimental Zoology Part A: Ecological Genetics and Physiology 307A, 600-610.

Thorp, R.W., 2000. The collection of pollen by bees. Pl Syst Evol 222, 211-223.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., 2002. Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol 3, 0081-0088.

Townsend, G.F., Lucas, C.C., 1940. The chemical nature of royal jelly. Biochemical Journal 34, 1155.

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J.E., Corces, V.G., 2012. Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. Genome research 22, 2176-2187.

Vickers, M., Gluckman, P., Coveny, A., Hofman, P., Cutfield, W., Gertler, A., Breier, B., Harris, M., 2005. Neonatal leptin treatment reverses developmental programming. Endocrinology 146, 4211-4216.

Wahlbeck, K., Forsén, T., Osmond, C., Barker, D.P., Eriksson, J.G., 2001. ASsociation of schizophrenia with low maternal body mass index, small size at birth, and thinness during childhood. Archives of General Psychiatry 58, 48-52.

Wang, Y., Jorda, M., Jones, P.L., Maleszka, R., Ling, X., Robertson, H.M., Mizzen, C.A., Peinado, M.A., Robinson, G.E., 2006. Functional CpG Methylation System in a Social Insect. Science 314, 645-647.

Wang, Y., Kocher, S.D., Linksvayer, T.A., Grozinger, C.M., Page, R.E., Amdam, G.V., 2012. Regulation of behaviorally associated gene networks in worker honey bee ovaries. The Journal of experimental biology 215, 124-134.

Weaver, I.C., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., Meaney, M.J., 2004. Epigenetic programming by maternal behavior. Nature neuroscience 7, 847-854.

West-Eberhard, M.J., 1989. Phenotypic Plasticity and the Origins of Diversity. Annual Review of Ecology and Systematics 20, 249-278.

Wojdacz, T.K., Dobrovic, A., 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic Acids Research 35, e41.

Wolffe, A.P., Matzke, M.A., 1999. Epigenetics: Regulation Through Repression. Science 286, 481-486.

Yamazaki, Y., Kiuchi, M., Takeuchi, H., Kubo, T., 2011. Ecdysteroid biosynthesis in workers of the European honeybee Apis mellifera L. Insect Biochemistry and Molecular Biology 41, 283-293.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Jin, P., Ren, B., He, C., 2012. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nature protocols 7, 2159-2170.

Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. Science 328, 916-919.

Zhan, Y., Maung, S.W., Shao, B., Myat, M.M., 2010. The bHLH Transcription Factor, Hairy, Refines the Terminal Cell Fate in the Drosophila Embryonic Trachea. PloS one 5, e14134.

Zidar, N., Čavić, M., Kenda, R.B., Koselj, M., Ferluga, D., an, s., 1998. Effect of intrauterine growth retardation on the clinical course and prognosis of IgA glomerulonephritis in children. Nephron 79, 28-32.