8-19-2015 12:00 AM

Determination of the phylogenetic range of Sex combs reduced activity in Drosophila melanogaster

Laura E. Garofalo, *The University of Western Ontario*

Supervisor: Anthony Percival-Smith, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Developmental Biology

DETERMINATION OF THE PHYLOGENETIC RANGE OF SEX COMBS REDUCED ACTIVITY IN *DROSOPHILA MELANOGASTER*

(Thesis format: Monograph)

by

Laura E. Garofalo

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

ABSTRACT

The *homeotic selector* (*Hox*) genes are required for body patterning in bilaterians. Sex combs reduced (SCR) is a HOX protein in *Drosophila melanogaster* with two activities: SCR^{T1} and SCR^{lab} activity required for patterning the prothorax (T1) and labia, respectively. SCR^{T1} is proposed to be conserved throughout bilaterians while the phylogenetic range of functional conservation of SCR^{lab} is comparatively unknown. The goal of this work was to elucidate the evolutionary time point at which SCR activity changed. CRISPR/Cas9 transgenesis was used to incorporate ϕC31 integrase recombination sites in Drosophila *Scr*. The ϕC31 integrase could then be used to replace Drosophila *Scr* with *Scr* orthologs to study their function. Here, two *Scr* specific CRISPR guide sequences and a donor template were created to facilitate CRISPR/Cas9 homology-directed repair. Addtionally, *Scr* orthologs from six phylogenetically diverse species were isolated and incorporated into vectors to facilitate their insertion at *Scr*.

Keywords: *Hox* gene, Sex combs reduced, functional conservation, CRISPR/Cas9, *Drosophila melanogaster*, ortholog, homology-directed repair

CO-AUTHORSHIP STATEMENT

I performed all experimental procedures described in this thesis – excepting the cloning of the donor template – and drafted the thesis. Dr. Anthony-Percival Smith assisted in cloning the donor template, made intellectual contributions to the experimental design, and proofread the thesis. Dr. Kathleen Hill also helped proofread the thesis.

DEDICATION

To my loving parents, John and Avelina Garofalo

ACKNOWLEDGMENTS

This work had been made possible by the contributions of numerous people. First and foremost, I want to thank my supervisor, Dr. Anthony Percival-Smith, for his guidance and support throughout my graduate work. I appreciate all your contributions of time, ideas, and expertise that resulted in a positive graduate experience. Your passion for genetics was both inspirational and motivational. Advice given by my advisors, Dr. Greg Kelly and Dr. Kathleen Hill, has been a great help in the development of this thesis. Thank you Dr. Greg Kelly for sharing your insights on cloning, experimental timelines, and other aspects of this project. Thank you Dr. Kathleen Hill for your helpful suggestions in the editing of this thesis. I also wish to acknowledge the mentorship provided by Dr. Tom Drysdale, Professor of the Developmental Biology Collaborative Program, during the early stages of this project.

I offer special thanks to the graduate students of the APS lab: Alaa Briek, Samantha Koot, and Anirban Banjeree for their help and support that extended into weekends and holidays. Your teamwork has been an integral component in the completion of this work. I also acknowledge the help provided by Stella Sloma, Claire Newman, and all other honors students and volunteers who offered their time to the APS lab. Thanks to all members of the Molecular Genetics Unit for your assistance and cooperation in the core. Outside of the Molecular Genetics Unit, I thank fellow graduate student, Michael Del Vasto, for daily laughs and constant companionship; my graduate experience would have been far less memorable without you.

Finally, I extend my deepest gratitude to parents, sister, and extended family for their support in all my pursuits throughout my life. Mom and Dad, you helped me solidify my decision to pursue graduate studies and serve as a constant reminder of the reason I strive for greatness. A heartfelt thank you to my future husband and dearest friend, Tomasz Paldino, for your never-ending support, faith, and love. Thank you for always believing in me and reminding me to stay positive, look forward, and dream big.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

aa	Amino acid
abd-A	*abdominal-A* (gene)
Abd-B	*Abdominal-B* (gene)
AEL	After egg laying
ANT-C	Antennapedia complex
Antp	*Antennapedia* (gene)
AP	Anterior-posterior
att	Attachment site
bp	Base pair
BX-C	Bithorax complex
Cas	CRISPR-associated
CDS	Coding sequence
chiRNA	Chimeric RNA
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CTD	Carboxy-terminal domain
Dfd	*Deformed* (gene)
DSB	Double-stranded break
dsDNA	Double-stranded DNA
evo-devo	Evolutionary developmental biology
ftz	*Fushi tarazu* (gene)
HD	Homeodomain
HDR	Homology-directed repair
HOM-C	Homeotic complex
Hox	*Homeotic selector* (gene)
lab	*labial* (gene)
NHEJ	Non-homologous end joining
nt	Nucleotide
PAM	Protospacer adjacent motif

pb	*proboscipedia* (gene)
PB	Proboscipedia (protein)
RMCE	Recombinase-mediated cassette exchange
Scr	*Sex combs reduced* (gene)
SCR	Sex combs reduced (protein)
SLiM	Short linear motif
TALEN	Transcription activator-like effector nuclease
tracrRNA	Trans-activating CRISPR RNA
UAS	Upstream activating sequence
Ubx	*Ultrabithorax* (gene)

1 INTRODUCTION

1.1 Evolutionary Developmental Biology

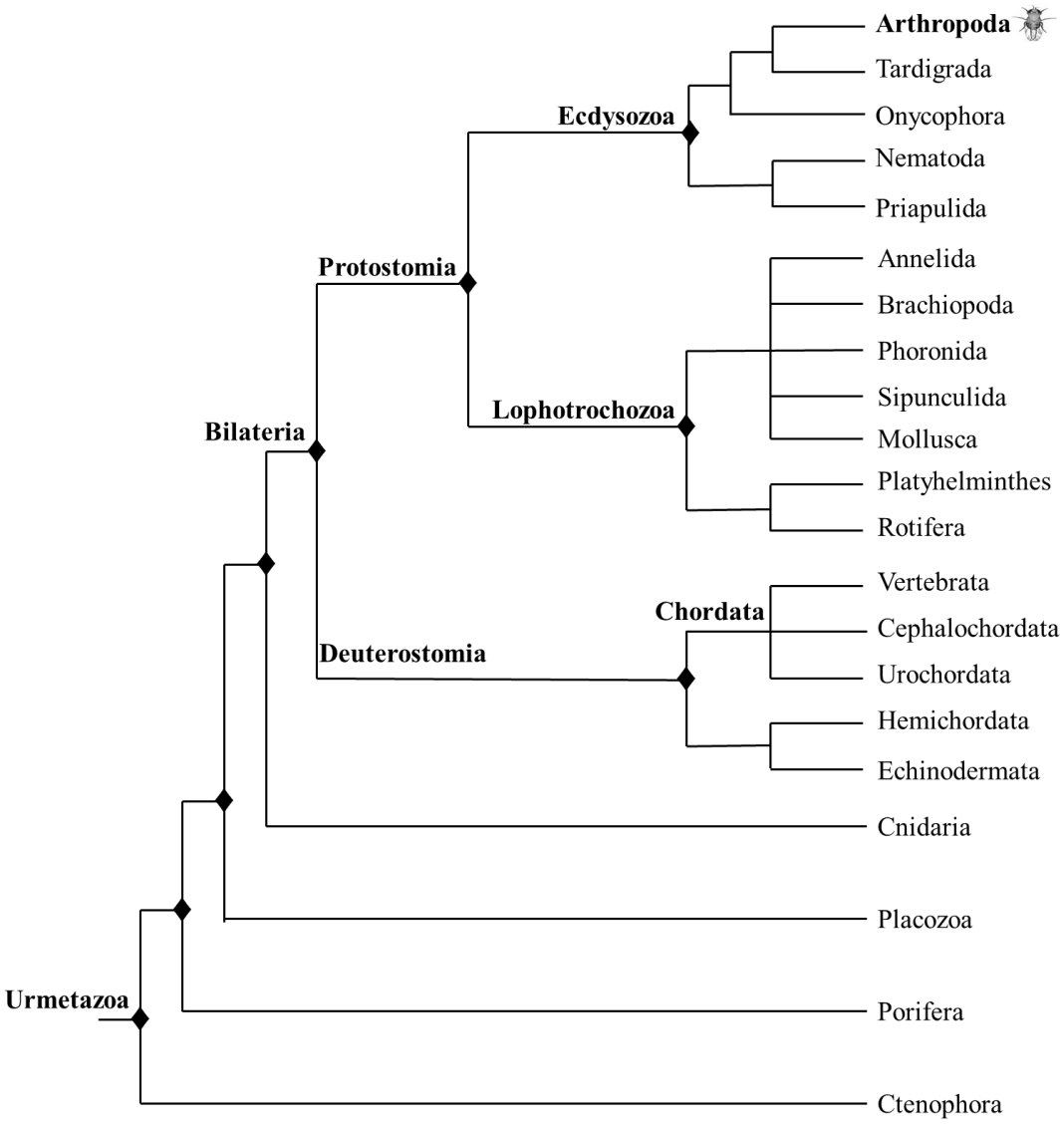
The goal of this work is to determine the phylogenetic range of Sex combs reduced activity in determining the body plan, which places this work centrally in the field of evolutionary developmental biology (evo-devo). All living organisms on Earth descended from the last universal common ancestor that lived 3.5 to 3.8 billion years ago (GLANSDORFF ET AL. 2008). Over evolutionary time, animal diversity increased through modifications to a shared set of developmental genes. The field of evo-devo focuses on comparisons of developmental processes between animals to determine their evolutionary relationship. Evo-devo demonstrates that there is a universal genetic recipe for development and that modifications to this ancient recipe gave rise to novel morphologies and animal diversity.

The earliest attempt to understand evolutionary relationships occurred in the early nineteenth century with the advent of comparative embryology. Morphological comparisons of animals revealed that the animal body is constructed from a shared set of homologous units (GEOFFROY ST. HILLAIRE 1818). Anatomically similar animals were grouped together on the sole basis of their body plan. Later in the mid-nineteenth century, Charles Darwin formulated the scientific theory of evolution and established three key principles: natural selection, heredity, and variation (DARWIN 1859). However, it was not until Gregor Mendel demonstrated the genetic basis of heredity that Darwin's theory could be extended (MENDEL 1866). The modern synthesis merges Darwin's theory of evolution with Mendelian classical genetics. It also integrates ideas from other scientific disciplines, such as taxonomy, biogeography, and embryology. Molecular biology and in particular, genome sequencing allowed new connections to be made between animal phyla once thought to be unrelated based on anatomical comparisons. Molecular biology also led to the identification of important developmental genes and the observation that there is a high degree of sequence similarity between developmental genes. Evo-devo with molecular biology techniques allows for the study of the divergence and conservation of developmental processes.

The accumulation of genomic DNA sequences has produced an accurate picture of the animal tree of life (**FIGURE 1.1**). At the base of the tree of life, is the hypothetical ancestor Urmetazoa, which diverged into five major urmetazoan lineages: Ctenophora (comb jellies), Porifera (sponges), Placozoa (basal invertebrates), Cnidaria (e.g. sea anemones, corals, hydroids, and jellyfish), and Bilateria (RYAN ET AL. 2013). The Porifera and Placozoa lack a true epithelium and symmetry. The Cnidaria and Ctenophora are diploblastic (with two epithelia, lacking mesoderm) and have radial symmetry. The Bilateria are triploblastic (with true endoderm, mesoderm, and ectoderm) and have bilateral symmetry. Bilaterians can be further divided into two groups – the protostomes and deuterostomes – based on the first opening that forms in development. In the protostomes, the first opening (the blastopore) to form develops into the mouth, whereas in deuterostomes, the blastopore develops into the anus. The protostomes encompasses two groups: Lophotrochozoa (soft tissued animals with cilia) and Ecdysozoa (exoskeleton-covered animals that molt), which include *Drosophila melanogaster* and all other arthropods. The deuterostomes include: Vertebrata (vertebrates), Cephalochordata (lancelets, *Amphioxus*), Urochordata (tunicates, ascidians), Hemichordata (e.g. acorn worms), and Echinodermata (e.g. starfish, sea urchins, sand dollars, and sea cucumbers).

All bilaterians descend from a common protostome-deuterostome ancestor, Urbilateria. The presence of conserved developmental genes in the genomes of protostomes and deuterostomes makes it likely that they were also present in Urbilateria. The *Pax6* gene, for example, is required for eye development in protostomes and deuterostomes. Loss-of-function mutations in the *Pax6* homologs – Drosophila *eyeless*, murine *Small eye*, and human *PAX6* – lead to severe eye defects or the absence of eyes altogether in each organism. *PAX6* loss-of-function phenotypes indicate that there is an evolutionary conservation of requirement of PAX6 for eye development. When Drosophila eyeless protein is expressed ectopically, that is, in cells that do not normally express PAX6, ectopic eyes form in non-retinal tissues on the fly (HALDER ET AL. 1995). Likewise, the ectopic expression of murine PAX6 protein leads to the formation of extra eyes in both Drosophila and *Xenopus*, members of the protostomes and deuterostomes, respectively (CHOW ET AL. 1999). It is unlikely that PAX6 evolved its function independently in both

FIGURE 1.1 The animal tree of life. Phylogenetic tree of selected, major taxa that descend from the hypothetical ancestor, Urmetazoa. Five major urmetazoan lineages have been identified: Ctenophora (comb jellies), Porifera (sponges), Placozoa (basal invertebrates), Cnidaria (e.g. sea anemones, corals, hydroids, and jellyfish), and Bilateria (RYAN ET AL. 2013). Bilateria encompass two groups: the protostomes (Protostomia) and the deuterostomes (Deuterostomia). The protostomes are further divided into two groups: Lophotrochozoa (soft tissued animals with cilia) and Ecdysozoa (exoskeleton-covered animals that molt), which includes *Drosophila melanogaster* (common fruit fly) and all other arthropods. The deuterostomes includes the animals of Chordata. The length of lines does not indicate evolutionary time, but rather, is an indication of relatedness.


Arthropoda
Tardigrada
Onycophora
Ecdysozoa
Nematoda
Priapulida
Protostomia
Annelida
Brachiopoda
Phoronida
Sipunculida
Lophotrochozoa
Bilateria
Mollusca
Platyhelminthes
Rotifera
Vertebrata
Chordata
Cephalochordata
Deuterostomia
Urochordata
Hemichordata
Echinodermata
Cnidaria
Placozoa
Urmetazoa
Porifera
Ctenophora

protostomes and deuterostomes as a result of convergent evolution. Instead, *Pax6* was most likely a gene in Urbilateria for eye development and its function has been universally conserved in bilaterians. The *Pax6* gene is only one example of a functionally conserved developmental gene. The function of the *homeotic selector* (*Hox*) genes has also been highly conserved across bilaterians.

The importance of *Hox* genes in bilaterian development is illustrated by the transformations that occur when HOX protein activity is reduced or mis-regulated. Homeotic transformations occur when one body segment is transformed into the likeness of another (BATESON 1894). Calvin Bridges was the first to identify a genetically heritable homeotic transformation in Drosophila: a mutant fly with a second pair of wings, which replaced its halteres (a balancing organ on the third thoracic segment). Bridges termed these four-winged flies, *bithorax* mutants, which possessed mutant alleles of the *Hox* gene, *Ultrabithorax.* The discovery of Bridges' *bithorax* mutant led to the systematic screening for other homeotic mutant flies. It took geneticists nearly half a century to identify eight, linked *Hox* genes in Drosophila (DUNCAN AND KAUFMAN 1975; LEWIS 1978; LEWIS ET AL. 1980A, B). These early observations of homeotic transformations led to the suggestion that these genes have a central role in determining segmental identity during development.

The isolation of the Drosophila *Hox* genes led to the isolation and study of their orthologs from other species (MCGINNIS ET AL. 1984; LAPPIN ET AL. 2006). Functional analyses of *Hox* genes has provided overwhelming evidence to support the fact that *Hox* genes play a critical role in bilaterian body patterning. Moreover, their role in body patterning has been universally conserved across bilaterians. Some *Hox* genes, however, are pleiotropic and may have acquired new functions over evolution. Despite our understanding of *Hox* gene function in many bilaterians, little is known about how HOX proteins with multiple functions evolved new activities while still maintaining conserved functions. Even well less understood is the specific time points in evolution that HOX protein function may have changed. My thesis aims to characterize the functional conservation of the *D. melanogaster* HOX protein, Sex combs reduced, and the evolutionary time point at which its activity changed.

1.2 Drosophila as a Model Organism

The first documented scientific use of Drosophila was in 1901 at a Harvard University laboratory by William E. Castle. However, it was not until 1909 when Thomas Hunt Morgan began to establish Drosophila as a model organism and popularize its use for genetic study after he fortuitously identified a white-eyed mutant fly. His discovery spurred a revolution in our understanding of the mechanistic basis of heredity. At Colombia University, Morgan admitted three principal students to his lab: Alfred H. Sturtevant, Calvin B. Bridges, and Hermann J. Müller. Over the next two decades, Morgan and his team performed controlled Drosophila crosses and counted mutant progeny. Their work led to some of the most influential scientific breakthroughs, including: the creation of the first genetic map of Drosophila by Sturtevant in 1913, the discovery of genetically inheritable homeotic mutants by Bridges in 1915, the introduction of the balancer chromosome by Müller in 1918, and the implementation of the Chromosomal Theory of Inheritance by Morgan. Morgan extended the laws of Mendelian Inheritance and built the foundation for the basic principles of heredity, establishing Drosophila as a key model organism in the study of genetics.

In order for a model organism to be useful for genetic study, it is important to be able to observe several generations with numerous individuals to furnish a reliable basis for conclusions (DEMEREC AND KAUFMAN 1996). Thus, geneticists favor organisms, like Drosophila, that breed rapidly, develop quickly, and are easy to observe. Drosophila has a short generation time of about 10 days from the point of egg laying to eclosion as an adult fly, and therefore several generations can be studied within a few weeks. Female flies are highly fecund and have a large brood size, laying up to 400 eggs each day, allowing for the daily collection of numerous embryos. Finally, Drosophila development begins externally after oviposition, making observations of embryogenesis easy relative to mammals.

Drosophila has a relatively small genome that encodes approximately 13 600 genes; in contrast, to the human genome that is made up of 23 000 genes (ADAMS ET AL. 2000; INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2004). The mammalian genome has undergone repeated duplications resulting in extensive genomic redundancy,

which is not found with the Drosophila genome. The complete Drosophila genome was sequenced and published in 2000 and is publicly available at online databases, such as *FlyBase* (ADAMS ET AL. 2000). Despite a relatively simple genome, Drosophila is a complex, multicellular organism that parallels human development in many aspects. Cross-genomic comparisons of Drosophila and humans have revealed that many of the genetic networks and molecular pathways of development in humans are conserved in flies. In fact, 77% of known human genetic diseases have a recognizable match in the Drosophila genome (REITER ET AL. 2001). Additionally, a portion of these orthologs in Drosophila are tagged by *P*-element insertions, allowing genetic manipulation and further study (REITER ET AL. 2001). Online databases, such as Homophila, are powerful intergenomic resources that link the human and fly genome in order to facilitate the study of human genetic diseases in the fly model (CHIEN ET AL. 2002).

1.3 The Manipulation of the Drosophila Genome

The ability to modify the Drosophila genome efficiently and precisely is essential for sophisticated genetic analysis. Geneticists often rely on random mutagenesis or traditional gene-targeting methods in order to create transgenic flies (WANG ET AL. 2013). *P*-element mediated, germ line transformation was introduced in 1981 as a means to incorporate DNA into Drosophila through gene transfer (SPRADLING AND RUBIN 1982). Modified transposable elements are randomly integrated into the genome, allowing for the engineering of genetically defined fly lines with regulated transgenes and techniques for generating genetic mosaics (RUBIN AND LEWIS 2000; VENKEN AND BELLEN 2007). *P*-element mediated transgenesis, however, is limited by the size of DNA that can be integrated and the inability to control the site of insertion (VENKEN AND BELLEN 2007). The availability of transposon-based transgenesis facilitated the development of an array of effective genetic techniques in Drosophila, many of which have since been adapted in other model systems (RUBIN AND LEWIS 2000). These techniques include the development of enhancer traps for genetic screens of expression patterns in 1987, large-scale insertional mutagenesis with modified transposable elements in 1988, site-specific recombination for creating chromosomal rearrangements in 1989, and the two-component, *GAL4-UAS* system for controlling ectopic gene expression in 1993 (RUBIN AND LEWIS 2000; BISCHOF AND BASLER 2008).

Traditional gene targeting methods in Drosophila are often expensive, time-consuming, labor intensive, and most importantly, inefficient in generating transgenic flies (BIBIKOVA ET AL. 2002; BEUMER ET AL. 2008, 2013; CHRISTIAN ET AL. 2010; CERMAK ET AL. 2011; LIU ET AL. 2012; TREEN ET AL. 2014). Modern alternative methods, like zinc-finger nucleases and transcription activator-like effector nucleases (TALENs), have been developed to accelerate the process of genome modification (LIU ET AL. 2012; CERMAK ET AL. 2013; BEUMER AND CARROLL 2014). These methods enable targeted modification through customizable, sequence-specific DNA nucleases that induce double-stranded breaks (DSBs) at target sites in an organism's DNA, prompting the host to repair it and thereby modify its genome in the process. However, these methods have been proven to have varying efficiency, specificity, and toxicity (WANG ET AL. 2013). Moreover, they entail a complex design process and no simultaneous gene targeting has yet been reported (WANG ET AL. 2013).

A new genome engineering tool, CRISPR (clustered regularly interspaced palindromic repeats), holds promise for efficient, highly specific, targeted genetic modifications to Drosophila (**FIGURE 1.2**). CRISPR is a component of the prokaryotic immune system that confers resistance to exogenous genetic elements (BHAYA ET AL. 2011). In type II CRISPR systems, a CRISPR RNA (crRNA) containing a sequence complementary to that of an invading genetic element, and a *trans*-activating CRISPR RNA (tracrRNA) interact with the CRISPR-associated (Cas) nuclease, Cas9, to direct sequence-specific, double-stranded cleavage of exogenous DNA (**FIGURE 1.3.A**; BHAYA ET AL. 2011). Target site recognition relies solely on the Watson-Crick base pairing between the spacer of crRNA and one strand of target DNA (protospacer), which is immediately followed by a "NGG" tri-nucleotide protospacer adjacent motif (PAM) on the opposite strand (JINEK ET AL. 2012; SEBO ET AL. 2013).

Recently, the RNA-guided nuclease Cas9 has been isolated from *Streptococcus pyogenes* and the fusion of crRNA and tracrRNA created a synthetic, chimeric RNA (chiRNA; **FIGURE 1.3.B**; JINEK ET AL. 2012; BASSETT ET AL. 2013; GRATZ ET AL. 2013). Together, Cas9 and chiRNA make CRISPR a simple, two-component system for creating targeted DSBs. The introduction of a custom chiRNA targeting a gene of interest into a host will

FIGURE 1.2 A Type II CRISPR locus in *Streptococcus pyogenes*. CRISPR (clustered regularly interspaced short palindromic repeats) is a component of the prokaryotic immune system that confers resistance to exogenous genetic elements. It is an acquired immunity, such that short segments of foreign DNA (or "spacers") are incorporated into the bacterial genome between CRISPR repeats, serving as a 'memory' of past exposures. CRISPR spacers are then used to recognize and silence exogenous genetic elements. A type II CRISPR locus, such as that in the bacterium *S. pyogenes*, contains an array of multiple, alternating spacers and short, palindromic direct repeats. The identical repeats range between 21 and 47 bp in different loci; the spacers are of constant length but are hypervariable in sequence, and derived from previously encountered DNA phages or plasmids. The entire array is transcribed as a single mRNA under the direction of a promoter located in the leader sequence. CRISPR-associated (*Cas*) genes – indicated by the grey, pointed boxes – can be of variable size and number. *Cas* genes encode the CAS proteins that add new spacer-repeat pairs, process the CRISPR transcript, and cleave the recognized foreign DNA (MALI ET AL. 2013).

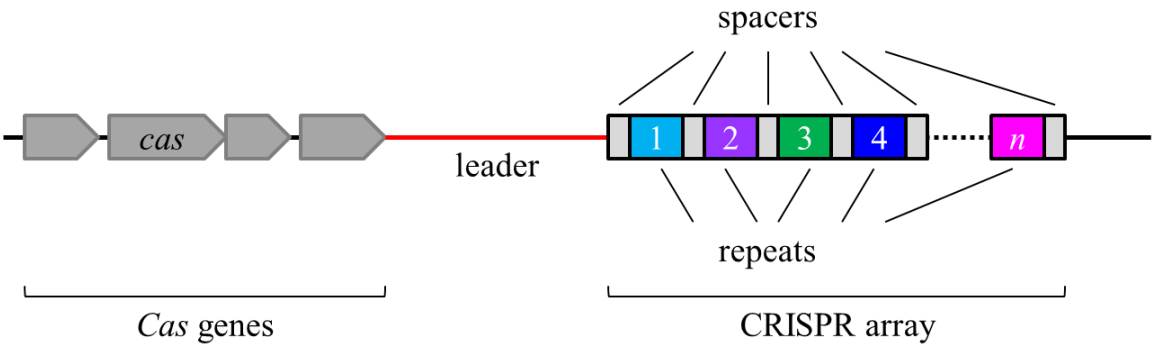
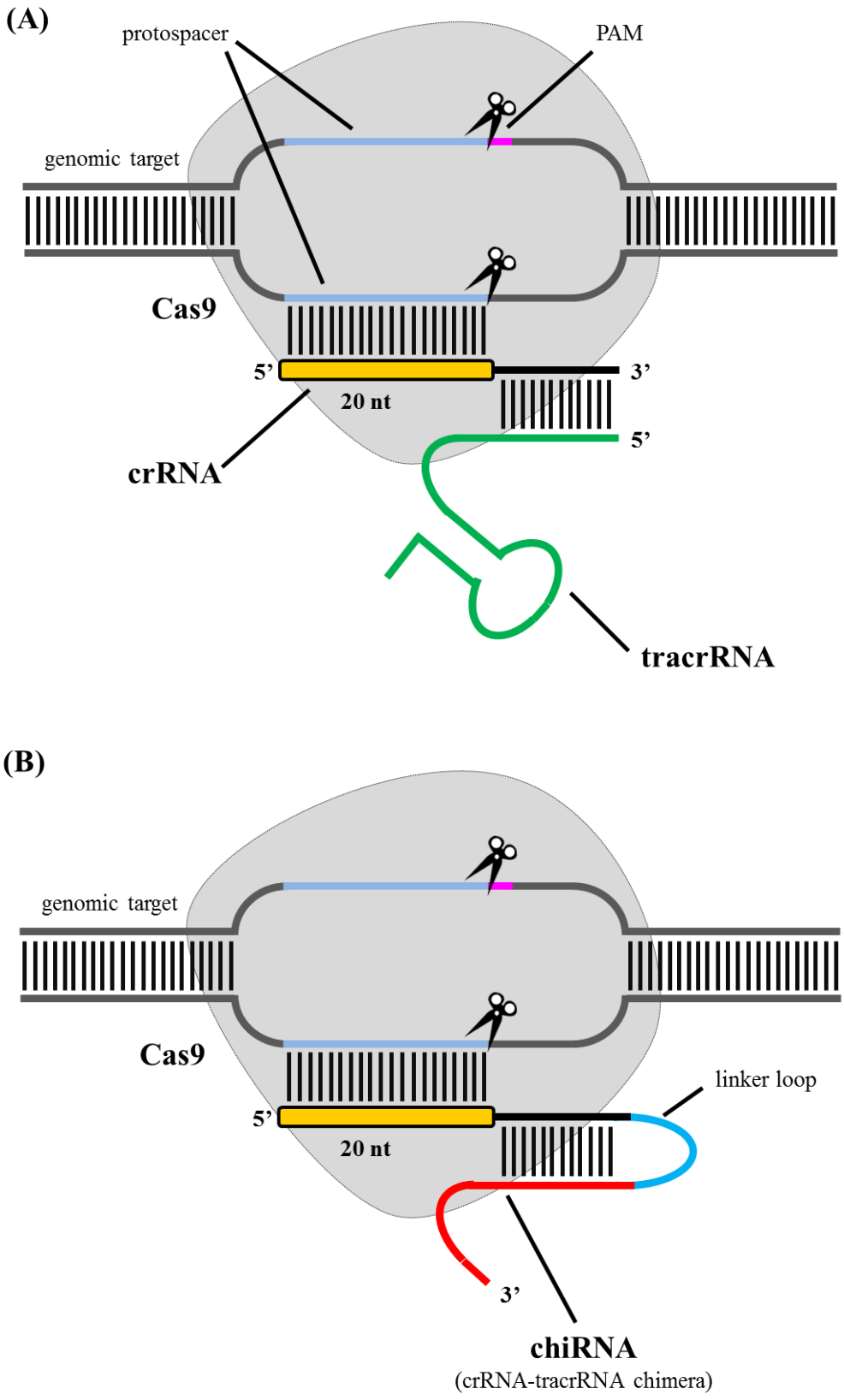

spacers
cas
leader
1
2
3
4
n
repeats
Cas genes
CRISPR array

FIGURE 1.3 The CRISPR/Cas9 system for generating targeted double-stranded breaks in DNA. The CRISPR/Cas9 system is a modern genome engineering tool for generating targeted double-stranded breaks (DSBs) in DNA. CRISPR is a component of the prokaryotic immune system that confers resistance to exogenous genetic elements. (A) In type II CRISPR systems, a CRISPR RNA (crRNA; yellow box) containing a sequence complementary to that of an invading genetic element, and a *trans*-activating CRISPR RNA (tracrRNA; green) interact with the CRISPR-associated (Cas) nuclease, Cas9, to direct sequence-specific, double-stranded cleavage of exogenous DNA. Target site recognition relies solely on the Watson-Crick base pairing between the spacer of crRNA and one strand of target DNA (protospacer; light blue), which is immediately followed by a "NGG" tri-nucleotide protospacer adjacent motif (PAM; pink) on the opposite strand. (B) Recently, the RNA-guided nuclease Cas9 has been isolated from *Streptococcus pyogenes* and the fusion of crRNA and tracrRNA created a synthetic, chimeric RNA (chiRNA; red). Together, Cas9 and chiRNA make CRISPR a simple, two-component system for creating targeted DSBs. The introduction of a custom chiRNA targeting a gene of interest into a host will guide Cas9 to a genomic target and induce a DSB in the host DNA (JINEK ET AL. 2012).

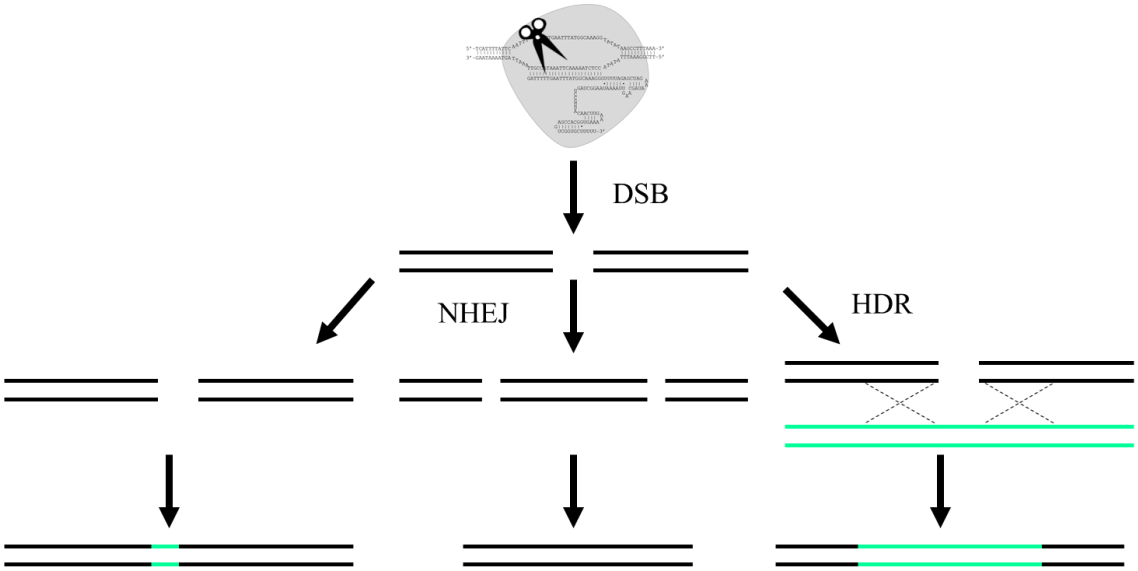

(A)
protospacer
PAM
genomic target
Cas9
5'
3'
20 nt
5'
crRNA
tracrRNA
(B)
genomic target
Cas9
linker loop
5'
20 nt
3'
chiRNA
(crRNA-tracrRNA chimera)

guide Cas9 to a genomic target and induce a DSB in the host DNA (WANG ET AL. 2013). The host's repair machinery responds to the DSB through non-homologous end joining (NHEJ) or homology-directed repair (HDR) (**FIGURE 1.4**; GRATZ ET AL. 2013). In NHEJ, DNA will reconnect from either side of a DSB where there is little to no sequence overlap for annealing (GRATZ ET AL. 2013). The joining of DNA ends often induces errors in the host genome in the form of insertions or deletions (GRATZ ET AL. 2013). In HDR, the host's repair machinery searches for a homologous DNA template and, if present, incorporates it into the genome at the point of the DSB (GRATZ ET AL. 2013). If the homologous DNA used to repair the DNA contains genetic modifications, these modifications will be incorporated at the point of the DSB, thereby editing the genome.

CRISPR/Cas9-mediated HDR can be used with subsequent genetic manipulation, like recombinase-mediated cassette exchange (RMCE). This combination of genetic strategies holds great potential for studying the functional conservation of developmental genes in Drosophila.

RMCE enables the repeated incorporation of large transgenic constructs at a single position in the genome (GROTH ET AL. 2004). In Drosophila, the recombinase, ϕC31 integrase, can catalyze the site-specific recombination of DNA between two non-identical recognition sites, *attP* and *attB* (**FIGURE 1.5**; GROTH ET AL. 2004; BISCHOF ET AL. 2007). These *att* recognition sites are short genetic motifs with partial inverted-repeat symmetry that flank a central crossover sequence at which synapsis occurs. The ϕC31 integrase catalyzes the recombination of *attB*-containing vectors into *attP*-containing genomic targets that have been introduced previously into a genome, by transgenic techniques like CRISPR (GROTH ET AL. 2004). The reciprocal exchange of a genetic cassette in a donor vector for an endogenous cassette in a genome occurs in a unidirectional manner through an energy-independent transesterification reaction (GROTH ET AL. 2004). A gene of interest can be designed and genetically modified in Drosophila with CRISPR/Cas9-mediated HDR to contain *attP* sites outside its coding sequence. The ϕC31 integrase can then facilitate the exchange of a genetic ortholog for the endogenous coding sequence at the *attP* sites (BATEMAN ET AL. 2006). This places the ortholog under the control of endogenous regulatory sequences, allowing the *in vivo* study of ortholog function.

FIGURE 1.4 Double-stranded breaks in DNA can be used to incorporate defined genomic modifications. Double-stranded breaks (DSBs) in DNA induced by the CRISPR/Cas9 system can be repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR). The repair of DSBs in DNA by NHEJ can result in small insertions or deletions at the target site (left) or deletion/inversion of large genomic regions when two DSBs occur (middle). The repair of DSBs in DNA can result in HDR if a donor template is present (right). HDR can result in a variety of genomic modifications (bottom; BASSETT AND LIU 2014).

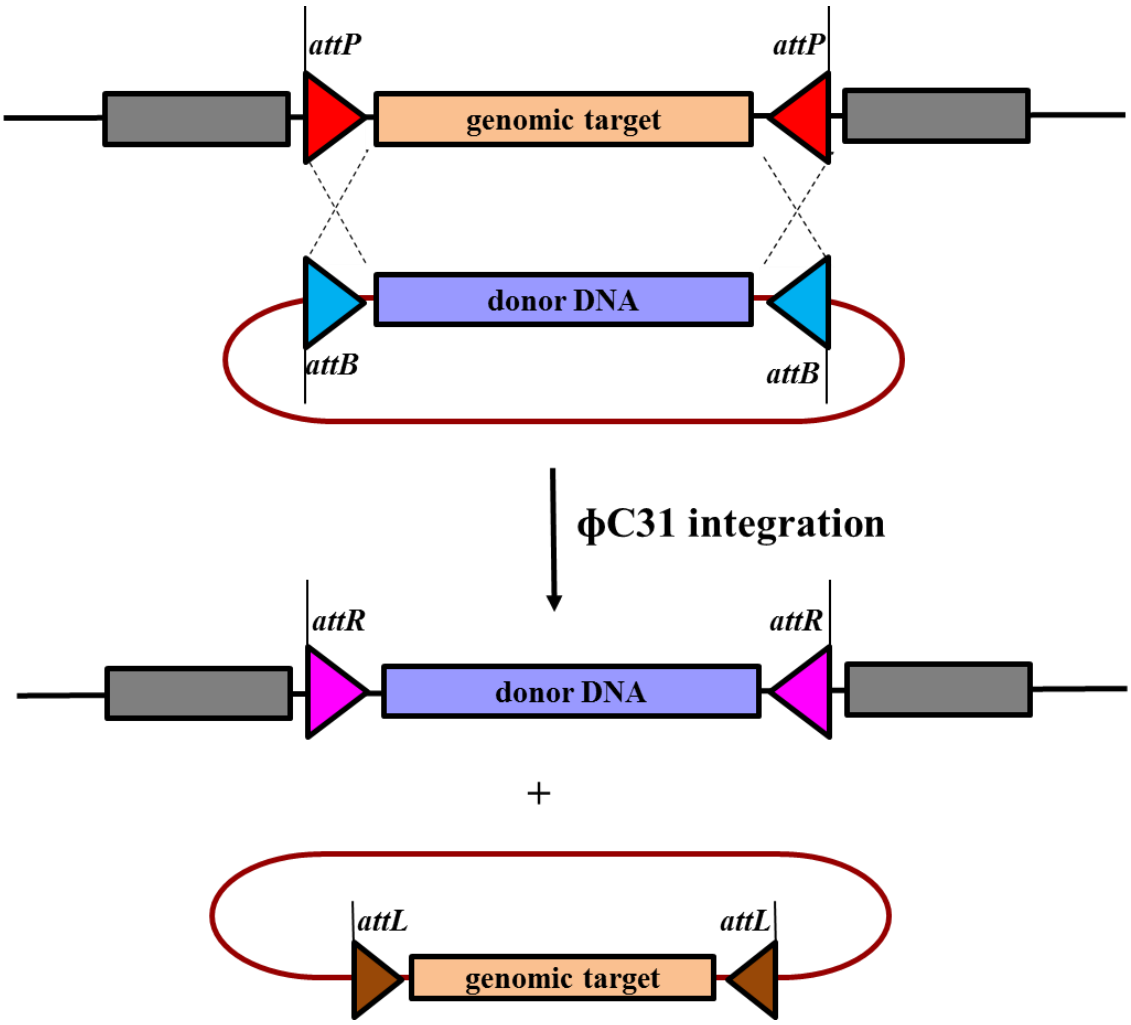


Single cut
Short deletions (1-100 nt)
Short insertions (1-20 nt)

Double cut
Long deletions (1 kb +)
Inversions

Homologous recombination
Gene replacements
Gene deletions
Protein or RNA tagging

FIGURE 1.5 Schematic of ϕC31 integrase-mediated cassette exchange. ΦC31 integrase-mediated cassette exchange enables the repeated, unidirectional incorporation of large transgenic constructs at a single position in the genome. The recombinase, ϕC31 integrase, can accurately catalyze the exchange of DNA cassettes between recombination sites (triangles). The ϕC31 integrase catalyzes the recombination of *attB*-containing vectors (*attB* sites, blue trianges; donor DNA, purple rectange) into *attP*-containing genomic targets (*attP* sites, red trianges; donor DNA, light orange rectange) that have been introduced previously into a host (e.g. Drosophila), by transgenic techniques like CRISPR. The *att* sites are modified during cassette exchange and after recombination takes place are called *attR* (pink trianges) in the host and *attL* (brown trianges) in the donor vector.


attP
attP
genomic target
donor DNA
attB
attB
ϕC31 integration
attR
attR
donor DNA
+
attL
attL
genomic target

1.4 The Development of *Drosophila melanogaster*

D. melanogaster is a holometabolous insect (i.e. undergoes complete metamorphosis) that has a life cycle divided into four distinct stages: egg, larva, pupa, and imago. Drosophila development involves a complex series of developmental stages and processes that begin after fertilization. Females are fertilized internally and lay eggs externally on the surface of fermenting fruit or other decaying organic matter. The fertilized fly egg gives rise to a segmented, fully-differentiated larva over the course of a 24 hour embryonic period. Embryonic development initiates following fertilization and the fusion of gametes to form the zygote. Embryonic nuclei undergo several rounds of rapid synchronous division in the centre of the embryo to form a multinucleated cell, or syncytium. By the end of the eighth nuclear division, most nuclei have migrated to the periphery of the embryo. However, a small group of about five to fifteen nuclei move towards the posterior pole of the embryo where they form the pole cells, which develop into the adult germ line. At this point, the embryo is referred to as a syncytial blastoderm because the somatic nuclei are still contained within a common plasma membrane. The accumulation of about 6 000 nuclei in the embryonic cytoplasm initiates the formation of the cellular blastoderm at the end of the thirteenth and final nuclear division. The plasma membrane begins slowly to invaginate, dividing the syncytium into individual somatic cells through the process of cellularization at 2:50 h after egg laying (AEL).

Gastrulation commences at 3 h AEL with the segregation of the presumptive germ layers – endoderm, mesoderm, and ectoderm – through a series of invaginations. Germ band elongation begins with the invagination of the posterior midgut primordium. Germ band elongation involves multiple cellular rearrangements and the redistribution of the three germ layer tissues. As the germ band elongates posteriorly, it wraps around the posterior end of the embryo such that the most posterior structures lie adjacent and dorsal to the cephalic primordia (RILEY ET AL. 1987). Further invagination of the posterior midgut forms the ventral furrow and brings mesodermal and endodermal primordia into the interior of the embryo (SWEETON ET AL. 1991). The pole cells are also internalized at this time during a separate invagination event. When the germ band reaches its full extension, the embryo undergoes the process of segmentation where it is divided into imaginal discs

(RILEY ET AL. 1987). The imaginal discs are groups of primordial cells that give rise to adult structures during pupal metamorphosis.

At 7:20 h AEL, the germ band begins to retract and shorten, moving posterior segments to their final position at the posterior end of the developing embryo. Cellular movements during germ band shortening define grooves of the embryonic segments. The germ band completes its contraction at 9:40 to 10:20 h AEL as the labial segments begin to migrate anteriorly initiating the process of head involution (MAHAFFEY AND KAUFMAN 1987). The remaining stages of embryonic development involve a series of morphogenetic movements that include the internalization of ectoderm-derived nervous system and organogenesis of the mesoderm.

The hatching of the first instar larva at 24 h AEL marks the completion of embryogenesis. Over the course of a week, the larva will undergo two molts progressing through three instar larval stages. The larva will pupate within its third and last larval skin and undergo metamorphosis, whereby adult structures form from the cells in the imaginal discs. Metamorphosis culminates at 10 days AEL with the eclosion of an adult fly, or imago. The entire process of Drosophila development, from the initial nuclear divisions to eclosion as an adult fly, is governed by a tightly controlled genetic network.

1.5 Gene Expression during Drosophila Embryogenesis

Bilaterians share a common body plan, which is often composed of repeated, metameric units. The Drosophila body plan is established during early embryogenesis through the hierarchical expression of five classes of genes. Two segmental registers exist in the developing fly embryo: the parasegmental register that is first visible early in embryogenesis during gastrulation and, the segmental register that is visible late in embryogenesis during germ band extension/retraction and later in the larva and imago. Parasegments mark the posterior compartment of one segment and the anterior compartment of the next segment (RILEY ET AL. 1987). Gene expression within the parasegments define the segments during the process of segmentation in development (MARTINEZ-ARIAS AND LAWRENCE 1985). The completion of segmentation in Drosophila results in a larval and adult fly partitioned into fifteen segments, each with unique

identity: three head segments – mandibular, maxillary, and labial – three thoracic segments, and nine abdominal segments.

Extensive genetic and molecular analyses of Drosophila embryogenesis has led to one of the best understood examples of a complex cascade of transcriptional regulation during development (TOMANCAK ET AL. 2007). Our initial understanding of Drosophila development is based on the isolation and characterization of developmental mutants by three Nobel Prize-winning scientists, Christiane Nüsslein-Volhard, Eric Wieschaus, and Ed Lewis (NÜSSLEIN-VOLHARD 1979; NÜSSLEIN-VOLHARD AND WIESCHAUS 1980; ANDERSON AND NÜSSLEIN-VOLHARD 1984; JÜRGENS ET AL. 1984; NÜSSLEIN-VOLHARD ET AL. 1984; WIESCHAUS ET AL. 1984). Nüsslein-Volhard and Wieschaus studied early embryogenesis while Lewis focused his efforts on late embryogenesis. Nüsslein-Volhard and Wieschaus attempted to ambitiously identify every gene required for early Drosophila body patterning by classifying phenotypes of recessive embryonic lethal mutations. This initiated the discovery that embryonic development of Drosophila is orchestrated by the hierarchical, regulated expression of five classes of genes: the first four classes of genes – maternal-effect, gap, pair-rule, and segment-polarity genes – establish segmentation, while the last class – *Hox* genes – determines the identity of each segment.

The genetic cascade that controls segmentation in the developing fly begins during oogenesis in the ovaries of the maternal fly. During oogenesis, the maternal-effect genes are transcribed and their mRNA is stored in the developing egg. After fertilization, maternal-effect mRNA is translated and begins to pattern the embryo. Some maternal-effect mRNAs encode morphogens that are expressed as protein gradients in the syncytial blastoderm. These morphogenetic gradients generate the AP and dorsal-ventral coordinates of the embryo. Loss-of-function mutations in maternal-effect genes produce malformed embryos with a defective anterior or posterior end. For example, *bicoid* mutant mothers give rise to embryos that lack anterior head and thorax structures, which are instead replaced by inverted posterior structures (DRIEVER AND NÜSSLEIN-VOLHARD 1988).

The maternal-effect genes activate the first class of zygotically transcribed genes, the gap genes. Expression of the gap genes transitions the embryo from one characterized by a basic system of coordinates to one with differential expression along the AP axis. The gap genes encode transcription factors that are expressed in wide, overlapping domains along the AP axis of the embryo. The gap proteins divide the embryo into broad domains – anterior, middle, and posterior – that encompass the progenitors of several contiguous segments. Mutations in gap genes lead to a phenotype in which the embryonic body plan has a gap due to missing adjacent segments. The gap gene *Krüppel*, for example, is centrally expressed in the embryo in the thoracic and abdominal segments. Loss-of function *Krüppel* mutants lack these regions, creating gaps in the developing embryonic body plan (PREISS ET AL. 1985).

The gap proteins activate of the next class of zygotic genes, the pair-rule genes. The pair-rule genes encode transcription factors that divide the embryo into periodic units, called parasegments. The pair-rule genes are expressed in an ON/OFF pattern of seven bands of cells along the AP axis and these ON/OFF patterns of pair-rule expression establish the 14 parasegments. Mutations in pair-rule genes, like *fushi tarazu* and *evenskipped*, delete portions of alternate segments resulting in an embryo with half the number of denticle bands (HUGHES AND KRAUSE 2001).

Pair-rule proteins control the transcription of the segment-polarity genes. Segment-polarity proteins generate AP polarity within each segment by defining fourteen parasegmental boundaries. Parasegments encompass the primordial cells of the posterior part of one segment and the cells of the anterior part of the next (MARTINEZ-ARIAS AND LAWRENCE 1985). Mutations in the segment-polarity genes, like *engrailed* and *gooseberry*, result in a mirror-image segmental transformation: either the anterior or posterior half is duplicated within each segment. *Engrailed* and *gooseberry* mutants are characterized by having the posterior part of each segment replaced by duplications of the anterior region of the adjacent segment (GILBERT 2000).

The three classes of zygotic segmentation genes – gap, pair-rule, and segment-polarity – collectively define a linear series of metameric units by using the maternal-effect morphogenetic gradients set up in the early-cleavage embryo. Together, the

segmentation genes control the expression of the final class of developmental genes, the *Hox* genes (INGHAM AND MARTINEZ-ARIAS 1992). *Hox* genes are the master regulators of body patterning, specifically for establishing segmental identity (MCGINNIS AND KRUMLAUF 1992). *Hox* genes encode transcription factors that control the expression of a subset of developmental genes in the primordia of each segment (MCGINNIS AND KRUMLAUF 1992). The expression of developmental genes in each segment results in the development of specific anatomical structures that define segmental identity (MCGINNIS AND KRUMLAUF 1992). The importance of *Hox* genes in development is illustrated by the phenotypes of homeotic transformations. *Hox* mutant alleles induce homeotic transformations in which one segment is transformed into the likeness of another. An *Antennapedia* gain-of-function mutant, for example, develops legs on its head in place of antennae; whereas an *Antennapedia* loss-of-function mutant develops ectopic antennae in place of its second leg pair (STRUHL 1981; FRISCHER ET AL. 1986; SCHNEUWLY ET AL. 1987).

Molecular and genetic analysis of Drosophila has led to a deep understanding of its developmental processes. Moreover, it has provided evidence that developmental processes have been highly conserved in many other animals, including humans. All animals appear to share a fundamental genetic recipe for body patterning that has been conserved for hundreds of millions of years.

1.6 Toolkit Genes and the Animal Body Plan

Embryonic body plan formation is evolutionarily conserved across bilaterally symmetric animals. Within the kingdom Animalia, bilaterians are distinguished by their bilateral symmetry and presence of differentiated cell types that are derived from the three germ layers during ontogeny. These characteristic traits are shared by many bilaterian animals, including the phyla of Chordata, Nematoda, and Arthropoda (BABENKO AND KRYLOV 2004). Each phylum is characterized by a unique body plan and shared set of morphologies, like the number and pattern of body segments. Bilaterian morphologies diversified at the advent of animal evolution, around 600 million years ago, through modifications to developmental genetic networks (KNOLL AND CARROLL 1999). All bilaterian genomes contain toolkit genes that determine the overall body plan and the

number, identity, and pattern of body parts (CARROLL ET AL. 2005). Toolkit genes can be identified based on four characteristics: they comprise a small fraction of the genome, they encode transcription factors or components of signaling pathways, their spatial and temporal expression correlates with the region in which they function, and they are highly conserved across bilaterians (CARROLL ET AL. 2005).

Toolkit genes can be classified into two families based on the proteins that they encode: transcription factors that regulate the expression of genes during embryogenesis or components of signaling pathways that mediate intercellular interactions (CARROLL 2005). Although bilaterians are remarkably divergent in form, toolkit genes are often structurally and functionally conserved (CARROLL 2005; HEFFER ET AL. 2010). This is largely due to the fact that significant changes in toolkit genes would be detrimental to development. Most evolutionary change between species is the result of mutations with minimal or no functional consequence (COPPER AND BROWN 2008). Mutations in coding or functional elements (e.g. exons, *cis*-regulatory elements) are likely to impair function, be deleterious to the organism, and subsequently be eliminated by purifying selection (COPPER AND BROWN 2008).

Although toolkit gene structure and function is highly conserved in bilaterians, toolkit gene expression pattern can vary greatly between or within a taxa. Changes in the expression pattern of toolkit genes, such as segmentation genes, is one way to explain the diversity of animal body plans observed in nature. The expression pattern of the segmentation gene, *fushi tarazu* (*ftz*), for example, has changed over the course of insect evolution. In Drosophila, *ftz* is expressed in a seven-striped pattern along the AP axis of the embryo. A similar expression pattern is observed in other holometabolous insects, like the red flour beetle *Tribolium castaneum*. Tribolium *ftz* is expressed in a pair-rule fashion, although it is expressed in a different register than Drosophila *ftz* (HEFFER ET AL. 2010). Another insect, the desert locust *Schistocerca gregaria*, lacks a striped pattern of expression entirely (HEFFER ET AL. 2010). Other segmentation genes, like *engrailed*, also have varying expression patterns in bilaterians. *Engrailed* is expressed in a series of transversal stripes in arthropods and vertebrates (MINELLI AND FUSCO 2004). In Drosophila, *engrailed* is expressed in the posterior portion of ectoderm-derived

metameres but in vertebrates, the *engrailed* ortholog is expressed in mesoderm-derived metameres (MINELLI AND FUSCO 2004).

Another way to explain the diverse array of animal body plans is through the duplication of toolkit genes followed by divergence (HEFFER ET AL. 2010). *Hox* genes, for example, have undergone multiple duplication events that have generated *Hox* clusters in early bilateral organisms (HEFFER ET AL. 2010). Duplication events allowed genes to diverge, either through the partitioning of existing functions or the acquisition of novel functions (HEFFER ET AL. 2010). One gene that has acquired a novel function is the Drosophila *ftz* gene, which has shifted function from a *Hox* gene to a pair-rule gene over evolutionary time (ALONSO ET AL. 2001; LÖHR ET AL. 2001). This functional change is attributed to a relaxation of constraints due to a functional overlap between *ftz* and the *Hox* genes, *Antennapedia* (*Antp*) or *Sex combs reduced* (HEFFER ET AL. 2010). This idea is supported by the finding that the expression of *ftz* insect orthologs induces *Antp*-like transformations in Drosophila (LÖHR ET AL. 2001). Moreover, nucleotide sequence alignments suggest that *ftz* and *Antp* are closely related (TELFORD 2005).

Comparative genetic analyses of developmental orthologs from vertebrate and invertebrate species has revealed that many genes of high order processes, like development, show a remarkably high degree of conservation. The *Hox* genes are a set of highly conserved toolkit genes, exhibiting a high degree of sequence and functional conservation across bilaterians. Cross-species comparisons of *Hox* orthologs provide evidence of a genetic program that is common between arthropods and vertebrates (MINELLI AND FUSCO 2004).

1.7 The *homeotic selector* genes

An evolutionarily conserved genetic strategy exists to coordinate bilaterian body plan patterning during embryogenesis. Although there are numerous bilaterian body plans, all bilaterians share symmetry along the AP axis that is patterned by a group of toolkit genes, the *homeotic selector* (*Hox*) genes (HEFFER ET AL. 2010). *Hox* genes are the master regulators of body patterning, specifically for establishing segmental identity in all bilaterian animals. Comparative analyses of *Hox* orthologs have weakened the argument

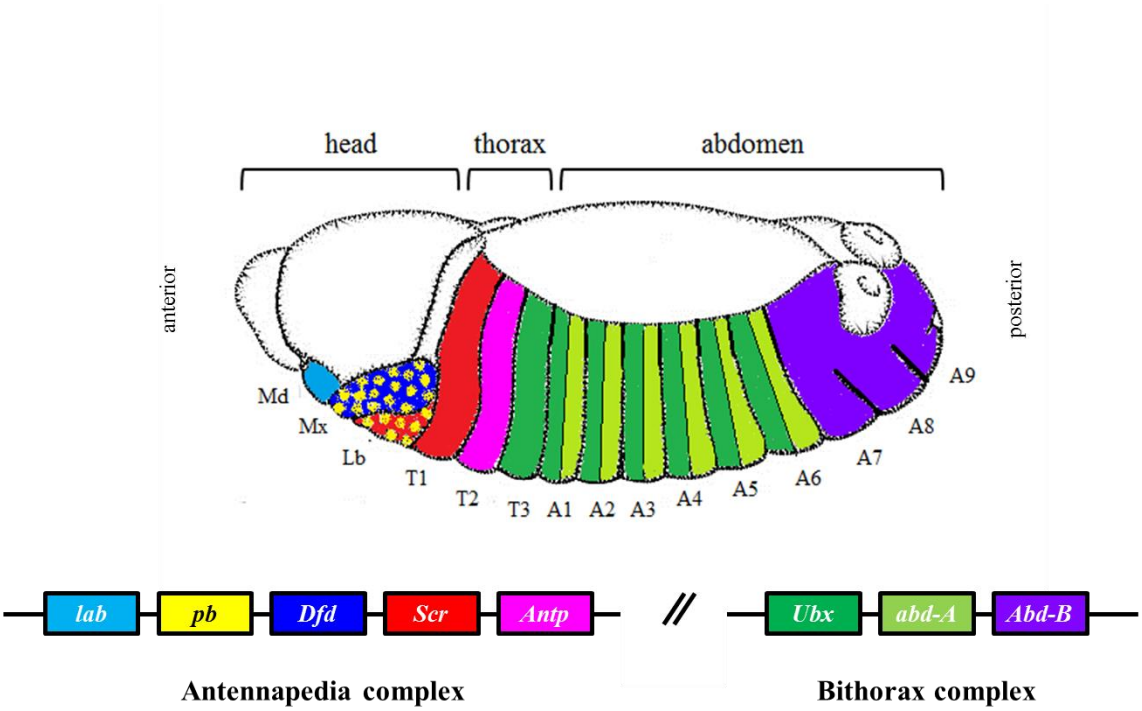
that bilaterian segmentation arose from convergent evolution. Instead, there appears to be a fundamental unity in the genetic control of segmentation in all bilaterians.

All *Hox* genes encode transcription factors that pattern the AP axis during embryonic development. HOX transcription factors regulate the expression of specific developmental genes required for determining segmental identity. Cells require positional information to ensure that naïve progenitor cells differentiate into tissues apposite to their location within the developing embryo. The expression of a specific combination of *Hox* genes within embryonic segments is thought to control the development of primordia into segment-specific structures (LAPPIN ET AL. 2006). As a key player in bilaterian segmentation and development, *Hox* genes exhibit four levels of conservation: conservation of structure, expression, requirement, and function.

All *Hox* genes contain a conserved homeobox sequence that encodes a 60 amino acid residue domain, known as the homeodomain (HD; MCGINNIS AND KRUMLAUF 1992). The HD is a DNA-binding domain of HOX transcription factors that controls gene expression during development. In Drosophila, the amino acid sequence of HOX HDs is highly conserved, with some variation between the eight HOX proteins. The homeobox sequence is also present in the genomes of vertebrates, including humans. The HD has been so well conserved throughout evolution that individual HOX proteins have been found to exhibit greater similarity to the corresponding HOX protein in another species than to HOX proteins encoded by adjacent genes within the same cluster of the same species. The sequence similarity between the *Hox* genes in different species has been attributed to the presence and importance of the conserved homeobox, which determines the specificity and function of each HOX protein. Thus, the conservation of *Hox* gene sequence and structure is directly related to the function of the encoded HOX protein (MCGINNIS ET AL. 1984).

In Drosophila, there are eight *Hox* genes that reside on the right arm of the third chromosome, organized into two gene clusters – the Antennapedia complex (*ANT-C*) and the Bithorax complex (*BX-C*) – that together constitute the homeotic complex (*HOM-C*; **FIGURE 1.6**). The *ANT-C* houses five *Hox* genes – *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), and *Antennapedia* (*Antp*) – that collectively

FIGURE 1.6 The colinear expression of *Hox* genes in the developing Drosophila embryo. *Hox* genes in *Drosophila melanogaster* are organized into two gene clusters: the Antennapedia complex (*ANT-C*) and the Bithorax complex (*BX-C*). The *ANT-C* houses five *Hox* genes – *labial* (*lab*; light blue), *proboscipedia* (*pb*; yellow), *Deformed* (*Dfd*; dark blue), *Sex combs reduced* (*Scr*; red), and *Antennapedia* (*Antp*; pink) – that collectively pattern the anterior segments of the fly. The *BX-C* houses three *Hox* genes – *Ultrabithorax* (*Ubx*; dark green), *abdominal-A* (*abd-A*; light green), and *Abdominal-B* (*Abd-B*) – that are responsible for specifying segmental identity in the posterior thoracic and abdominal segments. *Hox* genes are expressed in spatially restricted domains in the developing Drosophila embryo. The Drosophila embryo is divided into three head segments – mandibular (Md), maxillary (Mx) and labial (Lb) – three thoracic segments (T1-T3), and nine abdominal segments (A1-A9). The order of *Hox* genes within a gene cluster is directly related to their order of expression in the developing embryo along its anterior-posterior axis. This is termed "colinearity". Thus, 3' genes, like *labial*, are expressed more anteriorly and earlier than downstream genes, like *Abd-B.* Speckled regions in the head denote the co-expression of *pb* with *Dfd* in the maxillary segment and of *pb* with *Scr* in the labial segment. Stripped segments in the abdomen denote the overlap of *Ubx* and *abd-A* expression.


head
thorax
abdomen
anterior
posterior
Md
Mx
Lb
T1
T2
T3
A1
A2
A3
A4
A5
A6
A7
A8
A9
lab
pb
Dfd
Scr
Antp
Ubx
abd-A
Abd-B
Antennapedia complex
Bithorax complex

pattern the anterior segments of the fly. The *BX-C* houses three *Hox* genes – *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*) – that are responsible for specifying segmental identity in the posterior thoracic and abdominal segments (LEWIS 1978; KAUFMAN ET AL. 1980).

The arrangement of *Hox* genes in clusters is not unique to Drosophila, and in fact, appears to be a general rule in almost all animals. The vertebrate counterparts of the *HOM-C* genes are also concentrated in clusters: 39 *Hox* genes organized in four paralogous gene clusters (A-D) on four separate chromosomes. These *Hox* gene clusters presumably arose by duplication and divergence from a primordial *Hox* cluster (LAPPIN ET AL. 2006). Each cluster consists of 13 paralog groups that have been assigned on the basis of sequence similarity and relative position within the cluster (LAPPIN ET AL. 2006). Over evolutionary time, *Hox* gene clusters of various species, including those within the genus Drosophila, have undergone some degree of rearrangement. However, it appears that the order of *Hox* genes within a given complex is still conserved and gene expression remains largely unaffected, indicating that regulatory sequences have been preserved within each gene despite the structural reorganization (NEGRE ET AL. 2005).

The order of *Hox* genes within a gene cluster is directly related to their order of expression in the developing embryo along the AP axis. The relationship between the chromosomal arrangement of *Hox* genes and the order of their expression is termed 'colinerarity' and is conserved in bilaterians (GAUNT 1988; DUBOULE AND DOLLÉ 1989; GRAHAM ET AL. 1989; IZPISUA-BELMONTE ET AL. 1991). Thus, 3' *Hox* genes, like *labial/Hox1*, are expressed more anteriorly and earlier than downstream genes, like *Abd-B/Hox9-13* in Drosophila and vertebrates, respectively (GAUNT 1988; DUBOULE AND DOLLÉ 1989; GRAHAM ET AL. 1989; IZPISUA-BELMONTE ET AL. 1991). Some animals exhibit slight deviations from this general pattern, but the function of these genes has remained unchanged (LAPPIN ET AL. 2006). In the tunicate *Oikopleura dioica*, for example, *Hox* genes are unclustered and completely isolated from one another, yet they still exhibit colinear expression (SEO ET AL. 2004).

Hox genes also exhibit conservation of requirement. *Hox* mutant alleles induce homeotic transformations in which the identity of one segment is transformed into the likeness of

another. In Drosophila, the *Antp* mutant alleles trigger an antenna to leg transformation, whereas *Ubx* mutant alleles cause the haltere to transform into a part of the wing (WAKIMOTO 1984). Mutations in vertebrate *Hox* genes can also induce homeotic transformations similar to that observed in Drosophila. *HoxA11* (Drosophila *Abd-B* ortholog) mutant mice exhibit two transformations: a posteriorized thirteenth thoracic segment that has an extra lumbar vertebra, and an anteriorized sacrum that also has an extra lumbar vertebra (SMALL AND POTTER 1993).

Lastly, the function of HOX proteins in evolutionarily distant species has been highly conserved. Comparisons of *Hox* orthologs in Drosophila ectopic expression studies have demonstrated that *Hox* orthologs are functionally interchangeable between species. Many *Hox* orthologs have been shown to functionally complement their Drosophila counterparts, eliciting a comparable segmental transformation after ectopic expression. For example, the ectopic expression of onychophoran *Ubx* in Drosophila induces an antenna-to-leg and wing-to-haltere transformation; the same homeotic transformation observed with the misexpression of Drosophila *Ubx* (GALANT AND CARROLL 2002). The ectopic expression of vertebrate *Hox* orthologs can also reproduce similar phenotypes as observed with the ectopic expression of Drosophila *Hox* genes. For example, the ectopic expression of human *HOXB4* produces a phenotype similar to the misexpression of *Dfd* in Drosophila, indicating that it maintains many of the same regulatory and developmental functions (MCGINNIS ET AL. 1990). This signifies that a *Hox* ortholog can perform some of the same molecular and developmental functions as its Drosophila complement (GALANT AND CARROLL 2002). In fact, vertebrate *Hox* orthologs have been shown to rescue the development of *Hox* mutant flies. The human *HOXB4* gene, ortholog of Drosophila *Dfd*, is required for hindbrain development. Although flies lack the mammalian hindbrain, *HOXB4* can rescue the development of posterior head structures in *Dfd*-deficient flies that lack these structures (MALICKI ET AL. 1992).

Hox genes are a key member of the genetic toolkit of development. *Hox* genes play a crucial role in determining segmental identity along the AP axis of the developing embryo. As a result, many aspects related to *Hox* gene function have been widely conserved among different animal phyla. Functional comparisons of *Hox* orthologs

provide a means to understand the evolutionary relationship between invertebrates and vertebrates, and specifically the time points at which *Hox* gene function changed during bilaterian evolution.

1.8 The *Hox* gene, *Sex combs reduced*

The fourth *Hox* gene of the *ANT-C* in Drosophila is *Sex combs reduced* (*Scr*). *Scr* spans a DNA segment of over 70 kilobases – including three exons and two introns – proximal to the *Antp* locus (LEMOTTE ET AL. 1989). *Scr*, like all *Hox* genes, encodes a transcription factor that directly binds DNA to regulate the transcription of a specific set of developmental genes. Analyses of homeotic transformations in *D. melanogaster* show that the *Scr* is required for the determination of the labial and prothoracic segments in the larval and adult fly (STRUHL 1982; PATTATUCCI ET AL. 1991; PEDERSON ET AL. 1996). Drosophila SCR protein is proposed to have two activities: SCR^{T1} activity required for larval and adult prothorax (T1) structure and salivary gland development and SCR^{lab} activity required for proboscis development (PERCIVAL-SMITH ET AL. 2013). In the prothoracic segment, SCR is required for the formation of a full T1 larval beard during embryogenesis and later, patterns the T1 body wall, first leg bristles, and male sex comb on the fifth tarsal segment (WAKIMOTO AND KAUFMAN 1981; SATO ET AL. 1985). In the labial segment, SCR is required for the development of salivary glands during embryogenesis and later, during metamorphosis, SCR, in combination with Proboscipedia (PB), is required for the formation of the proboscis, the Drosophila feeding tube and labial segment derivative (PERCIVAL-SMITH ET AL. 1997; PERCIVAL-SMITH 2013). The dual requirement of SCR and PB to determine the adult proboscis is conserved in species of three insect orders: *D. melanogaster*, *Tribolium castaneum* and *Oncopeltus fasciatus* (BEEMAN ET AL. 1989; PERCIVAL-SMITH ET AL. 1997; HUGHES AND KAUFMAN 2000).

The phenotypes of loss- and gain-of-function *Scr* alleles illustrate the essential and pleiotropic role of *Scr* in the determination of labial and prothoracic identity. Null *Scr* alleles are embryonic lethal when homozygous and exhibit homeotic transformations of the labial segment to maxillary identity and the T1 segment to mesothorax (T2) identity (MAHAFFEY AND KAUFMAN 1987; RILEY ET AL. 1987; GLICKSMAN AND BROWER 1988). In

the labial segment of null *Scr* alleles, maxillary sense organs are partially duplicated and labial derivatives, like salivary glands, are lost (MAHAFFEY AND KAUFMAN 1987). In the prothoracic segment of null *Scr* alleles, T2 denticles replace the T1 denticles in the T1 denticle belt and the characteristic prothoracic derivative, the T1 beard, is reduced (MAHAFFEY AND KAUFMAN 1987). Loss-of-function viable hypomorphic *Scr* alleles induce similar transformations in the adult fly (RILEY ET AL. 1987). In the labial segment of *Scr* hypomorphs, the proboscis transforms towards maxillary palp identity, exhibiting a decreased number of pseudotracheal rows and the formation of maxillary palp-like bristles (PATTATUCCI ET AL. 1991). In the prothoracic segment of adult *Scr* hypomorphs, T1 transforms towards T2 identity, exhibiting a decreased number of sex comb bristles on the T1 legs of males (PATTATUCCI ET AL. 1991). Gain-of-function *Scr* alleles that result in ectopic expression further illustrate the developmental role of SCR. Ectopic SCR expression in the developing embryo induces the formation of ectopic salivary glands, ectopic T1 beards on posterior segments (T2 and T3) and the disruption of head involution (GIBSON ET AL. 1990; ZHAO ET AL. 1993). Ectopic SCR expression during larval development induces an arista (antenna) to tarsus (leg) transformation, the malformation of the mouth parts, a reduction in the size of the compound eye and ectopic sex combs on the T2 and T3 legs of males (GIBSON ET AL. 1990; ZHAO ET AL. 1993).

Within the class Insecta, *Scr* is expressed in the labial and prothoracic segment of holometabolous insects, like Drosophila and *Tribolium*, and hemimetabolous insects, like *Oncopeltus*, *Acheta* and *Thermobia* (ROGERS ET AL. 1997; DECAMILLIS ET AL. 2001). Although the localization of *Scr* expression to the labial and prothoracic segments is conserved in insects, its expression pattern can differ. For example, in hemimetabolous insects *Scr* expression is restricted to small patches, whereas in Drosophila *Scr* is expressed broadly in the T1 segment (ROGERS ET AL. 1997). The variability of *Scr* expression pattern within Insecta has directly affected the morphological evolution of insects by allowing for the specialization of unique labial and prothoracic characteristics (ROGERS ET AL. 1997). The homeotic transformations of *Scr* mutant alleles can also vary between insects. In Drosophila, *Tribolium*, and *Oncopeltus*, *Scr* and *pb* are required for determining labial identity. Loss of *pb* function in all three insects results in homeotic transformations of the labial appendages to legs (DECAMMILIS ET AL. 2001). In contrast,

loss of *Scr* function induces different homeotic transformations in each insect (HUGHES AND KAUFMAN 2000; DECAMMILIS ET AL. 2001). *Tribolium* undergoes a labium to antenna transformation, *Oncopeltu*s undergoes a labium to mixed antenna/leg transformation, and Drosophila undergoes a labium (i.e. proboscis) to maxillary transformation (HUGHES AND KAUFMAN 2000; DECAMMILIS ET AL. 2001). Despite differences in insect *Scr* expression patterns and homeotic transformations, the role of *Scr* in the determination of the labial and prothoracic segments is a conserved function.

The function of *Scr* is also conserved in higher order bilaterian phyla. The functional conservation of SCR activities is illustrated by the ectopic expression of the murine (Chordate) *Scr* ortholog, HoxA5, in Drosophila. Ectopic HoxA5 expression reproduces similar homeotic transformations in Drosophila as seen with the ectopic expression of Drosophila SCR protein. Ectopic HoxA5 expression during embryogenesis induces ectopic T1 beards on T2 and T3, in the first instar larvae. In the adult, ectopic HoxA5 expression induces a strong arista to T1-like leg transformation – with some legs displaying sex comb bristles – malformed mouth parts and reduced compound eyes (ZHAO ET AL. 1993). The ability of HoxA5 to induce ectopic T1 structures and transformations suggest that SCR^{T1} activity is universally conserved. SCR^{lab} activity, however, is not conserved in murine HoxA5 (PERCIVAL-SMITH ET AL. 2013). Co-ectopic expression of murine HoxA5 with PB cannot induce ectopic proboscises – like co-ectopic expression of Drosophila SCR and PB can – suggesting that Drosophila SCR has acquired this activity during the evolution of insects (PERCIVAL-SMITH ET AL. 2013). Thus, some functional properties of SCR and its cognates have been conserved over 600 million years since the divergence of Arthropods and Chordates: SCR^{T1} activity is universally conserved in bilaterian SCR orthologs, while SCR^{lab} may have been acquired at some point during insect evolution (PERCIVAL-SMITH ET AL. 2013).

1.9 Functional conservation of Sex combs reduced protein

The SCR protein contains multiple peptide domains and motifs that have been conserved at different taxonomic levels (**FIGURE 1.7** and **FIGURE 1.8**). The HOX5 class homeodomain (HD) and YPWM motif of SCR protein are sufficient for SCR^{T1} activity and are conserved in all bilaterian SCR orthologs, including the murine SCR ortholog,

FIGURE 1.7 Conserved regions in Drosophila Sex combs reduced protein. The structure of the SCR protein (417 amino acids) in *D. melanogaster*, indicating the taxonomic level of conserved peptide motifs and domains. The octapeptide motif, YPWM motif, homeodomain (HD), and KMAS motif are universally conserved in bilaterians. The LASCY, SCKY, PQDL, and NANGE motifs are conserved in all arthropods and protostomes. The DYTQL motif and C-terminal domain (CTD) are conserved insects. The YTPNL, DISPK, and NEAGS are conserved amongst Dipterans. The NDPVT, QSLAS, and VNVPM are found only in Drosophilids. The grey region of the protein indicates amino acid sequence in which conserved regions have not been identified.

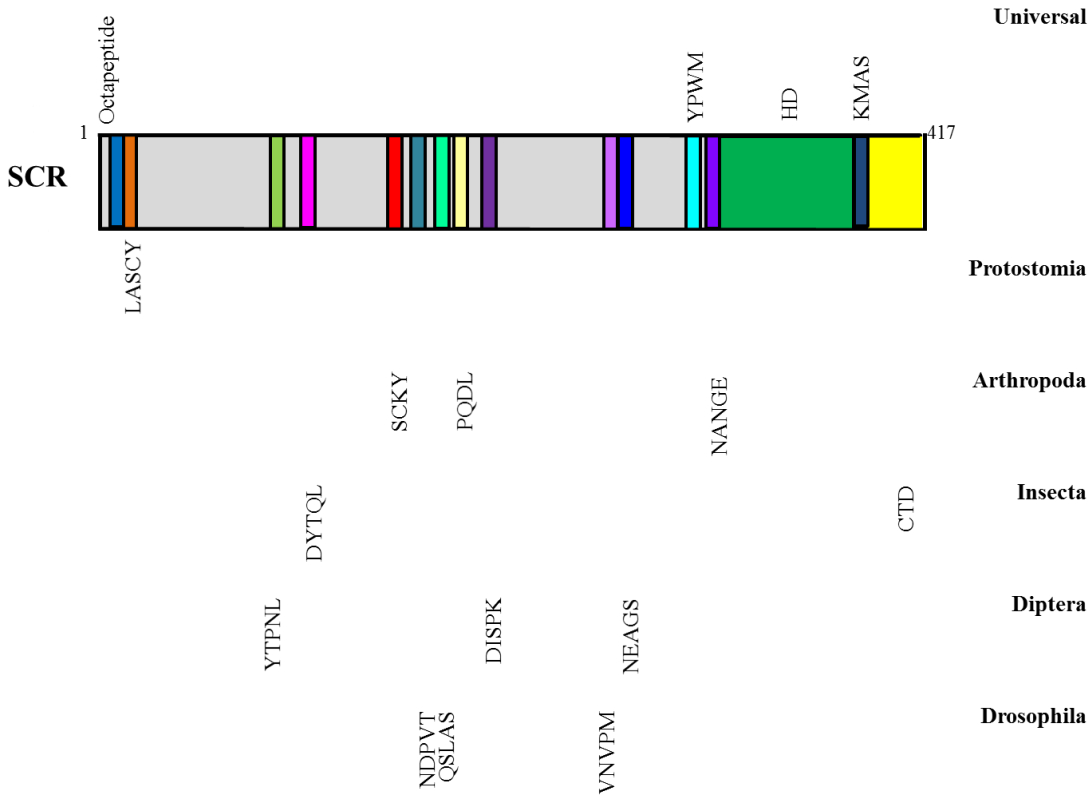
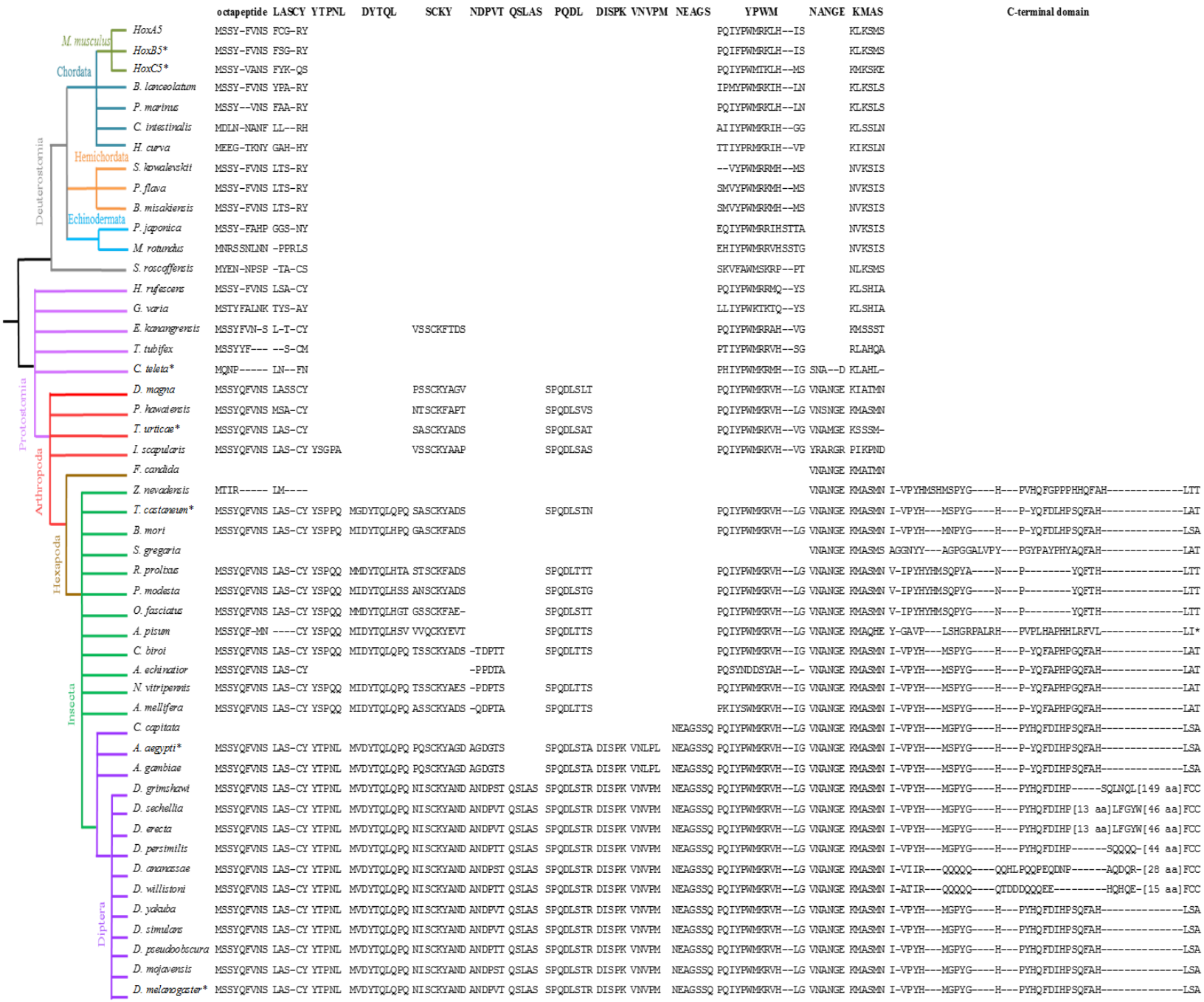

Universal
Octapeptide
YPWM
HD
KMAS
1
417
SCR
Protostomia
LASCY
Arthropoda
SCKY
PQDL
NANGE
Insecta
DYTQL
CTD
Diptera
YTPNL
DISPK
NEAGS
Drosophila
NDPVT
QSLAS
VNVPM

FIGURE 1.8 Alignment of representative sequences of bilaterian Sex combs reduced orthologs. There are multiple conserved regions in orthologous SCR proteins, outside of the homeodomain, which are conserved at different taxonomical levels. Sequence alignments of orthologous SCR proteins are shown for all conserved regions except the HD. Portions of the amino acid sequence in the C-terminal domain are condensed for simplicity, which is indicated by square brackets and the number of amino acids (aa) encompossed in the condensed portion. In this unrooted tree (left), the length of branch lines does not indicate evolutionary time, but rather, is an indication of taxonomic relatedness. Organisms belonging to the same taxonomic classification are grouped together, such that the taxonomic level of conservation of SCR protein regions is emphasized. Species with an asterisk (*) denote organisms, from which, *Scr* orthologs were isolated for use in the cloning of the reintegration-ortholog vectors. Accession numbers for SCR orthologs are listed in **APPENDIX 1.1**. A summary of conserved SCR regions can be found in **APPENDIX 1.2**. Primary protein sequences were aligned using ClustalW2 (EMBL-EBI) and MAFFT v.6 (KATOH ET AL. 2002).



	octapeptide	LASCY	YTPNL	DYTQL	SCKY	NDPVT	QSLAS	PQDL	DISPK	VNVPM	NEAGS	YPWM	NANGE	KMAS	C-terminal domain
HoxA5	MSSY-FVNS	FCG-RY										PQIYPWMRKLH--IS		KLKSMS	
*HoxB5**	MSSY-FVNS	FSG-RY										PQIFPWMRKLH--IS		KLKSMS	
*HoxC5**	MSSY-VANS	FYK-QS										PQIYPWMTKLH--MS		KMKSKE	
B. lanceolatum	MSSY-FVNS	YPA-RY										IPMYPWMRKIH--LN		KLKSLS	
P. marinus	MSSY--VNS	FAA-RY										PQIYPWMRKLH--LN		KLKSLS	
C. intestinalis	MDLN-NANF	LL--RH										AIIYPWMKRIH--GG		KLSSLN	
H. curva	MEEG-TKNY	GAH-HY										TTIYPRMKRIH--VP		KIKSLN	
S. kowalevskii	MSSY-FVNS	LTS-RY										--VYPWMRRMH--MS		NVKSIS	
P. flava	MSSY-FVNS	LTS-RY										SMVYPWMRKMH--MS		NVKSIS	
B. misakiensis	MSSY-FVNS	LTS-RY										SMVYPWMRKMH--MS		NVKSIS	
P. japonica	MSSY-FAHP	GGS-NY										EQIYPWMRRIHSTTA		NVKSIS	
M. rotundus	MNRSSNLNN	-PPRLS										EHIYPWMRRVHSSTG		NVKSIS	
S. roscoffensis	MYEN-NPSP	-TA-CS										SKVFAWMSKRP--PT		NLKSMS	
H. rufescens	MSSY-FVNS	LSA-CY										PQIYPWMRRMQ--YS		KLSHIA	
G. varia	MSTYFALNK	TYS-AY										LLIYPWKIKTQ--YS		KLSHIA	
E. kanangrensis	MSSYFVN-S	L-T-CY			VSSCKFTDS							PQIYPWMRRAH--VG		KMSSST	
T. tubifex	MSSYYF---	--S-CM										PTIYPWMRRVH--SG		RLAHQA	
*C. teleta**	MQNP-----	LN--FN										PHIYPWMKRMH--IG	SNA--D	KLAHL-	
D. magna	MSSYQFVNS	LASSCY			PSSCKYAGV			SPQDLSLT				PQIYPWMKRVH--LG	VNANGE	KIAIMN	
P. hawaiensis	MSSYQFVNS	MSA-CY			NTSCKFAPT			SPQDLSVS				PQIYPWMKRVH--LG	VNSNGE	KMASMN	
*T. urticae**	MSSYQFVNS	LAS-CY			SASCKYADS			SPQDLSAT				PQIYPWMRKVH--VG	VNAMGE	KSSSM-	
I. scapularis	MSSYQFVNS	LAS-CY	YSGPA		VSSCKYAAP			SPQDLSAS				PQIYPWMRKVH--VG	YRARGR	PIKPND	
F. candida													VNANGE	KMAIMN	
Z. nevadensis	MTIR-----	LM----											VNANGE	KMASMN	I-VPYHMSHMSPYG----H---PVHQFGPPPHHQFAH----------LTT
*T. castaneum**	MSSYQFVNS	LAS-CY	YSPPQ	MGDYTQLQPQ	SASCKYADS			SPQDLSTN				PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MSPYG----H---P-YQFDLHPSQFAH----------LAT
B. mori	MSSYQFVNS	LAS-CY	YSPPQ	MIDYTQLHPQ	GASCKFADS			SPQDLSTG				PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MNPYG----H---P-YQFDLHPSQFAH----------LSA
S. gregaria													VNANGE	KMASMS	AGGNYY---AGPGGALVPY---PGYPAYPHYAQFAH----------LAT
R. prolixus	MSSYQFVNS	LAS-CY	YSPQQ	MMDYTQLHTA	STSCKFADS			SPQDLTTT				PQIYPWMKRVH--LG	VNANGE	KMASMN	V-IPYHYHMSQPYA----N---P--------YQFTH----------LTT
P. modesta	MSSYQFVNS	LAS-CY	YSPQQ	MIDYTQLHSS	ANSCKYADS			SPQDLSTG				PQIYPWMKRVH--LG	VNANGE	KMASMN	V-IPYHYHMSQPYG----N---P--------YQFTH----------LTT
O. fasciatus	MSSYQFVNS	LAS-CY	YSPQQ	MMDYTQLHGT	GSSCKFAE-			SPQDLSTT				PQIYPWMKRVH--LG	VNANGE	KMASMN	V-IPYHYHMSQPYG----N---P--------YQFTH----------LTT
A. pisum	MSSYQF-MN	----CY	YSPQQ	MIDYTQLHSV	VVQCKYEVT			SPQDLTTS				PQIYPWMKRVH--LG	VNANGE	KMAQHE	Y-GAVP---LSHGRPALRH---PVPLHAPHHLRFVL----------LI*
C. biroi	MSSYQFVNS	LAS-CY	YSPQQ	MIDYTQLQPQ	TSSCKYADS	-TDPTT		SPQDLTTS				PQIYPWMKRVH--IG	VNANGE	KMASMN	I-VPYH---MSPYG----H---P-YQFAPHPGQFAH----------LAT
A. echinatior	MSSYQFVNS	LAS-CY				-PPDTA						PQSYNDDSYAH--L-	VNANGE	KMASMN	I-VPYH---MSPYG----H---P-YQFAPHPGQFAH----------LAT
N. vitripennis	MSSYQFVNS	LAS-CY	YSPQQ	MIDYTQLQPQ	TSSCKYAES	-PDPTS		SPQDLTTS				PQIYPWMKRVH--IG	VNANGE	KMASMN	I-VPYH---MSPYG----H---P-YQFAPHPGQFAH----------LAT
A. mellifera	MSSYQFVNS	LAS-CY	YSPQQ	MIDYTQLQPQ	ASSCKYADS	-QDPTA		SPQDLTTS				PKIYSWMKRVH--IG	VNANGE	KMASMN	I-VPYH---MSPYG----H---P-YQFAPHPGQFAH----------LAT
C. capitata											NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHP-----SQLNQL[149 aa]FCC
*A. aegypti**	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	PQSCKYAGD	AGDGTS		SPQDLSTA	DISPK	VNLPL	NEAGSSQ	PQIYPWMKRVH--IG	VNANGE	KMASMN	I-VPYH---MSPYG----H---P-YQFDIHPSQFAH----------LSA
A. gambiae	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	PQSCKYAGD	AGDGTS		SPQDLSTA	DISPK	VNLPL	NEAGSSQ	PQIYPWMKRVH--IG	VNANGE	KMASMN	I-VPYH---MSPYG----H---P-YQFDIHPSQFAH----------LSA
D. grimshawi	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPST	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHP-----SQLNQL[149 aa]FCC
D. sechellia	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPVT	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHP[13 aa]LFGYW[46 aa]FCC
D. erecta	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPVT	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHP[13 aa]LFGYW[46 aa]FCC
D. persimilis	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPTT	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHP------SQQQQ-[44 aa]FCC
D. ananassae	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPST	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VIIR---QQQQQ----QQHLPQQPEQDNP------AQDQR-[28 aa]FCC
D. willistoni	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPTT	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-ATIR---QQQQQ----QTDDDQQQEE---------HQHQE-[15 aa]FCC
D. yakuba	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPVT	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHPSQFAH----------LSA
D. simulans	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPVT	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHPSQFAH----------LSA
D. pseudoobscura	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPTT	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHPSQFAH----------LSA
D. mojavensis	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPST	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHPSQFAH----------LSA
*D. melanogaster**	MSSYQFVNS	LAS-CY	YTPNL	MVDYTQLQPQ	NISCKYAND	ANDPVT	QSLAS	SPQDLSTR	DISPK	VNVPM	NEAGSSQ	PQIYPWMKRVH--LG	VNANGE	KMASMN	I-VPYH---MGPYG----H---PYHQFDIHPSQFAH----------LSA

HoxA5 (PERCIVAL-SMITH ET AL. 2013). The HD is the only domain in SCR with an essential function: it is both necessary and sufficient for DNA-binding. The remainder of SCR protein is composed of a number of small, differentially conserved peptide motifs. The octapeptide and KMAS motif are universally conserved across all bilaterian SCR orthologs; the LASCY motif is conserved in protostome SCR orthologs; the SCKY, PQDL, and NANGE motifs are conserved in arthropod SCR orthologs; the DYTQL motif and C-terminal domain (CTD) are insect specific; the YTPNL, DISPK, and NEAGS are conserved in all Dipteran SCR orthologs; and the NDPVT, QSLAS, and VNVPM conserved only in Drosophila SCR orthologs (CURTIS ET AL. 2006; PERCIVAL-SMITH ET AL. 2013; SIVANANTHARAJAH AND PERCIVAL-SMITH 2015). Previous analysis suggests some of these motifs may contribute to SCR^{lab} activity; and from this functional analysis of SCR, it is suggested that three key events occurred during the evolution of insects: i) SCR acquired the negative regulatory DYTQL motif and CTD; ii) SCR acquired the activity to determine labial identity; and iii) PB expression shifted posteriorly to assist in the switch from the bilaterian conserved SCR^{T1} activity to the insect specific SCR^{lab} activity (PERCIVAL-SMITH ET AL. 2013).

A functional dissection of *Scr* resulted in two key discoveries: (1) *Scr* exhibits differential pleiotropy, and (2) all of the conserved SCR protein motifs – not including the HD, octapeptide, or CTD – may represent plastic sequence elements, called short linear sequence motifs (SLiMs; SIVANANTHARAJAH AND PERCIVAL-SMITH 2014, 2015). Differential pleiotropy is the observation that short, independently acting peptide elements each make small, additive tissue-specific contributions to SCR activity (CARROLL 2005; SIVANANTHARAJAH AND PERCIVAL-SMITH 2009; SIVANANTHARAJAH AND PERCIVAL-SMITH 2014). In a functional analysis of *Scr*, hypomorphic *Scr* alleles were ranked from weakest to strongest *Scr* phenotype in three tissues: the sex combs bristles, the proboscis, and the larval salivary glands (SIVANANTHARAJAH AND PERCIVAL-SMITH 2009). If every region in SCR was uniformly required in all tissues, the same allelic series would be expected for each tissue. However, there was a differential requirement of the octapeptide, DYTQL, NEAGS, YPWM, and CTD in all three tissues. This study concluded that SCR functions are distributed throughout the protein in small, additive functional motifs that are important, but not essential for SCR activities

(SIVANANTHARAJAH AND PERCIVAL-SMITH 2009). Some of these SCR motifs may be SLiMs, which are small motifs of 3-10 amino acids that act as effector binding sites with widespread cellular function (SIVANANTHARAJAH AND PERCIVAL-SMITH 2015). Many SLiMs in SCR appear to have seemingly non-essential functions. The YPWM motif, for example, is important but not essential for SCR function as it can be deleted with minimal phenotypic effect. The expression of Scr^{3}, a hypomorphic allele, which has a YPWM change to YLWM, has only a small effect on salivary gland development in Drosophila (JOSHI ET AL. 2010; SIVANANTHARAJAH AND PERCIVAL-SMITH 2009). Differential pleiotropy is a genetic observation which suggests that some SCR protein motifs are important, but not essential, for protein function and it provides a mechanism to facilitate *Scr* gene evolution by reducing the pleiotropy of a mutation (HITTINGER ET AL. 2005; MERABET ET AL. 2011; SIVANANTHARAJAH AND PERCIVAL-SMITH 2015).

SCR plays a crucial role in conferring labial and prothoracic segmental identity in Drosophila. Previous analysis of *Scr* orthologs indicate that SCR^{T1} activity is conserved in *Scr* orthologs. SCR-dependent prothoracic phenotypes, such as the T1 beards and male sex comb, generated by SCR^{T1} activity can be induced by ectopic expression of *Scr* orthologs, like murine HoxA5 (ZHAO ET AL. 1993). However, the emphasis of my research is to identify the time at which SCR^{lab} activity arose. One proposal suggests that labial activity is insect specific (PERCIVAL-SMITH ET AL. 2013). In this case, the rescue of SCR-dependent labial phenotypes would only be observed with insect *Scr* orthologs. However, my analysis is not limited to testing this specific hypothesis. It is possible that SCR^{lab} activity evolved prior to the common ancestor of insects and other arthropods. In this case, the rescue of SCR-dependent labial phenotypes would be observed with the *Scr* orthologs outside of Insecta. The completion of a comprehensive functional analysis of *Scr* orthologs in Drosophila using CRISPR and RMCE enables the determination of the phylogenetic range of SCR labial and prothoracic activities and the identification of the evolutionary time point at which SCR activity changed.

2 MATERIALS AND METHODS

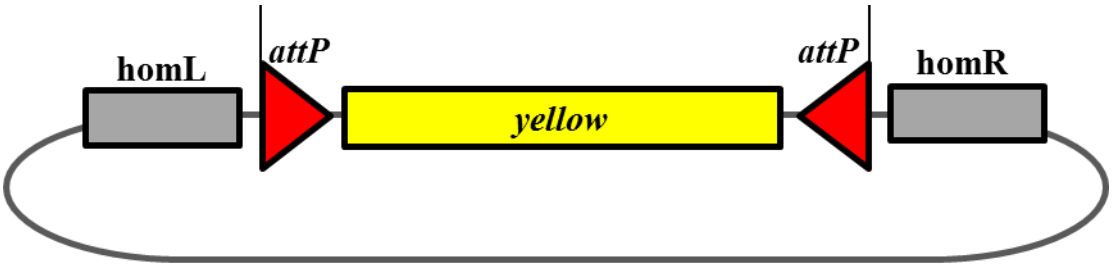
2.1 Acquisition and preparation of materials

Primers were obtained from Invitrogen Life Technologies (Burlington, Ontario, Canada). PCR for cloning was performed with high-fidelity *Taq* polymerase (Invitrogen Life Technologies). PCR purifications were performed with EZ-10 Spin Column DNA Gel Extraction kit (Bio Basic Inc., Toronto, Ontario, Canada) according to the manufacturer's instructions. Restriction enzymes and T4 DNA ligases were from New England BioLabs (Whitby, Ontario, Canada). The subcloning efficiency DH5α competent cells (Invitrogen Life Technologies) were used for bacterial transformation of ligations. Bacteria were grown on LB plates containing spectinomycin (pFus_A) or ampicillin (reintegration vector and all derivatives) at 100 µg/ml. Spectinomycin (pFus_A) plates also included 30 µL 8% X-gal and 30 µL 200 µM IPTG prior to the plating of bacteria. Plasmid purifications were performed using the Presto Mini Plasmid Kit (Geneaid Biotech Ltd., New Taipei City, Taiwan) according to the manufacturer's instructions. Plasmid purifications for constructs used in embryo injections were performed using QIAfilter Plasmid Midi Kit (QIAGEN Inc., Toronto, Ontario, Canada).

2.2 Construction of the donor template

The *Scr* locus has three exons, the first of which is non-coding. CRIPSR/Cas9-mediated HDR was used to replace the *Scr* coding sequence (i.e. second and third exons) with the *yellow* (y^+) body marker. The donor template, containing the DNA sequences from *Scr* exon 2 and *Scr* exon 3 plus y^+ flanked by inverted *attP* ϕC31 recombination sites, was constructed (**FIGURE 2.1**). *Scr* exon 2 sequence (1642 bp) was PCR amplified from *y w D. melanogaster* genomic DNA with primers X2-*Scr*-*Bsa*I-F and X2-*Scr*-*attP*-*Bsa*I-R (**APPENDIX 2**). *Scr* exon 3 (1762 bp) was PCR amplified from *y w D. melanogaster* genomic DNA with primers X3-*Scr*-*Bsa*I-*attP*-F and X3-*Scr*-*Bsa*I-R (**APPENDIX 2**). The primers used to amplify *Scr* exons 2 and 3 added the *attP* recombination site sequence (39 bp) and a *Bsa*I restriction site to the 3' and 5' end of exons 2 and 3, respectively. The y^+ gene was PCR amplified from *Not*I-digested MiMIC plasmid (GenBank plasmid #GU370067; VENKEN ET AL. 2011) with primers Y-*Bsa*I-F and Y-*Bsa*I-R (**APPENDIX 2**), which added *Bsa*I restriction sites to the 5' and 3' ends. The three DNA fragments – *Scr* exon 2 (with 3' *attP*), *Scr* exon 3 (with 5' *attP*), and y^+ – were digested with *Bsa*I, which

FIGURE 2.1 The donor template. The donor template can be used with the CRISPR/Cas9 system to induce homology-directed repair. The donor template contains three features: (1) a left and right homology arm (homL and homR, respectively; green boxes) with the sequences from *D. melanogaster Scr* exon 2 and 3, respectively, (2) inverted *attP* recombination sites for ϕC31 integrase-mediated cassette exchange, and (3) the Drosophila body marker *yellow*.



pFus_A

generates unique 5’ overhangs. *Scr* exon 2, *Scr* exon 3, and y^+ were purified and ligated together in an ordered assembly reaction into dephosphorylated, *Bsa*I-digested pFus_A (Addgene plasmid #31028; CERMAK ET AL. 2011), resulting in the donor template.

pFus_A contains a *lacZ* gene for blue-white selection that is excised after digestion with *Bsa*I. The successful ligation of *Scr* exon 2, *Scr* exon 3, and y^+ into pFus_A results in plasmid that entirely lacks *lacZ*. Bacterial colonies containing the correct donor template should appear as white and thus, were screened using colony PCR. Three screens were designed to amplify junctions within the donor template that would have been created if successful ligation occurred (**APPENDIX 3**). Screen A was used in a preliminary screen of all white colonies to identify any potentially correct clones, and then screens B and C were used in succession to further identify correct clones. Screen A amplified a 1936 bp fragment – containing *Scr* exon 2 – from pFus_A to y^+. Screen B amplified a 2049 bp fragment – containing *Scr* exon 3 – from the 3’ end of y^+ to pFus_A. Screen C amplified a 329 bp fragment from the 3’ end of y^+ to *Scr* exon 3. In all screens, a blue colony was used as a negative control and 1 µL the ligation was used as a positive control.

2.3 Construction of pU6-chiRNAs

Two chiRNAs were designed to recognize and target the coding region of *Scr*, in exons 2 and 3. For efficient target recognition, chiRNAs require 20 nt of complementary DNA to its genomic target, *Scr*, the first base pair of which must be a guanine (JINEK ET AL. 2012). Cleavage by Cas9 also requires that the 3’ end of the genomic target sequence contain diguanines (NGG), known as the proto-spacer adjacent motif (PAM) (JINEK ET AL. 2012). PAM sequences naturally occur in *D. melanogaster* and were identified with a sequence search of *Scr* in BLAST. The 5’ chiRNA targets the 5’ of *Scr* exon 2 in *Scr* and the 3’ chiRNA targets the 3’ of *Scr* exon 3. The target-specific sequences for both of the *Scr* chiRNAs were synthesized as 5’-phosphorylated oligonucleotides (**APPENDIX 4**), annealed, and ligated into the *Bbs*I sites of pU6-*Bbs*I-chiRNA (GRATZ ET AL. 2013). The *Bbs*I restriction sites were abolished after successful ligation, and therefore *Bbs*I failed to digest correct chiRNA clones. The 5’ and 3’ chiRNAs were verified by DNA sequencing at the Robarts DNA Sequencing Facility (London, Ontario, Canada).

2.4 Construction of the reintegration vectors for recombinase-mediated cassette exchange

The reintegration vector. RMCE, specifically ϕC31 integrase-mediated cassette exchange, requires *attB*-containing vectors for exchange into an *attP*-containing genomic target. A reintegration vector containing inverted *attB* ϕC31 recombination sites was generated for use in RMCE. *Xba*I and *Hind*III were used to excise the *GAL4* coding sequence from a pBS-KS derivative (Drosophila Genomics Resource Center, Vector 1325) that contains two inverted *attB* sequences. A multiple cloning sequence (MCS) – containing *EcoR*I, *Bgl*II, *Not*I and *Xho*I sites – was synthesized as 5'-phosphorylated oligonucleotides (**APPENDIX 5**), annealed, and ligated into the *Xba*I and *Hind*III sites of pBS-KS, resulting in the reintegration vector (**FIGURE 2.2**). The reintegration vector, containing the MCS between inverted *attB* sites, was used in subsequent cloning experiments to insert *Scr* orthologs. The reintegration vector was verified by DNA sequencing at the Robarts DNA Sequencing Facility.

The reintegration-ortholog vectors. The reintegration-ortholog vectors contain *Scr* orthologs flanked by inverted *attB* sites. These vectors can be used for ϕC31 integrase-mediated cassette exchange at *attP* sites in Drosophila *Scr* (**FIGURE 2.3**). *Scr* orthologs were isolated from *Aedes aegypti* (yellow fever mosquito), *Capitella teleta* (polychaete worm), *Mus musculus* (mouse), *Tetranychus urticae* (red spider mite), and *Tribolium casateneum* (red flour beetle). The two coding exons in the *Scr* orthologs were PCR amplified from the genomic DNA of each species with primers that added a restriction enzyme site found in the MCS of the reintegration vector (**APPENDIX 6**). The two coding exons of these *Scr* orthologs were then joined together by PCR (HO ET AL. 1989). All *Scr* ortholog amplicons begin at the ATG start codon in the first coding exon and end at the stop codon in the second coding exon, with the exception of *C. teleta* (*Hox5*) – which includes 14 nt of the 3' UTR. The *Scr* orthologs from *A. aegypti* (*Scr*), *C. teleta* (*Hox5*), *M. musculus* (*HoxB5*), *T. urticae* (*Scr*), and *T. casateneum* (*Cx*) were PCR amplified with primers that create *Not*I ends, resulting in the fragments: Aa-*Scr*-*Not*I, Ct-*Hox5*-*Not*I, Mm-*HoxB5*-*Not*I, Tu-*Scr*-*Not*I, and Tc-*Cx*-*Not*I. These *Scr* ortholog fragments were gel isolated and digested with *Not*I. Another *Scr* ortholog from *M. musculus* (*HoxC5*) was PCR amplified with primers that create *BamH*I ends, resulting in the fragment

FIGURE 2.2 The reintegration vector. RMCE, specifically ϕC31 integrase-mediated cassette exchange, requires *attB*-containing vectors for exchange into an *attP*-containing genomic target. The reintegration vector contains a multiple cloning sequence (MCS; sequence shown) – containing *EcoR*I, *Bgl*II, *Not*I and *Xho*I sites – between inverted *attB* sites. The reintegration vector was used in subsequent cloning experiments to insert *Scr* orthologs between the *attB* sites.

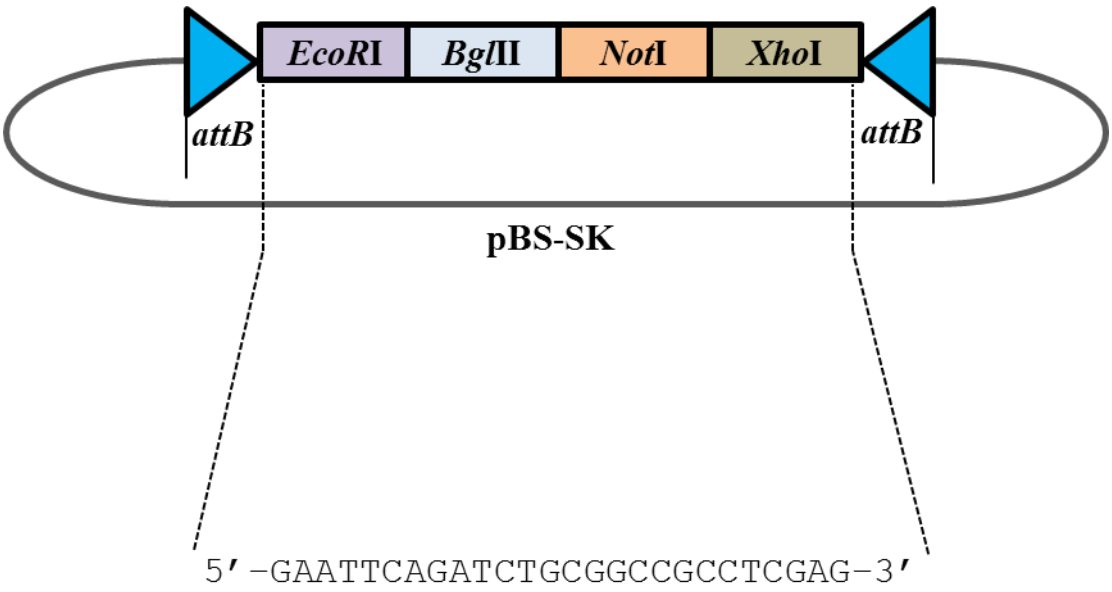
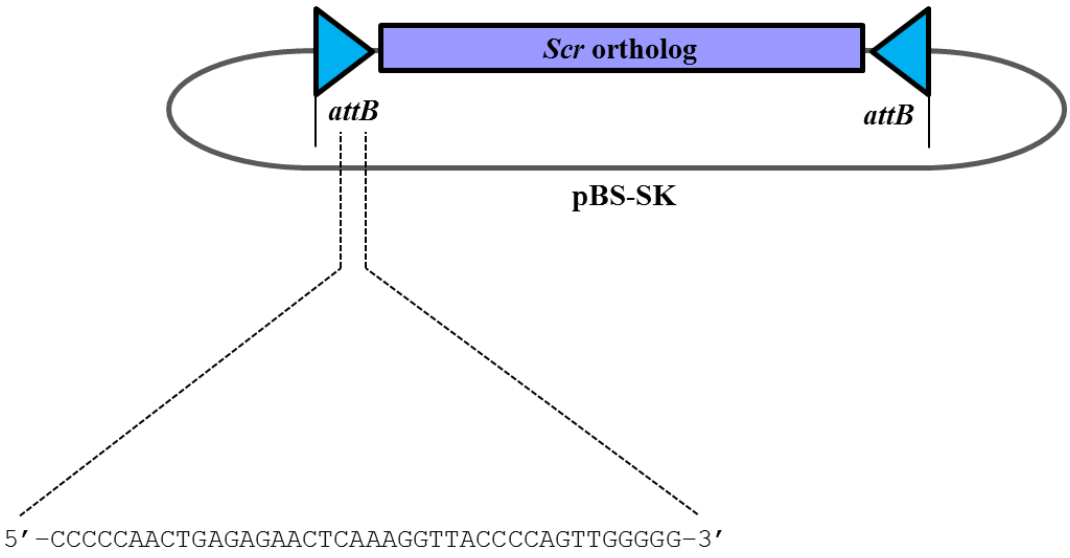

EcoRI
BglII
NotI
XhoI
attB
attB
pBS-SK
5′-GAATTCAGATCTGCGGCCGCCTCGAG-3′

FIGURE 2.3. Schematic of the reintegration-ortholog vector. The reintegration-ortholog vector contains a *Scr* ortholog flanked by inverted *attB* sites (*attB* sequence shown) and can be used for ϕC31 integrase-mediated cassette exchange at *attP* sites flanking *Scr* in the Drosophila genome. The *Scr* ortholog contains only exonic, coding sequence.


Scr ortholog
attB
attB
pBS-SK
5′–CCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGG–3′

Mm-*HoxC5-BamH*I. Mm-*HoxC5-BamH*I was gel isolated and digested with *BamH*I, creating *Bgl*II-compatible cohesive ends. The *D. melanogaster Scr* coding sequence was obtained by gel isolating pUAST-*Scr* digested with *EcoR*I and *Xho*I, resulting in the fragment Dm-*Scr-EcoR*I-*Xho*I. Dm-*Scr-EcoR*I-*Xho*I was digested with *EcoR*I and *Xho*I. Each *Scr* ortholog was ligated into a dephosphorylated reintegration vector cut with *Not*I, *Bgl*II or *EcoR*I/*Xho*I, generating seven reintegration-ortholog vectors. The reintegration-ortholog vectors were verified by DNA sequencing at the Robarts DNA Sequencing Facility and point mutations were identified with BLAST (**APPENDIX 7**).

2.5 Extraction of genomic DNA

Genomic DNA of *A. aegypti*, *D. melanogaster*, *T. urticae*, and *T. castaneum* was purified via phenol-chloroform extraction (**APPENDIX 8**). For all species, 25-40 specimens were homogenized in 300 µL lysis buffer (10% Tris HCl pH 9.15, 4% EDTA, 5% SDS and 1 µg/mL RNase A) and incubated on ice for 30 min. The mixture was centrifuged at 21 000x*g* for 15 min at 4 ºC, and the supernatant was transferred to a new tube. Equal volumes of phenol and chloroform were added and the mixture was centrifuged at 21 000x*g* for 7 min at room temperature. The supernatant was transferred to a new tube, 600 µL 100% ethanol was added and centrifuged at 21 000x*g* for 20 min at room temperature. The DNA pellet was washed with 400 µL 75% ethanol. The DNA pellet was re-suspended in 100 µL DNase-free water and 10 µL 3M sodium acetate and 250 µL 100% ethanol was added. The tube was inverted until DNA precipitation was visible, followed by centrifugation at 21 000x*g* for 3 min at room temperature. The supernatant was removed and the DNA pellet was washed with 400 µL 75% ethanol. The DNA pellet was desiccated for 10 min and re-suspended in 100 µL DNase-free water.

2.6 Fly stocks and culture

The fly strains used were obtained from the Bloomington Stock Center (Indiana University, Bloomington, Indiana). The genotypes of the stocks used were: *y w*, strain 5135 (y^1*; P{neo*^{r}*, FRT}82B P{*w^+*}*; Bloomington stock center, 5135), act-cas9 (y^1 *M{Act5c-cas9} ZH-2A* w^+; Bloomington stock center, 54590), and nos-cas9 (y^1 *P{nos-cas9} M{nos-Cas9}ZH-2A* w^+; Bloomington stock center, 54591) All fly stocks and

crosses were maintained on standard cornmeal food and kept at 23-25 ºC with 50 ± 5% relative humidity and a 12 h light/dark cycle.

2.7 Injection of embryos

Live embryos were collected on a yeasted apple or grape juice plate every 30 min at 23-25 ºC and dechorionated for 60 s with 3% sodium hypochlorite. Embryos were then washed with tap water and lined up with a dissecting needle on an apple or grape juice agar strip. Embryos were transferred onto a glass microscope slide with acid-free double-sided tape. Embryos were partially desiccated under a hairdryer for 4-5 min. Embryos were covered in halocarbon oil and transferred to a microscope equipped with a manual micromanipulator (Wilovert, Wetzlar, Germany). DNA was microinjected into the posterior pole of the embryo using a heat-pulled glass needle (capillary tubing, FHC Inc., Bowdoin, Maine, USA) attached to a halocarbon filled syringe. All injections were performed at room temperature, with typically 50-100 embryos injected 30 min AEL, at the syncytial blastoderm stage. The injection medium for CRISPR/Cas9-mediated HDR used on embryos that did not express cas9 (*y w* and strain 5135) was prepared to final concentrations: pHsp70-Cas9 500 ng/µl, chiRNA 500 ng/µl (each), donor template 120 ng/µl, 10% glycerol, and PBS. The injection medium for CRISPR/Cas9-mediated HDR used on Cas9-expressing embryos (act-cas9 and nos-cas9) was prepared to final concentrations: chiRNA 500 ng/µl, donor template 500 ng/µl, 10% glycerol, and PBS for the injection of a single chiRNA; and chiRNA 250 ng/µl (of each), donor template 500 ng/µl, 10% glycerol, and PBS for the injection of the 5’ and 3’ chiRNAs together.

2.8 Identification of successful CRISPR/Cas9 mutants

To assess the germline transmission of targeted modifications (generated by CRISPR/Cas9-mediated HDR), all G_0 adult flies that developed from injected embryos (hereafter, “survivors”) were crossed to *y w* flies (for *y w*, act-cas9, and nos-cas9 survivors) or strain 5135 flies (for strain 5135 survivors). The F1 progeny were screened for 10-14 days after the first flies emerged for progeny with brown or y^+ bodies (i.e. dark coloration of the adult cuticle), indicating transmission of the donor template.

3 RESULTS

3.1 A strategy for studying Sex combs reduced activities in Drosophila

3.1.1 The generation of a model fly using CRISPR/Cas9-mediated homology-directed repair

In order to study the functional conservation of SCR^{T1} and SCR^{lab} activities, my goal is to create flies with *Scr* orthologs expressed from *Scr* regulatory sequences. To do this, a manipulable *Scr* locus is needed. A manipulable *Scr* locus contains the selective body color marker y^+ flanked by inverted *attP* sites. The *Scr* locus has three exons, the first of which is non-coding; CRIPSR/Cas9-mediated HDR can replace the coding sequence of *Scr* with the y^+ sequence flanked by inverted *attP* sites (**FIGURE 3.1.A**). Thus, the donor template used for HDR contains the y^+ sequence flanked by inverted *attP* sites sequence; the donor template also contains regions of *Scr* coding sequence for accurate HDR. The CRISPR nuclease Cas9 requires guide sequences, chiRNAs, for targeted DNA cleavage. The two chiRNAs synthesized recognize and target the 5' end of exon 2 and the 3' end of exon 3 in *Scr* (**FIGURE 3.2**). The CRISPR components, Cas9 (excluding injections into act-cas9 and nos-cas9 flies) and the two chiRNAs, plus the donor template were injected into syncytial blastoderm embryos. Cas9, guided by chiRNAs, will induce two DSBs upstream each of the PAM sequences in *Scr* and HDR can then occur, inserting the donor template DNA carried on *pFus_A* (**FIGURE 3.1.B**). The donor template DNA should insert between exons 2 and 3 of *Scr*, essentially replacing *Scr* coding sequence and intron 2. Transformed flies ($y\ w;\ Scr^{attP\ y+\ attP}$) that have incorporated the *attP*-y^+ cassette can then be identified by the wild type (y^+) body marker (**FIGURE 3.1.C**).

3.1.2 The incorporation of *Sex combs reduced* orthologs into Drosophila using recombinase-mediated cassette exchange

ϕC31 integrase-mediated cassette exchange requires *attB*-containing vectors (reintegration vector) for exchange into an *attP*-containing genomic target. Transgenic fly lines that express a *Scr* ortholog from Drosophila *Scr* regulatory sequences can be generated through the co-injection of a reintegration-ortholog vector and ϕC31 integrase into syncytial blastoderm flies. Reintegration-ortholog vectors were generated that

FIGURE 3.1 Schematic of CRISPR/Cas9-mediated homology-directed repair at the *Sex combs reduced* locus. (A) The *Scr* locus in *D. melanogaster* prior to genetic manipulation contains three exons (boxes). Green denotes a coding sequence and grey denotes a non-coding sequence in the final SCR protein. (B) Two chiRNAs (not shown) recognize and bind exon 2 and exon 3 of *Scr*, and Cas9 nuclease (not shown) induces two double-stranded breaks (DSBs) in *Scr* upstream the PAM sequence (not shown). (C) CRISPR/Cas9-mediated homology-directed repair (HDR) occurs at the DSBs and donor template DNA – containing the *Scr* homology arms (homL and homR; green box), the *yellow* (y^+) gene (yellow box), and the *attP* sequences (red triangles) – is inserted between exons 2 and 3 of *Scr*. (D) The *Scr* locus in *D. melanogaster* after CRISPR/Cas9-mediated HDR. The coding sequence and second intron in the *Scr* locus has been replaced by two inverted *attP* recombination sites that flank the Drosophila body marker, y^+ (*y w;* $Scr^{attP\ y+\ attP}$).

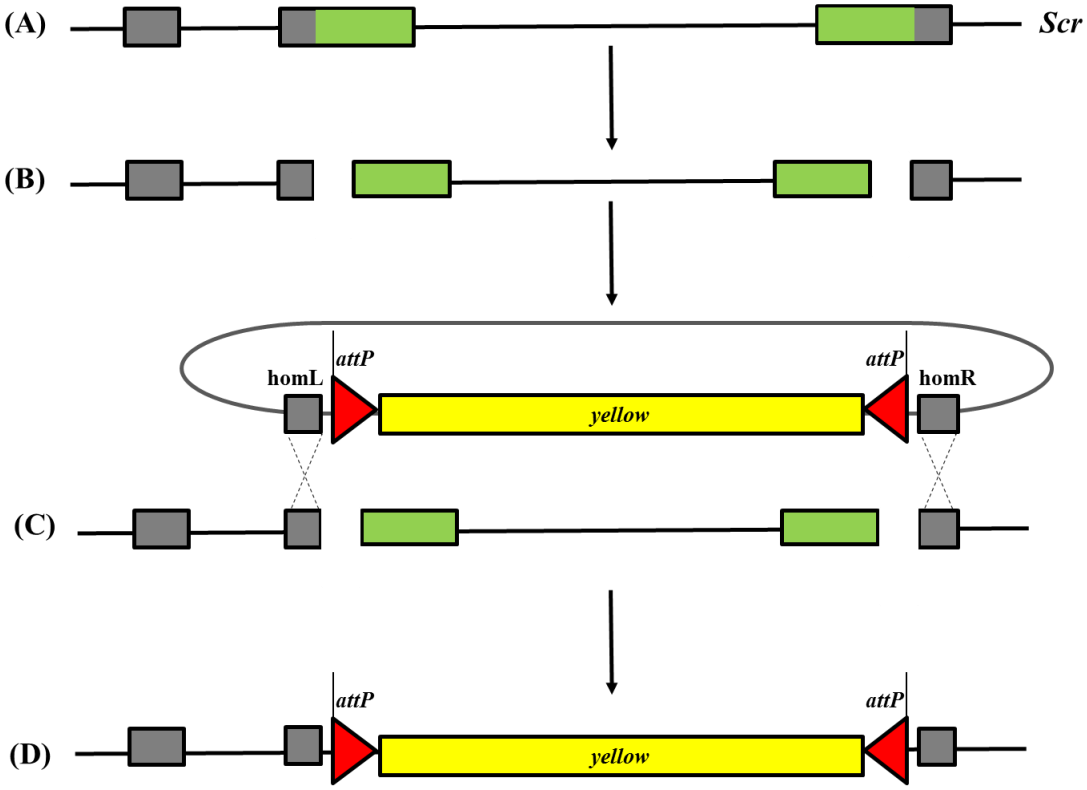
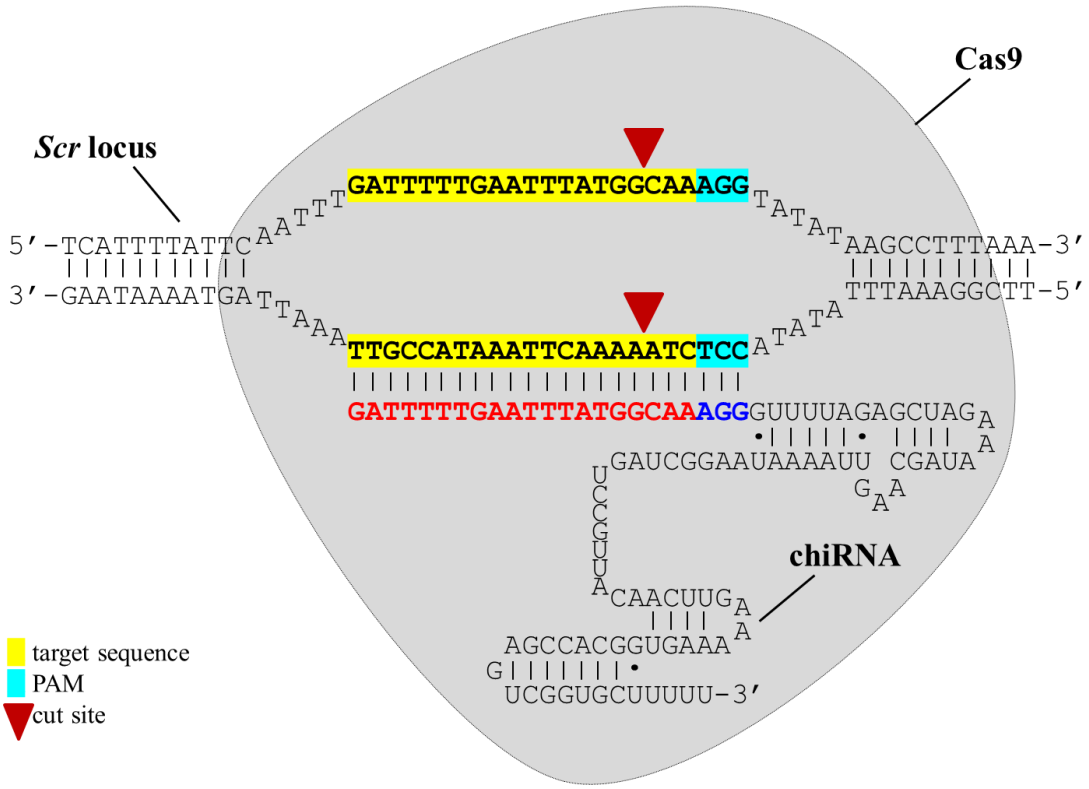

(A)
Scr
(B)
homL
attP
attP
homR
yellow
(C)
attP
attP
(D)
yellow

FIGURE 3.2 Schematic of the 5’ chiRNA and Cas9 complex at exon 2 of the *Sex combs reduced* locus. The 5’ chiRNA guides the CRISPR nuclease Cas9 (grey) to its complementary target DNA (yellow highlighted letters), the 5’ end of *Scr* exon 2. Target recognition requires 20 nt of complementary DNA (red letters) and a 3 bp PAM sequence, NGG (blue highlighted letters), at the 3’ end of the genomic target sequence. Cas9 cleaves (red arrowheads) the complementary and non-complementary DNA strands, resulting in a double-stranded break at the target site in *Scr*.


Cas9
Scr locus
GATTTTTGAATTTATGGCAAAGG
5′-TCATTTTATTC
3′-GAATAAAATGA
AAGCCTTTAAA-3′
TTTAAAGGCTT-5′
TTGCCATAAATTCAAAAATCTCC
GATTTTTGAATTTATGGCAAAGGGUUUUAGAGCUAG
GAUCGGAAUAAAAUU CGAUA
chiRNA
CAACUUG
AGCCACGGUGAAA
UCGGUGCUUUUU-3′
target sequence
PAM
cut site

contain the *Scr* ortholog from *A. aegypti*, *C. teleta*, *M. musculus* (HoxB5 and HoxC5), *T. urticae*, and *T. casateneum* flanked by inverted *attB* sites. The co-injection of reintegration-ortholog vector and ϕC31 integrase can replace the y^+ sequence with a *Scr* ortholog, allowing for the identification of successfully transformed flies by the loss of y^+ and hence a yellow body marker (*y w; Scr*$^{attR\ Scr\ \text{ortholog}\ attR}$**; FIGURE 3.3**).

A fly line expressing Drosophila *Scr* coding sequence is needed for use as a control in the functional assay. The control line can be generated through the co-injection of a reintegration vector – containing Drosophila *Scr* coding sequence flanked by inverted *attB* sites – and ϕC31 integrase into syncytial blastoderm flies (*y w; Scr*$^{attR\ Scr\ attR}$). If the control fly line is viable to adulthood, the experiment can proceed.

3.1.3 Assaying the functional conservation of Sex combs reduced ortholog activities

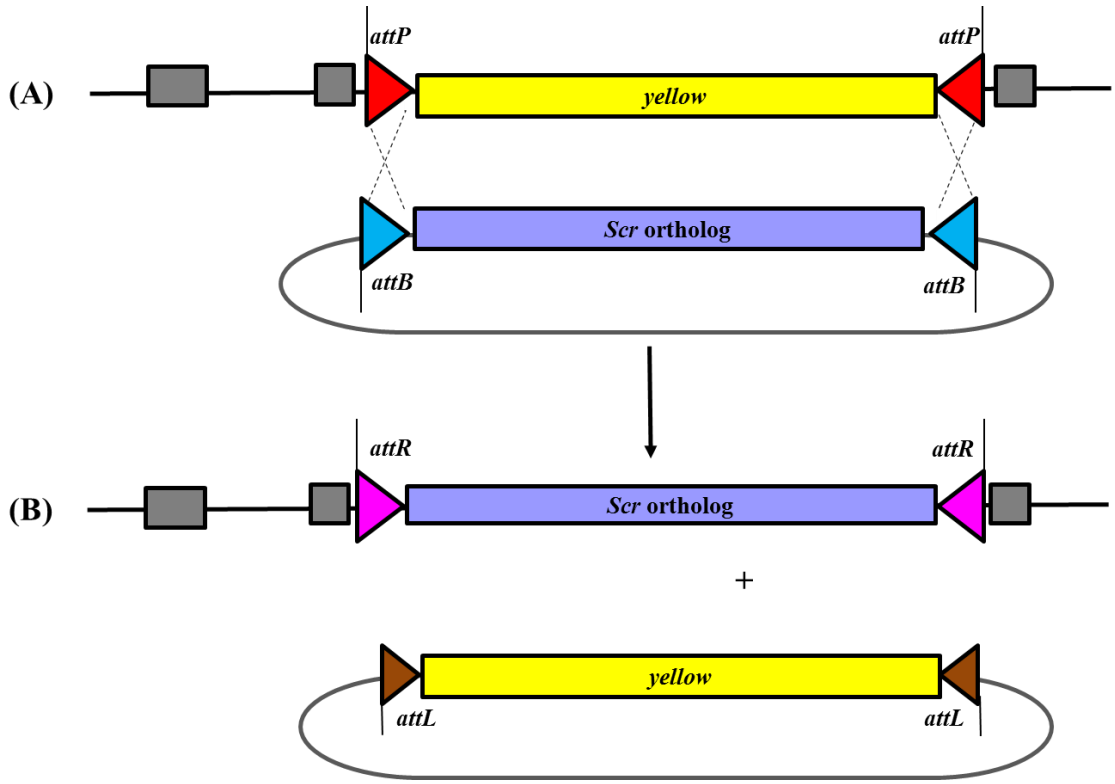
The functional assay aims to answer the question: can a given *Scr* ortholog rescue SCR-dependent labial and prothoracic phenotypes in the fly? To answer this question, SCR-dependent labial and prothoracic phenotypes can be quantified. If the *Scr* ortholog-containing Drosophila (*y w; Scr*$^{attR\ Scr\ ortholog\ attR}$) lines are homozygous viable, the following phenotypes can be quantified: (1) the number of salivary gland nuclei in larvae, (2) the number of pseudotracheal rows in the adult fly proboscis, and (3) the number of sex comb bristles on the T1 leg of adult male flies. If the *Scr* ortholog-containing Drosophila lines are not viable to adulthood the following first instar larval phenotypes can be quantified: (1) the number of T1 beard setae, and (2) the presence or absence of a labial segment to maxillary identity transformation. If *Scr* orthologs are expressed in CRISPR/Cas9 modified strain 5135 flies, FRT/FLP recombination can be used to induce clones of cells that express the *Scr* ortholog to assay the proboscis and T1 leg phenotypes in the adult fly. The controls for the functional assay are the labial and prothoracic phenotypes of the RMCE-generated fly line.

3.2 The CRISPR/Cas9 system did not induce homology-directed repair at *Sex combs reduced* in flies

3.2.1 The creation of the donor template

To create the donor template, a ligation reaction containing four DNA fragments with unique 5’ overhangs was transformed into bacteria. In the experiment that yielded the

FIGURE 3.3 Schematic of ϕC31 integrase-mediated cassette exchange at the *Sex combs reduced* locus. (A) The ϕC31 integrase (not shown) mediates the exchange of reintegration-ortholog vector DNA (containing a *Scr* ortholog flanked by inverted *attB* recombination sites) for the Drosophila body marker, *yellow*, at *attP* target sites in the fly. (B) The *Scr* locus after ϕC31 integrase-mediated cassette exchange has occurred in the fly (*y w; Scr*$^{attR\ Scr\ \text{ortholog}\ attR}$). A *Scr* ortholog has replaced the *yellow* body marker in the fly and *attP* sites have been transformed to *attR* sites in the fly genome and *attL* on the reintegration-ortholog vector.


(A)
attP
attP
yellow
Scr ortholog
attB
attB
(B)
attR
attR
Scr ortholog
+
yellow
attL
attL

correct donor template, 94 white colonies were identified by the loss of the *lacZ* gene from the vector. These 94 colonies were used in screen A, which identified 21/94 potentially correct donor templates. These 21 donor templates were then screened with screen B and C, which identified 3/21 potentially correct donors templates. Thus, of all the white colonies analyzed, only 3% (3/94) colonies tested positive in all three screens. Plasmid DNA was isolated from the three colonies that tested positive in all screens. The orientation of each of the three sequences proposed to be in the donor template was verified by restriction analysis in individual digestion reactions with *Hind*III, *BstE*II, *Bgl*II or *Xho*I (**FIGURE 3.4**).

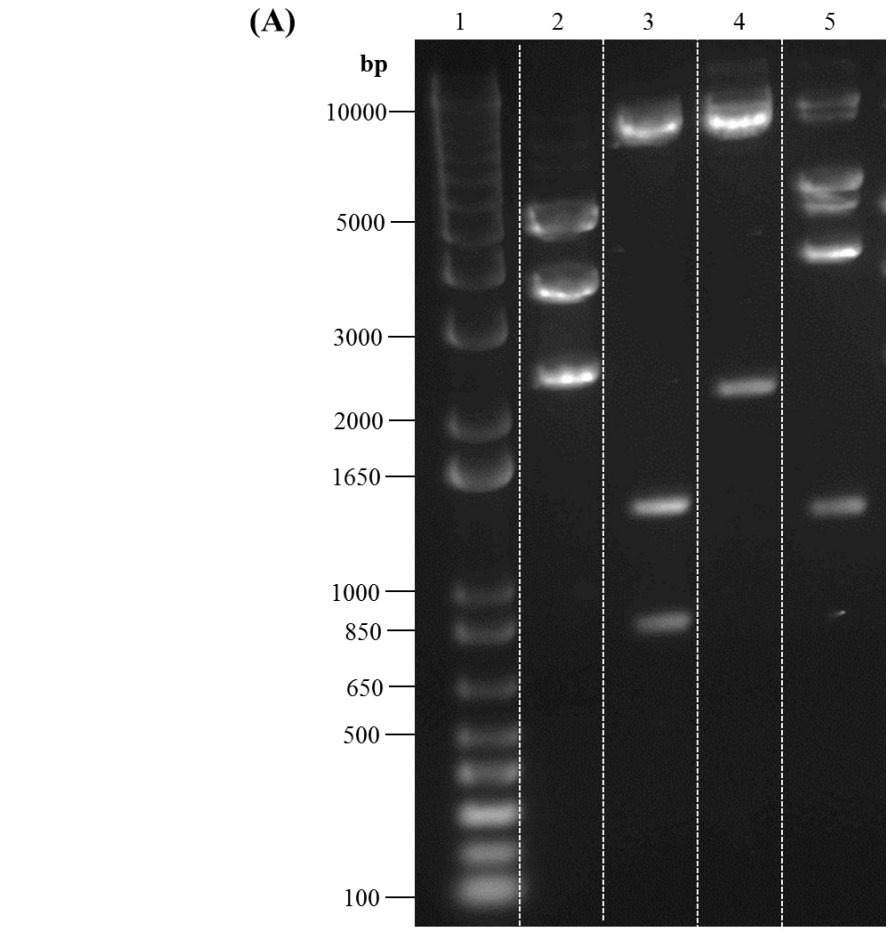
3.2.2 The CRISPR/Cas9 system did not induce homology-directed repair at *Sex combs reduced* in *y w* or strain 5135 flies

In an attempt to generate a working model fly with a manipulable *Scr* locus, *y w* and strain 5135 flies were injected with the CRISPR components, Cas9 and one or both chiRNAs, plus the donor template. Average embryonic survival rate of *y w* flies was 11.8% (94/800) with the 5' chiRNA, 8.3% (66/800) with the 3' chiRNA, and 6.6% (106/1600) with both chiRNAs (**TABLE 3.1**). Overall, average embryonic survival of *y w* flies was 8.9% (**TABLE 3.1**). The average sterility rate observed amongst *y w* flies was 48.1% (**TABLE 3.1**). Average embryonic survival rate of strain 5135 flies was 7.3% (233/3200) with the 5' chiRNA, 8.1% with the 3' chiRNA (390/4800), and 11.1% (266/2400) with both chiRNAs (**TABLE 3.1**). Overall, average embryonic survival of strain 5135 flies was 8.8% (**TABLE 3.1**). The average sterility rate observed amongst strain 5135 flies was 39.9% (**TABLE 3.1**). No F1 progeny with brown (y^+) bodies were observed and therefore, CRISPR/Cas9-mediated HDR did not occur at *Scr* in *y w* or strain 5135 in multiple trials.

3.3 The CRISPR/Cas9 system did not induce homology-directed repair at *Sex combs reduced* in Cas9-expressing flies

The inability to generate a working model fly by injecting the CRISPR components and donor template into *y w* and strain 5135 flies required a change in strategy. Transgenic *D. melanogaster* lines expressing the CRISPR nuclease Cas9 have been shown to be more

FIGURE 3.4 Restriction analysis of the donor template. (A) 1% agarose gel electrophoresis of restriction analysis of the correct donor template. Lane 1: 1 Kb Plus DNA Ladder (Invitrogen); lane 2: *BstE*II (2 377, 4 097, 5 193 bp); lane 3: *Hind*III (837, 1 358, 9 472 bp); lane 4: *Xho*I (2 123, 9 544 bp); lane 5: *Bgl*II (1 268, 3 989, 6 410 bp) with partially undigested plasmid at top. (B) Schematic of restriction analysis of the donor template. Restriction analysis of the correct donor template (11 667 bp; shown in linear form) with *BstE*II, *Hind*III, *Xho*I and *Bgl*II. Numbers in parentheses indicate the nucleotide at which the restriction enzyme cuts.


(A)
1
2
3
4
5
bp
10000
5000
3000
2000
1650
1000
850
650
500
100

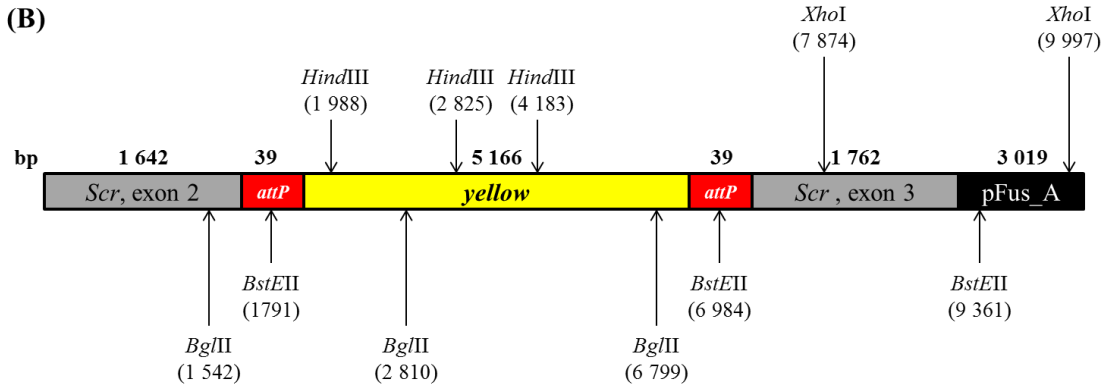

(B)
XhoI
(7 874)
XhoI
(9 997)
HindIII
(1 988)
HindIII
(2 825)
HindIII
(4 183)
bp
1 642
39
5 166
39
1 762
3 019
Scr, exon 2
attP
yellow
attP
Scr , exon 3
pFus_A
BstEII
(1791)
BstEII
(6 984)
BstEII
(9 361)
BglII
(1 542)
BglII
(2 810)
BglII
(6 799)

TABLE 3.1 Summary of CRISPR/Cas9 injection trials

Fly strain	chiRNAs	Number of survivors out of total number of embryos injected	Average survival[a]	Number of sterile survivors	Average sterility rate[b]
y w	5' and 3'	106 of 1600	8.9%	45	48.1%
	5'	94 of 800		43	
	3'	66 of 800		37	
5135	5' and 3'	266 of 2400	8.8%	90	39.9%
	5'	233 of 3200		90	
	3'	390 of 4800		184	
act-cas9	5' and 3'	14 of 200	6.2%	3	29.1%
	5'	22 of 400		9	
	3'	24 of 400		6	
nos-cas9	5' and 3'	18 of 200	8.2%	6	41.3%
	5'	18 of 200		8	
	3'	13 of 200		6	

[a]Average survival is calculated as the average of the number of survivors (i.e. surviving, CRISPR injected flies) divided by the approximate, total number of fly embryos injected in the 5' chiRNA, 3' chiRNA, and 5' & 3' chiRNA trials; [b]Average sterility rate is calculated as the average of the number of sterile survivors divided by the number of total survivors in the 5' chiRNA, 3' chiRNA, and 5' & 3' chiRNA injection trials

efficient in generating CRISPR/Cas9-mediated NHEJ and HDR mutants relative to wild-type flies (REN ET AL. 2013; GRATZ ET AL. 2014; PORT ET AL. 2014; REN ET AL. 2014A). Thus, embryos from two Cas9-expressing fly lines – act-cas9 and nos-cas9 – were injected with one or both chiRNAs and the donor template. Act-cas9 flies ubiquitously express Cas9 driven by the regulatory sequence of the *act5C* gene and nos-cas9 flies express Cas9, specifically in the germline, driven by the *nanos* promoter.

Average embryonic survival of act-cas9 flies was 5.5% (22/400) with the 5' chiRNA, 6% (24/400) with the 3' chiRNA, and 7% (14/200) with both chiRNAs (**TABLE 3.1**). Overall, average embryonic survival of act-cas9 flies was 6.2% (**TABLE 3.1**). The average sterility rate observed amongst act-cas9 flies was 29.1% (**TABLE 3.1**). Average embryonic survival of nos-cas9 flies was 9% (18/200) with the 5' chiRNA, 6.5% (13/200) with the 3' chiRNA, and 9% (18/200) with both chiRNAs (**TABLE 3.1**). Overall, average embryonic survival of nos-cas9 flies was 8.2% (**TABLE 3.1**). The average sterility rate observed amongst nos-cas9 flies was 41.3% (**TABLE 3.1**). No F1 progeny with brown (y^+) bodies were observed and therefore, CRISPR/Cas9-mediated HDR did not occur at *Scr* using act-cas9 or nos-cas9 flies.

4 DISCUSSION

4.1 The CRISPR/Cas9 system did not induce homology-directed repair at *Sex combs reduced* in Drosophila

The failure of the CRISPR/Cas9 system to induce HDR at *Scr* in Drosophila despite multiple trials indicates that the rate of HDR is extremely low. CRISPR/Cas9 relies on creating DSBs in DNA to engineer genes. Some characteristics of the *Scr* gene may have prevented its cleavage and thus, prevented modification by CRISPR/Cas9. CRISPR/Cas9-mediated HDR may not have occurred at *Scr* due to poor thermodynamic stability of the chiRNA-DNA heteroduplex, inaccessibility of the target sequence in *Scr* within the context of chromatin, or inaccessibility due to the epigenetic status of the locus (MCCLELLAND 1981; BASSETT AND LIU 2014). The fact that the sequence of bilaterian *Hox* genes, like *Scr* and its orthologs, has been so well conserved over evolution suggests

that *Hox* genes are stringently protected and as a result, may not be sensitive to modification by CRISPR/Cas9. So far, there have been no reports of CRISPR/Cas9 modification to any Drosophila *Hox* genes.

It is possible that the use of CRISPR/Cas9 to modify *Scr* may have resulted in toxicity or sterility. This is indicated by the overall, low survival (8.0%) and high sterility rate (39.6%) observed. High sterility rates are common in Drosophila CRISPR/Cas9 mutagenesis studies and can range from 5.6–78.4% (**TABLE 4.1**; BASSETT ET AL. 2013; GRATZ ET AL. 2013; SEBO ET AL. 2013; REN ET AL. 2014A; YU ET AL. 2013). Low survival, ranging from 3-11%, have also been reported in Drosophila (**TABLE 4.1**; BASSETT ET AL. 2013; GRATZ ET AL. 2014). Recent observations have suggested that overexpression of *actin*-driven *Cas9* alone can result in toxicity (PORT ET AL. 2014). This suggests that non-specific, off-target mutagenesis may occur even in the absence of a chiRNA (BASSETT AND LIU 2014). This may explain the lower survival of act-cas9 flies relative to that of *y w* or strain 5135 flies observed in this study. The average survival of nos-cas9 flies (8.2%) was slightly better than that of act-cas9 flies (6.2%), which may indicate that *nanos*-driven *Cas9* is less toxic than *actin*-driven Cas9, possibly due to more restricted expression of Cas9 in early embryos.

Besides the toxic effect of CRISPR/Cas9, low survival may also be a reflection of deleterious off-target cleavage due to the low targeting specificity of chiRNAs; although the targeting specificity of chiRNAs used in this study was not tested. Surprisingly, no relationship was found between the number of chiRNAs used in this study and the efficiency of mutagenesis. This is unlike previous reports where increasing the number of chiRNAs used from one to two, improved efficiency (GRATZ ET AL. 2013; REN ET AL. 2013). It is important to note that studies of chiRNA targeting specificity focus on NHEJ events, rather than HDR events, and therefore it is difficult to make direct comparisons between the results in this study to those in previous reports. It is unknown if the number of chiRNAs used in the injection trials in this study is related to the failure of the CRISPR/Cas9 system to induce HDR at *Scr*.

TABLE 4.1 Comparison of CRISPR/Cas9 transgenesis in the literature

Reference	Current study	Gratz et al. 2013	Bassett et al. 2013	Yu et al. 2013	Kondo and Ueda 2013	Sebo et al. 2013	Ren et al. 2014a	Port et al. 2014
Cas9 promoter	*hsp70, actin, nos*	*hsp70*	T7	Sp6	*nos*	*vasa*	*nos*	*act, nos*
Delivery	DNA injection	DNA injection	mRNA injection	mRNA injection	Transgenic	Transgenic	Transgenic	Transgenic
chiRNA promoter	*U6*	*U6*	T7	T7	*U6*	*U6*	*nos-mini, U6a, U6b,*	*U6:1, U6:2, U63*
Gene targeted	*Scr*	*yellow*	*yellow, white*	*yellow, K81, CG3708, CG9652, kl-3, light, RpL15*	*white,* neuropeptide genes (*Ast, capa, Ccap, Crz, Eh, Mip, npf*), *mir-219, mir-315*	*EGFP, mRFP*	*white*	*yellow, ebony*
Survival rate (%)[a]	6.2-8.9	50-69	nd	nd	nd	~50%	26.2	nd
Sterility rate (%)[b]	29.1-48.1	nd	26.9-62.5	5.6-78.4	N/A	45-68	7.1-70.6	nd
Mosaic survivors (%)[c]	0	6-66	4-88	35.7-80	N/A	0[f]	0[f]	0[f]
Germline transmission (%)[d]	0	5.9-20.7	0-79	35.7-100	0-100	35-71	0-100	25-100
G1 mutants (%)[e]	0	0.25-1.37	0-34.5	2.1-98.9	0-99.4	7.7-24.7	0-74.2	12.4-74.2

[a] Survival rate is defined s the number of (G_0) flies surviving after injection with the CRISPR/Cas9 components; [b] Sterility rate is defined as the number of (G_0) fertile flies as a proportion of surviving injected flies (survivors); [c] Percentage of flies that exhibit mosaic expression in the injected generation, either visibly with positive marker or using HMRA (high resolution melt analysis); [d] Proportion of surviving, fertile flies that produced at least one mutant offspring; [e] Total number of mutant G_1 offspring as a percentage of the total offspring; [f] No mosaic expression would be expected, due to germline expression of Cas9; N/A, not applicable to this technique; nd, not determined in this study. Note that all studies listed, excepting the current one, mutated flies via the NHEJ pathway (BASSETT AND LIU 2014).

Transgenic flies expressing the CRISPR nuclease Cas9 have been shown to be more efficient than *y w* flies injected with Cas9 (REN ET AL. 2013; GRATZ ET AL. 2014; PORT ET AL. 2014; REN ET AL. 2014A). In this study, CRISPR/Cas9-mediated HDR did not occur at *Scr* in *y w*, strain 5135, or in two Cas9-expressing fly lines (act-cas9 and nos-cas9). This contrasts with previous reports where a gene that could not be modified via HDR in *y w* flies was successfully modified at a rate of 70% when Cas9-expressing flies were used (GRATZ ET AL. 2014). The lack of success in modifying *Scr* by CRISPR/Cas9 in both non-Cas9 expressing (*y w* and strain 5135) and in Cas9-expressing flies (act-cas9 and nos-cas9), suggests that the failure is unrelated to the type of flies used for injection. Instead, the lack of success may be due to extremely low rates of CRISPR/Cas9-mediated HDR. If the rate of CRISPR/Cas9-mediated HDR is less than 1%, more then 100 fertile survivors are required in order to detect mutants. In these trials, 666 *y w* (141 flies) and strain 5135 (525 flies) fertile survivors were examined and therefore, it is unlikely that the lack of HDR is due to the number of survivors examined. However, only 71 fertile, Cas9-expressing – act-cas9 (42 flies) and nos-cas9 (29 flies) – survivors were examined and thus, more injection trials may be needed in order to detect a mutant if HDR is occurring an extremely low rate in Cas9-expressing flies.

The failure of CRISPR/Cas9 to induce mutations in *Hox* genes has been previously reported in the tunicate, *Ciona intestinalis*, and in the zebrafish, *Danio rerio*. Interestingly, some *Hox* genes that could not be mutated by CRISPR/Cas9 were shown to be mutable by another gene editing technology, TALENs (HWANG ET AL. 2013; TREEN ET AL. 2013; SASAKI ET AL. 2014). In *C. intestinalis*, CRISPR/Cas9 failed to introduce mutations at the *Hox12* locus, which was previously shown to be mutable by TALENs (TREEN ET AL. 2013; SASAKI ET AL. 2014). However, CRISPR/Cas9 was able to modify *Hox3* and SCR ortholog, *Hox5*, by NHEJ, which previously could not be modified by TALENs (SASAKI ET AL. 2014). This was also found in zebrafish, where CRISPR/Cas9 successfully induced mutations in genes that previously could not be altered by TALENs (HWANG ET AL. 2013). This difference in the capabilities of the two technologies may be derived from their different mechanisms of mutagenesis. CRISPR/Cas9 relies on the formation of a chiRNA-DNA duplex to direct cleavage by the Cas9 nuclease, whereas TALENs recognize their target through the formation of a protein-DNA complex and the

dimerization of the *FokI* nuclease domains (CHRISTIAN ET AL. 2010; BASSETT AND LIU 2013). It is possible that genomic loci that are resistant to one method of transgenesis may be sensitive to the other.

The inability of CRISPR/Cas9 to induce HDR at *Scr*, suggests that the failure is possibly related to aspects of the target locus itself rather than the sterility or toxicity of the CRISPR/Cas9 system. Moreover, it is unlikely that the failure is due to the inefficiency of CRISPR/Cas9 as multiple groups have proven that it is an efficient genetic engineering technology in Drosophila, despite its low survival and fertility rates (BASSETT ET AL. 2013; GRATZ ET AL. 2013; KONDO AND UEDA 2013; PORT ET AL. 2014; SEBO ET AL. 2013; REN ET AL. 2014A; YU ET AL. 2013). The majority of these studies focused on creating mutant flies by the NHEJ pathway. The studies that report successful CRISPR/Cas9 mutagenesis by the HDR pathway, find that it occurs at an extremely low rate (BAENA-LOPEZ ET AL. 2013; GRATZ ET AL. 2013; GRATZ ET AL. 2014; PORT ET AL. 2014; YU ET AL. 2014; XUE ET AL. 2014). Additionally, none of these studies report modification to any Drosophila *Hox* genes.

Some genes, like *Scr* in Drosophila, may not be sensitive to manipulation by the CRISPR/Cas9 system. In this study, manipulating the number of chiRNAs injected and changing the type of fly strain used did not improve the efficiency of CRISPR/Cas9-mediated transgenesis. Further investigation is required to understand why CRISPR/Cas9 could not modify *Scr* in Drosophila. Moreover, additional injections (with the chiRNAs and donor template) of Cas9-expressing flies may be required to detect mutants if CRISPR/Cas9-mediated HDR is occurring at an extremely low rate at *Scr*.

4.2 Limitations of the CRISPR/Cas9 system

As a relatively new technology, the CRISPR/Cas9 system is still in its infancy and its limitations are not yet fully understood. CRISPR/Cas9-mediated transgenesis relies on the creation of targeted DSBs and subsequent gene repair. Therefore, in order for CRISPR/Cas9 to be an effective gene editing tool, it must generate DSBs in both a specific and efficient manner. The rate at which DSBs are generated is determined by the targeting specificity of the CRISPR guide sequence, the chiRNA, which provides specificity by base pairing with a 20 nt complimentary sequence. Currently, the

CRISPR/Cas9 system is limited by its low targeting specificity, variable cleavage efficiency, and toxicity (BASSETT AND LIU 2014).

Specificity profiles of the CRISPR/Cas9 system are chiRNA-dependent. Off-target cleavage (and associated lethality or sterility) caused by the low targeting specificity of a chiRNA is a major limitation of the CRISPR/Cas9 system. The targeting specificity of a chiRNA is determined by a 20 nt target sequence (that starts with a guanine) and the requirement of a neighboring PAM sequence (NGG) in the genome. Recent observations suggest that these chiRNA design parameters are flexible: shortening the target sequence and introducing mismatches within it can be tolerated without increasing the frequency of off-target events or affecting mutagenesis efficiency. In fact, truncated chiRNAs, with 17 to 19 nt of target complementarity, have been shown to reduce off-target effects while retaining similar efficiency as the full-length (20 nt) chiRNAs (FU ET AL. 2014). In addition to shortening the chiRNA target sequence, some nucleotide mismatches in the target sequence can also be tolerated, but this is dependent on the number and position of the mismatch. Mismatches in the PAM-proximal region (12 nt closest to the PAM) are more likely to disrupt the chiRNA-DNA hybrid and have the greatest effect on mutagenesis efficiency (REN ET AL. 2014B). However, mismatches in the PAM-distal region (8 nt farthest from the PAM) are less critical for specificity than those in the remaining chiRNA and, in fact, up to five mismatches can be tolerated in PAM-distal nucleotides (REN ET AL. 2014B). The fact that mismatches in the PAM-distal region are more easily tolerated than those in the PAM-proximal region suggests that the requirement for a guanine at the beginning of the target sequence is somewhat relaxed (CONG ET AL. 2013; BEUMER & CARROLL 2014; REN ET AL. 2014B). In addition to chiRNA length and sequence, the requirement of target sequence adjacent to a PAM sequence (NGG) with diaguanines also appears to be flexible. Target sequences adjacent to an NAG PAM sequence can be cleaved at $1/5^{th}$ the efficiency of those adjacent to a canonical NGG PAM sequence in transformed cell lines, although this has not yet been observed in Drosophila (HSU ET AL. 2013). Together, these data indicate that chiRNA design is complex and the parameters are somewhat flexible.

In addition to chiRNA design, the efficiency of CRISPR/Cas9 transgenesis can be affected by the concentration, number, and orientation of chiRNAs. It has been observed that reducing the concentration of chiRNA injected into Drosophila embryos increases their survival, but at the cost of a drastic decrease in mutagenesis efficiency (BASSETT ET AL. 2013). More systematic testing is required to determine the optimal chiRNA concentration to yield the highest number of desired mutants without sacrificing survivorship. Mutagenesis rate can also be affected by the number of chiRNAs used to modify a gene. Improved mutagenesis efficiency has been repeatedly observed when the number of chiRNAs used was increased from one to two chiRNAs (GRATZ ET AL. 2013; REN ET AL. 2013). This improvement may reflect the fact that multiple chiRNAs provide more opportunities to induce DSBs in a target gene than a single chiRNA alone (GRATZ ET AL. 2013). If two chiRNAs are used, their relative orientation may also play a role in determining mutagenesis efficiency. Varying mutagenesis efficiencies were observed when the orientation between chiRNA pairs were manipulated (REN ET AL. 2014B). The reason for this difference in activity is unclear, but recent studies of the crystal structure of Cas9 suggest that it might be influenced by the positioning of the incoming DNA in the Cas9-chiRNA ribonucleoprotein complex (JINEK ET AL. 2014; NISHIMASU ET AL. 2014; REN ET AL. 2014B). It appears that the optimal pair of chiRNAs are those that target opposite strands of DNA – rather than the same strand of DNA – as they induce DSBs *in vitro* with the fewest off-target effects (MALI ET AL. 2013, B; RAN ET AL. 2013; REN ET AL. 2014B).

Another issue with the CRISPR/Cas9 system is variable cleavage efficiency. The efficiency of CRISPR/Cas9-mediated transgenesis varies significantly between loci and even between target sites within the same locus in Drosophila (BASSETT ET AL. 2013; GRATZ ET AL. 2013; REN ET AL. 2013; YU ET AL. 2013; GRATZ ET AL. 2014). Drosophila genes have been modified by CRISPR/Cas9 at a mutagenesis rate of 0-88% (BASSETT ET AL. 2013; GRATZ ET AL. 2013; KONDO AND UEDA 2013; PORT ET AL. 2014; SEBO ET AL. 2013; REN ET AL. 2014A; YU ET AL. 2013). Although the reason for this is unclear, it has been suggested that efficiency could be affected by secondary structures within the chiRNA, thermodynamic stability of the chiRNA-DNA duplex, or accessibility of the target sequence within the chromatin or epigenetic environment (BASSETT AND LIU

2014). Further investigation is required to determine parameters that affect cleavage efficiency.

Finally, the CRISPR/Cas9 system is limited by its toxic effect on developing Drosophila embryos. Survival can be as low as 3% when modifying wild-type Drosophila (BASSETT ET AL. 2013). Although Cas9-expressing flies exhibit improved mutagenesis efficiency, they also suffer from low survival. The overexpression of *Cas9* alone in transgenic, Cas9-expressing flies results in low survival and potentially, toxicity; this may indicate that some lethal, off-target cleavage is occurring even in the absence of chiRNA (BASSETT AND LIU 2014; PORT ET AL. 2014).

5 SUMMARY

The evolution of morphological diversity in bilaterian animals is a direct product of the evolution of developmental genetic networks (CARROLL 2005). *Hox* genes are a key player in the development of bilaterians. It is therefore beneficial to identify the evolutionary time points at which HOX protein function changed. The goal of this thesis was to determine the phylogenetic range of SCR activities, SCR^{T1} and SCR^{lab}, through a homologous replacement strategy.

I attempted to create a working, model fly using a recent transgenesis technology, the CRISPR/Cas9 system. A working, model fly has a manipulable *Scr* locus, at which phylogenetically distant *Scr* orthologs can be incorporated by RMCE to replace the endogenous *Scr* sequence. An analysis of SCR-dependent phenotypes following the expression of each *Scr* ortholog would lead to an understanding of the functional conservation of SCR^{T1} and SCR^{lab} activities.

Despite multiple trials over the course of one year, CRISPR/Cas9-mediated HDR did not generate the desired genetic modifications at *Scr* in Drosophila. The reason for this is unclear but may be due, in part, to the limitations of the CRISPR/Cas9 system. Currently, CRISPR/Cas9 is limited by its low targeting specificity, variable cleavage efficiency and toxicity (BASSETT AND LIU 2014). The failure of the system may also be due to factors at the target locus, such as chromatin environment and epigenetic status.

Functional comparisons of *Scr* orthologs provide a means to understand the conservation of SCR^{T1} and SCR^{lab} activities and the time point(s) at which SCR function changed during bilaterian evolution. The completion of a comprehensive functional analysis of *Scr* orthologs in Drosophila using a homologous replacement strategy will enable the determination of the phylogenetic range of conservation of SCR labial and prothoracic activities and identify the evolutionary time point at which SCR activity changed.

5.1 Future directions

To determine the phylogenetic range of SCR activities in Drosophila, a working model fly that can be used to express *Scr* orthologs is required. Here, CRISPR/Cas9 did not incorporate the necessary genetic modifications at *Scr* in Drosophila such that *Scr* ortholog function could be studied *in vivo*.

The targeting specificity of chiRNAs is a key determinant in the overall success of CRISPR/Cas9-mediated transgenesis. A test to determine if the 5' and 3' chiRNAs can induce DSBs at *Scr* would confirm if the chiRNAs designed here are candidates for modifying *Scr*. One previously described test involves the coinjection of chiRNAs into Cas9-expressing embryos and subsequent PCR to detect cleavage fragments (GRATZ ET AL. 2013). Targeting specificity can also be evaluated by the Surveyor assay or by Genome-wide Unbiased Identification of DSBs Evaluated by Sequencing (GUIDE-seq). The Surveyor assay utilizes the Cel-1 nuclease that recognizes and cleaves DNA mismatches that result from the hybridization of wild-type and mutant sequences (JINEK ET AL. 2013). GUIDE seq is a sensitive method for the global detection of off-target cleavage induced by CRISPR/Cas9 (TSAI ET AL. 2015). If the targeting specificity of chiRNAs used in this study is found to be low, the chiRNAs can be redesigned and tested for optimal specificity. Online tools are available to aid in the design of highly specific chiRNAs: CRISPR Optimal Target Finder, CRISPR Target, ZiFit target design tool, and E-CRISPR.

Once highly specific chiRNAs are identified, it may be worthwhile to generate flies that ubiquitously express the chiRNAs (PORT ET AL. 2014). One way to do this is by utilizing an established genetic techniques, like *P*-element mediated transgenesis or RMCE. Transgenic chiRNA flies could be injected with Cas9 and the donor template to induce

HDR at *Scr*. Alternatively, the chiRNA flies could be crossed with Cas9-expressing flies, like act-cas9 or nos-cas9, and their embryos injected with the donor template to induce HDR as previously described (PORT ET AL. 2014)

If CRISPR/Cas9 continues to fail, it would be of interest to determine if other gene editing technologies, like TALENs, could introduce the desired genetic modifications to *Scr*. Research shows that genes found to be immutable by CRISPR/Cas9 can be mutated by TALENs (HWANG ET AL. 2013; TREEN ET AL. 2013; SASAKI ET AL. 2014).

If/when a model fly with a manipulable *Scr* locus is generated, more *Scr* orthologs from other phylogenetically diverse species can be isolated and inserted into the fly by RMCE to further pinpoint the time at which SCR activity changed.

6 REFERENCES

ADAMS MD, CELNIKER SE, HOLT RA, EVANS CA, GOCAYNE JD, AMANATIDES PG, SCHERER SE, LI PW, HOSKINS RA, GALLE RF, ET AL. 2000. The genome sequence of *Drosophila melanogaster*. Science 287(5461):2185-95.

ALEXANDRE C, BAENA-LOPEZ A, VINCENT JP. 2014. Patterning and growth control by membrane-tethered *wingless*. Nature 505(7482):180-5.

ALONSO CR, MAXTON-KUECHENMEISTER J, AKAM M. 2001. Evolution of ftz protein function in insects. Curr Biol 11(18):1473-8.

ANDERSON KV AND NÜSSLEIN-VOLHARD. 1984. Information of the dorsal-ventral pattern of the Drosophila embryo is stored as maternal mRNA. Nature 311: 223-7.

BABENKO VN AND KRYLOV DM. 2004. Comparative analysis of complete genomes reveals gene loss, acquisition and acceleration of evolutionary rates in metazoa, suggests a prevalence of evolution via gene acquisition and indicates that the evolutionary rates in animals tend to be conserved. Nucleic Acids Res 32(17):5029-35.

BACHILLER D, MACIAS A, DUBOULE D, MORATA G. 1994. Conservation of a functional hierarchy between mammalian and insect *Hox/HOM* genes. EMBO J 13(8):1930-41.

BAENA-LOPEZ LA, ALEXANDRE C, MITCHELL A, PASAKARNIS L, VINCENT JP. 2013. Accelerated homologous recombination and subsequent genome modification in Drosophila. Development 140(23):4818-25.

BAENA-LOPEZ LA, ALEXANDRE C, MITCHELL A, PASAKARNIS L, VINCENT JP. 2013. Accelerated homologous recombination and subsequent genome modification in Drosophila. Development 140(23):4818-25.

BASSETT AR, TIBBIT C, PONTING CP, LIU JL. 2013. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep 4(1):220-8.

BATEMAN JR, LEE AM, WU CT. 2006. Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. Genetics 173(2):769-77.

BATESON W. 1984. Materials for the study of variation: Treated with special regard to discontinuity in the origin of species. Ithaca, NY: Macmillan and Co.

BEEMAN RW, STUART JJ, HAAS MS, DENELL RE. 1989. Genetic analysis of the homeotic gene complex (HOM-C) in the beetle *Tribolium castaneum*. Dev Biol 133(1):196-209.

BEUMER KJ AND CARROLL D. 2014. Targeted genome engineering techniques in Drosophila. Methods 68(1):29-37.

BEUMER KJ, TRAUTMAN JK, MUKHERJEE K, CARROLL D. 2013. Donor DNA utilization during gene targeting with zinc-finger nucleases. G3 (Bethesda).

BEUMER KJ, TRAUTMAN JK, BOZAS A, LIU JL, RUTTER J, GALL JG, CARROLL D. 2008. Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A 105(50):19821-6.

BHAYA D, DAVISON M, BARRANGOU R. 2011. CRISPR-cas systems in Bacteria and Archaea: Versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 45:273-97.

BIBIKOVA M, GOLIC M, GOLIC KG, CARROLL D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161(3):1169-75.

BISCHOF J AND BASLER K. 2008. Recombinases and their use in gene activation, gene inactivation, and transgenesis. Methods Mol Biol 420:175-95.

BISCHOF J, MAEDA RK, HEDIGER M, KARCH F, BASLER K. 2007. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A 104(9):3312-7.

CARROLL SB, GRENIER JK, WEATHERBEE SD. 2005. From DNA to diversity: Molecular genetics and the evolution of animal design. Oxford: Blackwell Publishing.

CERMAK T, DOYLE EL, CHRISTIAN M, WANG L, ZHANG Y, SCHMIDT C, BALLER JA, SOMIA NV, BOGDANOVE AJ, VOYTAS DF. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39(12):e82.

CHIEN S, REITER LT, BIER E, GRIBSKOV M. 2002. Homophila: Human disease gene cognates in Drosophila. Nucleic Acids Res 30(1):149-51.

CHOW RL, ALTMANN CR, LANG RA, HEMMATI-BRIVANLOU A. 1999. Pax6 induces ectopic eyes in a vertebrate. Development 126(19):4213-22.

CHRISTIAN M, CERMAK T, DOYLE EL, SCHMIDT C, ZHANG F, HUMMEL A, BOGDANOVE AJ, VOYTAS DF. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186(2):757-61.

CONG L, RAN FA, COX D, LIN S, BARRETTO R, HABIB N, HSU PD, WU X, JIANG W, MARRAFFINI LA, ET AL. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819-23.

COOPER GM AND BROWN CD. 2008. Qualifying the relationship between sequence conservation and molecular function. Genome Res 18(2):201-5.

CURTIS CD, BRISSON JA, DECAMILLIS MA, SHIPPY TD, BROWN SJ, DENELL RE. 2001. Molecular characterization of *Cephalothorax*, the *Tribolium* ortholog of *Sex combs reduced*. Genesis 30(1):12-20.

DARWIN 1859. The origin of species by means of natural selection. London: John Murray. 1959.

DECAMILLIS MA, LEWIS DL, BROWN SJ, BEEMAN RW, DENELL RE. 2001. Interactions of the *Tribolium* Sex combs reduced and Proboscipedia orthologs in embryonic labial development. Genetics 159(4):1643-8.

DEMEREC M AND KAUFMAN BP. 1996. Drosophila guide: Introduction to the genetics and cytology of *Drosophila melanogaster*. 10th ed. Carnegie Institution of Washington.

DRIEVER W AND NUSSLEIN-VOLHARD C. 1988. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. Cell 54(1):95-104.

DUBOULE D AND MORATA G. 1994. Colinearity and functional hierarchy among genes of the homeotic complexes. Trends Genet. 10(1):358–64.

DUNCAN I AND KAUFMAN TC. 1975. Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: mapping of the proximal portion of the right arm. Genetics 80(4): 733-52.

EMERALD BS AND ROY JK. 1997. Homeotic transformation in Drosophila. Nature 389(6652):684.

FRISCHER LE, HAGEN FS, GARBER RL. 1986. An inversion that disrupts the *Antennapedia* gene causes abnormal structure and localization of RNAs. Cell 47(6):1017-23.

FU Y, SANDER JD, REYON D, CASCIO VM, JOUNG JK. 2014. Improving CRISPR-cas nuclease specificity using truncated guide RNAs. Nat Biotechnol 32(3):279-84.

GALANT R AND CARROLL SB. 2002. Evolution of a transcriptional repression domain in an insect Hox protein. Nature 415(6874):910-3.

GAUNT SJ. 1988. Mouse homeobox gene transcripts occupy different but overlapping domains in embryonic germ layers and organs: a comparison of *Hox-3.1* and *Hox-1.5.* Development 103:135–144

GEOFFROY ST. HILLAIRE. 1818. Philosophie anatomique (tome premiere). Paris: J.B. Ballière. 1918.

GIBSON G, SCHIER A, LEMOTTE P, GEHRING WJ. 1990. The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. Cell 62(6):1087-103.

GILBERT SF. 2000. Developmental biology. 6th ed. Sunderland, MA: Sinauer Associates.

GLANSDORFF N, XU Y, LABEDAN B. 2008. The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. Biol Direct 3:29,6150-3-29.

GLICKSMAN MA AND BROWER DL. 1988. Misregulation of homeotic gene expression in Drosophila larvae resulting from mutations at the *Extra sex combs* locus. Dev Biol 126(2):219-27.

GRAHAM A, PAPALOPULU N, KRUMLAUF R. 1989. The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell 57(3):367–78.

GRATZ SJ, HARRISON MM, WILDONGER J, O'CONNOR-GILES KM. 2015. Precise genome editing of Drosophila with CRISPR RNA-guided Cas9. Methods Mol Biol 1311:335-48.

GRATZ SJ, UKKEN FP, RUBINSTEIN CD, THIEDE G, DONOHUE LK, CUMMINGS AM, O'CONNOR-GILES KM. 2014. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics 196(4):961-71.

GRATZ SJ, CUMMINGS AM, NGUYEN JN, HAMM DC, DONOHUE LK, HARRISON MM, WILDONGER J, O'CONNOR-GILES KM. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194(4):1029-35.

GROTH AC, FISH M, NUSSE R, CALOS MP. 2004. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics 166(4):1775-82.

HALDER G, CALLAERTS P, GEHRING WJ. 1995. Induction of ectopic eyes by targeted expression of the *eyeless* gene in Drosophila. Science 267(5205):1788-92.

HITTINGER CT, STERN DL, CARROLL SB. 2005. Pleiotropic functions of a conserved insect-specific HOX peptide motif. Development 132(23):5261-70.

HEFFER A, SHULTZ JW, PICK L. 2010. Surprising flexibility in a conserved *Hox* transcription factor over 550 million years of evolution. Proc Natl Acad Sci U S A 107(42):18040-5.

HO SN, HUNT HD, HORTON RM, PULLEN JK, PEASE LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77(1):51-9.

HSU PD, SCOTT DA, WEINSTEIN JA, RAN FA, KONERMANN S, AGARWALA V, LI Y, FINE EJ, WU X, SHALEM O, ET AL. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31(9):827-32.

HUGHES CL AND KAUFMAN TC. 2000. RNAi analysis of Deformed, Proboscipedia and Sex combs reduced in the milkweed bug *Oncopeltus fasciatus*: novel roles for *Hox* genes in the hemipteran head. Development 127(17):3683-94.

HUGHES SC AND KRAUSE HM. 2001. Establishment and maintenance of parasegmental compartments. Development 128(7):1109-18.

HWANG WY, FU Y, REYON D, MAEDER ML, TSAI SQ, SANDER JD, PETERSON RT, YEH JR, JOUNG JK. 2013. Efficient genome editing in zebrafish using a CRISPR-cas system. Nat Biotechnol 31(3):227-9.

INGHAM PW AND MARTINEZ-ARIAS A. 1992. Boundaries and fields in early embryos. Cell 68(2): 221-35.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. 2004. Finishing the euchromatic sequence of the human genome. Nature 431(7011):931-45.

IZPISUA-BELMONTE JC, FALKENSTEIN H, DOLLÉ P, RENUCCI A, DUBOULE D. 1991. Murine genes related to the Drosophila *AbdB* homeotic genes are sequentially expressed during development of the posterior part of the body. EMBO J 10(8):2279–89.

JINEK M, EAST A, CHENG A, LIN S, MA E, DOUDNA J. 2013. RNA-programmed genome editing in human cells. Elife 2:10.7554/eLife.00471.

JINEK M, CHYLINSKI K, FONFARA I, HAUER M, DOUDNA JA, CHARPENTIER E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816-21.

JOSHI R, SUN L, MANN R. 2010. Dissecting the functional specificities of two HOX proteins. Genes Dev 24(14):1533-45.

JÜRGENS G, WIESCHAUS E, NÜSSLEIN-VOLHARD C AND KLUDING H. 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the third chromosome. Roux's Arch of Dev Biol 193(5): 283-95.

KATOH K, MISAWA K, KUMA K, MIYATA T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res 30(14):3059-66.

KAUFMAN TC, LEWIS R, WAKIMOTO B. 1980. Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: The homoeotic gene complex in polytene chromosome interval 84a-B. Genetics 94(1):115-33.

KNOLL AH AND CARROLL SB. 1999. Early animal evolution: Emerging views from comparative biology and geology. Science 284(5423):2129-37.

KONDO S AND UEDA R. 2013. Highly improved gene targeting by germline-specific Cas9 expression in drosophila. Genetics 195(3):715-21.

LAPPIN TR, GRIER DG, THOMPSON A, HALLIDAY HL. 2006. HOX genes: Seductive science, mysterious mechanisms. Ulster Med J 75(1):23-31.

LEMOTTE PK, KUROIWA A, FESSLER LI, GEHRING WJ. 1989. The homeotic gene *Sex combs reduced* of Drosophila: gene structure and embryonic expression. EMBO J 8(1):219-27.

LEWIS EB. 1978. A gene complex controlling segmentation in Drosophila. Nature 276(5688):565-70.

LEWIS RA, KAUFMAN TC, DENELL RE, TALLERICO P. 1980. Genetic analysis of the Antennapedia gene complex (ANT-C) and adjacent segments 84B-D. Genetics 95(2):367-81.

LEWIS RA, WAKIMOTO BT, DENELL RE, KAUFMAN TC. 1980b. Genetic analysis of the Antennapedia gene complex (ANT-C) and adjacent chromosomal regions of *Drosophila melanogaster*. II. polytene chromosome segments 84A-84B1,2. Genetics 95(2):383-97.

LIU J, LI C, YU Z, HUANG P, WU H, WEI C, ZHU N, SHEN Y, CHEN Y, ZHANG B, ET AL. 2012. Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. J Genet Genomics 39(5):209-15.

LOHR U, YUSSA M, PICK L. 2001. Drosophila *fushi tarazu*: a gene on the border of homeotic function. Curr Biol 11(18):1403-12.

MAHAFFEY JW AND KAUFMAN TC. 1987. Distribution of the *Sex combs reduced* gene products in *Drosophila melanogaster*. Genetics 117(1):51-60.

MALI P, ESVELT KM, CHURCH GM. 2013A. Cas9 as a versatile tool for engineering biology. Nat Methods 10(10):957-63.

MALICKI J, CIANETTI LC, PESCHLE C, MCGINNIS W. 1992. A human HOX4B regulatory element provides head-specific expression in Drosophila embryos. Nature 358(6384):345-7.

MARTINEZ-ARIAS A AND LAWRENCE PA. 1985. Parasegments and compartments in the Drosophila embryo. Nature 313(6004):639-42.

MCCLELLAND M. 1981. The effect of sequence specific DNA methylation on restriction endonuclease cleavage. Nucleic Acids Res 9(22):5859-5866.

MCGINNIS N, KUZIORA MA, MCGINNIS W. 1990. Human *Hox-4.2* and Drosophila *Deformed* encode similar regulatory specificities in Drosophila embryos and larvae. Cell 63(5):969-76.

MCGINNIS W AND KRUMLAUF R. 1992. Homeobox genes and axial patterning. Cell 68(2):283-302.

MCGINNIS W, GARBER RL, WIRZ J, KUROIWA A, GEHRING WJ. 1984. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell 37(2):403-8.

MENDEL, GJ. 1866. Versuche über Pflanzen-Hybriden [Experiments Concerning Plant Hybrids]. Verhandlungen des naturforschenden Vereines in Brünn [Proceedings of the Natural History Society of Brünn] IV(1865): 3-47.

MERABET S, LITIM-MECHERI I, KARLSSON D, DIXIT R, SAADAOUI M, MONIER B, BRUN C, THOR S, VIJAYRAGHAVAN K, PERRIN L, ET AL. 2011. Insights into HOX protein function from a large scale combinatorial analysis of protein domains. PLoS Genet 7(10):e1002302.

MINELLI A AND FUSCO G. 2004. Evo-devo perspectives on segmentation: Model organisms, and beyond. Trends Ecol Evol 19(8):423-9.

NASSIF N, PENNEY J, PAL S, ENGELS WR, GLOOR GB. 1994. Efficient copying of nonhomologous sequences from ectopic sites via *P*-element-induced gap repair. Mol Cell Biol 14(3):1613-25.

NEGRE B, CASILLAS S, SUZANNE M, SANCHEZ-HERRERO E, AKAM M, NEFEDOV M, BARBADILLA A, DE JONG P, RUIZ A. 2005. Conservation of regulatory sequences and gene expression patterns in the disintegrating Drosophila *Hox* gene complex. Genome Res 15(5):692-700.

NISHIMASU H, RAN FA, HSU PD, KONERMANN S, SHEHATA SI, DOHMAE N, ISHITANI R, ZHANG F, NUREKI O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156(5):935-49.

NÜSSLEIN-VOLHARD C. SUBTELNEY G AND KOENIGSBERG IR, editors. 1979. Maternal effect mutations that alter the spatial coordinates of the embryo of *Drosophila melanogaster.* In determination of spatial organization. New York: Academic Press. p. 185-211.

NÜSSLEIN-VOLHARD C AND WEICHAUS E. 1980. Mutations affecting segment number and polarity in Drosophila. Nature 287: 795-801.

NÜSSLEIN-VOLHARD C, WEICHAUS E AND KLUDING H. 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. I. Zygotic loci on the second chromosome. Roux's Arch Dev Biol 193(5): 267-82.

PATTATUCCI AM, OTTESON DC, KAUFMAN TC. 1991. A functional and structural analysis of the *Sex combs reduced* locus of *Drosophila melanogaster*. Genetics 129(2):423-41.

PEDERSON JD, KIEHART DP, MAHAFFEY JW. 1996. The role of HOM-C genes in segmental transformations: reexamination of the Drosophila *Sex combs reduced* embryonic phenotype. Dev Biol 180(1):131-42.

PERCIVAL-SMITH A, SIVANANTHARAJAH L, PELLING JJ, TEFT WA. 2013. Developmental competence and the induction of ectopic proboscises in *Drosophila melanogaster*. Dev Genes Evol 223(6):375-87.

PERCIVAL-SMITH A, WEBER J, GILFOYLE E, WILSON P. 1997. Genetic characterization of the role of the two HOX proteins, Proboscipedia and Sex combs reduced, in determination of adult antennal, tarsal, maxillary palp and proboscis identities in *Drosophila melanogaster*. Development 124(24):5049-62.

PORT F, CHEN HM, LEE T, BULLOCK SL. 2014. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S A 111(29):E2967-76.

PREISS A, ROSENBERG UB, KIENLIN A, SEIFERT E, JACKLE H. 1985. Molecular genetics of kruppel, a gene required for segmentation of the Drosophila embryo. Nature 313(5997):27-32.

PRUD'HOMME B, MINERVINO C, HOCINE M, CANDE JD, AOUANE A, DUFOUR HD, KASSNER VA, GOMPEL N. 2011. Body plan innovation in treehoppers through the evolution of an extra wing-like appendage. Nature 473(7345):83-6.

REITER LT, POTOCKI L, CHIEN S, GRIBSKOV M, BIER E. 2001. A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. Genome Res 11(6):1114-25.

REN X, YANG Z, MAO D, CHANG Z, QIAO HH, WANG X, SUN J, HU Q, CUI Y, LIU LP, ET AL. 2014A. Performance of the Cas9 nickase system in *Drosophila melanogaster*. G3 (Bethesda) 4(10):1955-62.

REN X, YANG Z, XU J, SUN J, MAO D, HU Y, YANG SJ, QIAO HH, WANG X, HU Q, ET AL. 2014B. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. Cell Rep 9(3):1151-62.

REN X, SUN J, HOUSDEN BE, HU Y, ROESEL C, LIN S, LIU LP, YANG Z, MAO D, SUN L, ET AL. 2013. Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9. Proc Natl Acad Sci U S A 110(47):19012-7.

RILEY PD, CARROLL SB, SCOTT MP. 1987. The expression and regulation of Sex combs reduced protein in Drosophila embryos. Genes Dev 1(7):716-30.

ROGERS BT, PETERSON MD, KAUFMAN TC. 1997. Evolution of the insect body plan as revealed by the Sex combs reduced expression pattern. Development 124(1):149-57.

RUBIN GM AND LEWIS EB. 2000. A brief history of Drosophila's contributions to genome research. Science 287(5461):2216-8.

RYAN JF, PANG K, SCHNITZLER CE, NGUYEN AD, MORELAND RT, SIMMONS DK, KOCH BJ, FRANCIS WR, HAVLAK P, NISC COMPARATIVE SEQUENCING PROGRAM, ET AL. 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science 342(6164):1242592.

SASAKI H, YOSHIDA K, HOZUMI A, SASAKURA Y. 2014. CRISPR/Cas9-mediated gene knockout in the ascidian *Ciona intestinalis*. Dev Growth Differ 56(7):499-510.

SATO T, HAYES PH, DENELL RE. 1985. Homeosis in Drosophila: roles and spatial patterns of expression of the *Antennapedia* and *Sex combs reduced* loci in embryogenesis. Dev Biol 111(1):171-92.

SCHNEUWLY S, KLEMENZ R, GEHRING WJ. 1987. Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene *Antennapedia*. Nature 325(6107):816-8.

SCOTT MP AND WEINER AJ. 1984. Structural relationships among genes that control development: Sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of Drosophila. Proc Natl Acad Sci U S A 81(13):4115-9.

SEBO ZL, LEE HB, PENG Y, GUO Y. 2014. A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. Fly (Austin) 8(1):52-7.

SEO HC, EDVARDSEN RB, MAELAND AD, BJORDAL M, JENSEN MF, HANSEN A, FLAAT M, WEISSENBACH J, LEHRACH H, WINCKER P, ET AL. 2004. *Hox* cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. Nature 431(7004):67-71.

SIVANANTHARAJAH L AND PERCIVAL-SMITH A. 2014. Differential pleiotropy and HOX functional organization. Dev Biol 398(1):1-10.

SIVANANTHARAJAH L AND PERCIVAL-SMITH A. 2009. Analysis of the sequence and phenotype of Drosophila *Sex combs reduced* alleles reveals potential functions of conserved protein motifs of the Sex combs reduced protein. Genetics 182(1):191-203.

SMALL KM AND POTTER SS. 1993. Homeotic transformations and limb defects in Hox A11 mutant mice. Genes Dev 7(12A):2318-28.

SPRADLING AC AND RUBIN GM. 1982. Transposition of cloned *P*-elements into Drosophila germ line chromosomes. Science 218(4570):341-7.

STRUHL G. 1982. Genes controlling segmental specification in the Drosophila thorax. Proc Natl Acad Sci U S A 79(23):7380-4.

STRUHL G. 1981. A gene product required for correct initiation of segmental determination in Drosophila. Nature 293(5827):36-41.

SWEETON D, PARKS S, COSTA M, WIESCHAUS E. 1991. Gastrulation in Drosophila: the formation of the ventral furrow and posterior midgut invaginations. Development 112(3):775-89.

TELFORD MJ. 2000. Evidence for the derivation of the Drosophila *fushi tarazu* gene from a *Hox* gene orthologous to lophotrochozoan *Lox5*. Curr Biol 10(6):349-52.

THORPE HM AND SMITH MC. 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc Natl Acad Sci U S A 95(10):5505-10.

TOMANCAK P, BEATON A, WEISZMANN R, KWAN E, SHU S, LEWIS SE, RICHARDS S, ASHBURNER M, HARTENSTEIN V, CELNIKER SE, ET AL. 2002. Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol 3(12):RESEARCH0088.

TOMANCAK P, BERMAN BP, BEATON A, WEISZMANN R, KWAN E, HARTENSTEIN V, CELNIKER SE, RUBIN GM. 2007. Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol 8(7):R145.

TREEN N, YOSHIDA K, SAKUMA T, SASAKI H, KAWAI N, YAMAMOTO T, SASAKURA Y. 2014. Tissue-specific and ubiquitous gene knockouts by TALEN electroporation provide new approaches to investigating gene function in *Ciona*. Development 141(2):481-7.

TSAI SQ, ZHENG Z, NGUYEN NT, LIEBERS M, TOPKAR VV, THAPAR V, WYVEKENS N, KHAYTER C, IAFRATE AJ, LE LP, ET AL. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-cas nucleases. Nat Biotechnol 33(2):187-97.

TURNER FR AND MAHOWALD AP 1979. Scanning electron microscopy of *Drosophila melanogaster* embryogenesis. III. Formation of the head and caudal segments. Dev Biol 68(1): 96-109.

VENKEN KJ AND BELLEN HJ. 2007. Transgenesis upgrades for *Drosophila melanogaster*. Development 134(20):3571-84.

VENKEN KJ, SCHULZE KL, HAELTERMAN NA, PAN H, HE Y, EVANS-HOLM M, CARLSON JW, LEVIS RW, SPRADLING AC, HOSKINS RA, ET AL. 2011. MiMIC: A highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. Nat Methods 8(9):737-43.

WAKIMOTO BT AND KAUFMAN TC. 1981. Analysis of larval segmentation in lethal genotypes associated with the *Antennapedia* gene complex (ANT-C) in *Drosophila melanogaster*. Dev Biol 81(1):51-64.

WAKIMOTO BT, TURNER FR, KAUFMAN TC. 1984. Defects in embryogenesis in mutants associated with the antennapedia gene complex of *Drosophila melanogaster*. Dev Biol 102(1):147-72.

WANG QT. 2012. Epigenetic regulation of cardiac development and function by Polycomb group and Trithorax group proteins. Dev Dyn 241(6):1021-33.

WANG H, YANG H, SHIVALILA CS, DAWLATY MM, CHENG AW, ZHANG F, JAENISCH R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153(4):910-8.

WEISCHAUS E, NÜSSLEIN-VOLHARD C, AND JÜRGENS G. 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. III. Zygotic loci on the X-chromosome and fourth chromosome. Roux's Arch Dev Biol 193(5): 296-307.

XUE Z, REN M, WU M, DAI J, RONG YS, GAO G. 2014. Efficient gene knock-out and knock-in with transgenic Cas9 in Drosophila. G3 (Bethesda) 4(5):925-9.

YU Z, REN M, WANG Z, ZHANG B, RONG YS, JIAO R, GAO G. 2013. Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. Genetics 195(1):289-91.

ZHAO JJ, LAZZARINI RA, PICK L. 1993. The mouse *Hox-1.3* gene is functionally equivalent to the Drosophila *Sex combs reduced* gene. Genes Dev 7(3):343-54.

APPENDICES

APPENDIX 1.1 Accession numbers of species used in the Sex combs reduced ortholog alignment

Genus species	Common name	Accession number	Database
Acromyrmex echinatior	Panamanrian leafcutter ant	XP_011056349 (partial)	NCBI Protein
Acyrthosiphon pisum	Pea aphid	XP_008182528	NCBI Protein
*Aedes aegypti**	Yellow fever mosquito	XP_001660497	UniProt/TrEMBL
Anopheles gambiae	Malaria mosquito	AAC31944	UniProtKB
Apis mellifera	Western honey bee	XP_623903	NCBI Protein
Balanoglossus misakiensis	Acorn worm	BAH23874	UniProtKB
Bombyx mori	Silkmoth	NP_001037339	NCBI Protein
Branchiostoma lanceolatum	Common lancelet	ACJ74385	UniProtKB
*Capitella teleta **	Polychaete worm	ABY67956	UniProtKB
Cerapachys biroi	Clonal raider ant	EZA52179	NCBI Protein
Ceratitis capitata	Medfly	W8B8R9	UniProtKB
Ciona intestinalis	Vase tunicate/Sea squirt	NP_001027665	UniProtKB
Daphnia magna	Freshwater water flear	BAJ05331.1	UniProtKB
Drosophila ananassae	Fruit fly	XP_001953504	UniProtKB
Drosophila erecta	Fruit fly	XP_001979124	UniProtKB
Drosophila grimshawi	Fruit fly	XP_001990655	UniProtKB
*Drosophila melanogaster**	Common fruit fly	AAS65103	UniProtKB
Drosophila mojavensis	Fruit fly	XP_001999776	UniProtKB
Drosophila persimilis	Fruit fly	XP_002016960	UniProtKB
Drosophila pseudoobscura	Fruit fly	XP_001359213	UniProtKB
Drosophila sechellia	Fruit fly	XP_002038616	UniProtKB
Drosophila simulans	Fruit fly	XP_002102414	UniProtKB
Drosophila willistoni	Fruit fly	XP_002070807	UniProtKB
Drosophila yakuba	Fruit fly	XP_002096729	UniProtKB
Euperipatoides kanangrensis	Velvet worm	CCK73373	UniProtKB
Folsomia candida	Springtail	AAK51914 (partial)	UniProtKB
Gibbula varia	Sea snail	HM136797	UniProtKB
Haliotis rufescens	Red abalone	AAF78248	UniProtKB
Herdmania curvata	Tunicate	AAF60347	UniProtKB
Ixodes scapularis	Deer tick	XP_002406405	UniProtKB
Metacrinus rotundus	Japanese sea lily	BAF43724	UniProtKB
Mus musculus (HoxA5)	House mouse	NP_034583	UniProtKB
*Mus musculus (HoxB5)**	House mouse	NP_032294	UniProtKB
*Mus musculus (HoxC5)**	House mouse	P32043	UniProtKB
Nasonia vitripennis	Parasitic wasp	NP_001128396	NCBI Protein
Oncopeltus fasciatus	Milkweed bug	ACZ60640	UniProtKB
Parhyale hawaiensis	Amphipod crustacean	AGC12527	UniProtKB

APPENDIX 1.1 Accession numbers of species used in the Sex combs reduced ortholog alignment (continued)

Genus species	Common name	Accession number	Database
Peronella japonica	Sand dollar	BAO57698	UniProtKB
Petromyzon marinus	Sea lamprey	AFZ94990	UniProtKB
Ptychodera flava	Acorn worm	AAR07636	UniProtKB
Publilia modesta	Modest treehopper	ADZ56089	UniProtKB
Rhodnius prolixus	Assassin bug	ACN43631	UniProtKB
Saccoglossus kowalevskii	Acorn worm	NP_001158410	UniProtKB
Schistocerca gregaria	Desert locust	CAA52159 (partial)	UniProtKB
Symsagittifera roscoffensis	Mint-sauce worm	ACM69152	UniProtKB
*Tetranychus urticae**	Red spider mite	tetur20g02540	UniProtKB
*Tribolium castaneum**	Red flour beetle	NP_001034523	UniProtKB
Tubifex tubifex	Sludge worm	BAN14798	UniProtKB
Zootermopsis nevadensis	Dampwood termite	KDR19415	UniProtKB

Species are listed in alphabetical order; all accession numbers refer to a full length SCR protein sequence unless otherwise indicated; an asterisk (*) indicates a species from which a *Scr* ortholog was isolated for use in reintegration-ortholog vector cloning

APPENDIX 1.2 Summary of conserved regions of Drosophila Sex combs reduced protein

SCR region	Sequence	Length (aa)	AA position in SCR	Conservation
Octapeptide	MSSYQFVNS	9	8-16	Bilateria
Homeodomain	TKRQRTSYTRYQTLELEKEFHFNRYLTRRRR IEIAHALCLTERQIKIWFQNRRMKWKKEH	60	324-383	Bilateria
LASCY	LASCY	5	17-21	Protostomia
YTPNL	YTPNLYPNTPQAHYANQ	17	82-98	Diptera
DYTQL	MVDYTQLPQLRL	12	109-120	Insecta
SCKY	NSCKYA	6	164-170	Arthropoda
NDPVT	NDPVTPGGSGGGG	13	171-183	Drosophila
QSLAS	QSLAS	8	198-205	Drosophila
PQDL	SPQDLSTR	8	206-213	Arthropoda
DISPK	DISPKLSPSSVVESVARSL	18	214-232	Diptera
VNVPM	VNVPMHSPGGGDSDSES	17	266-282	Drosophila
NEAGS	DSGNEAGSSQ	10	283-292	Diptera
YPWM	PQIYPWMKRVHLGTS	4 to 11	302-316	Bilateria
NANGE	TVNANGE	7	317-323	Arthropoda
KMAS	KMASMN	5	384-389	Bilateria
C-terminal domain	IVPYHMGPYGHPYHQFDIHSQFAHLSA	27	389-417	Insecta

aa/AA, amino acid

APPENDIX 2 Primers used to isolate DNA fragments used in the construction of the donor template

Amplicon	Amplification Oligonucleotides (5' to 3')		PCR template
	Forward	Reverse	
Scr , exon 2	X2-*Scr*-*Bsa* I-F: CAGCTAGGTCTCGCTATACCTGGG GGCAAGTTTACAATATTTC	X2-*Scr*-*attP* -*Bsa* I-R: CAGCTAGGTCTCCCATGCCCCCA ACTGAGAGAACTCAAAGGTTACC CCAGTTGGGGTAAAGCCAGGGGT CGTTGTCGTG	*y w*⁻ *D. melanogaster* genomic DNA
yellow (y+)	Y-*Bsa* I-F: CAGCTAGGTCTCCCATGCGACTAT TAAATGATTATCGCC	Y-*Bsa* I-R: CAGCTAGGTCTCGGTCCTCGACCT GCAGGTCAACGGATC	MiMIC (Venken et al. 2011)
Scr , exon 3	X3-*Scr*-*Bsa* I-*attP* -F: CAGCTAGGTCTCGGGACCCCCCA ACTGAGAGAACTCAAAGGTTACC CCAGTTGGGGGACGCGTGGCACT TTTCGGGTAC	X3-*Scr*-*Bsa* I-R: CAGCTAGGTCTCGCGCCGATTTGT TTTCTCTAAAATT	*y w*⁻ *D. melanogaster* genomic DNA

Underlined sequence delineates the sequence of the *attP* recombination site

APPENDIX 3 Primers used in colony PCR to identify correct, donor templates

Screen	Amplification Oligonucleotides (5' to 3')		Length of amplicon (bp)
	Forward	Reverse	
A	CTGGCAGTTCCCTACTCTCG	GGTAAATCAGCGGGCTGCGTTCG	1936
B	CAGGGAAAGTTCAACTTAATCGC	CTGTCCTGGCTGGTCTAGACGTC	5166
C	Same as Screen B, forward (above)	GAGCCGCCACCAATTGGACC	329

bp, base pairs

APPENDIX 4 Oligonucleotides used in the cloning of pU6-chiRNAs

chiRNA	Oligonucleotides (5' to 3')	
	Forward	**Reverse**
5' chiRNA	chiRNA-X2-F: CTTCGATTTTTGAATTTATGGCAA	chiRNA-X2-R: AAACTTGCCATAAATTCAAAAATC
3' chiRNA	chiRNA-X3-F: CTTCGCGTGGCACTTTTCGGGTAC	chiRNA-X3-R: AAACGTACCCGAAAAGTGCCACGC

APPENDIX 5 Oligonucleotides used in the cloning of the reintegration vector

Oligonucleotides (5' to 3')	
Forward	**Reverse**
CTAGGAATTCAGATCTGCGGCCGCCTCGAG	AGCTCTCGAGGCGGCCGCAGATCTGAATTC

APPENDIX 6 Oligonucleotides used to isolate *Sex combs reduced* orthologs

Species	Amplified region	Restriction site added	Oligonucleotides (5' to 3')	
			Forward	Reverse
Aedes aegypti	*Scr* , exon 2	*Not* I	CAGCTAGCGGCCGCATGGATGCCAATAATTTTGTG	CGTTTACTGTACTCTCGGGCATATGCACTCGTTTC
	Scr , exon 3		CATATCGGGCAGAGTACAGTAAACGCCAATGGAG	CAGCTAGCGGCCGCCTATGCACTAAGATGCGCAAATTG
Capitella teleta	*Hox5* , exon 1	*Not* I	CAGCTAGCGGCCGCATGAGTTCGTACTTTGTGAATTC	CATTCGATGTATCGTGTCCGATATGCATCCGC
	Hox5 , exon 2		ATATCGGACACGATACATCGAATGCAGATAACAAG	CAGCTAGCGGCCGCCAATTATGTGGTCACTATG
Mus musculus	*HoxB5* , exon 1	*Not* I	CAGCTAGCGGCCGCATGAGCTCGTACTTTGTAAAC	GCCCAGTCATATCGTGGCTGATGTGAAGCTTC
	HoxB5 , exon 2		ACATCAGCCACGATATGACTGGGCCAGACGGAAAAAG	CAGCTAGCGGCCGCTCAAGGTTGGAAGGCGCTGC
	HoxC5 , exon 1	*BamH* I	CAGCTAGGATCCATGAGCTCCTACGTAGCCAATTC	TGCCATCCGTCTCGTGGCTCATGTGCAGTTTG
	HoxC5, exon 2		ACATGAGCCACGAGACGGATGGCAAGCGGTCC	CAGCTAGGATCCTAAAGAGCTTCTTTGCTCTTC
Tetranychus urticae	*Scr* , exon 1	*Not* I	CAGCTAGCGGCCGCATGAGCTCTTATCAATTTGTTAATC	TTGCATTGACTCTTTGACCGACATGTACTTTC
	Scr , exon 2		ATGTCGGTCAAAGAGTCAATGCAATGGGTGAAAC	CAGCTAGCGGCCGCTTATGCTTTTGTTTCTCCATG
Tribolium castaneum	*Cx* , exon 1	*Not* I	CAGCTAGCGGCCGCATGAGCTCCTACCAGTTCGTC	CGTTCACAGTACTTTGGCCGAGATGAACTCTC
	Cx exon 2		ATCTCGGCCAAAGTACTGTGAACGCAAATGGC	CAGCTAGCGGCCGCCTAAGTAGCGAGATGGGCAAATTG

APPENDIX 7 Summary of errors in the reintegration-ortholog vectors identified by BLAST

Species of *Scr* Ortholog	Length of *Scr* CDS in vector	Position of mutated nucleotide in CDS[a]	Nucleotide mismatch	AA change	SCR protein length (aa)	AA position in SCR protein	Mutation type	Variant
Drosophila melanogaster	1254	15 345 747 933	T to C A to G C to T	Cys to Cys Leu to Leu Gly to Gly Val to Val	417	5 115 249 311	Silent Silent Silent Silent	*Sex combs reduced* , Isoform A [ref|NP_524248.2]
Aedes aegypti	1113	616 649 801 803 804 821	A to G A to T C to G G to C G to C C to A	Asn to Asp Asn to Tyr Ile to Met Gly to Pro Gln to Glu Ala to Asp	370	206 217 267 268 268 274	Missense Missense Missense Missense Missense Missense[c]	No No No No No No
Tribolium castaneum	939	247	C to T	Pro to Ser	312	83	Missense	*Cx-E* allele, exon 1 [gb|AY057859.1|AT057859S1]
Tetranychus urticae	1050	432	G to A	Glu to Glu	349	144	Silent	No
Capitella teleta	911[b]	608 858	C to T T to C	Ala to Val His to His	298	203 286	Missense Silent	No No
Mus musculus, HoxB5	810	498	G to A	Thr to Thr	269	166	Silent	No
Mus musculus, HoxC5	668	94 125	A to G G to A	Arg to Gly Gly to Glu	222	32 42	Missense Missense	No No

[a] position of nucleotide mismatch is relative to the start codon (ATG); [b] The CDS of *C. teleta* (897 nt) includes 14 nt of 3' UTR); [c] This missense mutation falls within a conserved region of SCR, the NANGE motif (note that all other silent and missense mutations for all orthologs listed do not fall within conserved SCR regions); CDS, coding sequence; aa/AA, amino acid

APPENDIX 8 Sources of genomic DNA used in the isolation of *Sex combs reduced* orthologs

Species	Person(s)	Institution	Sample donated
Aedes aegypti	Dr. Fiona Hunter	Brock University (Saint Catherines, Ontario, Canada)	frozen specimens
Capitella teleta	Dr. Valery Forbes	University of Nebraska-Lincoln (Lincoln, Nebraska, USA)	frozen specimens
	Dr. Elaine Seaver	Whitney Laboratory for Marine Bioscience (St. Augustine, Florida, USA)	
Mus musculus	Dr. Kathleen Hill	University of Western Ontario (London, Ontario, Canada)	genomic DNA
Tetranychus urticae	Drs. Vojislava and Miodrag Grbic	University of Western Ontario (London, Ontario, Canada)	live specimens
Tribolium castaneum	Dr. Paul Fields	Agriculture and Agri-Food Canada (London, Ontario, Canada)	live specimens

CURRICULUM VITAE

Name	Laura E. Garofalo
Post-secondary Education and Degrees	University of Toronto Mississauga, Ontario, Canada 2007-2011 **H.B.Sc.**
	University of Western Ontario London, Ontario, Canada 2013-2015 **M.Sc.**
Honors and Awards	Western Graduate Research Scholarship 2013-2015
	Graduation with Distinction (University of Toronto) 2011
Related Work Experience	Teaching Assistant (Genetics in Everyday Life, Scientific Methods in Biology, Plants as a Human Resource, Advanced Genetics, Education in the Life Sciences) University of Western Ontario 2013-2015
Academic Associations	The Society of Biology Graduate Students of the University of Western Ontario (SOBGS) The Society of Graduate Students of the University of Western Ontario (SOGS)