SEQUENCING AND CHARACTERIZATION OF A MATERNAL-EFFECT SEX DETERMINING AUTOSOMAL INVERSION IN THE HESSIAN FLY

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Karen A. Vellacott-Ford

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Jeffrey Stuart, Chair

Department of Entomology

Dr. Jody Banks

Department of Botany

Dr. Michael Gribskov

Department of Biology

Dr. Steven Scofield

Department of Agronomy

Approved by:

Dr. Stephen Cameron

Head of the graduate program

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Jeff Stuart, for his support, feedback, and extreme patience over the many years I spent on this work. Although my late co-advisor, Rich Shukle, sadly could not be with us in the last few years, his positive attitude and advice were invaluable to me during the difficult early phase of my project. I had the impression that he would directly say whatever he was thinking, yet he managed to be tactful and not laugh too much when I brought up some of my more ridiculous ideas.

I almost certainly would not have made it through this PhD program without my excellent committee members: Jody Banks, Mike Gribskov, Steve Scofield, and the late Virginia Ferris. In addition to providing insight and feedback related to my research and presentations, they each spent extra time listening to my concerns about various problems and giving helpful advice on both science and life in general. Im sad that Virginia Ferris is no longer with us but thankful for the memories of our conversations and her great stories.

I have been fortunate to work with people who not only did great work but were supportive and helpful to everyone around them. Chaoyang Zhao patiently taught me all about working with the Hessian fly when I joined the lab and was understanding of my mistakes. Lucio Navarro Escalante also took the time to help me in the lab and give me advice while always maintaining a positive attitude. Whenever I feel pressure to act like a professional, I pause and try to imagine what Lucio would do. Andrew Katz, who worked in our lab as an undergrad, helped me set up many fly crosses for my experiments. He was enthusiastic about the science and kindly tolerated my stream-of-consciousness explanations. Verónica Campos Medina had to spend the most time with me, as we started our PhD work together and shared an office. She helped me to deal with many stressful moments in my life and I often benefited from her wisdom and advanced emotional intelligence. Whenever I wanted to start an argument for fun, she was there for me. Sue Cambron helped me with all of my greenhouse-related problems and made me feel comfortable taking to her about anything.

Shubha Subramanyam has given me a lot of great advice and I've enjoyed our conversations about science and many other topics. Alisha Johnson was also helpful to me, including with finding a work-around for my bizarre primer issue.

I owe Andrew Schneider thanks for more than I can write about in this space, including discussing my project with me many times, helping me learn to code, and helping me navigate both grad school and life in general.

Finally, I must thank my family and friends for their love and support. My parents and my twin sister, Suzanna, have been especially helpful in the last couple of years and have gone out of their way to help me minimize the stress in my life. Suzanna's husband, Kevin Coppock, has also been extremely helpful and great at making cookies. I would like to thank Bradley Smith for emotional support, listening to me talk about my project all the time, making helpful suggestions on all of the writing I've sent him, and always having something nice to say.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

ABSTRACT

Vellacott-Ford, Karen A. PhD, Purdue University, May 2020. Sequencing and Characterization of a Maternal-Effect Sex Determining Autosomal Inversion in the Hessian Fly. Major Professor: Jeffrey J. Stuart.

The unusual sex-determination system of the Hessian fly provides an excellent opportunity to investigate early sex chromosome evolution, sex determination, and chromosome behavior. The female Hessian fly has two copies of each X chromosomeone copy from each parentwhereas the male has only the copies contributed by his mother. However, the Hessian fly has no heterogametic sex; both the mother and father contribute a copy of each somatic chromosome (two X chromosomes and two autosomes) to each of their gametes. The sex-determining karyotype in males is established through elimination of paternal X chromosomes during early embryogenesis. Whether an embryo discards its paternal X chromosomes, resulting in male development, or retains these chromosomes, resulting in female development, depends on the genotype of the mother. An inversion on the long arm of Autosome 1 (A1) has suppressed recombination around a sex-determining master switch, causing it to take on the role of a maternal-effect neo-W chromosome. In a ZW sex-determination system, sex is determined by the female gamete; the female is the heterogametic sex (ZW) whereas the male is homogametic (ZZ). We refer to A1 with the sex-determining inversion as W′ (prime representing the maternal effect) and A1 lacking the inversion as Z. Female-producing females (ZW′) contribute W′ to half of their offspring, which become female-producing females, and Z to the other half, which become male-producing females (ZZ). The presence of W′ in the mother prevents the elimination of paternal X chromosomes in her offspring, resulting in the female karyotype. The offspring of mothers lacking W are typically all male; however, there is a third form of A1 that results in the production of both male and female offspring. The sequence of W′, its evolutionary history, and the mechanism by which it prevents paternal

X chromosome elimination are unknown. As a first step in addressing these unknowns, W′ and Z from both a New World and an Old World Hessian fly population were sequenced and characterized. The W′ and Z sequences reveal that the inversion occurred prior to the divergence of the New World and Old World populations and that relatively few changes have since accumulated within the inversion sequence of either population. Additionally, a region of A1 outside of the inversion on Scaffold A1.36 has been identified in which recombination between Z and W′ has been suppressed; this region may also have a role in sex determination. Genes were annotated for both the inversion scaffolds and Scaffold A1.36. Candidate genes for the sex-determination master switch have been selected from these regions based on their predicted functions and differences between the Z and W′ sequences of the genes.

1. INTRODUCTION

The Hessian fly, Mayetiola destructor, is an economically significant worldwide pest of wheat belonging to the family Cecidomyiidae, the gall midges; it was transported to the United States during the Revolutionary War. The cecidomyiids have traditionally been placed within the suborder Nematocera, a paraphyletic group of flies having filamentous, multi-segmented antennae. Sequencing of the Hessian fly genome and comparison of its gene sequences with those of other insects, however, has revealed it to be within a sister group to the drosophilids [1]. The Hessian fly has a gene-for-gene interaction with wheat, its main food source. The larva sends a signal to its host plant, in the form of salivary proteins, which shuts down the host's immune response and redirects its metabolism from plant growth to production of nutrients for the insect; in this way, the entire plant is transformed into an unusual gall. The Hessian fly life cycle lasts approximately one month. The female fly lays her eggs between the veins of wheat seedling leaves. Upon hatching, the first instar larva crawls to the base of the leaf where it establishes a feeding site by attaching its mouthparts to the cell wall. The second instar is stationary at its feeding site and continues to feed until it transitions to the third instar, at which point its skin turns brown, loosens, and becomes the pupal case. After a few days of pupation, the adult emerges. In addition to its agricultural pest status and gene-for-gene interaction with wheat, the Hessian fly has an unusual chromosome cycle and sex determination system that provide interesting research opportunities.

1.0.1 Sex-determining maternal-effect autosomal inversion

The Hessian fly has an established XO sex chromosome system in which individuals with two X chromosomes develop as females while those with only one develop as males. Additionally, one of its autosomes has taken on the role of a maternal-effect neo-W chromosome.

In a ZW sex-determination system, the female is the heterogametic sex (ZW) whereas the male is homogametic (ZZ) and sex is determined by the female gamete. An inversion on autosome A1 has given it properties of a maternal-effect neo-W chromosome; the inversion, which suppresses recombination around a sex determining region, is inherited heterozygously in the female and is not carried in the male. A second inversion, present only in female producers, has been identified in many but not all Hessian fly populations and is thought to be involved in extending the recombination suppression of the first [2]; both inversions are depicted in 1.2. We refer to A1 with the sex-determining inversion as W-prime, or W′,(prime representing the maternal effect) and A1 lacking the inversion as Z. When W′ is present in the mother, her offspring will attain the female-determining karyotype which includes copies of the two X chromosomes from both parents. If the mother lacks W′, her offspring will typically undergo a paternal X chromosome elimination event during embryogenesis, resulting in the male-determining karyotype (shown in Figure 1.1). A third form of A1 exists which allows the mother to produce a mixture of male and female offspring. This form, which lacks the inversions and was sequenced as part of the reference genome, is dominant to Z and recessive to W′.

1.1 Chromosome cycle and sex determination of the Hessian fly

The gall midges, including the Hessian fly [3], have germline-specific (E) chromosomes which are eliminated from future somatic cells during early embryogenesis and from male germline cells during spermatogenesis. These E chromosomes are derived from somatic chromosomes [1] and are thought to be involved in oogenesis, though their exact role is unknown. The Hessian fly's somatic chromosomes include two autosomes, A1 and A2, and two sex chromosomes, X1 and X2. Sex determination in the Hessian fly depends on an additional chromosome elimination event in early embryogenesis: the loss of paternal X chromosomes from future males, which is controlled by the maternal genotype [3].

1.1.1 Chromosome elimination: early embryogenesis

Germline-limited (E) chromosomes

During or soon after the segregation of future germ cells in early embryogenesis, E chromosomes are eliminated from future somatic nuclei. Chromosome elimination is prevented in future germ cell nuclei by the polar granules, which slow their rate of division. The exact time of chromosome elimination and the behaviors of chromosomes to be eliminated vary among the gall midges and the precise details of the elimination process are unclear. Depending on the species, elimination may occur one or more divisions after germ cell segregation and may be either spread out over a range of divisions or restricted to a specific division. In the Hessian fly and all other investigated members of its subfamily, Cecidomyiinae, elimination has been observed only during the fifth mitotic division. Mitotic divisions prior to the elimination appear normal and chromosomes to be eliminated appear similar in morphology to those that will be retained until anaphase of the elimination division. In this division, all chromosomes appear to enter anaphase and begin traveling toward the poles; however, those that will be eliminated return to the metaphase plate rather than enter daughter nuclei [4].

paternal X chromosomes and sex determination

Sex determination mechanisms vary among the gall midges; several species in addition to the Hessian fly exhibit monogeny [5], [6], which is linked to elimination of paternal chromosomes in the embryo. In all observed cases, the males and females are the same in germline chromosome number but differ somatic chromosome number, which determines the sex [6], [3] . The paternal X chromosomes may be eliminated in the same time and manner as the E chromosomes, though they could not be visually distinguished from the E chromosomes during this observation. While Hessian fly E chromosomes are always eliminated at the fifth division, the division at which paternal X chromosomes are eliminated in future males is unknown. No chromosome elimination has been observed at any other time of embryogenesis in the Hessian fly. In some other gall midges, the paternal X chromosomes are eliminated at a

separate time from the E chromosomes. The behavior of the E and paternal X chromosomes during elimination may also differ [6].

The Sciarids, or fungus gnats, are a sister group to the Cecidomyiids. Both belong to the superfamily Sciaroidea and share many features of their sex determination mechanisms. In both, the genotype of the mother determines the sex of the offspring through control of paternal sex chromosome elimination during early embryogenesis. In an abnormal mitotic anaphase, sister chromatids fail to separate and become included in the daughter nuclei. In both sciarids and cecidomyiids, the precise mechanism of paternal chromosome elimination versus retention is unknown. In the sciarids, however, one paternal X chromosome is eliminated in future females while two are eliminated in future males.

1.2 Sex determination pathway: comparison to other flies

The chromosome behavior of cecidomyiids and sciarids is not the only unusual feature of their sex determination. The genetic basis of sex determination in insects is best characterized in Drosophila. The number of sex chromosomes provides the signal which initates a splicing-regulation cascade ending in sex-specific somatic and germline development and dosage compensation [7]. Genes at the bottom of the cascade are more highly conserved than those at the top and evolution of the pathway is thought to have begun with the most downstream genes in the pathway with later incorporation of upstream regulators [8].

The master switch at the top of the Drosophila sex determination cascade, sex lethal (sxl), is spliced into multiple forms which regulate both its own splicing and the splicing of other genes involved in sex determination. Default splicing of sxl produces the male-specific form, which encodes a non-functional protein. In the embryo, female-specific splicing of sxl is initially signaled by the karyotype, resulting in a functional protein which proceeds to regulate continued female-specific splicing of sxl throughout development [9]. The role of sxl in sex determination appears to be restricted to Drosophila; in other insects, including the Hessian fly and sciarids, sxl does not undergo sex-specific splicing [10], [11].

The next step in the sex-determination cascade is the sex-specific splicing of transformer (tra), which is controlled by sxl in Drosophila. Only the female-specific form of tra encodes a functional protein, which in turn participates in the female-specific splicing of doublesex [12]. In several insects including the medfly, housefly, wasp, and honeybee, tra fulfills a role similar to that of Drosophila sxl by responding to the primary sex determination signal and regulating its own splicing in a positive feedback loop [7]. While tra varies among insects in its sequence and interactions with upstream regulators, its role in the splicing of doublesex is more highly conserved [12]. Tra has not be identified in the Hessian fly [10] or in Sciara, though attempts have been made to do so using highly conserved parts of the sequence.

Transformer-2 (tra-2) cooperates with tra in the female-specific splicing of doublesex (dsx) [13]. Tra-2 does not generally undergo sex-specific splicing; this is true in both the Hessian fly and Sciara, as well. Tra-2 has an RNA binding domain and RS domain required for its doublesex-splicing function. The RNA-binding domain is conserved in Hessian fly and sciarid tra-2 while the RS domains vary more from those of other insects [10]. Sciara tra-2 is able to participate in the splicing of Drosophila doublesex, though it is less effective than Drosophila tra-2 [14]. Functional analysis of this gene has not been done in the Hessian fly.

Doublesex (dsx) is a transcription factor that in most insects controls sex-specific differentiation at the end of the sex determination cascade [8]. The male form of the dsx protein, DSXM, activates expression of genes required for male differentiation while the female form, DSXF, does the opposite [15]. In the Hessian fly, dsx is spliced into sex-specific forms similar to those typically found in other insects [10]. Sciaria doublesex expression is atypical; all splice variants are present in both sexes, though the quantities are sex-specific. Only the female-specific protein DSXF is translated and is present in both sexes. For these reasons, dsx is unikely to play a discriminatory role in sciarid sex determination as it does in most insects [16].

The Hessian fly is an economically significant worldwide pest of wheat belonging to the family Cecidomyiidae, the gall midges. It is thought to have evolved in the fertile crescent and was transported to the United States during the Revolutionary War. The Hessian fly has a gene for gene interaction with wheat plants. Its life cycle lasts approximately 28 days.

The female Hessian fly lays her eggs between the veins of wheat seadling leaves. The first instars hatch and crawl to the base of the leaf where they establish a feeding site. The larva sends a signal to the host plant in the form of salivary proteins which redirects the plant's metabolism from plant growth to production of nutrients on which the larva feeds. The second instar continues to feed until it transitions to the third instar, at which point its skin turns brown, loosens, and becomes the pupal case. After about X days of pupation, the adult emerges. [17].

1.3 Chromosome cycle of the Hessian fly

The gall midges, including the Hessian fly [3], have germline-specific (E) chromosomes which are eliminated from future somatic cells during early embryogenesis. These E chromosomes are thought to be involved in oogenesis and are derived from somatic chromosomes. The Hessian fly's somatic chromosomes include two autosomes, A1 and A2, and two sex chromosomes, X1 and X2. Sex determination in the Hessian fy, discussed in more detail in the next section, depends on an additional chromosome elimination event in early embryogenesis: the loss of paternal X chromosomes from future males.

1.3.1 Gametogenesis

During male gametogenesis, both the paternal copies of each somatic chromosome and the germline-limited chromosomes are eliminated; for this reason, male flies pass on only their maternal copy of each S chromosome to their offspring. Recombination does not occur during male meiosis. Imprinting is used to differentiate among maternal and paternal chromosomes; only the maternally derived chromosomes are modified during spermatogenesis. Female gametes receive both the maternally and paternally derived copies of each chromosome in addition to E chromosomes.

1.3.2 Chromosome elimination: early embryogenesis

During or soon after the segregation of future germ cells in early embryogenesis, E chromosomes are eliminated from future somatic nuclei. Chromosome elimination is prevented in future germ cell nuclei by the polar granules, which slow the rate of division. The exact time of chromosome elimination and the behaviors of chromosomes to be eliminated vary among the gall midges and the precise details of the elimination process are unclear. Depending on the species, elimination may occur one or more divisions after germ cell segregation and may be spread out over a range of divisions or be restricted to a specific one. In the Hessian fly and all other investigated members of its subfamily, Cecidomyiinae, elimination has been observed only during only the fifth mitotic division. Mitotic divisions prior to the elimination appear normal and chromosomes to be eliminated appear similar in morphology to those that will be retained until anaphase of the elimination division. The behavior of chromosomes during the elimination anaphase varies among the gall midges; sister chromatids may be separated only at the kinetochores or along the entire length of the chromatids, may be joined at the ends, and may stay near the equator or move toward the poles. For the Hessian fly, all chromosomes appear to enter anaphase and begin travelling toward the poles, but those to be eliminated return to the metaphase plate rather than entering daughter nuclei [4]. The E chromosomes then form a mass which becomes a part of the yolk and disappears over the course of embryogenesis.

1.3.3 Paternal sex chromosome elimination and sex determination

An individual Hessian fly having two copies of each X chromosome will develop as a female whereas one having only one copy of each will develop as a male. This sex-determining karyotype is established by the genotype of the mother, which typically produces unisexual broods. Less commonly, a female may produces offspring of both sexes. Exceptional males may occur due to non-disjunction, resulting in some eggs with fewer X chromosomes.

Sex determination mechanisms vary among the gall midges; several species in addition to the Hessian fly exhibit monogeny [5], [6], which appears to be linked to elimination of

paternal chromosomes in the embryo. The reason for evolution of monogeny in Cecidomyiids is unknown. Within the subfamily Cecidomyiinae, that of the Hessian fly, in which several species are capable of monogeny, the males and females differ in somatic chromosome numbers but are the same in germline chromosome numbers in all observed cases. The resulting number of X chromosomes determines sex. [6], [3] In many other species, establishment of the sex determining karyotype does not involve sex chromosome elimination. For example, members of Lestremiinae, which is considered to be the least derived and most general subfamily of Cecidomyiidae, typically reproduce via paedogenesis, in which female larvae parthogenetically produce either male or female offspring.

While Hessian fly E chromosomes are always eliminated at the fifth division, the division at which paternal X chromosomes are eliminated in future males is unknown. No chromosome elimination has been observed at any other time of embryogenesis in the Hessian fly. In some other gall midges, the paternal X chromosomes are eliminated at a separate time from the E chromosomes. The behavior of the E and paternal X chromosomes during elimination may also differ. For example, elimination of X1 and X2 in the 7th mitotic division has been observed in *Wachtliella persicariae* whereas its E chromosomes are eliminated in the fourth. The elimination of these S chromosomes differs from the E chromosome elimination in that the ends of the S sister chromatids never separate and they do not move from the equator during anaphase [6] .

1.4 Sex-determining maternal-effect autosomal inversion

In a ZW sex-determination system, the female is the heterogametic sex (ZW) whereas the male is homogametic (ZZ) and sex is determined by the female gamete. An inversion on autosome A1 has given it properties of a maternal-effect neo-W chromosome; the inversion, which suppresses recombination around a sex determining region, is inherited heterozygously in the female and is not carried in the male. A second inversion, present only in female producers, has been identified in many but not all Hessian fly populations and is thought to be involved in extending the recombination suppression of the first [2]. We refer to

A1 with the sex-determining inversion as W-prime (prime representing the maternal effect) and A1 lacking the inversion as Z. When W′ is present in the mother, her offspring will attain the female-determining karyotype which includes copies of the two X chromosomes from both parents. If the mother lacks W′, her offspring will typically undergo a paternal X chromosome elimination event during embryogenesis, resulting in the male-determining karyotype. A third form of A1 exists which allows the mother to produce a mixture of male and female offspring. This form, which lacks the inversions and was sequenced as part of the reference genome, is dominant to Z and recessive to W′.

1.5 Similarity to Sciarids

The Sciarids, or fungus gnats, are a sister group to the Cecidomyiids, both belonging to the superfamily Sciaroidea, which has evolved a sex determination system with many similarites to that of the Hessian fly. The genotype of the mother determines the sex of the offspring through control of paternal sex chromosome elimination during early embryogenesis. In an abnormal mitosis, sister chromatids fail to separate and these eliminated chromosomes have different histone phosphorylation patterns from those which are retained. In both sciarids and cecidomyiids, the precise mechanism of paternal chromosome elimination versus retention is unknown. In both, it is clear that the eliminated chromosomes are modified differently from those to be retained, but the identity and role of the maternal effect responsible for the retention are unknown.

1.6 Sex determination pathway: comparison to other flies

The chromosome behavior of cecidomyiids and sciarids is not the only unusual feature of their sex determination. The genetic basis of sex determination in insects is best characterized in Drosophila. The number of sex chromosomes provides the signal which initates a splicing-regulation cascade ending in sex-specific somatic and germline development and dosage compensation [7]. Genes at the bottom of the cascade are more highly conserved

than those at the top and evolution of the pathway is thought to have begun with the most downstream genes in the pathway with later incorporation of upstream regulators [8].

The master switch at the top of the Drosophila sex determination cascade, sex lethal (sxl), is spliced into multiple forms which regulate both its own splicing and the splicing of other genes involved in sex determination. Default splicing of sxl produces the male-specific form, which encodes a non-functional protein. In the embryo, female-specific splicing of sxl is initially signaled by the karyotype, resulting in a functional protein which proceeds to regulate continued female-specific splicing of sxl throughout development [9]. The role of sxl in sex determination appears to be restricted to Drosophila; in other insects, including the Hessian fly and sciarids, sxl does not undergo sex-specific splicing [10], [11].

The next step in the sex-determination cascade is the sex-specific splicing of transformer (tra), which is controlled by sxl in Drosophila. Only the female-specific form of tra encodes a functional protein, which in turn participates in the female-specific splicing of doublesex [12]. In several insects including the medfly, housefly, wasp, and honeybee, tra fulfills a role similar to that of Drosophila sxl by responding to the primary sex determination signal and regulating its own splicing in a positive feedback loop [7]. While tra varies among insects in its sequence and interactions with upstream regulators, its role in the splicing of doublesex is more highly conserved [12]. Tra has not be identified in the Hessian fly [10] or in Sciara, though attempts have been made to do so using highly conserved parts of the sequence.

Transformer-2 (tra-2) cooperates with tra in the female-specific splicing of doublesex (dsx) [13]. Tra-2 does not generally undergo sex-specific splicing; this is true in both the Hessian fly and Sciara, as well. Tra-2 has an RNA binding domain and RS domain required for its doublesex-splicing function. The RNA-binding domain is conserved in Hessian fly and sciarid tra-2 while the RS domains vary more from those of other insects [10]. Sciara tra-2 is able to participate in the splicing of Drosophila doublesex, though it is less effective than Drosophila tra-2 [14]. Functional analysis of this gene has not been done in the Hessian fly.

Doublesex (dsx) is a transcription factor that in most insects controls sex-specific differentiation at the end of the sex determination cascade [8]. The male form of the dsx protein, DSXM, activates expression of genes required for male differentiation while the female form,

DSXF, does the opposite [15]. In the Hessian fly, dsx is spliced into sex-specific forms similar to those typically found in other insects [10]. Sciaria doublesex expression is atypical; all splice variants are present in both sexes, though the quantities are sex-specific. Only the female-specific protein DSXF is translated and is present in both sexes. For these reasons, dsx is unikely to play a discriminatory role in sciarid sex determination as it does in most insects [16].

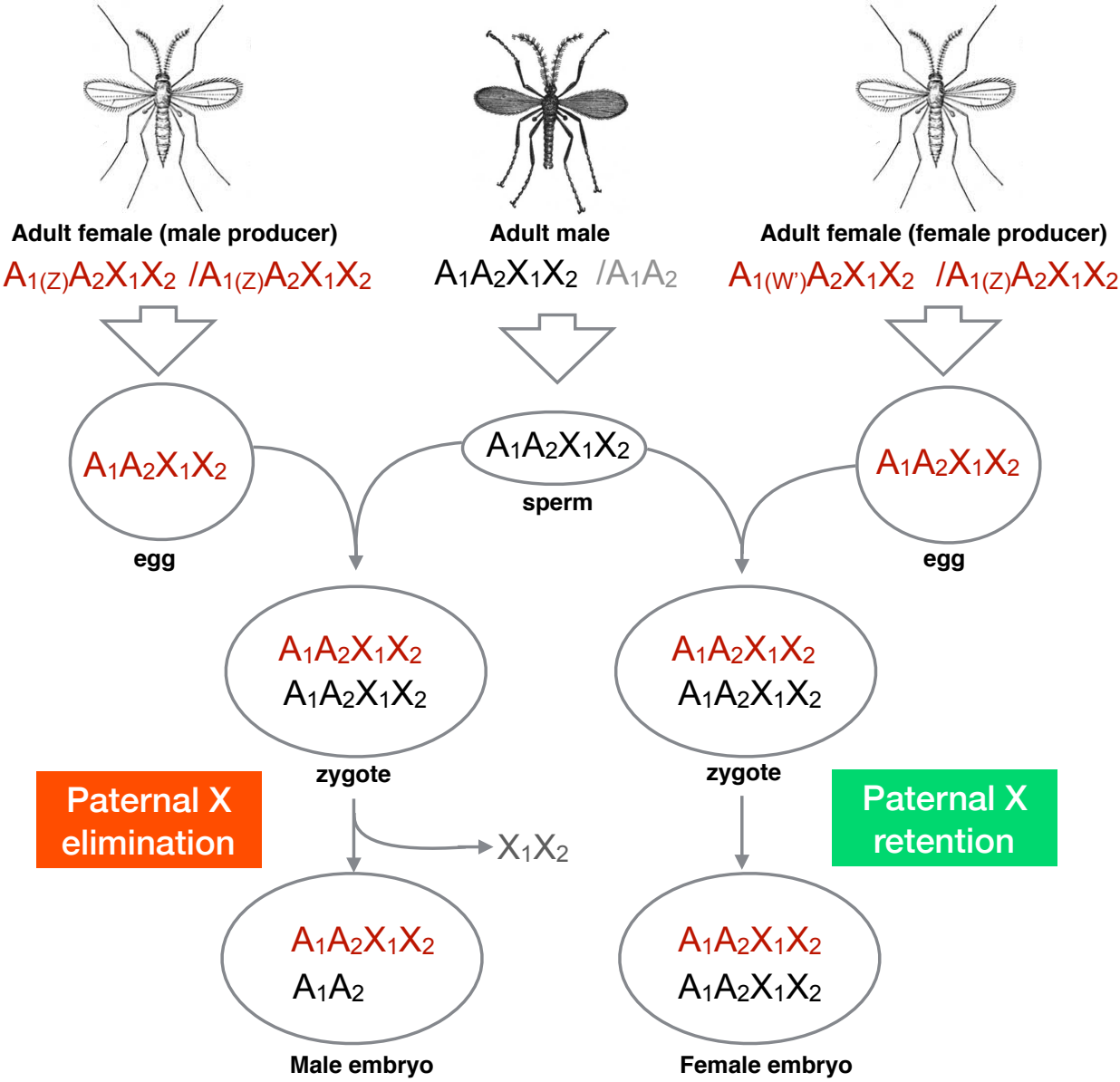

Fig. 1.1. Sex determination in the Hessian fly. A male fly is shown with male-producing and female-producing female flies along with their karyotypes. Both females have two copies of each X chromosome while the male only has his maternal copy of each. The male can only pass on his maternally-derived chromosomes (shown in black); paternally-derived copies are shown in grey. A female-producing female has the W′ form of A1 while a male-producing female has the Z form of A1. Each parent contributes a copy of each somatic chromosome to the zygote. During embryogenesis in the offspring of a male-producing female, the paternal copies of the X chromosomes are eliminated, resulting in the male genotype. In the offpsring of a female-producing female, the paternal X chromosomes are retained, resulting in the female karyotype.

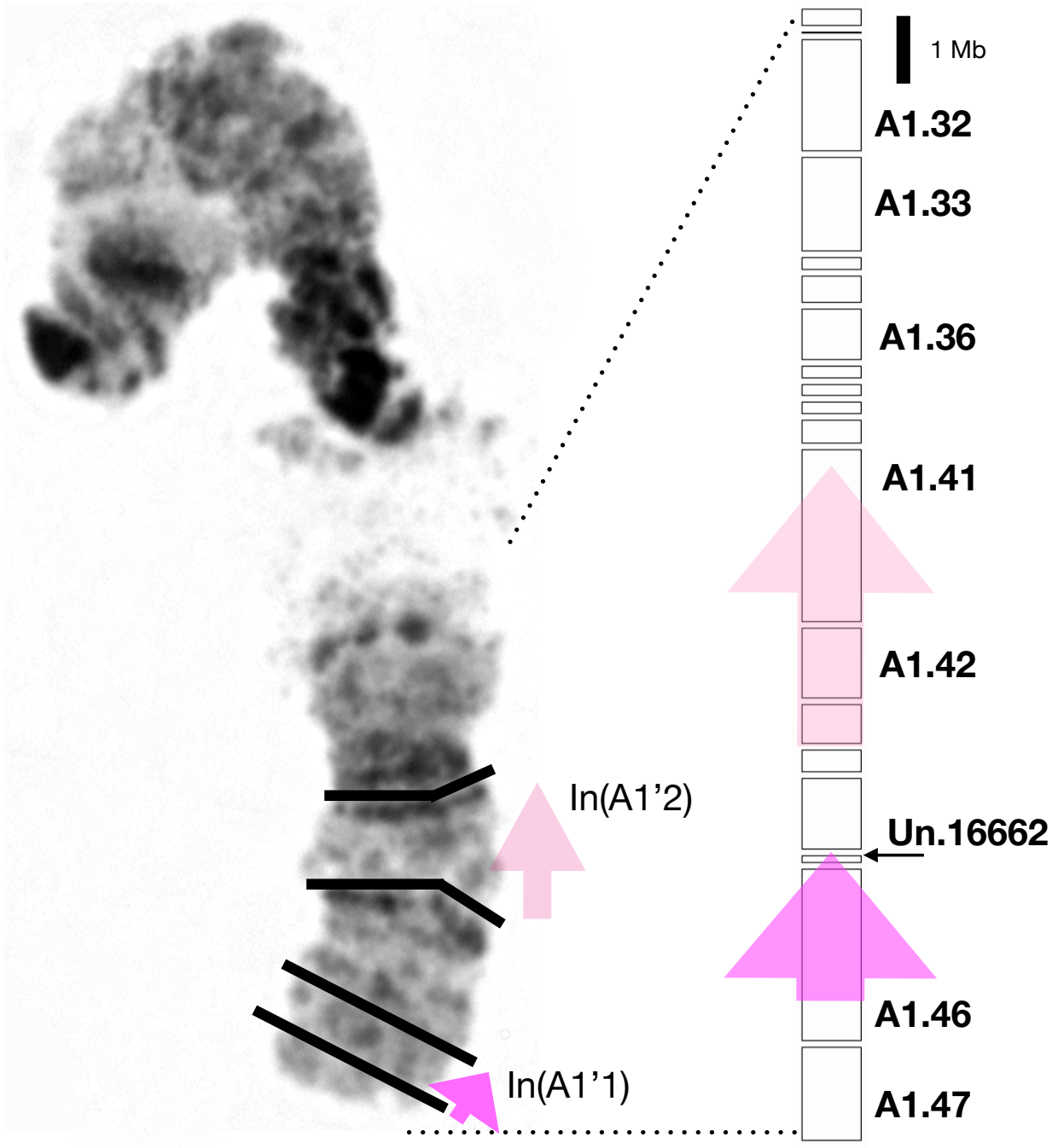

Fig. 1.2. Autosome 1 and its inversions associated with female producers, represented by arrows. A1 scaffolds are represented by rectangles. The sex-determining inversion, In(A1q1), is located near the distal end of the long arm of A1 and includes the majority of Scaffold A1.46 in addition to the small scaffold, Un.16662. The inversion In(A1q2) is located proximally to In(A1q1) and includes scaffolds A1.42, A1.43, and part of A1.41. Its exact breakpoints are unknown.

2. SEQUENCING AND CHARACTERIZATION OF A MATERNAL-EFFECT SEX DETERMINING AUTOSOMAL INVERSION

2.1 Introduction

Evolution of heteromorphic sex chromosomes has occurred independently many times in eukaryotes, including in vertebrates, plants, and insects. While established sex chromosomes have been extensively researched in model organisms including Drosophila and several mammals [18], little is known about the forces that drive the evolution of new heteromorphic sex chromosomes from autosomes. The unususal sex determination system of the Hessian fly provides an opportunity to investigate early sex chromosome evolution, sex determination, and chromosome behavior. The Hessian flys somatic chromosomes include two autosomes, A1 and A2, and two sex chromosomes, X1 and X2 [4]. An individual Hessian fly having two copies of each X chromosome will develop as a female whereas one having only a single copy of each will develop as a male. This sex-determining karyotype is established by the genotype of the mother [3], which typically produces unisexual broods. In a ZW sex-determination system, sex is determined by the female gamete and the female is the heterogametic sex (ZW) whereas the male is homogametic (ZZ). An inversion on autosome A1 in the Hessian fly, In(A1q1), has given it properties of a maternal-effect neo-W chromosome; the inversion, which suppresses recombination around a sex determining region, is inherited heterozygously in the female and is not carried in the male [2]. A second inversion present only in female producers, In(A1q2), has been identified in many but not all Hessian fly populations and is thought to have occured after In(A1q1), extending its suppression of recombination [2]. Sex chromosomes commonly evolve from autosomes through a series of inversion events which suppress recombination in a sex-determining region [19]. We refer to A1 with the sex-determining inversion as W-prime (prime representing the maternal effect)

and A1 lacking the inversion as Z. When W′ is present in the mother, her offspring will attain the female-determining karyotype which includes copies of the two X chromosomes from both parents. If the mother lacks W′, her offspring will typically undergo a paternal X chromosome elimination event during embryogenesis, resulting in the male-determining karyotype [2]. A third form of A1 exists which allows the mother to produce a mixture of male and female offspring. This form, which lacks the inversions and was sequenced as part of the reference genome [1], is dominant to Z and recessive to W′. The the sequence of W′, its evolutionary history, and the mechanism by which it prevents paternal X chromosome elimination are unknown. As a first step in addressing these unknowns, the W′ and Z forms of A1 from both a New World and an Old World Hessian fly population have been sequenced and characterized. The W′ and Z sequences reveal that the inversion occurred prior to the divergence of the New World and Old World Hessian fly populations and that relatively few changes have accumulated within the inversion sequence of either population since then. An additional region of A1 in which the W′ sequences are unusually similar to each other and different from the Z sequences has been identified proximal to both In(A1q1) and In(A1q2). This region may also have a role in sex determination and include a third inversion which has not yet been identified.

2.2 Methods

2.2.1 Material and crosses

Crosses were designed to produce four half sisters: one female-producing and one male-producing daughter from mothers of both Old World (Magen, Israel) and New World (Indiana) populations. The father was taken from a third population, GP (Kansas), to ensure that each half sister would share a copy of A1 that would differ in sequence from the maternal copy. These sequence differences allow identification of each Z and W′ form of A1 from both Israel and White-eye.

Crosses were set up as follows: A GP male was placed in a covered pot of wheat seedlings with a White-eye female. After being allowed to mate, the male was transferred to a separate

covered pot containing an Israel female. Pots were numbered according to the male and population of the mother. Females were then allowed to lay eggs on the wheat seedlings in the same pots. For the crosses in which both the Israel and White-eye mothers produced female offspring, these female offspring were allowed to mate and lay eggs, each in a separate covered pot.

2.2.2 DNA extraction and sequencing

Each female was collected after being allowed to lay eggs and was identified as either a female producer or male producer by the sex of her offspring. Four of these females sharing the same father were selected for sequencing (as shown in Figure 2.1). A syringe needle was used to slice off the abdomen of each female to prevent any contaminating sperm DNA from being sequenced along with the DNA of the female. The remaining body was then placed in a 1.5 mL tube, frozen in liquid nitrogen, and stored at -80 degrees C. Genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue kit and sequenced by the Purdue Genomics Core facility.

2.2.3 Read filtering and alignment

Genome sequencing reads were trimmed and filtered using Trimmomatic: reads fewer than 50 bases in length were discarded after bases with a quality score of less than 30 were trimmed from the $3'$ end of each read. Using Bowtie2, reads were then aligned to the reference genome.

The resulting BAM files were then filtered using Samtools to remove reads of low-quality mapping (score of less than 20), unmapped and unpaired reads, and duplicates. The GATK IndelRealigner tool was used to reduce errors resulting from incorrect alignment of reads representing false indels. The alignments were viewed in Integrated Genome Viewer (IGV). Mapping statistics were calculated using Samtools stats.

2.2.4 Variant calling and assignment of variants to chromosome of origin

A combined variant call format (VCF) file for all four genotypes was made using Samtools mpileup and bcftools. Positions at which read coverage was less than three for any of the four genotypes were excluded. The VCF file was used to assign SNPs and indels to their chromosome of origin (example of this process shown in Figure 2.2). If an allele was present for at least one genotype but was not in all four, that allele was assigned to the maternal chromosome(2) (either Israel or White-eye Z or W′). If the alternate allele at that position was present in all four genotypes, it was assigned to the paternal (GP Z) chromosome. If the same two variants were present in all four samples, they were sorted using a custom Python script (using Pysam 0.11.1) and the alignment files for each sample. Reads already classified as maternal or paternal variants were used to classify the unknown variants. For each of the four genotypes, all reads that aligned to a position with an assigned allele were added to one of two lists: maternal reads or paternal (GPZ) reads, depending on the sequence of the read at that position. At each position with unclassified variants, reads intersecting the position which were in the maternal/paternal list were used to identify these variants. Once these unknown variants were classified, the reads containing these variants were added to the maternal/paternal read lists and used to classify more unknown variants. If at any position there were reads from the maternal copy of A1 with more than one variant (only one is possible), the most frequent variant was considered to be the real one while the others were ignored. At least two reads were required to represent the variant for it to be counted.

2.2.5 Comparison of frequencies of variants between chromosome A1 scaffolds near the inversion and the rest of A1

Comparisons of frequency of variants in individual chromosome A1 scaffolds versus the rest of chromosome A1 were made using the Wilcoxon rank sum (Mann-Whitney) test in R. In summary, the number of SNPs within each 10 kb region for the scaffold (sample 1) and the rest of chromosome A1 (sample 2) were ranked. These were pooled for the two samples and sorted in ascending order. The ranks (positions in sorted list) were then summed for each

sample. The W statistic for each sample was calculated as R - n(n + 1)/2, where R is the rank sum and n is the sample size. P values were calculated based on a normal distribution of W.

2.3 Results

2.3.1 Identifying and assembling inversion sequences

Differences between the maternal and paternal copies of chromosome A1 were used to identify both the White-eye and Israel inversion sequences. I identified inversion-specific reads as those that contained at least one variant that was present in the female producer and missing in at least one of the other three genotypes. I made the assumption that W′ and Z sequences are most likely identical in regions for which no W′ or Z-specific variants could be identified; however, this may not be the case for regions of low coverage that fail to truly represent both sequences.

Fortunately, maternal-specific variants were frequent enough to identify the majority of the inversion sequences for both populations. The reference sequence for the main inversion scaffold, A1.46, has several gaps–regions of unknown sequence, each less than 10 kb–which combined make up about 7.5% of the scaffold. Excluding these gap positions, the majority (82% for Israel and 87% for White-eye sequences) of Scaffold A1.46 positions are covered by at least three maternal reads. The vast majority of variants identified from both Israel and White-eye W′ sequences (97%) were supported by at least 3 reads. Although part of the inversion sequence remains unknown due to gaps and, to a lesser extent, low coverage, each of these regions are estimated to be only a few kb in length and are likely to include only a small number of the inversion genes.

2.3.2 Inversion In(A1q1) region compared with the rest of the genome

Although the Israel and US populations have been geographically separated for at least 200 years, the female producer sequences of the Israel and White-eye populations should have

some similarities to each other if the same sex-determining inversion is present in both. For the entire genome of both Israel and White-eye populations, the female producer sequence has a similar number of SNPs and indels compared to the male producer sequence from its own population (Table 2.2). Within the main In(A1q1) scaffold, A1.46, fewer SNPs and indels were found between the female producer sequences of each population than between the male producer sequences (Table 2.3).

In(A1q1), which is always present in female-producers, is located on near the end of the long arm of Autosome 1 (Figure 1.2). A second inversion which is associated with, but not required for, female-producers, In(A1q2), is found several Mb proximal to In(A1q1) when present. Only a few A1 scaffolds, all located on the long arm, have a higher frequency of variants shared between the two W′ sequences when compared to the rest of A1 (Table 2.4). These include A1.46 and its nearest large scaffolds, A1.45 and A1.47. Unexpectedly, a region several Mb proximal to In(A1q1), spanning scaffolds A1.33 to A1.36, also has an unusually high frequency of variants shared by the W′ sequences. When counting only the variants that are shared by the two W′ sequences and found in neither Z sequence, the similarity between A1.46 and A1.36 in particular is more obvious. Compared to the rest of the chromosome, scaffolds A1.33 (W = 299790, p ¡ 0.001) and A1.36 (W = 213490, p ¡ 0.001) have a high frequency of variants shared only by the two W′ sequences.

Within In(A1q1), the high frequency of variants shared between the W′ sequences of the Israel and White-eye populations is expected due to a lack of recombination in this region; the variants shared between the two W′ sequences cannot be lost over time through recombination. Consistent with this is the low frequency of variants shared by W′ and Z sequences of each population for these scaffolds (Tables 2.5, 2.6). As new mutations accumulate on Z, they cannot be introduced into W′ via recombination. As in the In(A1q1) scaffolds, A1.36 has an unusually low frequency of variants shared between Z and W′ for both the Israel (W = 47128, p ¡ 0.001) and White-eye (W = 46949, p ¡ 0.001) sequences. Two scaffolds in the In(A1q2) region also have a high frequency of these variants compared to the rest of the chromosome, though much lower than that of A1.46 or A1.33-A1.36. This would not be so unusual if In(A1q2), which appears to suppress recombination through part

of A1.36, were present in these populations; however, this inversion has not been identified in either the Israel population or in the Indiana population [2], from which the White-eye population was derived (S. Cambron, personal communication). The scaffolds overlapping with both In(A1q1) and In(A1q2), as well as the A1.33-A1.36 region, are described in more detail in the following sections.

2.3.3 Female producer versus male producer sequences throughout In(A1q1) scaffolds

Compared to most of the A1 scaffolds, both inversion scaffolds–A1.46 and Un.16662–have a high frequency of variants that are shared between the W′ sequences of each population (Figure 2.3). That the vast majority of total W′ variants are shared between the Israel and White-eye sequences throughout the entire region indicates that very few changes have occured within the inversion since these populations have been separated. In contrast, the two Z sequences share a much lower proportion of their variants with each other (Figure 2.5). When considering only those variants that are shared between the W′ sequences and absent in both Z sequences, the highest frequency is found between positions 1700000-1750000 of A1.46 (Figure 2.4). Within this region, these W′-specific variants also make up a higher proportion of the total variants than they do for most of A1.46. Although this information is insufficient to determine whether this region has a sex-determining function, it would make sense for fewer of the W′ variants to be shared with either Z sequence in the gene that determines whether males or females are produced.

2.3.4 InA1q2 not present in Israel or White-eye

The inversion In(A1q2), whose exact breakpoints are unknown, spans at least the distal end of A1.41 to the proximal end of A1.43 for the populations in which it has been identified. If this inversion is also present in both the Israel and White-eye populations, the Z and W′ sequences of each population should share relatively few variants while the two W′ sequences should have more variants in common than in a typical A1 scaffold. If In(A1q2) is present

only in the White-eye population, the White-eye female producer sequence in this region may not be similar to the Israel female producer sequence; however, due to supression of recombination there should be fewer variants shared between the White-eye male producer and female producer sequences in this region compared to the scaffolds outside of the inversions.

For both White-eye and Israel sequences, the majority of variants are shared between Z and W′ for regions of several hundred kb within scaffolds A1.41 to A1.43 (Figure 2.9). It appears that recombination has occurred between the Z and W′ chromosomes within both populations, though less than has occured on the short arm of A1. There are also smaller regions which have a higher than typical frequency of W′ variants shared between Israel and White-eye sequences (Figure 2.7). This may be a result of incomplete recombination suppression by In(A1q1).

2.3.5 Region identified outside of inversion with high similarity between female producers

The Israel and White-eye female producer sequences are unusually similar to each other within two large scaffolds outside of In(A1q1), A1.33 and A1.36 (Figure 2.10). However, in the more proximal scaffold (A1.33), the majority of White-eye variants are also shared between male producer and female producer sequences (Figure 2.12). For part of A1.33, the frequency of variants shared only between the two W′ sequences is higher than in most other non-inversion scaffolds but lower than that of the inversion; the two W′ sequences are similar to each other wheareas the Z sequences vary. Scaffolds on either side of A1.33 appear more similar to the short arm of A1, far from the inversion; for this reason it is unlikely that its proximity to the centromere or to either known inversion is responsible for the similarity between the two female-producer sequences. As in the inversion sequence, the vast majority of W′ variants are shared between Israel and White-eye sequences for A1.36. Although no inversion has been identified in this region of A1, the comparison of Z and W sequences within A1.36 is reminiscent of that within In(A1q1); few variants are shared between Z and

W while most are shared between Israel and White-eye W′ sequences. Recombination may be suppressed in this region due to a smaller undiscovered inversion.

2.4 Discussion

2.4.1 The sex-determining inversion sequence

Prior to recombination suppression around a sex-determining master switch, most of the sequence between the male-determining and female-determining chromosomes should be similar. Suppression of recombination allows mutations that are beneficial only to the heterogametic sex or detrimental to the homogametic sex to accumulate on the non-recombining chromosome as a group. Over time, the non-recombining sex chromosome is also expected to undergo asexual decay; genes unrelated to sex determination are lost, and a dosage compensation mechanism must evolve to balance gene expression between the sexes. Because the W′ sequence of the Hessian fly is carried only by female producers, this sequence is able to accumulate mutations that are detrimental to both males and male-production. The Hessian fly W′ has not yet undergone much asexual decay, as individuals homozygous for the inversion are able to survive [2]. Throughout the entire inversion sequence, though, the W′ sequence varies from the Z sequences. Much of this variation should be unrelated to sex determination; differences between the sequences that would later become Z and W′ likely existed prior to the inversion event and were captured and preserved by the inversion along with the sex-determining gene. In some regions within the inversion scaffolds, the number of differences between the Israel and White-eye Z sequences is comparable to that between the Z and W′ sequences. The similarity of the two W′ sequences indicates that the sex-determining inversion and most of its differences from the male-producer sequence occurred prior to the separation of the Old World and New World populations. Within the inversion scaffolds and Scaffold A1.36, there are few differences between the W′ sequences of the Israel and White-eye populations compared with the Z sequences. In heteromorphic sex chromosome systems, the effects of recombination suppression may result in lower variability for the non-recombining chromosome. This has been observed in the W chromosome

in birds [20] and the Y chromosome in humans [21]. Factors that may contribute to this reduced variability in the non-recombining chromosome are discussed below.

2.4.2 Low variability in W′ sequence

In Drosophila [22] and in humans [23], levels of nucleotide diversity increase with recombination rate. It has been proposed that in regions of low recombination, hitchhiking following selective fixation of advantageous mutations is partially responsible for low variability [24]. In regions of high recombination, sequence diversity may increase as a result of mutagenic double strand break repair [24]. Because the Hessian fly probably does not have a dosage compensation mechanism for W′ and Z, deleterious mutations in vital genes may result in the loss of the entire inversion sequence of that individual from the population, taking all unique mutations with it. The W′ chromosome has a lower effective population size than that of the Z chromosome; for each copy of W′ present in the population, there are five copies of Z. Lower effective population size should increase the effect of genetic drift on the W′ chromosome [14], which may have resulted in loss of variability between the W′ sequences prior to the geographic separation of the old and new world populations. If genetic drift has contributed to the loss of W′ variability prior to the separation of the Old World and New World populations, it may eventually result in greater differences between the populations as new mutations are fixed. For chromosomes that are carried more frequently in males, variability may be higher due to male mutation bias [20]. Gametogenesis in males involves many more replications than in females, giving male-determining chromosomes more opportunities to acquire replication-induced mutations. In a typical ZW system, the Z chromosome will spend two thirds of its time in males while the W chromosome spends all of its time in females. In the Hessian fly, the Z sequence spends about four sevenths of its time in males while the rest of Autosome 1 and Autosome 2 spend about half of their time in males. If male mutation bias has an effect on the Hessian fly, the Z sequence and remaining autosome sequences should have a similarly increased mutation rate when compared with the W′ chromosome. The extent of male mutation bias is affected by the difference in number of eggs

and sperm produced in addition to life history traits such as the age of the male at time of reproduction. For Drosophila species, the numbers of sperm and eggs produced are similar and no male bias has been observed, with the exception of Drosophila miranda [25].

2.4.3 Scaffold A1.36 may also have a female producer-specific inversion

The sequence within Scaffold A1.36 has many features of In(A1q1): a high frequency of differences between the male producer and female producer sequences, few differences between the two female producer sequences, and low recombination between the male producer and female producer sequences within each population. The scaffold is too far from In(A1q1) for its recombination to be suppressed by this inversion, as recombination has occurred within scaffolds closer to the inversion. Sequence closer to the centromere, where recombination may be less frequent, also has undergone more recombination between the Z and W′ sequences than in this scaffold. The recombination suppression may be a result of a third inversion that has not been discovered previously due to its small size. An alternative explanation is that the order of this scaffold has been misplaced relative to the others and its sequence is actually closer to the inversion. This seems unlikely, as the lack of recombination extends to the two smaller scaffolds placed on either side of Scaffold A1.36. The second inversion associated with female producers, In(A1q2), is not present in either the Israel or White-eye female producer sequences. Recombination between the male producer and female producer sequences has occurred in both populations within the scaffolds that are associated with In(A1q2). If Scaffold A1.36 has an inversion, it is likely present in both Old World and New World populations and is older than In(A1q2). This scaffold may also have genes with a role in female production or development.

2.4.4 Conclusions and Future work

The sex-determining inversion, In(A1q1), and most of its differences from the male-producer sequence occurred prior to the separation of the Old World and New world populations. An undiscovered inversion and genes involved in female production or female

development may also be present on Scaffold A1.36. The second inversion, In(A1q2), is not present in the Israel and Indiana populations and probably originated in a different US population. This inversion may extend the recombination suppression of both In(A1q1) and the potential A1.36 inversion. The next step in this project, identification of inversion genes and comparison of their male producer and female producer-specific sequences, is discussed in Chapter 3. Confirming the presence or absence of an inversion in Scaffold A1.36 and comparing its age to that of In(A1q1) may provide more insight into the evolution of the W′ chromosome. Additionally, if the presence of In(A1q1) can be confirmed in a more distant Old World population, its sequence may be useful in identifying regions within the inversion that first began to accumulate differences from the male producer sequence.

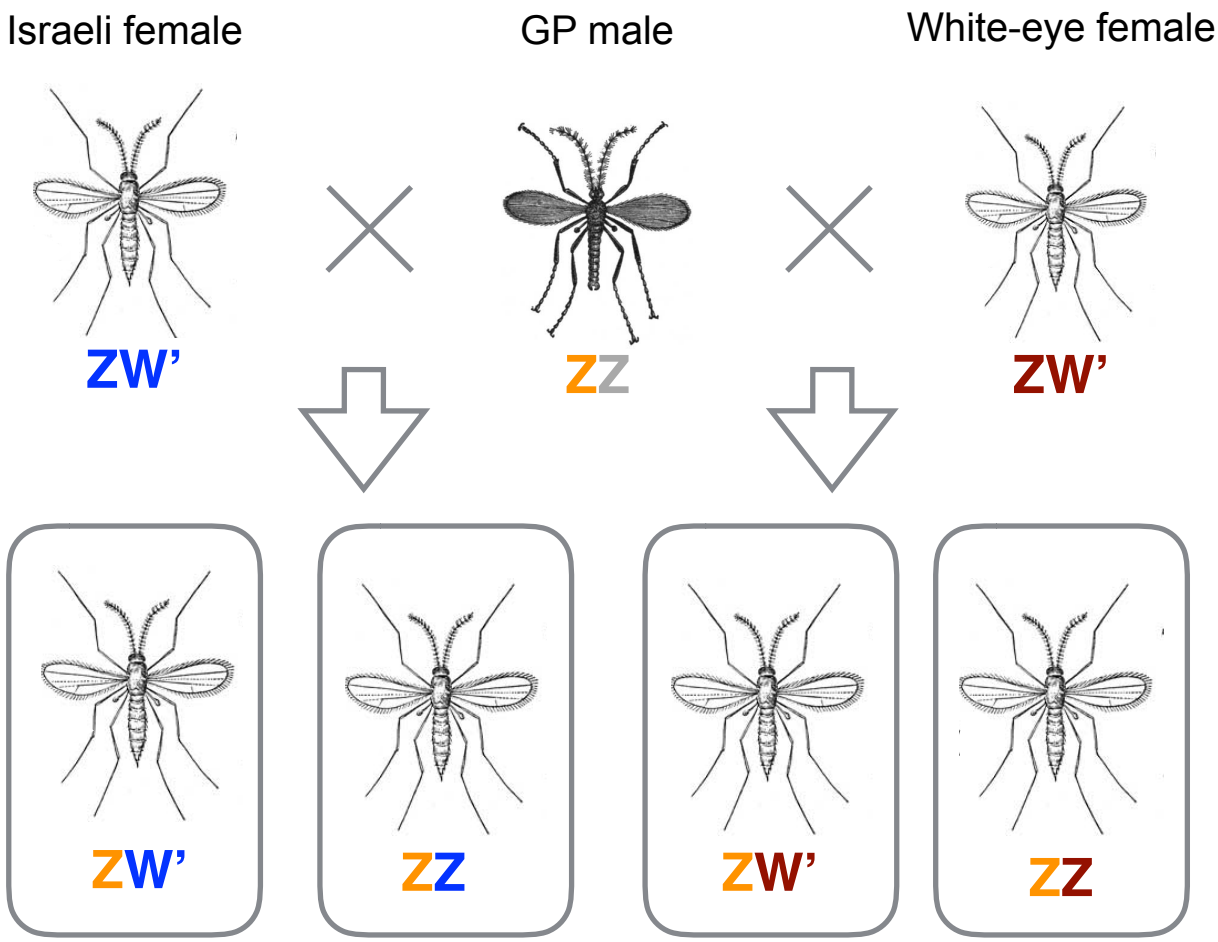

Fig. 2.1. Crosses for sequencing Z and W′ from Israel and White-eye. The crosses shown above were repeated with several GP males and pairs of females. The offspring chosen to be sequenced, one female producer and one male producer from each mother, are shown in squares with their genotypes. Chromosomes are color-coded for parent of origin: blue for Israel, red for White-eye, orange for the GP father's maternal copy of Z, and grey for the GP male's paternal copy of Z, which does not get passed on to the offspring.

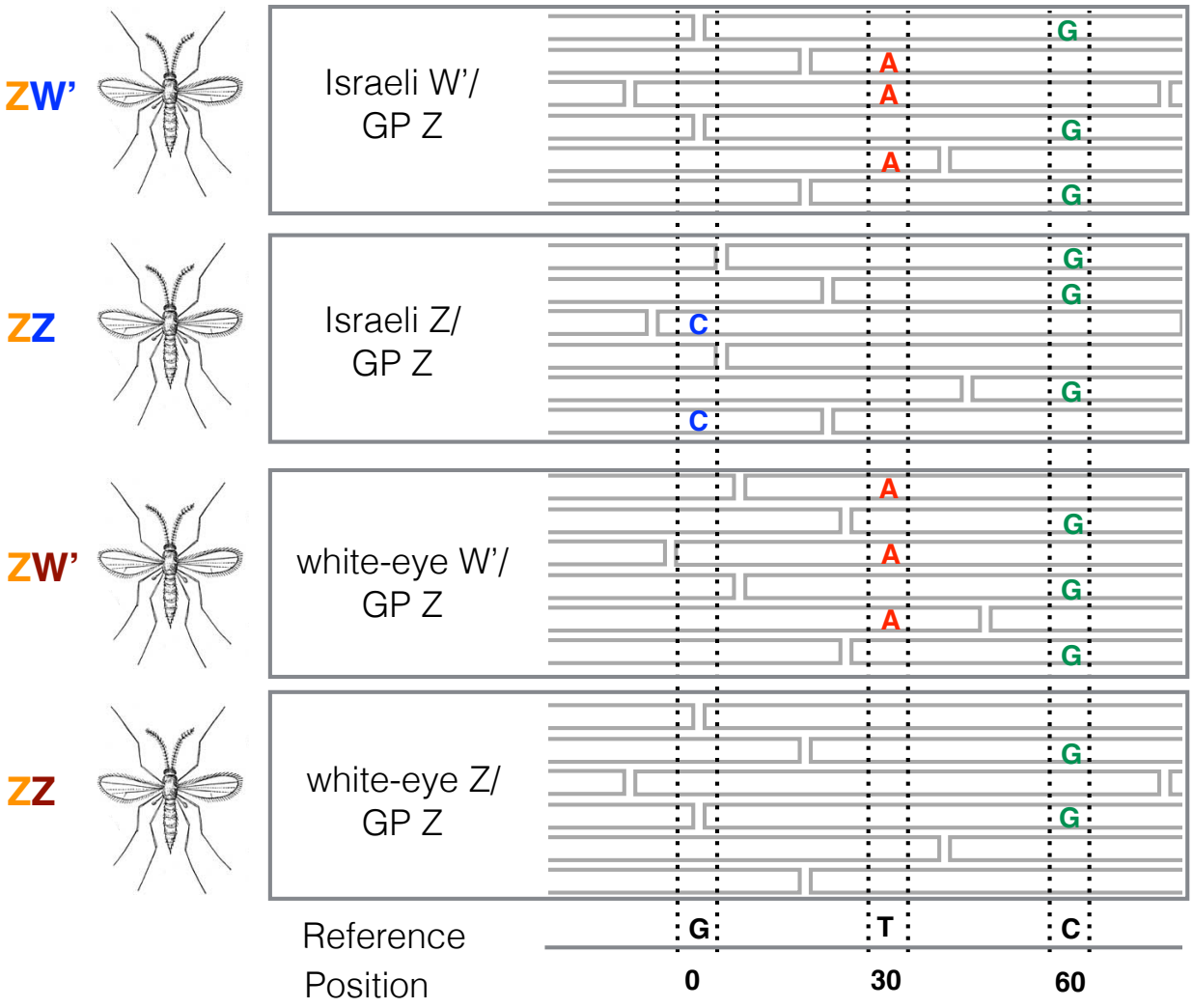

Fig. 2.2. Example sequence alignments for the Israel and White-eye male producers and female producers. Reads are represented by rectanges with variants from the reference shown above reference position and sequence. At position 0, C belongs to Israel Z whereas the other forms of A1 all match the reference sequence, G. At position 30, both W′ sequences are A while the Z sequences all match the reference sequence, T. At position 60, all genotypes have the same two alleles, making this position ambiguous. However, reads overlapping with position 30 can be used to identify the maternal (C) and paternal (G) alleles at this position.

Table 2.1.
Inversion sequence variants and length excluding gaps, low coverage areas, and positions at which reads could not be assigned to the inversion

sequence	length[1]	% of ref[2]	variants[3]	% of total variants[4]
Israel W′	1948146	82	43905	96.5
White-eye W′	2075845	87	43828	96.5

[1] Total number of positions within In(A1q1) scaffold A1.46 covered by at least three sequencing reads which could be assigned to the inversion

[2] Length of identified inversion sequence as a percentage of the reference sequence (excluding gaps in the reference)

[3] Total number of inversion-specific SNPs and indels supported by at least three inversion-specific reads

[4] Percentage of total inversion-specific SNPs and indels supported by at least three inversion-specific reads

Table 2.2.
Total number of whole genome SNPs and indels with respect to the reference sequence in Israel and White-eye female producer and male producer sequences

	total		exon only	
genotype	SNPs	indels	SNPs	indels
Israel W′/GPZ	1084511	300684	187292	23897
Israel Z/GPZ	1077583	298390	185128	23529
White-eye W′/GPZ	1004665	283797	172849	22567
White-eye Z/GPZ	1001597	283760	172795	22651

Table 2.3.
Total number of Scaffold A1.46 SNPs and indels with respect to the reference sequence in Israel and White-eye female producer and male producer sequences

	total		exon only	
genotype	SNPs	indels	SNPs	indels
Israel W′/GPZ	37022	8464	7687	718
Israel Z/GPZ	29132	6785	5865	541
White-eye W′/GPZ	36961	8447	7611	720
White-eye Z/GPZ	28253	6678	5668	551

Table 2.4.: Mean and Standard deviation of Israel and White-eye W′ SNPs and indels per 50kb of chromosome A1 scaffolds

	Israel W′[1]		White-eye W′[2]		Both W′[3]		Only W′[4]	
scaffold	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**
A1.1	174	65	160	79	28	14	0	1
A1.3	173	99	160	66	43	48	1	1
A1.4	268	145	132	106	29	28	1	1
A1.5	60	42	136	21	8	9	0	0
A1.6	204	135	71	17	12	10	0	0
A1.7	132	104	116	60	23	25	2	2
A1.8	223	72	38	44	6	8	0	0
A1.10	232	123	16	9	2	3	0	1
A1.11	199	54	144	81	23	15	3	7
A1.12	258	155	156	56	58	50	1	1
A1.13	171	131	144	128	33	29	0	0
A1.14	86	51	266	200	36	45	1	1
A1.15	177	101	193	122	79	46	2	3
A1.16	238	84	173	91	65	54	2	2
A1.17	212	81	199	75	56	25	0	1
A1.18	220	125	163	44	27	21	0	1
A1.19	207	31	192	57	28	12	0	0
A1.20	229	109	273	133	44	27	0	0
A1.21	152	54	171	31	30	10	0	0
A1.22	180	56	146	62	17	20	0	1
A1.23	263	90	272	82	76	49	1	1
A1.24	190	82	202	70	15	12	0	0

Continued on next page

Table 2.4 – *Continued from previous page*

	Israel W′[1]		White-eye W′[2]		Both W′[3]		Only W′[4]	
scaffold	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**
A1.25	193	52	221	67	5	3	0	0
A1.26	140	68	140	40	12	25	0	0
A1.27	64	76	60	71	24	25	1	1
A1.28	104	79	97	58	8	17	0	1
A1.29	136	94	174	102	15	16	1	1
A1.30	92	20	211	140	15	13	0	0
A1.32	133	147	224	145	20	25	1	1
A1.33	292	173	340	155	243	155	46	56
A1.34	223	80	200	81	104	52	34	46
A1.35	172	78	187	86	77	72	1	1
A1.36	336	112	337	113	318	107	178	79
A1.37	240	45	217	25	151	59	3	2
A1.38	57	17	200	83	17	11	3	4
A1.39	8	7	6	6	4	5	0	0
A1.40	20	14	23	19	17	12	0	1
A1.41	304	146	269	176	109	90	13	30
A1.42	227	115	226	104	44	35	1	1
A1.43	231	169	205	170	71	115	9	23
A1.44	254	165	348	158	84	60	7	10
A1.45	226	104	226	108	79	77	19	35
A1.46	523	112	522	110	494	108	287	75
A1.47	287	130	243	133	114	117	36	59
A1 [5]	223	100	218	104	96	97	33	53
A1 excl. A1.46 [6]	197	85	192	85	62	63	11	28

[1]Israel W′ sequence SNP and indel positions with respect to the reference sequence

[2]White-eye W′ sequence SNP and indel positions with respect to the reference sequence

[3]SNP and indel positions shared by Israel and White-eye W′ sequences

[4]SNP and indel positions shared by Israel and White-eye W′ sequences, excluding any positions at which the SNP or indel is shared by either Z sequence

[5]Means and standard deviations were calculated using all A1 scaffolds listed above

[6]Means and standard deviations were calculated using all A1 scaffolds excluding A1.46

Table 2.5.: Mean and Standard deviation of Israel Z and W′ SNPs and indels per 50kb of chromosome A1 scaffolds

	Israel W′[1]		Israel Z[2]		Both IS[3]		Only IS [4]	
scaffold	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**
A1.1	174	65	174	67	168	63	140	59
A1.3	173	99	174	99	163	94	121	101
A1.4	268	145	270	149	254	139	224	129
A1.5	60	42	60	40	56	39	49	31
A1.6	204	135	206	126	195	127	180	119
A1.7	132	104	132	108	122	101	101	96
A1.8	223	72	223	72	213	69	207	68
A1.10	232	123	231	125	220	119	218	120
A1.11	199	54	194	53	188	51	167	55
A1.12	258	155	258	161	243	156	185	188
A1.13	171	131	169	126	161	119	132	93
A1.14	86	51	86	53	76	51	43	38
A1.15	177	101	179	103	160	101	84	102
A1.16	238	84	237	83	223	84	161	116
A1.17	212	81	214	79	206	80	151	64
A1.18	220	125	221	123	212	117	184	103
A1.19	207	31	206	31	198	30	171	33
A1.20	229	109	231	108	218	107	173	91
A1.21	152	54	152	49	143	50	112	56
A1.22	180	56	180	56	176	54	159	42
A1.23	263	90	261	94	252	91	177	44
A1.24	190	82	190	81	184	81	169	73
A1.25	193	52	190	62	183	64	181	64

Continued on next page

Table 2.5 – *Continued from previous page*

	Israel W′[1]		Israel Z[2]		Both IS[3]		Only IS [4]	
scaffold	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**
A1.26	140	68	128	61	125	61	116	63
A1.27	64	76	62	75	59	73	35	50
A1.28	104	79	102	79	98	78	90	77
A1.29	136	94	135	94	130	92	117	84
A1.30	92	20	92	19	86	18	71	27
A1.32	133	147	206	165	98	125	84	123
A1.33	292	173	223	149	69	70	3	3
A1.34	223	80	268	92	64	53	4	2
A1.35	172	78	199	81	71	28	18	16
A1.36	336	112	273	84	86	32	3	2
A1.37	240	45	188	60	143	81	11	9
A1.38	57	17	127	83	23	2	8	8
A1.39	8	7	6	4	5	4	0	0
A1.40	20	14	21	17	15	11	0	1
A1.41	304	146	416	144	123	88	67	78
A1.42	227	115	228	116	217	113	153	117
A1.43	231	169	304	158	92	86	30	31
A1.44	254	165	314	130	76	43	29	26
A1.45	226	104	232	106	61	45	8	13
A1.46	523	112	324	106	102	44	6	7
A1.47	287	130	307	112	87	52	19	16
A1[5]	223	100	219	92	131	80	89	88
A1 excl. A1.46 [6]	197	85	210	90	134	81	96	87

[1]Israel W′ sequence SNP and indel positions with respect to the reference sequence

[2]Israel Z sequence SNP and indel positions with respect to the reference sequence

[3]SNP and indel positions shared by Israel W′ and Z sequences

[4]SNP and indel positions shared by Israel W′ and Z sequences, excluding any positions at which the SNP or indel is shared by either White-eye sequence

[5]Means and standard deviations were calculated using all A1 scaffolds listed above

[6]Means and standard deviations were calculated using all A1 scaffolds excluding A1.46

Table 2.6.: Mean and Standard deviation of White-eye Z and W′ SNPs and indels per 50kb of chromosome A1 scaffolds

	White-eye W′[1]		White-eye Z[2]		Both WE[3]		Only WE[4]	
scaffold	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**
A1.1	160	79	160	78	154	78	126	70
A1.3	160	66	164	67	134	70	108	68
A1.4	132	106	132	106	122	101	95	79
A1.5	136	21	135	22	130	17	122	9
A1.6	71	17	72	17	65	16	54	16
A1.7	116	60	116	60	109	58	85	59
A1.8	38	44	38	43	35	41	29	36
A1.10	16	9	16	9	15	9	12	9
A1.11	144	81	142	80	133	79	114	75
A1.12	156	56	159	52	143	56	87	79
A1.13	144	128	144	134	139	127	107	96
A1.14	266	200	268	203	255	198	218	205
A1.15	193	122	195	121	177	120	98	131
A1.16	173	91	170	87	156	85	94	96
A1.17	199	75	197	75	191	76	136	50
A1.18	163	44	164	47	153	46	128	51
A1.19	192	57	191	57	184	55	155	49
A1.20	273	133	274	133	260	130	216	118
A1.21	171	31	172	30	162	30	134	30
A1.22	146	62	146	64	142	61	126	57
A1.23	272	82	275	80	265	80	190	31
A1.24	202	70	201	70	197	71	182	70

Continued on next page

Table 2.6 – *Continued from previous page*

	White-eye W′[1]		White-eye Z[2]		Both WE[3]		Only WE[4]	
scaffold	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**	**m**	**st dev**
A1.25	221	67	220	68	218	67	211	65
A1.26	140	40	145	34	137	42	122	62
A1.27	60	71	59	69	54	67	31	44
A1.28	97	58	98	57	90	59	85	59
A1.29	174	102	174	102	169	101	154	96
A1.30	211	140	214	139	205	136	190	126
A1.32	224	145	226	144	216	144	153	128
A1.33	340	155	358	160	214	157	39	91
A1.34	200	81	287	86	56	22	2	3
A1.35	187	86	172	69	48	29	10	9
A1.36	337	113	216	103	62	36	2	2
A1.37	217	25	175	84	93	62	33	40
A1.38	200	83	112	78	71	43	21	23
A1.39	6	6	9	5	5	5	0	0
A1.40	23	19	21	16	17	14	1	1
A1.41	269	176	283	185	186	165	81	107
A1.42	226	104	171	95	92	68	76	64
A1.43	205	170	200	152	44	49	13	18
A1.44	348	158	372	118	148	76	50	33
A1.45	226	108	234	122	117	72	18	35
A1.46	522	110	312	100	95	42	4	3
A1.47	243	133	239	132	68	52	13	15
A1[5]	218	104	199	89	131	80	89	88
A1 excl. A1.46[6]	192	85	189	86	139	70	91	66

[1]White-eye W′ sequence SNP and indel positions with respect to the reference sequence

[2]White-eye Z sequence SNP and indel positions with respect to the reference sequence

[3]SNP and indel positions shared by White-eye W′ and Z sequences

[4]SNP and indel positions shared by White-eye W′ and Z sequences, excluding any positions at which the SNP or indel is shared by either Israel sequence

[5]Means and standard deviations were calculated using all A1 scaffolds listed above

[6]Means and standard deviations were calculated using all A1 scaffolds excluding A1.46

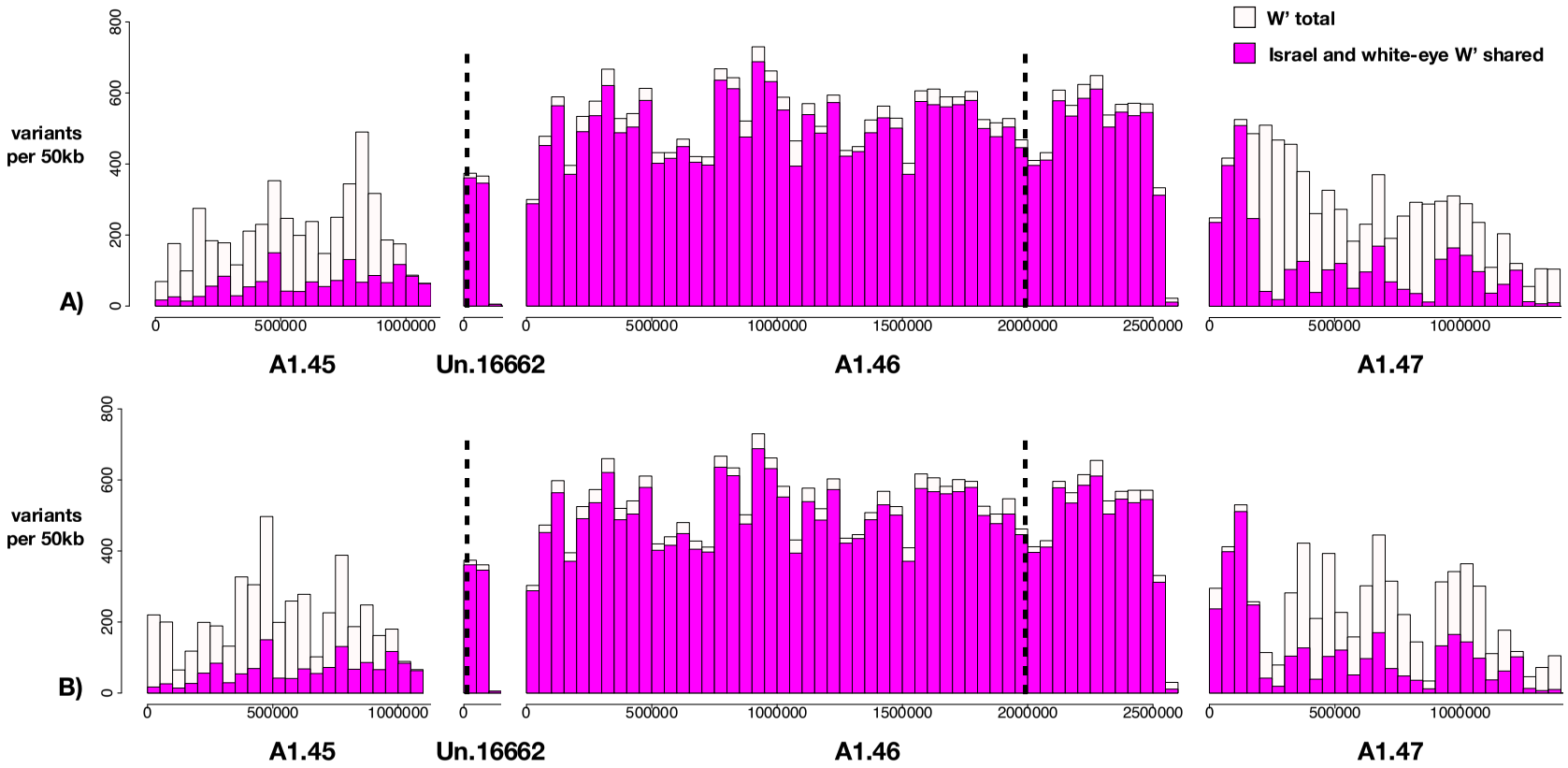

Fig. 2.3. Number of positions per 50 kb in the In(A1q1) region at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Israel and White-eye W′ sequences are shown in pink. Inversion breakpoints are indicated by dotted lines. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis.

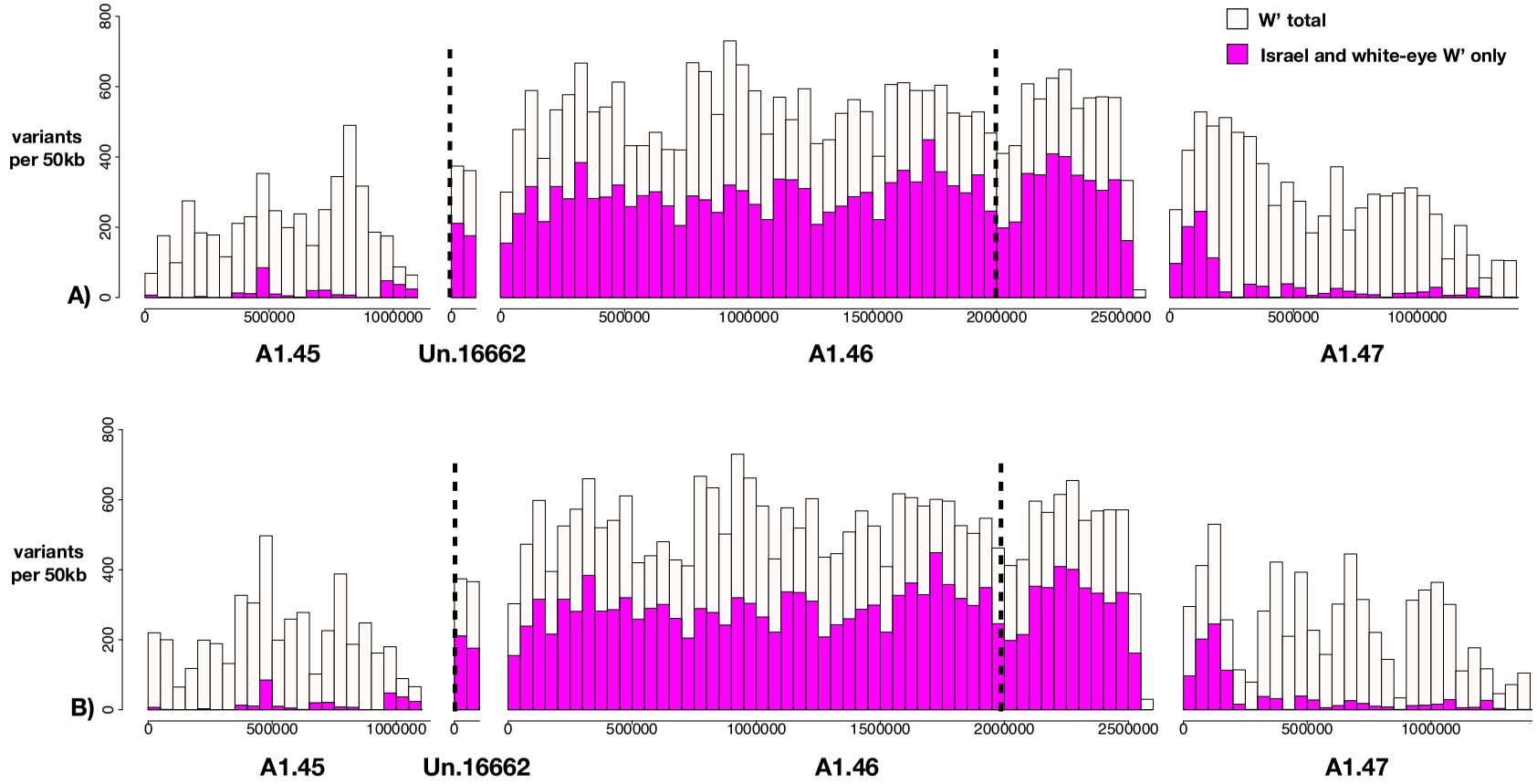

Fig. 2.4. Number of positions per 50 kb in the In(A1q1) region at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Israel and White-eye W′ sequences, excluding positions at which the SNP or indel is shared with either Z sequence, are shown in pink. Inversion breakpoints are indicated by dotted lines. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis.

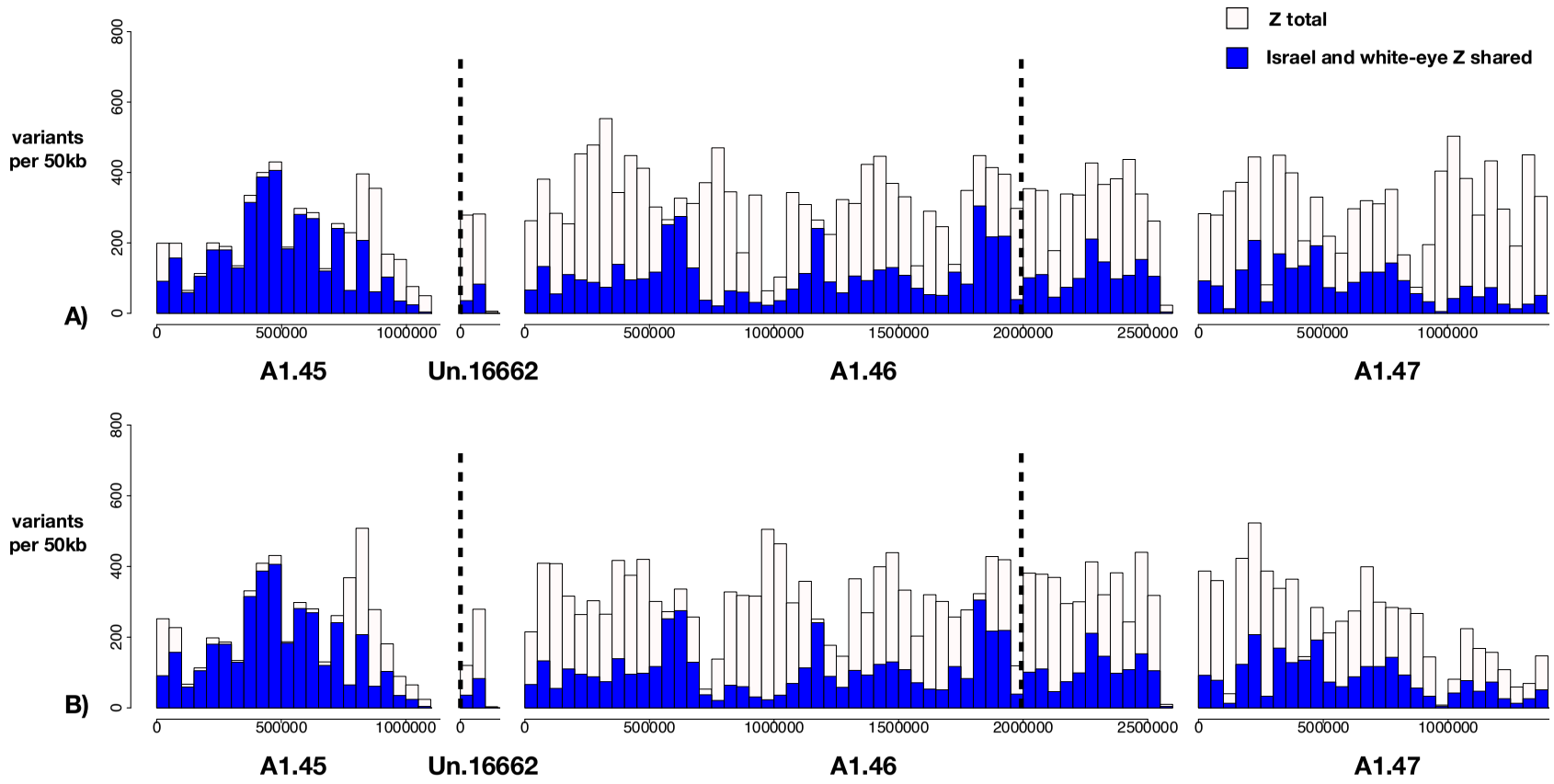

Fig. 2.5. Number of positions per 50 kb in the In(A1q1) region at which the Z sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Israel and White-eye population Z sequences are shown in blue. Inversion breakpoints are indicated by dotted lines. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis.

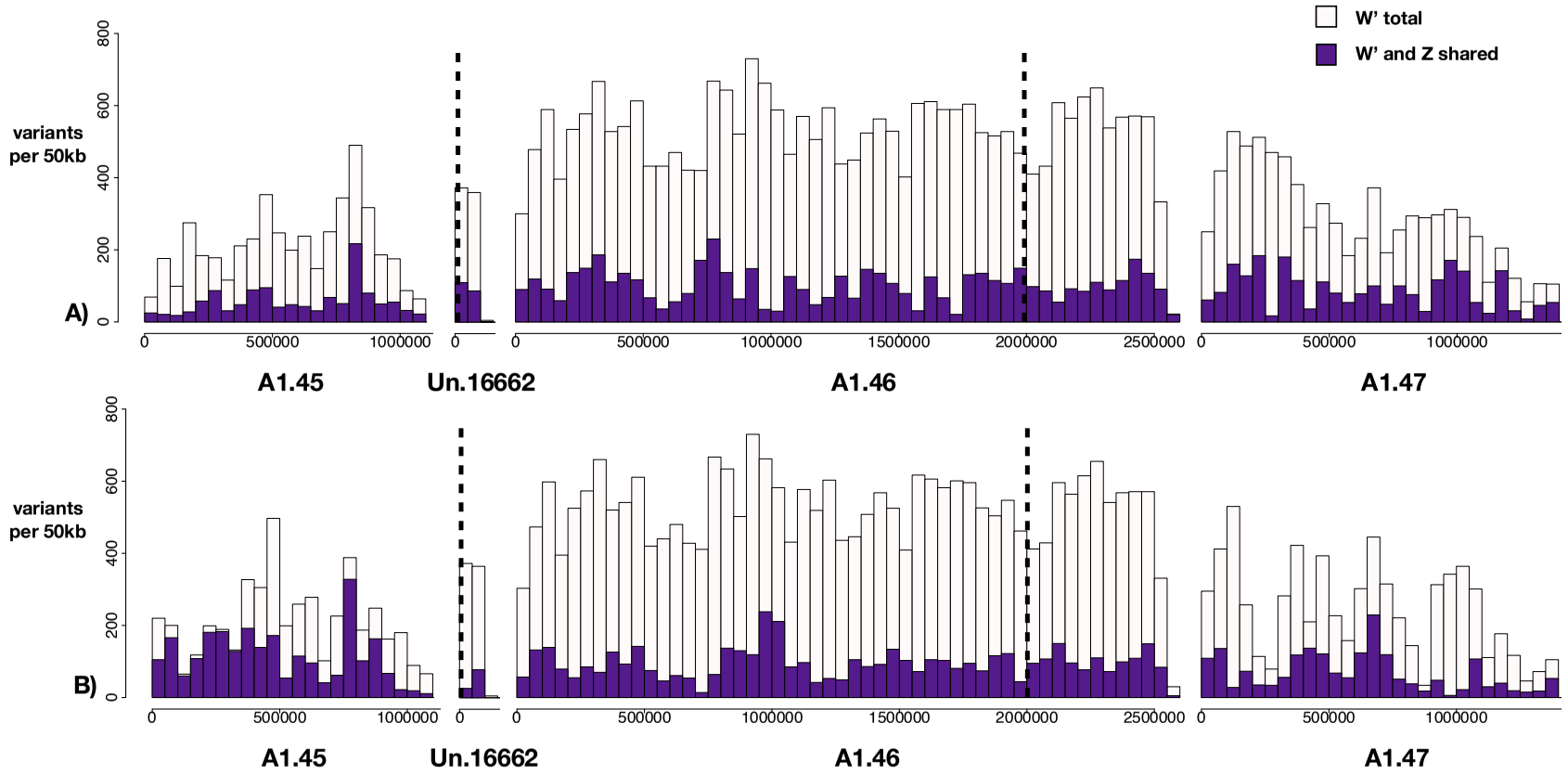

Fig. 2.6. Number of positions per 50 kb in the In(A1q1) region at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Z and W′ sequences within each population are shown in purple. Inversion breakpoints are indicated by dotted lines. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis

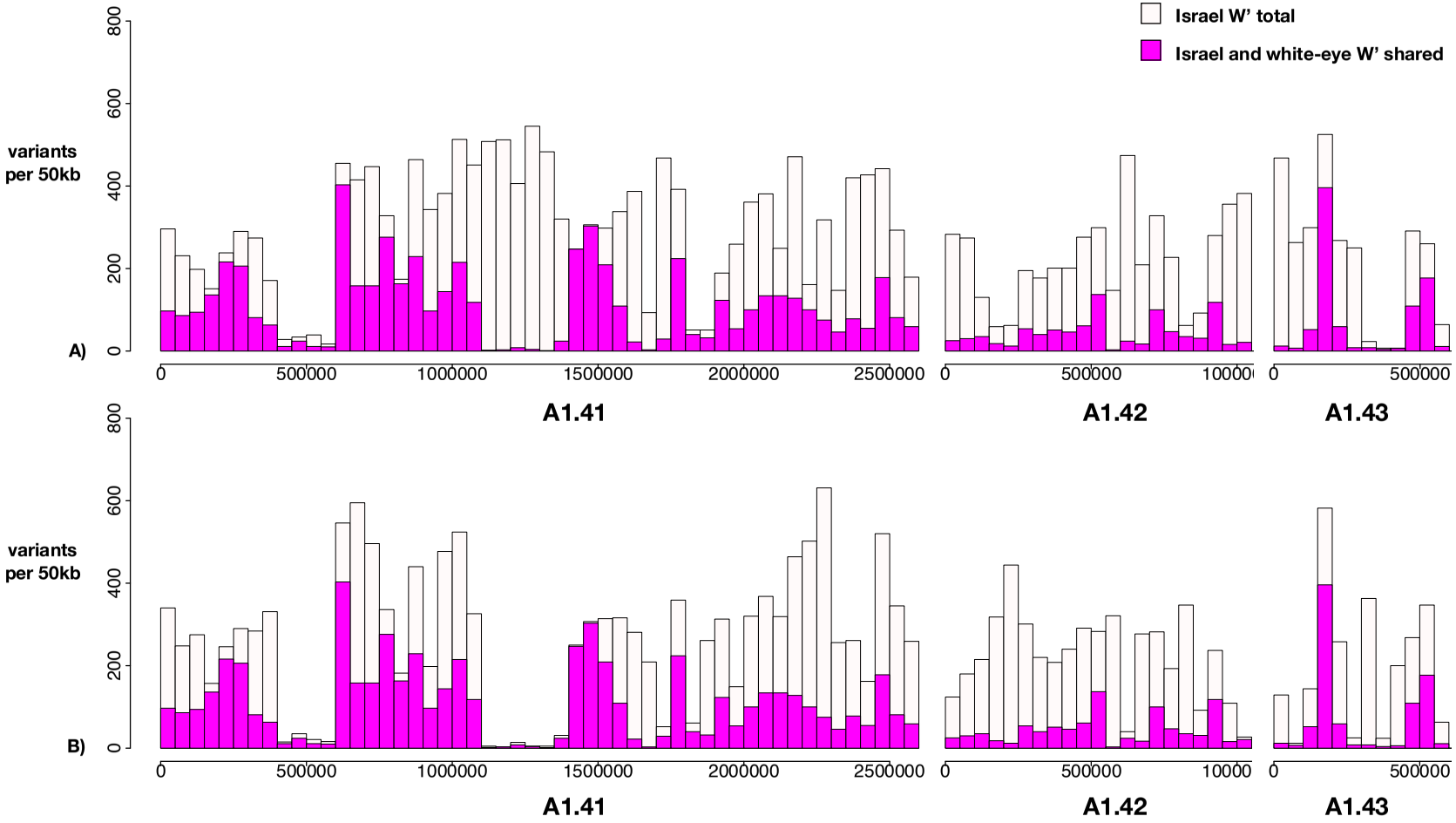

Fig. 2.7. Number of positions per 50 kb in the In(A1q2) region at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Israel and White-eye W′ sequences are shown in pink. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis.

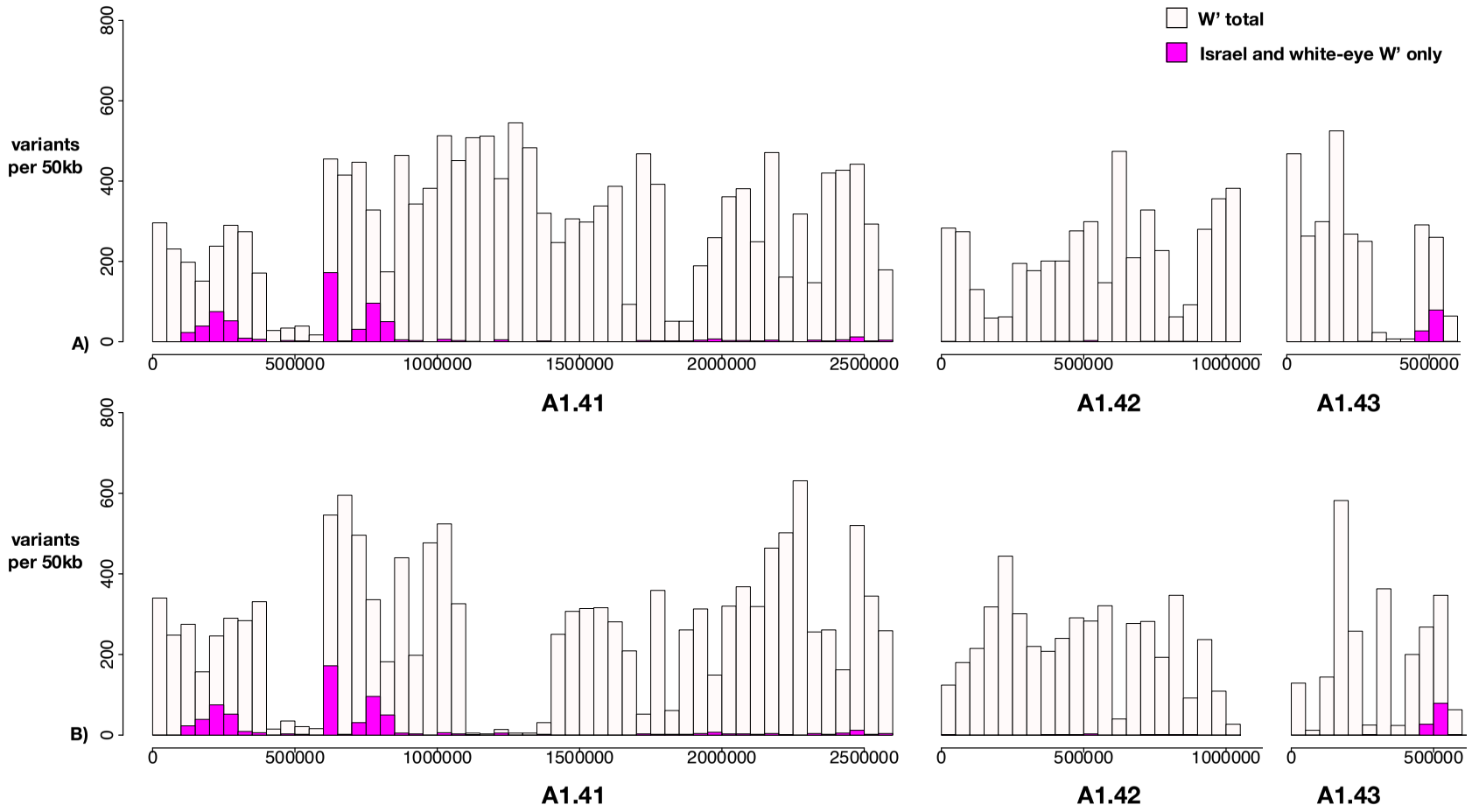

Fig. 2.8. Number of positions per 50 kb in the In(A1q2) region at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Israel and White-eye W′ sequences, excluding positions at which the SNP or indel is shared with either Z sequence, are shown in pink. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis.

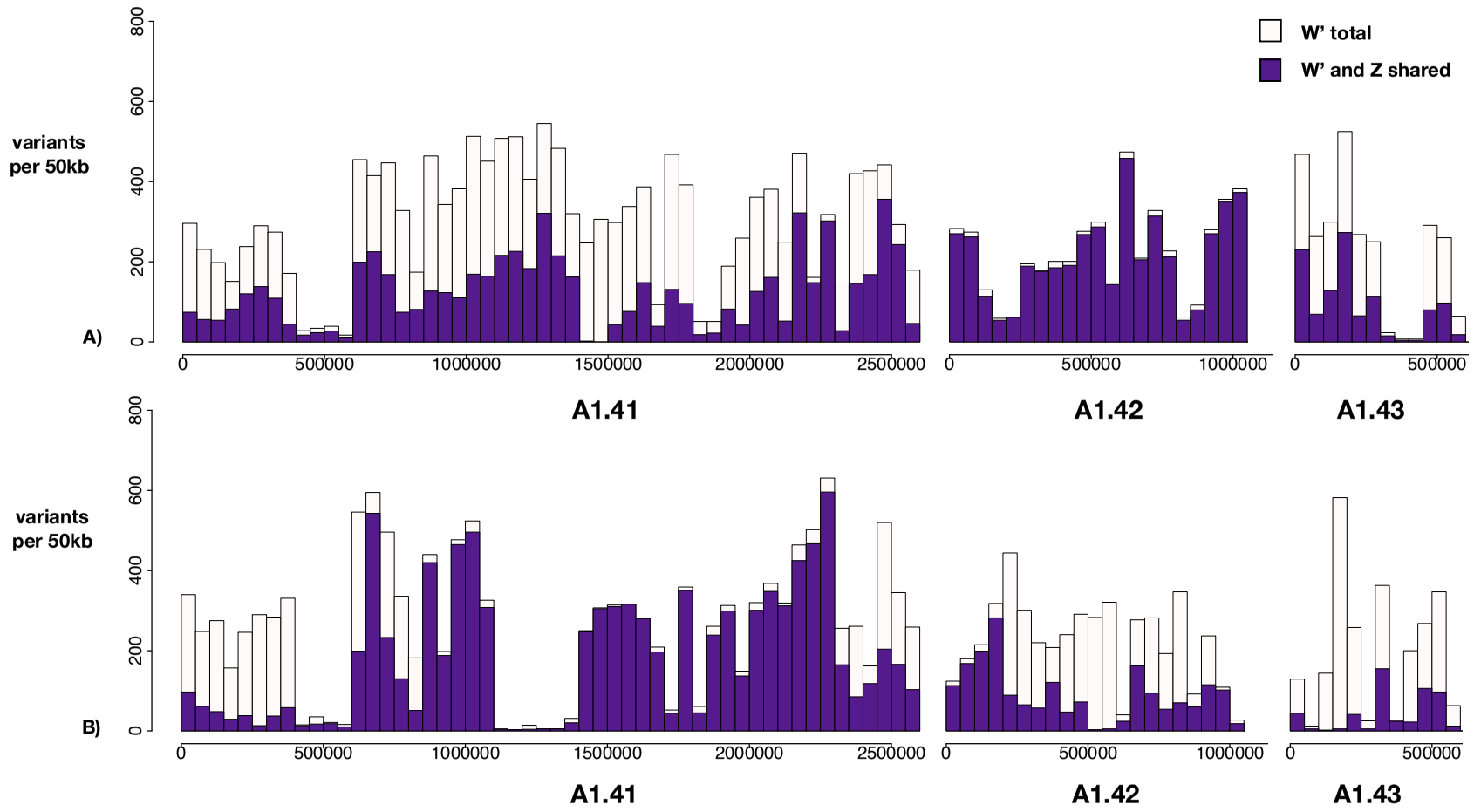

Fig. 2.9. Number of positions per 50 kb in the In(A1q2) region at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Z and W′ sequences are shown in purple. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis.

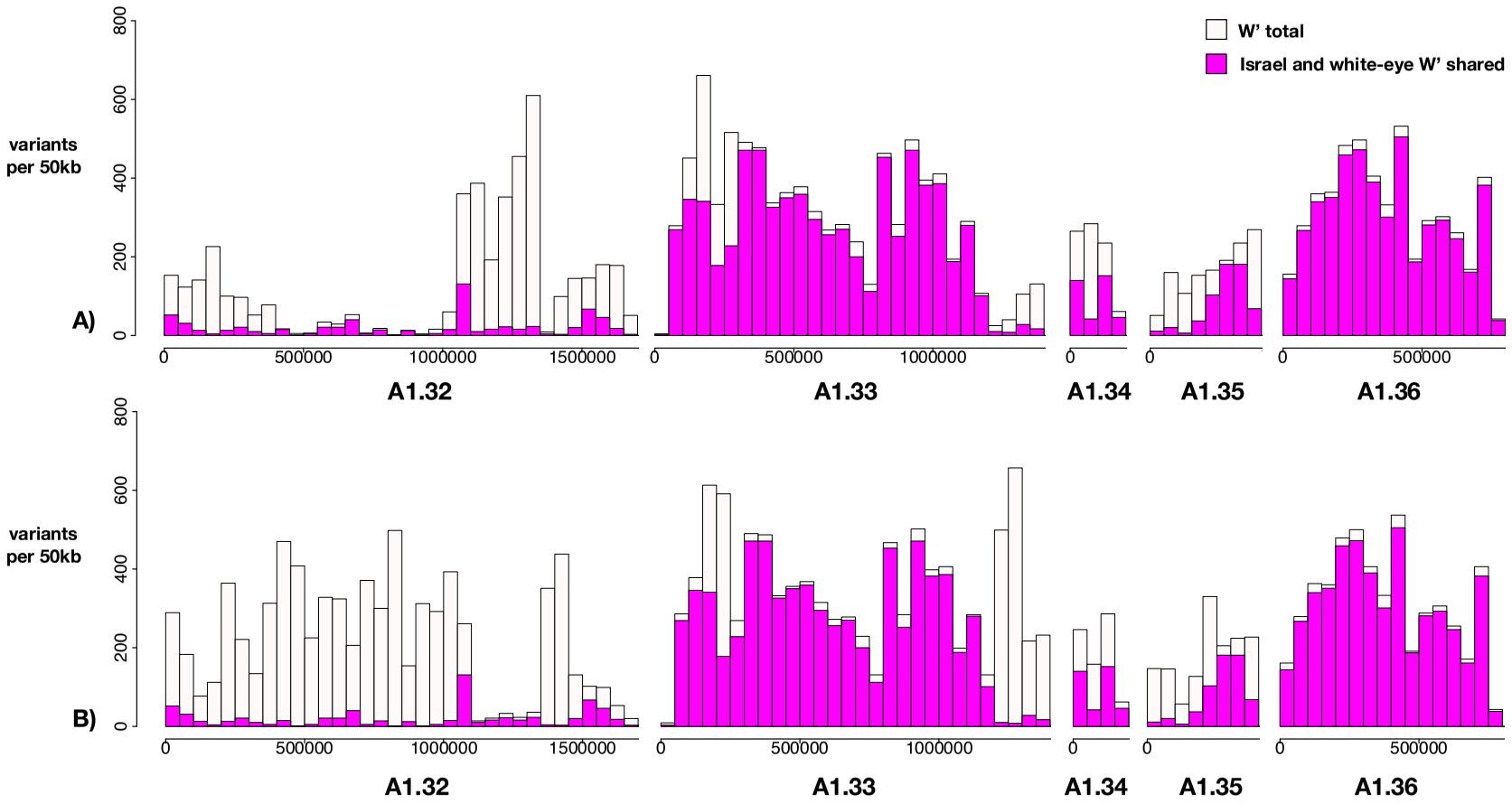

Fig. 2.10. Number of positions per 50 kb in scaffolds A1.32–A1.36 at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Israel and White-eye W′ sequences are shown in pink. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis.

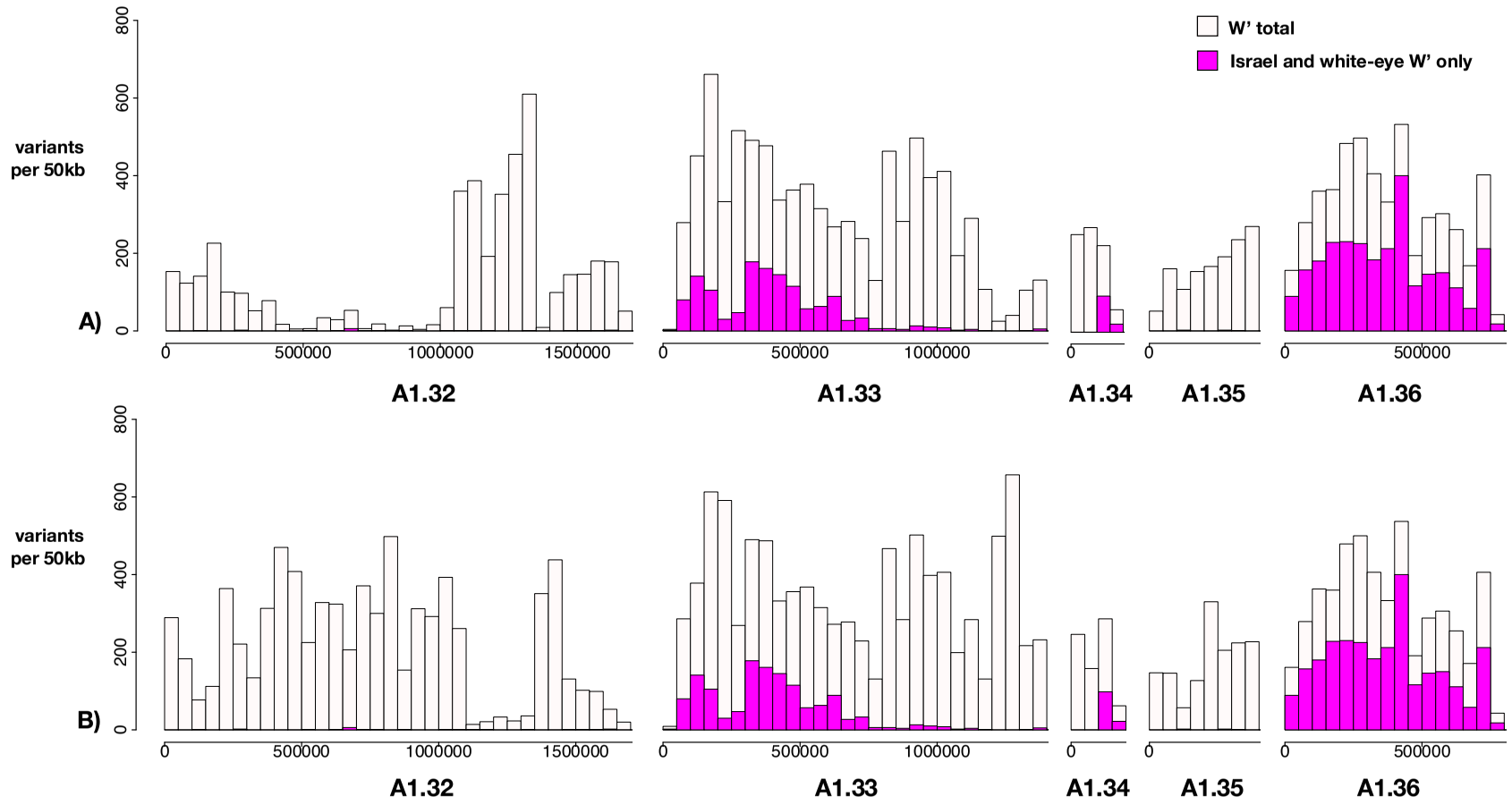

Fig. 2.11. Number of positions per 50 kb in scaffolds A1.32–A1.36 at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Israel and White-eye W′ sequences, excluding positions at which the SNP or indel is shared with either Z sequence, are shown in pink. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis.

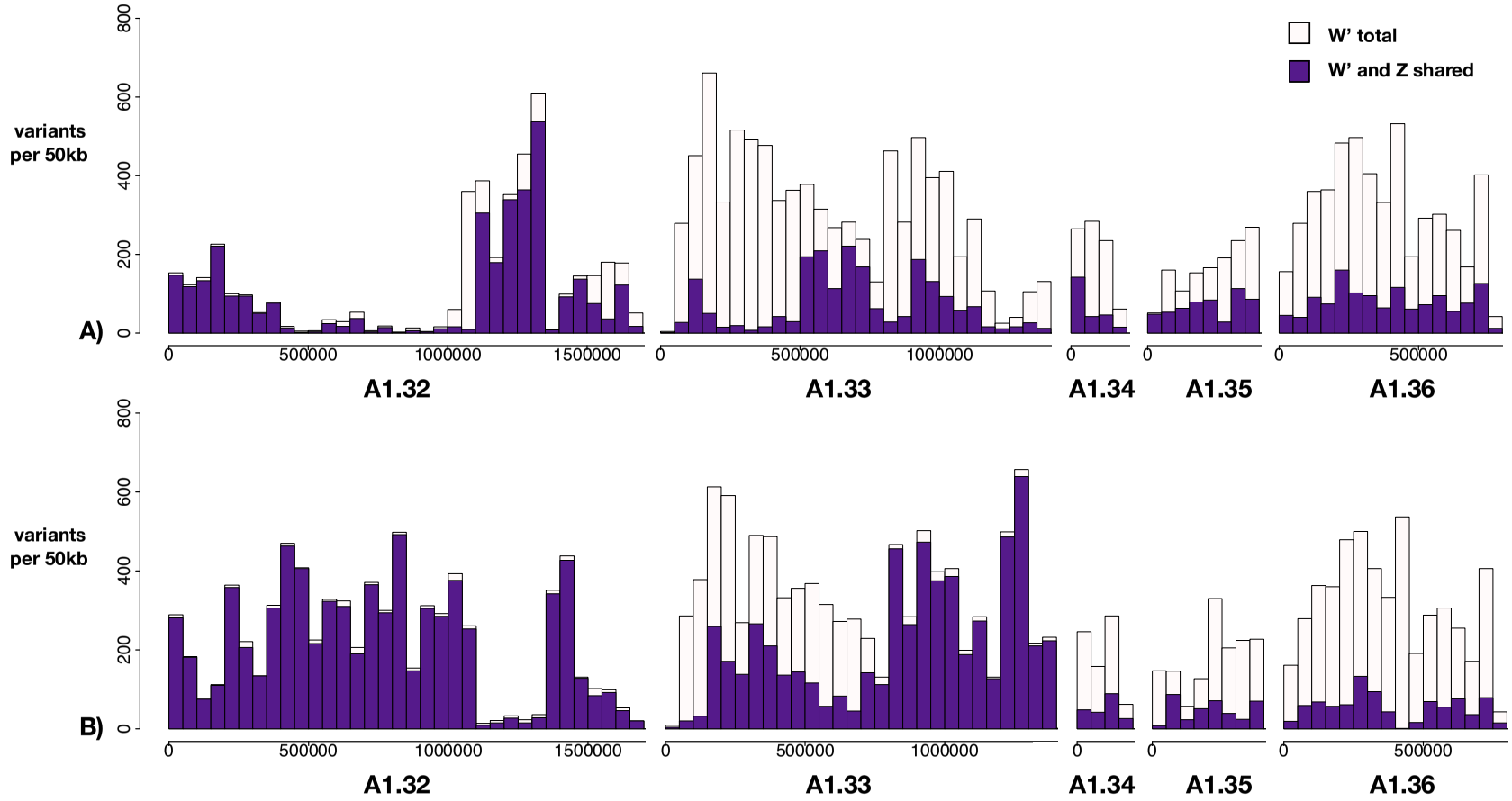

Fig. 2.12. Number of positions per 50 kb in scaffolds A1.32–A1.36 at which the W′ sequence has a SNP or indel with respect to the reference genome sequence for A) Israel and B) White-eye population sequences. The subset of these positions that are shared between the Z and W′ sequences within each population are shown in purple. Each histogram bar width represents 50 kb. Positions within each scaffold are numbered on the x-axis

3. INVERSION GENES

3.1 Introduction

The unusual sex chromosome system of the Hessian fly provides an excellent opportunity to investigate both the evolution of sex determination mechanisms and chromosome behavior. The female Hessian fly has two copies of each X chromosome whereas males have only the maternal copies. However, there is no heterogametic sex; both the male and female gamete contribute a copy of each X chromosome to the zygote [3]. The sex-determining karyotype of the Hessian fly is established during early embryogenesis through elimination of paternal X chromosomes in future males and retention of these chromosomes in future females [3]. The genotype of the mother determines whether her offspring will become male or female; if she has the sex-determining autosomal inversion In(A1q1), her offspring will all be female [2]. If she lacks the inversion, she will most likely produce all male offspring; a small percentage of females lacking the inversion are bisexual, having a form of A1 that results in production of both male and female offspring. The inversion acts as a maternal-effect neo-W chromosome, suppressing recombination around at least one sex-determining gene [2]. It is present only in females and inherited heterozygously so that about half of the females in the population are female producers (ZW′) and the other half are male producers (ZZ). The Hessian fly reference genome was sequenced from individuals homozygous for the bisexual form of A1 [1], which is recessive to the inversion and dominant to the male-producing form.

The mechanism by which paternal X chromosomes are rescued in future females is unknown. The sciarids, or fungus gnats, are a sister group to the cecidomyiids and have a similiar sex determination system in which the genotype of the mother determines the sex of the offspring by allowing either the elimination or retention of paternal sex chromosomes. In sciarids, however, chromosome elimination occurs in both sexes; future females lose only one paternal X chromosome while future males lose two [26].

As in the Hessian fly, the maternal factor that rescues paternal X chromosomes in sciarids is unknown. Two models have been proposed: In the one-factor model, a maternal factor (MF) interacts with paternal X chromosomes, facilitating their elimination. In the two-factor model, MF prevents X chromosome elimination by binding a chromosome factor (CF) and preventing it from interacting with the chromosome. In both models, it is assumed that the number of paternal X chromosomes eliminated depends on the amount of maternal factor present [27]. Because all or none (rather than one or two) of the paternal X chromosomes are eliminated in the Hessian fly, it is not necessary to assume that the maternal factor must reach a specific concentration to fulfill its role; its presence or absence alone may be sufficient to make the difference between chromosome elimination and retention.

Sex determination cascade

The chromosome behavior of cecidomyiids and sciarids is not the only unusual feature of their sex determination. The genetic basis of sex determination in insects is best characterized in Drosophila. The number of sex chromosomes provides the signal which initates a splicing-regulation cascade ending in sex-specific somatic and germline development and dosage compensation [7]. Genes at the bottom of the cascade are more highly conserved than those at the top and evolution of the pathway is thought to have begun with the most downstream genes in the pathway with later incorporation of upstream regulators [8].

The master switch at the top of the Drosophila sex determination cascade, sex lethal (sxl), is spliced into multiple forms which regulate both its own splicing and the splicing of other genes involved in sex determination. Default splicing of sxl produces the male-specific form, which encodes a non-functional protein. In the embryo, female-specific splicing of sxl is initially signaled by the karyotype, resulting in a functional protein which proceeds to regulate continued female-specific splicing of sxl throughout development [9]. The role of sxl in sex determination appears to be restricted to Drosophila; in other insects, including the Hessian fly and sciarids, sxl does not undergo sex-specific splicing [10], [11].

The next step in the sex-determination cascade is the sex-specific splicing of transformer (tra), which is controlled by sxl in Drosophila. Only the female-specific form of tra encodes a functional protein, which in turn participates in the female-specific splicing of doublesex [12]. In several insects including the medfly, housefly, wasp, and honeybee, tra fulfills a role similar to that of Drosophila sxl by responding to the primary sex determination signal and regulating its own splicing in a positive feedback loop [7]. While tra varies among insects in its sequence and interactions with upstream regulators, its role in the splicing of doublesex is more highly conserved [12]. Tra has not be identified in the Hessian fly [10] or in Sciara, though attempts have been made to do so using highly conserved parts of the sequence.

Transformer-2 (tra-2) cooperates with tra in the female-specific splicing of doublesex (dsx) [13]. Tra-2 does not generally undergo sex-specific splicing; this is true in both the Hessian fly and Sciara, as well. Tra-2 has an RNA binding domain and RS domain required for its doublesex-splicing function. The RNA-binding domain is conserved in Hessian fly and sciarid tra-2 while the RS domains vary more from those of other insects [10]. Sciara tra-2 is able to participate in the splicing of Drosophila doublesex, though it is less effective than Drosophila tra-2 [14]. Functional analysis of this gene has not been done in the Hessian fly.

Doublesex (dsx) is a transcription factor that in most insects controls sex-specific differentiation at the end of the sex determination cascade [8]. The male form of the dsx protein, DSXM, activates expression of genes required for male differentiation while the female form, DSXF, does the opposite [15]. In the Hessian fly, dsx is spliced into sex-specific forms similar to those typically found in other insects [10]. Sciaria doublesex expression is atypical; all splice variants are present in both sexes, though the quantities are sex-specific. Only the female-specific protein DSXF is translated and is present in both sexes. For these reasons, dsx is unikely to play a discriminatory role in sciarid sex determination as it does in most insects [16].

Behavior of eliminated chromosomes in the Hessian fly and insects with similar sex determination systems

In the Hessian fly, paternal X chromosomes in future male embryos are eliminated during an abnormal anaphase. During anaphase of the fifth mitotic division, the chromosomes to be eliminated appear to begin moving toward the poles (a small space is visible between the sister chromatids) but return to the metaphase plate while the somatic (S) chromosomes move to the poles [4]. The paternal X chromosomes in future males are probably eliminated at the same time as the germline-limited (E) chromosomes; they could not be visibly distinguished, but no chromosome elimination has been observed at any time [3]. Additionally, both E chromosomes and all paternally-derived S chromosomes are eliminated in male germ cells during spermatogenesis whereas these chromosomes are retained in female gametes [3].

The sciarids are a sister group to the cecidomyiids, the family of the Hessian fly, and have a similar sex determination mechanism in which the mother determines the sex of her offspring through elimination of paternal X chromosomes during early embryogenesis [28]. Both germline-limited and paternal X chromosomes are eliminated in early mitotic divisions, but these eliminations occur separately. In both elimination events, the behavior of chromosomes to be eliminated is similar to that of the Hessian flys eliminated chromosomes; the chromosomes appear to enter anaphase but are left behind while the chromosomes to be retained are able to reach the poles [26]. Failure of the eliminated chromosomes to reach the poles appears to result from incomplete detachment of sister chromatids [29]. Sciarids also eliminate germline-limited chromosomes and all of the paternally-derived S chromosomes during male gametogenesis [26].

In mealybugs, sex determination is also controlled by maternal factors. Rather than elimination of the paternal X chromosomes, though, the entire set of paternally-derived chromosomes in future males is silenced through conversion to heterochromatin during the seventh mitotic division [30]. Male mealybugs also discard their paternally-derived chromosome set during gametogenesis; the spindle microtubules attach only to euchromatic ma-

ternal chromosomes while the heterochromatic paternal chromosomes are left behind and eliminated [31].

Imprinting and Chromosome Elimination

The term imprinting, epigenetic modification of chromosomes based on parental origin, was coined to describe differences in morphology and behavior between the maternally and paternally-derived chromosomes of sciarids [32]. Imprinting occurs either during gametogenesis and remodeling of paternal chromatin in the fertilized egg, when the two parental genomes are physically separate [33]. Markers of imprinting are maintained through the cell cycle but must be reversible, as maternally-derived chromosomes in male gametes must be recoded as paternally-derived in the next generation. Both DNA cytosine methylation and histone modifications are associated with imprinting, though DNA methylation is typically a long-term mark whereas histone modification is more dynamic and dependent on context [34], [35].

Differences between packaging of maternal and paternal chromosomes begins in gametogenesis. In male gametes, chromatin is repackaged with sperm-specific proteins called protamines to allow it to be tightly packed to fit into the sperm nucleus [36]. Successful passage through the cell cycle requires remodeling of paternal chromatin upon fertilization by maternal factors in the oocyte, which remove protamines and supply histones [37], [38]. Mutations in maternal effect genes in Drosophila maternal haploid and sesame impair remodeling of paternal chromosomes, resulting in their loss during the first mitotic division [39], [40]. A Drosophila zygote resulting from the union of a Wolbachia-infected sperm and uninfected egg will also lose its paternal chromosomes in the first mitotic division, possibly due to delayed replacement of sperm protamines with H3.3 and H4 [41]. By an unknown mechanism, only eggs that are also infected with Wolbachia are able to rescue the paternal chromosomes. In plants, uniparental chromosome loss due to hybrid incompatibility frequently involves the inability of paternal chromosomes to be modified with maternal CENH3, a centromere-specific variant of histone H3 involved in chromosome segregation [42], [43], [44].

In cecidomyiids and sciarids, all chromosomes successfully complete the first few mitotic divisions before they are eliminated. Mealybugs complete six mitotic divisions before silencing their paternal chromosomes. These chromosomes may be modified in a way that allows them to be distinguished from maternally-derived chromosomes but requires a specific context to have an impact their behavior. During mitosis, bookmarking is used to preserve the identity of different chromosomal regions and allow them to resume their functions outside of mitosis; while some histone modifications must be removed, other modifications or protein complexes may temporarily hold their place [45]. Histone modifications also may recruit protein complexes that further modify chromatin in the region or change its gene expression [46].

Chromosome behavior during mitosis

Beginning in prophase, chromosomes must become condensed to allow sister chromatids to be distinguished and to be transported quickly and without damage during anaphase [47]. The serine-threonine kinase Aurora B, as part of the Chromosomal Passenger Complex (CPC) [48], phosphorylates multiple sites on histone H3 [49]. Phosphorylation of H3 is linked to removal of heterochromatin protein 1 (HP1) [50] and recruitment of condensin [51], which is used to organize the chromosomes into their mitotic structures [52]. Monomethylation at H4K20 is also required for cell cycle progression and chromosome condensation in Drosophila, humans, and fission yeast [53]. The centromere has its own highly conserved variants of histone H3 [54] and a distinctive pattern of histone modifications [55]. The centromere must be clearly defined both as the site of sister chromatid attachment and assembly of the kinetochore [56], a large multilayered protein complex that binds mitotic spindle microtubules to direct chromosome segregation [57].

During prometaphase of the early syncytial mitotic divisions of Drosophila, the nuclear membrane breaks down only around the spindle poles [58] and is remodeled into a spindle membrane [59]. In the spindle membrane, lamins continue to interact with mitotic chromosomes [60].

After nuclear envelope breakdown, microtubules of the mitotic spindle radiating from centrosomes at opposite poles of the cell begin to interact with the kinetochores of the attached sister chromatids. Binding of a microtubule to a kinetochore stabilizes the microtubule and orients the kinetochore of the sister chromatid toward the opposite pole [61]. When both kinetochores of a pair of sister chromatids are bound to spindle microtubules from opposite poles, they begin to move toward the metaphase plate [61].

The spindle checkpoint does not allow anaphase to begin until every pair of sister chromatids is aligned at the metaphase plate with proper spindle attachments [62]. Kinetochores that are not properly bound to microtubules have checkpoint proteins associated with them, keeping the checkpoint active for the entire cell so that none of the chromosomes may begin movement [63]. These proteins sense the presence of an attachment between the kinetochore and microtubules and the amount of tension between the kinetochores of the sister chromatids [64].

Once the spindle checkpoint is cleared, the APC allows the centromeric cohesin to be removed so that the sister chromatids can begin moving to opposite poles [65]. Anaphase delay may be caused by attachment of a single kinetochore to microtubules of opposite poles [66]. These attachments are usually corrected by Aurora B during metaphase, but are occasionally able to bypass the spindle checkpoint. Cleavage of cohesin is sufficient to allow sister chromatid separation, but sister chromatids will not successfully migrate to the poles unless other cell cycle events associated with anaphase proceed on time [67]. If the centromere isnt activated for anaphase, the lack of tension at kinetochores resulting from sister chromatid separation allows spindle checkpoint proteins to reassociate with the kinetochore, blocking anaphase.

Failure of sister chromatids to completely separate, which is thought to be responsible for paternal X chromosome elimination in sciarids [29], will also prevent them from reaching the poles. During a normal anaphase, the phosphorylation on histone H3 that is required for initial chromosome compaction begins to decrease [68]. In sciarids, phosphorylation of H3 is elevated on the arms of the paternal X chromosomes to be eliminatedwhich undergo

delayed segregation compared to the rest of the chromosomesduring anaphase of the early mitotic divisions of embryogenesis [69].

To identify genes with a potential role in the fate of paternal X chromosomes or in other aspects of sex determination, genes within the inversion were annotated and the sequences were compared between the translated Z and W forms. Genes were also annotated in scaffold A1.36, which has recently been identified as a region in which the female producer sequences differ from the male producer sequences. Several genes involved in chromosome behavior and modification were identified and are discussed as potential candidates for the maternal factor that determines whether the paternal X chromosomes are eliminated, resulting in male offspring.

3.1.1 Methods

Blast2go annotation

Blast2go and Interproscan were used to annotate genes within the inversion and A1.36 scaffolds with the Hessian fly official gene set (OGS), which was downloaded from i5k. BLASTP was used with the translated reference gene sequences using the nr database. The OGS was made previously by others using: the CEGMA gene set, Hessian fly cufflinks output and ESTs, and protein homology from fruit fly, mosquito, honey bee, and Tribolium to train Maker 2.22. Maker output was then used to train Augustus and SNAP to make the gene models, which were manually edited by the community [1].

Cluego gene enrichment analysis

EggNOG mapper was used to assign GO terms to Hessian fly translated gene sequences from the official gene set. The Diamond mapping mode was used. The taxonomic scope was restricted to the insects. The gene IDs with their corresponding GO terms were then sent to the ClueGO team to make the reference files for ClueGO analysis.

Cluego version 2.5.1 was used to group GO terms based on their shared genes and test for enrichment of the terms and groups from scaffolds A1.46 and A1.36. All evidence codes and the following ontologies were used: Biological Process, Cellular Component, Molecular Function, and Immune System Process. GO terms from levels three to eight were included in the analysis. A minimum of two genes genes per GO term and four percent of the total number of genes associated with each term were required to come from the scaffold of interest (A1.36 or A1.46) for the term to be included in the analysis. A right-sided hypergeometric test was used to test for enrichment of GO terms and groups. A Bonferroni step down procedure was used to correct p values for multiple testing. The GO fusion option was used to reduce redundancy from GO terms with parent-child relationships. Groups including at least two terms, sharing at least half of their genes and half of their terms, and having a Kappa score of at least 0.4 were iteratively merged.

Comparing Z, W′ and reference translated gene sequences

OGS (nucleotide) sequences from the inversion scaffolds were edited by replacing reference alleles with the chromosome-specific alleles of the W′ and Z sequences of the Israel and white-eye populations. Sequences were then translated into all potential open reading frames using EMBOSS Transeq. The longest potential coding sequence was then used to compare the differences among these four sequence and the reference sequence.

3.2 Results

3.2.1 GO term enrichment in inversion scaffolds and in scaffold A1.36

Of the 20,163 total genes in the official gene set, 14,371 were assigned results from the EggNOG mapper. Of these, 7,272 were assigned GO terms; 5,730 were assigned terms from the Molecular Function ontology, 5,195 from the Cellular Component ontology, 6,822 from the Biological process ontology, and 358 from the Immune System Process ontology. These genes served as the reference set for the hypergeometric tests for GO term and group

enrichment. GO terms and groups with a corrected p value less than 0.05 will be referred to as enriched.

Scaffold A1.46

For Scaffold A1.46, 94 genes were assigned GO terms. After filtering and grouping GO terms according to criteria listed in the methods, the remaining GO termsrepresented by a total of 45 geneswere organized into 10 groups and 5 ungrouped terms (Figure 3.1 and (Table 3.1)).

Three of these GO terms were enriched: morphogenesis of a polarized epithelium, which was mapped to nine genes; adherens junction, mapped to eight genes, and photoreceptor cell fate commitment, mapped to six genes.

The only enriched groups were those numbered 5 (with five terms) and 7 (with ten terms), which consisted of terms related to cell-cell adhesion, the plasma membrane, Notch signaling, and development. Both included the enriched adherens junction term. The largest group, 9, was not enriched and included 65 GO terms, most of which were related to cell fate, signaling, and organ development (Table 3.1).

Scaffold A1.36

For Scaffold A1.36, 48 genes were assigned GO terms. Most of these terms each only mapped to one gene. After filtering and grouping, 12 GO termsarranged into two groups and three ungrouped termswere represented by 11 genes. Almost all of these terms had p values less than 0.05 but were represented by only two genes (Figure 3.1, Table 3.2).

Group 1 was enriched and consisted of five terms related to histone modification and methyltransferase activity, all of which were mapped to three genes: histone-lysine N-methyltransferase pr-set7, N-lysine methyltransferase KMT5A-B, and tRNA (cytosine-5-)-methyltransferase. Group 0, which included the enriched term sex determination, establishment of X:A ratio was represented by only two genes, which were annotated as segmentation Runt isoform X1 and segmentation Runt-like.

3.2.2 Differences between Z and W′ in genes that may be involved in paternal X chromosome retention

Genes involved in sister chromatid cohesion: SMC3, PDS5, and BRCA2

Gene 61 (1192 residues) and gene 123 (1135 residues) of scaffold A1.46 were annotated as SMC3. Structural Maintenance of Chromosomes 3 (SMC3) forms part of the ring structure of cohesin, which physically tethers sister chromatids together until anaphase [70]. Both genes have a RecF/RecN/SMC N-terminal domain and a flexible hinge domain, which is required for sister chromatid cohesion [71]. For gene 61, the white-eye and Israel W′ sequences are different from the Z sequences only at position 1059 (glutamate in the W′ sequence and aspartate in the Z sequence), within the RecF/RecN/SMC N-terminal domain. Gene 123 is missing the ABC domain, which is required to open and close the cohesin ring; for this reason, gene 123 is probably not involved in sister chromatid cohesion as part of the cohesin ring. For gene 123, both W′ sequences were different from the Z sequences at 11 positions (Table 3.4). Most of these differences appear to be conservative, though there is one position (653) at which the Z residue is hydrophobic (leucine) and the W′ residue is polar (serine). There were no differences between the two Z sequences.

Sister chromatid cohesion protein PDS5 is involved in both stabilization of cohesin binding [72] and release of cohesin from sister chromatids [73]. Gene 150 (1244 residues) of scaffold A1.46 was annotated as PDS5. Both of the W′ sequences are different from the Z sequences at one position within the Armadillo fold domain. For both the Z and W′ sequences, though, the residue at this position (301) is polar and non-charged (serine in W′, threonine in Z). The W′ sequences (threonine) were also different from the Z sequences (alanine) at position 2, outside of the domain.

Gene 53 of scaffold A1.36 (1076 residues) was annotated as Breast Cancer Type 2 Susceptibility Protein (BRCA2), a tumor suppressor protein with roles in DNA repair, cell cycle checkpoint regulation, and sister chromatid cohesion [73]. Domains identified for this gene include a single BRCA2 repeat, a helical domain, and an oligonucleotide/oligosaccharide-binding domain. At 37 positions, both Israel and white-eye W′ sequences have different

residues from the Z sequences (Table 3.8); three of these positions are within the helical domain (positions 786-949). There are no differences between the Israel and white-eye Z sequences for this gene. At position 841, the Z residue is leucine and the W′ residue is proline. The structure of proline allows it to introduce turns in the amino acid chain and kinks in alpha helices [74]. Substitution of any other amino acid with a proline in the BRCA2 helical domain may disrupt its structure and interfere with its ability to form interactions. At the C terminal end, from positions 1070-1076, the W′ sequence is different from the Z sequence due to a frame shift caused by two indels. The W′ sequence also has an additional residue (alanine) as a result of the frame shift. Although these differences occur outside of the domains, they may disrupt the folding of the protein in a way that disrupts interactions of domains near the C terminal.

Genes that may target paternal X chromosomes for modification

nesprin-1

A1.46 gene 147, annotated as nesprin-1, is unusual in its high frequency of differences between the Z and W′ sequences and low frequency of differences between the two Z sequences.

This gene has two models, designated A and B, which each have 39 exons spanning approximately 80 kb (Figure 3.3). The first exon of model A is located upstream of the first exon of model B; it is very small in both models, with no overlap. The remaining exons are shared between the two models. Expressed sequence tag (EST) sequences from the Hessian fly data used to create the gene models were obtained from the i5k database and can be viewed there using gbrowse. Several of the largest ESTs aligning with gene 147 are represented in figures 3.3 and 3.4. Although no EST spans the entire gene, the largest ESTs aligning to each end of the gene overlap for several exons. Of the 39 gene model exons, 35 are present in at least one EST. The four exons lacking EST representation are among the smallest in the gene model. The largest exon, overlapping Scaffold A1.46 positions 1800000-1810000, is not completely used in any of the ESTs; different parts of this exon are present

in each of the ESTs, with part of the middle region of the exon missing in two of them. Some of the ESTs additionally have small exons aligning within intron regions of the gene model.

The frequency of differences (SNPs or indels) between the Z and W′ sequences of both Israel and White-eye populations, ranging from 0-20 per kb, is shown for the Scaffold A1.46 region of gene 147 in Figure 3.3. These differences were most frequent within the large intron regions near the 5′ end of the gene; they were least frequent in regions overlapping exons near both ends of the gene–the second and third exons from the 5′ end and the third and fourth exons from the 3′ end. The frequency of differences between the two Z sequences, ranging from 0-30 per kb, in this region is shown in Figure 3.4. There are no differences between the two Z sequences in regions containing the first and fourth exons of both gene models. Although these gene model exons are not present in any of the ESTs shown, two small ESTs indicate that other transcripts are expressed from this region. At the 3′ end of the gene, beginning with the last 1.5 kb of the largest exon, the frequency of differences between the two Z sequences drops to only 0-4 per kb. Four of the ESTs align primarily within this region that has few differences between the two Z sequences, especially when compared with the number of differences between the Z and W′ sequences.

The translated Z, W′, and reference sequences of the nesprin-1 gene model A were compared. In the first 13000 positions of the 17259-residue translated sequence, there are many differences between the Israel and white-eye Z sequences as well as differences between the Z and W sequences. However, the differences between the Z sequences dramatically decrease from position 13000 to the C terminal end, where differences between the Z and W′ sequences are most frequent. In this part of the sequence, there are 64 differences between Z and W′ for both Israel and white-eye population sequences and only five differences between the two Z sequences.

The top BLAST hits for the translated gene sequence include Muscle-specific protein 300 (Msp-300), an N-terminal isoform of nesprin-1 in *Drosophila melanogaster*. Its predicted domains include those characteristic of nesprin-1: two N-terminal actin-binding calponin homology (CH) domains, a large number of spectrin repeats, and a C-terminal KASH domain. Nesprin-1 gene family signatures were also identified within the C-terminal and N-terminal

regions. The only BLAST hit confirmed to be a nesprin-1, Msp-300, aligned with just the N-terminal part of the sequence with 55 percent sequence identity. Predicted nesprin-1 sequences from other flies, however, aligned with the C-terminal region of the Hessian fly gene sequence as well. These included a predicted nesprin-1 from *Drosophila biarmipes*, an uncharacterized protein from *Ceratitis capita*, and predicted uncharacterized proteins from *Zeugodacus cucurbitae* and *Musca domestica*. These sequences were more similar to D. melanogaster Msp-300, with 79 percent and 82 percent sequence identity between Msp-300 and the *M. domestica* and *C. capita* sequences, respectively. Predicted domains of the four sequences were compared with those of the Hessian fly gene using Interproscan.

In the Hessian fly sequence, the majority of the spectrin repeats (40 of the 52) are clustered into a group within the first (N terminal) 8100 residues. The sequences from the other four flies also have two CH domains at the N terminal followed by 40 to 48 spectrin repeats, all within a region of about 8000 residues. In this region, differences between the Z and W′ sequences that occur within spectrin repeats are most common within positions 6000-8000. Differences between the two Z sequences also appear to fall within specific spectrin repeats (such as SRs 15 and 37-39) while others appear to be more conserved between the male producer sequences (SRs 22-29).

A large middle region of the gene (positions 8123-15703) mostly lacks predicted domains, with the exception of two isolated spectrin repeats at positions 13419-13523 and 14961-15063. Disordered regions were predicted within positions 8341-10215 and 13533-14733. The other fly sequences also have large regions (ranging in size from 5000 residues in *D. biarmipes* to 12500 in the *M. domestica* sequence) separating the two main groups of spectrin repeats. Similar to the Hessian fly sequence, the D. biarmipes sequence has two isolated spectrin repeats within this large domain, closer to the C terminal end. One isolated spectrin repeat is present in a similar location in the *M. domestica* sequence. None are present in this region of the *C. capita* or *Z. cucurbitae* sequences; oddly, the *C. capita* sequence alone has a single Ataxin-2 domain within this region.

There are many differences between the Israel and White-eye Z sequences in positions 8300-12700; from position 12700 to the C terminal, there are only five differences between

them. Differences between Z and W′ sequences are most frequent within positions 12200-13300 and 14000-14900. In several of these positions, one sequence (either Z or W′) is hydrophobic while the other is polar. There are also a few positions at which one sequence has a proline and the other has either a polar or hydrophobic residue. At most of these positions at which the Z and W′ residues have different chemical properties, the reference sequence is identical to the W′ sequence. The reference genome was sequenced from a subset of GP individuals homozygous for a form of A1 that allows females to produce a mixture of male and female offspring. This form of A1 does not have the sex-determining inversion and therefore should be able to undergo recombination with the Z form. While the Z sequences of Israel and white-eye are identical to each other in this region, the reference sequence is more similar to the W′ sequences from positions 14119 to 16210. There are few differences between Z and W′ within and near SRs 41 and 42 compared to the rest of the 12200-15000 region. Each has a single difference between the Z (threonine) and W′ (lysine in SR 41 and arginine in SR 42) sequences.

In the Hessian fly sequence and those of the other four flies, the last 1600 residues have a small group of spectrin repeats near the C-terminal KASH domain. The Hessian fly sequence is different from the others in its arrangement of spectrin repeats. In the C terminal group, only six are present in the D. biarmipes, C. capita, and Z. cucurbitae sequences; these share similar locations relative to the KASH domain. The Hessian fly sequence appears to have the same six repeats (SRs 44, 46, 47, 48, 50, and 51) in addition to one also present in the M. domestica sequence (SR 49) and two unique to its own sequence (SRs 45 and 50). The M. domestica sequence is also unusual, lacking a spectrin repeat that is present in the other four sequences (SR 47) and having one additional SR not present in any of the other four.

In the Hessian fly sequence, SRs 45 and 47-50 have differences between the Z and W′ sequences. In SRs 48-50, the substitutions are with amino acids that have similar chemical properties. In SR 45, two of the positions have a polar residue (glutamine) in the W′ sequence while charged residues, glutamate and lysine, are present in the Z sequence at these positions. In SR 47, there are three differences between the Z and W′ sequences. At positions 16341,

16378, and 16393, the Z sequence residues are histidine, alanine, and glutamine while the W′ residues are glutamine, threonine, and histidine.

The KASH domain for the Hessian fly gene sequence is also unusual; among other differences from the other four fly sequences, it is missing a highly conserved glycine residue near the C terminal end of the domain. This glycine is shared by the most diverse sequences used to build the KASH domain model, both in other insects and in vertebrates. The Z, W′, and reference sequences are identical within this domain. Between the KASH domain and SR 52, however, there are three differences between the Z and W′ sequences. The W′ sequence has hydrophobic residues, isoleucine and methionine, while the Z sequence has charged and polar residues, arginine and threonine, at positions 17168 and 17170. At position 17170, the Z sequence is proline and the W′ sequence is serine.

split ends

Gene 194 (4629 residues) of scaffold A1.46 was annotated as split ends. Split ends family members are RNA-binding proteins that take part in multiple developmental processes in Drosophila and in mammals, including the Notch and Wnt signaling pathways, through transcriptional silencing [75], [76]. These proteins are able to recruit chromatin-modifying protein complexes to specific chromosomes [77] and are involved in the silencing of maternal X chromosomes in mammals [78].

Three N-terminal SHARP RNA-recognition motifs domains and a C-terminal SPOC domain (involved in developmental signaling) were identified. Within these domains, there were no differences among the Z and W′ sequences. Outside of the SHARP and SPOC domains, the sequence is not highly conserved among members of the split ends family [76]. Many disordered regions were predicted between the N-terminal SHARP domains and C-terminal SPOC domain.

At 19 positions, both the Israel W′ and white-eye W′ sequences were different from the Z sequences; nine of these were outside of the predicted disordered regions (Table 3.9). At positions 812 and 898, the Z residue is proline while the W′ residue is serine. At position 700,

the Z residue is threonine while the W′ residue is alanine. Because the Z and W′ residues at these positions have different chemical properties, they may affect the function of this region (positions 700-1010); however, its function is unknown. In the N-terminal half of the sequence, which includes this region, differences between the two Z sequences appear to be conservative (both residues are either hydrophobic or polar).

MSL2

Gene 145 (604 residues) of scaffold A1.46 was annotated as male-specific lethal 2 (MSL2), an E3 ubiquitin ligase. In Drosophila, MSL2 is regulated by Sxl and is expressed only in males, where it controls dosage compensation by targeting the dosage compensation complex to the X chromosome where it modifies histones to increase gene expression [79]. In the Hessian fly, MSL2 is expressed in both males and females [10]; whether it has a role in dosage compensation is unknown. Two domains were identifieda zinc RING finger domain and a CXC domain. The zinc RING finger domain of MSL2 binds MSL1 to start formation of the dosage compensation complex [80] and the CXC domain recognizes and binds the X chromosome during dosage compensation in Drosophila [81]. The Israel and white-eye W′ sequences were different from both Z sequences at six positions, all of which are outside of predicted domains (Table 3.11). At four of these positions, either the Z or W′ residue is polar while the other is hydrophobic. At one position (211), the W′ residue is cysteine and the Z residue is arginine. There were no differences between the Israel and white-eye Z sequences.

Tudor domain-containing protein

Gene 124 (642 residues) of scaffold A1.46 was annotated as a Tudor domain-containing protein. Tudor domain-containing proteins bind methylated lysine or arginine residues and are involved in chromatin remodeling and RNA processing during development [82]. In Drosophila, Tudor domain-containing proteins are classified into four groups (Ying and Chen, 2012). Group 1 is involved in chromatin regulation through binding of methylated histone

tails. Groups 2-4 are involved in synthesis of piRNAs, microRNAs, and snRNPs, respectively. The majority of Tudor domain-containing proteins belong to group 4. Two Tudor domains and four SNase-like OB fold domains were identified in gene 124. At 12 positions, both the Israel and white-eye W′ sequences were different from the Z sequences (Table 3.10). Six of these were within Tudor domains. At positions 508 and 546, the Z sequence is serine while the W′ sequence is a hydrophobic residue. The rest of the differences between Z and W′ sequences appear to be more conservative. The Israel and white-eye Z sequences were identical within the domains; there are five positions outside of these domains where the two Z sequences are different.

3.3 Discussion

3.3.1 GO term enrichment

The largest group of functionally related GO terms in Scaffold A1.46 consists of terms related to early development, including developmental signaling cascades, cell fate and differentiation, and organ development. This group is not considered to be enriched, probably because many of the terms in this group were represented by the same small number of genes. The most significantly enriched terms were adherens junction", photoreceptor cell fate commitment, establishment of body hair or bristle planar orientation, and morphogenesis of polarized epithelium. The adherens junction term forms two enriched groups which include additional terms related to intercellular junctions. One group also has the the terms imaginal disc growth and renal system development while the other has terms related to Notch signaling regulation. Intercellular junctions both maintain connections between cells and participate in signaling during development. Notch signaling is required for embryonic development and cell fate specification. In Drosophila, Notch signaling is also regulated by Sex lethal to effect sex-specific somatic development [83].

On Scaffold A1.36, two GO term groups and three individual GO terms are enriched according to the Cluego analysis. The group including the term Sex determination, establishment of X:A ratio is only represented by two genes, annotated as segmentation runt and

segmentation runt-like. On Scaffold A1.36, six genes (numbers 25,28,29,30,34, and 38) have been annotated as runt-related transcription factors. For all insects in which runt has been investigated, segmentation runt and runt-related transcription factors are conserved and found grouped closely together [84]. In Drosophila, runt acts as a numerator element; its presence on the X chromosome allows the cell to translate the number of X chromosomes into a signal to activate sex-specific splicing beginning with Sex lethal [85]. After dosage compensation is implemented in the Drosophila embryo, runt fulfills more general developmental roles including segmentation [85]. In other insects, runt is not restricted to the sex chromosomes and has no identified role in sex determination [84]. The other enriched group for Scaffold A1.36 is represented by three genes with methyltransferase activity, annotated as histone-lysine N-methyltransferase pr-set7, N-lysine methyltransferase KMT5A-B, and tRNA (cytosine-5-)-methyltransferase. Both KMT5A-B and pr-set7 monomethylate Histone 4 at Lysine 20. Methylation of H4K20 is associated with chromatin compaction, silent chromatin, and suppression of histone acetylation [86]. The tRNA(cytosine-5-)-methyltransferase gene is able to add methyl groups to both tRNA and DNA cytosines; methyl groups on DNA cytosines act as epigenetic markers [87].

Within a very small group of genes, any highly specific GO term will likely pass the enrichment test, especially if the group includes a homolog of the gene. For example, Sex determination, establishment of X:A ratio was assigned to only three genes in the entire Hessian fly genome and two of these are segmentation runt genes on Scaffold A1.36. Similarly, histone H4-K20 monomethylation is represented by only two genes, both belonging to the small group of Scaffold A1.36 genes that were assigned GO terms. This makes it difficult to determine from the gene enrichment analysis alone that this region is specialized for sex determination or epigenetic modification. The other enriched GO terms for this scaffold are tRNA modification, lipid localization, and positive regulation of antimicrobial peptide biosynthetic process, which are more commonly associated with genes outside of the scaffold and have no obvious connection to sex determination.

For both Scaffolds A1.46 and A1.36, the majority of GO terms that are represented by multiple genes are related to early embryonic development. This is unsurprising for scaffolds

involved in Hessian fly sex determination. The genes controlling early embryonic development are maternally supplied, as are the genes that determine the fate of the Hessian flys paternal X chromosomes and subsequent sex-specific development. Genes with developmental roles unrelated to sex are often given additional roles in the regulation of genes in the sex determination cascade.

3.3.2 Candidate genes for the maternal factor that determines the fate of paternal X chromosomes

Of the genes compared, nesprin-1 and BRCA2 appear to be the best candidates for the maternal factor that determines whether paternal X chromosomes are eliminated or retained. Nesprin-1 has a large C-terminal region with many differences between the Z and W′ sequences and few differences between the two Z sequences. BRCA2 also has a high frequency of differences between the Z and W′ sequences compared to the other genes and no differences between the two Z sequences. Although a mechanism for paternal chromosome retention isn't obvious based on their known functions, both genes are able to interact with chromosomes during mitosis and can take on context-specific roles and interactions with other proteins which are able to modify chromosomes. The other two genes involved in sister chromatid cohesion, SMC3 and PDS5, have nearly identical Z and W′ sequences. The other genes that may target specific chromosomes for modification–split ends, tudor domain-containing protein, and MSL2–have several differences between the Z and W′ sequences. These genes are probably not the master switch but may be more important (or have regions that are more important) in males or male-producing females. Alternatively, the differences between the Z and W′ sequences in these genes may have simply accumulated because they are not deleterious. In the Tudor domain-containing protein, most of the differences are conservative. In the case of split ends, these occur outside of the conserved domains in a region that may be more tolerant of amino acid substitutions; however, a large part of this region lacks differences between the Z sequences. The differences between the Z and W′ sequences of MSL2 are less conservative and may compromise the function of the protein.

3.3.3 nesprin-1

Nesprins (nuclear envelope spectrin repeat proteins) are components of the linker of nucleoskeleton and cytoskeleton (LINC) complex, which forms a bridge across the nuclear envelope through connections between KASH and SUN domain proteins [88]. Among the roles of the LINC complex are control of nuclear shape and positioning and influence of chromatin organization and transcription [89].

The sequence of nesprin-1 contains multiple start and stop sites from which different transcripts are produced for specialized functions, tissues, cell locations, and developmental stages in vertebrates and in flies [90]. The largest form of the nesprin-1 protein has both the N-terminal actin-binding calponin homology domains, over 50 spectrin repeats, and the C-terminal KASH domain which allow it to make connections between the cytoskeleton to the nucleus. The spectrin repeats of nesprins form modular domains that are able to dimerize, interact with other nesprins, and act as scaffolds for various protein complexes [91], depending on their sequences and combinations. It has been suggested that smaller nesprin isoforms in the nucleus act as a scaffolds for regulatory complexes that associate with heterochromatin [92].

Investigation of nesprin-1 structure and function has mainly been carried out in vertebrates due to its association with human diseases. Emery-Dreifuss muscular dystrophy is caused by mutations in the C-terminal region of nesprin-1 as well as mutations in two of its binding partners, the vertebrate nuclear lamina proteins emerin and Lamin A/C [93]. The drosophila ortholog of nesprin-1 was first identified by its N-terminal isoform, Msp-300, which is involved in nuclear migration during muscle morphogenesis [94]. The nesprin-1 sequences are most highly conserved between vertebrates and drosophila in the C and N-terminal regions [95], which have the actin-binding and nuclear membrane domains.

Most small isoforms of nesprin-1 come from the C-terminal and typically include the KASH domain along with a small number of spectrin repeats. In vertebrates, these isoforms are able to localize to the inner nuclear membrane and interact with emerin and Lamin A/C through their last two spectrin repeats, which are highly conserved [96]. Though short

isoforms appear to be expressed from the C-terminal of nesprin-1 in Drosophila as well, these have not yet been characterized. Near the C terminal of nesprin-1 in flies, the two spectrin repeats closest to the KASH domain appear to be among the most conserved and may have a similar role to those in vertebrates, allowing interactions with proteins in the nuclear lamina.

The KASH domain in the Hessian fly sequence may not be completely functional in either male producers or female producers due to its lack of the highly conserved C-terminal glycine residue. If this is the case, C-terminal isoforms from this gene may still interact with nuclear lamina proteins through its spectrin repeats. Vertebrate nesprin-1 isoforms lacking KASH domains are able to localize to various subcellular compartments and structures including the nucleolus, heterochromatin, nuclear matrix, and centrosome [90]. In the Hessian fly sequence, the three differences in the Z and W′ sequences between the KASH domain and the last spectrin repeat may impact its ability to localize to specific regions or interact with other proteins if this part of the sequence is included after splicing.

The region between spectrin repeats 41 and 42 may contain binding sites for interactions with specific proteins. The spectrin repeats themselves may be equally functional in both the Z and W′ forms, each having only a single substitution of a polar residue for a charged one. Between these two SRs, many positions among and within predicted disordered regions have different structural and chemical properties between the W′ and Z sequences. At most of these positions, the sequence has either a proline or another hydrophobic residue in one of the sequences and a polar residue in the other.

Differentially spliced proteins with disordered regions are able to form diverse interactions within regulatory and signaling complexes [97]. Disordered sequences generally have more polar and charged residues and fewer hydrophobic resides (with the exception of proline) when compared with ordered regions. The unique structure of proline is useful in maintaining protein conformations that allow binding. The relatively few hydrophobic residues that are present in disordered regions, in addition to prolines, typically make contact with more ordered proteins during protein-protein interactions.

If this region does have a binding site, its function may be compromised by the substitution of these residues in the W′ sequence. The reference sequence, which comes from

bisexual-producers, shares the W′ sequence in this region. Unlike the W′ chromosome, the form of A1 from which the reference genome was sequenced does not have the sex-determining inversion and should be free to undergo recombination with the Z chromosome. However, while both the reference and white-eye Z sequences come from United States populations, the white-eye Z sequence is dramatically different from the reference sequence but nearly identical to the Israel Z sequence in this part of the gene. The similarity between the female producer and bisexual-producer sequences in this region which appears to be very highly conserved in male producers suggests that it is important for consistent elimination of the paternal X chromosomes. On the N-terminal side of SR 41, there are also a lot of differences between the Z and W′ sequences in a region that is highly conserved between the two Z sequences. In this region, however, the reference sequence is closer to the Z sequence. Disruption of this sequence in addition to the regin between SRs 41 and 42 may be necessary for paternal X chromosomes to be consistently retained, resulting in production of all female offspring.

3.3.4 BRCA2

BRCA2 is a tumor suppressor gene that is able to form complexes with many different binding partners [98], taking on multiple roles in the cell cycle including DNA repair, mitotic checkpoint regulation, and sister chromatid cohesion. BRCA2 loss of function mutations result in chromosome instability and cause defects in gonadal development in vertebrates and in Drosophila [99]. In vertebrates and in Drosophila, BRCA2 forms a complex with Rad51 to repair double stranded breaks through homologous recombination during both mitosis and meiosis [100]. BRCA2 is also able to form a complex with PDS5 that targets DNA to the nuclear lamina for recombination [101]. At least in vertebrates, BRCA2 regulates chromosome-spindle attachments by acetylating the mitotic checkpoint protein BubR1 [102], [103]. In Drosophila, BRCA2 also regulates sister chromatid cohesion during DNA replication by inhibiting binding of the cohesin subunit SA at origins of replication; this counteracts the role of PDS5, which increases binding of SA to these sites [73]. If BRCA2 is as important

for fertility in the Hessian fly as it is in Drosophila and other organisms, the male producer sequence may be highly conserved so that individual flies always carry at least one functional copy. The frequency of differences between the male producer and female producer sequences suggest that relative to other genes in the region, it probably began to accumulate mutations either before or soon after the inversion occurred.

3.4 Conclusions and Future Work

Scaffolds A1.36 and A1.46 were enriched for genes involved in early development, including histone H4K20 methylation. Because the master switch gene for sex determination in the Hessian fly has a maternal effect, it makes sense for genes in these scaffolds to be involved in these processes, which require maternally supplied proteins. Additionally, sex determination pathways often make use of proteins which previously had more general roles in early development. Because sex in the Hessian fly is determined by the elimination or retention of paternal X chromosomes, candidate genes were selected from those with potential roles in chromosome behavior. Of these genes, twoannotated as nesprin-1 and brca2have a large number of differences between the female producer and male producer sequences in both Old World and New World populations. Although the master switch could be a transcription or splicing factor that produces sex-specific transcripts rather than a gene that interacts directly with the chromosomes to be eliminated, the large number of differences between the female producer and male producer sequences combined with the conservation of male producer-specific sequence between Old World and New World populations suggest that these genes have sex-specific functions. Further work will be needed to understand the mechanism of paternal X chromosome retention in future females, including functional analysis of candidate genes and validation of the gene models. The nesprin-1 gene in particular has multiple isoforms, none of which match the full gene model perfectly. Expression analysis is needed to determine whether the part of the gene with the major differences between male producer and female producer sequence is expressed in the mother or is present in the eggs. To determine whether any candidate gene is sufficient for female production, the W

sequence of the gene can be swapped with the Z sequence using crispr. Because mothers that produce a mixture of male and female offspring do not have the inversion, the gene responsible for production of bisexual offspring can be mapped. Identification of this gene may provide clues about how unisexual production evolved and whether multiple genes are required for production of all-female offspring. Investigation of histone modifications on the eliminated chromosomes versus those to be retained may reveal part of the sex determination mechanism. Identifying changes in these patterns resulting from knockout of a candidate gene may help to clarify the function of the gene in a sex determination role.

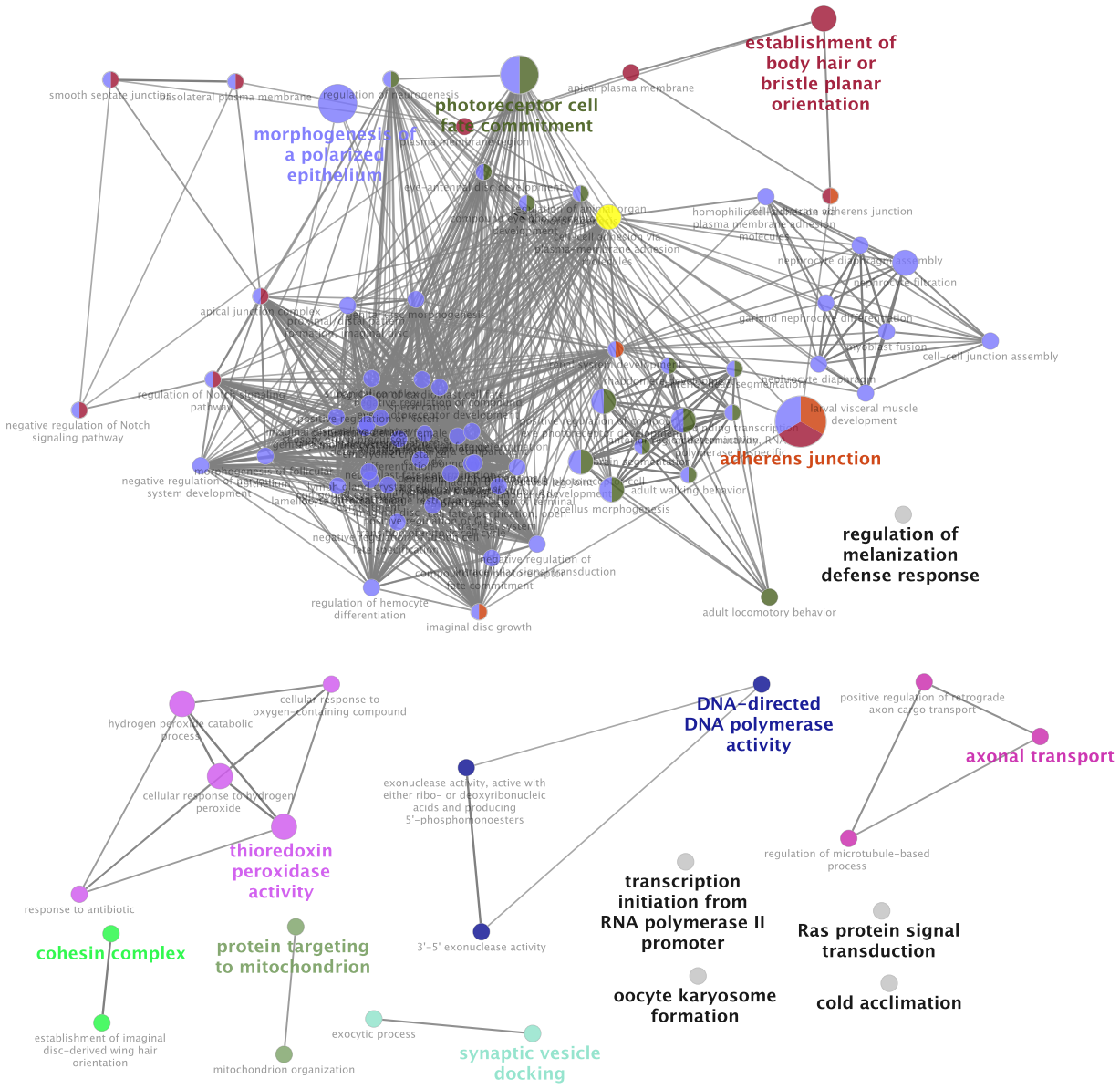

Fig. 3.1. Scaffold A1.46 GO term groups. GO terms are represented by circles, with larger circles representing GO terms with smaller p values. GO terms that have been grouped together and connected by lines and color-coded.

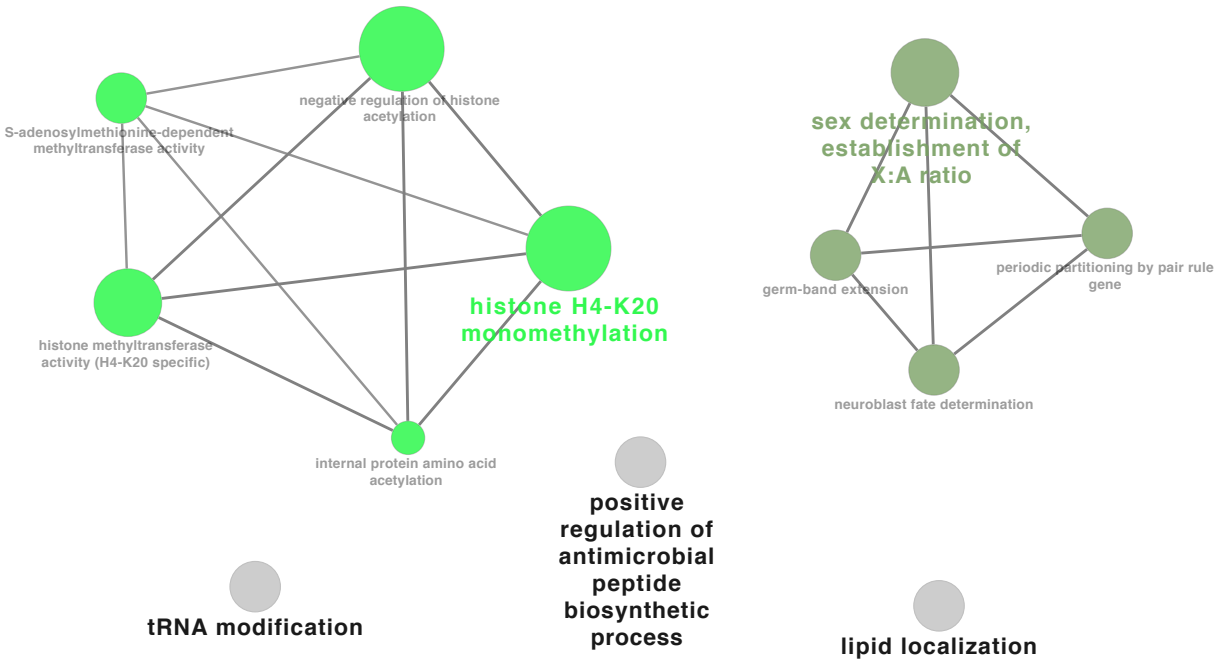

Fig. 3.2. Scaffold A1.36 GO term groups. GO terms are represented by circles, with larger circles representing GO terms with smaller p values. GO terms that have been grouped together and connected by lines and color-coded.

Table 3.1.: Scaffold A1.46 GO terms and groups, term and group significance, percentage of total genes associated with each term found in A1.46, and number of genes within A1.46 associated with each term.

Group	**Term**	**Term P val**	**Group P val**	**percent genes**	**no. of genes**
0	mitochondrion organization	0.80	0.11	4.49	4
	protein targeting to mitochondrion	0.40	0.11	18.18	2
1	cohesin complex	0.29	0.09	22.22	2
	establishment of imaginal disc-derived wing hair orientation	0.60	0.09	6.45	2
2	exocytic process	0.78	0.11	5.26	3
	synaptic vesicle docking	0.45	0.11	6.25	2
3	DNA-directed DNA polymerase activity	0.61	0.10	13.33	2
	3'-5' exonuclease activity	0.70	0.10	8.00	2
	exonuclease activity, active with either RNA or DNA and producing 5'-phosphomonoesters	0.64	0.10	6.90	2
4	regulation of microtubule -based process	0.67	0.11	4.62	3
	positive regulation of retrograde axon cargo transport	0.40	0.11	18.18	2
	axonal transport	0.21	0.11	12.00	3
5	adherens junction	0.00	0.05	7.55	8
	cell-substrate adherens junction	0.59	0.05	7.32	3

Continued on next page

Table 3.1 – *Continued from previous page*

Group	Term	Term P val	Group P val	percent genes	no. of genes
	cell-cell adhesion via plasma -membrane adhesion molecules	0.10	0.05	7.35	5
	imaginal disc growth	0.83	0.05	5.66	3
	renal system development	0.80	0.05	4.35	4
6	response to antibiotic	0.45	0.11	6.15	4
	cellular response to oxygen-containing compound	0.63	0.11	4.41	3
	hydrogen peroxide catabolic process	0.06	0.11	50.00	2
	cellular response to hydrogen peroxide	0.06	0.11	50.00	2
	thioredoxin peroxidase activity	0.06	0.11	50.00	2
7	adherens junction	0.00	0.01	7.55	8
	apical junction complex	0.74	0.01	4.88	4
	cell-substrate adherens junction	0.59	0.01	7.32	3
	plasma membrane region	0.51	0.01	4.08	6
	smooth septate junction	0.23	0.01	25.00	2
	basolateral plasma membrane	0.24	0.01	4.35	2
	apical plasma membrane	0.35	0.01	4.05	3
	regulation of Notch signaling pathway	0.82	0.01	4.21	4
	negative regulation of Notch signaling pathway	0.60	0.01	6.45	2
	establishment of body hair or bristle planar orientation	0.09	0.01	40.00	2

Continued on next page

Table 3.1 – *Continued from previous page*

Group	**Term**	**Term P val**	**Group P val**	**percent genes**	**no. of genes**
8	DNA-binding transcription activator activity, RNA polymerase II-specific	0.31	0.10	4.76	2
	adult locomotory behavior	0.73	0.10	4.82	4
	adult walking behavior	0.56	0.10	14.29	2
	regulation of animal organ morphogenesis	0.72	0.10	4.00	5
	eye-antennal disc development	0.38	0.10	5.21	5
	brain segmentation	0.23	0.10	25.00	2
	anterior head segmentation	0.51	0.10	15.38	2
	regulation of neurogenesis	0.54	0.10	4.00	6
	ocellus morphogenesis	0.06	0.10	50.00	2
	photoreceptor cell fate commitment	0.03	0.10	7.41	6
	anterior region determination	0.09	0.10	40.00	2
	ocellus photoreceptor cell development	0.09	0.10	40.00	2
	rhabdomere development	0.12	0.10	4.26	2
	compound eye photoreceptor development	0.55	0.10	4.55	5
	positive regulation of compound eye photoreceptor development	0.09	0.10	40.00	2
9	DNA-binding transcription activator activity, RNA polymerase II-specific	0.31	0.10	4.76	2

Continued on next page

Table 3.1 – *Continued from previous page*

Group	**Term**	**Term P val**	**Group P val**	**percent genes**	**no. of genes**
	negative regulation of JNK cascade	0.66	0.10	7.41	2
	R8 cell development	0.35	0.10	20.00	2
	adherens junction	0.00	0.10	7.55	8
	apical junction complex	0.74	0.10	4.88	4
	nephrocyte diaphragm	0.29	0.10	22.22	2
	adult walking behavior	0.56	0.10	14.29	2
	regulation of hemocyte differentiation	0.81	0.10	5.56	3
	cell-cell adhesion via plasma -membrane adhesion molecules	0.10	0.10	7.35	5
	smooth septate junction	0.23	0.10	25.00	2
	cell-cell junction assembly	0.63	0.10	4.41	3
	homophilic cell adhesion via plasma membrane adhesion molecules	0.38	0.10	9.38	3
	imaginal disc growth	0.83	0.10	5.66	3
	muscle cell fate determination	0.65	0.10	12.50	2
	basolateral plasma membrane	0.24	0.10	4.35	2
	development of primary male sexual characteristics	0.66	0.10	7.41	2
	negative regulation of nervous system development	0.78	0.10	4.17	4
	renal system development	0.80	0.10	4.35	4
	nephrocyte filtration	0.09	0.10	40.00	2

Continued on next page

Table 3.1 – *Continued from previous page*

Group	Term	Term P val	Group P val	percent genes	no. of genes
	morphogenesis of a polarized epithelium	0.01	0.10	6.08	9
	myoblast fusion	0.66	0.10	7.69	2
	regulation of Notch signaling pathway	0.82	0.10	4.21	4
	morphogenesis of follicular epithelium	0.55	0.10	4.29	3
	subapical complex	0.31	0.10	5.56	2
	negative regulation of intracellular signal transduction	0.67	0.10	5.00	3
	regulation of animal organ morphogenesis	0.72	0.10	4.00	5
	larval visceral muscle development	0.56	0.10	14.29	2
	negative regulation of terminal cell fate specification, open tracheal system	0.45	0.10	16.67	2
	negative regulation of fusion cell fate specification	0.40	0.10	18.18	2
	eye-antennal disc development	0.38	0.10	5.21	5
	negative regulation of Notch signaling pathway	0.60	0.10	6.45	2
	positive regulation of Notch	0.12	0.10	4.26	2

Continued on next page

Table 3.1 – *Continued from previous page*

Group	Term	Term P val	Group P val	percent genes	no. of genes
	signaling pathway				
	formation of a compartment boundary	0.23	0.10	25.00	2
	proximal/distal pattern formation, imaginal disc	0.38	0.10	9.38	3
	dorsal/ventral lineage restriction, imaginal disc	0.83	0.10	9.52	2
	germarium-derived female germ-line cyst encapsulation	0.71	0.10	11.76	2
	brain segmentation	0.23	0.10	25.00	2
	anterior head segmentation	0.51	0.10	15.38	2
	nephrocyte diaphragm assembly	0.29	0.10	22.22	2
	regulation of neurogenesis	0.54	0.10	4.00	6
	glial cell fate determination	0.23	0.10	25.00	2
	genital disc morphogenesis	0.13	0.10	14.29	3
	ocellus morphogenesis	0.06	0.10	50.00	2
	embryonic crystal cell differentiation	0.29	0.10	22.22	2
	photoreceptor cell fate commitment	0.03	0.10	7.41	6
	neuroblast fate determination	0.24	0.10	4.35	2
	sensory organ precursor cell fate determination	0.70	0.10	8.00	2
	garland nephrocyte differentiation	0.13	0.10	33.33	2

Continued on next page

Table 3.1 – *Continued from previous page*

Group	**Term**	**Term P val**	**Group P val**	**percent genes**	**no. of genes**
	epithelial cell proliferation involved in Malpighian tubule morphogenesis	0.75	0.10	11.11	2
	positive regulation of G1/S transition of mitotic cell cycle	0.56	0.10	14.29	2
	compound eye cone cell fate commitment	0.81	0.10	8.70	2
	lymph gland crystal cell differentiation	0.51	0.10	15.38	2
	lamellocyte differentiation	0.81	0.10	8.70	2
	anterior region determination	0.09	0.10	40.00	2
	imaginal disc-derived leg joint morphogenesis	0.71	0.10	11.76	2
	imaginal disc-derived male genitalia morphogenesis	0.76	0.10	10.53	2
	regulation of cardioblast cell fate specification	0.72	0.10	8.33	2
	ocellus photoreceptor cell development	0.09	0.10	40.00	2
	rhabdomere development	0.12	0.10	4.26	2
	compound eye photoreceptor fate commitment	0.51	0.10	4.23	3
	compound eye photoreceptor development	0.55	0.10	4.55	5
	positive regulation of compound	0.09	0.10	40.00	2

Continued on next page

Table 3.1 – *Continued from previous page*

Group	**Term**	**Term P val**	**Group P val**	**percent genes**	**no. of genes**
	eye photoreceptor development				
	negative regulation of compound	0.56	0.10	14.29	2
	eye photoreceptor development				
None	transcription initiation from	0.73	Na	5.08	3
	RNA polymerase II promoter				
None	Ras protein signal transduction	0.75	Na	4.08	4
None	cold acclimation	0.23	Na	25.00	2
None	oocyte karyosome formation	0.40	Na	6.06	2
None	regulation of melanization	0.81	Na	10.00	2
	defense response				

Table 3.2.: Scaffold A1.36 GO terms and groups, term and group significance, percentage of total genes associated with each term found in A1.36, and number of genes within A1.36 associated with each term.

Group	**Term**	**Term P val**	**Group P val**	**percent genes**	**no. of genes**
0	sex determination, establishment of X:A ratio	0.00	0.07	66.67	2
	germ-band extension	0.02	0.07	12.50	2
	periodic partitioning by pair rule gene	0.01	0.07	28.57	2
	neuroblast fate determination	0.04	0.07	4.35	2
1	S-adenosylmethionine-dependent methyltransferase activity	0.02	0.04	6.12	3
	internal protein amino acid acetylation	0.06	0.04	4.65	2
	histone H4-K20 monomethylation	0.00	0.04	100.00	2
	negative regulation of histone acetylation	0.00	0.04	100.00	2
	histone methyltransferase activity (H4-K20 specific)	0.00	0.04	40.00	2
None	positive regulation of antimicrobial peptide biosynthetic process	0.04	Na	4.35	2
None	tRNA modification	0.02	Na	14.29	2
None	lipid localization	0.03	Na	4.69	3

Table 3.3.: Genes of inversion scaffold A1.46

gene	model	position and strand	description
1	A	A1.46:11776-13417 (+)	neurogenic locus Notch
2	A	A1.46:35984-41720 (+)	neurogenic locus Notch
3	A	A1.46:43530-52515 (-)	NA
4	A	A1.46:53725-54523 (-)	C19orf12 homolog
5	A	A1.46:55567-60689 (-)	chondroitin sulfate glucuronyltransferase
6	A	A1.46:61086-64710 (+)	actin-related 2 isoform X1
7	A	A1.46:64703-66022 (-)	NIF3 1
8	A	A1.46:66103-67846 (+)	seipin
9	A	A1.46:68189-72279 (-)	homeobox OTX2-like
10	A	A1.46:88381-89994 (-)	homeotic ocelliless isoform X1
11	A	A1.46:106291-106536 (-)	AGAP000215-PA,partial
12	A	A1.46:118380-120433 (+)	NA
13	A	A1.46:151351-151909 (-)	NA
14	A	A1.46:157337-159674 (+)	ras GTPase-activating-binding 2
15	A	A1.46:160510-160791 (+)	NA
16	A	A1.46:181895-190075 (-)	irregular chiasm C-roughest
16	B	A1.46:181895-190075 (-)	(see previous description)
17	A	A1.46:184960-185907 (+)	NA
18	A	A1.46:217045-223266 (-)	uncharacterized protein
19	A	A1.46:232212-235711 (-)	irregular chiasm C-roughest
20	A	A1.46:296009-296214 (+)	NA
21	A	A1.46:315087-319824 (+)	NA
21	B	A1.46:315087-319824 (+)	NA
22	A	A1.46:361642-363101 (+)	SSGP Family F

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
23	A	A1.46:361896-362880 (-)	NA
24	A	A1.46:363651-369999 (+)	SSGP Family B
24	B	A1.46:363651-369999 (+)	SSGP Family E
25	A	A1.46:370343-373200 (-)	NA
25	B	A1.46:370343-373200 (-)	NA
26	A	A1.46:374836-375067 (+)	NA
27	A	A1.46:377154-380007 (+)	NA
28	A	A1.46:396137-401988 (+)	NA
29	A	A1.46:457616-459706 (-)	uncharacterized protein
30	A	A1.46:470178-477409 (+)	irregular chiasm C-roughest
31	A	A1.46:480457-484313 (+)	tektin-3-like isoform X1
32	A	A1.46:480741-481800 (-)	serine protease snake
33	A	A1.46:484186-488584 (-)	regulator of nonsense transcripts 2
34	A	A1.46:488761-493186 (+)	neutral alpha-glucosidase AB
35	A	A1.46:493074-493865 (-)	Mediator of RNA polymerase II transcription subunit 10
36	A	A1.46:494049-495908 (+)	nudC domain-containing 1
37	A	A1.46:495943-497344 (-)	epimerase family SDR3901
38	A	A1.46:497478-502140 (+)	peptidyl-prolyl cis-trans isomerase
39	A	A1.46:502085-503338 (-)	peridoxin posttranslational modification
40	A	A1.46:504065-506064 (+)	mitochondrial cardiolipin hydrolase
41	A	A1.46:537063-537253 (-)	NA

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
42	A	A1.46:550888-551537 (-)	NA
43	A	A1.46:577829-578220 (+)	NA
44	A	A1.46:581043-581442 (-)	NA
45	A	A1.46:583148-583858 (-)	NA
46	A	A1.46:584245-584461 (-)	NA
47	A	A1.46:590165-590773 (-)	NA
48	A	A1.46:593056-597145 (-)	sodium-dependent neutral amino acid transporter B(0)AT3
49	A	A1.46:597866-598967 (-)	NA
50	A	A1.46:599024-599713 (+)	40S ribosomal S28
51	A	A1.46:644788-646426 (+)	Secreted E
52	A	A1.46:646794-648675 (+)	Secreted F
53	A	A1.46:658342-685579 (+)	ADAMTS 1 isoform X2
53	B	A1.46:658342-685579 (+)	ADAMTS 3 isoform X1
54	A	A1.46:661336-662857 (-)	Secreted F
55	A	A1.46:680822-682344 (-)	NA
56	A	A1.46:698764-700046 (+)	ADAMTS 1
57	A	A1.46:713869-714193 (+)	NA
58	A	A1.46:722145-722561 (+)	NA
59	A	A1.46:722343-722698 (-)	NA
60	A	A1.46:731101-733629 (-)	aldehyde dehydrogenase
61	A	A1.46:734097-739722 (-)	structural maintenance of chromosomes 3
62	A	A1.46:734250-735026 (+)	kDa salivary
63	A	A1.46:740116-748635 (+)	N-acetylgalactosaminyl -transferase 7

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
64	A	A1.46:748560-754040 (-)	Peroxisome proliferator -activated receptor gamma coactivator-related 1
65	A	A1.46:754570-757755 (+)	translation initiation factor eIF-2B subunit epsilon
66	A	A1.46:757722-761089 (-)	carbohydrate sulfotransferase 5
67	A	A1.46:761544-763016 (-)	Mediator of RNA polymerase II transcription subunit 27
68	A	A1.46:763048-765256 (+)	ww domain-binding 11
69	A	A1.46:765152-768327 (-)	sideroflexin-1
70	A	A1.46:768724-769958 (-)	uncharacterized protein
71	A	A1.46:770546-770978 (-)	NA
72	A	A1.46:781758-794528 (-)	uncharacterized protein
73	A	A1.46:786021-790329 (+)	NA
74	A	A1.46:795580-795780 (-)	NA
75	A	A1.46:796327-796840 (-)	ER degradation-enhancing alpha-mannosidase-like 2
76	A	A1.46:797706-799663 (+)	NA
77	A	A1.46:800150-802166 (-)	peroxiredoxin 1
77	B	A1.46:800150-803321 (-)	(see previous description)
78	A	A1.46:803776-805336 (-)	conserved hypothetical protein
79	A	A1.46:805660-808567 (-)	transferrin precursor
80	A	A1.46:809022-812125 (-)	ubiquitin carboxyl-terminal hydrolase 30 homolog
81	A	A1.46:812534-813324 (-)	Casein kinase I isoform alpha
82	A	A1.46:814571-827880 (+)	dynein heavy chain axonemal

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
83	A	A1.46:827972-840154 (-)	cerebellar degeneration-related 2-like isoform X3
84	A	A1.46:834230-835117 (+)	transmembrane protein
85	A	A1.46:864092-868086 (-)	TAR DNA-binding 43 isoform X4
86	A	A1.46:867959-880644 (+)	membrane-associated progesterone receptor component 1
87	A	A1.46:868338-871016 (-)	hypothetical protein
88	A	A1.46:880892-882640 (-)	NA
89	A	A1.46:884105-885608 (-)	NA
90	A	A1.46:889644-896678 (-)	mucin-5ac
91	A	A1.46:900200-900598 (-)	vegetative cell wall
92	A	A1.46:931160-931433 (-)	NA
93	A	A1.46:936639-937012 (+)	NA
94	A	A1.46:998358-1009083 (-)	NA
95	A	A1.46:1008051-1008684 (+)	NA
96	A	A1.46:1054337-1054477 (-)	NA
97	A	A1.46:1088934-1089698 (+)	NA
98	A	A1.46:1132274-1133565 (+)	NA
99	A	A1.46:1134043-1137101 (+)	serine arginine repetitive matrix 1-like isoform X1
100	A	A1.46:1143614-1143933 (+)	NA
101	A	A1.46:1160130-1160582 (-)	NA
102	A	A1.46:1174036-1179288 (+)	chascon mitochondrion
103	A	A1.46:1181427-1193907 (+)	probable ATP-dependent DNA

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
			helicase HFM1
103	B	A1.46:1181450-1193907 (+)	(see previous description)
104	A	A1.46:1184844-1185523 (-)	cuticle 19
105	A	A1.46:1200352-1212841 (-)	mediator of RNA polymerase II transcription subunit 26 isoform X2
106	A	A1.46:1246279-1247039 (-)	myelin transcription factor 1
107	A	A1.46:1248465-1250385 (-)	alpha kinase 1 isoform X2
108	A	A1.46:1291458-1292089 (+)	TPA: Ci-Rhysin2 Deltex3-a
109	A	A1.46:1292185-1294336 (-)	heat shock
110	A	A1.46:1350723-1351492 (+)	NA
111	A	A1.46:1352898-1355525 (+)	NA
111	B	A1.46:1352898-1355525 (+)	NA
111	C	A1.46:1352898-1355576 (+)	NA
112	A	A1.46:1353471-1355052 (-)	NA
113	A	A1.46:1382018-1383733 (+)	SSGP Family B
114	A	A1.46:1384172-1385952 (+)	SSGP Family F
115	A	A1.46:1395506-1395681 (-)	NA
116	A	A1.46:1428815-1429063 (+)	NA
117	A	A1.46:1430556-1430858 (-)	NA
118	A	A1.46:1458209-1465463 (+)	rho GTPase-activating 190 isoform X1
119	A	A1.46:1475075-1477237 (-)	focal adhesion
119	B	A1.46:1475075-1477237 (-)	(see previous description)
120	A	A1.46:1477940-1480341 (-)	focal adhesion
121	A	A1.46:1552231-1556688 (-)	NA

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
122	A	A1.46:1580003-1580228 (-)	NA
123	A	A1.46:1588559-1592687 (+)	Structural maintenance of chromosomes 3
124	A	A1.46:1597626-1599809 (+)	tudor domain-containing
125	A	A1.46:1632695-1634790 (-)	NA
126	A	A1.46:1635489-1636482 (-)	NA
127	A	A1.46:1638767-1639831 (-)	mitochondrial inner membrane protease subunit 1
128	A	A1.46:1640039-1643996 (+)	ubiquitin conjugation factor E4 A
129	A	A1.46:1643942-1645481 (-)	Peroxisomal membrane PEX16
130	A	A1.46:1645448-1647578 (+)	molybdopterin synthase catalytic subunit
130	B	A1.46:1645763-1647578 (+)	(see previous description)
130	C	A1.46:1645763-1647578 (+)	(see previous description)
131	A	A1.46:1647577-1648494 (-)	exosome complex component CSL4
132	A	A1.46:1649452-1649880 (+)	NA
133	A	A1.46:1654091-1685048 (+)	von Willebrand factor type EGF and pentraxin domain-containing
133	B	A1.46:1683605-1683966 (+)	NA
134	A	A1.46:1684279-1685051 (-)	uncharacterized protein
135	A	A1.46:1685602-1686888 (-)	uncharacterized protein
136	A	A1.46:1691256-1694119 (+)	UDP-glucuronosyltransferase 2B13
137	A	A1.46:1698937-1710869 (+)	synaptotagmin 1 isoform X1
138	A	A1.46:1702291-1702984 (-)	NA
139	A	A1.46:1715921-1716666 (-)	NA

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
140	A	A1.46:1718130-1730016 (-)	juvenile hormone binding in insects
141	A	A1.46:1730715-1734820 (+)	AN1-type zinc finger 2A
142	A	A1.46:1736088-1745753 (-)	lethal(2) giant larvae isoform X1
143	A	A1.46:1741003-1741246 (+)	NA
144	A	A1.46:1746296-1748512 (+)	lactosylceramide 4-alpha-galactosyltransferase
144	B	A1.46:1746339-1748512 (+)	(see previous description)
145	A	A1.46:1748574-1751146 (-)	E3 ubiquitin- ligase MSL2 isoform X2
146	A	A1.46:1751600-1754321 (-)	Gawky isoform X1
147	A	A1.46:1754054-1830704 (+)	Nesprin-1
147	B	A1.46:1754607-1830704 (+)	(see previous description)
148	A	A1.46:1829681-1835825 (-)	UDP-glucuronosyl -transferase 2B7-like
149	A	A1.46:1836575-1836975 (+)	proteasome inhibitor PI31 subunit
150	A	A1.46:1838899-1844629 (+)	sister chromatid cohesion PDS5 homolog B isoform X2
151	A	A1.46:1845026-1846721 (+)	serine protease gd-like
152	A	A1.46:1846793-1851582 (-)	PIH1 domain-containing 1
153	A	A1.46:1851777-1862101 (+)	multidrug resistance-associated 1
153	B	A1.46:1851777-1862101 (+)	(see previous description)
154	A	A1.46:1862915-1874181 (-)	very long-chain specific acyl-CoA dehydrogenase, mitochondrial
154	B	A1.46:1862915-1874208 (-)	(see previous description)
155	A	A1.46:1874633-1876546 (+)	carboxymethyl transferase
156	A	A1.46:1876957-1879877 (+)	uncharacterized protein

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
157	A	A1.46:1879943-1884414 (-)	serine protease inhibitor
158	A	A1.46:1884749-1893142 (-)	a kinase anchor
159	A	A1.46:1929147-1929798 (+)	NA
160	A	A1.46:1935076-1937446 (+)	sphingoid base -phosphate phosphatase
161	A	A1.46:1941915-1942885 (-)	metallo-beta-lactamase domain-containing 1
162	A	A1.46:1944892-1948061 (+)	CTP synthase
163	A	A1.46:1948079-1949106 (-)	dynactin subunit 5
164	A	A1.46:1949369-1957611 (+)	ubiquitin conjugation factor E4 B
165	A	A1.46:1960929-1962009 (+)	serine protease inhibitor dipetalogastin
166	A	A1.46:1962948-1963887 (+)	(see previous description)
166	B	A1.46:1963141-1963957 (+)	(see previous description)
167	A	A1.46:1964658-1966282 (+)	(see previous description)
168	A	A1.46:1965452-1968776 (-)	NA
169	A	A1.46:1968009-1968599 (+)	hypothetical protein
170	A	A1.46:1970416-1971267 (+)	IWS1 homolog isoform X1
171	A	A1.46:1971589-1973067 (+)	YIPF1
172	A	A1.46:1973052-1981066 (-)	AMP deaminase 2 isoform X1
172	B	A1.46:1973052-1983251 (-)	(see previous description)
172	C	A1.46:1973052-1983268 (-)	(see previous description)
172	D	A1.46:1973052-1983385 (-)	(see previous description)
173	A	A1.46:1977121-1977985 (+)	NA
174	A	A1.46:1988699-1992783 (-)	transmembrane 94 isoform X3
175	A	A1.46:1992859-1994507 (+)	uncharacterized Golgi apparatus

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
			membrane CG5021 isoform X1
176	A	A1.46:1994954-1995896 (+)	adenylate kinase
			isoenzyme 6 homolog
177	A	A1.46:1997048-1997560 (+)	UPF0047
178	A	A1.46:2000275-2003933 (+)	2,5-phosphodiesterase 12
179	A	A1.46:2003020-2006789 (-)	DNA polymerase subunit
			gamma-mitochondrial
180	A	A1.46:2007146-2011067 (+)	DNA primase large subunit
181	A	A1.46:2016003-2022413 (+)	otopetrin-2 like
182	A	A1.46:2064310-2067936 (-)	otopetrin-2 isoform X2
183	A	A1.46:2073179-2076073 (+)	otopetrin-2 isoform X2
184	A	A1.46:2076074-2077289 (+)	RING-box 1A
185	A	A1.46:2077328-2078472 (-)	NA
186	A	A1.46:2078909-2084947 (+)	histone-lysine N-methyltransferase
			Suv4-20
187	A	A1.46:2084715-2085804 (-)	hypothetical protein
188	A	A1.46:2085822-2089140 (+)	probably RNA-binding 19
189	A	A1.46:2089253-2090014 (-)	dna damage-regulated autophagy
			modulator 1
190	A	A1.46:2091725-2092201 (-)	lethal(2)essential for life-like
191	A	A1.46:2093899-2100418 (-)	Heat shock 27
192	A	A1.46:2101859-2102422 (-)	Heat shock 27
193	A	A1.46:2103047-2103628 (+)	Heat shock 23
194	A	A1.46:2104606-2121724 (-)	split ends isoform X1
194	B	A1.46:2104606-2121724 (-)	(see previous description)
194	C	A1.46:2104606-2121724 (-)	(see previous description)

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
195	A	A1.46:2140725-2140896 (+)	NA
196	A	A1.46:2146351-2146984 (+)	NA
197	A	A1.46:2153762-2156067 (+)	Nucleoporin amo1
197	B	A1.46:2153762-2158669 (+)	serine threonine kinase RIO3
198	A	A1.46:2158557-2159827 (-)	vesicle transport SFT2C
199	A	A1.46:2159908-2161007 (-)	uncharacterized protein
200	A	A1.46:2161012-2166712 (-)	NIK and IKK binding
201	A	A1.46:2166982-2168045 (+)	glycine-rich cell wall structural
202	A	A1.46:2168649-2170714 (-)	disulfide-isomerase a6
202	B	A1.46:2168649-2171049 (-)	(see previous description)
203	A	A1.46:2174400-2175841 (+)	juvenile hormone acid O-methyltransferase
204	A	A1.46:2178498-2180065 (+)	Uncharacterized protein
205	A	A1.46:2220779-2224161 (-)	toll, partial
206	A	A1.46:2239019-2241787 (+)	actin binding
207	A	A1.46:2247906-2248256 (+)	NA
208	A	A1.46:2251020-2251931 (+)	NA
209	A	A1.46:2253279-2253565 (-)	NA
210	A	A1.46:2254178-2255708 (+)	leucine-rich repeat-containing 15-like
211	A	A1.46:2291892-2293400 (+)	casein
212	A	A1.46:2304759-2310382 (-)	dachshund homolog 2 isoform X1
213	A	A1.46:2324542-2324662 (+)	NA
214	A	A1.46:2343656-2344498 (-)	dachshund homolog
215	A	A1.46:2344972-2345536 (+)	NA
216	A	A1.46:2356082-2356810 (+)	NADH dehydrogenase

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
			[ubiquinone] 1 subunit C2
217	A	A1.46:2357295-2360292 (-)	galectin-4-like isoform X1
218	A	A1.46:2360944-2362516 (-)	amnionless
219	A	A1.46:2365217-2367069 (-)	NA
220	A	A1.46:2368185-2368480 (-)	NA
221	A	A1.46:2369060-2392046 (-)	cell adhesion molecule 4
222	A	A1.46:2403106-2405632 (-)	NA
223	A	A1.46:2405929-2408453 (-)	NA
224	A	A1.46:2408939-2414271 (-)	WD repeat-containing
			26 homolog
224	B	A1.46:2408939-2414271 (-)	(see previous description)
225	A	A1.46:2415330-2430464 (-)	venus kinase receptor
226	A	A1.46:2415352-2420340 (+)	metallophosphoesterase
			1 isoform X2
227	A	A1.46:2433707-2435356 (-)	ubiquitin thioesterase
			otubain-like
228	A	A1.46:2436353-2438234 (+)	60S ribosomal L9
229	A	A1.46:2438235-2440395 (-)	Mitochondrial import
			receptor subunit TOM70
229	B	A1.46:2438235-2440707 (-)	(see previous description)
230	A	A1.46:2441255-2443874 (+)	rac GTPase-activating 1
231	A	A1.46:2443825-2445990 (-)	syntaxin-4 isoform X1
232	A	A1.46:2446446-2446499 (-)	NA
233	A	A1.46:2447585-2449322 (-)	inosine-uridine preferring
			nucleoside hydrolase
233	B	A1.46:2449221-2452811 (-)	(see previous description)

Continued on next page

Table 3.3 – *Continued from previous page*

gene	model	position and strand	description
234	A	A1.46:2454076-2458491 (+)	kinesin associated kap
235	A	A1.46:2458774-2459211 (-)	uncharacterized protein
236	A	A1.46:2461965-2470742 (+)	Tyramine beta-hydroxylase
237	A	A1.46:2470917-2471944 (-)	mediator of RNA polymerase II transcription subunit 8
238	A	A1.46:2472450-2480742 (+)	amyloid beta A4 precursor-binding family A member 2-like isoform X8
239	A	A1.46:2482759-2486338 (-)	LIM domain-containing jub isoform X1
240	A	A1.46:2487965-2488826 (+)	ADP-ribosylation factor 2-binding
241	A	A1.46:2488687-2493215 (-)	Acidic repeat containing
242	A	A1.46:2493546-2494748 (+)	Uncharacterized protein
243	A	A1.46:2494787-2499187 (-)	Sentrin-specific protease 1
243	B	A1.46:2494787-2499187 (-)	(see previous description)
243	C	A1.46:2494787-2499187 (-)	(see previous description)
244	A	A1.46:2499182-2508222 (+)	ATP-binding cassette sub-family D member 3
245	A	A1.46:2508520-2527119 (-)	methyltransferase bin3
245	B	A1.46:2508520-2527119 (-)	(see previous description)
245	C	A1.46:2508520-2527404 (-)	(see previous description)
246	A	A1.46:2528605-2533373 (+)	lysophosphatidylcholine acyltransferase isoform X1
247	A	A1.46:2562873-2565314 (-)	NA
248	A	A1.46:2566664-2566914 (+)	NA

Table 3.4.: Genes of inversion scaffold Un.16662

gene	model	position and strand	description
1	A	Un.16662:15441-20306 (+)	rho GTPase-activating gacF
2	A	Un.16662:20506-21896 (+)	NA
3	A	Un.16662:22077-23330 (-)	ubiquitin-conjugating enzyme E2 G2
3	B	Un.16662:22077-23591 (-)	(see previous description)
4	A	Un.16662:27285-27884 (-)	NA
5	A	Un.16662:36948-40751 (+)	glutamine synthetase 2 cytoplasmic
6	A	Un.16662:50261-52899 (+)	DEAD-box helicase Dbp80
7	A	Un.16662:52872-54402 (-)	ribonuclease P subunit p20
7	B	Un.16662:52872-54817 (-)	(see previous description)
8	A	Un.16662:55388-56613 (+)	NA
8	B	Un.16662:55537-66981 (+)	uncharacterized protein
8	C	Un.16662:55537-66981 (+)	(see previous description)
9	A	Un.16662:67727-67864 (-)	NA
10	A	Un.16662:68006-75797 (-)	glycogen phosphorylase
11	A	Un.16662:78903-80106 (-)	tetratricopeptide repeat 1
12	A	Un.16662:80487-82190 (-)	fatty acid-binding muscle isoform X2
13	A	Un.16662:84872-85747 (-)	myelin P2 isoform X2
14	A	Un.16662:86362-90365 (-)	vacuolar sorting-associated 18 homolog
15	A	Un.16662:99657-100001 (+)	sterol carrier-2

Table 3.5.: Genes of scaffold A1.36

gene	model	position and strand	description
1	A	A1.36:1084-1376 (-)	NA
2	A	A1.36:2175-3196 (+)	NA
3	A	A1.36:6590-20244 (-)	NA
3	B	A1.36:6590-20244 (-)	NA
3	C	A1.36:6590-20244 (-)	NA
3	D	A1.36:6590-20244 (-)	NA
3	E	A1.36:6590-20244 (-)	NA
3	F	A1.36:6590-20244 (-)	NA
3	G	A1.36:6590-20244 (-)	NA
4	A	A1.36:23587-25763 (+)	NA
5	A	A1.36:24138-27564 (-)	NA
6	A	A1.36:29461-30646 (+)	NA
7	A	A1.36:37291-41450 (-)	phosphatidylinositol 3,4,5-triphosphate 3-phosphatase and dual-specificity phosphatase PTEN
8	A	A1.36:50273-52587 (-)	Hsp70/Hsp90 organizing (Hop)
9	A	A1.36:52961-54285 (+)	splicing factor U2af 38 kDa subunit
10	A	A1.36:55305-56665 (+)	seIT
11	A	A1.36:57118-57820 (+)	NA
12	A	A1.36:58065-58852 (+)	NA
13	A	A1.36:58873-60104 (+)	uncharacterized protein
14	A	A1.36:61704-62211 (-)	PITH domain-containing GA19395

Continued on next page

Table 3.5 – *Continued from previous page*

gene	model	position and strand	description
15	A	A1.36:64871-70649 (+)	NA
16	A	A1.36:66362-66807 (-)	NA
17	A	A1.36:80944-87228 (-)	histone-lysine N-methyltransferase pr-set7
18	A	A1.36:113462-116994 (+)	espin isoform X1
19	A	A1.36:130670-131194 (+)	Histone H3.3
20	A	A1.36:136988-142367 (+)	N-lysine methyltransferase KMT5A-B
21	A	A1.36:142499-147335 (+)	tRNA (cytosine-5-) -methyltransferase
22	A	A1.36:157377-159210 (-)	leucine-rich repeats and immunoglobulin-like domains 3
23	A	A1.36:163995-166083 (-)	venom allergen
24	A	A1.36:178230-179228 (-)	Sptzl 1B
25	A	A1.36:205572-208353 (-)	segmentation Runt isoform X1
26	A	A1.36:216471-217242 (+)	NA
27	A	A1.36:234908-238230 (+)	antichymotrypsin-2-like isoform X1
28	A	A1.36:239772-241589 (-)	segmentation Runt-like
29	A	A1.36:261608-262162 (+)	runt-related transcription factor 2 isoform X1
30	A	A1.36:273474-278681 (+)	runt-related transcription factor 3 isoform X2
31	A	A1.36:282466-284045 (+)	NA
32	A	A1.36:282545-284387 (-)	SSGP Family F
33	A	A1.36:286517-287119 (-)	NA

Continued on next page

Table 3.5 – *Continued from previous page*

gene	model	position and strand	description
34	A	A1.36:317954-318244 (-)	segmentation Runt
35	A	A1.36:322763-323218 (+)	NA
36	A	A1.36:333685-334085 (-)	NA
37	A	A1.36:339926-348014 (+)	mitogen-activated kinase kinase kinase 7-like
37	B	A1.36:339926-348014 (+)	(see previous description)
38	A	A1.36:352696-359925 (+)	runt-related transcription factor 1-like
39	A	A1.36:374024-378497 (+)	CWC15 homolog
40	A	A1.36:377716-378342 (-)	methylated-DNA–[]-cysteine S-methyltransferase
41	A	A1.36:379407-380556 (+)	immunoglobulin A1 protease autotransporter
42	A	A1.36:382822-384423 (+)	arrestin homolog
43	A	A1.36:384834-388811 (+)	dnaJ homolog subfamily C member 7
44	A	A1.36:389158-389739 (-)	Heat shock 27
45	A	A1.36:391311-391865 (-)	heat shock 23-like
46	A	A1.36:392065-396591 (-)	splicing factor 1
47	A	A1.36:396778-399821 (+)	nucleolar 14 homolog
48	A	A1.36:398540-403170 (-)	glucose-induced degradation 8 homolog
48	B	A1.36:398540-403170 (-)	(see previous description)
49	A	A1.36:403229-403819 (+)	39S ribosomal mitochondrial
50	A	A1.36:403927-405478 (-)	p21-activated kinase -interacting 1-like

Continued on next page

Table 3.5 – *Continued from previous page*

gene	model	position and strand	description
51	A	A1.36:405826-410812 (+)	ATP-dependent RNA helicase p62
51	B	A1.36:405826-410812 (+)	(see previous description)
52	A	A1.36:410126-410792 (-)	histidine triad nucleotide-binding 1
53	A	A1.36:412435-415928 (-)	breast cancer type 2 susceptibility-like
54	A	A1.36:419262-419880 (-)	NA
55	A	A1.36:419705-420389 (+)	NA
56	A	A1.36:420216-421985 (-)	zinc metalloproteinase nas-4-like
57	A	A1.36:422874-423159 (+)	SSGP-11C family
58	A	A1.36:431894-432447 (-)	NA
59	A	A1.36:433385-434502 (-)	NA
60	A	A1.36:444754-445300 (-)	NA
61	A	A1.36:454309-460044 (-)	Niemann-Pick C1
62	A	A1.36:460944-462303 (+)	transmembrane channel 5 isoform X2
63	A	A1.36:470054-470757 (+)	ubiquitin-like 5
64	A	A1.36:470699-471880 (-)	probable tRNA(His) guanylyltransferase
65	A	A1.36:474130-480256 (-)	neuroglian isoform X1
66	A	A1.36:494918-501559 (-)	NA
67	A	A1.36:501257-501519 (+)	NA
68	A	A1.36:503410-505659 (+)	adenosylhomocysteinase
68	B	A1.36:503410-506978 (+)	(see previous description)
69	A	A1.36:506821-514105 (-)	cytoplasmic dynein 1

Continued on next page

Table 3.5 – *Continued from previous page*

gene	model	position and strand	description
			light intermediate chain 1
70	A	A1.36:522423-523269 (-)	NA
71	A	A1.36:525318-526055 (-)	BTB POZ domain-containing
			KCTD12
72	A	A1.36:529066-532484 (-)	apoptotic chromatin
			condensation inducer
			in the nucleus
73	A	A1.36:532907-535311 (+)	actin-related 2 3
			complex subunit 2
73	B	A1.36:532907-535311 (+)	(see previous description)
74	A	A1.36:535015-535368 (-)	NA
75	A	A1.36:537095-539245 (-)	gastrula zinc finger
76	A	A1.36:539907-542072 (+)	very-long-chain-3-oxoacyl
			reductase-like
77	A	A1.36:552021-554193 (+)	zinc finger 239-like
78	A	A1.36:555829-556835 (-)	NA
79	A	A1.36:557025-559919 (+)	gastrula zinc finger-like
79	B	A1.36:557025-559919 (+)	zinc finger 699-like isoform X1
79	C	A1.36:557025-559919 (+)	gastrula zinc finger-like
80	A	A1.36:566400-568552 (+)	nucleoplasmin isoform X2
81	A	A1.36:569986-574998 (+)	4-hydroxybutyrate coenzyme
			A transferase
82	A	A1.36:575204-578953 (-)	alpha-1,6-mannosyl-glyco 2-beta-
			N-acetylglucosaminyltransferase
83	A	A1.36:585655-589180 (-)	exonuclease 1
83	B	A1.36:585655-589180 (-)	(see previous description)

Continued on next page

Table 3.5 – *Continued from previous page*

gene	model	position and strand	description
84	A	A1.36:593980-595971 (+)	forkhead box K2 isoform X1
85	A	A1.36:600695-602723 (-)	TPR-containing DDBG0280363
86	A	A1.36:619984-620240 (+)	NA
87	A	A1.36:637417-637661 (+)	NA
88	A	A1.36:649097-650754 (+)	mediator of RNA polymerase II transcription subunit 4
89	A	A1.36:650775-652379 (-)	CAS1 domain-containing 1
90	A	A1.36:657410-659171 (+)	mpv17 2
91	A	A1.36:659295-661349 (-)	nuclear envelope integral membrane 1
92	A	A1.36:663199-667795 (-)	inositol monophosphatase 3
92	B	A1.36:663199-671553 (-)	inositol monophosphatase 3
92	C	A1.36:663199-671553 (-)	inositol monophosphatase 3
93	A	A1.36:664880-665825 (+)	transmembrane 17B-like
94	A	A1.36:675024-675720 (+)	bombyxin B-1 homolog
95	A	A1.36:676933-679778 (+)	coatomer subunit beta
96	A	A1.36:679927-682861 (-)	cleavage and polyadenylation specificity factor 73
97	A	A1.36:683127-687108 (+)	zinc finger CCCH domain -containing
98	A	A1.36:706851-707191 (-)	NA
99	A	A1.36:713363-727975 (-)	Dopamine receptor 1
100	A	A1.36:754091-756039 (-)	nose resistant to fluoxetine 6-like

Table 3.6.: Translated A1.46 gene 147 (nesprin-1) predicted domains and differences between Z and W′ sequences

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
60	H	H	H	H	Q	disordered region
71	V	L	V	L	L	disordered region
245	R	H	R	H	R	disordered region
380	V	I	V	I	V	NA
396	N	S	N	S	N	NA
529	T	N	T	N	T	NA
800	A	A	T	A	A	CH domain
911	G	G	S	G	G	NA
958	Q	Q	Q	Q	K	spectrin repeat
1016	L	L	L	L	S	spectrin repeat
1189	A	T	A	T	A	NA
1228	Y	C	C	C	Y	NA
1305	S	T	T	T	T	spectrin repeat
1508	I	T	T	T	T	spectrin repeat
1516	V	Y	Y	V	Y	spectrin repeat
1560	E	E	Q	E	E	spectrin repeat
1561	N	N	D	N	N	spectrin repeat
1636	L	F	F	L	F	NA
1662	S	A	A	A	A	NA
1889	D	D	E	D	E	spectrin repeat
1892	H	R	H	R	H	spectrin repeat
1976	K	R	K	K	K	spectrin repeat
2106	N	N	Q	N	Q	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
2224	T	T	S	T	T	spectrin repeat
2395	S	S	N	S	S	NA
2483	S	G	S	G	S	NA
2526	Q	Q	Q	Q	E	NA
2655	V	V	A	V	V	NA
2665	T	I	T	I	T	NA
2785	N	N	D	N	N	NA
2853	E	G	E	G	E	NA
2923	D	N	D	N	D	NA
3498	N	N	N	N	S	NA
3515	T	T	T	T	M	spectrin repeat
3532	E	E	G	E	E	spectrin repeat
3882	V	I	V	I	V	NA
3975	A	S	A	S	S	spectrin repeat
3984	Y	Y	Y	Y	H	spectrin repeat
3994	I	V	I	V	V	spectrin repeat
4017	L	L	L	L	V	spectrin repeat
4064	E	E	Q	E	E	NA
4070	A	T	A	T	A	NA
4128	E	D	E	D	E	NA
4216	A	A	A	A	V	spectrin repeat
4221	A	V	A	V	A	spectrin repeat
4326	G	E	G	E	G	NA
4413	V	V	V	V	A	spectrin repeat
4611	V	V	V	V	A	spectrin repeat
4657	I	V	I	V	V	spectrin repeat

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
4658	V	I	V	V	V	spectrin repeat
4735	N	S	N	S	S	NA
4773	A	S	A	S	A	NA
4791	M	M	M	M	I	NA
4921	S	N	S	N	S	spectrin repeat
5044	M	M	I	M	M	NA
5063	L	L	F	L	L	NA
5130	S	S	A	S	S	NA
5191	P	P	Q	P	P	spectrin repeat
5954	A	G	A	G	A	spectrin repeat
6051	N	S	N	S	N	spectrin repeat
6056	A	V	A	V	A	spectrin repeat
6105	A	V	A	V	A	spectrin repeat
6273	F	L	F	L	F	NA
6325	P	S	S	S	P	spectrin repeat
6338	L	D	H	D	L	spectrin repeat
6461	A	S	A	S	A	spectrin repeat
6556	V	A	A	A	V	NA
6609	H	H	L	H	H	NA
6815	S	T	S	T	S	spectrin repeat
6834	V	V	V	V	I	spectrin repeat
6877	V	V	V	V	A	NA
6898	A	V	A	V	A	NA
6926	A	A	A	A	V	NA
6952	S	T	S	T	S	spectrin repeat
6955	S	S	C	S	C	spectrin repeat

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
6972	R	R	R	R	K	spectrin repeat
7110	S	S	S	S	A	spectrin repeat
7114	L	S	L	S	L	spectrin repeat
7172	E	D	D	D	E	NA
7183	K	R	K	R	R	NA
7215	R	R	R	R	H	NA
7245	G	C	G	C	G	NA
7463	Q	K	Q	K	Q	spectrin repeat
7514	E	A	E	A	E	NA
7599	K	K	K	E	K	spectrin repeat
7650	V	L	V	L	L	spectrin repeat
7652	H	Y	H	Y	H	spectrin repeat
7703	R	R	H	R	R	spectrin repeat
7707	R	R	R	R	C	spectrin repeat
7735	I	I	T	I	I	spectrin repeat
7736	V	D	D	D	V	spectrin repeat
7760	L	L	H	L	L	spectrin repeat
7866	Q	Q	H	Q	Q	NA
7944	V	I	V	I	V	spectrin repeat
7970	R	R	H	R	R	spectrin repeat
7978	V	V	I	V	V	spectrin repeat
8258	L	T	T	T	L	NA
8324	R	H	R	H	R	NA
8341	S	C	S	C	S	disordered region
8382	V	V	I	V	V	disordered region
8419	A	V	A	V	A	disordered region

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
8443	V	V	M	V	V	disordered region
8456	D	N	D	N	D	NA
8518	L	M	L	M	L	NA
8531	K	K	E	K	K	NA
8536	I	I	F	I	I	NA
8541	I	V	V	V	I	NA
8560	Q	K	K	K	Q	NA
8614	D	N	N	N	D	NA
8658	L	L	V	L	L	NA
8698	Q	Q	E	Q	Q	NA
8710	K	E	E	E	K	NA
8719	I	T	T	T	I	NA
8739	E	K	K	K	E	NA
8744	D	D	G	D	D	NA
8779	F	F	L	F	F	NA
8803	E	E	D	E	E	NA
8814	A	A	V	A	A	NA
8824	T	S	T	S	T	NA
8831	S	N	N	N	S	NA
8897	K	Q	Q	Q	K	disordered region
8910	N	S	S	S	N	disordered region
8920	R	R	K	R	R	disordered region
8928	S	P	S	P	S	disordered region
8929	A	T	A	T	A	disordered region
8945	T	N	T	N	T	NA
8964	P	P	L	P	P	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
8965	Q	Q	E	Q	Q	NA
9014	A	T	T	T	A	disordered region
9035	G	E	E	E	G	disordered region
9043	Q	K	K	K	Q	NA
9071	E	E	G	E	E	disordered region
9077	E	K	E	K	E	disordered region
9162	G	G	V	G	G	NA
9176	T	T	K	T	T	NA
9181	C	Y	C	Y	C	NA
9229	I	M	I	M	I	NA
9232	V	V	A	V	V	NA
9266	M	I	M	I	M	NA
9272	R	K	R	K	R	NA
9292	D	D	E	D	D	NA
9359	E	D	D	D	E	NA
9360	Y	Y	D	Y	Y	NA
9420	E	E	E	G	E	NA
9427	E	K	E	K	E	NA
9453	E	Q	E	Q	E	NA
9558	I	I	V	V	I	NA
9810	A	S	A	S	A	NA
9898	Q	H	H	H	Q	NA
9907	T	I	I	I	T	NA
9921	I	N	I	N	I	NA
9924	E	G	G	G	E	NA
9938	E	E	G	E	E	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
9944	I	V	V	V	I	NA
9979	T	K	T	K	T	NA
9983	L	L	S	L	L	NA
9995	S	S	N	S	S	NA
10006	Q	Q	K	Q	Q	NA
10014	I	T	I	T	I	NA
10020	S	L	S	L	S	NA
10027	T	T	I	T	T	NA
10039	P	Q	P	Q	P	NA
10040	E	Q	Q	Q	E	NA
10101	E	V	E	V	E	NA
10124	V	A	A	A	V	NA
10144	G	E	E	E	G	NA
10502	N	N	S	N	N	NA
10509	E	Q	E	Q	E	NA
10538	G	E	E	E	G	NA
10626	V	I	I	I	V	NA
10636	L	S	S	S	L	NA
10656	I	I	N	I	I	NA
10661	T	I	T	I	T	NA
10662	E	D	D	D	E	NA
10675	M	V	V	V	M	NA
10750	Q	E	Q	E	Q	NA
10790	Y	F	Y	F	Y	NA
10806	Q	P	Q	P	Q	NA
10845	I	V	I	V	I	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
10881	E	E	G	E	E	NA
10923	I	I	V	I	I	NA
10933	E	E	Q	E	E	NA
10935	K	E	E	E	K	NA
10936	G	R	G	R	G	NA
10972	Q	Q	P	Q	Q	NA
10978	E	K	E	K	E	NA
11012	D	N	D	N	D	NA
11023	T	T	N	T	T	NA
11024	K	K	E	K	K	NA
11030	K	K	N	K	K	NA
11198	Q	K	K	K	Q	NA
11289	A	T	T	T	A	NA
11290	E	E	E	D	E	NA
11370	Q	K	Q	K	Q	NA
11383	E	K	E	K	E	NA
11415	S	L	S	L	S	NA
11416	P	S	P	S	P	NA
11450	Q	H	H	H	Q	NA
11461	L	S	S	S	L	NA
11484	N	S	N	S	N	NA
11488	P	P	S	P	P	NA
11495	E	K	E	K	E	NA
11497	L	S	L	S	L	NA
11499	R	R	C	R	R	NA
11500	L	S	L	S	L	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
11516	T	T	I	T	T	NA
11533	E	E	D	E	E	NA
11629	L	L	P	L	L	NA
11722	P	L	L	L	P	NA
11772	E	K	K	K	E	NA
11789	I	I	V	I	I	NA
11805	M	T	T	T	M	NA
11808	V	F	V	F	V	NA
12002	T	A	A	A	T	NA
12137	I	I	M	I	I	NA
12190	A	V	A	V	A	NA
12211	E	G	E	G	E	NA
12221	S	N	S	N	S	NA
12233	P	P	Q	P	P	NA
12244	E	D	E	D	E	NA
12268	A	V	A	V	A	NA
12272	K	T	K	T	K	NA
12299	I	I	V	I	I	NA
12306	Q	Q	K	Q	Q	NA
12364	V	L	V	L	V	NA
12365	E	G	E	G	E	NA
12422	D	E	E	E	D	NA
12430	F	S	S	S	F	NA
12443	M	T	M	T	M	NA
12461	D	H	H	H	D	NA
12465	D	V	V	V	D	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
12469	N	K	N	N	N	NA
12549	H	R	H	R	H	NA
12568	C	Y	Y	Y	C	NA
12572	L	H	H	H	L	NA
12618	L	S	S	S	L	NA
12685	S	T	S	T	S	NA
12703	H	R	R	R	H	NA
12710	K	E	E	E	K	NA
12732	M	L	M	L	M	NA
12735	R	S	R	S	R	NA
12756	L	I	L	I	L	NA
12801	F	C	F	C	F	NA
12806	S	P	S	P	S	NA
12812	I	V	I	V	I	NA
12863	E	Q	E	Q	E	NA
12878	P	L	P	L	P	NA
12881	P	L	P	L	P	NA
12895	M	V	M	V	M	NA
12999	K	R	K	R	K	NA
13031	R	S	R	S	R	NA
13069	S	F	S	F	S	NA
13074	E	K	E	K	E	NA
13100	I	M	I	M	M	NA
13130	N	D	N	D	N	NA
13166	I	T	I	T	I	NA
13207	T	K	T	K	T	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
13218	T	A	T	A	T	NA
13387	R	R	C	R	C	NA
13464	K	K	T	K	T	spectrin repeat
13553	T	S	T	S	T	NA
13700	V	I	V	I	V	NA
13743	N	S	N	S	N	NA
13800	E	E	G	E	G	NA
13989	E	G	G	G	E	NA
13993	K	E	K	E	K	NA
14021	P	A	P	A	P	NA
14041	I	I	M	I	M	NA
14085	Q	K	K	K	K	NA
14089	T	A	A	A	A	NA
14109	V	A	V	A	V	NA
14119	F	F	S	F	S	NA
14128	V	V	V	V	I	NA
14133	L	L	S	L	S	NA
14216	P	P	S	P	S	NA
14277	K	K	E	K	E	NA
14305	T	T	I	T	I	NA
14368	D	D	D	E	D	NA
14407	P	P	S	P	S	NA
14430	S	S	P	S	P	NA
14471	E	E	D	E	D	NA
14476	Q	Q	P	Q	P	NA
14481	V	A	V	A	V	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	ref	IS W	IS Z	WE W	WE Z	domain
14491	L	L	S	L	S	NA
14517	V	I	V	I	V	NA
14519	L	F	L	F	L	NA
14535	N	H	N	H	N	NA
14560	V	I	V	I	V	NA
14566	M	M	T	M	T	NA
14595	E	E	K	E	K	NA
14606	Q	Q	E	Q	E	NA
14663	I	I	A	I	A	NA
14716	Q	L	Q	L	Q	NA
14771	K	K	E	E	E	NA
14773	Q	Q	Q	K	Q	NA
14906	S	S	N	S	N	NA
15019	T	R	T	R	T	spectrin repeat
15036	Q	R	R	R	R	spectrin repeat
15080	D	N	N	N	D	NA
15087	Q	K	K	K	Q	NA
15090	L	L	F	L	L	NA
15127	Q	Q	E	E	E	NA
15158	I	T	T	T	T	NA
15243	N	N	S	N	S	NA
15314	T	T	K	T	K	NA
15333	L	S	L	S	L	NA
15334	G	A	A	A	A	NA
15358	L	I	L	I	L	NA
15383	T	T	I	T	I	NA

Continued on next page

Table 3.6 – *Continued from previous page*

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
15446	T	T	I	T	I	NA
15508	S	T	S	T	S	NA
15611	S	S	L	S	L	NA
15634	L	M	L	M	L	NA
15686	A	A	A	L	A	NA
15808	Q	Q	E	Q	E	spectrin repeat
15847	Q	Q	K	Q	K	spectrin repeat
15880	A	A	V	A	V	spectrin repeat
16207	S	S	N	S	N	NA
16210	A	A	T	A	T	spectrin repeat
16341	H	Q	H	Q	H	spectrin repeat
16378	A	T	A	T	A	spectrin repeat
16393	Q	H	Q	H	Q	spectrin repeat
16452	M	M	I	M	I	spectrin repeat
16613	R	K	R	K	R	spectrin repeat
16625	I	V	I	V	I	spectrin repeat
16635	L	L	I	L	I	spectrin repeat
16899	V	V	I	V	I	spectrin repeat
17168	I	I	R	I	R	NA
17170	P	S	P	S	P	NA
17173	T	M	T	M	T	NA

Table 3.7.: Translated A1.46 gene 123 (SMC3) predicted domains and differences between Z and W′ sequences

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
154	E	Q	E	Q	E	RecF/RecN/SMC N terminal domain
260	M	V	M	V	M	RecF/RecN/SMC N terminal domain
374	E	K	E	K	E	RecF/RecN/SMC N terminal domain
390	R	R	R	Q	R	RecF/RecN/SMC N terminal domain
400	E	G	E	E	E	RecF/RecN/SMC N terminal domain
615	L	F	L	F	L	RecF/RecN/SMC N terminal domain
616	N	S	N	S	N	RecF/RecN/SMC N terminal domain
626	N	S	N	S	N	RecF/RecN/SMC N terminal domain
653	L	S	L	S	L	RecF/RecN/SMC N terminal domain
683	Q	H	Q	H	Q	RecF/RecN/SMC N terminal domain
849	V	L	V	L	V	RecF/RecN/SMC N terminal domain
1049	A	V	A	V	A	NA
1058	Q	R	Q	R	Q	NA

Table 3.8.: Translated A1.36 gene 53 (BRCA2) predicted domains and differences between Z and W′ sequences

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
54	P	S	P	S	P	disordered region
57	E	D	E	D	E	disordered region
71	V	I	V	I	V	NA
216	V	L	V	L	V	NA
226	L	S	L	S	L	NA
279	K	T	K	T	K	NA
409	T	A	T	A	T	NA
457	Q	H	Q	H	Q	NA
463	E	D	E	D	E	NA
467	V	A	V	A	V	NA
554	N	Q	N	Q	N	NA
572	F	L	F	L	F	NA
616	S	N	S	N	S	NA
636	V	A	V	A	V	NA
667	L	I	L	I	L	NA
672	F	V	F	V	F	NA
695	H	N	H	N	H	NA
720	K	E	K	E	K	NA
723	P	S	P	S	P	NA
727	L	M	L	M	L	NA
734	G	R	G	R	G	NA
741	V	D	V	D	V	NA
761	I	M	I	M	I	NA
779	N	D	N	D	N	NA

Continued on next page

Table 3.8 – *Continued from previous page*

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
808	T	S	T	S	T	helical domain
841	L	P	L	P	L	helical domain
861	N	H	N	H	N	helical domain
1051	Y	H	Y	H	Y	NA
1067	N	I	N	I	N	NA
1068	L	Y	L	Y	L	NA
1070	H	I	H	I	H	NA
1071	C	V	C	V	C	NA
1072	C	G	C	G	C	NA
1073	V	L	V	L	V	NA
1074	A	I	A	I	A	NA
1075	N	I	N	I	N	NA
1076	C	I	C	I	C	NA
1077	-	A	-	A	-	NA

Table 3.9.: Translated A1.46 gene 194 (split ends) predicted domains and differences between Z and W′ sequences

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
688	I	V	I	V	V	NA
694	T	S	T	S	S	NA
700	T	A	T	A	T	NA
812	P	S	P	S	P	NA
821	L	L	L	L	V	NA
898	P	S	P	S	P	NA
1010	H	L	H	L	H	NA
1030	V	I	V	I	I	disordered region
1058	Q	Q	Q	Q	H	disordered region
1248	V	A	V	A	V	NA
1265	S	S	S	L	S	NA
1269	T	T	T	A	T	NA
1337	H	Q	H	Q	H	disordered region
1711	T	S	T	S	T	NA
1734	E	D	E	D	E	disordered region
1735	E	D	E	D	E	disordered region
1780	I	M	I	M	M	disordered region
1903	G	S	G	S	G	disordered region
1951	T	S	T	S	T	disordered region
1954	S	P	S	P	S	disordered region
2248	S	P	S	P	P	NA
2253	S	N	S	S	S	NA
2256	L	S	L	L	L	NA

Continued on next page

Table 3.9 – *Continued from previous page*

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
2485	N	N	N	N	I	disordered region
2782	A	A	A	A	T	disordered region
2944	G	G	G	G	E	disordered region
2982	S	L	S	L	S	disordered region
2995	L	L	L	L	P	disordered region
3011	S	T	S	T	S	disordered region
3083	I	V	I	V	V	NA
3098	P	A	P	A	A	disordered region
3109	P	P	P	P	L	disordered region
3394	V	A	V	A	A	NA
3475	A	S	A	S	S	NA
3490	I	V	I	V	I	NA
3508	V	A	V	A	A	NA
3563	T	A	T	A	A	NA
3571	N	T	N	T	T	NA
3598	I	M	I	M	I	disordered region
3635	A	T	A	T	A	disordered region
3637	S	S	S	S	P	disordered region
3740	V	I	V	I	I	NA
3764	I	I	I	I	M	NA
3790	T	S	T	S	S	NA
3810	L	M	L	M	L	NA
3871	S	A	S	A	A	NA
4261	Q	H	Q	H	Q	NA

Table 3.10.: Translated A1.46 gene 124 (Tudor-domain containing protein) predicted domains and differences between Z and W′ sequences

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
20	F	L	F	L	L	NA
35	I	M	I	M	M	NA
149	V	V	V	A	A	NA
151	S	S	S	C	S	NA
193	S	P	S	P	S	NA
197	L	F	L	F	L	NA
226	A	A	A	A	T	NA
240	K	N	K	N	N	NA
277	H	N	H	N	H	Tudor domain, OB fold
301	A	V	A	V	A	Tudor domain, OB fold
311	V	M	V	M	V	Tudor domain, OB fold
386	N	D	N	D	N	OB fold
438	A	S	A	S	A	NA
505	T	T	N	T	N	Tudor domain, OB fold
508	S	L	S	L	S	Tudor domain, OB fold
546	S	A	S	A	S	Tudor domain, OB fold
554	N	S	S	N	S	Tudor domain, OB fold
596	L	M	L	M	L	OB fold
606	K	T	K	T	K	OB fold

Table 3.11.: Translated A1.46 gene145 (MSL2) predicted domains and differences between Z and W′ sequences

position	**ref**	**IS W**	**IS Z**	**WE W**	**WE Z**	**domain**
131	N	S	N	S	N	NA
182	I	T	I	T	I	NA
206	T	A	T	A	T	NA
211	R	C	R	C	R	NA
507	S	L	S	L	S	NA
563	A	T	A	T	A	NA

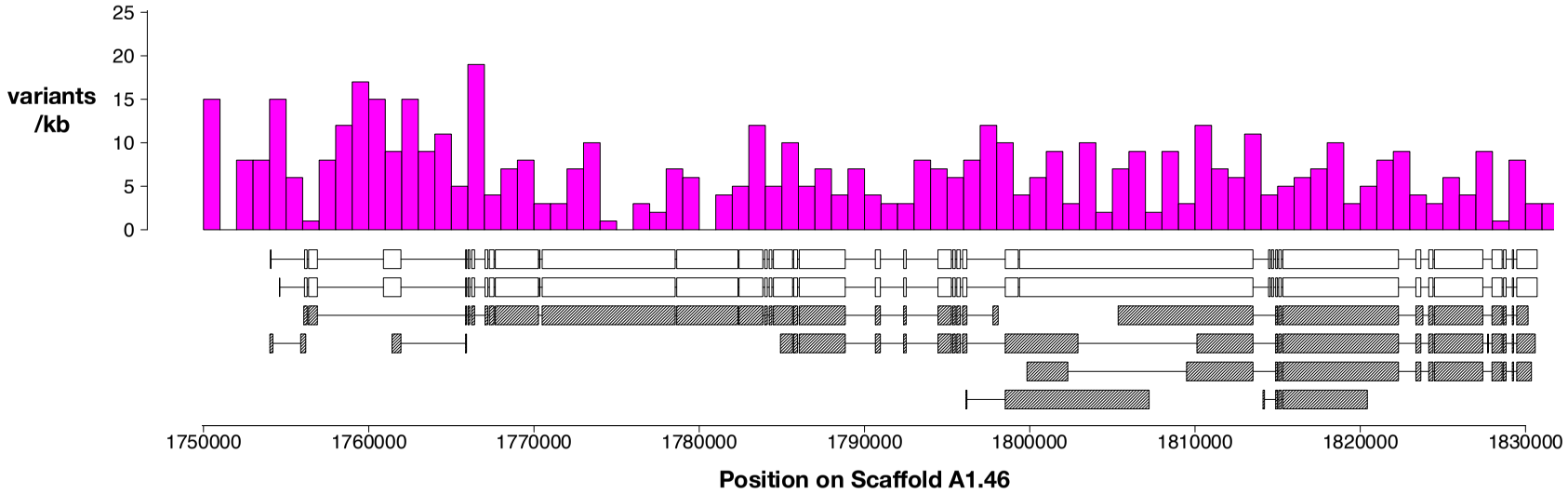

Fig. 3.3. Number of positions per kb in region of Nesprin-1 gene at which the W′ sequence has a SNP or indel with respect to the Z sequence for both Israel and White-eye populations. Each histogram bar width represents one kb. Positions within Scaffold A1.46 are numbered on the x-axis. Gene models (white) and ESTs (grey) are represented below the histogram. Exons and introns are shown as rectangles and lines, respectively.

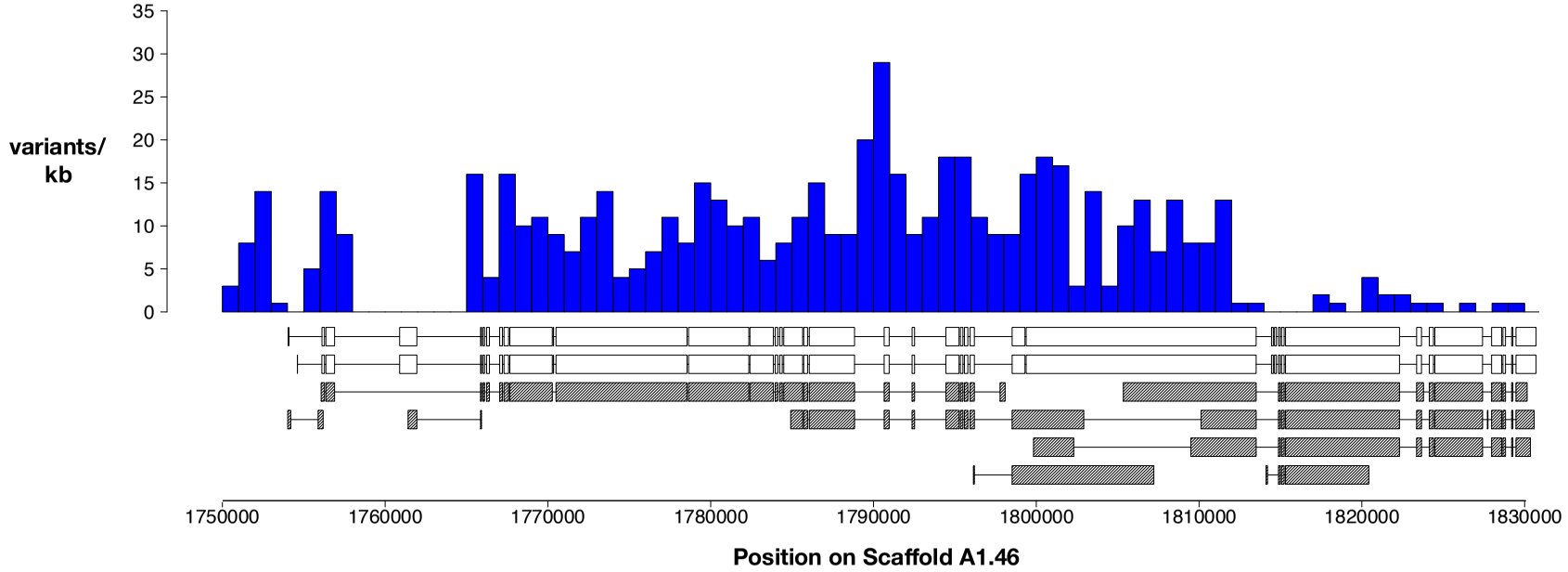

Fig. 3.4. Number of positions per kb in region of Nesprin-1 gene at which the Z sequence of the Israel population has a SNP or indel with respect to the Z sequence of the White-eye population. Each histogram bar width represents one kb. Positions within Scaffold A1.46 are numbered on the x-axis. Gene models (white) and ESTs (grey) are represented below the histogram. Exons and introns are shown as rectangles and lines, respectively.

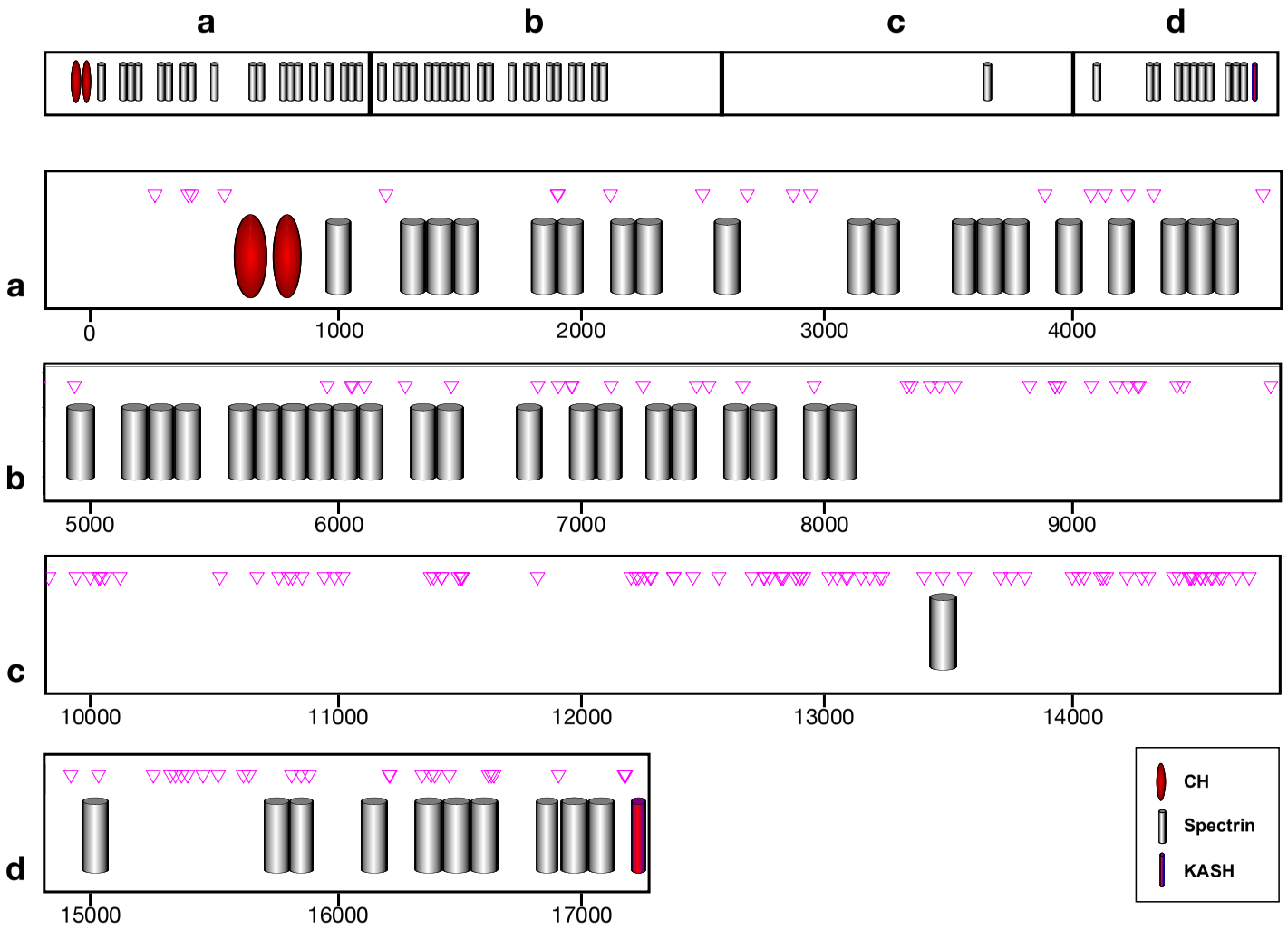

Fig. 3.5. A1.46 gene 147 (probable nesprin-1) domains and positions at which the W′ and Z translated gene sequences are different for both Israel and white-eye populations. Differences between the Z and W′ sequences are marked with pink triangles.

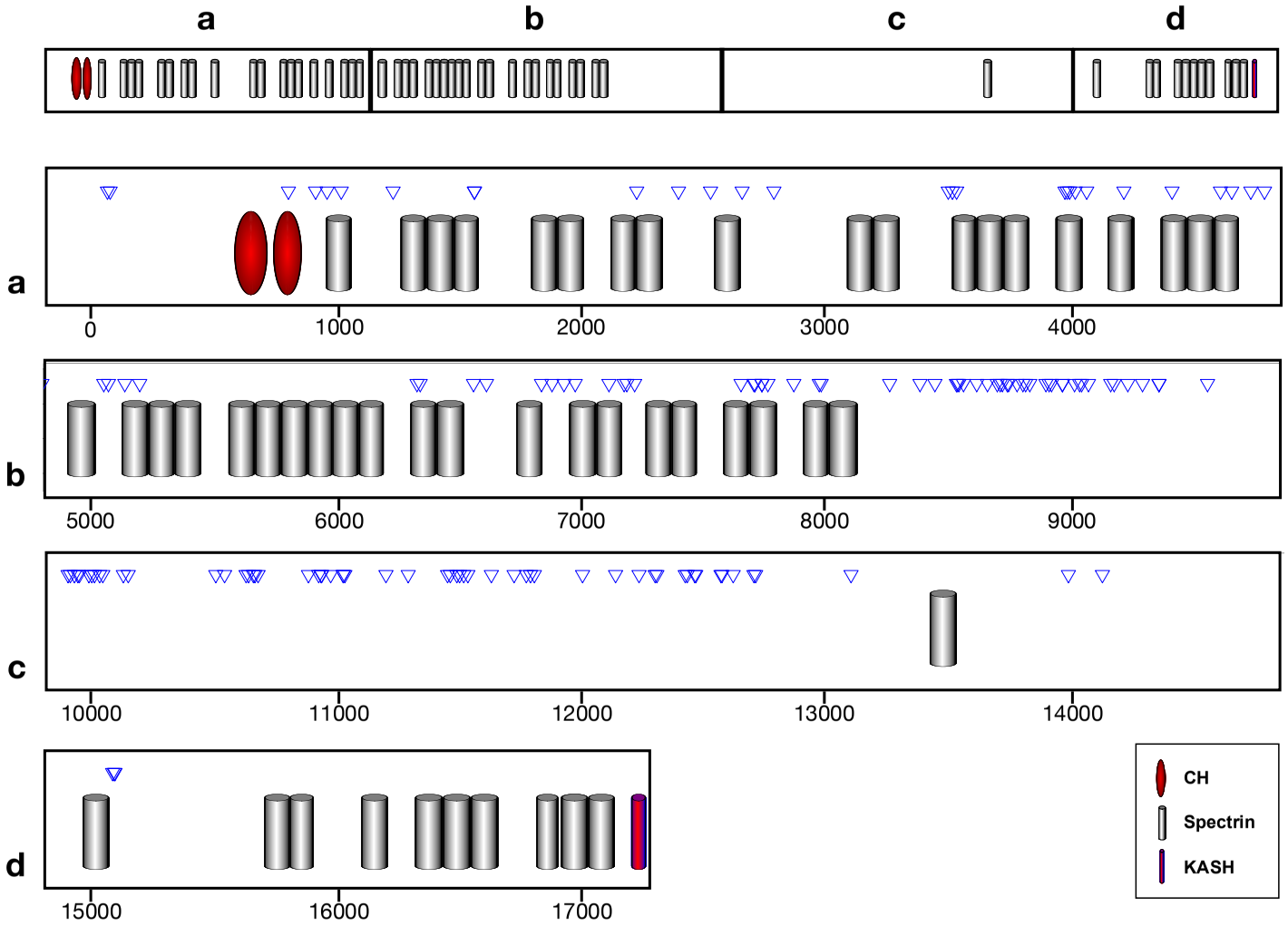

Fig. 3.6. A1.46 gene 147 (probable nesprin-1) domains and positions at which the Israel and white-eye Z translated gene sequences are different from each other. Differences between the Israel and white-eye Z sequences are marked with blue triangles.

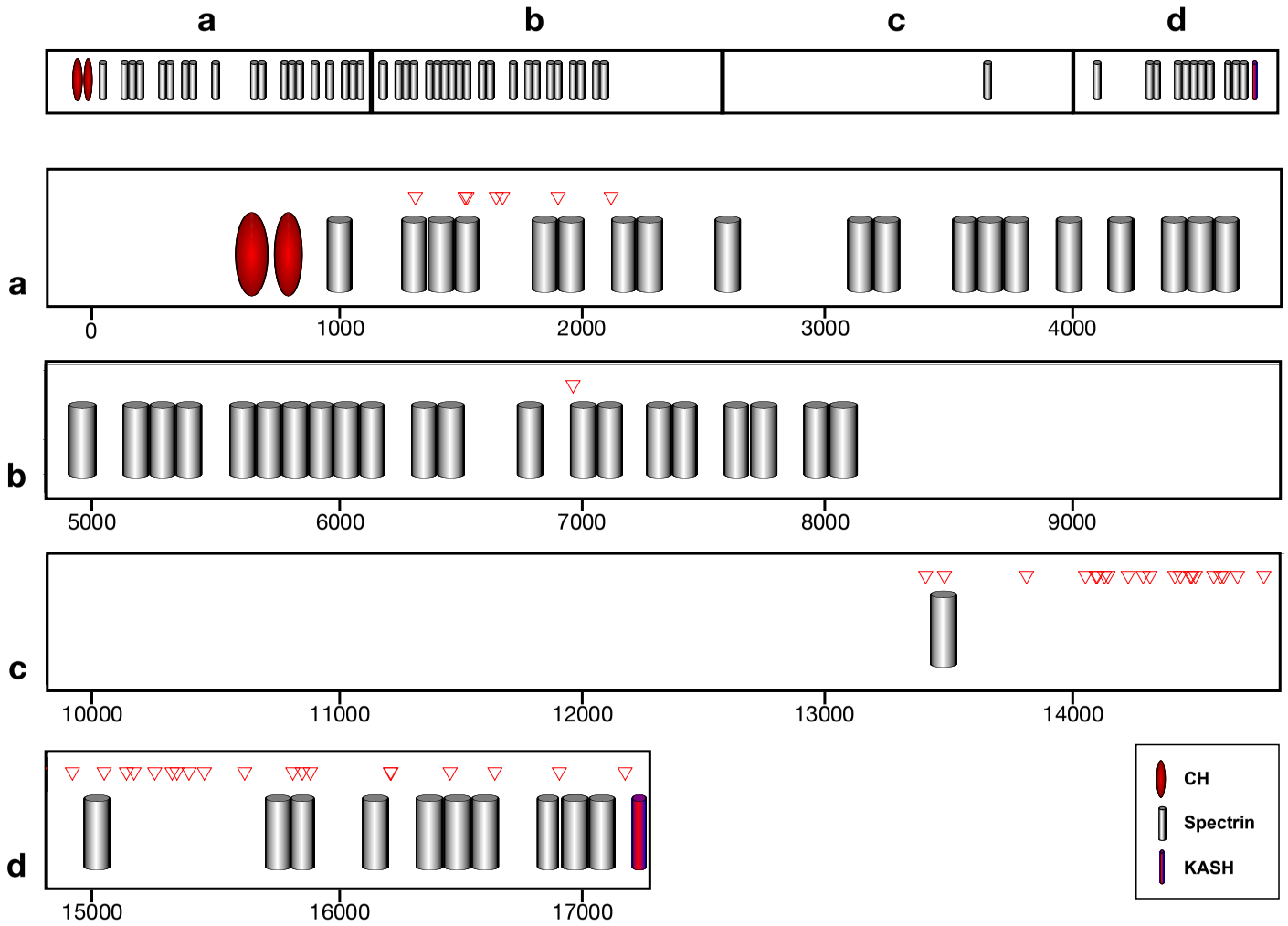

Fig. 3.7. A1.46 gene 147 (probable nesprin-1) domains and positions at which both the Israel and white-eye Z translated gene sequences are different from the translated reference gene sequence. Differences between the Z and reference sequences are marked with red triangles.

Fig. 3.8. A1.46 gene 147 (probable nesprin-1) KASH domain and nearby spectrin repeat positions compared with those of some of its top BLAST hits. Spectrin repeats are represented by grey cylinders. The KASH domain is represented by the red cylinder.

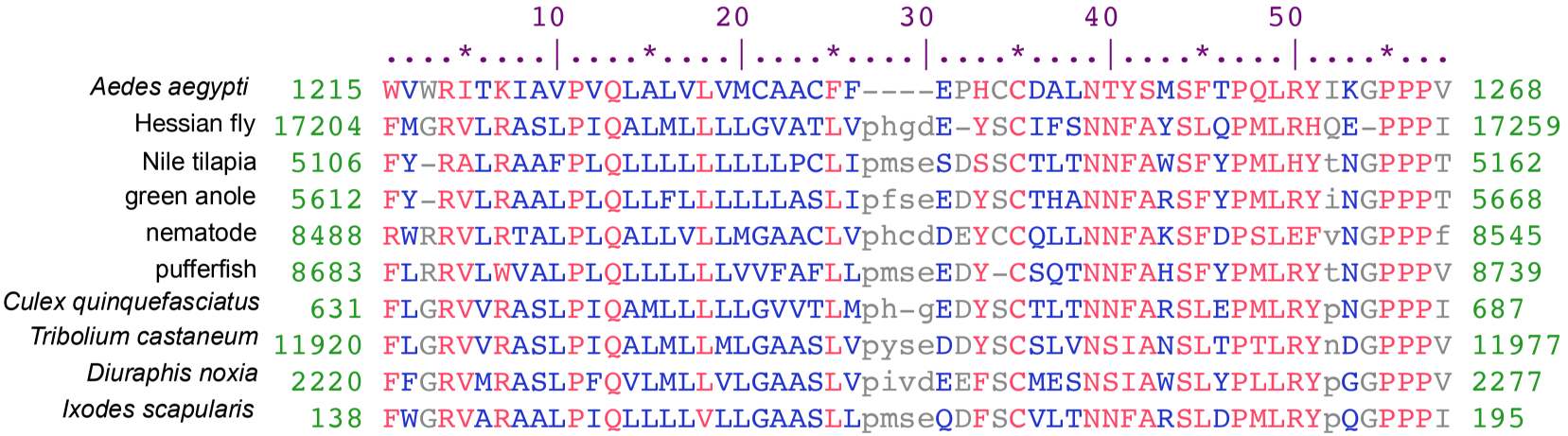

Fig. 3.9. A1.46 gene 147 (probable nesprin-1) KASH domain alignment with nine of the most diverse sequences used to build the KASH domain model (NCBI Conserved Domains Database). The most conserved residues are in red and unaligned residues are in grey.

REFERENCES

[1] C. Zhao, L. N. Escalante, H. Chen, T. R. Benatti, J. Qu, S. Chellapilla, R. M. Waterhouse, D. Wheeler, M. N. Andersson, R. Bao, M. Batterton, S. K. Behura, K. P. Blankenburg, D. Caragea, J. C. Carolan, M. Coyle, M. El-Bouhssini, L. Francisco, M. Friedrich, N. Gill, T. Grace, C. J. Grimmelikhuijzen, Y. Han, F. Hauser, N. Herndon, M. Holder, P. Ioannidis, L. Jackson, M. Javaid, S. N. Jhangiani, A. J. Johnson, D. Kalra, V. Korchina, C. L. Kovar, F. Lara, S. L. Lee, X. Liu, C. Löfstedt, R. Mata, T. Mathew, D. M. Muzny, S. Nagar, L. V. Nazareth, G. Okwuonu, F. Ongeri, L. Perales, B. F. Peterson, L.-L. Pu, H. M. Robertson, B. J. Schemerhorn, S. E. Scherer, J. T. Shreve, D. Simmons, S. Subramanyam, R. L. Thornton, K. Xue, G. M. Weissenberger, C. E. Williams, K. C. Worley, D. Zhu, Y. Zhu, M. O. Harris, R. H. Shukle, J. H. Werren, E. M. Zdobnov, M.-S. Chen, S. J. Brown, J. J. Stuart, and S. Richards, "A massive expansion of effector genes underlies gall-formation in the wheat pest mayetiola destructor," *Current Biology*, vol. 25, no. 5, pp. 613–620, March 2015.

[2] R. A. C. Z. J. G. W. M.-S. C. S. E. C. B. J. S. J. J. S. Thiago R. Benatti, Fernando H. Valicente, "A neo-sex chromosome that drives postzygotic sex determination in the hessian fly (mayetiola destructor)," *Genetics*, vol. 183, no. 3, pp. 769–777, March 2010.

[3] J. H. J.J. Stuart, "Cytogenetics of the hessian fly: Ii. inheritance and behavior of somatic and germ-line-limited chromosomes," *Journal of Heredity*, no. 79, pp. 190–199, 1988.

[4] C. Bantock, "Experiments on chromosome elimination in the gall midge, mayetiola destructor," *Journal of embryology and experimental morphology*, vol. 24, no. 2, pp. 257–286, September 1970.

[5] A. McClay, "Unisexual broods in the gall midge cystiphora sonchi (bremi) (diptera: Cecidomyiidae)," *Canadian entomologist*, vol. 128, no. 4, pp. 775–776, 1996.

[6] B. Matuszewski, *Animal Cytogenetics*, J. Bernard, Ed. Gebrüder Borntraeger, 1982, vol. 3.

[7] H. K. Salz, "Sex determination in insects: a binary decision based on alternative splicing," *Current opinion in genetics & development*, vol. 21, no. 4, pp. 395–400, 2011.

[8] J. Shukla and J. Nagaraju, "Doublesex: a conserved downstream gene controlled by diverse upstream regulators," *Journal of genetics*, vol. 89, no. 3, pp. 341–356, 2010.

[9] L. R. Bell, J. I. Horabin, P. Schedl, and T. W. Cline, "Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in drosophila," *Cell*, vol. 65, no. 2, pp. 229–239, 1991.

[10] C. Zhao, "Genomics of a gall midge: Avirulence and sex determination in the hessian fly (mayetiola destructor)," Ph.D. dissertation, Purdue University, 2011.

[11] E. Serna, E. Gorab, M. F. Ruiz, C. Goday, J. M. Eirín-López, and L. Sánchez, "The gene sex-lethal of the sciaridae family (order diptera, suborder nematocera) and its phylogeny in dipteran insects," *Genetics*, vol. 168, no. 2, pp. 907–921, 2004.

[12] E. C. Verhulst, L. van de Zande, and L. W. Beukeboom, "Insect sex determination: it all evolves around transformer," *Current opinion in genetics & development*, vol. 20, no. 4, pp. 376–383, 2010.

[13] K. Hoshijima, K. Inoue, I. Higuchi, H. Sakamoto, and Y. Shimura, "Control of doublesex alternative splicing by transformer and transformer-2 in drosophila," *Science*, vol. 252, no. 5007, pp. 833–837, 1991.

[14] I. Martín, M. F. Ruiz, and L. Sánchez, "The gene transformer-2 of sciara (diptera, nematocera) and its effect on drosophila sexual development," *BMC developmental biology*, vol. 11, no. 1, p. 19, 2011.

[15] K. T. Coschigano and P. C. Wensink, "Sex-specific transcriptional regulation by the male and female doublesex proteins of drosophila." *Genes & development*, vol. 7, no. 1, pp. 42–54, 1993.

[16] M. F. Ruiz, M. Alvarez, J. M. Eirín-López, F. Sarno, L. Kremer, J. L. Barbero, and L. Sánchez, "The gene doublesex of dipteran sciara does not follow the expression pattern observed in insects." *GENETICS*, p. 176701, 2015.

[17] R. S. M. O. H. Jeff J. Stuart, Ming-Shun Chen, "Gall midges (hessian flies) as plant pathogens," *Annual Review of Phytopathology*, no. 50, pp. 339–357, May 2012.

[18] J. K. Abbott, A. K. Nordén, and B. Hansson, "Sex chromosome evolution: historical insights and future perspectives," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 284, no. 1854, 2017. [Online]. Available: http://rspb.royalsocietypublishing.org/content/284/1854/20162806

[19] R. Bergero and D. Charlesworth, "The evolution of restricted recombination in sex chromosomes," *Trends in Ecology & Evolution*, vol. 24, no. 2, pp. 94–102, 2009.

[20] E. Axelsson, N. G. Smith, H. Sundstro m, S. Berlin, and H. Ellegren, "Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and turkey," *Molecular Biology and Evolution*, vol. 21, no. 8, pp. 1538–1547, 2004.

[21] M. A. W. Sayres, K. E. Lohmueller, and R. Nielsen, "Natural selection reduced diversity on human y chromosomes," *PLoS genetics*, vol. 10, no. 1, p. e1004064, 2014.

[22] D. J. Begun and C. F. Aquadro, "Levels of naturally occurring dna polymorphism correlate with recombination rates in d. melanogaster," *Nature*, vol. 356, no. 6369, pp. 519–520, 04 1992. [Online]. Available: http://dx.doi.org/10.1038/356519a0

[23] M. J. Lercher and L. D. Hurst, "Human snp variability and mutation rate are higher in regions of high recombination," *Trends in Genetics*, vol. 18, no. 7, pp. 337 – 340, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168952502026690

[24] R. J. Kulathinal, S. M. Bennett, C. L. Fitzpatrick, and M. A. Noor, "Fine-scale mapping of recombination rate in drosophila refines its correlation to diversity and divergence," *Proceedings of the National Academy of Sciences*, vol. 105, no. 29, pp. 10 051–10 056, 2008.

[25] K. D. M. Melissa A. Wilson Sayres, "Genome analyses substantiate male mutation bias in many species," *Bioessays*, vol. 33, no. 12, pp. 938–945, December 2011.

[26] M. R. E. Clara Goday, "Chromosome elimination in sciarid flies," *BioEssays*, vol. 23, no. 3, pp. 242–250, February 2001.

[27] L. Sánchez, "Sex-determining mechanisms in insects based on imprinting and elimination of chromosomes," *Sexual development*, vol. 8, no. 1-3, pp. 83–103, 2014.

[28] A. L. P. P. Lucas Sanchez, "Sex determination in sciarid flies: a model for the control of differential x-chromosome elimination," *Journal of Theoretical Biology*, vol. 197, no. 2, pp. 247–259, March 1999.

[29] B. de Saint Phalle and W. Sullivan, "Incomplete sister chromatid separation is the mechanism of programmed chromosome elimination during early sciara coprophila embryogenesis," *Development*, vol. 122, no. 12, pp. 3775–3784, 1996.

[30] S. Khosla, G. Mendiratta, and V. Brahmachari, "Genomic imprinting in the mealybugs," *Cytogenetic and genome research*, vol. 113, no. 1-4, pp. 41–52, 2006.

[31] E. Bonnefoy, G. A. Orsi, P. Couble, and B. Loppin, "The essential role of drosophila hira for de novo assembly of paternal chromatin at fertilization," *PLoS genetics*, vol. 3, no. 10, p. e182, 2007.

[32] H. V. Crouse, "The controlling element in sex chromosome behavior in sciara," *Genetics*, vol. 45, no. 10, p. 1429, 1960.

[33] A. C. Ferguson-Smith and M. A. Surani, "Imprinting and the epigenetic asymmetry between parental genomes," *Science*, vol. 293, no. 5532, pp. 1086–1089, 2001.

[34] A. Henckel, K. Nakabayashi, L. A. Sanz, R. Feil, K. Hata, and P. Arnaud, "Histone methylation is mechanistically linked to dna methylation at imprinting control regions in mammals," *Human molecular genetics*, vol. 18, no. 18, pp. 3375–3383, 2009.

[35] R. Pathak and R. Feil, "Oocyte-derived histone h3 lysine 27 methylation controls gene expression in the early embryo," *Nature Structural and Molecular Biology*, vol. 24, no. 9, p. 685, 2017.

[36] R. L. Kanippayoor, J. H. Alpern, and A. J. Moehring, "Protamines and spermatogenesis in drosophila and homo sapiens: a comparative analysis," *Spermatogenesis*, vol. 3, no. 2, p. e24376, 2013.

[37] S. D. Perreault, "Chromatin remodeling in mammalian zygotes," *Mutation Research/Reviews in Genetic Toxicology*, vol. 296, no. 1-2, pp. 43–55, 1992.

[38] S. J. Raja and R. Renkawitz-Pohl, "Replacement by drosophila melanogaster protamines and mst77f of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus," *Molecular and cellular biology*, vol. 25, no. 14, pp. 6165–6177, 2005.

[39] X. Tang, J. Cao, L. Zhang, Y. Huang, Q. Zhang, and Y. S. Rong, "Maternal haploid, a metalloprotease enriched at the largest satellite repeat and essential for genome integrity in drosophila embryos," *Genetics*, pp. genetics–117, 2017.

[40] F. B. P. C. Benjamin Loppin, Mylène Docquier, "The maternal effect mutation sésame affects the formation of the male pronucleus in drosophila melanogaster," *Developmental Biology*, no. 222, pp. 392–404, 2000.

[41] F. Landmann, G. A. Orsi, B. Loppin, and W. Sullivan, "Wolbachia-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus," *PLoS pathogens*, vol. 5, no. 3, p. e1000343, 2009.

[42] M. Sanei, R. Pickering, K. Kumke, S. Nasuda, and A. Houben, "Loss of centromeric histone h3 (cenh3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids," *Proceedings of the National Academy of Sciences*, vol. 108, no. 33, pp. E498–E505, 2011.

[43] M. Ravi and S. W. Chan, "Haploid plants produced by centromere-mediated genome elimination," *Nature*, vol. 464, no. 7288, p. 615, 2010.

[44] S. Maheshwari, E. H. Tan, A. West, F. C. H. Franklin, L. Comai, and S. W. Chan, "Naturally occurring differences in cenh3 affect chromosome segregation in zygotic mitosis of hybrids," *PLoS genetics*, vol. 11, no. 1, p. e1004970, 2015.

[45] S. K. Zaidi, D. W. Young, M. A. Montecino, J. B. Lian, A. J. Van Wijnen, J. L. Stein, and G. S. Stein, "Mitotic bookmarking of genes: a novel dimension to epigenetic control," *Nature reviews genetics*, vol. 11, no. 8, p. 583, 2010.

[46] A. J. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications," *Cell research*, vol. 21, no. 3, p. 381, 2011.

[47] J. R. Swedlow and T. Hirano, "The making of the mitotic chromosome: modern insights into classical questions," *Molecular cell*, vol. 11, no. 3, pp. 557–569, 2003.

[48] M. Carmena, M. Wheelock, H. Funabiki, and W. C. Earnshaw, "The chromosomal passenger complex (cpc): from easy rider to the godfather of mitosis," *Nature reviews Molecular cell biology*, vol. 13, no. 12, p. 789, 2012.

[49] F. Wang and J. M. Higgins, "Histone modifications and mitosis: countermarks, landmarks, and bookmarks," *Trends in cell biology*, vol. 23, no. 4, pp. 175–184, 2013.

[50] T. Hirota, J. J. Lipp, B.-H. Toh, and J.-M. Peters, "Histone h3 serine 10 phosphorylation by aurora b causes hp1 dissociation from heterochromatin," *Nature*, vol. 438, no. 7071, p. 1176, 2005.

[51] R. Giet and D. M. Glover, "Drosophila aurora b kinase is required for histone h3 phosphorylation and condensin recruitment during chromosome condensation and to organize the central spindle during cytokinesis," *The Journal of cell biology*, vol. 152, no. 4, pp. 669–682, 2001.

[52] R. A. Oliveira, S. Heidmann, and C. E. Sunkel, "Condensin i binds chromatin early in prophase and displays a highly dynamic association with drosophila mitotic chromosomes," *Chromosoma*, vol. 116, no. 3, pp. 259–274, 2007.

[53] Y. Wang and S. Jia, "Degrees make all the difference: the multifunctionality of histone h4 lysine 20 methylation," *Epigenetics*, vol. 4, no. 5, pp. 273–276, 2009.

[54] M. Schuh, C. F. Lehner, and S. Heidmann, "Incorporation of drosophila cid/cenp-a and cenp-c into centromeres during early embryonic anaphase," *Current Biology*, vol. 17, no. 3, pp. 237–243, 2007.

[55] B. A. Sullivan and G. H. Karpen, "Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin," *Nature structural & molecular biology*, vol. 11, no. 11, p. 1076, 2004.

[56] J. Fang, Y. Liu, Y. Wei, W. Deng, Z. Yu, L. Huang, Y. Teng, T. Yao, Q. You, H. Ruan *et al.*, "Structural transitions of centromeric chromatin regulate the cell cycle-dependent recruitment of cenp-n," *Genes & development*, vol. 29, no. 10, pp. 1058–1073, 2015.

[57] I. M. Cheeseman, "The kinetochore," *Cold Spring Harbor perspectives in biology*, vol. 6, no. 7, p. a015826, 2014.

[58] M. F. Wolfner, "Nuclear envelope dynamics in drosophila pronuclear formation and in embryos," *Nuclear Envelope Dynamics in Embryos and Somatic Cells*, pp. 131–142, 2003.

[59] K. R. Katsani, R. E. Karess, N. Dostatni, and V. Doye, "In vivo dynamics of drosophila nuclear envelope components," *Molecular biology of the cell*, vol. 19, no. 9, pp. 3652–3666, 2008.

[60] A. Harel, E. Zlotkin, S. Nainudel-Epszteyn, N. Feinstein, P. A. Fisher, and Y. Gruenbaum, "Persistence of major nuclear envelope antigens in an envelope-like structure during mitosis in drosophila melanogaster embryos," *Journal of Cell Science*, vol. 94, no. 3, pp. 463–470, 1989.

[61] H. Maiato, J. DeLuca, E. Salmon, and W. C. Earnshaw, "The dynamic kinetochore-microtubule interface," *Journal of cell science*, vol. 117, no. 23, pp. 5461–5477, 2004.

[62] A. Amon, "The spindle checkpoint," *Current opinion in genetics & development*, vol. 9, no. 1, pp. 69–75, 1999.

[63] A. Musacchio and K. G. Hardwick, "The spindle checkpoint: structural insights into dynamic signalling," *Nature reviews Molecular cell biology*, vol. 3, no. 10, p. 731, 2002.

[64] E. Logarinho, H. Bousbaa, J. M. Dias, C. Lopes, I. Amorim, A. Antunes-Martins, and C. E. Sunkel, "Different spindle checkpoint proteins monitor microtubule attachment and tension at kinetochores in drosophila cells," *J Cell Sci*, vol. 117, no. 9, pp. 1757–1771, 2004.

[65] J.-M. Peters, "The anaphase-promoting complex: proteolysis in mitosis and beyond," *Molecular cell*, vol. 9, no. 5, pp. 931–943, 2002.

[66] J. Gregan, S. Polakova, L. Zhang, I. M. Tolić-Nørrelykke, and D. Cimini, "Merotelic kinetochore attachment: causes and effects," *Trends in cell biology*, vol. 21, no. 6, pp. 374–381, 2011.

[67] R. A. Oliveira and K. Nasmyth, "Getting through anaphase: splitting the sisters and beyond," 2010.

[68] F. Hans and S. Dimitrov, "Histone h3 phosphorylation and cell division," *Oncogene*, vol. 20, no. 24, p. 3021, 2001.

[69] M. C. Escribá and C. Goday, "Histone h3 phosphorylation and elimination of paternal x chromosomes at early cleavages in sciarid flies," *J Cell Sci*, vol. 126, no. 14, pp. 3214–3222, 2013.

[70] S. H. Harvey, M. J. Krien, and M. J. O'Connell, "Structural maintenance of chromosomes (smc) proteins, a family of conserved atpases," *Genome biology*, vol. 3, no. 2, pp. reviews3003–1, 2002.

[71] B. Robison, V. Guacci, and D. Koshland, "A role for the smc3 hinge domain in the maintenance of sister chromatid cohesion," *Molecular biology of the cell*, pp. mbc–E17, 2017.

[72] S. Vaur, A. Feytout, S. Vazquez, and J.-P. Javerzat, "Pds5 promotes cohesin acetylation and stable cohesin–chromosome interaction," *EMBO reports*, vol. 13, no. 7, pp. 645–652, 2012.

[73] Z. Misulovin, M. Pherson, M. Gause, and D. Dorsett, "Brca2, pds5 and wapl differentially control cohesin chromosome association and function," *PLoS genetics*, vol. 14, no. 2, p. e1007225, 2018.

[74] M. J. Betts and R. B. Russell, "Amino acid properties and consequences of substitutions," *Bioinformatics for geneticists*, vol. 317, p. 289, 2003.

[75] K. Mace and A. Tugores, "The product of the split ends gene is required for the maintenance of positional information during drosophila development," *BMC developmental biology*, vol. 4, no. 1, p. 15, 2004.

[76] H. Su, Y. Liu, X. Zhao *et al.*, "Split end family rna binding proteins: Novel tumor suppressors coupling transcriptional regulation with rna processing," *Cancer translational medicine*, vol. 1, no. 1, p. 21, 2015.

[77] L. H. Jin, J. K. Choi, B. Kim, H. S. Cho, J. Kim, J. Kim-Ha, and Y.-J. Kim, "Requirement of split ends for epigenetic regulation of notch signal-dependent genes during infection-induced hemocyte differentiation," *Molecular and cellular biology*, vol. 29, no. 6, pp. 1515–1525, 2009.

[78] I. Pinheiro and E. Heard, "X chromosome inactivation: new players in the initiation of gene silencing," *F1000Research*, vol. 6, 2017.

[79] E. R. Smith, A. Pannuti, W. Gu, A. Steurnagel, R. G. Cook, C. D. Allis, and J. C. Lucchesi, "The drosophila msl complex acetylates histone h4 at lysine 16, a chromatin modification linked to dosage compensation," *Molecular and cellular biology*, vol. 20, no. 1, pp. 312–318, 2000.

[80] K. Copps, R. Richman, L. M. Lyman, K. A. Chang, J. Rampersad-Ammons, and M. I. Kuroda, "Complex formation by the drosophila msl proteins: role of the msl2 ring finger in protein complex assembly," *The EMBO journal*, vol. 17, no. 18, pp. 5409–5417, 1998.

[81] S. Zheng, R. Villa, J. Wang, Y. Feng, J. Wang, P. B. Becker, and K. Ye, "Structural basis of x chromosome dna recognition by the msl2 cxc domain during drosophila dosage compensation," *Genes & development*, vol. 28, no. 23, pp. 2652–2662, 2014.

[82] M. Ying and D. Chen, "Tudor domain-containing proteins of drosophila melanogaster," *Development, growth & differentiation*, vol. 54, no. 1, pp. 32–43, 2012.

[83] H. Salz and J. W. Erickson, "Sex determination in drosophila: The view from the top," *Fly*, vol. 4, no. 1, pp. 60–70, 2010.

[84] E. J. Duncan, M. J. Wilson, J. M. Smith, and P. K. Dearden, "Evolutionary origin and genomic organisation of runt-domain containing genes in arthropods," *Bmc Genomics*, vol. 9, no. 1, p. 558, 2008.

[85] J. B. Duffy and J. P. Gergen, "The drosophila segmentation gene runt acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene sex-lethal." *Genes & Development*, vol. 5, no. 12a, pp. 2176–2187, 1991.

[86] K. Nishioka, J. C. Rice, K. Sarma, H. Erdjument-Bromage, J. Werner, Y. Wang, S. Chuikov, P. Valenzuela, P. Tempst, R. Steward *et al.*, "Pr-set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone h4 and is associated with silent chromatin," *Molecular cell*, vol. 9, no. 6, pp. 1201–1213, 2002.

[87] M. Schaefer, J. P. Steringer, and F. Lyko, "The drosophila cytosine-5 methyltransferase dnmt2 is associated with the nuclear matrix and can access dna during mitosis," *PloS one*, vol. 3, no. 1, p. e1414, 2008.

[88] G. G. Luxton and D. A. Starr, "Kashing up with the nucleus: novel functional roles of kash proteins at the cytoplasmic surface of the nucleus," *Current opinion in cell biology*, vol. 28, pp. 69–75, 2014.

[89] A. Rothballer and U. Kutay, "The diverse functional lincs of the nuclear envelope to the cytoskeleton and chromatin," *Chromosoma*, vol. 122, no. 5, pp. 415–429, 2013.

[90] D. Rajgor, J. A. Mellad, F. Autore, Q. Zhang, and C. M. Shanahan, "Multiple novel nesprin-1 and nesprin-2 variants act as versatile tissue-specific intracellular scaffolds," *PloS one*, vol. 7, no. 7, p. e40098, 2012.

[91] K. Djinovic-Carugo, M. Gautel, J. Ylänne, and P. Young, "The spectrin repeat: a structural platform for cytoskeletal protein assemblies," *FEBS letters*, vol. 513, no. 1, pp. 119–123, 2002.

[92] K. G. Young and R. Kothary, "Spectrin repeat proteins in the nucleus," *Bioessays*, vol. 27, no. 2, pp. 144–152, 2005.

[93] Q. Zhang, C. Bethmann, N. F. Worth, J. D. Davies, C. Wasner, A. Feuer, C. D. Ragnauth, Q. Yi, J. A. Mellad, D. T. Warren *et al.*, "Nesprin-1 and-2 are involved in the pathogenesis of emery–dreifuss muscular dystrophy and are critical for nuclear envelope integrity," *Human molecular genetics*, vol. 16, no. 23, pp. 2816–2833, 2007.

[94] Q. Zhang, C. Ragnauth, M. J. Greener, C. M. Shanahan, and R. G. Roberts, "The nesprins are giant actin-binding proteins, orthologous to drosophila melanogaster muscle protein msp-300," *Genomics*, vol. 80, no. 5, pp. 473–481, 2002.

[95] J. G. Simpson and R. G. Roberts, "Patterns of evolutionary conservation in the nesprin genes highlight probable functionally important protein domains and isoforms," 2008.

[96] C. Zhou, L. Rao, C. M. Shanahan, and Q. Zhang, "Nesprin-1/2: roles in nuclear envelope organisation, myogenesis and muscle disease," *Biochemical Society Transactions*, vol. 46, no. 2, pp. 311–320, 2018.

[97] M. M. Babu, "The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease," *Biochemical Society Transactions*, vol. 44, no. 5, pp. 1185–1200, 2016.

[98] J. S. Martinez, C. Baldeyron, and A. Carreira, "Molding brca2 function through its interacting partners," *Cell Cycle*, vol. 14, no. 21, pp. 3389–3395, 2015.

[99] A. Weinberg-Shukron, M. Rachmiel, P. Renbaum, S. Gulsuner, T. Walsh, O. Lobel, A. Dreifuss, A. Ben-Moshe, S. Zeligson, R. Segel *et al.*, "Essential role of brca2 in ovarian development and function," *New England Journal of Medicine*, vol. 379, no. 11, pp. 1042–1049, 2018.

[100] M. Klovstad, U. Abdu, and T. Schüpbach, "Drosophila brca2 is required for mitotic and meiotic dna repair and efficient activation of the meiotic recombination checkpoint," *PLoS genetics*, vol. 4, no. 2, p. e31, 2008.

[101] T. Kusch, "Brca2–pds5 complexes mobilize persistent meiotic recombination sites to the nuclear envelope," *J Cell Sci*, vol. 128, no. 4, pp. 717–727, 2015.

[102] E. Choi, P.-G. Park, H.-o. Lee, Y.-K. Lee, G. H. Kang, J. W. Lee, W. Han, H. C. Lee, D.-Y. Noh, S. Lekomtsev *et al.*, "Brca2 fine-tunes the spindle assembly checkpoint through reinforcement of bubr1 acetylation," *Developmental cell*, vol. 22, no. 2, pp. 295–308, 2012.

[103] M. A. Lampson and T. M. Kapoor, "The human mitotic checkpoint protein bubr1 regulates chromosome–spindle attachments," *Nature cell biology*, vol. 7, no. 1, p. 93, 2005.