11-24-2017 10:00 AM

Characterizing Mekk1: Candidate Behavioural Isolation Gene

Caryn Dooner, *The University of Western Ontario*

Supervisor: Dr. Amanda Moehring, *The University of Western Ontario*
Co-Supervisor: Dr. Susanne Kohalmi, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

Recommended Citation

Dooner, Caryn, "Characterizing Mekk1: Candidate Behavioural Isolation Gene" (2017). *Electronic Thesis and Dissertation Repository*. 5138.
https://ir.lib.uwo.ca/etd/5138

Abstract

Behavioural isolation can occur due to divergence in aspects of courtship and mating, and can contribute to reproductive isolation. The purpose of this study is to determine how a gene, *Mekk1,* contributes to female rejection behaviour between *D. melanogaster* and *D. simulans*. Unique variants were identified within *D. simulans Mekk1* that could contribute to behaviour, most of which are non-coding. Both transcripts of *Mekk1* appear to be expressed at similar levels in *D. simulans* and *D. melanogaster*. These data also indicate that *Mekk1* may be expressed in a specific region of the brain called the mushroom body, which has previously been linked to female rejection behaviour in *D. simulans*. It is possible that *Mekk1* in *D. simulans* is acting through the mushroom body to mediate female rejection behaviour. This study will aid in understanding molecular contributions to behavioural isolation between species of *Drosophila*.

Keywords

Evolution, behavioural isolation, *Drosophila*, gene expression

Acknowledgements

There are many people I would like to acknowledge that have contributed significantly to this thesis. Thank-you to my co-supervisor, Dr. Susanne Kohalmi, and to my supervisor Dr. Amanda Moehring. Dr. Kohalmi you contributed in so many ways and I can't thank you enough. Thank-you to Dr. Moehring, I learned a lot from you.

I would like to thank my advisors, Dr. Graham Thompson and Dr. Anne Simon. Thank-you for your valuable feedback on my work throughout my degree, and thank-you to Dr. Thompson and for always being supportive. Thank-you to Dr. Anthony Percival-Smith for being a reliable sounding board and for fielding so many of my silly questions. I also want to thank the coolest fly lady I know, Dr. Tara Edwards. I would never have learned how to read a genotype or dissect a brain without your help. Thank-you to Dr. Brent Sinclair for letting me borrow some of your equipment and for always being encouraging. Thank-you to Dr. Jamie Kramer for giving me copies of some fly stocks and for giving me some great advice on how to improve my protocols. Thank-you to Dr. Andre Lachance for helping me improve my grammar as well as my understanding of evolution. Thank-you to Dr. Ben Rubin for all of your help with stats, especially R.

Thank-you to BioRad and BioRad representatives Michael Wakem and Claude Lachance for all of your patience and assistance with anything from gel doc troubles to ddPCR. Thank-you to Geneious and Geneious representative Daisy Wang. Thank-you to Dr. Gurpreet Dhami for patiently showing me how to use so much of the equipment in the MGU Core. Thank-you to the Biotron Experimental Climate Change Research Centre (UWO) for allowing me to use both the Insect and Microscopy Modules. Thank-you to Karen Nygard, Carrie Hamilton, and Steve Bartlett.

Thank-you to Jess Pardy for showing me the molecular “ropes” and for always being around for science rants. Thank-you to Jacqueline Lebenzon and Susan Anthony for always being willing to share some take-out on weekends or late lab nights, grabbing early morning coffees, and being my outreach buddies. Thanks to Jalina Bielaska-Da Silva for answering all of my behaviour questions, Heather Ward for making copious amounts of fly food on the regular, and Pria Mahabir for always chatting about fly brains. Thank-you to Dova Brennman and Rachel Chambers for all of the constructive science chats.

Thank-you to my Kohalmi labmates – Emily Clayton, Emily Cornelius, Sara Rad, and Jessica Belcastro. I never felt like the odd one out.

Finally, thank-you to my family for supporting me through this process. It would not have been possible without you.

Table of Contents

Abstract i

Keywords i

Acknowledgements ii

List of Tables vii

List of Figures viii

List of Appendices x

Abbreviations xii

1: Introduction 1

1.1 Speciation 1

1.2 Sexual selection and behaviour 6

1.3 *Drosophila* as a model organism for speciation 9

1.4 *Drosophila* behaviour and species isolation 13

1.5 *Mekk1*: previous general characterization 18

1.6 *Mekk1*: previous behavioural characterization 20

1.7 Hypotheses and Objectives 23

1.7.1 Overall Objective 23

1.7.2 Sequencing characterization 23

1.7.3. Characterizing Mekk1 transcript expression 24

1.7.4. Tissue-specific expression characterization 27

1.7.5. Determining tissue-specificity of Mekk1-associated behaviour 28

1.7.6. Determining if the known function of Mekk1 contributes to behaviour 28

2: Material and Methods 30

2.1 *Drosophila* stocks 30

2.2 *Drosophila* crosses 30

2.3 DNA isolation 36

2.4 Primer design and testing 36

2.5 DNA preparation for sequencing 42

2.6 Sequence analysis 42

2.7 Tissue preparation for RNA isolation 43

2.8 RNA isolation 44

2.9 RNA quantification 44

2.10 cDNA synthesis 44

2.11 RT-PCR 45

2.12 *Mekk1* visualization 46

2.13 Generating the genotypes for the behavioural rescue experiment 46

2.14 Genotyping 47

2.15 Mating assays 48

3: Results 50

3.1 Comparative sequence analysis of *Mekk1* gene region 50

3.2 Coding region analysis of *Mekk1* 58

3.3 Analysis of proximal regulatory regions of *Mekk1* 59

3.4 Transcript confirmation 62

3.5 Expression analysis of *Mekk1* transcripts 64

3.6 Behavioural assays: rescuing *Mekk1* in a tissue-specific manner 66

3.7 Behavioural assays: disrupting the coding region of *Mekk1* 74

3.8 *Mekk1* visualization 75

4: Discussion 81

4.1 Multiple unique variants may contribute to behaviour 81

4.2 *Mekk1* expression is similar between *D. simulans* and *D. melanogaster* 83

4.3 Evidence that *Mekk1* may be expressed in the mushroom body....................... 85

4.4 Potential mechanisms of *Mekk1* as a behavioural isolation gene....................... 86

4.5 Conclusions and future directions.. 88

References 92

CV 190

List of Tables

Table 1.1. Summary of objectives and hypotheses. 25

Table 2.1 List of *Drosophila* strains with source and genotype. 31

Table 2.2 *D. melanogaster* primers for sequencing. 38

Table 2.3 *D. simulans* primers for sequencing. 40

Table 2.4 Primers for semi-quantitative RT-PCR. 41

Table 3.1 Filter system showing how sequence differences were categorized. 52

Table 3.2 Percent similarity for each strain of *D. simulans* as compared to *D. melanogaster* (BJS) within *Mekk1* region. 54

Table 3.3 Statistical values for *Mekk1* RT-PCR by lifecycle stage and tissue type. 67

List of Figures

Figure 1.1. Behavioural phenotypes associated with hybrid females and candidate behavioural isolation gene disruptions. 10

Figure 1.2. Schematic comparing genotypes tested during deficiency mapping of chromosomes for candidate genes contributing to behavioural isolation. 16

Figure 1.3. Principle of deficiency mapping. 17

Figure 1.4. Locations of *P*-element insertions in previously characterized *Mekk1* deficiency strains. 21

Figure 2.1. Crossing scheme (1) illustrating how a *P*-element disruption and a UAS element linked to *Mekk1* were recombined onto a single homologous chromosome 32

Figure 2.2. Crossing scheme (2) shows how stocks were generated for later use in the behavioural rescue experiments. 33

Figure 2.3. Crossing scheme (3) showing inter- and intra-species crosses for *Mekk1* behavioural assay experiments. 34

Figure 2.4. Crossing scheme (4) showing how various genotypes were combined and used to visualize tissue-specific expression of *Mekk1*. 35

Figure 2.5. Relative position of *D. melanogaster* primers for sequencing. 37

Figure 2.6. Relative position of *D. simulans* primers for sequencing. 39

Figure 3.1. *Mekk1* gene region showing the relative positions of variants unique to *D. simulans* (FC). 53

Figure 3.2. Diagram of initiator element (Inr) in *D. melanogaster* compared to *D. simulans*. 60

Figure 3.3. Diagram showing alternative start of putative transcripts in *D. melanogaster* and *D. simulans* 63

Figure 3.4. RT-PCR shows presence of two transcripts in both *D. melanogaster* and *D. simulans*. 65

Figure 3.5. Comparison of relative *Mekk1* expression levels determined by RT-PCR between species across lifecycle stages and tissue types. 68

Figure 3.6. Effect of tissue-specific *Mekk1* rescue on mating frequency in various genotypes of *Drosophila* hybrid females. 71

Figure 3.7. Courtship latency of males from *D. melanogaster* (BJS) and *D. melanogaster* (GFP) paired with females from both *D. melanogaster* (BJS) and *D. melanogaster* (GFP). 72

Figure 3.8. Courtship latency of males from *D. melanogaster* (BJS) and *D. melanogaster* (GFP) paired with females from both *D. melanogaster* (BJS) and *D. melanogaster* (GFP). 73

Figure 3.9. Mating frequency for deletion mapping of functional kinase domain of *Mekk1*. 76

Figure 3.10. Visualizing expression using GFP 77

Figure 3.11. Images of *Mekk1* expression pattern using GFP in the female brain. 80

Figure 4.1. Schematic for a potential *Mekk1*-mediated pathway that influences female receptivity. 89

List of Appendices

Appendix A: Modifications to RNA isolation procedure. 101

Appendix B: Sequence comparison between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216) for *Mekk1* gene region from upstream gene to 3’ UTR. 103

Appendix C: Coverage map of the data used to generate the *Mekk1* consensus sequence for each of the four strains, including electropherograms in areas of uncertainty. 123

Appendix D: Thesis data compared to unedited independent re-analysis. 130

Appendix E: Sequence comparison for transcript RB of *Mekk1* between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216). 162

Appendix F: Sequence comparison for transcript RD/RC of *Mekk1* between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216). 170

Appendix G: Sequence comparison for primary structure of RB protein of *Mekk1* between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216). 177

Appendix H: Sequence comparison for primary structure of RD/RC protein of *Mekk1* between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216). 180

Appendix I: Gel images from RT-PCR comparing relative levels of *Mekk1* expression in females between *D. simulans* (FC) and *D. melanogaster* (BJS). 183

Appendix J: Genotypes of flies from behavioural assays examining tissue-specific rescue of *Mekk1* and its effect on behaviour. 187

Appendix K: Fluorescent microscopy examining female adult ovaries of the strain with an enhancer trap within *Mekk1* in *D. melanogaster*............................ 188

Appendix L: Fluorescent microscopy examining female adult brains of the strain with an enhancer trap within *Mekk1* in *D. melanogaster*............................ 189

Abbreviations

SI Units not listed

ANOVA	Analysis of Variance
Bal	Balancer
BDSC	Berkeley Drosophila Stock Center
BJS	Dr. Brent J. Sinclair
BSC	Biological Species Concept
cDNA	Complimentary deoxyribonucleic acid
CO_2	Carbon Dioxide
ddH_2O	Double Distilled Water (H_2O)
Def	Deficiency
Del	Deletion
Dis	Disruption
DNA	Deoxyribonucleic acid
DVM	Dominant Visible Marker
E.T.	Enhancer Trap
EDTA	Ethylenediaminetetraacetic acid
EMBL-EBI	European Molecular Biology Laboratory-European Bioinformatics Institute
EST	Expressed Sequence Tag
FC	Florida City
FRET	Fluorescence Resonance Energy Transfer
FITC	Fluorescein isothiocyanate

gDNA	Genomic deoxyribonucleic acid
GFP	Green Flourescent Protein
Indel	Insertion/deletion
M-	*Mekk1* deficiency
MAPKKK	Mitogen Activated Protein Kinase Kinase Kinase
MEL	*D. melanogaster*
MMLV	Mouse Murine Lukemia Virus
M-X	*Mekk1* deletion
N/A	Not applicable
NaCl	Sodium Chloride
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
ProK	Proteinase K
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
SE	Standard error
SIM	*D. simulans*
SNV	Single Nucleotide Variant
TrisHCL	Hydroxymethylaminomethane hydrochloride
UAS	Upstream Activating Sequence
UTR	Untranslated Region

Wt	Wild-type

1: Introduction

1.1 Speciation

The range of biodiversity on our planet can be explained through the process of speciation and evolution (Katzourakis et al. 2001; Noor & Feder 2006; Sobel et al. 2014). Speciation is the process wherein new species are formed, and evolution is change at the population level over time (Darwin 1859; Mayr 1942). Selection refers to the process of certain individuals having the ability to out-reproduce one another, based on their ability to survive or differential ability to secure mates.

The most broad example of selection is natural selection (Darwin 1859), compared to other forms of selection such as sexual and artificial selection. If individuals are able to survive, they will have the opportunity to reproduce (Darwin 1859). Natural selection is the process through which individuals with traits not well-suited to their environment will not live to reproduce. Within natural selection, a process termed sexual selection specifically refers to an individual's ability to out-reproduce other individuals of the same species (Darwin 1859; Fisher 1930). In sexual selection, if a single individual has reproductive characteristics that confer an advantage over other individuals of the same species, it can increase the single individual's reproductive success.

Individuals having traits not well-suited to the environment they live in are less fit, and consequently have fewer offspring. Fewer offspring likely means fewer chances for their traits and genes to be passed on to the third generation. Inversely, individuals with characteristics or traits that are well-suited to the environment are likely more fit, will produce more offspring, and therefore have more of their traits passed on to the next generation. There are many different environments, each of which can lead to the selection of unique traits that are well-suited to a specific environment but not well-suited to other environments (specific traits; Sobel et al. 2014).

Individuals of the same species living in the same environment (a subpopulation of a species) may have the same specific traits. If a subpopulation of a species is left in an unique environment for several generations, this subpopulation may develop traits specific enough that they can no longer interbreed with individuals from other subpopulations that have been living in other environments that select for different specific traits. Specific traits that are well-suited to a specific environment can contribute to differences between populations and is one mechanism through which speciation can occur. The Biological Species Concept (BSC) defines separate species as groups of individuals that cannot breed with one another. The inability of individuals of distinct populations (and species according to the BSC) to reproduce with one another is referred to as reproductive isolation (Mayr 1942; Dobzhansky 1937).

Speciation, or the formation of reproductively isolated species, can include both allopatric and sympatric speciation (reviewed by Coyne and Orr 2004). When speciation occurs in allopatry (allopatric speciation), subpopulations of a given species are geographically isolated and live separated from one another without any contact (Mayr 1942). As a result, the independent populations become well-suited to their respective environments over time. When subpopulations lose contact with one another, there is a physical barrier that contributes to their separation which also prevents gene flow between the subpopulations. Lack of gene flow means that there is no exchange of genetic material between subpopulations. While the populations are isolated from one another, they can accumulate differences in their genome – in this case through selective pressures from the distinct environments. These differences that accumulate over time may result in physically separated populations being so genetically different that they are now reproductively isolated from one another. Hence in allopatric speciation, geographic isolation can lead to the formation of new species by creating physical separation and preventing gene flow. Once populations are reproductively isolated from one another, these populations are considered within the Biological Species Concept to be different species that diverged in allopatry (Mayr 1942, Dobzhansky 1937).

In the example above, the differences between two species that diverged in allopatry are mediated by different environments and by differential selection acting on geographically isolated subpopulations, which can lead to ecological adaptation (Sobel et al. 2014). In contrast to allopatric speciation, sympatric speciation explains how new species arise in the same geographic location (Mayr 1942). In other words, in sympatric subpopulations are not geographically separated and are exposed to the same environment (reviewed by Coyne and Orr 2004). Not only are the subpopulations exposed to the same environment, but the lack of a geographic barrier means that there is continual exposure of the diverging subpopulations to one another as well as a continuous opportunity for gene flow to occur.

There are a couple of examples that outline different means through which species develop reproductive isolation in sympatry (Savolainen et al. 2006, Barluenga et al. 2006). Savolainen et al. (2006) found evidence in Australia that species of *Howea*, or palm trees, diverged in sympatry (Savolainen et al. 2006). Evidence suggests that their flowering times differ by a matter of weeks and that they vary in their ability to thrive in soils with acidic pH. However, Savolainen et al. (2006) are unsure as to which of these differences contributed to a speciation event or if one of these differences arose after the speciation event. There is also evidence of sympatric speciation between species of *Amphilophus*, cichlid fish, in Cameroon (Barluenga et al. 2006). The sympatric species of cichlid are not only reproductively isolated but vary in coloration, with *A. citrinellus* being a much lighter beige compared to the darker brown *A. zaliosus*. Barluenga et al. (2006) hypothesized that these differences arose due to stratification in the cichlids crater. Two species of *Drosophila, Drosophila simulans* and *D. melanogaster,* exist in sympatry and remain reproductively isolated (Carracedo et al. 2003) which will be discussed further (section 1.2).

Aside from mechanisms leading to reproductive isolation and speciation events (including allopatry and sympatry), types of reproductive isolation can be classified as either post-zygotic or prezygotic (reviewed by Coyne and Orr 2004). Post-zygotic isolation refers to mechanisms that occur after the formation of a zygote: including maladapted hybrids, hybrid sterility, and hybrid inviability. Hybrid offspring formed from two diverging subpopulations having lower fitness is an example of post-zygotic isolation (Mayr 1963; Orr 1996). In some instances, sympatric populations may exhibit post-zygotic isolation: when one parent is from subpopulation A while the other parent is from the subpopulation B, their union forms a hybrid (AB) with decreased fitness relative to offspring purely from one subpopulation or the other (AA or BB). The previous example of post-zygotic isolation means that some matings do not result in post-zygotic issues, which indicates that the diverging subpopulations are not completely reproductively isolated. The maladapted hybrid offspring will reinforce the species divergence as individuals that have a preference not to mate with the diverging subpopulation will have more fit offspring, and pass on more of their traits (Butlin 1989). This is one manner of post-zygotic isolation and how it can contribute to reproductive isolation.

Prezygotic refers to reproductive isolating mechanisms that occur prior to fertilization and the formation of a zygote. Prezygotic isolation has a diverse array of mechanisms and can be used to explain speciation in sympatry. Species that exist in sympatry typically have stronger prezygotic isolating mechanisms compared to species that exist in allopatry, as discussed by Coyne and Orr (2004). This is thought to be related to the likelihood of exposure: populations in allopatry will rarely (if ever) be exposed to their diverged counterparts in another geographical region, whereas populations in sympatry will continually be exposed to one another. Prezygotic isolating mechanisms prevent mating earlier in an interaction, which is advantageous if individuals are continually exposed to potential mates that they are not reproductively compatible with. If individuals who are not reproductively compatible do mate successfully, it may result in maladapted hybrids which does not benefit either individual.

One mechanism of prezygotic isolation includes behavioural isolation (Dobzhansky 1937; Mayr 1947; Coyne and Orr 2004). Behavioural isolation applies to many species that have courtship and other mating rituals that are necessary to initiate reproduction to occur (Dobzhansky 1937). When subpopulations diverge, and courtship (or another behaviour) is altered slightly from one subpopulation to the next such that the next subpopulation does not recognize the first subpopulations courtship, then mating will not take place in a form of prezygotic reproductive isolation.

For example, male bowerbirds decorate their bower elaborately (Borgia 1995). If a male is unable to make an elaborate nest or makes the ‘wrong type’ because his behaviour has been altered, then the female bowerbird will not recognize that there is a male ready to mate and subsequent mating will not occur (Uy and Borgia 2000). The genetic basis of reproductive isolation may contribute to speciation, thus understanding the roles and regulation of genes and proteins involved in this process will increase our knowledge about speciation events.

1.2 Sexual selection and behaviour

Sexual selection is a form of natural selection (Darwin 1859), and can contribute to the process of speciation. Sexual selection results from an individual’s differing ability to secure a mate and in the propagation of traits that the opposite sex finds attractive (Mayr 1949; Birkhead 2000). Sexual selection can act on secondary sexual characteristics, which includes traits not directly involved in sexual reproduction (ex: coloration or behaviour). For a trait to be selected, one sex must have a preference for that trait. Females may be choosy for a variety of reasons, depending on the species. For example, female *Gryllus bimaculatus* (crickets) may be more choosy than males due to detriments incurred at mating (Green and Tregenza 2009). In species such as the cricket, females may incur detriments such as a decrease in longevity from mating with a male. In the case of the cricket, males may have secondary sexual characteristic that allow the female to determine the negative impact she may experience by mating with him.

There are hypotheses addressing the reason(s) the attractiveness of secondary sexual characteristics, that can be grouped into direct and indirect benefit hypotheses (Fisher, 1930). The direct benefit hypothesis states that an individual cannot only have elaborate traits, but must also be able to provide a female with a benefit such as protection or availability of food or access to a larger territory. Indirect benefit hypotheses can be broken down into the “sexy son’s” hypothesis and the “good genes” hypothesis. “Sexy son’s” hypothesis refers to the fact that because this male has attractive traits (remembering that females are generally the choosy ones), the male offspring of her union will also have attractive traits and produce lots of their own offspring propagating their mothers genes (Fisher 1930). The “good genes” hypothesis states that as this individual has good (attractive) traits, their offspring will have high fitness from inheriting their fathers’ good traits. Females typically have a preference for secondary sexual characteristics that may correlate with male fitness, such as his ability to provide direct benefits (such as food) for the female (Byrne and Rice 2005) or because the male will be able to pass on his “good” genes to their offspring (Fisher 1930).

For example, the male peacock tail is quite elaborate being large in size with a bright and diverse array of colors (Darwin 1871). Females find the elaborate male tail attractive, so females that prefer an elaborate tail will mate with males that have a more elaborate tail. As females that have a preference for the trait and males that have the trait will increase in frequency in the population, it is possible for the preference and trait to be inherited together and become linked (Kirkpatrick 1982). The term linkage disequilibrium refers to traits that are not linked and therefore not inherited together. Sexual selection in combination with linkage disequilibrium can result in multiple traits being inherited in a non-random manner, eventually leading to significant divergence between populations such that they are now reproductively isolated and considered separate species (Servedio 2009). If sexual selection is strong enough and occurs with preferences for two opposing extremes, it can potentially lead to speciation even in sympatry (Higashi et al. 1999). For example, speciation between *D. simulans* and *D. melanogaster* is believed to have occurred in sympatry due to female preference mediating sexual selection on the subpopulations (Carracedo et al. 2003).

Within sexual selection, antagonistic co-evolution is a form of sexual conflict between the sexes that can lead to evolution (Rice 1996). Sexual antagonism describes a phenomenon where one sex will develop an adaptation that is detrimental to the opposite sex, and this initial adaptation will select for a counter-adaptation in the opposite sex (Rice 1996). The adaptation may contribute to increasing the control one sex has over reproductive success (Arnqvist and Rowe 2002; Rice 1996). It is possible for female choice, the focus of this thesis, to result from sexual antagonism. Sexual antagonism can lead to females evolving an adaptation that increases their level of control over their reproductive success, which may help to increase the female's fitness. Sexual antagonism may also influence a female's preference for males of the same species *versus* males of another species (Parker and Partridge 1998). Sexual conflict, such as sexual antagonism, has been shown to contribute to speciation in insects (Arnqvist et al. 2000). Females choosing conspecific males to avoid heterospecific males and their detriments would facilitate the formation and maintenance of species barriers.

Hall (1994) examined *Drosophila* courtship, during which females are exposed to a male's courtship and must choose whether to accept or reject him. Females of certain *Drosophila* species have been shown to mate more quickly when exposed to a species-specific song which indicates that song can play a role in reproductive isolation (Ritchie et al. 1999). A male that cannot produce a song typical of the female's species will have diminished likelihood of mating with the female, resulting in the female and male being behaviourally isolated from one another in a form of reproductive isolation. With respect to reproductive isolation, the importance of early isolating mechanisms cannot be overstated (Coyne and Orr 2004; Noor and Feder 2006). Simple things that appear small yet compound or are precursor to larger isolating mechanisms are very important, even though they may not seem obvious. Something as subtle as a change in the pulse rate of a courtship song could be an initial behavioural divergence that contributed to reproductive isolation and subsequent speciation (Ritchie et al. 1999).

1.3 *Drosophila* as a model organism for speciation

Drosophila has a broad geographical range (Lachaise et al. 1986; Carracedo et al. 2003), and various species of *Drosophila* inhabit different ecological niches giving an ideal opportunity to study multiple modes of speciation and divergence (Coyne and Orr 2004). *Drosophila simulans* and *D. melanogaster* are both species with a broad geographical range that currently exist in sympatry (Lachaise et al. 1986; Carracedo et al. 2003). Interestingly, females of *D. simulans* reject *D. melanogaster* males, but females of *D. melanogaster* will mate with *D. simulans* males. This asymmetric heterospecific mating tendency sets up an scenario to study the genetic basis for female rejection behaviour (Laturney and Moehring 2012a; Fig. 1.1). In this case, the asymmetry is indicative of one species having increased discrimination compared to the other species. This thesis intends to capitalize on that unique discrimination and characterize it to determine how *D. simulans* female choice is mediated against a heterospecific male.

Drosophila are common model organisms for several reasons, including the ease of rearing in a laboratory and the fact that they have a short generation time; hence, it is

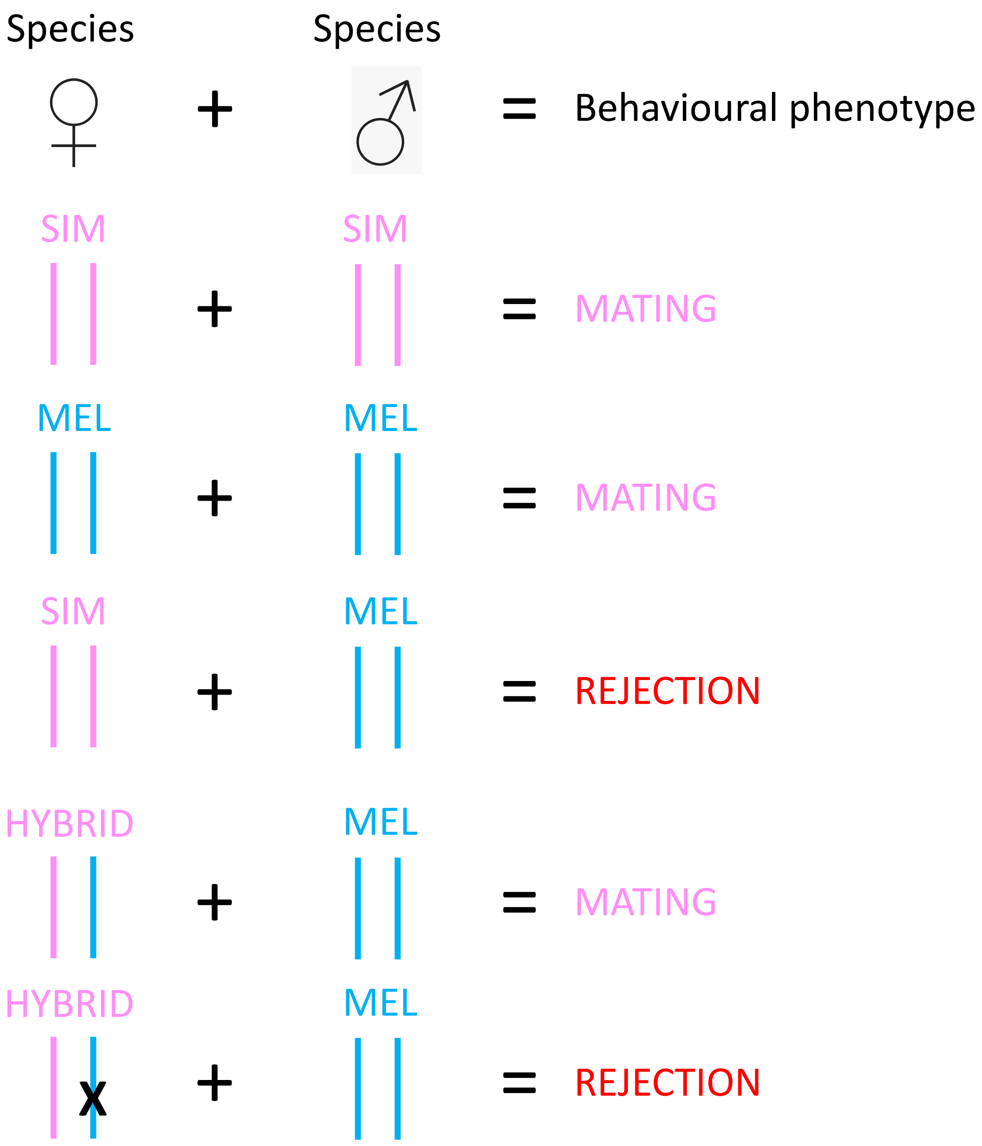


Figure 1.1. Behavioural phenotypes associated with hybrid females and candidate behavioural isolation gene disruptions. Species are identified as SIM (*D. simulans*) or MEL (*D. melanogaster*), "X" indicates a gene disruption. Female individuals (left) are paired with males (middle) and the behavioural phenotype that the females exhibit is shown (MATING or REJECTION; on the right); phenotypes illustrated based on Laturney and Moehring (2012a).

possible to impose artificial selection on transgenic strains. *Drosophila* are ideal for genetic studies because of the plethora of genetic tools that are available. For example, *Drosophila* can be used for deficiency mapping (Pasyukova et al. 2000) which will be discussed in more detail in section 1.4. The *Drosophila* Gene Disruption project has targeted 95% of *Drosophila* genes using site-specific transposons (Bellen et al. 2011), which has contributed to the development of databases such as Flybase (Gramates et al. 2017). Other databases that are commonly used for information on *Drosophila* are Flyatlas (Chintapalli et al. 2007) and Flymine (Lyne et al. 2007). There are many tools that allow genetic alterations to be introduced into the genome (Spradling et al. 1999; Bellen et al. 2004; Thibault et al. 2004; Metaxakis et al. 2005; Bellen et al. 2011; Bischof et al. 2013). For previous studies in *Drosophila* a construct was inserted into the genome to disrupt the function of a specific gene, and to subsequently observe the effect on phenotype. Two examples of transposons are called *P*-elements or *Minos* elements which can be used to cause gene disruptions (Metaxakis et al. 2005). The phenotype that the fly has when a disruption, such as a transposon, is present helps scientists infer the function of the disrupted genetic region.

For example, a modified *Minos* element has been used to analyze tissue-specific gene expression (Venken et al., 2011). Expression patterns can be monitored because a fluorescent marker, Green Fluorescent Protein (GFP), is part of this modified *Minos* element. The GFP expression pattern allows visualization of the expression of the genomic region into which *Minos* integrates. Another way to examine tissue-specific gene expression is using a combination of an Upstream Activating Sequence (UAS)-GFP element and an enhancer trap (Duffy 2002). An enhancer trap contains a transposable element with a reporter gene (in this case GAL4), and it allows one to visualize where a regulatory element is being expressed by “hijacking” the genes enhancer (a regulatory element; Bellen et al. 1989).

While *D. melanogaster* is well-characterized and has numerous genetic tools available, other *Drosophila* species are lacking those advantages. While *D. simulans* is less characterized compared to *D. melanogaster*, comparative analysis has been performed between *D. simulans* and *D. melanogaster* at both the DNA and the transcript expression levels (specifically by Begun et al. 2007). For example, there is an inversion in the right arm of the third chromosome, which contains the gene of interest (*Mekk1*). *Drosophila simulans* and *D. melanogaster* shared a common ancestor roughly 4.3mya, around the same time the two strains (subpopulations) of *D. simulans* are thought to have differentiated (Cutter 2008). Hence, it is likely that the genomic sequences differ between the two *Drosophila* species (González et al. 1982), as well as between strains of *D. simulans* (Choudhary and Singh 1987).

Many different modes of genetic regulation have been described for *Drosophila* genes including promoters, enhancers, and insulator sequences (Wittkopp et al. 2004; Nègre et al. 2010; Arnold et al. 2014). As seen for other organisms, promoter sequences located directly upstream of a gene regulate gene expression. It is possible that the manner in which genes are regulated between species, including and outside of direct promoter regulation, underlies the reproductive isolation exhibited between *Drosophila* species and is influenced by sequence divergence between species. As there are many ways to regulate gene expression and some are more difficult than others to characterize (or have shown to be conserved between species), the proximal promoter region will be the regulatory focus in this thesis.

1.4 *Drosophila* behaviour and species isolation

The field of behavioural genetics focuses on understanding the heritable basis of behaviours. Examples of genes that influence behaviour include *fruitless* (*fru*; Hall 1978)and *Viola* (Balakireva et al. 1998), which have both been shown to influence male sexual orientation (Balakireva et al. 1998; Hall 1978). As reviewed by Coyne and Orr (2004), males typically court females at a higher frequency than females will accept the males offer. Therefore, many relevant male behaviours have been analyzed to identify associated genes. For example, the *period* (*per*) gene contributes to the courtship song of a *Drosophila* male (Wheeler et al. 1991), as does *fru* (Clyne and Miesenböck 2008). Combined with the fact that Ritchie et al. (1999) found females to have a species-specific preference for male song, it is possible that the period gene contributes to behavioural isolation between *Drosophila* species.

A review by Bakker and Pomiankowski (1995) discusses female mate choice in the context of male traits. An important but understudied area of behavioural genetics includes genes that influence female receptivity (Laturney and Moehring 2012b), despite evidence supporting female choice as the main isolating barrier between the two species of interest in this thesis (Izquierdo et al. 1992). During courtship in *Drosophila*, the male is typically the “sender” as he emits a behaviour or signal which is received by a female who needs to interpret the behaviour before accepting or rejecting him as a mate (Ryan 1998). Genes that alter the rejection behaviour of a female will influence how the male’s behaviour is interpreted. Although *Drosophila* females exhibit choice and can reject

D. melanogaster males by flying away, kicking him off during courtship, or extruding her ovipositor to prevent mating (Hall 1994), the males are not usually as selective. As mating involves two individuals, it is important to consider that the behaviour of interest is not the only behaviour relevant in the courtship interaction.

Drosophila courtship was described in detail by Hall (1994). First, the male detects the female and subsequently taps the female's abdomen. The male follows the female, and then extends his wing and vibrates it in a "love song". Components of the sounds are among the more species-specific elements of *Drosophila* courtship. Females of certain Drosophila species have been shown to mate more quickly when exposed to a species-specific song which indicates that song can play a role in reproductive isolation (Ritchie et al. 1999). A male that cannot produce a song typical of the female's species will have decreased likelihood of mating with the female. After the wing song, the male will extend his proboscis and lick the female's genitalia, followed immediately by the first copulation attempt which involves abdominal bending by the male (Hall 1994). *Drosophila* female's "choice" (ex: rejection by flying away) can occur at any point throughout the courtship (Hall 1994).

To determine which genes are involved in female mating behaviour, female hybrids (*D. melanogaster* x *D. simulans*) were examined with and without certain regions disrupted (Laturney and Moehring 2012a). As a result, candidate regions for behavioural isolation genes were identified in *D. simulans* females. The paper (Laturney and Moehring 2012a) further outlines that the behavioural phenotype of the female depends on the genetic background (pure species *vs* hybrid), and it was shown that the rejection phenotype seen in pure species *D. simulans* females is recessive to *D. melanogaster* acceptance (as *D. simulans* will reject *D. melanogaster* males and hybrids will mate). If loci responsible for *D. simulans* rejection are “unmasked”, then the behavioural phenotype in the hybrid will change from acceptance to rejection. Rejection may be unmasked by disrupting the dominant *D. melanogaster* locus that is a “factor” contributing to acceptance of *D. melanogaster* males. The “factor” contributing to acceptance could be anything present in DNA, coding or non-coding.

The principle of hybrid deficiency mapping is to compare the behavioural phenotype of individuals with and without disruptions in the *D. melanogaster* half of their genome, and observe whether or not they have the same behavioural phenotype (Fig. 1.1; Laturney and Moehring 2012a). If individuals with or without a genomic disruption have different phenotypes, such that the individuals with a disruption reject and individuals without a disruption accept *D. melanogaster* males, then the genomic region that is disrupted contains at least one or more candidate loci that contribute to behavioural isolation (Fig. 1.1, Fig. 1.2). Deficiency mapping involves using multiple different strains of transgenic flies that have certain genomic regions disrupted or knocked out, regions that are characterized at the cytological level (Fig. 1.3; discussed in Laturney and Moehring 2012a). Deficiency-mapping was used by Laturney and Moehring (2012a) to identify five regions along the right arm of the third chromosome, which was refined to three specific candidate genes (using fine-mapping), that contribute to behavioural isolation between *D. simulans* and *D. melanogaster*: *Mekk1*, *fru*, and *kat60* (Moehring lab, unpublished). Consistent with the resuts of Merrill et al. (2011), who examined behavioural isolation between species of *Heliconius* butterflies, the candidate

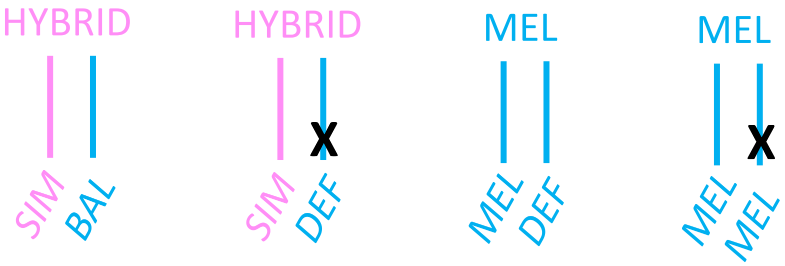


Figure 1.2. Schematic comparing genotypes tested during deficiency mapping of chromosomes for candidate genes contributing to behavioural isolation. The *BAL* (balancer) and *DEF* (deficiency) are tested in both a HYBRID (female hybrid) and MEL (*D. melanogaster*), with “X” indicating a gene disruption. For each individual, the two lines represent the right arm of the third chromosome (two homologs per individual) with either a wild-type *D. melanogaster* or *D. simulans* chromosome paired with either a deficiency or a balancer chromosome. Blue=*D. melanogaster*, pink=*D. simulans*.

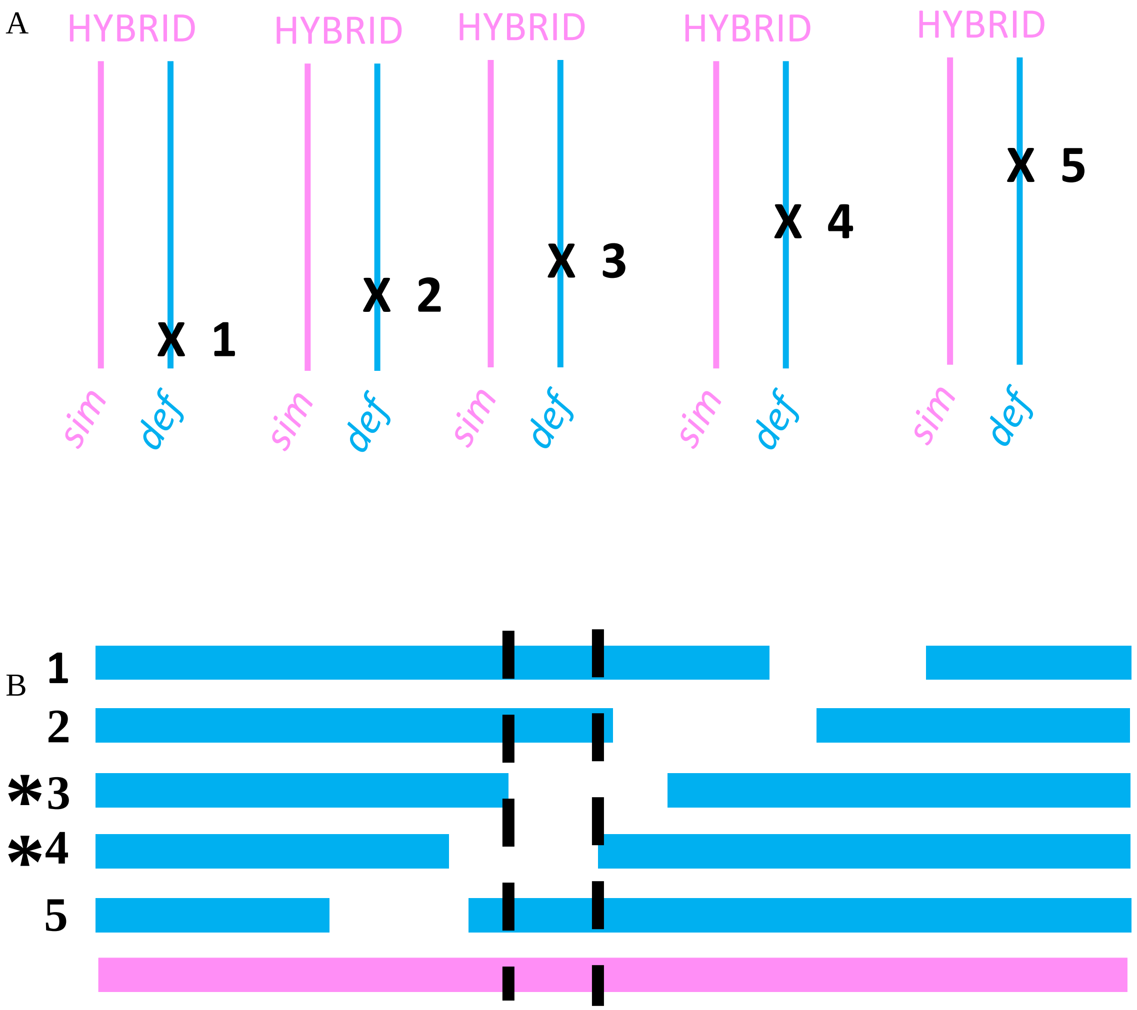


Figure 1.3. Principle of deficiency mapping. (A) A population is generated where individuals have a disruption or deletion, some of which will span overlapping regions of the genome (disruptions one through five indicated by an "X" on the *D. melanogaster* homolog, indicated in blue). (B) The disruptions (numbers corresponding to (A)) are lined up relative to the *D. simulans* homolog. Regions that overlap and show the behaviour (indicated by the dotted black lines) contain candidate loci contributing to behavioural isolation. (*) indicates disruption lines that show behaviour of interest.

behavioural isolation genes were within a chromosomal inversion and were not sex-linked.

1.5 *Mekk1*: previous general characterization

This particular study will focus on *Mekk1*, a gene that contributes to *D. simulans* rejection of *D. melanogaster* males (Laturney and Moehring 2012b). However, previous characterizations of *Mekk1* indicate pleiotropy, where *Mekk1* has functions outside of behaviour (Inoue et al. 2001). *Mekk1* codes for a Mitogen-Activated Protein Kinase Kinase Kinase (MAPKKK; MAP3K; MEKK1), and has been implicated in stress response in *D. melanogaster*. *Mekk1*-deficient flies (having a deletion spanning the kinase domain) cannot activate an effective stress response, specifically to chemical stress or heat shock (Gramates et al. 2017; Inoue et al. 2001). *Mekk1* is one of four homologous MAP3Ks in *Drosophila* that are involved in stress response (Zhuang et al. 2006).

Drosophila Mekk1 has homologs in other species. Interestingly, *Mekk1* is referred to by a different name in other species such as mice and humans: *Mekk4* (Ryabinina et al 2006). To avoid confusion in this thesis, *Mekk1* will only refer to *Drosophila Mekk1*, while *Mekk4* will be used in conjunction with a common species name for clarification. The mammalian *Mekk4* is involved in mediateing stress response as it acts upstream of the p38 and JNK pathway (Takekawa et al. 1997). The protein product of *Mekk1* also acts upstream of the JNK pathway in *Drosophila* (Ryabinina et al. 2006). With respect to courtship behaviour, it is possible that being exposed to a male of a different species results in perceived stress by the female, and this leads to activation of the p38 and/or JNK pathways and subsequent rejection of the male. Alternatively, or possibly in addition to the previous proposition, the protein product of *Mekk4* in mice is necessary for normal hearing development (Haque et al. 2016). The auditory system of *Drosophila* has been shown to be similar to the auditory system of vertebrates (Albert and Göpfert 2015). It is therefore possible that *Mekk1* also impacts hearing development in *Drosophila*. For example, *Mekk1* could be required for the female's auditory perception of the male's species-specific wing-song. As a result, the altered perception of the wing-song may influence whether the female will accept or reject the male.

Mekk1 is required for wild-type phosphorylation of p38 which regulates the transcription of genes associated with stress response, and a deletion to the *Mekk1* kinase domain impairs wild-type stress response (Inoue et al. 2001; Zhuang et al. 2006). Although there is some redundancy in the MAP3K functionality, Zhuang et al. (2006) showed that MEKK1 is necessary for wild-type stress response mediated by p38. *Mekk1* has also been shown to contribute to negative regulation of Hippo signaling, which contributes to organ size throughout development via apoptosis (Huang et al. 2016). Aside from Hippo signaling MEKK1 is also important in JNK-mediated apoptosis (Zhang et al. 2016). Developmentally, *Mekk1* is also important (and affiliated with p38) in determining apical-basal polarity as part of the junctional polarity complex regulating cytoskeletal elements (Schimizzi et al. 2016).

According to the Flybase annotation (Gramates et al. 2017), *Mekk1* is transcribed as two transcripts which are initiated at two different transcription start sites, with one longer transcript (RB) and one shorter transcript (RD). Graze et al. (2009) generated expression data that included *Mekk1*, however the transcripts were not distinguishable and data were presented under the gene name (rather than two separate transcripts). Aside from this paper, no comprehensive expression data for *Mekk1* has been presented other than Expressed Sequence Tag evidence (EST; Humbert et al. 1993). Both transcripts of *Mekk1* were supported by EST data in *D. melanogaster*. Interestingly, the longer transcript was also identified in *D. simulans* but there was no evidence found for the second shorter transcript. Two protein products have been identified in both species for the *Mekk1* gene (Inoue et al. 2001), and the smaller protein has a size consistent with the RD transcript.

1.6 *Mekk1*: previous behavioural characterization

Hybrid females with the *D. melanogaster Mekk1* allele disrupted, and one intact *D. simulans* allele, did not mate with *D. melanogaster* males (Fig. 1.1; Laturney and Moehring 2012b). Therefore, the *D. melanogaster* allele of *Mekk1* (acceptance) is dominant to the *D. simulans* allele of *Mekk1* (rejection). *Mekk1* is a candidate gene that is thought to contribute to behavioural isolation between *D. simulans* and *D. melanogaster*, located within an inversion within the right arm of the third chromosome (Gramates et al. 2017). Two *Mekk1* disruptions were originally used to characterize *Mekk1* as a candidate gene for behavioural isolation between *D. melanogaster* and *D. simulans* (Laturney and Moehring 2012b). Both disruptions are *P*-element insertions 10,872 base pairs (bp) in length, roughly the same length as *Mekk1* (Fig. 1.4; Bellen et al. 2004). One disruption was in the 5’ untranslated region upstream of the first putative transcript, RB. The other disruption was upstream of the 5’ UTR of the second putative transcript, RD, and in the first intron of the first transcript (Gramates et al. 2017). Neither disruption was in the coding region of *Mekk1*.

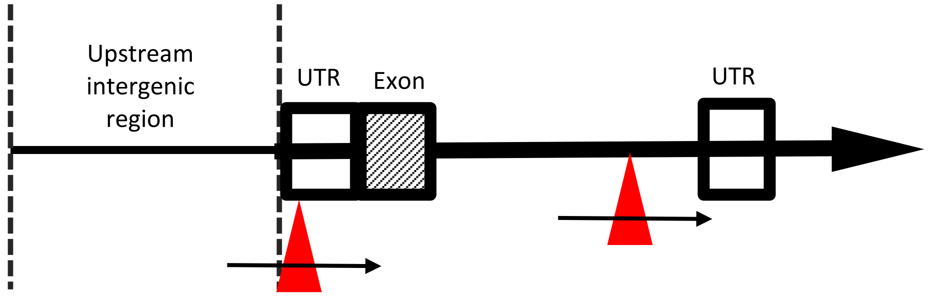


Figure 1.4. Locations of *P*-element insertions in previously characterized *Mekk1* deficiency strains. Open box: untranslated region (UTR); hashed box: exon (RB); thick solid line: intronic region; thin solid line: intergenic region; red triangles: location of the *P*-element insertions; black arrows: orientation of the *P*-element. Schematic only shows 5’ end of *Mekk1*. Locations of insertions determined based on information from Gramates et al. (2017).

The deficiency mapping was performed using three strains of *D. simulans*: FC, 199, and 216 (Moehring Lab, unpublished). Only one strain, *D. simulans* (FC), showed the behaviour associated with *Mekk1* (Fig. 1.1), while the two strains 199 and 216 of *D. simulans* did not. When hybrid females with the *D. melanogaster* disrupted *Mekk1* copy were created with *D. simulans* (199 or 216), these females showed the same behaviour as hybrids without the disruption, in other words they readily accepted *D. melanogaster* males. The phenotype changed if a disruption was present in *Mekk1* and the parental male was *D. simulans* (FC). Therefore, *D. simulans* (199 and 216) can be used to comparatively eliminate differences within *Mekk1* that may influence female behaviour in *D. simulans* (FC). Although the location and dominance of the species-specific alleles are known, further characterization of *Mekk1* is necessary to elucidate its expression and how it contributes to the female's behavioural phenotype.

It is evident that *Mekk1* is involved in stress response (Inoue et al. 2001), and is also developmentally important (Huang et al. 2016). Although the physiological stresses (including chemical and heat stress) are not the same as behavioural stress (being exposed to an individual of another species), it is possible that *Mekk1* contributes to female rejection behaviour in *D. simulans* through its characterized stress response pathway (Inoue et al. 2001). As *Mekk1* is involved in regulating Hippo signaling during development (Huang et al. 2016), it is also possible that *Mekk1* is differentially regulated between species and contributes to differences in organ growth. If these differences in organ growth influence adult behaviour in a way that affects mate discrimination, it is possible that the role of *Mekk1* during development also influences adult female rejection behaviour in *Mekk1*. While being involved in development (Huang et al. 2016) and stress response (Inoue et al. 2001) may explain mechanisms of how *Mekk1* contributes to adult behaviour, there is also high-throughput expression data indicating that *Mekk1* is highly expressed in the adult female ovary (Gramates et al. 2017). In rats, physiological feedback from organs other than the brain can influence behaviour (Levine and Mullins 1964). Expression of *Mekk1* at high levels in the ovary may contribute to female rejection behaviour in *D. simulans* (FC).

1.7 Hypotheses and Objectives

1.7.1 Overall Objective

The objectives of this study can be broken down into characterization and behavioural assays (Table 1.1). Both objectives contributed to the goal of understanding how *Mekk1* influences female behaviour, and how *Mekk1* contributes to reproductive isolation between two species of *Drosophila*. Objectives will be outlined, as well as hypotheses for each specific goal (sections 1.7.2-1.7.7).

1.7.2 Sequencing characterization

Originally, the *Mekk1* hybrid behavioural assays were carried out with two specific strains of *Drosophila*: *D. simulans* (FC) and *D. melanogaster* (BJS). These two strains were sequenced for *Mekk1*, and *Mekk1* was also sequenced in *D. simulans* (199 and 216). The sequencing data may yield insight into whether certain regions of the gene are diverged between *D. simulans* (FC) and the other three strains (*D. simulans* 199 and 216, and *D. melanogaster* BJS).

The sequencing analyses encompassed the gene region as a whole, focusing on specific regions from the 3’ end of the upstream gene (*CG7718* in *D. melanogaster* and *GD20153* in *D. simulans*; Gramates et al. 2017) to within the 3’ UTR of *Mekk1*. The coding regions were analyzed and compared to determine if there may be coding differences (which correspond to differences in amino acid sequence) that are responsible for the behaviour. Non-coding regions, including the promoter, were analyzed to determine if there is a regulatory element contributing to the behaviour.

1.7.3. Characterizing *Mekk1* transcript expression

It is possible that *Mekk1* transcript expression is regulated by something other than a cis-element characterized through sequencing (such as the promoter region). I used RT-PCR to measure relative overall transcript levels of *Mekk1*. It is also possible that even if there are large differences in a promoter region sequence that these differences do not contribute to a difference in gene expression levels. In a previous study by Graze et al. (2009) there were statistically insignificant differences between the pure species, and hybrid *Mekk1* expression when assessed by qPCR. The first reason *Mekk1* expression will be revisited is to examine if the trends of transcript regulation between species are consistent with Graze et al. (2009), specifically looking at the two strains of interest for this study. It is clear that the difference in *Mekk1* that contributes to female rejection behaviour in *D. simulans* is strain-specific (Moehring Lab, unpublished), and the study by Graze et al. (2009) did not use the same strains as those being characterized here.

A secondary issue is that Graze et al. (2009) did not examine each of the two *Mekk1* transcripts individually. The data were pooled and presented as one gene (*Mekk1*) rather than each of the transcripts (RB and RD/RC; Gramates et al. 2017). Examining each transcript individually may give a more informative idea of how *Mekk1* is regulated. The difference in behaviour may be related to overall *Mekk1* expression, or related to just one of the transcripts. In female pupae and adults, expression levels of *Mekk1* in the head

Table 1.1. Summary of objectives and hypotheses.

Overall objectives	Specific goals[1]	Hypotheses
To characterize *Mekk1*	Sequencing characterization (1.7.2)	If there are sequence differences unique to *D. simulans* (FC) then these unique sequences may contribute to *D. simulans* female rejection behaviour
	Characterizing *Mekk1* transcript expression (1.7.3)	If there are differences in the amount of *Mekk1* transcript between *D. simulans* and *D. melanogaster*, then these transcript level differences may contribute to *D. simulans* (FC) female rejection behaviour
	Tissue-specific characterization of *Mekk1* expression (1.7.4)	If there is a tissue-specific pattern associated with *Mekk1*, then this may indicate potential tissues through which *Mekk1* is mediating female rejection behaviour
To examine behavioural phenotypes associated with *Mekk1*	Determining tissue-specificity of *Mekk1*-associated behaviour (1.7.5)	If *Mekk1* can be rescued successfully in a tissue-specific manner, then *Mekk1* can be classified as a behavioural isolation gene
	Examining contribution of kinase domain of *Mekk1* to behaviour (1.7.6)	If the kinase domain in the *Mekk1* protein product is disrupted and this causes a different behavioural phenotype from a non-disrupted individual, then the kinase domain in *Mekk1* is responsible for mediating female rejection behaviour

[1] The number in brackets beneath each specific goal corresponds to a section of the introductory chapter.

and body were quantified. Examining the female adult lifecycle stage is relevant as that is when the behaviour is observed. The pupal lifecycle stage involves development of various regions of the brain, as well as other systems including the reproductive tract (Belote and Bakert 1987; Kimura et al. 2005). Levels of expression during development may be important in influencing brain development, and therefore, the development of the ability to receive signals such as those from male courtship. One particular region of the brain, the mushroom body, develops during pupation (Technau and Heisenberg 1982). The mushroom body has previously been linked to decision-making in adult *Drosophila* (Zhang et al. 2007).

The third issue with the Graze et al. (2009) paper, which ties into the second issue, is that they only examined the heads of adult *Drosophila*. Graze et al. (2009) did not examine adult *Drosophila* bodies, nor did they examine pupa Although it would be a logical assumption that the head (and specifically the brain) may contribute to any behaviour, previous evidence has shown that physiological feedback from other organs in the body can influence behaviour, supported in rats by Levine and Mullins (1964). It is possible that expression in the body of the fly may contribute to rejection behaviour in females. It is also possible that expression of *Mekk1* may be important in stages of development (Technau and Heisenberg 1982) and/or certain tissue types in the head and specifically the brain (eg: the mushroom body; Zhang et al. 2007). Two lifecycle stages and two tissue types for each *Mekk1* transcript were examined to shed light on how *Mekk1* may influence female rejection behaviour in *D. simulans* (FC) in a more informative manner than Graze et al. (2009). If one transcript is differentially regulated between species, that alone does not definitively explain how *Mekk1* mediates behaviour, but it may provide further hypotheses to test pertaining to behaviour and shed light on what tissues may mediate female rejection behaviour.

1.7.4. Tissue-specific expression characterization

Aside from quantifying expression levels using molecular methods, there are other ways to observe tissue-specific gene expression. Two strains of *D. melanogaster* with a GFP construct within the *Mekk1* gene region were observed to determine if there is GFP expression (Gramates et al. 2017; Venken et al. 2011). A third strain containing an enhancer trap construct (in combination with UAS-GFP) was observed to determine if there is GFP expression (Duffy 2002), and also compared to the first two strains. I compared the fluorescence pattern for each of the three strains to that of a wild-type *Drosophila*. This allowed me to determine if the GFP expression pattern differs from a wild-type fly with autofluorescence (tissue that naturally fluoresces; Meerwaldt et al. 2005). I compared the amount of fluorescence for tissues that have previously characterized *Mekk1* expression levels (specifically the ovaries; Gramates et al. 2017). If the ovaries of the strains with GFP constructs have higher levels of fluorescence compared to a wild-type fly, then it is likely that the strain is expressing GFP in a *Mekk1*-specific manner. If that is the case and GFP is expressed in a *Mekk1*-specific manner, then the brains of the *Drosophila* will be observed to determine if there is a pattern of interest present.

Examining *Mekk1*-specific expression patterns may inform gathering further expression-level data and the tissue-specificity of subsequent rescue experiments. Expression throughout the entire brain may result in a detectable difference in expression levels of *Mekk1* between species. However, if *Mekk1* is only expressed in a subset of cells or a region of the brain, this could mean that the difference in expression between species may lie in a small number of cells between species. If that is the case, it is possible that a method such as RT-PCR may not be sensitive enough to detect such a small difference in relative expression.

1.7.5. Determining tissue-specificity of *Mekk1*-associated behaviour

There are typically two parts to characterizing a gene as being responsible for a phenotype: knockout and rescue. As it pertains to *Mekk1*, the “knockout” portion was completed through deficiency- and fine-mapping (Laturney and Moehring 2012b). The rescue portion will be attempted through the use of the GAL4/UAS system (Fischer et al. 1988). There is a UAS construct linked to a *Mekk1* transcript (Bischof et al. 2013) that, when used in combination with a *P*-element disruption, will be transcribed when GAL4 is present to rescue *Mekk1* expression in a tissue-specific manner.

The GAL4 sequence will be under the control of two different tissue-specific promoters, being expressed in a ubiquitous manner as well as the brain (Gramates et al. 2017). The experiment served two purposes: 1) to confirm that *Mekk1* is a behavioural isolation gene, and 2) to inform which tissues mediate female rejection behaviour by *Mekk1*. If the behavioural phenotype is successfully rescued to a wild-type level using the GAL4/UAS system, then it will not only confirm that *Mekk1* is a behavioural isolation gene, but also shed light on which tissues within the female are important in *Mekk1* rejection behaviour.

1.7.6. Determining if the known function of *Mekk1* contributes to behaviour

The protein product of the *Mekk1* gene is a kinase. The kinase portion of the protein is the only characterized functional domain. The initial disruptions that were used to characterize *Mekk1* as a behavioural isolation gene were in non-coding regions. There is a strain of *Drosophila* that has a deletion spanning the kinase domain, leading the protein to have decreased kinase function. This deletion strain was used to directly determine if the behaviour of interest is mediated through the kinase function of MEKK1. If the results indicate that the behaviour of interest is not mediated through the kinase function of *Mekk1,* it is possible that the behavioural differences are mediated through gene expression rather than the protein functionality of MEKK1.

2: Material and Methods

2.1 *Drosophila* stocks

The majority of *Drosophila* strains were obtained from three different stock centers: Bloomington *Drosophila* Stock Center, Kyoto Stock Center, and the *Drosophila* Species Stock Center (Table 2.1). The strain UAS-Mekk1 was received from the founders of the FlyORF Zürich ORFeome project. Wild-type strains were either pure isofemale *D. melanogaster* (BJS, London ON; from Dr. B. Sinclair) or *D. simulans* (Florida City, USA; from Dr. J. Coyne). Two alternative strains of *D. simulans* were obtained from the *Drosophila* Species Stock Center: strains 216 (stock number: 14021-0251.216) and 199 (stock number: 14021-0251.199). All fly strains and crosses were maintained on standard cornmeal media in 30mL plastic food vials and housed in incubators with a standard 14:10 hour light-dark cycle, 24ºC and 75% relative humidity.

2.2 *Drosophila* crosses

Anaesthesia by CO_2 was used to collect virgin flies 0-8 hours after eclosion, and the virgins were then separated by sex and aged 5-7 days to sexual maturity. Upon reaching maturity these flies were then placed into food vials and paired with individuals of a specific genotype of the opposite sex. Specific crossing schemes and example crossing schemes for behavioural assays, molecular work and for visualization studies are given in Fig. 2.1-2.4. Unless otherwise stated it doesn’t matter which genotype is male or female, so whatever sex was most readily available was used for the cross.

Table 2.1 List of *Drosophila* strains with source and genotype.

Species[1]	Stock[2]	Strain[3]	*GFP*[4]	Genotype	From[5]	Number[6]
D. mel	BJS	Wt	No	Wt	N/A	1
D. sim	FC	Wt	No	Wt	N/A	2
D. sim	199	Wt	No	Wt	N/A	
D. sim	216	Wt	No	Wt	N/A	
D. mel	19991	Dis	No	y^{1} w^{67c23}; P{EPgy2}Mekk1EY02276	BDSC	3
D. mel	41489	Dis	Yes	y^{1} w*; Mi{MIC}Mekk1MI05281	BDSC	
D. mel	13748	Dis	No	y^{1} w^{67}c^{23}; ry^{506} P{SUPor-P}Mekk1KG02510	BDSC	
D. mel	60227	Dis	Yes	y^{1} w*; Mi{PT-GFSTF.2}Mekk1MI05281-GFSTF.2/TM6C, Sb^{1} Tb^{+}	BDSC	
D. mel	109805	Del	No	w*; Mekk1Ur^{36} / TM3, Sb^{1} Ser^{1}	Kyoto	4
D. mel	UAS	UAS	No	y^{-}w^{-} M{eGFP.vas-int.Dm}ZH-2A; +; M{3x-pUASMekk1-HA.attB}ZH- 86Fb	Zurich	5
D. mel	5137	UAS	Yes	y^{1} w*; P{UAS-mCD8::GFP.L}LL5, P{UAS-mCD8::GFP.L}2	BDSC	6
D. mel	25374	Gal4	No	y^{1} w*; P{Act5C-GAL4-w}E^{1}/CyO	BDSC	7
D. mel	55851	Gal4	No	w*; P{GAL4-da.G32}2; MKRS/TM6B, Tb^{1}	BDSC	8
D. mel	8765	Gal4	No	P{GAL4-elav.L}2/CyO	BDSC	9
D. mel	105282	Gal4$_{M}$	No	y* w*; P{w+mW.hs=GawB}Mekk1NP6602 / TM6, P{w-=UAS-lacZ.UW23-1}UW23-1	Kyoto	10
D. mel	3703	Bal	No	w^{1118}/Dp(1;Y)y^{+}; CyO/nub^{1} b^{1} sna^{Sco} lt^{1} stw^{3}; MKRS/TM6B, Tb^{1}	BDSC	11

[1] *D. mel = D. melanogaster*, *D .sim = D. simulans*

[2] Stock numbers correspond to the original order number from the corresponding stock center

[3] Strain: Type of transgenic construct in each strain; Wt=wild type, Dis=disruption; Del=deletion of the kinase region; UAS=upstream activating sequence; Gal4=GAL4 (to be paired with UAS strains; subscript M indicates GAL4 within *Mekk1* gene region); Bal=balancer stock, subscript M indicates GAL4 within *Mekk1*

[4] GFP category indicates if strain contains a *GFP* construct.

[5] N/A = see section 2.1; BDSC=Bloomington Drosophila Stock Center; Kyoto=Kyoto Stock Center (Japan); Zurich=FlyORF Zurich ORFeome Project founders

[6] Number Corresponds to strains found in crossing schemes (Fig. 2.1-2.4). .

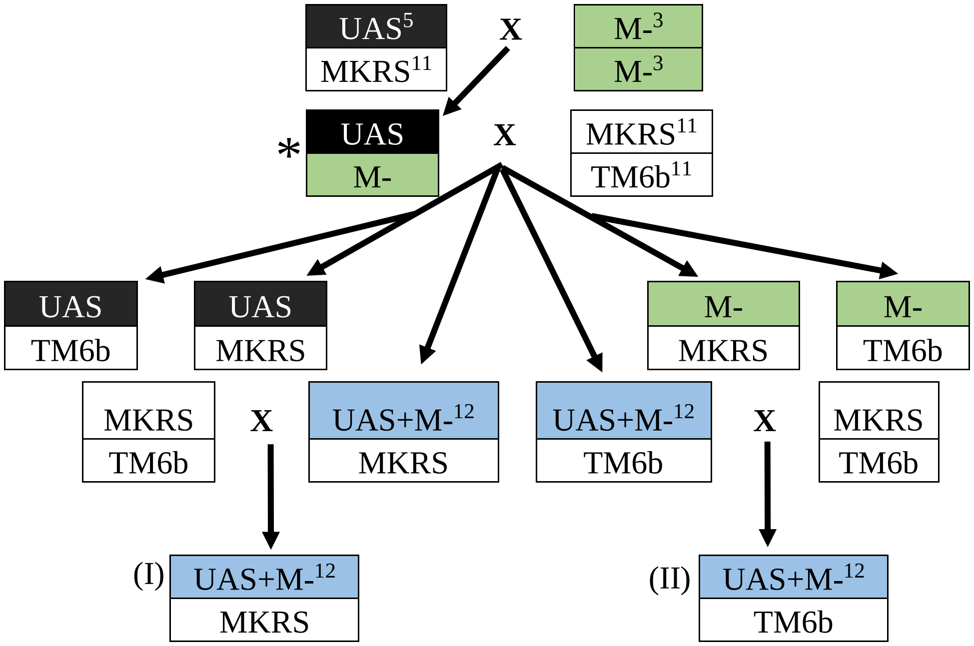


Figure 2.1. Crossing scheme (1) illustrating how a *P*-element disruption and a UAS element linked to *Mekk1* were recombined onto a single homologous chromosome. M-: *Mekk1* disruption (strain 19991), UAS=*UAS-Mekk1* (RB; strain UAS). (*) indicates a female individual. Total generations shown: 4. Superscript numbers 1-11 are defined in Table 2.1, while #12 shows recombined homolog (generated in this scheme from two separate genotypes). Originally the recombined homolog (I & II) was maintained over a balancer (MKRS; I), however if one of the constructs recombined out, an alternative would be to maintain the chromosome over a different balancer (TM6b; II). Single sets of stacked boxes indicate homologs of the third chromosome in one individual. Colors of the boxes and lettering indicate different types of homologs or transgenic constructs. Uncolored boxes with black print indicate balancer chromosomes. Generations are roughly three weeks. Arrows indicate offspring. “+” within a single box indicate constructs on the same homolog. “X” indicates crossing two individuals together. Genotype of second chromosome of (I) and (II): *CyO*/*nub1 b1 snaSco lt1 stw3*.

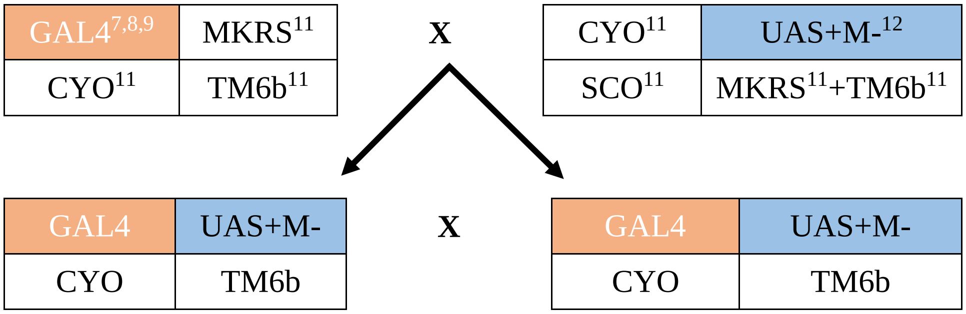


Figure 2.2. Crossing scheme (2) shows how stocks were generated for later use in the behavioural rescue experiments. UAS/M-: stock generated through crossing scheme (1); several GAL4 strains were used: 55851, 8765, and 25374. This crossing scheme was repeated separately for each GAL4 driver. There are two possible genotypes for the UAS/M- individual, which is due to the alternative approach explained in Figure 2.1. Total generations shown: 2. Two sets of stacked boxes illustrate homologs of the second and third chromosomes in single individuals. Colors as in Figure 2.1.

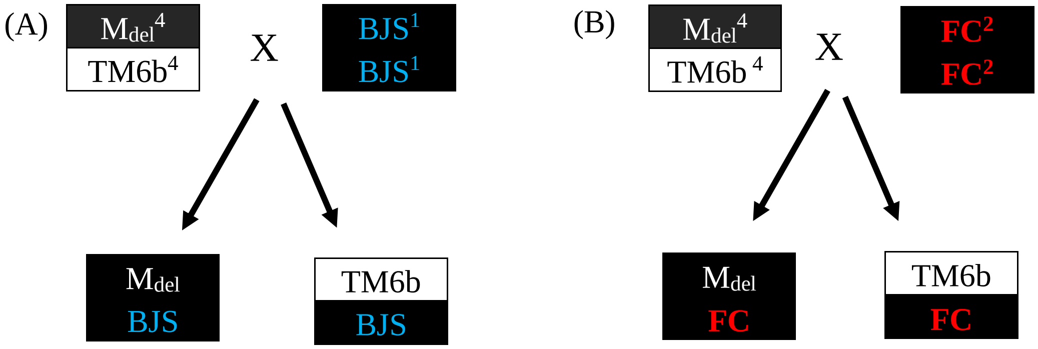


Figure 2.3. Crossing scheme (3) showing inter- and intra-species crosses for *Mekk1* behavioural assay experiments. (A) shows an intraspecies cross (all offspring are *D. melanogaster*), while (B) shows an interspecies cross (all offspring are hybrids of *D. melanogaster* and *D. simulans*). M_{del} homolog contains a deletion in the functional domain of the *Mekk1* protein product. The four genotypes in the second generation correspond to Figure 1.4. Total generations illustrated: 2. Colors as in Figure 2.1.

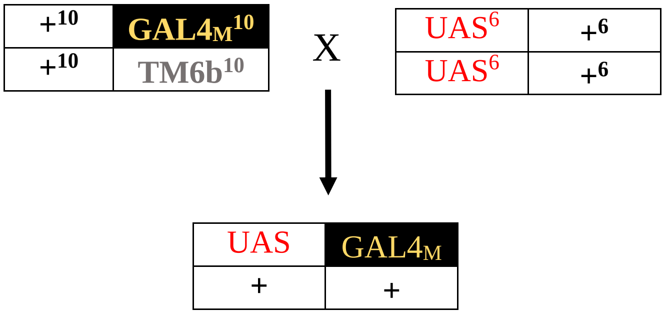


Figure 2.4. Crossing scheme (4) showing how various genotypes were combined and used to visualize tissue-specific expression of *Mekk1*. Cross between 105282 (containing an enhancer trap with GAL4) and 5137 (containing UAS-*GFP*), resulting in an enhancer trap within *Mekk1* activating a *UAS-GFP* (expressing GFP in the pattern of *Mekk1*). Description of boxes in Figure 2.2 and colors as in Figure 2.1.

2.3 DNA isolation

DNA extractions were performed using Squishing Buffer (100µL 0.5M EDTA, 250µL 5M NaCl, 500µL 1M TrisHCl pH 8.2, 49.15mL ddH_2O) and proteinase K (ProK; 0.02g proteinase K from *Tritrachium album* by Sigma-Aldrich, prepared in 0.6mL 10mM Tris pH 7.6, 0.4mL glycerol); modified from a protocol by Gloor et al. (1991). An anaesthetized single fly was placed in a solution containing 0.5µL ProK and 49.5µL Squishing Buffer, then homogenized using a pestle and incubated at 37°C for 30 minutes. To inactivate the ProK the samples were heated to 95°C for two minutes.

2.4 Primer design and testing

Initially primers were designed based on the *D. melanogaster* sequence (FlyBase; Gramates et al. 2017) and tested on DNA from both *D. melanogaster* (BJS) and *D. simulans* (FC). Primers were tested in this manner both for sequencing with genomic DNA (gDNA) and for RT-PCR using complimentary DNA (cDNA). As data became available through sequencing, primers were designed based on the new sequence in combination with the reference sequence. Once the primers successfully amplified *D. simulans* (FC) sequences they were also used to amplify DNA from the other two strains of *D. simulans* (199 and 216). Sequencing primers and their respective locations are listed in Figure 2.5 and Table 2.2 for *D. melanogaster* (BJS) and in Figure 2.6 and Table 2.3 for *D. simulans* (FC). Primers for RT-PCR are listed in Table 2.4.

Prior to sequencing, primers were tested pairwise to ensure that they bind to unique sites and amplified PCR fragments of a predicted size. The initial primer testing used *Paq*5000 DNA polymerase (Agilent, London ON) and a modified touchdown PCR protocol to decrease non-specific amplification as follows: one cycle of three minutes at

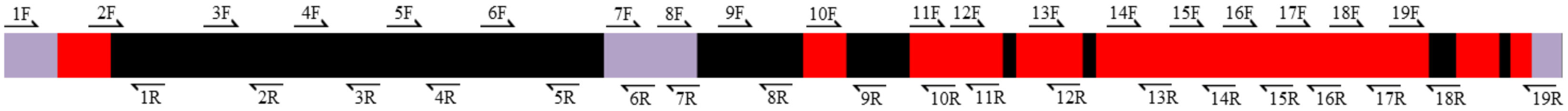


Figure 2.5. Relative position of *D. melanogaster* primers for sequencing. Numbers correspond to Primer Set number in Table 2.2, F: forward primer, R: reverse primer. Purple: UTRs; red: exons; black: intronic sequences.

Table 2.2 *D. melanogaster* primers for sequencing.

Primer Set[1]	Forward primer (5’ → 3’)	Reverse primer (5’ → 3’)
1	CGACTTGAGTGGCTCGAAGA	GTGAAATGCACAGCGAGAGA
2	GAGGATGTTGCAAGGTGAGC	GAAGGGTGCTCTCTTGTAAG
3	GCACTGCAAAAATTGGATCA	GCGAGCAGCAGACACATAAA
4	TGCGACAAACATCCATTTTC	AACATACGAAAGCCCCAAAA
5	CGCTCTTGCAACATTCGTATT	GACCAGCGGTGCGTATACTT
6	ACTCATCGAATTTCCAATGC	ATGGGTTATGCACTATCTCC
7	TGGCGAATAGATGGGGAATA	CAGCGAGACGAGAGAGAACC
8	TGGCTCGCGTCTAAAGTTCT	AACTCTGGGGAGGAGGACAT
9	ATTAGGGCATAATCGCAGG	AGATGGCACTATCAGATGC
10	TCCAGTGCAGATGACTCG	AACGAGTTCTTCATCTCACG
11	CCCTGCTCAACGAGATGTTT	TAATGTGGGTGAGCAGATCG
12	AAGCTCGGTAGCGTAGATCG	TCTACACCCGATGCTGATGA
13	AGCATCGGGTGTAGAGAACG	GAATGTACGACGGCAGGTTT
14	GAGTCGGATGCAGAGTCCAT	CTCCTCGTTGTCGTCGGTAT
15	ATGCTACGGTGCGAAAGACT	CATCCAAATCTCGCATGTTG
16	GCGAGATTTGGATGAGCAAG	TGTCAACGGGACTGTACGAA
17	AATACGCCGCCACTACAGTC	CAGTCAGCTCCACCAGTGAC
18	ACGGCGGTGAACAATAACAC	CCCCATACATCGAATTGCTT
19	GGTGAGTGCTTCGCATTGTA	TAAAACCAGACGAGGGCATT

[1] Primer set number corresponds to numbers in Figure 2.5.

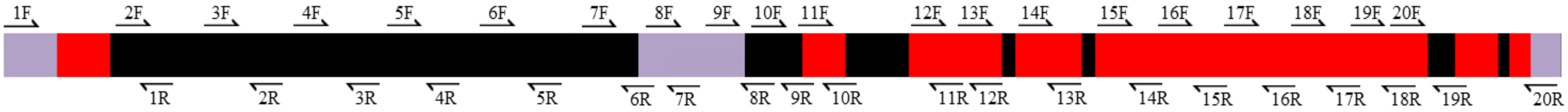


Figure 2.6. Relative position of *D. simulans* primers for sequencing. Numbers correspond to Primer Set number in Table 2.3, F: forward primer, R: reverse primer. Purple: UTRs; red: exons; black: intronic sequences.

Table 2.3 *D. simulans* primers for sequencing.

Primer Set[1]	**Forward primer (5'→ 3')**	**Reverse primer (5'→3')**
1	TAGGACAGGCTGCTTACC	AACAACAAGATAACGTGG
2	CATTTCCTCTCCCCCTCTTT	AGTCCACCGTCAGCGTTATC
3	TTACAACAGAGCGCCCTTC	AAGTCAAATGAGAACACTGC
4	TTAGCCGAGATTTATGTG	TACACGCCATAATTATGC
5	CGCTCTTGCAACATTCGTATT	GACCAGCGGTGCGTATACTT
6	TCCGCATATTGCACACG	ACCAGCGGTGCGTATACTTC
7	TGCCTCATACATAAATGG	ATAACGCGCAGCACTTGTCG
8	TGGCGAATAGATGGGGAATA	CAGCGAGACGAGAGAGAACC
9	TGGCTCGCGTCTAAAGTTCT	AACTCTGGGGAGGAGGACAT
10	ATGTCCTCCTCCCCAGAGTT	CTGGTACGTGGCGGAGTAGT
11	AGCGAAACAATTAGGACAAGC	AAACATTTCGTTGAGCAGG
12	ACCAGGACGGCCAGATCC	ATCGGAATGAATGCCACTATCC
14	AAGCTCGGTAGCGTAGATCG	TCTACACCCGATGCTGATGA
15	CAAGCAGAACAAGTTGAAGC	TACAAAGGTAGGAATGTACG
16	GAGTCGGATGCAGAGTCCAT	CTCCTCGTTGTCGTCGGTAT
17	ATGCTACGGTGCGAAAGACT	CATCCAAATCTCGCATGTTG
18	GCGAGATTTGGATGAGCAAG	TGTCAACGGGACTGTACGAA
19	ACACCAACACGGACTCCAAT	GCACCAGGTTTTTGTGCTTT
20	ACGGCGGTGAACAATAACAC	CCCCATACATCGAATTGCTT
21	GGTGAGTGCTTCGCATTGTA	TAAAACCAGACGAGGGCATT

[1] Primer set number corresponds to numbers in Figure 2.6.

Table 2.4 Primers for semi-quantitative RT-PCR.

Transcript	Gene[1]	F Sequence (5’ → 3’)[2]	R Sequence (5’ → 3’)[3]
RB	*Mekk1*	CATTCCGCCCATACGATTAC	CTGGTACGTGGCGGAGTAGT
RD	*Mekk1*	AATGCGTCGGAAGAAAGTTG	GTAGTGCCGAAGGCCTCATA
RPL	*Rpl32*	GACGCTTCAAGGGACAGTATCTG	AAACGCGGTTCTGCATGAG

[1] *Mekk1* is the gene of interest with two transcripts, *Rpl32* is the control gene to which *Mekk1* expression was normalized.
[2] Forward primer sequence
[3] Reverse primer sequence

95°C; two cycles of 30 seconds at 95°C, 30 seconds at 59°C, 30 seconds at 72°C; followed by two cycles of 30 seconds where the annealing temperature was lowered to 57°C, followed by 30 cycles where the annealing temperature was lowered to 55°C, and a final holding step at 4°C.

2.5 DNA preparation for sequencing

To generate template DNA for sequencing, standard PCR was performed using *Taq* DNA polymerase (Life Technologies) as per manufacturers instructions. A 1% agarose gel was used to size separate the PCR amplicons. The DNA was gel extracted using the GenepHlow Gel/PCR Kit from Geneaid according to manufacturer's guidelines. To determine the integrity and concentration of the extracted DNA sample, a small aliquot (5µL) was size-separated on a 1% agarose gel. Samples which showed off-target amplification after gel extraction were electrophoresed for a second time and gel purification was repeated. DNA sequencing was performed from both ends of each amplicon (forward and reverse) at the Robarts Research Institute, London (ON) using MiSeq (Illumina).

2.6 Sequence analysis

The consensus sequences for *D. melanogaster* (BJS) and *D. simulans* (FC) were generated with MUSCLE (Edgar 2004) implemented through Geneious (R9) using default settings (Kearse et al. 2012). The sequence reads were compiled and mapped to the appropriate reference sequence from FlyBase (Gramates et al. 2017). The reads were aligned by the program based on their sequence similarity to the reference sequence. Once compiled, the chromatograms were examined to ensure that the consensus sequence generated was accurate. If the regions of the consensus sequence were generated from one read or less, the contig was re-examined for quality. Any region with more than three non-sensical ‘N’ nucleotides per 100 bp (3 N / 100 bp) was left unedited in the consensus sequence. The consensus sequence was exported and run through the EMBOSS Needle program (EMBL-EBI) which uses a Needleman-Wünsch algorithm to generate a measurement of percent identity. Percent identity was used as a measure of similarity between the consensus sequences from both *D. melanogaster* (BJS) and *D. simulans* (FC).

2.7 Tissue preparation for RNA isolation

For RT-PCR RNA was isolated from four different tissue types of females: adult heads, adult bodies, pupal heads and pupal bodies. Adults were sorted within 0-8 hours of eclosion to ensure virginity and aged 5-7 days prior to tissue preparation. Regardless of lifecycle stage or tissue type, five individuals were used per sample, and each RT-PCR experiment was repeated with three independent samples. Third instar larvae were sexed twice over the course of a day and tissues were isolated on day 2-3 of the pupal life cycle stage. Each of the four sample types were collected in a Falcon tube and flash frozen for two minutes in liquid nitrogen. The adult samples were vortexed to separate heads from bodies, and the heads and bodies were separated to be processed separately. The vortexed samples were then flash frozen again for two minutes in liquid nitrogen. To collect heads and bodies of pupae, the individual pupae were cut with a razor blade between their thoracic and head segments as these can be visualized from day 2-3. The two segments were then sorted into separate tubes and flash frozen a second time. All of the flash frozen samples were kept in liquid nitrogen until the RNA isolation procedure began.

2.8 RNA isolation

All Eppendorf tubes used in this procedure were RNase-Free, and the procedures were performed at an RNase-Free work bench. Tissue samples were removed from the liquid nitrogen one at a time, and homogenized using a motorized pestle for one minute, or until all the tissue appeared to be homogenized. Upon homogenization the samples were placed on ice until the remaining samples were homogenized. RNA was isolated according to a protocol by Bertucci and Noor (2001; modifications in Appendix A).

2.9 RNA quantification

For RT-PCR, the resuspended RNA sample was analyzed for concentration and purity using a nanophotometer (IMPLEN) according to the manufacturer's specifications. The concentration was determined in ng/µL using lid 50, and the purity was determined by examining A260/280 and A260/230. If both of the absorbance ratios were greater than 1.8 samples were considered pure and used for cDNA synthesis.

2.10 cDNA synthesis

Synthesis of cDNA was performed using an MMLV-reverse transcriptase (Life Tech) kit according to the manufacturer's specifications using oligo(dT)18 primer (Thermoscientific). Once RNA was quantified, 1000µg RNA was used in the cDNA reaction (20µL total reaction volume). The RT-PCR was semi-quantitative as the expression of the transcripts of interest (*Mekk1*) and all values were normalized to a control gene (*Rpl32*) to reduce any bias introduced by variability of RNA quantification.

2.11 RT-PCR

Once cDNA was generated, PCR was performed using *Taq* DNA Polymerase (Invitrogen). The reactions were amplified with a lower number of cycles compared to PCR performed for sequencing as reactions needed to be stopped in the exponential phase to allow for visualization of differences in expression levels. As the control gene, *Rpl32* (Ponton et al. 2011), was expressed at a higher level compared to the *Mekk1* transcripts the PCR for the control gene was analyzed with a fewer number of cycles compared to the *Mekk1* transcripts (primer sequences in Table 2.5). The same thermal cycler PCR protocol was used as described previously (section 2.4), with cycle number modifications: 32 cycles for both *Mekk1* transcripts (RB and RD/RC), and 20 cycles for *Rpl32* (control). Any decrease in cycle number was performed at the end of the reaction (by shortening the overall protocol as opposed to removing cycles from the beginning). A modified touchdown protocol was used in an effort to ensure a lack of nonspecific amplification, despite the fact that the primers were tested for off-target effects and found to have none as long as there was no genomic contamination present in the sample. Biological replicates were performed using separate samples of five individuals each. RT-PCR reactions were performed on each biological replicate in an effort to produce more robust data.

The PCR fragments were size-separated on a 2% agarose gel. The samples were loaded into the gel with Xylene Cyanol (Sigma-Aldrich) for 45 minutes at 100V. The gel was imaged using a Gel Doc™ XR+ Gel Documentation System (BioRad) and the images were analyzed using semi-quantitative densitometry in QuantityOne (BioRad). A densitometry measurement was taken for each band of interest (each *Mekk1* transcript and the control, *Rpl32*), and the density of each *Mekk1* transcript was normalized to the control to calculate semi-quantitative relative expression. Once these measurements were obtained, the relative expression levels were compared between species within each transcript for each sample type using a two-way Analysis of Variance (ANOVA) in Microsoft Excel (2013).

2.12 *Mekk1* visualization

Various transgenic strains were used in an attempt to visualize tissue-specific *Mekk1* expression. The genotypes used to examine tissue-specific expression of *Mekk1* were as follows: (*P{UAS-mCD8::GFP.L}LL5;P{w*$^{+mW.hs}$*=GawB}Mekk1*^{NP6602}), (*Mi{PT-GFSTF.1}CG9492*$^{\mathbf{MI09168\text{-}GFSTF.1}}$*/TM3***,** *Sb*^{1} *Ser*^{1}), and (*Mi{MIC}Mekk1*$^{\mathbf{MI05281}}$). The enhancer trap (*GawB*) was crossed with a UAS-*GFP* (Table 2.1). The specimens were then examined using a FITC filter cube in an Upright Nikon Eclipse C*i* Fluorescent Microscope under 10x magnification. Images were taken using a Nikon DS-Fi1 camera and analyzed using ImageJ software (Schneider et al. 2012).

2.13 Generating the genotypes for the behavioural rescue experiment

Two GAL4 strains (one expressed highly in many tissues and one restricted to the brain) were used to rescue *Mekk1* expression in a tissue-specific manner with a UAS linked to the coding sequence of transcript RB (*Drosophila* strains outlined in Table 2.1; Bischof et al. 2013). The UAS and *P*-element insertion (*P{EPgy2}Mekk1*$^{\mathbf{EY02276}}$, which is inside the 5’ untranslated region of RB) were recombined onto a single homologous chromosome and maintained over a balancer to allow the rescue experiments to be performed in hybrid individuals (*D. melanogaster* x *D. simulans*; Fig. 2.1). Two balancers were attempted and the final experiment was performed using TM6b as the third chromosome balancer.

Once the *D. melanogaster* stock genotypes were established according to Figure 2.1, they were crossed to *D. simulans* (FC) to create hybrids (Fig. 2.3). The four genotypes assayed for behaviour included: 1) *GAL4*/*D. sim.*; *UAS-Mekk1*,*M-*/*D. sim.* 2) *sco*/*D. sim.*; *UAS-Mekk1*,*M-*/*D. sim.* 3) *GAL4*/*D. sim.*; *TM6B, Tb*^{1}/*D. sim.* 4) *D. mel.*/*D. sim*; *M-*/*D. sim.* (outlined in Fig. 2.3). The GAL4 driver in the four test genotypes depends on the tissue-specificity of the rescue, and a species abbreviation (either *D. mel.* or *D. sim.*) indicates a wild-type chromosome from either *D. melanogaster* or *D. simulans,* respectively. The fourth genotype described would be created by crossing the *Mekk1* disruption stock (19991) directly to *D. simulans*. The main genotype of interest is the first genotype described (*GAL4*/*D. sim*; *UAS-Mekk1*,*M-*/*D. sim*) which is directly compared to the fourth genotype described (*D. mel*/*D. sim*; *M-*/*D. sim*) to determine if their behavioural phenotypes are different, and if the GAL4 is rescuing *Mekk1* in a manner relevant to female receptivity to a heterospecific male. The second and third genotypes described (*sco*/*D. sim*; *UAS-Mekk1*,*M-*/*D. sim* and *GAL4*/*D. sim*; *TM6B, Tb1*/*D. sim*) are controls for transgenic GAL4 and *UAS* elements.

2.14 Genotyping

To have the *UAS-Mekk1* element and the *Mekk1 P*-element disruption on the same chromosome it was necessary to screen for recombinant chromosomes. Females with one element (*UAS-Mekk1* element and *Mekk1 P*-element) on either chromosome were mated to a male with balancers (stock 3703) on each of his third chromosomes (mating 1). The progeny of this mating were then mated to individuals from the 3703 stock in order to maintain any potentially recombinant chromosomes (mating 2). Once the flies had been maintained in a vial together for a week, the parent in mating 2 that was generated via mating 1 was genotyped for the two elements of interest. PCR was run to the same specifications as in section 2.4, except the reaction was a multiplex with two sets of primers: one for the *UAS-Mekk1* element (F: GATAGCGATTGCAGCTCTCC; R: GAACGCAAACAGAAGGTGGT), and one for the *Mekk1 P*-element (F: TTTGGGAGTTTTCACCAAGG; R: TAAGGTGGTCCCGTCGATAG). The products were run on a 3% gel to achieve better separation compared to section 2.4.

2.15 Mating assays

Mating assays were performed at a consistent time throughout the day within each experiment, specifically within 1-2 hours of lights on. All four fly genotypes (for both the rescue and deficiency mapping) were assayed at the same time. Pairs of flies (one male and one female) were aspirated into standard food vials (section 2.1) and observed for one hour for courtship and mating. The males used in these assays had sperm with GFP tags. After one hour, assays containing *D. melanogaster* female flies were discarded, and hybrid female flies (with their male counterparts) were placed into an incubator for 23 hours. Once the 24 hour assay period was complete, the flies were frozen (-20ºC) for no more than two months. Females from each assay that were known to have mated were kept for each set of dissections as a positive control of fluorescence. The female reproductive tracts were dissected and analyzed using fluorescence microscopy (same as described in section 2.12) to determine if the female mated during the assay, which would indicate that she accepted the male she was paired with. Once each female had been scored for presence/absence of GFP-tagged sperm in the reproductive tract, pairwise comparisons were made between genotypes using a student t-test.

3: Results

3.1 Comparative sequence analysis of *Mekk1* gene region

A previous study examined the genetic distance between *D. simulans* and *D. melanogaster* strains and found that there is variation between these species and between different strains of the same species (González et al., 1982). Therefore, it was important to sequence *Mekk1* in the strains initially used for the experiments that characterized *Mekk1* as a behavioural isolation gene (Moehring Lab, unpublished). When *Mekk1* was characterized as a candidate behavioural isolation gene in hybrid *Drosophila* females, *Mekk1* was disrupted in the *D. melanogaster* half of the genome and the *D. simulans* copy of *Mekk1* was intact. Three different strains of hybrid females with a wild-type *D. simulans* copy of *Mekk1* were originally tested (FC, 199, and 216; Moehring Lab, unpublished). Only one strain (*D. simulans* FC) showed *D. simulans* rejection behaviour associated with *Mekk1*. While the other two *D. simulans* strains (199 and 216) hybrid females continued to accept *D. melanogaster* males, indicating that in strains 199 and 216, *Mekk1* does not contribute to *D. simulans* female rejection of males from a different species. To explore the relationship between sequence variation and variation in female rejection behaviour, I sequenced one strain of *D. melanogaster* (BJS) and three strains of *D. simulans* (Fc, 199, and 216) for *Mekk1* from the upstream intergenic region to the 3’ UTR.

The genes were sequenced and aligned using a paired-end method and a species-specific reference sequence from FlyBase (Gramates et al. 2017). The reference sequence was used to effectively annotate the coding regions of the four newly sequenced strains.

Once the sequences were annotated, *Mekk1* for all four newly sequenced strains were compared.

The three strains of *D. simulans* (FC, 199, and 216) all differ from the reference sequence (Gramates et al. 2017; data not shown), as well as from one another and from *D. melanogaster* (BJS; Appendix B). For the analysis to be effective, differences that are irrelevant to strain-specific behaviour needed to be filtered out. Most of the variants are species-specific but not strain-specific as the sequence differed between *D. melanogaster* and *D. simulans* (indels, SNVs, and repeats), but did not differ among *D. simulans* strains. In the analysis that follows, only variants that are unique to *D. simulans* (FC) are analyzed (Table 3.1, Fig. 3.1). Each distinct region within the *Mekk1* sequence was compared between all three strains, and data are shown for the comparison of each strain of *D. simulans* to *D. melanogaster* (BJS; Table 3.2). There are two putative transcripts in each species created through alternative transcriptional start sites (FlyBase), and the reference sequence was used to annotate the coding region of the four sequenced strains.

Each *Mekk1* gene region was roughly 11 kb in length (from the upstream intergenic region to the end of the seventh exon), and these regions were aligned using Geneious R9 (Biomatters Ltd 2015-2016; Appendix B) and % similarity was used as a measure of comparison (EMBOSS; EMBL-EBI), and also to generate a coverage map (indicating how many times each region was sequenced to generate the consensus; Appendix C). Along with the coverage map, electropherograms are included for regions of the consensus that were generated using less than two reads (Appendix C). I calculated the number of non-sensical ("N") nucleotides per 100bp based the low coverage region

Table 3.1 Filter system showing how sequence differences were categorized.

Types of Sequence Differences[1]	*D. melanogaster* (BJS)	*D. simulans* (FC)[2]	*D. simulans* (199)	*D. simulans* (216)
No differences	GGGATCA	GGGATCA	GGGATCA	GGGATCA
Species-specific differences	TTTAGAA	TTCAGAA	TTCAGAA	TTCAGAA
Unique to strain FC (strain-specific differences)	CGGTGGA	CGGAGGA	CGGTGGA	CGGTGGA
Unique to species and FC (species- and strain-specific differences)	CCACGCT	CCACCCT	CCACACT	CCACACT

[1] Types of sequence differences; grey backgrounds are considered types that are potentially relevant to the behaviour of interest. Highlighted letters correspond to nucleotides that differ between strains; teal: genotype of strain of interest; yellow and red: genotype variants. The top row in grey corresponds to changes identified as blue lines in Fig. 3.1, while the bottom row in grey corresponds to those shown as red lines in Fig 3.1. Sequences are examples taken from Appendix B.

[2] *D. simulans* (FC) is the strain that exhibits the behaviour of interest; the other three strains are used comparatively to determine loci that may be relevant to the behaviour of interest.

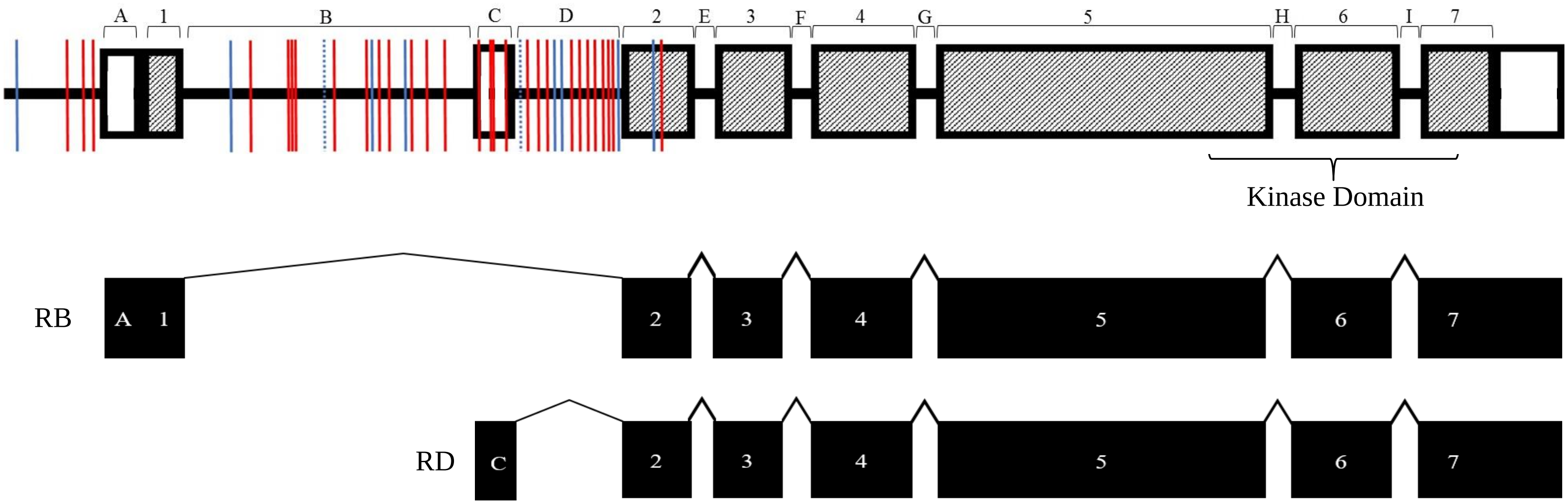


Figure 3.1. *Mekk1* gene region showing the relative positions of variants unique to *D. simulans* (FC). Non-coding regions are drawn as a bold horizontal line, UTRs as open boxes, exons as hashed boxes. The kinase domain, the only characterized functional domain of MEKK1, is indicated by a bracket below. The *Mekk1* gene is shown in 5’-->3’ orientation relative to the sense strand. Blue/red lines (top/bottom grey rows, respectively) and alphanumeric labels defined in Table 3.1. Solid vertical colored lines show SNVs, dotted lines show nucleotide repeat length differences. The two transcripts, RB and RD, are indicated with alphanumeric labels in white, black boxes indicate exons and UTRs present in the transcript with spliced introns indicated by black tented lines.

Table 3.2 Percent similarity for each strain of *D. simulans* as compared to *D. melanogaster* (BJS) within *Mekk1* region.

Region Type[1]	Region Designation[2]	Strain FC[7]	Strain 199	Strain 216
UTR	A	96.1	96.1	96.1
	C[3]	88.6	88.9	89.0
Intron	B[3,4]	86.2	86.5	86.6
	D[3,5,6]	83.9	N/A	N/A
	E	81.6	81.6	81.6
	F	92.1	92.1	92.1
	G	90.0	90.0	90.0
	H[3]	46.7	46.0	46.0
	I	89.4	89.4	89.4
Exon	1	95.5	95.5	95.5
	2[3]	98.0	98.5	98.5
	3	97.4	97.4	97.4
	4	97.9	97.9	97.9
	5[3]	96.7	96.6	96.7
	6	98.3	98.3	98.3
	7	98.9	98.9	98.9

[1] Transcript RB is comprised of all regions (as annotated). Transcript RD is comprised of UTR C, Introns D-I, and Exons 2-7.

[2] Corresponds to the gene regions of *Mekk1* in Figure 3.1. Grey indicates a region with nucleic acid variants unique to *D. simulans* (FC).

[3] These regions contain differences that are not present in strain FC but are present in 199 and 216.

[4] Contains a dinucleotide repeat variation (FC has two more nucleotides at this site compared to the other two strains of *D. simulans*; shown as the first blue hashed line in Fig. 3.1)

[5] Region D contains a region of low confidence as indicated by "N/A"

[6] Contains a single nucleotide repeat variation (FC has three more nucleotides at this site compared to the other two strains of *D. simulans*; shown by the second blue hashed line in Fig. 3.1)

[7] If the % similarity is lowest in the strain FC column (compared to the strain 199 and 216 columns), then there is a variant unique to strain FC in this region.

and included that in Appendix C as well as an image of the electropherogram or sequence for any regions with a coverage of 1 or 0 in the consensus. If the region with low coverage had more than 3 N / 100 bp, then the sequence was left unedited in the consensus sequence comparison (Appendix B), and the edited sequence (from Appendix C) was used for comparison in Table 3.2.

For any region with two or more reads, the consensus and electropherograms used to generate the consensus were examined manually to ensure accuracy. If there were only two reads that disagreed on what nucleotide was called, the one with higher accuracy overall (with fewer N's per 100bp) was used in the consensus. With three or more reads per region, the nucleotide called the most was used in the consensus (for example: if C was called three times and T was called once, then C was used in the consensus sequence). If there was an even number of reads and two different bases were called the same number of times (for example: two reads with T and two reads with G) then whichever reads were the most accurate (based on the average of their two accuracies in N/100bp) were used in the consensus sequence. In some cases, the software was unable to detect a peak. In the event there was a distinguishable single peak, the sequence generated by the software was manually edited. In that scenario, the base pair corresponding to the color of the highest peak was inserted at the locus that previously had an "N". Despite this analysis, there was still uncertainty in the data so a third party independent re-analysis of the data was conducted. Once the third party compiled a consensus, it was compared to the consensus sequence generated in this thesis (Appendix D). The data were found to be very similar, although the third party analysis used a different process to generate the consensus. Any differences that persisted between the data from this thesis and the third party analysis is largely due to the third party strictly adhering to the electropherogram's intrinsic threshold of detection (Appendix C). As the differences between the consensus generated in this thesis and the consensus generated by the third party were not easily reconcilable, the sequencing data should be interpreted with caution.

Mekk1 is shorter in *D. melanogaster* (BJS; 10788 bp) compared to *D. simulans*, but the *D. simulans* sequences are almost identical in length with those from strain 216 and FC at 10,983bp compared to 10,982bp in strain 199. As expected, the percent (%) similarity is higher when comparing exons than when comparing introns (Table 3.2; Appendix C). Region H (an intron) has the lowest similarity between *D. melanogaster* (BJS) and all three strains of *D. simulans* at roughly 46-47%. The low value is due to a species-specific insertion/deletion (indel) that is 156bp in length that is absent in *D. melanogaster* and present in *D. simulans*. However, as this indel is species-specific rather than line-specific, it is unlikely to cause the observed difference in behaviour among *D. simulans* strains.

There are 39 variants that are unique to *D. simulans* (FC) and they are found in the upstream intergenic region up to the second exon in the RB transcript (Table 3.2). As % similarity refers to the measure of comparison of an entire region (as defined in Table 3.2; Appendix C), variant refers to a difference in sequence at a particular locus within a specific region and a SNV is a variant that affects a single nucleotide, compared to an indel or nucleotide repeat variation that affects more than one nucleotide. Regions that contain sequences unique to *D. simulans* (FC) have a lower % similarity in strain FC compared to strain 199 and strain 216 (Table 3.2; shown in grey; Appendix C). Of the 39 unique variants 37 are SNVs, one is a single nucleotide repeat length variation and one is a dinucleotide repeat length variation (Fig. 3.1). Both length variations were identified in intronic regions (Fig. 3.1; Appendix C). Twenty-nine of the 39 unique variants are unique to *D. simulans* (FC) with the sequence of the other strains being identical (BJS, 199, 216; third row in Table 3.1; Appendix C). The remaining ten sequence differences vary between strains and *D. simulans* (FC) has a unique variant (last row in Table 3.1; Appendix C). One intronic region (region D in Table 3.2; Appendix C) has a region of 50bp that I was not able to sequence in *D. simulans* strains 199 and 216. This region of low confidence does contain two nucleotides that are potentially unique to *D. simulans* (FC), the strain of interest (Appendix B; Appendix C).

Some of the comparisons in Tables 3.2 indicate lower % similarity in *D. simulans* (FC) compared to the other three strains, which points to regions that potentially contribute to the behavioural phenotype as there is a nucleotime in that region that has a variant unique to *D. simulans* (FC; region highlighted in grey in Table 3.1; Appendix C). The majority of these nucleotide differences (37/39) are in non-coding regions as they are present in either an intron, or the 5’-UTR. Only two of the SNVs are present in a coding region. Although these differences unique to *D. simulans* (FC) are in a coding region, they do not result in any amino acid substitutions in the corresponding protein product (Appendices G & H). As only two nucleotides are altered within the coding region, and not in a UTR, it is also unlikely that these coding sequence differences have a significant or biologically relevant effect on translational efficiency (Shen et al. 1999). Shen et al. (1999) determined the SNVs can potentially affect the secondary structure of a transcript leading to an effect on translational efficiency, but this is documented to occur through sequence differences in UTRs (not coding regions; Chatterjee and Pal 2009). Therefore, based on this sequence analysis, it was concluded that the difference in behaviour between *D. melanogaster* and *D. simulans* is likely due to a change in a non-coding regulatory element. While I am confident in this analysis, the sequencing data should be interpreted with caution due to some instances of low coverage (identified in Table 3.1; Appendix C). Regardless of the uncertainty in certain regions of the sequence, the interpretation remains the same: it is likely a non-coding regulatory region contributing to female rejection behaviour in *D. simulans* (FC).

3.2 Coding region analysis of *Mekk1*

Aside from differences between strains, it was interesting to look at species-specific differences. As sequencing identified variants between strains and between species, it was important to look at the putative amino acid sequence for *Mekk1* to determine if there were any coding sequence differences. Differences in the coding region may indicate that the protein has an altered function in *D. simulans* compared to *D. melanogaster*, and this altered function could potentially contribute to the observed difference in behaviour.

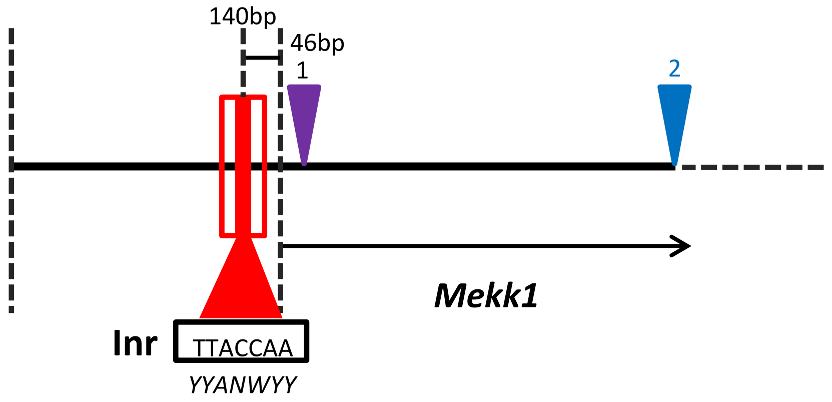
There is a difference in the length of the transcripts and protein products between *D. simulans* and *D. melanogaster*. For the RB transcript the length is 4,722 bp in all three strains of *D. simulans*, while *D. melanogaster* (BJS) RB transcript is 6 bp shorter (4,716 bp; Appendix E). The difference in length is not in the region coding for the kinase but rather in the first exon of the RB transcript. For transcript RD, all of the transcripts in all four strains are 4,494 bp in length (Appendix F). Therefore, the size difference of the RB compared to the RD/RC protein product is species-dependent and is due to the indel in the first exon of the RB transcript (*D. melanogaster* has 74 amino acids while all three strains of *D. simulans* have 76 amino acids).

Although not necessary for the line-specific comparisons, there are sequence differences within species between MEKK1 isoforms created through alternative translational start sites. Between the RB and RD/RC isoforms within either species (*D. melanogaster* or *D. simulans*), 1,492 amino acids at the C-terminus of the protein are common to both. There are five amino acids at the N-terminus of the RD/RC protein product that are unique compared to the RB protein product, which has a longer unique sequence at the N-terminal end compared to RD/RC.

Consistent with the differences in transcript length, all of the *D. simulans* strains code for a *Mekk1* protein (MEKK1-B) from the RB transcript that is 1,573 amino acids while the one from *D. melanogaster* is shorter with 1,571 amino acids in length (Appendix G). The RD/RC protein product (MEKK1-D) is 1,497 amino acids in length in all four strains of *Drosophila* (Appendix H). Therefore, the differences between transcripts and protein products is species-specific rather than line-specific, indicating that this difference in length does not likely contribute to the female rejection behaviour.

3.3 Analysis of proximal regulatory regions of *Mekk1*

It is possible that the *Mekk1* transcripts are differentially regulated in the different species causing the behavioural differences observed. In the sequenced promoter region a putative initiator element (Inr) with the sequence TTACCAA (consensus Inr sequence: YYANWYY; Y=pyrimidine, N=any nucleic acid, W=weak nucleic acid; Smale and Baltimore 1989) that is present in *D. melanogaster* (BJS) and absent in all three strains of *D. simulans* was identified (Fig. 3.2). The putative Inr is located less than 200 bp


Drosophila melanogaster

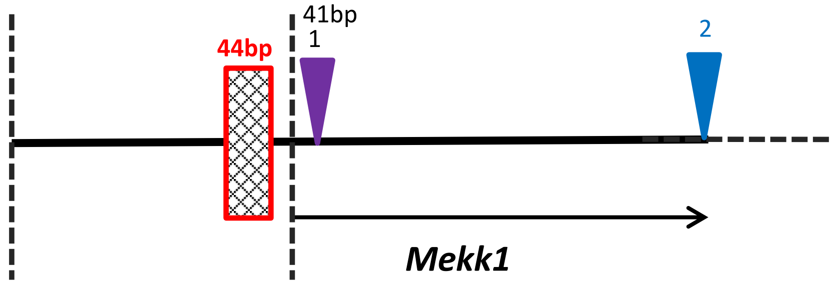

Drosophila simulans

Figure 3.2. Diagram of initiator element (Inr) in *D. melanogaster* compared to *D. simulans*. Purple and blue triangles: putative transcript start sites. Dotted vertical lines: outline upstream intergenic region. Arrows: *Mekk1* gene region. Red boxes indicate region containing Inr in *D. melanogaster* (top) and deleted region in *D. simulans* (bottom). Inr is 140bp upstream of the putative transcriptional start site in *D. melanogaster* and deletion is 44bp upstream of the putative transcriptional start site in *D. simulans*.

upstream of *Mekk1*. However, it is unlikely that the initiator element contributes to female rejection behaviour as it is a species-specific sequence rather than a strain-specific sequence. The proximal promoter of *Mekk1* is directly upstream of *Mekk1* in the intergenic region. Four nucleotide changes were identified in the upstream proximal promoter that are unique to *D. simulans* (FC; Appendix B).

The translational start site (TLS) and the transcriptional start site (TSS) were identified from the reference sequence on FlyBase (Gramates et al. 2017). The TLS is between the 5’ UTR and the first exon of the RB transcript (Fig. 3.1). The TSS (designated as +1) is where the 5’ UTR starts, which is 359bp upstream of the TLS. To examine the predicted TSS, the intergenic region from the last nucleotide of the upstream gene to the first nucleotide of *Mekk1* was analyzed using software from the Berkley *Drosophila* Genome Project (Reese 2000). Searching the intergenic region for a predicted TSS was performed as the TSS from FlyBase were putative; it is possible that the TSS was actually farther upstream or downstream. The software (Reese 2000) produces a number between zero and one indicating the strength (s) of the predicted TSS (the closer to one, the stronger the predicted TSS). For the three strains of *D. simulans* this encompasses 418 nucleotides directly upstream of *Mekk1*. For *D. melanogaster* (BJS) the sequence spans positions -468 to -1, and the length is longer due to a species-specific indel. There is a single predicted TSS for the proximal promoter region for *D. melanogaster* (BJS), *D. simulans* (199) and *D. simulans* (216; Reese 2000). TSSs are at-294 (s: 0.96), -129 (s: 0.91), and -129 (s: 0.91), respectively. However, no predicted TSS was identified for *D. simulans* (FC).

When the computer analysis is changed to include the 5’ UTR of the first transcript for *Mekk1* (RB), the sequence used for all three strains of *D. simulans* spans positions -418 to +359, while the sequence used for *D. melanogaster* (BJS) spans positions -468 to +359. The strongest putative TSS for all four strains was predicted for the same position between all four strains within the 5’ UTR. In this case, the TSS is predicted to be identical between the three strains of *D. simulans* (FC, 199 and 216) and is located at position +46 (all have a strength of 1). *D. melanogaster* (BJS) has a similar predicted TSS located at position +45 (s: 1). Ideally, when using the Berkley *Drosophila* Genome Project Neural Network Promoter Prediction software (Reese 2000), the predicted TSS should be relatively consistent regardless of the length of sequence used for prediction. If they are consistent, then this acts as a stronger indicator that the putative TSS are accurately predicted. Unfortunately, this was not the case for *Mekk1* and the predicted TSS and the predicted TSS depends on the sequence used for prediction. It is possible that the TSS is different between *D. simulans* (FC) and the other three strains of *Drosophila*. If *D. simulans* (FC) has a unique TSS, it may be the underlying cause of the observed female rejection behaviour.

3.4 Transcript confirmation

Although two transcripts (RB and RD/RC per species) were putatively predicted through the use of alternative start sites (Fig. 3.3), one of the transcripts in *D. simulans* did not have prior experimental support (Humbert et al. 1993). After examining the annotated *Mekk1* reference sequence from FlyBase (Gramates et al. 2017), there are sequences common to both transcripts in both species, and there are also regions unique to each transcript in both species (Fig. 3.1). RT-PCR was performed on cDNA prepared

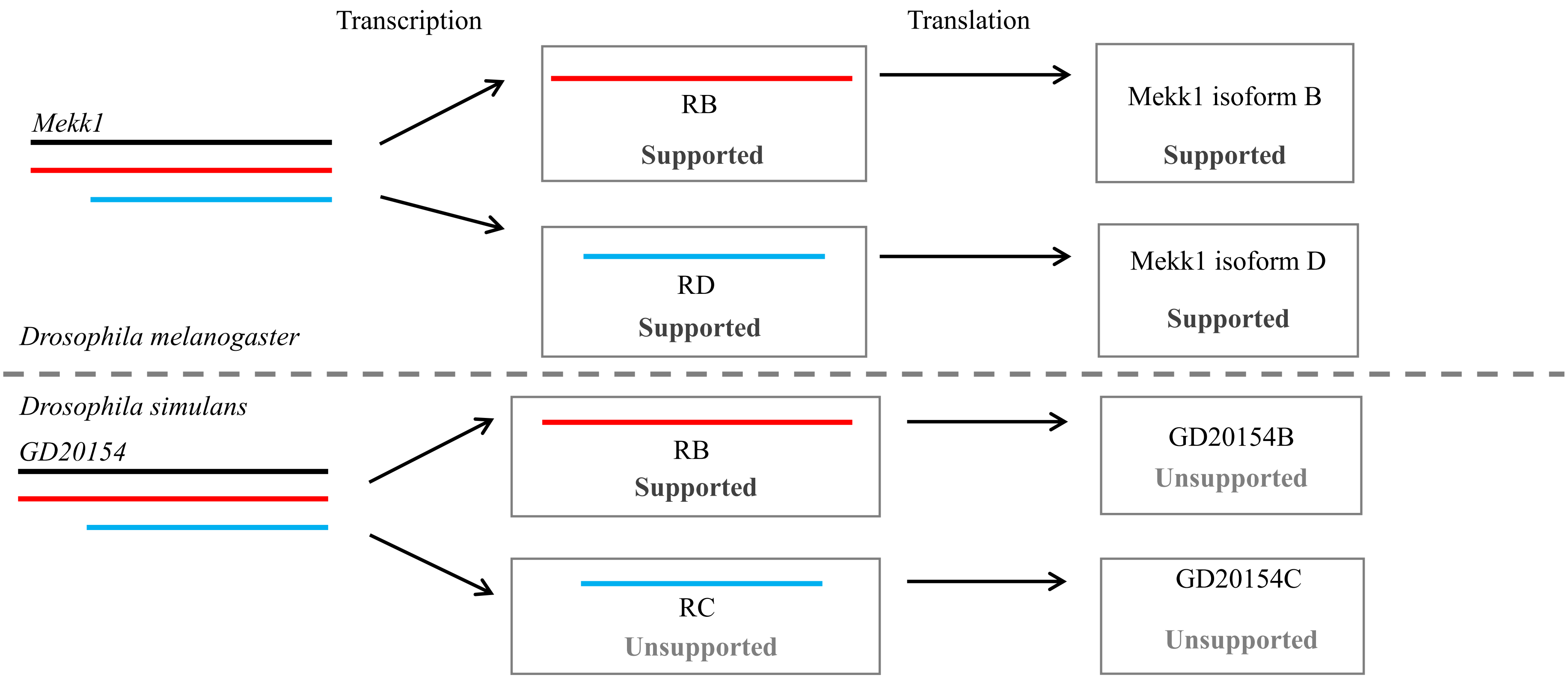


Figure 3.3. Diagram showing alternative start of putative transcripts in *D. melanogaster* and *D. simulans*[1].

Previous experimental evidence described as “Supported” and putative transcript/protein as “Unsupported”. Red lines symbolize the RB transcript (including introns), the blue lines the RB/RC transcript (including introns), and the black line the complete gene region (5’→ 3’ relative to the sense strand), with names indicated for each gene/transcript/protein isoform.

---

[1] According to FlyBase (Gramates et al. 2017) there was no experimental evidence for *Mekk1* in *D. simulans*. However, the study by Graze et al. (2009) did examine *Mekk1* in a high throughput study and found evidence that RB was expressed in *D. simulans*, providing experimental evidence for *Mekk1* being expressed in both species.

from several whole flies of either *D. melanogaster* (BJS) or *D. simulans* (FC). The amplicon size for the RB transcript was predicted to be 198bp and 144bp for transcript RD/RC. Bands were present at the predicted sizes (Fig. 3.4), and bands other than the products of interest are genomic contamination based on amplicon size. As both type of transcript (RB and RD/RC) were detected in both species, the difference in behaviour cannot be explained by one of the transcripts being entirely present/absent in one of the species. It was important to determine if the transcripts were expressed in a differential tissue-specific manner.

3.5 Expression analysis of *Mekk1* transcripts

As the second shorter transcript (RD/RC) had not been detected prior to this study in *D. simulans*, there was no published expression data for the RD/RC transcript in this species. Hence, RT-PCR was performed using RNA samples isolated from females at different lifecycle stages and from different tissue types: adult heads, adult bodies, pupal heads, and pupal bodies. The four tissue types were collected for various reasons. RNA from adult females was examined as this is the lifecycle stage that the behaviour is observed, while brain development occurs during pupation. Various regions of the brain develop throughout pupation (Belote and Bakert 1987), specifically a region known as the mushroom body (Technau and Heisenberg 1982), which has been linked to decision-making in adult *Drosophila* (Zhang et al. 2007). RNA from heads was examined as the trait of interest is a behaviour which could be controlled by the nervous system, while the body potentially is engaged in feedback loops that could also contribute to behaviour (Levine and Mullins 1964).

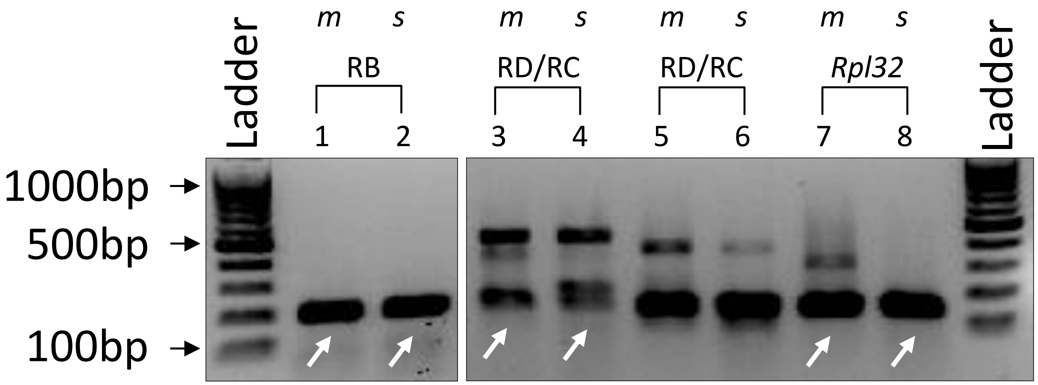


Figure 3.4. RT-PCR shows presence of two transcripts in both *D. melanogaster* and *D. simulans*. A 100bp ladder is shown in the outside lane. Brackets above indicate pairs of lanes that contain samples amplified with the same primer set but differ by templates *m* = *D. melanogaster* (BJS) and *s* = *D. simulans* (FC). RB or RD *Mekk1* transcripts; *Rpl32* a control. White arrow: band of interest. Primers for lanes 5 & 6 differed from lanes 3 & 4, and were not used moving forward.

The transcripts were amplified independently of one another and results were compared between species within each sample type (Fig. 3.5; Table 3.3). Although there were trends in the data, none of these differences were statistically significant (Table 3.3, Appendix I), potentially due to the high amount of variation across the technical and the biological replicates. In both species both transcripts were detected in all four female RNA samples (Fig. 3.4), indicating that the behaviour is not likely due to a transcript being present or absent in one species' head or body compared to the other.

3.6 Behavioural assays: rescuing *Mekk1* in a tissue-specific manner

To confirm the role of *Mekk1* in female behaviour and determine if *Mekk1* influences behaviour in a tissue-specific manner, a rescue experiment using the GAL4/UAS system was performed. To perform this rescue experiment, the gene of interest (*Mekk1*) was disrupted. Initially, a *P*-element insertion within *Mekk1* was used in a hybrid (*D. melanogaster* x *D. simulans*) background and the behavioural phenotype was observed (Moehring Lab, unpublished). These females had one functional copy of *Mekk1* from *D. simulans* and one disrupted copy from *D. melanogaster*. After the trait of interest, in this case female rejection of *D. melanogaster* males, was observed in individuals that carried a *Mekk1* disruption, a rescue experiment determines if adding a wild-type *Mekk1* restores the wild-type phenotype.

The GAL4 driver under the control of a tissue-specific promoter (Duffy 2002) was used to turn on a *Mekk1* transcript linked to a UAS element (Bischof et al. 2013). The GAL4/UAS system will be used in a strain with the *Mekk1* disruption to rescue *Mekk1* expression, and secondly to avoid overexpression of *Mekk1* wherever the GAL4/UAS system expresses *Mekk1*. This experiment can also determine if *Mekk1* acts

Table 3.3 Statistical values for *Mekk1* RT-PCR by lifecycle stage and tissue type.

Transcript	**Lifecycle stage and tissue type**	***F*-value**	***p*-value**
RB	Pupae Bodies	$F_{1,4}=0.85$	0.41
	Pupae Heads	$F_{1,4}=0.30$	0.61
	Adult Bodies	$F_{1,4}=0.16$	0.71
	Adult Heads	$F_{1,4}=0.01$	0.91
RD	Pupae Bodies	$F_{1,4}=0.04$	0.84
	Pupae Heads	$F_{1,4}=0.07$	0.81
	Adult Bodies	$F_{1,4}=0.05$	0.84
	Adult Heads	$F_{1,4}=0.00$	0.99

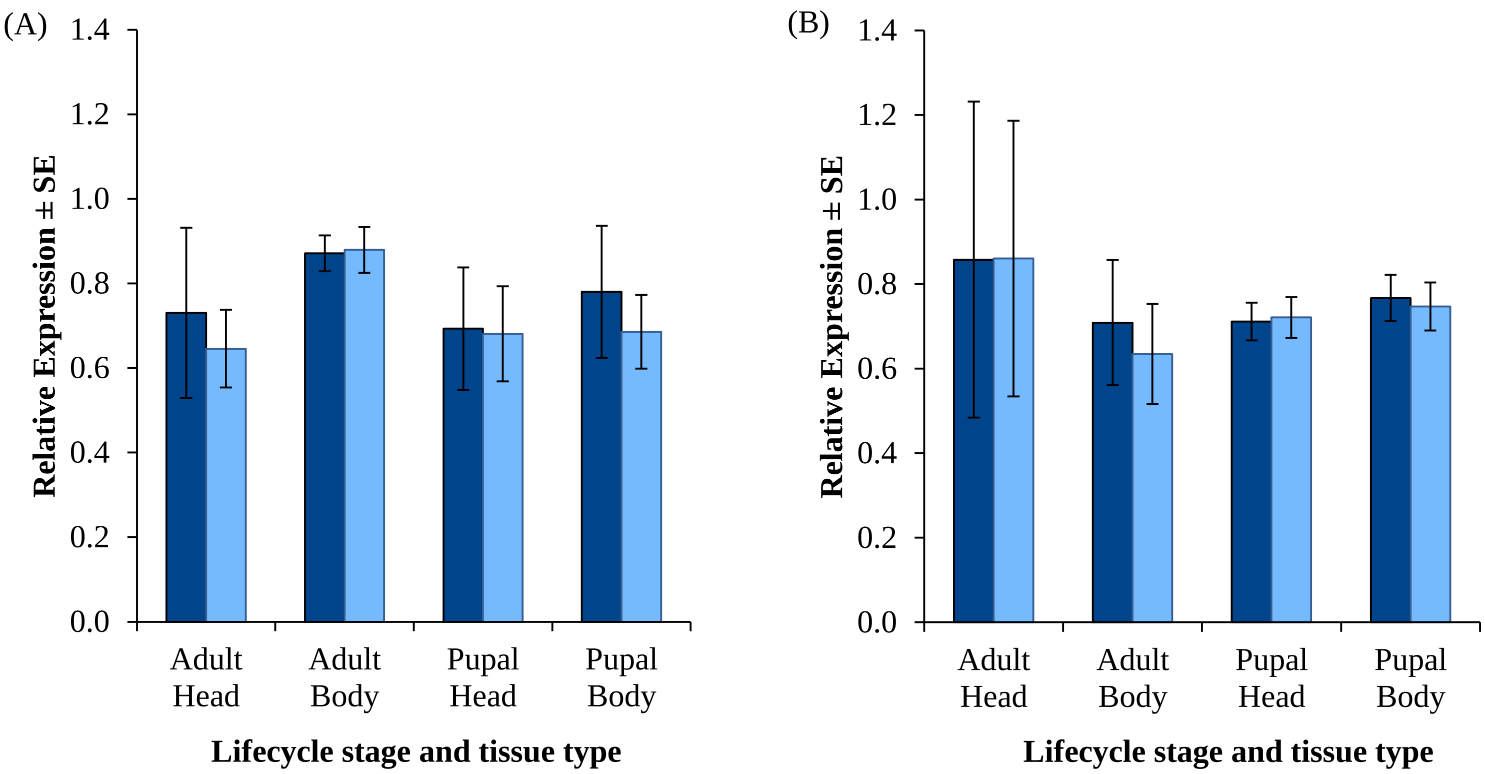


Figure 3.5. Comparison of relative *Mekk1* expression levels determined by RT-PCR between species across lifecycle stages and tissue types. (A) Transcript RB; (B) transcript RD/RC of *Mekk1*. Data shown were normalized to control gene (*Rpl32*). There were no significant differences between species for any of the sample types of either transcript (n=3; test statistics in Table 3.2). Dark blue: *D. melanogaster* (BJS), light blue: *D. simulans* (FC). Error bars are ± SE for biological replicates.

in a tissue-specific manner, and if *Mekk1* is indeed responsible for the behaviour of interest.

In hybrids with a disruption in *Mekk1*, females will reject *D. melanogaster* males. In wild-type hybrids, females will mate with *D. melanogaster* males. It is possible that *Mekk1* only needs to be expressed in a specific subset of tissues in order to mediate female rejection behaviour. Thus, if *Mekk1* is expressed in a manner consistent with it's wild-type expression pattern, then flies with all three transgenic elements (GAL4, UAS, and a *Mekk1* disruption) should have the same high mating frequency as a wild-type hybrid female with no *Mekk1* disruption. By utilizing GAL4 drivers that turn on *Mekk1* in different subsets of tissues, it may help narrow down which tissues are important in mediating female rejection behaviour.

The controls for the UAS and GAL4 driver respectively require that four separate assays be set up simultaneously, each assay containing a female of a different genotype. Individuals with a UAS and a *Mekk1* disruption should be unable to express *Mekk1* and have a low mating frequency/reject *D. melanogaster* males. If they accept *D. melanogaster* males, this indicates that either the genotype being assayed is incorrect or the UAS element is expressing *Mekk1* in the absence of a GAL4 driver. An individual with a GAL4 driver and no other transgenic elements should act the same as a wild-type hybrid female, meaning that it will accept *D. melanogaster* males.

The first GAL4 driver used was under the control of an *Actin-5C* promoter, a ubiquitously expressing promoter (Sieber & Thummel 2009). Interestingly, expressing *Mekk1* ubiquitously was lethal (data not shown). Therefore, a second GAL4 driver was used under the control of a *daughterless* promoter. *daughterless* is highly expressed in many tissues (Rohrbough et al. 2004), making the promoter the most similar GAL4 driver to the ubiquitous but lethal *Actin-5C* driver. The other GAL4 driver that was used was under the control of an *elav* promoter which causes high expression but only in neural tissues (Dimitroff et al. 2012).

Once flies of the appropriate genotype were sexually mature, behavioural assays were performed. The results were examined as the experiment was in progress after a sample size of 13 had been achieved for the group utilizing GAL4-*da* and 9 for the group utilizing GAL4-*elav*. The controls did not act as expected (Fig. 3.6); however, I continued my assays to a sample size of 17 for the GAL4-*da* group and 13 for the GAL4-*elav* group. As the individuals were already prepared and isolated prior to assaying, it made sense to proceed to a slightly larger sample size in case there was some sampling error.

To address the unexpected results, two experiments were performed. The first experiment was performed to ensure that the males in the assays were acting as anticipated (Fig. 3.7, Fig. 3.8). The original study that characterized *Mekk1* as a behavioural isolation gene used *D. melanogaster* males of a different strain (BJS; Moehring Lab, unpublished) compared to this study, which used *D. melanogaster* (GFP). The courtship of these two strains of *D. melanogaster* had been previously compared and found to be the same (Personal Communication; Tabashir Chowdhury, Moehring Lab). In order to ensure there hadn’t been any deviation in how comparable the two strains courtship is, the male courtship latency of both strains was examined.

The assays showed that the courtship latency for the two strains of *D. melanogaster* are different (Fig. 3.7, Fig. 3.8). The *D. melanogaster* (GFP) strain had

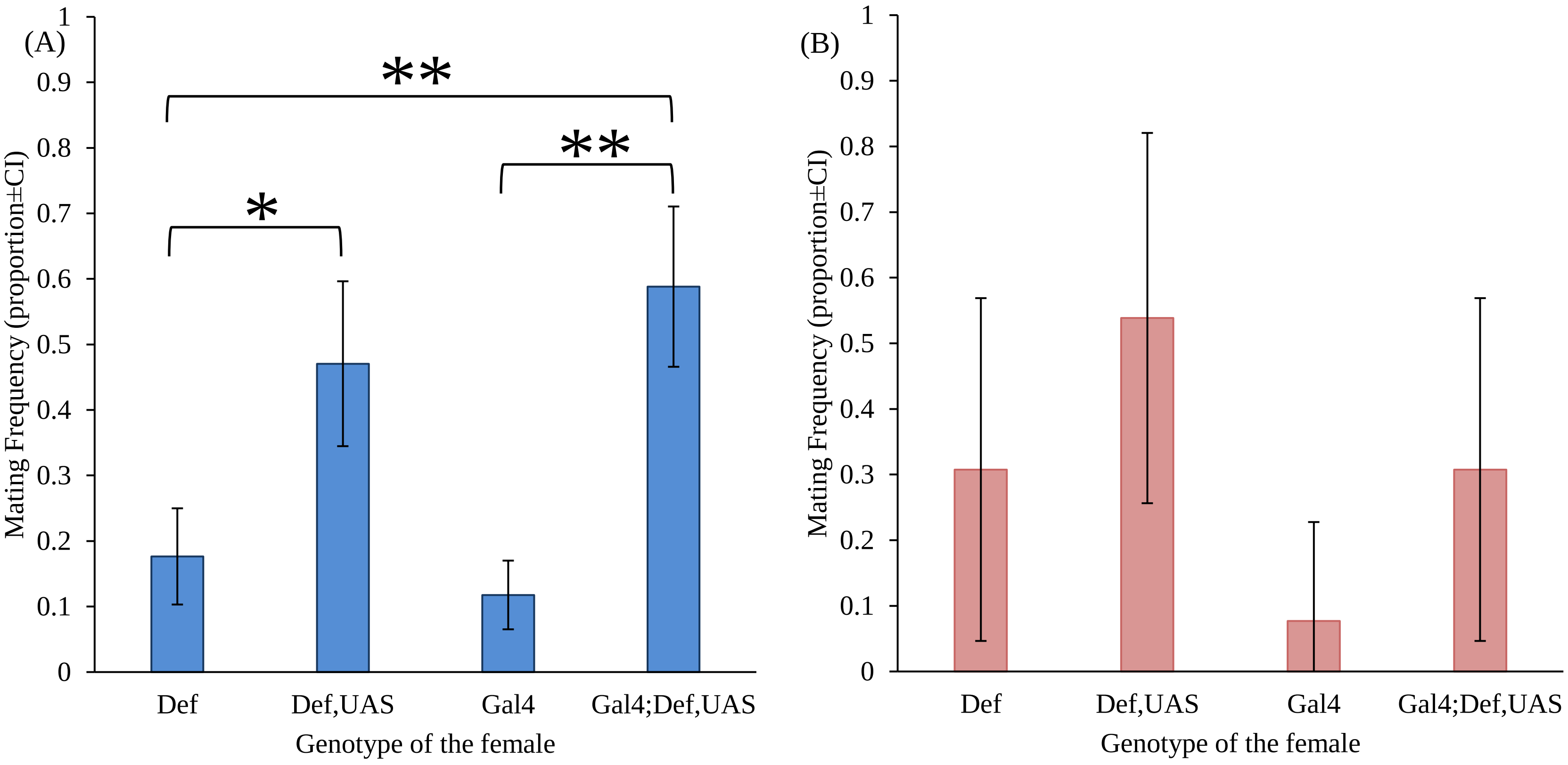


Figure 3.6. Effect of tissue-specific *Mekk1* rescue on mating frequency in various genotypes of *Drosophila* hybrid females. Mating frequency (proportion ± confidence interval) shown for the following genotypes: Def=*P*-element deficiency, UAS= *UAS-Mekk1*; Gal4=tissue-specific GAL4 driver. Blue (left) indicates rescue in all tissues (A; 55851, *daughterless-GAL4*), pink (right) indicates rescue in neural tissues (B; 8765, *elav-GAL4*). *=p<0.05, **=p<0.01

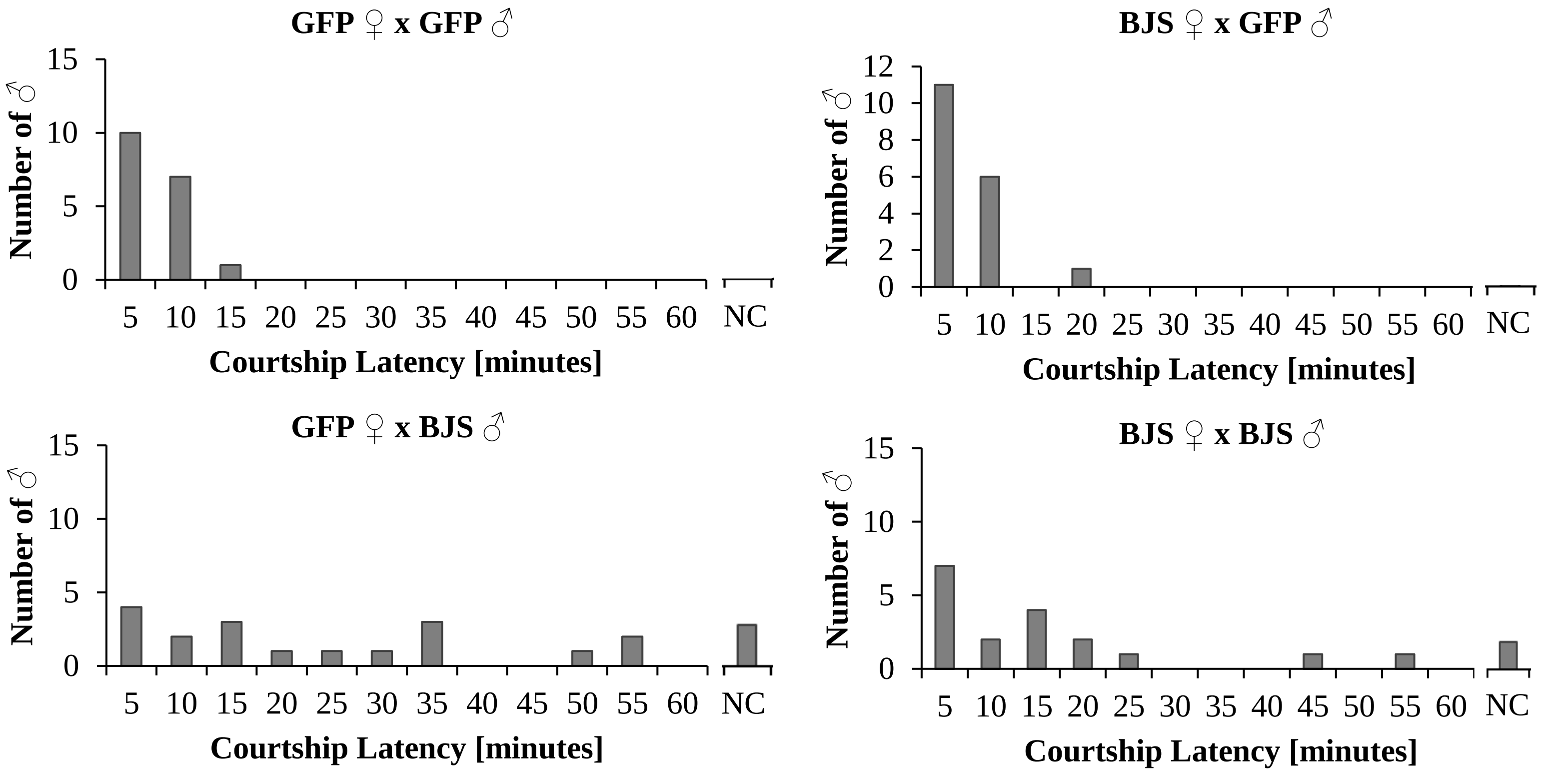


Figure 3.7. Courtship latency of males from *D. melanogaster* (BJS) and *D. melanogaster* (GFP) paired with females from both *D. melanogaster* (BJS) and *D. melanogaster* (GFP). Courtship latency (minutes±SD) measures the time from the start of the assay to when the male begins courtship with the female. For each assay 18 pairings were observed. However, not all pairings actually resulted in courtship. The number of males that actually courted are as follows: n=18 for GFP♂xGFP♀ and GFP♂xBJS♀; n=15 for BJS♂xGFP♀**;** n=16 for BJS♂xBJS♀.

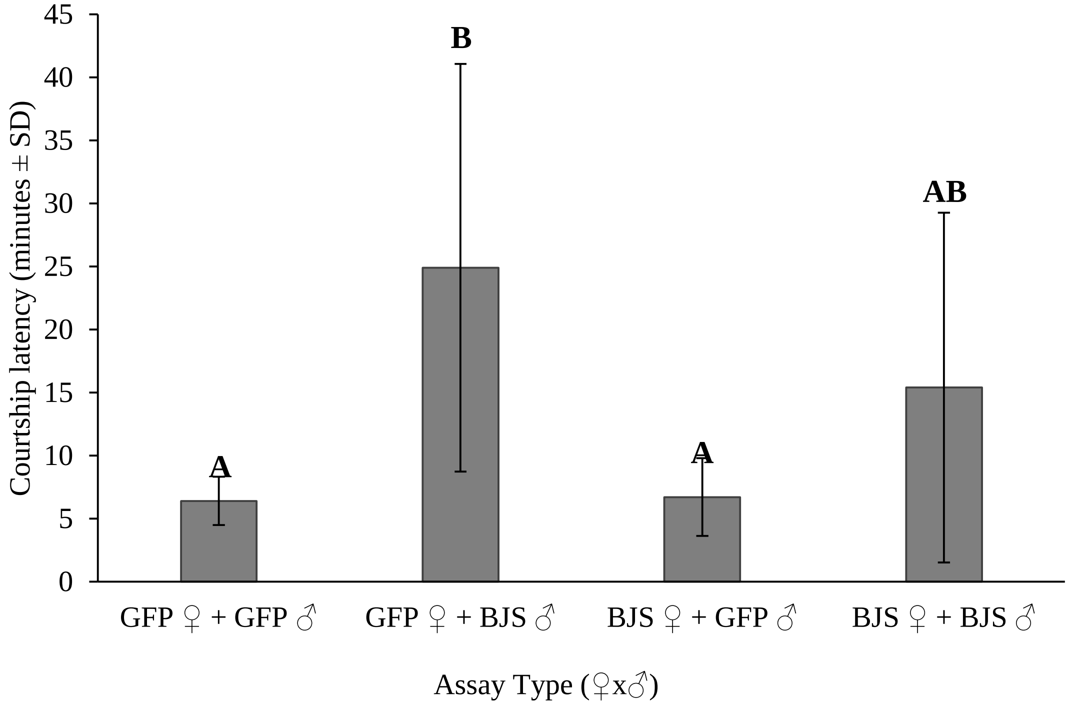


Figure 3.8. Courtship latency of males from *D. melanogaster* (BJS) and *D. melanogaster* (GFP) paired with females from both *D. melanogaster* (BJS) and *D. melanogaster* (GFP). Courtship latency (minutes±SD) indicates the time from the start of the assay to when the male begins courtship. There was a significant difference between courtship latency as determined by Kruskal-Wallis (H(3)=24.3, $p=2.0 \times 10^{-5}$), letters indicate significant differences (n=18).

significantly shorter courtship latency ($p<0.05$) compared *D. melanogaster* (BJS) males (Fig. 3.8), specifically when courting females of their own strain (6.4 ± 1.9 and 15.4 ±13.9, respectively). Most obviously, the distribution of courtship latency differs (Fig. 3.7), emphasizing that male courtship behaviour is not identical between the two strains of *D. melanogaster*. Although the initial experiment and the one performed for this thesis are different, the difference in courtship latency found here does highlight that the behaviour of the male is not consistent with that of the initial male *D. melanogaster* strain.

To ensure the individuals with both a UAS element and a *Mekk1* disruption were of the correct genotype, eight flies of the UAS-*Mekk1*,*Mekk1* stock were genotyped. Upon genotyping individuals from the assays, it became apparent that the *Mekk1* disruption was not present and the flies that were support to have UAS elements and *Mekk1* disruptions only had a UAS element (Appendix J). As the controls in the experiment were not correct, the results of the rescue experiment (for both GAL4 drivers) cannot be used. Potential explanations for the unexpected trends in mating frequency could include recombination or stock contamination.

3.7 Behavioural assays: disrupting the coding region of *Mekk1*

During the initial deficiency mapping, a disruption strain (19991) was used to test *Mekk1* as a candidate behavioural isolation gene (Moehring Lab, unpublished). The *P*-element disruptions used to initially characterize *Mekk1* as a candidate behavioural isolation gene were both in non-coding regions (Moehring Lab, unpublished; Fig.1.4), and did not affect the sequence of the final protein product. The deletion strain (109805) tested here has a deletion spanning a portion of the kinase domain (289 amino acids; Fig.

3.1), which is the only characterized functional domain present in the MEKK1 protein products (RB or RD/RC). The deletion causes the kinase domain to have significantly decreased function in the context of stress response (Inoue et al., 2001). Unlike the *P*-element insertions, the deletion strain has a functional portion of the MEKK1 protein product removed.

The assay of the deletion strain showed that there was no significant effect of the deletion on mating behaviour (Fig. 3.9). The mating frequency of the deficiency compared to the balancer genotype in the hybrid genetic background (Fig. 3.9) was similar, indicating that the deletion does not have an effect on female behaviour. The *P*-element disruptions previously tested did show an effect on female behaviour. It is therefore likely that the female rejection behaviour influenced by *Mekk1* is mediated through something other than the kinase domain.

3.8 *Mekk1* visualization

Three strains of *Drosophila* were used in an attempt to visualize tissue-specific *Mekk1* expression. Two strains of *Drosophila* have a construct containing a GFP tag within the *Mekk1* gene region (Fig. 3.10B, F, J; C, G, K; Venken et al. 2011). Both of these two constructs are both within an intron. The third strain (Fig. 3.10D, H, L) has an enhancer trap within *Mekk1* being used to express GFP from a UAS-GFP construct elsewhere in the genome. The three strains containing GFP will be compared to a wild-type strain of *Drosophila* to control for tissue autofluorescence (Fig. 3.10 A, E, I).

If the constructs function as anticipated, the pattern of fluorescence should mimic that of *Mekk1* expression. It is possible that the construct containing *GFP* introduced into the *Mekk1* gene region will not fully recapitulate *Mekk1* expression. *Mekk1* antibodies

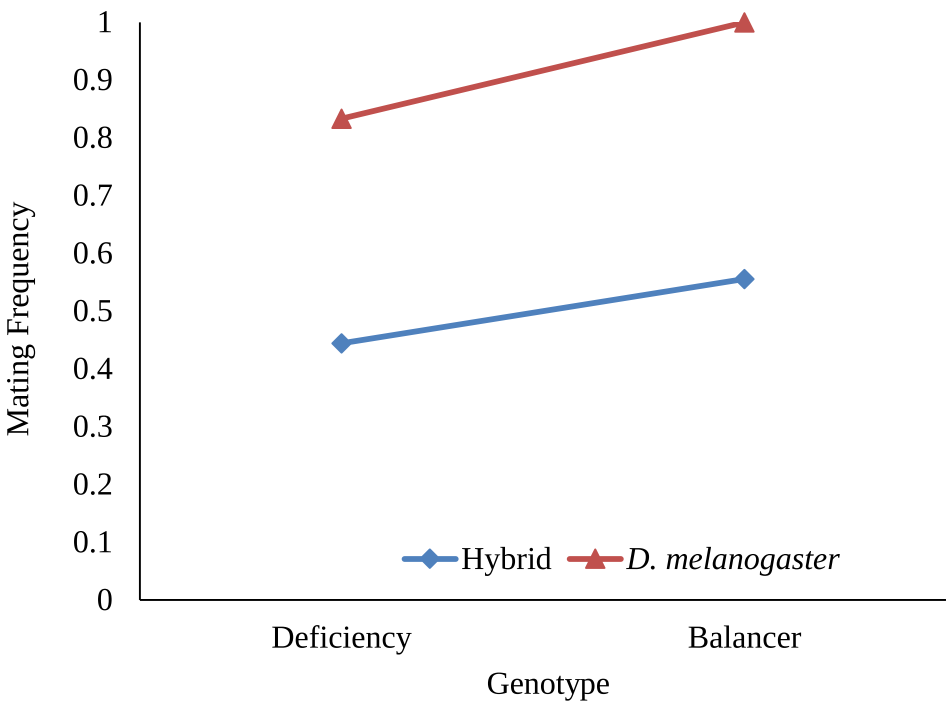


Figure 3.9. Mating frequency for deletion mapping of functional kinase domain of *Mekk1*. Blue line: hybrid (*D. simulans* x *D. melanogaster*); red line: *D. melanogaster*. Genotype indicates whether individual had a *Mekk1* deficiency (deletion in the kinase domain) or a balancer on the third chromosome (G=0.004802, df=1, p=0.945).

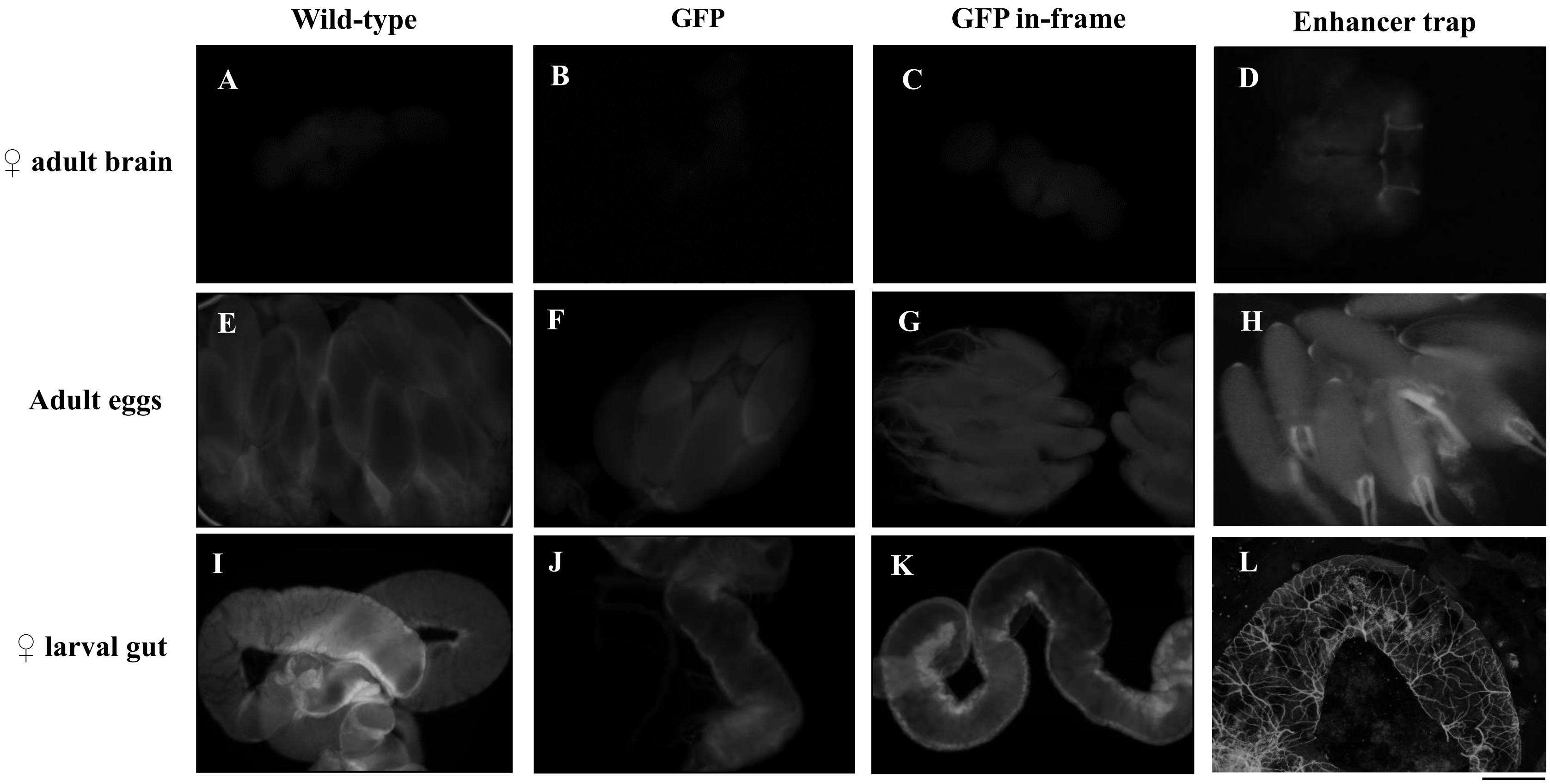


Figure 3.10. Visualizing expression using GFP. Wild-type (*D. melanogaster* BJS) was used as a comparison to gauge autofluorescence of the various tissue types compared to the strains that were being examined for expression due to inserts in the *Mekk1* gene. The level of fluorescence is similar between strains within tissue types (comparing A-C, E-G, I-K) for the first three strains (wild-type, GFP in-frame (stock #: 60227), GFP (stock #: 41489)). Scale bar: 50µm.

were explored as an option, however there are none commercially available specific to *Drosophila* and these strains (containing *GFP* or GAL4 in the *Mekk1* gene region) were the next available option. Aside from potential misexpression, a separate issue is that many tissues in *Drosophila* exhibit autofluorescence, which means that the tissues exhibit high levels of fluorescence similar to the fluorophore that one is looking for (Meerwaldt et al. 2005). Previous data has indicated that there is high *Mekk1* expression in the ovaries compared to the whole adult fly, and slightly increased expression of *Mekk1* in the larval gut as compared to the whole larval fly (Lyne et al., 2007), so the larval gut and adult ovaries were observed. The images in Figure 3.10E-H show eggs imaged from portions of adult ovaries. Taking this into consideration, control tissues from wild-type flies (UAS-*GFP* flies without the enhancer trap) were used to determine whether the constructs were indeed expressing *GFP* in a pattern different from autofluorescence for these tissues. For the first two strains with *GFP* constructs in an intron, there was no qualitative difference between autofluorescence and the level of fluorescence in the two of the strains of interest (Fig. 3.10 A-C, E-G, I-K). As the ovaries showed increased expression in the enhancer trap strain (Fig. 3.10 D, H, L; Appendix K) compared to the control, this was taken as an indication that GFP was being expressed in eggs. The larval gut images (specifically Fig. 3.10L) showed a different pattern of fluorescence compared to the control images (data not shown), which may also be indicative of the construct expressing GFP in the larval gut.

Once the enhancer trap in combination with the UAS-GFP was determined to be expressing *GFP* in a pattern similar to previously characterized *Mekk1* expression, female adult brains were dissected and examined (n=5; Fig. 3.10D; Appendix L). As *Mekk1* contributes to a behaviour in the female adult fly, it seemed logical to examine the brain and determine if *GFP* was being expressed, which may indicate that *Mekk1* is also being expressed. Upon imaging brains of female flies with both of the constructs (enhancer trap and UAS-GFP) and comparing those brains to females with negative controls (only containing UAS-GFP without the presence of the enhancer trap; Appendix L), it appears that the female flies with both constructs are expressing GFP in a specific region of the brain: the mushroom body (Fig. 3.10 D; Fig. 3.11). Interestingly, it appears that only certain regions of the mushroom body are expressing GFP. One of the regions expressing GFP is the calyx. Aside from the calyx, either the α/β lobes or the α'/β' lobes are expressing GFP (Fig. 3.11). The mushroom body has been heavily implicated in behaviour and decision making (Zhang et al. 2007), so it is interesting that a gene (*Mekk1*) implicated in female rejection behaviour is expressed in a region previously characterized as influencing behaviour.

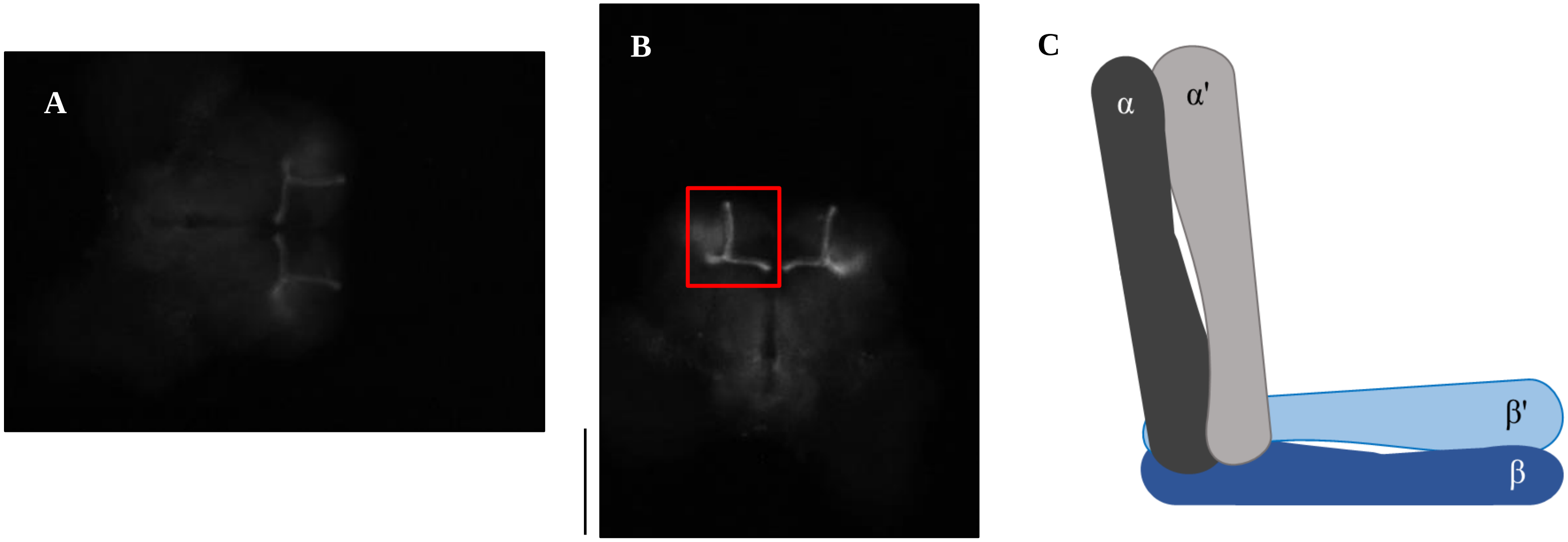


Figure 3.11. Images of *Mekk1* expression pattern using GFP in the female brain. Image A: same as Fig. 3.10 D; Image B: Image A rotated and altered by sharpness +50%, brightness +40%, and contrast +50%; C: diagram of a *Drosophila* mushroom body. Image B was enhanced to facilitate viewing the tissue-specific expression of GFP within the brain, unaltered images are available in Appendix J. The structure outlined by a red box in B appears to be a portion of the mushroom body corresponding to the calyx and either the α/β lobes, or the α'/β' lobes shown in C. The calyx (present in both hemispheres) projects posteriorly, with the β and β' lobes projecting medially from the anterior aspect of the calyx. The α/α' lobes both project dorsally from the anterior aspect of the calyx, connecting to the same point near calyx as the β and β' lobes. Scale bar: 50µm.

4: Discussion

4.1 Multiple unique variants may contribute to behaviour

To determine how *Mekk1* contributes to behaviour, *Mekk1* was sequenced in four lines of *Drosophila* (one line of *D. melanogaster* and three of *D. simulans*). Once the majority of the *Mekk1* gene region was sequenced in each of these four lines, the sequences were annotated. The sequences were then compared to determine which variants, if any, might contribute to female rejection behaviour in *D. simulans* (FC; Fig. 3.1, Table 3.2). Based on the coverage map (Appendix C) it is necessary to use caution when interpreting these data, specifically in regions of low confidence (Table 3.1)

Based on the annotated *Mekk1* sequences, the coding regions were compared (Table 3.2). The genomic DNA (gDNA) sequence indicated that in *D. melanogaster* and *D. simulans* there were two predicted transcripts (Fig. 3.3; RB and RD; one longer and one shorter, respectively; Gramates et al. 2017). The coding regions for each transcript were compared between four lines of *Drosophila*. Each transcript leads to a distinct protein (MEKK1B and MEKK1D), and each protein has a characterized kinase domain (Gramates et al. 2017; Fig. 3.1; Appendix C). All changes in the coding region were synonymous (Appendices B, E, & F), which means that the *Mekk1* primary amino acid sequence is identical and not responsible for female rejection behaviour in *D. simulans* (FC). Based on the annotated gDNA, the kinase domain is identical between three lines of *D. simulans* and the line of *D. melanogaster* examined here. This prediction was directly tested by performing complementation testing with a line of *Drosophila* that had a portion of the *Mekk1* kinase domain deleted in the gDNA (Fig. 3.9). When the kinase domain is disrupted, it does not have an effect on female rejection behaviour. The mechanism of how *Mekk1* that is contributing to female rejection behaviour in *D. simulans* (FC) is likely outside of *Mekk1*'s characterized function as a kinase.

The *P*-element disruptions in Figure 1.4 affect behaviour as characterized in previous experiments (Moehring Lab, unpublished), one might predict that the behaviour is mediated through the characterized kinase function of *Mekk1* and perhaps it being expressed at different levels between species. Based on the complementation testing experiment (Fig. 3.9), that prediction is incorrect. *Mekk1* has two functional protein products of roughly 1500 amino acids in length (see section 3.2), of which 262 amino acids contribute to the kinase domain near the C-terminus of the protein. There are approximately 1200 amino acids that could contribute to an uncharacterized functional domain at the N-terminus of the protein. Many protein kinases acting in the MAPK cascade that require *Mekk1* function have characterized docking motifs (Sharrocks et al. 2000; Tanoue et al. 2001). If a docking domain influences an interaction between a kinase (such as MEKK1) and another protein (any interactor of MEKK1), it's possible that this could affect the other protein's downstream interactions. Hence, a docking domain in MEKK1 is a potential and currently uncharacterized motif that may contribute to behaviour. Using *in silico* analyses to examine homology could be misleading, as the paper by Tanoue et al. (2001) found that the primary amino acid sequences in docking motifs can differ between proteins although conferring the same function. Hence, potential docking motifs need to be confirmed experimentally.

As the primary amino acid sequence is identical between the three strains of *D. simulans*, it is likely that the behaviour is mediated by SNV(s) in non-coding regions. Any differences in the coding regions are species-specific. When comparing the non-coding regions, there are SNVs in the 5' UTR of the second transcript (RD) in *D. simulans* (FC; Fig. 3.1; Appendix C). It is possible that SNVs in a UTR affect the folding of the mRNA differentially such that the secondary structure acts as a method of gene regulation (Shen et al. 1999). The mRNA may form a secondary structure after transcription but prior to translation that is unique to *D. simulans* (FC) caused by SNVs in the 5' UTR. It is also possible that some combination of these variants (shown in Fig .3.1; Appendix C), and that the combination of multiple variants both coding and non-coding have an effect on female rejection behaviour.

4.2 Mekk1* expression is similar between *D. simulans* and *D. melanogaster

There are two transcripts coded for by the *Mekk1* gene in *D. melanogaster* and *D. simulans*. According to the sequence of the gDNA the transcripts and protein products are very similar, but there was still a possibility that the gene is differentially regulated between species. Differential regulation can result in different phenotypes between species (Jeong et al., 2008). For example, the genes *yellow* and *tan* are differentially expressed between two species of *Drosophila* and result in different body pigmentation patterns. It is possible that *Mekk1* may be differentially regulated leading to *D. simulans* females having a unique behavioural phenotype compared to *D. melanogaster*. This thesis provides the first evidence that two transcripts are indeed expressed in both species of *Drosophila* (Fig. 3.4).

Once there was support for two transcripts being expressed in whole flies, it became relevant to examine the relative expression level of each transcript in *D. melanogaster* and *D. simulans* in subsets of tissues and at different lifecycle stages. As the trait of interest is a behaviour, the heads and bodies of female flies were separated in an effort to narrow down potential tissues that may mediate female rejection behaviour. It was also important to consider that *Mekk1* may be developmentally important (Connell-Crowley et al. 2007; Su et al. 1998), which means that expression may differ at the pupal and adult stages. For example, Cyclin-dependent Kinase 5 (CDK5) is important in neuronal development and is differentially expressed across lifecycle stages (Connell-Crowley et al., 2007). Another kinase, produced from the gene *misshapen*, is also important in neuronal development and has varying expression patterns between embryogenesis and adulthood in *Drosophila* (Su et al. 1998). Tissue samples from heads and bodies of flies were collected at both the adult and pupal stages. Two transcripts, RB and RD, were present in the head and body of both adult and pupal *Drosophila*. However, there was no significant difference in expression of either transcript between species within each tissue sample (Fig. 3.5). The RT-PCR data was highly variable, and this variance may have made it difficult to detect biologically-relevant differences. It is possible that *Mekk1* expression is highly variable, and a more sensitive technology (such as quantitative real-time PCR or droplet digital PCR) would be able to detect any biologically-relevant differences.

While there were no significant differences in the relative expression of the two *Mekk1* transcripts between species (Fig. 3.5), these data rule out the possibility that the female rejection behaviour is due to one transcript being entirely missing in one species and leading to a behavioural difference. The expression levels were determined using RT-PCR, it is possible that the analysis was not precise enough to pick up any subtle but biologically relevant differences in expression (Bastien et al. 2008; Paiva-Cavalcanti et al. 2010). It is possible that a subset of tissue (in the brain for example) was important in mediating female rejection behaviour, but was not detectable due to the large amount of other tissues present in the sample.

4.3 Evidence that *Mekk1* may be expressed in the mushroom body

The RT-PCR analysis showed no significant difference in *Mekk1* expression between species. However, when examining the head of a *Drosophila*, there are many different tissues that are included (brain, fat bodies, ocular anatomy; Gramates et al. 2017). The brain is relatively small compared to the head, and although there were no significant differences in expression between the two species, it seemed pertinent to examine if *Mekk1* may be expressed in the brain or a subset of tissues within the brain. To determine if *Mekk1* may be expressed in the brain or a subset of tissues within the brain, an enhancer trap in conjunction with an UAS linked to a GFP marker was used to examine *Mekk1* expression.

The fluorescent images show that *Mekk1* is expressed in part of the *Drosophila* brain: the mushroom body (Fig. 3.10 and Fig. 3.11). The mushroom body functions in courtship and mating (de Belle & Heisenberg, 1994). The region fluorescing in Fig. 3.10D appears to be the mushroom body, and specifically the calyx. Aside from the calyx, it is possible that either the α and β lobes or the α' and β' lobes of the mushroom body are expressing *Mekk1* (Fig. 3.10 – 3.10). As it is unclear which lobes express *Mekk1*, it is difficult to narrow to a specific function that *Mekk1* might have in the mushroom body (O'Dell et al. 1995).

To determine what tissues may be important in mediating female rejection behaviour, a rescue experiment was attempted. Previously, gene disruptions were used to map what loci may be responsible for female rejection behaviour in *D. simulans* (FC), and *Mekk1* was found to mediate this behaviour (Moehring Lab, unpublished). To confirm that this gene was indeed responsible for the behaviour observed, a GAL4/UAS experiment was used to rescue *Mekk1* in a tissue-specific manner. A ubiquitously expressed promoter (for *Actin-5c*) was used; however, driving *Mekk1* ubiquitously was lethal (results not shown). A widely (but not ubiquitously) expressed promoter, for the *daughterless* gene, was used to express GAL4 in a broad pattern.

While rescuing *Mekk1* expression in a broad pattern may serve the purpose of confirming that *Mekk1* is responsible for mediating female *D. simulans* (FC) rejection behaviour, it does not inform which tissues may mediate the behaviour. A promoter expressed in the brain (for the gene *elav*) was used to express GAL4 and rescue *Mekk1* expression in the brain. To perform this experiment, it was necessary to recombine the *Mekk1* disruption (*P*-element insertion) and the UAS-*Mekk1* coding sequence onto the same chromosome (*3R*; Gramates et al. 2017). Interestingly, one of the two constructs did not stably integrate, again indicating that both constructs cannot coexist on the same chromosome (Appendix J). Hence, this study was not able to successfully use GAL4/UAS to examine the effect of tissue-specific *Mekk1* expression on female behaviour.

4.4 Potential mechanisms of *Mekk1* as a behavioural isolation gene

Based on the results previously discussed, it is possible that *Mekk1* is acting through the mushroom body. Interestingly another gene, *katanin-60*, characterized by the Moehring Lab (personal communication) has been shown to potentially mediate female rejection behaviour (the same behaviour associated with *Mekk1*) through the mushroom body. Previous characterizations of *kat60* have shown that it is a microtubule severing protein associated with neuronal development (Mao et al., 2014). As both *Mekk1* and *katanin-60* (*kat60*) influence female rejection behaviour and both potentially act through the mushroom body, it is possible that these two genes either directly or indirectly influence one another. However, there are no previously characterized direct interactions between the protein products of *kat60* and *Mekk1* in *Drosophila* or in other species.

Aside from a direct interaction between the protein products of *Mekk1* (MEKK1) and *kat60* (KAT60), it is possible there is an indirect relationship between MEKK1 and KAT60. One common interactor of both MEKK1 and KAT60 is CDK5 (Ghosh et al. 2012; Orchard et al., 2014; Szklarczyk et al., 2015). CDK5 influences axon guidance during neural development (Connell-Crowley et al. 2000; Connell-Crowley et al., 2007), and is upstream of the MAPK pathway *Mekk1* is involved in (Zhuang et al. 2006). The MAPK pathway is largely regulated via phosphorylation, reaction that both MEKK1 and CDK5 have in common. CDK5 phosphorylates MEKK1 at an amino acid site separate from the kinase domain of MEKK1.

It is possible CDK5 regulates both MEKK1 and KAT60, and has an untested role in behaviour. Interestingly, KAT60 may influence behaviour through the mushroom body (Calhoun 2017), so CDK5 has the potential to act on and influence neuronal development through KAT60. There is also the potential for CDK5 to influence MEKK1 in a similar manner to KAT60. As MEKK1 and KAT60 may be expressed in the same region of the brain, and when disrupted lead to the same behavioural phenotype (Laturney and Moehring 2012b), it is possible that the upstream effector CDK5 influence both downstream genes resulting in the behaviour of interest (Fig. 4.1). I propose that CDK5 may interact with MEKK1 through an uncharacterized domain, and this interaction may mediate behaviour through a mechanism outside of the kinase cascade that MEKK1 is involved in (Fig. 4.1).

4.5 Conclusions and future directions

In summary, there are potentially 39 variants that are unique to *D. simulans* (FC) and may contribute to the behaviour of interest, most of which are non-coding (Fig. 3.1). It is likely that regulation of the gene contributes to behaviour as the coding regions are identical between strains of *D. simulans*. Any of these loci or some combination of them could contribute to female rejection behaviour. Although there were no differences found in *Mekk1* expression between species of *Drosophila* (Fig. 3.5), it is possible that the technology used was either not sensitive enough to detect differences relevant to the behaviour of interest or that the gene is differentially regulated at a lifecycle stage not examined here (Hindson et al. 2011; Hindson et al. 2013).

To address the issue of the technological sensitivity, cell separation techniques could be used to isolate cells of interest (Henry et al. 2012). For example, UNC-84 (a protein) could be used to isolate cells specific to the mushroom body (specifically Kenyon cells). Isolating specific cells would allow *Mekk1* expression to be quantified in subsets of tissues to determine if there are biologically relevant differences that were unable to be detected by RT-PCR. A previous study found evidence of SNV-dependent gene regulation, which suggests the possibility that the non-coding SNVs unique to *D. simulans* (FC) may contribute to differential gene regulation of *Mekk1* that was not detectable via RT-PCR (Nicoloso et al. 2010). The four SNVs in the intergenic region

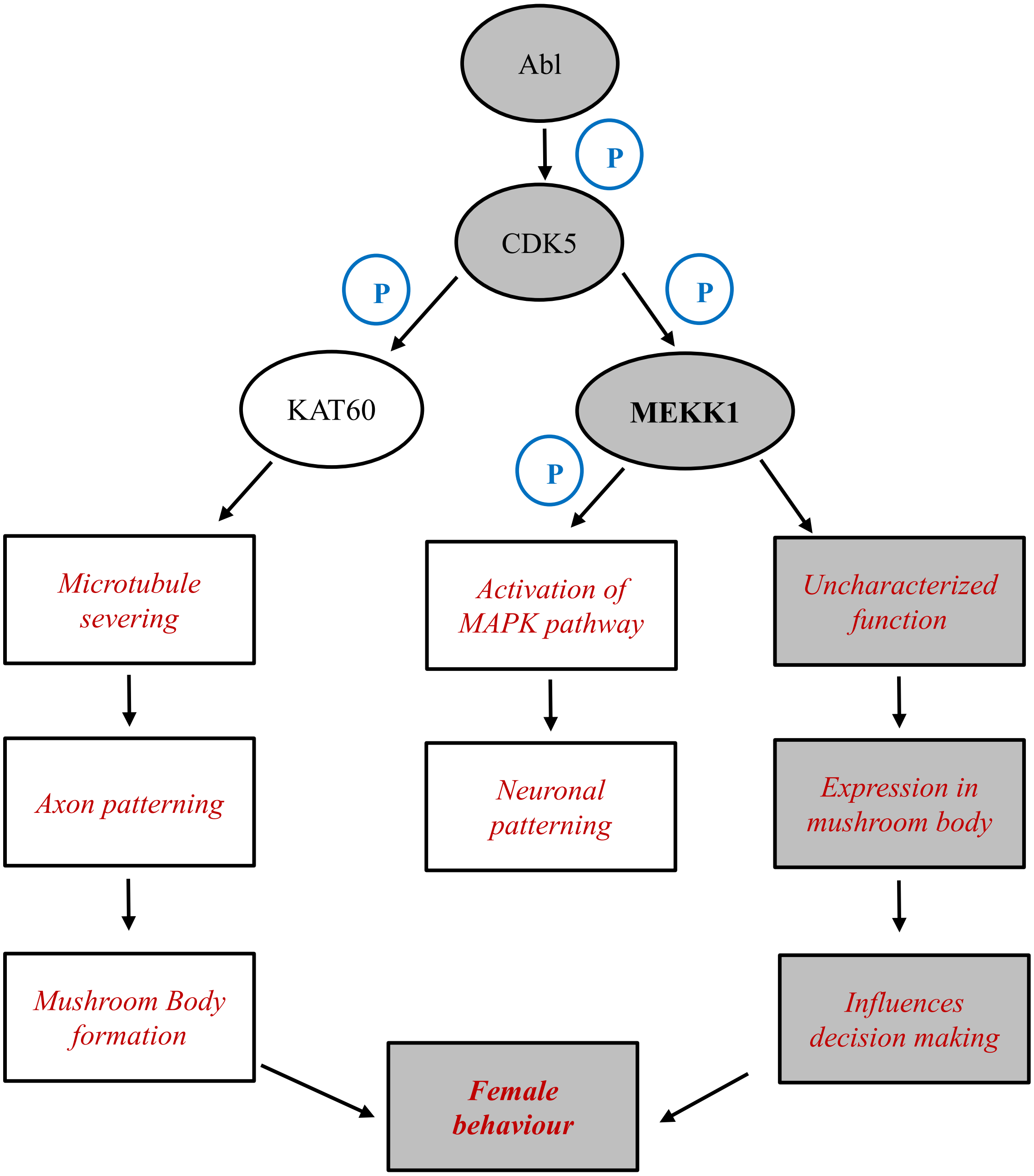


Figure 4.1. Schematic for a potential *Mekk1*-mediated pathway that influences female receptivity. Functions are indicated in red italics, proteins indicated in black, gene of interest and behaviour indicated by bolded lettering. Arrows indicate direction of mechanism of action. Blue 'P' indicates mechanism of action (phosphorylation).

upstream of *Mekk1* seem to influence whether or not a putative promoter is present in *D. simulans* (FC; section 3.3). It is possible one or a combination of all four SNVs may contribute to female rejection behaviour in *D. simulans* (FC) through differential gene regulation. To confirm that MEKK1 is indeed expressed in the mushroom body, regardless of potential expression differences between species, immunohistochemistry could be used. Finding or generating an antibody with a common immunogen between species would also decrease the cost of this assay.

Although it is unclear what cellular cascade *Mekk1* acts upon to influence behaviour, the fact that *Mekk1* may be expressed in the mushroom body indicates that it may influence decision making (Zhang et al. 2007). To determine which lobes of the mushroom body are expressing *Mekk1*, an immunohistochemistry experiment staining for FASII could be performed (Crittenden et al. 1998). If *Mekk1* and FASII are co-expressed, then *Mekk1* is expressed in the α and β lobes of the mushroom body.

A previous study showed that abolishing certain cell types in the mushroom body, a physically small portion of the head, affected the behaviour of adult *Drosophila* (de Belle & Heisenberg, 1994). This experiment specifically abolished odorant-based learning in adult flies of both sexes. Semi-quantitative RT-PCR was adequate to show that two *Mekk1* transcripts are expressed in the *Drosophila* head (Fig 3.4), however the amount of tissue expressing *Mekk1* in a manner relevant to behaviour may have been too small to detect (Bastien et al. 2008; Paiva-Cavalcanti et al. 2010). It's possible that a more sensitive technology, such as ddPCR, could determine if there are significant differences between species (Hindson et al. 2011; Hindson et al. 2013).

Further studies need to be done to characterize *Mekk1* and solidify its candidacy as a behavioural isolation gene as well as it's mechanism of action with respect to behaviour. Specifically, an experiment completing the "rescue" aspect to ensure that turning *Mekk1* back on after it has been disrupted actually rescues female rejection behaviour. To further test the hypotheses generated in Figure 4.1, interaction studies would need to be performed to confirm that CDK5 does interact in an upstream manner with both MEKK1 (product of *Mekk1*) and p60 (product of *kat60*). Flourescence resonance energy transfer (FRET) would be a potential study to examine interactions between these three proteins. Determining the signaling cascade through which *Mekk1* influences behaviour may lead to other testable candidate genes that influence behaviour, such as CDK5 (Laturney and Moehring 2012a; Ghosh et al. 2012). Understanding which genes contribute to behaviour as well as how they contribute to behaviour will provide insight into the molecular processes that underlie speciation.

References

Albert, J. T., & Göpfert, M. C. (2015). Hearing in Drosophila. *Current Opinions in Neurobiology*, *34*, 79-85.

Arnold, C. D., Gerlach, D., Spies, D., Matts, J. A., Sytnikova, Y. A., Pagani, M., Lau, N. C., & Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis - regulatory evolution. *Nature Genetics*, *46*(7), 685–692.

Arnqvist, G., Edvardsson, M., Friberg, U., Nilsson, T., & Alexander, R. D. (2000). Sexual conflict promotes speciation in insects. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(19), 10460–10464.

Arnqvist, G., & Rowe, L. (2002). Antagonistic coevolution between the sexes in a group of insects. *Letters to Nature*, *415*(6873), 787–790.

Bakker, T. C. M., & Pomiankowski, A. (1995). The genetic basis of female mate preferences. *Journal of Evolutionary Biology*, *8*(2), 129–171.

Balakireva, M., Stocker, R. F., Gendre, N., & Ferveur, J.-F. (1998). *Voila*, a new *Drosophila* courtship variant that affects the nervous system: behavioral, neural, and genetic characterization. *The Journal of Neuroscience*, *18*(11), 4335–4343.

Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature, 439(7077), 719.

Bastien, P., Procop, G. W., & Reischl, U. (2008). Quantitative Real-Time PCR is not more sensitive than conventional "PCR". *Journal of Clinical Microbiology*, *46*(6), 1897–1900.

Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y. P., Hahn, M. W., Nista, P. M., Jones, C. D., Kern, A.D., & Langley, C. H. (2007). Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*, *5*(11), 2534–2559.

Bellen, H. J., Kane, C. J. O., Wilson, C., Grossniklaus, U., Pearson, R. K., & Gehring, W. J. (1989). P-element-mediated enhancer detection : a versatile method to study development in *Drosophila*. *Genes & Development*, *3*(9), 1288–1300.

Bellen, H. J., Levis, R. W., He, Y., Carlson, J. W., Evans-Holm, M., Bae, E., Kim, J., Metaxakis, A., Savakis, C., Schulze, K. L., Hoskins, R. A., & Spradling, A. C. (2011). The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. *Genetics*, *188*(3), 731–743.

Bellen, H. J., Levis, R. W., Liao, G., He, Y., Carlson, J. W., Tsang, G., Evans-Holm, M., Heisinger, P. R., Schulze, K. L., Rubin, G. M., Hoskins, R. A., & Spradling, A. C. (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics*, *167*(2), 761–781.

Belote, J. M., & Bakert, B. S. (1987). Sexual behavior: its genetic control during development and adulthood in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(November), 8026–8030.

Bertucci, L. A., & Noor, M. A. (2001). Single fly total RNA preparations for RT-PCR. *Drosophila Information Service*, 84, 166-168.

Birkhead, T. (2000). Promiscuity: an evolutionary history of sperm competition. *Harvard University Press*. Cambridge, MA

Bischof, J., Björklund, M., Furger, E., Schertel, C., Taipale, J., & Basler, K. (2013). A versatile platform for creating a comprehensive UAS-ORFeome library in *Drosophila*. *Development (Cambridge, England)*, *140*(11), 2434–42.

Boguski, M. S., Lowe, T. M., & Tolstoshev, C. M. (1993). dbEST—database for “expressed sequence tags”. *Nature genetics*, *4*(4), 332-333.

Borgia, G. (1995). Complex male display and female choice in the spotted bowerbird: specialized functions for different bower decorations. *Animal Behaviour*, *49*, 1291–1301.

Byrne, P. G., & Rice, W. R. (2005). Remating in *Drosophila melanogaster*: An examination of the trading-up and intrinsic male-quality hypotheses. *Journal of Evolutionary Biology*, *18*(5), 1324–1331.

Calhoun, R. (2017). Genetics of female interspecific mate rejection in species of *Drosophila*. Doctor of Philosophy. *University of Western Ontario*. Department of Biology.

Chatterjee, S., & Pal, J. K. (2009). Role of 5’- and 3’-untranslated regions of mRNAs in human diseases. *Biology of the Cell*, *101*(5), 251–262.

Chintapalli, V. R., Wang, J., & Dow, J. A. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nature Genetics*, 39(6), 715.

Choudhary, M., & Singh, R. S. (1987). A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. III. Variations in genetic structure and their causes between *Drosophila melanogaster* and its sibling species *Drosophila simulans. Genetics*, *117*(4), 697–710.

Clyne, J. D., & Miesenböck, G. (2008). Sex-specific control and tuning of the pattern generator for courtship song in *Drosophila*. *Cell*, *133*(2), 354–363.

Connell-Crowley, L., Le Gall, M., Vo, D. J., & Giniger, E. (2000). The cyclin-dependent kinase Cdk5 controls multiple aspects of axon patterning *in vivo*. *Current Biology*, *10*(10), 599–602.

Connell-Crowley, L., Vo, D., Luke, L., & Giniger, E. (2007). *Drosophila* lacking the Cdk5 activator, p35, display defective axon guidance, age-dependent behavioral deficits and reduced lifespan. *Mechanisms of Development*, *124*(5), 341–349.

Coyne, J. A., & Orr, H. A. (2004). Speciation. Sunderland, MA.

Crittenden, J. R., Skoulakis, E. M. C., Han, K.-A., Kalderon, D., & Davis, R. L. (1998). Tripartite mushroom body architecture revealed by antigenic markers. *Learning & Memory*, *5*(1), 38–51.

Cutter, A. D. (2008). Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Molecular Biology and Evolution*, *25*(4), 778–786.

Darwin, C. (1859). On the origin of the species by natural selection.

Darwin, C. (1871). The descent of man.

de Belle, J. S., & Heisenberg, M. (1994). Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies. *Science*, *263*(5147), 692–695.

Dimitroff, B., Howe, K., Watson, A., Campion, B., Lee, H., Zhao, N., O'Connor, M. B., Neufeld, T. P., & Selleck, S. B. (2012). Diet and energy-sensing inputs affect TorC1-mediated axon misrouting but not TorC2-directed synapse growth in a *Drosophila Model* of tuberous sclerosis. *PLos ONE*, *7*(2), e30722.

Dobzhansky, T. (1937). Genetic nature of species differences. *The American Naturalist*, 71(735), 404-420.

Duffy, J. B. (2002). GAL4 system in *Drosophila*: a fly geneticist’s Swiss army knife. *Genesis (New York, N.Y. : 2000)*, *34*(1–2), 1–15.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *32*(5), 1792–1797.

Fischer, J. a, Giniger, E., Maniatis, T., & Ptashne, M. (1988). GAL4 activates transcription in *Drosophila. Nature*, 332(6167), 853-856.

Fisher, R. A. (1930). The genetical theory of natural selection. *Oxford University Press*. Warminster, Wilts.

Ghosh, D. K., Dasgupta, D., & Guha, A. (2012). Models, regulations, and functions of microtubule severing by Katanin. *International Scholarly Research Nnotes: Molecular Biology*, *2012*(Table 2), 1–14.

Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R., & Engels, W. R. (1991). Targeted gene replacement in *Drosophila* via *P* element-induced gap repair. *Science*, 253(5024), 1110-1117.

González, A. A. M., Cabrera, V. M., Larruga, J. M., Gullón, A., Evolution, S., & May, N. (1982). Genetic distance in the sibling species *Drosophila melanogaster*, *Drosophila simulans* and *Drosophila mauritiana*. *Evolution*, *36*(3), 517–522.

Gramates, L. S., Marygold, S. J., Santos, G. D., Urbano, J. M., Antonazzo, G., Matthews, B. B., Rey, A. J., Tabone, C. J., Crosby, M. A., Emmert, D. B., & Falls, K. (2017). FlyBase at 25: looking to the future. *Nucleic Acids research*, 45(D1), D663-D671.

Graze, R. M., McIntyre, L. M., Main, B. J., Wayne, M. L., & Nuzhdin, S. V. (2009). Regulatory divergence in *Drosophila melanogaster* and *D. simulans*, a genomewide analysis of allele-specific expression. *Genetics*, *183*(2), 547–561.

Green, K., & Tregenza, T. (2009). The influence of male ejaculates on female mate search behaviour, oviposition and longevity in crickets. *Animal Behaviour*, *77*(4), 887–892.

Hall, J. C. (1978). Courtship among males due to a male-sterile mutation in *Drosophila melanogaster*. *Behavior Genetics*, *8*(2), 125–141.

Hall, J. C. (1994). The mating of a fly. *Science*, *264*(5166), 1702–1714.

Haque, K., Pandey, A. K., Zheng, H. W., Riazuddin, S., Sha, S. H., & Puligilla, C. (2016). MEKK4 signaling regulates sensory cell development and function in the mouse inner ear. *Journal of Neuroscience*, *36*(4), 1347-1361.

Henry, G. L., Davis, F. P., Picard, S., & Eddy, S. R. (2012). Cell type-specific genomics of *Drosophila* neurons. *Nucleic Acids Research*, *40*(19), 9691–9704.

Higashi, M., Takimoto, G., & Yamamura, N. (1999). Sympatric speciation by sexual selection. *Nature*, *402*, 523–526.

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., Bright, I. J., Lucero, M. Y., Hiddessen, A. L., Legler, T. C., Kitano, T. K., & Colston, B. W. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical Chemistry*, *83*(22), 8604–8610.

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., & Hindson, B. J. (2013). Absolute quantification by droplet digitcal PCR versus analog real-time PCR. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Huang, D., Li, X., Sun, L., Huang, P., Ying, H., Wang, H., Wu. J., & Song, H. (2016). Regulation of Hippo signalling by p38 signalling. *Journal of Molecular Cell Biology*, *8*(4), 328–337.

Inoue, H., Tateno, M., Fujimura-Kamada, K., Takaesu, G., Adachi-Yamada, T., Ninomiya-Tsuji, J., Irie, K., Nishida, Y., & Matsumoto, K. (2001). A *Drosophila* MAPKKK, D-MEKK1, mediates stress responses through activation of p38 MAPK. *The EMBO Journal*, *20*(19), 5421–30.

Izquierdo JI, Carracedo MC, Pineiro R, C. P. (1992). Response to selection for increased hybridization between *Drosophila melanogaster* females and *D. simulans* males. *Journal of Heredity*, *83*(2), 100–104.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., & Carroll, S. B. (2008). The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell*, *132*, 783–793.

Katzourakis, A., Purvis, A., Azmeh, S., Rotherow, G., & Gilbert, F. (2001). Macroevlution of Hoverflies (Diptera:*Syrphidae*): the effect of the use of higher level taxa in studies of biodiversity and correlates of species richness. *Journal of Evolutionary Biology*, *14*, 219–227.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Kimura, K., Ote, M., Tazawa, T., & Yamamoto, D. (2005). Fruitless specifies sexually dimorphic neural circuitry in the *Drosophila* brain. *Letters to Nature*, *438*(November), 229–233.

Kirkpatrick, M. (1982). Sexual selection and the evolution of female choice. *Evolution*, *36*(1), 1–12.

Lachaise, D., David, J. R., Lemeunier, F., Tsacas, L., & Ashburner, M. (1986). The reproductive relationships of *Drosophila sechellia* with *D. mauritiana*, *D. simulans*, and *D. melanogaster* from the Afrotropical region. *Evolution*, 40(2), 262-271.

Laturney, M., & Moehring, A. J. (2012b). Fine-scale genetic analysis of species-specific female preference in *Drosophila simulans*. *Journal of Evolutionary Biology*, *25*(9), 1718–1731.

Laturney, M., & Moehring, A. J. (2012a). The genetic basis of female mate preference and species isolation in *Drosophila*. *International Journal of Evolutionary Biology*, *2012*, 1–13.

Levine, S., & Mullins, R. (1964). Estrogen administered neonatally affects adult sexual behavior in male and female rats, *144*(3615), 185–187.

Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji, W., McLaren, P., North, P., Rana, D., & Micklem, G. (2007). FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics. *Genome Biology*, *8*(7), R129.

Mao, C.-X., Xiong, Y., Xiong, Z., Wang, Q., Zhang, Y. Q., & Jin, S. (2014). Microtubule-severing protein Katanin regulates neuromuscular junction development and dendritic elaboration in *Drosophila*. *Development*, *141*(5), 1064–

Mayr, E. (1942). Systematics and the origin of species, from the viewpoint of a zoologist. *Harvard University Press.*

Mayr, E. (1947). Ecological factors in speciation. *Evolution*, 1(4), 263-288.

Mayr, E. (1949). Speciation and selection. *Proceedings of the American Philosophical Society*, 93(6), 514-519.

Mayr, E. (1963). Animal speciation and evolution. *Harvard University Press,* Cambridge, MA.

Meerwaldt, R., Links, T., Graaff, R., Thorpe, S. R., Baynes, J. W., Hartog, J., Gans, R., & Smit, A. (2005). Simple noninvasive measurement of skin autofluorescence. *Annals of the New York Academy of Sciences*, *1043*, 290–298.

Merrill, R. M., Van Schooten, B., Scott, J. A., & Jiggins, C. D. (2011). Pervasive genetic associations between traits causing reproductive isolation in *Heliconiu*s butterflies. *Proceedings of the Royal Society B*, *278*(September 2010), 511–518.

Metaxakis, A., Oehler, S., Klinakis, A., & Savakis, C. (2005). Minos as a genetic and genomic tool in *Drosophila melanogaster*. *Genetics*, *171*(2), 571–581.

Nègre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., Feng, X., Ahmad, K., Russell, S., Stein, L., White, K. P. (2010). A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genetics*, *6*(1), e1000814.

Nicoloso, M. S., Sun, H., Spizzo, R., Kim, H., Wickramasinghe, P., Shimizu, M., Wojcik, S. E., Ferdin, J., Kunej, T., Kiao, L., Manoukian, S., & Calin, G. A. (2010). Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Molecular and Cellular Pathobiology, 70*(7), 2789–2799.

Noor, M. A. F., & Feder, J. L. (2006). Speciation genetics: evolving approaches. *Nature Reviews Genetics, 7*(November), 851–862.

O'Dell, K. M., Armstrong, J. D., Yang, M. Y., & Kaiser, K. (1995). Functional dissection of the *Drosophila mushroom* bodies by selective feminization of genetically defined subcompartments. *Neuron*, *15*(1), 55–61.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell,. N. H., Chavali, C., Del-Toro, N., Duesbury, M., & Hermjakob, H. (2014). The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, *42*(D1), 358–363.

Orr, M. R. (1996).Life-history adaptation and reproductive isolation in a grasshopper hybrid zone. *Evolution*, 50(2), 704-716.

Paiva-Cavalcanti, M., Regis-da-silva, C., & Gomes, Y. (2010). Comparison of real-time PCR and conventional PCR for detection of Leishmania (Leishmania) infantum infection: a mini-review. *The Journal of Venomous Animals and Toxins Including Tropical Diseases*, *16*(4), 537–542.

Parker, G. A., & Partridge, L. (1998). Sexual conflict and speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1366), 261–274.

Pasyukova, E. G., Vieira, C., & Mackay, T. F. C. (2000). Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Genetics*, *156*(3), 1129–46.

Ponton, F., Chapuis, M., Pernice, M., Sword, G. A., & Simpson, S. J. (2011). Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in *Drosophila melanogaster*. *Journal of Insect Physiology*, *57*(6), 840–850.

Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Computers & chemistry*, 26(1), 51-56.

Rice, W. R. (1996). Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature*, *381*(6579), 232–234.

Ritchie, M., Halsey, E., & Gleason, J. (1999). Drosophila song as a species-specific mating signal and the behavioural importance of Kyriacou & Hall cycles in *D. melanogaster* song. *Animal Behaviour*, *58*(3), 649–657.

Rohrbough, J., Rushton, E., Palanker, L., Woodruff, E., Matthies, H. J. G., Acharya, U., Acharya, J. K., & Broadie, K. (2004). Ceramidase regulates synaptic vesicle exocytosis and trafficking. *The Journal of Neuroscience*, *24*(36), 7789–7803.

Ryan, M. J. (1998). Sexual selection, receiver biases, and the evolution of sex differences, *Science, 281*(5385), 1999–2003.

Savolainen, V., Anstett, M. C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., Powell, M. P., Springate, D., Salamin, N., & Baker, W. J. (2006). Sympatric speciation in palms on an oceanic island. *Nature*, 441(7090), 210.

Schimizzi, G. V., Maher, M. T., Loza, A. J., & Longmore, G. D. (2016). Disruption of the Cdc42/Par6/aPKC or Dlg/Scrib/Lgl polarity complex promotes epithelial proliferation via overlapping mechanisms. *PLoS ONE*, *11*(7), 1–16.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature*, *9*(7), 671–675.

Servedio, M. R. (2009). The role of linkage disequilibrium in the evolution of premating isolation. *Heredity*, *102*(1), 51–56.

Sharrocks, A. D., Yang, S. H., & Galanis, A. (2000). Docking domains and substrate-specificity determination for MAP kinases. *Trends in Biochemical Sciences*, *25*(9), 448–453.

Shen, L. X., Basilion, J. P., & Stanton, V. P. (1999). Single-Nucleotide Polymorphisms Can Cause Different Structural Folds of mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(14), 7871–7876.

Sieber, M. H., & Thummel, C. S. (2009). The DHR96 nuclear receptor controls triacylglycerol homeostasis in *Drosophila*. *Cell Metabolism*, *10*(6), 481–490.

Smale, S. T., & Baltimore, D. (1989). The "Initiator" as a transcription control element. *Cell*, *57*(1), 103–113.

Sobel, J. M., & Chen, G. F. (2014). Unification of methods for estimating the strength of reproductive isolation. *Evolution*, 68(5), 1511-1522.

Spradling, A. C., Stern, D., Beaton, A., Rhem, E. J., Laverty, T., Mozden, N., Misra, S., & Rubin, G. M. (1999). The Berkeley *Drosophila* Genome Project gene disruption project: Single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics*, *153*(1), 135–177.

Su, Y., Treisman, J. E., & Skolnik, E. Y. (1998). The *Drosophila* Ste20-related kinase misshapen is required for embryonic dorsal closure and acts through a JNK MAPK module on an evolutionarily conserved signaling pathway. *Genes & Development*, *12*(15), 2371–2380.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tasfou, K. P., Kuhn, M., & Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(D1), D447–D452.

Takekawa, M., Posas, F., & Saito, H. (1997). A human homolog of the yeast Ssk2/Ssk22 MAP kinase kinase kinases, MTK1, mediates stress-induced activation of the p38 and JNK pathways. *The EMBO Journal*, *16*(16), 4973-4982.

Tanoue, T., Maeda, R., Adachi, M., & Nishida, E. (2001). Identification of a docking groove on ERK and p38 MAP kinases that regulates the specificity of docking interactions. *The EMBO journal*, 20(3), 466-479.

Technau, G., & Heisenberg, M. (1982). Neural reorganization during metamorphosis of the corpora pedunculata in *Drosophila melanogaster*. *Nature*, *295*(5848), 405–407.

Thibault, S. T., Singer, M. a, Miyazaki, W. Y., Milash, B., Dompe, N. a, Singh, C. M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H. L., Ryner, L. A., & Margolis, J. (2004). A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. *Nature Genetics*, *36*(3), 283–7.

Uy, J. A. C., & Borgia, G. (2000). Sexual selection drives rapid divergence in Bowerbird display traits. *Evolution*, *54*(1), 273–278.

Venken, K. J. T., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., & Evans-Holm, M. (2011). MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Wheeler, D. A., Kyriacou, C. P., Greenacre, M. L., Yu, Q., Rutila, E., Rosbash, M., & Hall, J. C. (1991). Molecular transfer of a species-Specific behaviour from *Drosophila simulans* to *Drosophila melanogaster*. *Science*, *251*(4997), 1082–1085.

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, *430*(6995), 85–8.

Zhang, K., Guo, J., Peng, Y., Xi, W., & Guo, A. (2007). Dopamine-mushroom body circuit regulates saliency-based decision-making in *Drosophila*. *American Association for the Advancement of Science, 316*(5833), 1901–1904.

Zhang, Y., Cui, C., & Lai, Z.-C. (2016). The defender against apoptotic cell death 1 gene is required for tissue growth and efficient N-glycosylation in *Drosophila melanogaster*. *Developmental Biology*, *420*(1), 1–10.

Zhuang, Z., Zhou, Y., Yu, M., Silverman, N., & Ge, B. (2006). Regulation of *Drosophila* p38 activation by specific MAP2 kinase and MAP3 kinase in response to different stimuli. *Cellular Signalling*, *18*, 441–448.

Appendix A: Modifications to RNA isolation procedure.

The appropriate volume (250µL) of TRIzol Reagent (LifeTech) was added to each sample, and the sample was homogenized again for 15-20 seconds. The sample(s) were kept on ice (4°C) until each of the subsequent samples were homogenized. As the volume of tissue present for five heads is smaller than the volume of tissue in five bodies there were issues with yield. Use of a 21-gauge needle attached to a 5mL syringe via lure tip doubled the yield which enabled consistent and successful RNA isolations from head samples to be used for cDNA synthesis. The entire sample (consisting of the electric pestle-homogenized tissue and TRIzol) was flushed through the syringe three times, which takes 5-10 minutes per sample as the tissue can sometimes clog the syringe. An important point to note is that scaling the volume of TRIzol based on tissue volume (such that a differential volume was used between head and body samples) did not solve the yield issue, but altering tissue homogenization by using a syringe did. The use of columns (PureLink RNA Mini Kit – Thermo Fisher Scientific) is not advisable, specifically for isolation of RNA from fly heads, due to the fact that so much of the RNA will be lost during isolation. The sample(s) were then moved to room temperature (~22°C) and incubated for five minutes, then 50µL of chloroform was added. The samples were shaken vigorously for 20-30 seconds, and incubated for another five minutes at room temperature. The Eppendorf tube containing the sample was then centrifuged for 15 minutes at maximum speed (21,000xg) to separate the phases, and the aqueous phase was collected into a new Eppendorf tube.

The supernatant collected from TRIzol isolation (outlined above) was resuspended in 250µL nuclease-free water. Glycoblue (Ambion), a reagent consisting of glycogen covalently linked to a blue dye, was used as a RNA co-precipitant (such that 2.5µL were added to the resuspended RNA). Although the manufacturer suggests adding Glycoblue to a higher concentration than used in this protocol, this protocol succeeded in obtaining high yield and purity. The RNA was precipitated using 500µL 2-propanol (isopropanol), and incubated for 10 minutes prior to cold centrifugation (4ºC for 15 minutes at 21,000xg). The supernatant was discarded and the pellet washed loose with 250µL 75% ethanol and centrifuged at 4ºC for 15 minutes at 21,000xg. The supernatant was discarded and a pipette was used to remove excess ethanol. The pellet was resuspended in 50µL nuclease-free water.

A DNase treatment was performed using TURBO DNA-free kit (Ambion) according to the manufacturers guidelines. As the RNA was thought to be highly concentrated (and contaminated), the reagents were scaled as follows: 5µL DNase I buffer, 2.5µL DNase I, and 12µL DNase Inactivation Reagent. Once the DNase treatment was completed, the supernatant was removed by pipetting and transferred to a clean Eppendorf tube. The RNA was again precipitated using isopropanol (500µL) with 2.5µL Glycoblue, then incubated for 10 minutes at RT. The RNA was centrifuged for five minutes at 21,000xg at 4ºC and the supernatant was discarded. The pellet was washed with 1000µL 75% ethanol and again centrifuged for five minutes at 21,000xg at 4ºC. The supernatant was discarded and excess ethanol removed by pipetting. The pellet was air dried for anywhere from 5-30 minutes (depending on size/yield), and resuspended in 10µL nuclease-free water. After resuspension the RNA was incubated for five minutes at RT and gently mixed.

Appendix B: Sequence comparison between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216) for *Mekk1* gene region from upstream gene to 3’ UTR.

KEY: X = variant unique to strain FC; X = 5′ nucleotide of UTR; X = 3′ nucleotide of UTR; X = 5′ nucleotide of exon; X = 3′ nucleotide of exon. Labels above sequence illustrate coding regions (introns and intergenic regions unlabeled).

```
       1        10        20        30        40        50        60
       |        |         |         |         |         |         |
 FC    CCACCCTGTGTTCTTATCACCAGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA
BJS    CCACGCTGTGTTCTAATCACCAGGGGGATCATCATTTTAGAATCAAGTGCATATGTTATA
199    CCACACTGTGTTCTTATCACCAGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA
216    CCACACTGTGTTCTTATCACCAGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA

 FC    AATATATGTACTATCTATATGTACTTTTATTCATTTTTAATATTTGTTGAT----ATAAT
BJS    AATATTTGTACTATCTATATGTACTTTCATTTAGTTTTAATAATTTTTGGTGCTGATAAT
199    AATATATGTACTATCTATATGTACTTTTATTCATTTTTAATATTTGTTGAT----ATAAT
216    AATATATGTACTATCTATATGTACTTTTATTCATTTTTAATATTTGTTGAT----ATAAT

 FC    AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG
BJS    AATCATGCGTTCAAAATTAAATATCCCGCCCATCTGTGAATAGTGATGGACGCCTA-ACG
199    AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG
216    AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG

 FC    ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGACCTATGTTTTTAGCTATTTG
BJS    ATGGTTTACCGCTTTGGAGAGGCAGAAATAATCGGATAGACCAATGTTTTTAGCTATTTT
199    ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGACCTATGTTTTTAGCTATTTG
216    ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGACCTATGTTTTTAGCTATTTG

 FC    CAAGAATGTATTTCGAACTTTAATAAACAAAGGTGCAATGTGATGTTTAACTAATATTTT
BJS    CAAGAATGTATTTCGAACTTTGTTGAAAAAAGTTGCAATTT---GTTTAATTATTATTTT
199    CAAGAATGTATTTCGAACTTTAATAAAAAAAGGTGCAATGTGATGTTTAACTAATATTTT
216    CAAGAATGTATTTCGAACTTTAATAAAAAAAGGTGCAATGTGATGTTTAACTAATATTTT

 FC    TAGATGTTCTC--------------------------------------------TGACC
BJS    CAGATGTCCTCAATGTGTTTACCAATTTTGTTAAAAGAAATTTAAGAGCAAAGTCTAAGC
199    TAGATGTTCTC--------------------------------------------TGACG
216    TAGATGTTCTC--------------------------------------------TGACG

 FC    CAAAATTTAAATCTTTGACATACATTGTTCAGTTCTAAGCTTTTAACCAATTGTTGAACG
BJS    AAAAATTTAACTCTTCGAAATACTCTGCTAAGTT-TAAGCTTTTAACCAATTGTTGAACG
199    CAAAATTTAAATCTTTGACATACATTGCTCAGTTCTAAGCTTTTAACCAATTGTTGAACG
216    CAAAATTTAAATCTTTGACATACATTGCTCAGTTCTAAGCTTTTAACCAATTGTTGAACG

                                                       5’ UTR
 FC    AAAAGCCTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT
BJS    AAAAGTCCCAGTTTGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTT-TTTCACT
199    AAAAGTCTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT
216    AAAAGTCTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT
```

```
                               5' UTR
 FC   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG
BJS   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAGACTTTAGTTAGCTGCTTTTTTTTGG
199   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG
216   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG
                               5' UTR
 FC   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
BJS   CCCGGTAAGACGTACGACAAAGTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
199   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
216   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
                               5' UTR
 FC   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
BJS   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTGATTATCGCCAAAGGCAA
199   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
216   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
                               5' UTR
 FC   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
BJS   AACAAAAAGAAAAATAAAAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
199   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
216   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
                               5' UTR
 FC   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA
BJS   TGGATTGCGATAGTGTTTGTGCACTGGCGGCTGGCTGAGCATGTGTGCGTGTGAGCAGCA
199   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA
216   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA
                       5' UTR                             Exon 1
 FC   AATTCCAAGAATATAACAAAAGCAGAAAAAAACTGAGTTTTGCAAAAATGTCAAACAGGCG
BJS   AATTCCAAGAATATAACAAAAGCAGAAAAAAACAGAGTTTTGCAAAAATGTCAAACAGGCG
199   AATTCCAAGAATATAACAAAAGCAGAAAAAAACTGAGTTTTGCAAAAATGTCAAACAGGCG
216   AATTCCAAGAATATAACAAAAGCAGAAAAAAACTGAGTTTTGCAAAAATGTCAAACAGGCG
                               Exon 1
 FC   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
BJS   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
199   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
216   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
                               Exon 1
 FC   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA
BJS   GGCAGCAACAACAAATGCAGAGGAGCAGGTGG------CAAGGGAGGAGGAGAATGGCAA
199   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA
216   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA
                               Exon 1
 FC   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
BJS   TGGGCATCATAGTACCGTTACAGCTGAAACGCCACCCACACCACCCATCCCGCCCATTCC
199   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
216   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
```

Exon 1

```
 FC    GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
BJS    GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
199    GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
216    GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA

 FC    CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
BJS    CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
199    CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
216    CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC

 FC    ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTCAACATAT
BJS    ATTGGCACGTAACTAACAAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTTAACATAT
199    ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTCAACATAT
216    ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTCAACATAT

 FC    GTATCTCTGTCGCGGCTCGGCTCTCAAAGCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT
BJS    ----CTCTGTCGCGGCTCTGCTCTCAAAGCCATATTTGCGGGGTTTCCCTCTCATTCGCT
199    GTATCTCTGTCGCGGCTCGGCTCTCAA-GCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT
216    GTATCTCTGTCGCGGCTCGGCTCTCAAAGCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT

 FC    CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
BJS    CTCGCATTTCTTCTCTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
199    CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
216    CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA

 FC    CGTGTTTTG-GGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT
BJS    CGTGTTTTTTGACCACGTTATCTTGTTGTTGGCCTG-GGTGGCTCAACGAACCCCAGAAT
199    CGTGTTTTG-GGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT
216    CGTGTTTTG-GGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT

 FC    TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
BJS    TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
199    TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
216    TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA

 FC    AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG
BJS    AATTAAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCAAAAAATTGG
199    AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG
216    AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG

 FC    TTAACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC
BJS    ATCACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTTAGAAATTTTTGTGGAA
199    TTAACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC
216    TTAACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC

 FC    ATTTTATGAGATTGAGATTGAAA-----TCTT----------TGCTGCTTTATGTAACAA
BJS    ATTTTAGTAGATTGAGATTGAAACCAAATCTTGAAGTCTAAACGCTCCTTTATGTAACAG
199    ATTTTATGAGATTGAGATTGAAA-----TCTT----------TGCTGCTTTATGTAACAA
216    ATTTTATGAGATTGAGATTGAAA-----TCTT----------TGCTGCTTTATGTAACAA

 FC    TAAAGATTGTTTTCAGTGATCAACAAGTTTTCGCACTTACAACAGAGCGCCCTTCATT--
BJS    TAACGATTGTTTTCAGTGATCAACAAGTTTTCTCACTTACAAGAGAGCACCCTTCATAAA
199    TAAAGATTGTTTTCAGTGATCAACAAGTTTTCGCACTTACAACAGAGCGCCCTTCATT--
216    TAAAGATTGTTTTCAGTGATCAACAAGTTTTCGCACTTACAACAGAGCGCCCTTCATT--
```

```
 FC    --TTTTAAACTGATTTTAATGTGTTGAAATATTTTCAATGGAAAATCCCTAATTGCCCGT
BJS    TATTTTAAACTGATTTTAATGTGTTCAAATA---------AAAAAGCAATAATTGCTAGG
199    --TTTTAAACTGATTTTAATGTGTTGAAATATTTTCAATGGAAAATCCCTAATTGCACGT
216    --TTTTAAACTGATTTTAATGTGTTGAAATATTTTCAATGGAAAATCCCTAATTGCACGT

 FC    TATTAAATCGGGTATACAGTATAAAAACTAAGTTTCTTTATAATGCATGTTATATACGCT
BJS    TATTA-----------CATTACAAAAACTAAGTTTCTTTATAATGCATGTTATATTCGCT
199    TATTAAATCGGGTATACAGTATAAAAACTAAGTTTCTTTATAATGCATGTTATATACGCT
216    TATTAAATCGGGTATNCAGNANNAAAANTAAGTTTCTTTATNATNCANGNTANATNNNNT

 FC    CTGCGTCGTTTTTCGGAGTATTTCGGCATTTATCATAGTCATATTT--GTGTGGGGCCGC
BJS    CTGCGTCGTTTTTCGTAGTATTTCGGCATTTATCATAGTCATATTTTTGTGTGGGGCCGC
199    CTGCGTCGTTTTTCGGAGTATTTCGGCATTTATCATAGTCATATTT--GTGTGGGGCCGC
216    NNGNGTCGTTTTTNGGNNNNTTTNGGCATTTNTCATAGTCANATTT--GNGNGGGGCCGC

 FC    CTCAAGTCTTCCAAGATATTTGCTGATAACGCTGACGGAGGACTGATAACACATACACAC
BJS    CTCAAGTCTTCCAAGATATTTGCTGATAACGCTGACGGTGGACTGATAACACATACACAC
199    CTCAAGTCTTCCAAGATATTTGCTGATAACGCTGACGGTGGACTGATAACACATACACAC
216    CNCAAGTCNTNCAAGANATTTGCNGANAACGCNGNCGGTGGNCTGATAACNCATNCACAC

 FC    GCTCACCTTGAGGGAATTACGGATACGTTTTCCGCTGTTTTTTTTTTGTTGTTGTGTCTT
BJS    GCTCACCTTGAGAGAATTACGGATACGTTTTCCGCTGTTTTTTTTTTTT-----TGTCTT
199    GCTCACCTTGAGGGAATTACGGATACGTTTTCCGCTGTTTTTTTTTTGTTGTTGTGTCTT
216    NCTCNCCTTGAGGGAATTNCGGANNCGTTTTCCGCTGTTTTTTTTTTGTTGTTGTGTCTT

 FC    TTGCGCGGGGGCTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG
BJS    ATGCGCAGGGGCTTATTCATAATACACTTCATTTTTTTTGATTTGATTTTTATGGGATAG
199    TTGCGCGGGG-CTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG
216    TTGCGCGGGG-CTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG

 FC    GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA
BJS    GTGAGTGTTCA--------TCGTTTTGCGGGAAAACATGATTGTCGTATGTTAGCAGGAA
199    GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA
216    GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA

 FC    ATTTGCGGCAAATATCCATTTTCTTT-GGCACCTGTTTCTATGCCAACTAAGTGAAATTA
BJS    ATTTGCGACAAACATCCATTTTCTTTTGGCACCTGTTTCTATGGCAACTAAGTGAAATTA
199    ATTTGCGGCAAATATCCATTTTCTTT-GGCACCTGTTTCTATGCCAACTAAGTGAAATTA
216    ATTTGCGGCAAATATCCATTTTCTTT-GGCACCTGTTTCTATGCCAACTAAGTGAAATTA

 FC    GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA
BJS    GCCGAGATTTATGTGTCTGCTGCTCGCTCTTTTTTTCGTTGGTCTCT--ATATTTAATA
199    GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA
216    GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA

 FC    CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATATT
BJS    CCTGTTAATTGATAAGTTGATTGGTTCGCTGGAAAAATGGCAACGTATAGATATAATATT
199    CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATATT
216    CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATATT

 FC    A----TGTTATTTTTTT----AAAGCATCGTATA------GATACAAGCAGCGATTCTTT
BJS    ACATATGTTTTTTTTTTTATTAAAGCATCGTATATGCTAAGATACAAGCAGTGATTCTTT
199    A----TGTTATTTTTTT----AAAGCATCGTATA------GATACAAGCAGCGATTCTTT
216    A----TGTTATTTTTTT----AAAGCATCGTATA------GATACAAGCAGCGATTCTTT
```

```
 FC    AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACAGCATTTATAG
BJS    AGGAAAGGTTACGAAAAAGT-------TTTTGACTTTAAAGGTTTAAACAGCATTTATAG
199    AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACAGCATTTATAG
216    AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACAGCATTTATAG

 FC    GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTTGACTTTATCTCTCTTTTA
BJS    GGTAATCGGTAGTTCGGACATCTGTAAGCATTGTTCTCATTTGA----------------
199    GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTTGACTTTATCTCTCTTTTA
216    GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTTGACTTTATCTCTCTTTTA

 FC    ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGCAGGAAAAAAAA
BJS    -TGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGGAGGGAAGAAGA
199    ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGCAGGGAAGAAGA
216    ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGCAGGGAAGAAGA

 FC    AGAGGAGACGCAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
BJS    AGAAGAGACGGAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
199    AGAGGAGACGCAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
216    AGAGGAGACGCAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA

 FC    CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCAAA
BJS    CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCATA
199    CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCAAA
216    CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCAAA

 FC    TCTCCGTCTCATAATGTTTTGCGGTTGTCACCTGGCTAACTGAATGGAATTCCGTTGATT
BJS    TCTCCGTC-CATAATGTTTTGTGGTTGTCACCTGGCTAACTGAATGGAATTCCGATGATT
199    TCTCCGTCTCATAATGTTTTGCGGTTGTCACCTGGCTAACTGAATGGAATTCCGTTGATT
216    TCTCCGTCTCATAATGTTTTGCGGTTGTCACCTGGCTAACTGAATGGAATTCCGTTGATT

 FC    GAAGGTCTCTGGCCTAACTGTTGTTGCCTGGGGCTAAGAGAGCGAGAGATCGCGCTCTTG
BJS    GAAGGTCTCTGGCCTAACTGTTGTTGGCTTTGGTGAAGAGAGCGGGAGCTCGCGCTCTTG
199    GAAGGTCTCTGGCCTAACTGTTGTTGCCTGGGGCTAAGAGAGCGAGAGATCGCGCTCTTG
216    GAAGGTCTCTGGCCTAACTGTTGTTGCCTGGG-CTAAGAGAGCGAGAGATCGCGCTCTTG

 FC    CAACATTCGTATTTGTACTTGAACTTGCTGGTTATCTCTCTCTCTCTCGCTCTCTCTTCT
BJS    CAACATTCGTATTTGTACTTGAACTTGCAGGTTATCTTTCTCTCTC----TCTCTCTTTT
199    CAACATTCGTATTTGTACTTGAACTTGCTGGTTATCTCTCTCTCTC--GCTCTCTCTTCT
216    CAACATTCGTATTTGTACTTGAACTTGCTGGTTATCTCTCTCTCTC--GCTCTCTCTTCT

 FC    GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG
BJS    GCTCCACGTTCTAGACCTC-------------------------------AGTTCAGTTG
199    GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG
216    GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG

 FC    GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA
BJS    GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATT--------------GTACA
199    GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA
216    GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA

 FC    CACATATCAGAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC
BJS    CACATATCAGAAAACATTTCCGCATATTACACACGCACGCGCCCACA-------------
199    CACATATCAGAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC
216    CACATATCAGAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC
```

```
 FC   ACACACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAATTTTCTAATGTGTTTT
BJS   -------CACACACGTGCCACTGCCGTTTGCACTAATTTTAGAAGTTTCTAATGTGTTTT
199   ACACACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAATTTTCTAATGTGTTTT
216   ACACACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAATTTTCTAATGTGTTTT

 FC   TAAATGCCATGTTTACCGAAGAAAACAGCATAAACATGAGATTCAGAGATTGTGTATGAT
BJS   TAAATGCCATGTTTACAGAAGAAAACAGCATAAACATGAGATTCACAGATTGTGTATGAT
199   TAAATGCCATGTTTACCGAAGAAAACAGCATAAACATGAGATTCAGAGATTGTGTATGAT
216   TAAATGCCATGTTTACCGAAGAAAACAGCATAAACATGAGATTCAGAGATTGTGTATGAT

 FC   TTGTTATTTCTATTTATTTAAAA--TACAAACCTCTCAAAACCTATTCGATTACTCTTGT
BJS   TTGTTATTTCCATTTATTTAAAAATTGAAAACCTCTCAAAATCCATTTGATTACTATTTT
199   TTGTTATTTCTATTTATTTAAAA--TACAAACCTCTCAAAACCTATTCGATTACTCTTGT
216   TTGTTATTTCTATTTATTTAAAA--TACAAACCTCTCAAAACCTATTCGATTACTCTTGT

 FC   TCATTTTTGTGCCTTAAATCGGCAAAAAGTAAGCTAAGAAACTTTGAAAAATAAAACTAA
BJS   ACATTTTTAACTTTTTACTT---------TACTTTAAGAAGCATTGAAAGATAAAACTAA
199   TCATTTTCGTGCCTTAAATTGGCAAAAACTAAGCTAAGAAACATTGAAAAATAAAACTAA
216   TCATTTTCGTGCCTTAAATTGGCAAAAACTAAGCTAAGAAACATTGAAAAATAAAACTAA

 FC   AAACTATGCTTGTTT-TCCACATTTGTTTACAAGAAATCATTTTGATTGGGTGTCTGATG
BJS   AAACTATGGTTGCATATTAACATTTGTTTACATTTAATAATCTTGATTGGGTGTCTGATT
199   AAACTATGCTTGTTT-TCCACATTTGTTTACAAGAAATAATTTTGATTGGGTGTCTGATG
216   AAACTATGCTTGTTT-TCCACATTTGTTTACAAGAAATAATTTTGATTGGGTGTCTGATG

 FC   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG
BJS   GTCTGTTTCTTATTACAACTCACTTTGTTTTGGACGAACTCGTCCTATCAGTCGATCTGA
199   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG
216   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG

 FC   TAAATTGATCGGTTAGTCATCGAATTTCCAATGCCTCATACATAAATGGGGACCACGGTG
BJS   TAAATTGATCGCTTAGTCATCGAATTTCCAATGCTTCATACATAAATGGGGACCACGGTG
199   TAAATTGATCGGTTAGTCATCGAATTTCAAATGCCTCATACATAAATGGGGACCACGGTG
216   TAAATTGATCGGTTAGTCATCGAATTTCAAATGCCTCATACATAAATGGGGACCACGGTG

 FC   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
BJS   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
199   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
216   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG

 FC   ACAATAGATTTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTCGAATCTATTTTTT
BJS   ACCAAAGATTTT-CATATAGAGCCAATGAAGTTGATTGACTTGATTTGAATCGATTTTCA
199   ACAATAGATTTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTCGAATCTATTTTTT
216   ACAATAGATTTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTCGAATCTATTTTTT

 FC   TCAATCTATTTTATGCTGCTCGCTTGTAGAAAAATACATACACTCAATAGCTTAAGTAAA
BJS   T-----------ATTCTGACCGTTTGTAGAAAAATACCTACACTCAAAAGCTTAAGTAAA
199   TGAATCTATTTTATTCTGCTCGCTTGTAGAAAAATACATACACTCAATAGCTTAAGTAAA
216   TGAATCTATTTTATTCTGCTCGCTTGTAGAAAAATACATACACTCAATAGCTTAAGTAAA

 FC   CACATTGCTTATTAGCTAAC----GCGAACATAATAAAATGCCATAAATTAATAATAAAT
BJS   CACATTGCTTATTAGCCAACAAGCGGGTACATAATAAAATGCCATAAATTAATAATAAAT
199   CACATTGCTTATTAGCTAAC----GCGAACATAATAAAATGCCATAAATTAATAATAAAT
216   CACATTGCTTATTAGCTAAC----GCGAACATAATAAAATGCCATAAATTAATAATAAAT
```

```
 FC   ATTCGCGCGGTTAAGACAATTGCAAACCACAAACAAACGCAATACGTTCCACAAGCGATC
BJS   ATTCTCGCGGCTAAGACAATTGCGAACCACAAACAAACGCAATGCGCTCCAGAAACGATC
199   ATTCGCGCGGTTAAGACAATTGCAAACCACAAACAAACGCAATACGTTCCACAAGCGATC
216   ATTCGCGCGGTTAAGACAATTGCAAACCACAAACAAACGCAATACGTTCCACAAGCGATC

 FC   GAATACCCCTTTGCGTTCCACAACAAACCGTATCTCGGTGGTGCGTCTCTCTTTCTGGCT
BJS   GAATACCCCTTTGCGTTCCACAACAAACCGTATCACGGTGGTGTGTCTCTCTTTCTTGTG
199   GAATACCCCTTTGCGTTCCACAACAAACCGTATCTCGGTGGTGCGTCTCTCTTTCTGGCT
216   GAATACCCCTTTGCGTTCCACAACAAACCGTATCTCGGTGGTGCGTCTCTCTTTCTGGCT

 FC   CTCTTAGCTTCCATACCTCTCTGGCTCATCTTTTTTGCGTCTCCTCTCTCTGAGATGCCA
BJS   CTCTTAGTTTCCATACCTCTGTTGCTCATCTTATT-GCGTCTCCTCTCTCTGAGATACCA
199   CTCTTAGCTTCCATACCTCTCTGGCTCATCTTTTTTGCGTCTCCTCTCTCTGAGATGCCA
216   CTCTTAGCTTCCATACCTCTCTGGCTCATCTTTTTTGCGTCTCCTCTCTCTGAGATGCCA

 FC   CACAAAAATGGCTATAGTTTA--ATGGCAGATATTCTATGTATTGTAATCTTTAGGCTTC
BJS   TACACAAATGGCTATAGTTTATAGTGGCAGGTATTGTATTTATTGTAAACCGTAGGCTTC
199   CACAAAAATGGCTATAGTTTA--ATGGCAGATATTCTATGTATTGTAATCTTTAGGCTTC
216   CACAAAAATGGCTATAGTTTA--ATGGCAGATATTCTATGTATTGTAATCTTTAGGCTTC

 FC   ACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCCCAATCGAGTTGT
BJS   ACAATTTATATTTAATATGTTATGCATACTTTTACACTTTCTGTGCCACACTCCAGTAGT
199   ACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCACAATCGAGTTGT
216   ACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCACAATCGAGTTGT

 FC   TTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGCA
BJS   TTCTGACTGAAATGACCACTGGTAAATGTTAAACATAACTATTTAATAGCTAAACCGGCA
199   TTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGCA
216   TTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGCA

 FC   TCTCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAACCAATCGGA
BJS   TCCCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGGGCCGTAAACAATCGGA
199   TCTCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAAACAATCGGA
216   TCTCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAAACAATCGGA
                                                       5' UTR

 FC   TGCGTCGGTCGGGCGCTCGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATAG
BJS   TGCGTCGGTCGGGCGGTCGTCGCTCGCTCTTTTCAGTTCGCCGCGATTGGCAGTGAATAG
199   TGCGTCGGTCGGGCGCTCGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATAG
216   TGCGTCGGTCGGGCGCTCGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATAG
                                 5' UTR

 FC   CAGTCGTTTTGGGGGGGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCTTTAC
BJS   CAGTCGGTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGCTAC
199   CAGTCGTTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGCTAC
216   CAGTCGTTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGCTAC
                                 5' UTR

 FC   ATACAGATCCGA-TACTATACGCTTTGGCGAATAGATGGGGAATATTCTTGTTAAAAGTG
BJS   ATACAAAACCGTATACTGTACGCATTGGCGAATAGATGGGGAATATTCTTGTTAAAAGTG
199   ATACAGATCCGT-TACTATACGCTTTGGCGAATAGATGGGGAATATTCTTGT-AAAAGTG
216   ATACAGATCCGT-TACTATACGCTTTGGCGAATAGANGGGGAATATNCNNNNNNAAAGTG
```

```
                                 5’ UTR
 FC    CTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGGG
BJS    CTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGGG
199    CTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGGG
216    CTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGGG
                                 5’ UTR
 FC    TTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG-
BJS    TTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCAATTCGTCGTGATAAACCAGAA
199    TTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG-
216    TTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG-
                                 5’ UTR
 FC    -------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCACC
BJS    AACCAGCCAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCATAACCCACC
199    -------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCACC
216    -------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCACC
                                 5’ UTR
 FC    CCACCAGTGA-GATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCGG
BJS    CCACCAGTAAAGATTCCATCGATCCGAGAAGGGAATCGGCAGTCG-------CATACCGG
199    CCACCAGTGA-GATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCGG
216    CCACCAGTGA-GATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCGG
                                 5’ UTR
 FC    GGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGCT
BJS    GGATCGGTCATCCAGCCCCCAGGCCAGT------------CCAGTCATCGTTCGCTGGCT
199    GGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGCT
216    GGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGCT
                                 5’ UTR
 FC    CGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTTT
BJS    CGCGTCTAAAGTTCTC------TCCAGCTCTACCAGCTTTTAATTGTGAGCTGTCGCTTT
199    CGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTTT
216    CGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTTT
                                 5’ UTR
 FC    TTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTTG
BJS    TTATTTGCTCTCAGATCGTTCCGCCGCTGTTTGTT-TTTTGTCCCCAGAGTGGGTTGTTG
199    TTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTTG
216    TTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTTG
                                 5’ UTR
 FC    TGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGAC
BJS    TGGTGTACGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGAC
199    TGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGAC
216    TGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGAC
                                 5’ UTR
 FC    TCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCATT
BJS    TCTGAAAGTAACACTAGCCCTCGATTGCGGGGA--------AGTTTAAAACAGCGGCATT
199    TCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCATT
216    TCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCATT
```

```
                                 5' UTR
FC     ATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCACTC--CAAGGTTCTCTCTCGTCT
BJS    ATCAACCAGCCAGTTGAGAGTAAAAAAGAGGATTGCACTCTCCAAGGTTCTCTCTCGTCT
199    ATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCACTC--CAAGGTTCTCTCTCGTCT
216    ATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCACTC--CAAGGTTCTCTCTCGTCT
                                 5' UTR
FC     CGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGAT
BJS    CGCTGCGACTC-GCTGATTAGTGCCGGTTTTGGAGCTATTGGACTGCGCAAACAGT-GAT
199    CGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGAT
216    CGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGAT
             5' UTR          Exon 1 (unique to RD/RC)
FC     TCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGAACCTAGCCAATAGGCT
BJS    TGCCCAGGAGGAGGTCTGCAAAATGCGTCGGAAGAAGTGAGTGATCCAAGCCAATGGGCT
199    TCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGAACCTAGCCAATAGGCT
216    TCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGAACCTAGCCAATAGGCT

FC     GATGAAAAATGTGACGCGATTTGCACACAACAATCA--CAGCCAATTTTGGGCTTTTTTT
BJS    GATGAAAAGTGTGACGCGATTTGCACACAACAATCAAACGAGTAATTGCCAGCCAATTTT
199    GATGAAAAATGTGACGCGATTTGCACACAACAATCA--CAGCCAATTTTGGGCTTTTTTT
216    GATGAAAAATGTGACGCGATTTGCACACAACAATCA--CAGCCAATTTTGGGCTTTTTTT

FC     ---TTTT---GGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAATCGCAGGCG
BJS    GGGCTTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAATCGCAGGCG
199    ---TTTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAATCGCAGGCG
216    ---TTTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAATCGCAGGCG

FC     TGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGTCCTCCTCCC
BJS    TGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGTCCTCCTCCC
199    TGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGTCCTACTCCC
216    TGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGTCCTACTCCC


FC     CAGAGTTCGCGGCTTTCACACCCTTGGATACACCGACAAATTTCCATGTGATGCACACGA
BJS    CAGAGTTTGCCGCTTTCACTCCCTTGGATACACCGACAAATTTGTATGTGATGCACACGG
199    CAGCGTTCGCCGCTTTCACACCCTTGGATACACCGACAAATTTCCATGTGATGCACGCGA
216    CAGCGTTCGCCGCTTTCACACCCTTGGATACACCGACAAATTTCCATGTGATGCACGCGA

FC     AAAACAGCGCGAAATTTGTTTAAAAGTTGAAAGGAAAAATTTAAAAGTTAGTTCGGGATT
BJS    AGAACAGCGTGAAATTTGGTTGAAAGTTCAAATAAAAAATTTTAATGTTAGTTCGGGATA
199    AAAACAGCGCGAAATTTGGTTGAAAGTTTAAAAGAAAAATTTTAAAGTTAGTTCGGGATA
216    AAAACAGCGCGAAATTTGGTTGAAAGTTTAAAAGAAAAATTTTAAAGTTAGTTCGGGATA

FC     ACTGTTG------AATACCTTCTTGATCTCGTTGAAAACGCGATTCAAACGCTCATTGAA
BJS    ACTGTTATTTATCAACTGATAATTGATTTCGCTAAAAACTCGATTCAAACACTCATTGAT
199    ACTGTTG------AATTCCTTCTTGATCTCGTTGAAAACGCGATTCAAACGCTCATTGAA
216    ACTGTTG------AATTCCTTCTTGATCTCGTTGAAAACGCGATTCAAACGCTCATTGAA

FC      AAAGTTCTAAAAAACCAA--------------GACAACGTGAAAATTTATACATATTAAC
BJS    ATAGTCCTAAAAAAAAAGATATATTATATTGCGACAATATGAAAATTAATATACATTAAC
199    AAAGTTCTAAAAAACCAA--------------GACAACGTGGAAATTTATACATATTAAC
216    AAAGTTCTAAAAAACCAA--------------GACAACGTGGAAATTTATACATATTAAC

FC     TTAGCAATTTAATCACTTTATCGCAGCTGTAAAAAATTAT-------AACTTCCTTGTTG
```

```
BJS    TTAGCAATTTAATCACTTTATCCCAGCTGTAAAATATATTGAGTGAAAAATTCCTTGTTC
199    TTAGCAATTTAATCACTTTATCGCAGATGTAAAAAATTTT-------AACTTCCTTGTTG
216    TTAGCAATTTAATCACTTTATCGCAGCTGTAAAAAATTTT-------AACTTCCTTGTTG

 FC    AATAAACGACAATC-ATAATACTAAAGCTATTTTCTGAGTGCAGTAGGTATATTTATCTG
BJS    AATAAATTACACACCGAAATACTAAAGCTTTTTTCTGAGTGCAGTAGGTA--TTTATCTG
199    AATAAACGACACTCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN????????
216    AATAAACGACACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN?????????????

 FC    GGTTATCGGCGATTAAGCGGCACTTGTCCGCCGTTGCAATTAGCGAAACAATTAGGACAA
BJS    GGTTATCGGCGATTAAGCGGCACTTGTCCGCCGTTGCAATTAGCGAAACAATTAGGACAA
199    ??????????????????????????????????????NNNNNNGAACCAAATAGGACAA
216    ?????????????????????????????????????NNNNNNNNNNNNNNNNNNNNNNN

 FC    GCCCCTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAACGTCATCATCCA
BJS    GCCCCTAGATTAATGCAAATATACTATACTATCAACCATGACATTTAACGTCATCATCTA
199    GCCCCTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAACGTCATCATCCA
216    NNCCCTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAACGTCATCATCCA
                                                Exon 2

 FC    CTCTTTTT-----GGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCC
BJS    CTCTTTTTTTTTTGGATCTTTTAGAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCC
199    CTCTTTTT-----GGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCC
216    CTCTTTTT-----GGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCC
                                    Exon 2

 FC    AGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCGCCCGAGGATGA
BJS    AGTGCAGATGACTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCTCCCGAAGATGA
199    AGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCCCCCGAGGATGA
216    AGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCCCCCGAGGATGA
                                    Exon 2

 FC    GTTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAA
BJS    GCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAA
199    GCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAA
216    GCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAA
                              Exon 2

 FC    GAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGAGTATCCA----
BJS    GAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGAGTATCAGTACA
199    GAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGAGTATCCA----
216    GAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGAGTATCCA----

 FC    --CAATCAATACATACGGGGTGCTCCATATGGATAAGGCATCAGGATGGACGCGATAAGG
BJS    GTCAATCAGTAAATACGGTGTGCTCCATATGGATAAGGCATCAGGAAGGCCGCGATAAGG
199    --CAATCAATACATACGGGGTGCTCCATATGGATAAGGCATCAGGATGGACGCGATAAGG
216    --CAATCAATACATACGGGGTGCTCCATATGGATAAGGCATCAGGATGGACGCGATAAGG

 FC    GTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGACCCACCAAATGA
BJS    GTTTGCGTCCGTGCTAGGTACGTGATCGGGAATAGTGACGAACGCAAACTCACCAAATGA
199    GTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGACCCACCAAATGA
216    GTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGACCCACCAAATGA

 FC    CCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTATGGCTGATGTC
BJS    CCATATGCTACCATATGCCGCATCTGAT---------AGTGCCATCTCTATGGCTGATGTC
199    CCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTATGGCTGATGTC
216    CCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTATGGCTGATGTC
```

```
 FC   ATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACCTTGAAGTATTT
BJS   ATTTTCGTAATCACTGTTATTA-GAAAATGGAAATAATTTTAAATAACCTTGAAGTACTT
199   ATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACCTTGAAGTATTT
216   ATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACCTTGAAGTATTT

 FC   AATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTAGTCTGTGTAAAAATT
BJS   AATATCCATTTACAAATAAATATGTTTCAAGAATTGCCCATTTTAGACAGT-TTAAAATT
199   AATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTAGTCTGTGTAAAAATT
216   AATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTAGTCTGTGTAAAAATT

 FC   GGATTTTAATATAGATTA-AAAAATCTCATTTAATTATTTGGAAATTCATATCTTTAAAT
BJS   TGATTTTAATATAAATTACAAAGATCTCAGATAATTATTTGGAAATTAATATCCAGAAAT
199   GGATTTTAATATAGATTA-AAAAATCTCATTTAATTATTTGGAAATTCATATCTTTAAAT
216   GGATTTTAATATAGATTA-AAAAATCTCATTTAATTATTTGGAAATTCATATCTTTAAAT

 FC   TCTAAATATTTAG------AACCGCAAGAATATGCAAATAACAGTCGTCTTTTTAATCTA
BJS   ATATAGAATTTAAGGGGGTAACCTCAAGAATATGCAAATAACAGTAGTCTTTTTATTATA
199   TCTAAATATTTAG------AACCGCAAGAATATGCAAATAACAGTCGTCTTTTTAATCTA
216   TCTAAATATTTAG------AACCGCAAGAATATGCAAATAACAGTCGTCTTTTTAATCTA

 FC   ATAAAAAACATGCAAAGTTTTAGAATTCTGAGAAAAATGCATACTC--------ATAAGC
BJS   ATAAAAAA-ATGCAGATCTCCTGTGTTGTTAATGGTATCCGCTTTTTTTTTTATATTAGC
199   ATAAAAAACATGCAAAGTTTTAGAATTCTGAGAAAAATGCATACTC--------ATAAGC
216   ATAAAAAACATGCAAAGTTTTAGAATTCTGAGAAAAATGCATACTC--------ATAAGC
                                                        Exon 3

 FC   GACCACTTTCTAAACGTTTTTTCTCAGTTCCGCAGACTTGCCGGCCAGTTTGGGTGGCAC
BJS   GACCACTTTCTAAACGCTTTTCCTCAGTTCCGCAGACTTACCGGCCAGTGTGGGTGGCAC
199   GACCACTTTCTAAACGTTTTTTCTCAGTTCCGCAGACTTGCCGGCCAGTTTGGGTGGCAC
216   GACCACTTTCTAAACGTTTTTTCTCAGTTCCGCAGACTTGCCGGCCAGTTTGGGTGGCAC
                                   Exon 3

 FC   ACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATGGACTGCGC
BJS   CCCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATGGGCT-CTC
199   ACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATGGACTGCGC
216   ACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATGGACTGCGC
                                   Exon 3

 FC   CCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGGCTGCAGTC
BJS   CCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGGCTGCAGTC
199   CCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGGCTGCAGTC
216   CCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGGCTGCAGTC
                                   Exon 3

 FC   ACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCGGTGGCTGC
BJS   ACTCCTTCGCGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCGGTGGCTGC
199   ACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCGGTGGCTGC
216   ACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCGGTGGCTGC
                                   Exon 3

FC    TCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGACCAGATCAT
BJS   TCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCGCTGGAGTCTCTGGACCAGATCAT
199   TCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGACCAGATCAT
216   TCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGACCAGATCAT
```

```
                                 Exon 3
 FC    GCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAGGTGGTGGA
BJS    GCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAGGTGGTGGA
199    GCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAGGTGGTGGA
216    GCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAGGTGGTGGA
                                 Exon 3
 FC    GAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCCTCGGCGCC
BJS    GAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCCTCGGCGCC
199    GAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCCTCGGCGCC
216    GAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCCTCGGCGCC
                                 Exon 3
 FC    GCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGAAGAGCAAC
BJS    GCCGTTGGCTTTCGCTGCCCAACTTCCCGCTGGAATGCTAAGGAGTGATGGAAGAGCAAC
199    GCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGAAGAGCAAC
216    GCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGAAGAGCAAC
                                 Exon 3
 FC    GCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATTAAGCTCGG
BJS    ACCTGGGTTGGGGAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATTAAGCTCGG
199    GCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATTAAGCTCGG
216    GCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATTAAGCTCGG
                                 Exon 3
 FC    TAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAGACTGAGCT
BJS    TAGCGTAGATCGACAGGACGCCAAACTATCACAGGAGGAGCACACCTGGCAGACTGAGCT
199    TAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAGACTGAGCT
216    TAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAGACTGAGCT
                                 Exon 3
 FC    GAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTGGAGCAGCA
BJS    GAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACTGTGGAGCAGCA
199    GAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTGGAGCAGCA
216    GAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTGGAGCAGCA
                                 Exon 3
 FC    GGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCACATTATAAA
BJS    GGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCACATTATAAA
199    GGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCACATTATAAA
216    GGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCACATTATAAA
                                 Exon 3
 FC    CTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGATAGTGGCAT
BJS    CTACAAGTTCCAGCCGCGTTATCGCCGCGAGCCGAGTCTGATAAGCTTGGATAGTGGCAC
199    CTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGATAGTGGCAT
216    CTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGATAGTGGCAT
                    Exon 3
 FC    TCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCAAAAGTAAACCC
BJS    TCATTCCGATAGCAATTCTAATGCCAGCTGTATGTGAAACCATTCTTCAAAAGTAAACCC
199    TCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCAAAAGTAAACCC
216    TCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCAAAAGTAAACCC
```

```
                                                      Exon 4
 FC    CTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAGATGTGCCAAGG
BJS    CTTATTATTTATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAGATGTGCCAAGG
199    CTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAGATGTGCCAAGG
216    CTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAGATGTGCCAAGG
                               Exon 4

 FC    CTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGAGAGGT
BJS    ATGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGAGAAGT
199    CTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGAGAGGT
216    CTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGAGAGGT

                               Exon 4

 FC    AGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAGGCCAT
BJS    GGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCACAGGCCAT
199    AGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAGGCCAT
216    AGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAGGCCAT

                               Exon 4

 FC    GGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCCATGTG
BJS    GGGCGCCCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCCATGTG
199    GGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCCATGTG
216    GGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCCATGTG
                               Exon 4

 FC    CTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAAATTCT
BJS    CTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTTTAGGAAAAATTCT
199    CTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAAATTCT
216    CTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAAATTCT
                               Exon 4

 FC    GGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATCGCCTC
BJS    GGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATCGCCAC
199    GGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATCGCCTC
216    GGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATCGCCTC
                               Exon 4

 FC    CGACAGCGGTAGTTCATCAGCATCGGG-TGTAGAGAACGATGATTCTGCTGTTAACTCAA
BJS    CGACAGCGGTAGTTCATCAGCATCGGGGTGTAGAGAACGATGATTCGGCTGTTAACTCAA
199    CGACAGCGGTAGTTCATCAGCATCGGG-TGTAGAGAACGATGATTCTGCTGTTAACTCAA
216    CGACAGCGGTAGTTCATCAGCATCGGG-TGTAGAGAACGATGATTCTGCTGTTAACTCAA
                               Exon 4

 FC    TGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGACACCGT
BJS    TGGATTCATCAAAGCCGCCCAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGACACCGT
199    TGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGACACCGT
216    TGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGACACCGT
                               Exon 4

 FC    GCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATCAAGTA
BJS    GCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATCAAGTA
199    GCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATCAAGTA
216    GCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATCAAGTA
```

```
         Exon 4
FC     GCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATATAACATACATTTT
BJS    GCAATGAGTAAGTATATACGAAGTAATACGAAATTGAAGAAAAGAAATAATATACATTTT
199    GCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATATAACATACATTTT
216    GCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATATAACATACATTTT
                                  Exon 5
FC     CCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAGCCATA
BJS    CCTTCAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAGCCATA
199    CCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAGCCATA
216    CCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAGCCATA
                              Exon 5
FC     TGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTACGTGCA
BJS    TGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTACGTGCA
199    TGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTACGTGCA
216    TGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTACGTGCA
                              Exon 5
FC     GCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAATGTGC
BJS    GCAGCAATGGGACAGGAGAGCAATCCACTGGGCTGTATCGCAAGTTCATTGAAAATGTGC
199    GCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAATGTGC
216    GCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAATGTGC
                              Exon 5
FC     TCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGTGGCAT
BJS    TCAAAAGTCGAGGACTGGCCAAGTCGCTGGCCTTTCTACACAAGCTGCACAACGTGGCAT
199    TCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGTGGCAT
216    TCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGTGGCAT
                              Exon 5
FC     TGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGAGTCGG
BJS    TGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTCGATTACGAGTCGG
199    TGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGAGTCGG
216    TGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGAGTCGG
                              Exon 5
FC     ATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGAACAGG
BJS    ATGCAGAGTCCATTGAAGAAGACGTACCAAGGCTAGATCCTGAAATCAGCCGCGAACAGG
199    ATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGAACAGG
216    ATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGAACAGG
                              Exon 5
FC     TTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCTGCCGT
BJS    TTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAACCTGCCGT
199    TTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCTGCCGT
216    TTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCTGCCGT
                              Exon 5
FC     CGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGAGTTCT
BJS    CGTACATTCCCACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGAGTTCT
199    CGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGAGTTCT
216    CGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGAGTTCT
```

```
                                  Exon 5
 FC   TGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACAGTTGA
BJS   TGCGCATGCGTCTCGAAACCAGACCCGTTCGACCCAATCCACTGAGCCTGGAACAGTTGA
199   TGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACAGTTGA
216   TGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACAGTTGA
                                  Exon 5
 FC   TGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCAGCGGC
BJS   TGAAGGAGCTTAGGGAGGGCTTGACCCTAGCCTTGACCCATCGGGAGCGATACCAGCGGC
199   TGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCAGCGGC
216   TGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCAGCGGC
                                  Exon 5
 FC   ACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCATTTTAA
BJS   ACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTGGGCAGTTATATTAGCATTTTAA
199   ACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCATTTTAA
216   ACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCATTTTAA

                                  Exon 5
 FC   ACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGATCAAC
BJS   ACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTATTTGGAGTACATCGATCAAC
199   ACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGATCAAC
216   ACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGATCAAC
                                  Exon 5
 FC   TAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATGGATGT
BJS   TAGTGCTGGTTGCCGTTCCGGAAGGCAACCAGAAGTCCGTCCTTGAGAAGGAATGGATGT
199   TAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATGGATGT
216   TAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATGGATGT
                                  Exon 5
 FC   TTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAAATTCT
BJS   TTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAAATTCT
199   TTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAAATTCT
216   TTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAAATTCT
                                  Exon 5
 FC   GTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTACCGTTG
BJS   GTGGCATAATCAGCAAACTTTTGCGGAGCATTTCCGAACGCCTGGTGAAGCGCACCGTTG
199   GTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTACCGTTG
216   GTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTACCGTTG
                                  Exon 5
 FC   AGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAAATGGC
BJS   AGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAAATGGC
199   AGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAAATGGC
216   AGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAAATGGC
                                  Exon 5
 FC   AACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCGCTCCA
BJS   AGCTGCTGACCATTTGCAGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCGCTCCA
199   AACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCGCTCCA
216   AACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCGCTCCA
```

Exon 5

```
 FC   TTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTTTCATC
BJS   TTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTCGAGACGACGGATTTTCATC
199   TTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTTTCATC
216   TTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTTTCATC
```

Exon 5

```
 FC   GCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGACGTCA
BJS   GCGAGCACTATGAACATGACGTGGCCAACCAACAGCATGACTTCATCTGTTCGGACGTAA
199   GCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGACGTCA
216   GCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGACGTCA
```

Exon 5

```
 FC   AGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGACGGCAA
BJS   AGGCGGCGTTTAAACTCTTGCAGCAGGACGTTCTGCAAGTGCGCAACAAGTTGACGGCAA
199   AGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGACGGCAA
216   AGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGACGGCAA
```

Exon 5

```
 FC   TAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTGGATGAGCAAG
BJS   TAATTGAAGGAGTACAAAAACGGTGCTGTTTGAGCAACATGCGAGATTTGGATGAGCAAG
199   TAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTTTATGAGCAAG
216   TAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTGGATGAGCAAG
```

Exon 5

```
 FC   ACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATTTGGAT
BJS   ACAAACAGGCTGTGCTGTCGCGCACTCGAGAGATTTTGCATCAGGGGTACAAATTTGGAT
199   ACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATTTGGAT
216   ACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATTTGGAT
```

Exon 5

```
 FC   TCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAAGGACA
BJS   TTGAATATCACAAGGACGTCATTCGGTTGTTCGAGCAGAAGATTATGGACCAAAAGGACA
199   TCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAAGGACA
216   TCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAAGGACA
```

Exon 5

```
 FC   GCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTGGATGC
BJS   GCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTGGATGC
199   GCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTGGATGC
216   GCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTGGATGC
```

Exon 5

```
 FC   ATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTCCCAGG
BJS   ATTTCGTAATGGAGCGTTGCGAACGTGGACGAGGAATGCGTCCGCGTTGGGCCTCCCAGG
199   ATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTCCCAGG
216   ATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTCCCAGG
```

Exon 5

```
 FC   GTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGACGACG
BJS   GTCTGGAGTTTCTGATTCTTGCCTGTGATCCACAAATCACCCAGCACTTGGATGATGACG
199   GTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGACGACG
216   GTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGACGACG
```

```
                                   Exon 5
 FC    AGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCATCACTT
BJS    AGTTTGAGGCGCTAAAGCAGCAAATGGATCGCTGTATTTCGCACGTGATTGGCATCACCT
199    AGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCATCACTT
216    AGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCATCACTT
                                   Exon 5
 FC    CGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTCATCGC
BJS    CGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGCAAGACTTCATCGC
199    CGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTCATCGC
216    CGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTCATCGC
                                   Exon 5
 FC    CGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGTTCTTA
BJS    CGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGTCCTTA
199    CGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGTTCTTA
216    CGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGTTCTTA
                                   Exon 5
 FC    ATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCAGTTCA
BJS    ATCCTAATACGCCGCCACTGCAGTCGCCACCGTACAACAAGTTACTGCATCCGCAGTTCA
199    ATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCAGTTCA
216    ATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCAGTTCA
                                   Exon 5
 FC    GTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTATGTCG
BJS    GCTTAAAGGAGGATGTGTCAGGAACTTCGTACAGTCCCGTTGACAGTTCAGACTATGTTG
199    GTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTATGTCG
216    GTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTATGTCG
                                   Exon 5
 FC    ACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGACGCCTC
BJS    ACACTCCGTGCCAAAGGAGTGCCAACGGCGAGCTGCGTCTGCTGGTGCCCCAGACGCCGC
199    ACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGACGCCTC
216    ACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGACGCCTC
                                   Exon 5
 FC    CGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCGACAGG
BJS    CGACTCCAGCATCTCCCGGAAAGAGCAGCCTCGAAAGTACACCGCTGGCCTTGCGGCAGG
199    CGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCGACAGG
216    CGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCGACAGG
                                   Exon 5
 FC    AACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCGCGAAC
BJS    AACGTGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTACGCGAAC
199    AACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCGCGAAC
216    AACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCGCGAAC
                                   Exon 5
 FC    GAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAGGGCGC
BJS    GAAGGTTGATTGGGCAGGTCAAGTCTCTTAATTCCAGCGACAAGGTGCATATAAGGGCGC
199    GAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAGGGCGC
216    GAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAGGGCGC
```

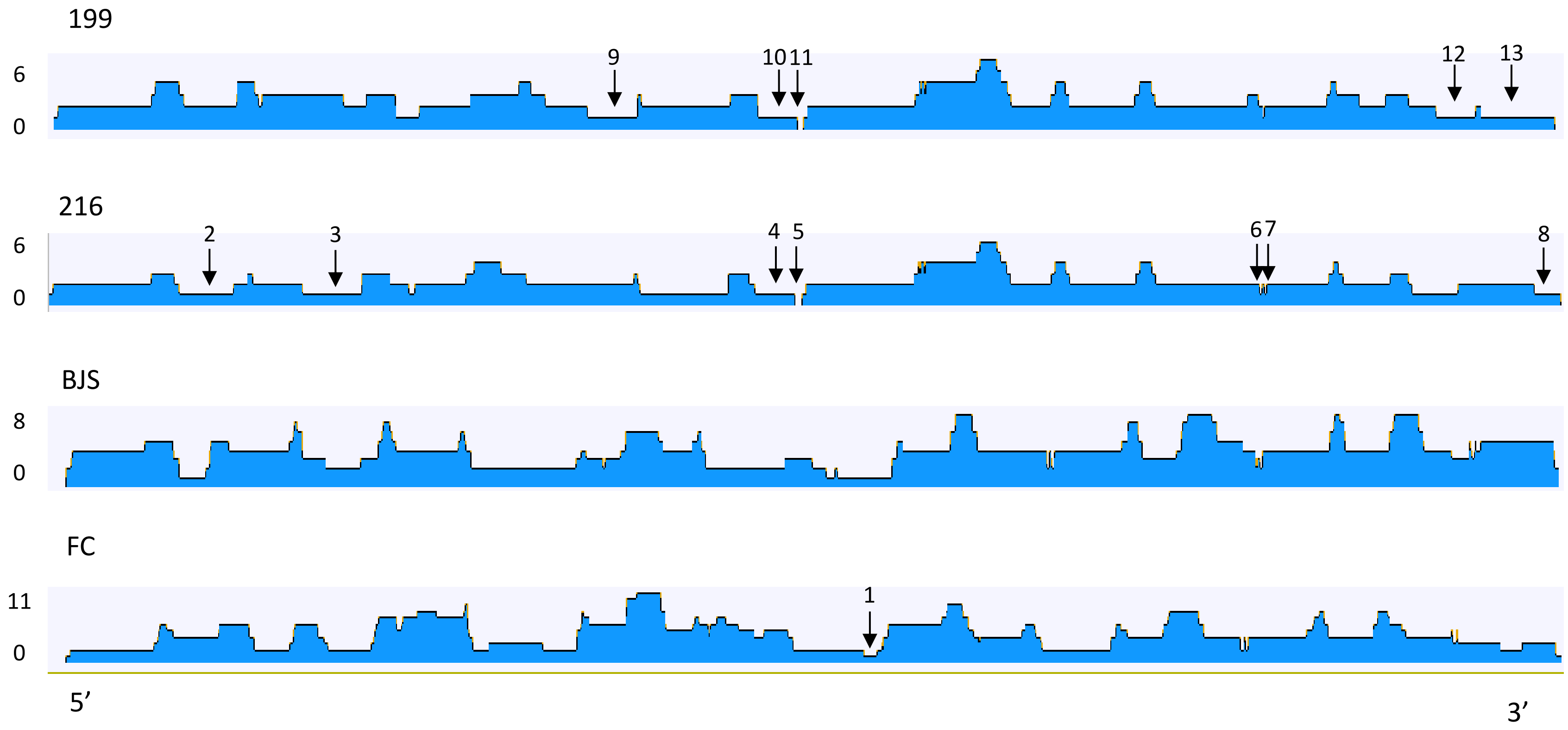
```
                                    Exon 5
 FC    GTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGGCAAGG
BJS    GTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGGCAAGG
199    GTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGGCAAGG
216    GTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGGCAAGG
                                    Exon 5
 FC    TGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGCAATCC
BJS    TATACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGCAATCC
199    TGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGCAATCC
216    TGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGCAATCC
                                    Exon 5
 FC    AGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGAAGGAA
BJS    AGCCGGGAGAGACGCGAGCACTCAAGAATGTGGCCGAAGAGCTAAAGATCTTGGAAGGAA
199    AGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGAAGGAA
216    AGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGAAGGAA
                                    Exon 5
 FC    TAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCTGCTCA
BJS    TAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAGGAGCTGCTTA
199    TAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCTGCTCA
216    TAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCTGCTCA
                                    Exon 5
 FC    TTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGGTAATC
BJS    TTTTCATGGAGCTGTGCTCTGAGGGCACCCTTGAGTCACTGGTGGAGCTGACTGGTAATC
199    TTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGGTAATC
216    TTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGGTAATC
                                    Exon 5
 FC    TGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGAGCTGC
BJS    TGCCGGAGGCGCTTACGCGACGTTTCACCGCCCAGCTGTTGTCGGGCGTGTCTGAGCTGC
199    TGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGAGCTGC
216    TGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGAGCTGC
                                    Exon 5
 FC    ACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGACGGCA
BJS    ACAAGCATGGCATTGTGCATCGCGACATCAAGACTGCTAACATATTCCTCGTGGACGGCA
199    ACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGACGGCA
216    ACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGACGGCA
                                    Exon 5
 FC    GCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACACCACCG
BJS    GCAACAGCCTAAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAAGCGCACACCACTG
199    GCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACACCACCG
216    GCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACACCACCG
                                    Exon 5
 FC    TGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGTGTTCA
BJS    TGCCCGGCGAGCTGCAGGGCTATGTAGGCACGCAGGCCTATATGGCGCCTGAGGTTTTCA
199    TGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGTGTTCA
216    TGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGTGTTCA
```

```
                                 Exon 5
 FC    CAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGG
BJS    CAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGG
199    CAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGG
216    CAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGG
                 Exon 5

 FC    TTGTAGAGATGGCCTCGGGCAAGGTGAGTGCTTCACATTGTAGATAAGCAATTCGATGTA
BJS    TTGTCGAGATGGCCTCGGGCAAGGTGAGTGCTTCGCATTGTAGATAAGCAATTCGATGTA
199    TTGTAGAGANGGCCTCGGGCAAGNTGAGTGCTTCNCATTGTAGANAAGCAATGCNATGTG
216    TTGTAGAGATGGCCTCGGGCAAGGTGAGTGCTTCACATTGTAGATAAGCAATGCGATGTG

 FC    TGGGGCGGGGTATACAGTCCATTTCACAATGTTGGATTATTAGTGCTTTTCTTTGGGTGA
BJS    TGGGGCGGGGTATACAGTCCATTTCACAATGTTGGATTATTAGTGCTTTTCTTTGGGTGA
199    TGGGGCGGGNNATACAGTCCATTTCACAATGTTGGATTATTANTGCTTTTCTTTGGNTGA
216    TGGGGCGGGGTATACAGTCCATTTCACAATGTTGGATTATTAGTGCTTTTCTTTGGGTGA

 FC    ATATTTTGCATTATACGCAGTCTAAAGTAATTTACACTAATTTTATTTTTTGAATGAATG
BJS    A-----------------------------------------------------------
199    ATATTTTGCATTATACGCAGNNTAANGNAATTTACANCTANNTTNATTTTTTGAATGAATG
216    ATATTTTGCATTATACGCAGTCTAAAGTAATTTACACTAATTTTATTTTTTGAATGAATG

 FC    CAGTACTCTTTTACTCTTTTCCTTTTAATACTAAATTTGTATAAAAATATTAAAATATTT
BJS    ----------------------------------------------ATTTTGAAATGTTT
199    CAGTACTCTTTTACTCTTTTCCTTTTAATACTAAATTTGTATAAAAATATTAAAATATTT
216    CAGTACTCTTTTACTCTTTTCCTTTTAATACTAAATTTGTATAAAAATATTAAAATATTT

 FC    ACACTTAATGAAAATACTGTGAATTAGGCTACTAAAACCTTTGTTAAATATGTGTAATAT
BJS    A-----------------------------------ACATTTAATATATAT---------
199    ACACTTAATGAAAATACTGTGAATTAGGCTACTAAAACCTTTGTTAAATATGTGTAATAT
216    ACACTTAATGAAAATACTGTGAATTAGGCTACTAAAACCTTTGTTAAATATGTGTAATAT
                                             Exon 6

 FC    TATTTAGTTTTGGTTTCTTTACAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCAGATC
BJS    -------TTATGGTTTCTTTACAGCGTCCTTGGGCCCAGTTTGATTCCAACTTCCAGATC
199    TATTTAGTTTTGGTTTCTTTACAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCAGATC
216    TATTTAGTTTTGGTTTCTTTACAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCAGATC
                                 Exon 6

 FC    ATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGAGGGT
BJS    ATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCACCGGAGAGCCTATCCCAGGAGGGT
199    ATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGAGGGT
216    ATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGAGGGT
                                 Exon 6

 FC    CACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGTGGAA
BJS    CACGACTTCATCGATCATTGTCTGCAGCATGATCCCAAGAGGCGTTTAACGGCAGTGGAA
199    CACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGTGGAA
216    CACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGTGGAA
                Exon 6

 FC    CTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCCCGCTGCCAAATGTT
BJS    CTGTTGGAGCACAATTTCTGCAAGGTGCGTATGCCTGTGGTTAGCTCGCTGCCAAATGTT
199    CTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCCCGCTGCCAAATGTT
216    CTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCCCGCTGCCAAATGTT
```

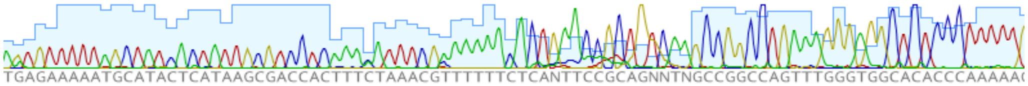
```
                                                Exon 7
FC    GCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTT
BJS   GCTAATTGATTTTCGTTTACATGTGCAGTACGGTCGAGACGAGTGCAGCAGTGAGCAGTT
199   GCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTT
216   GCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTT
                              Exon 7                      3’ UTR
                                          F
FC    GCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAGATGCA
BJS   GCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAGGTGCT
199   GCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAGATGCA
216   GCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAGATGCA
                                3’ UTR

FC    ACGTGGTGTCTACTTAACTAGCACCCGGTTTTTTTT-AGGCCCACTCCTACATATGATGC
BJS   ACGTGGTGTCTACTTAACTAGCTTGCGGTTTTTTTTTAGGCGCACCCCTACATATGATGC
199   ACGTGGTGTCTACTTAACTAGCACCCGGTTTTTTTT-AGGCCCACTCCTACATATGATGC
216   ACGTGGTGTCTACTTAACTAGCACCCGGTTTTTTTT-AGGCCCACTCCTACATATGATGC
                3’ UTR

FC    ATATGTACATCGGTGTGCATTTTGTTTAG
BJS   ATATGTACATCGGTGTGGATTTTGTTTAG
199   ATATGTACATCGGTGTGCATTTTGTTTAG
216   ATATGTACATCGGTGTGCATTTTGTTTAG
```

Appendix C: Coverage map of the data used to generate the *Mekk1* consensus sequence for each of the four strains, including electropherograms in areas of uncertainty.

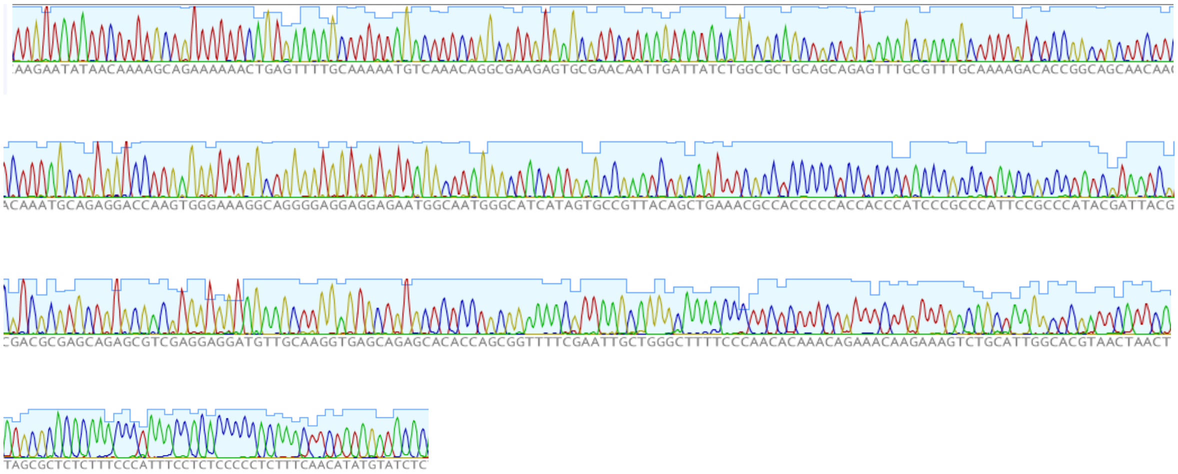
The strain is indicated above the coverage map on the left; numbers on the left side of the map indicate the number of reads; gene orientation (5’ → 3’) indicated at the bottom; numbers with each arrow correspond to electropherograms or sequences of the region the arrow points to which will follow the coverage map. Electropherograms described by number of N per 100bp average for the region.



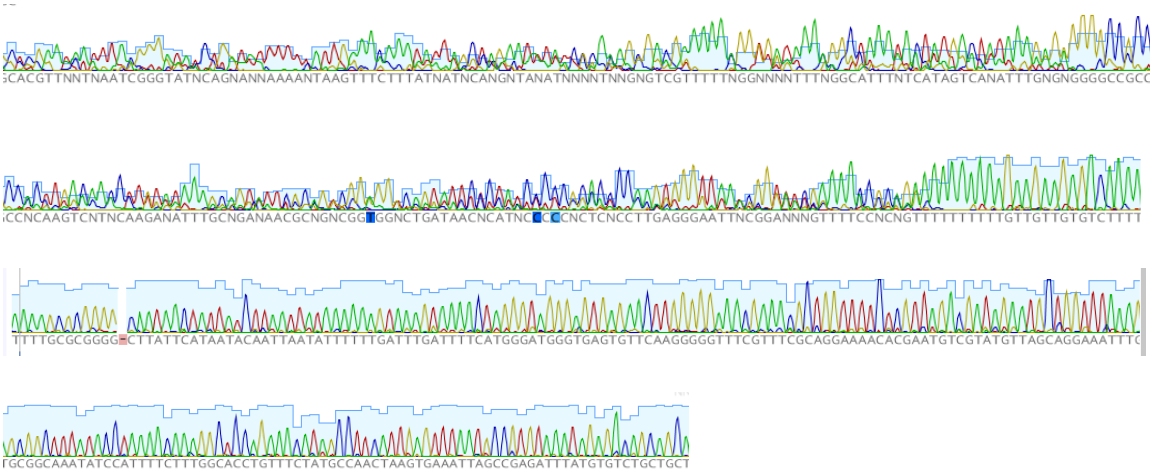
Region 1 (Line FC): Less than 3 N / 100 bp



Region 2 (Line 216): Less than 3 N / 100 bp



Region 3 (Line 216): 11 N / 100 bp



Region 3 edited sequence:

ACGTTATTAAATCGGGTATACAGTATAAAAACTAAGTTTCTTTATAATGCATGTTATATACGCTCTGCGTCGTTT

TTCGGAGTATTTCGGCATTTATCATAGTCATATTTGTGTGGGGCCGCCTCAAGTCTTCCAAGATATTTGCTGAT

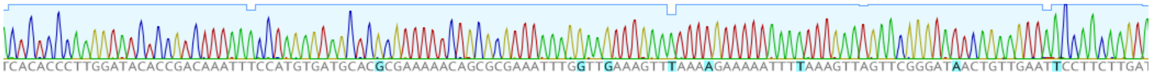
AACGCTGACGGTGGACTGATAACACATACACACGCTCACCTTGAGGGAATTACGGATACGTTTTCCGCTGTTT

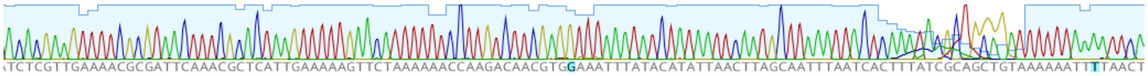
TTTTTTTGTTGTTGTGTCTTTTGCGCGGGGCTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGA

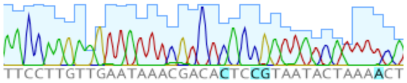
TGGGTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAAATTTGCGGC

AAATATCCATTTTCTTTGGCACCTGTTTCTATGCCAACTAAGTGAAATTAGCCGAGATTTATGTGTCTGCTGCTT

Region 4 (Line 216): Less than 3 N / 100bp







Region 5 (Line 216): Unable to be sequenced. Sequences of reads for that region:

Read 1:

NNNNNNNNNNNNNNNNNNNNNNNATTTTGGGCTTTTTTTTTTTTTGGGNGGGNAAAANNNNNNCCNNN

GNNNNNNNNNNNNNNNNNNNNNNGGGGGGNCNNGCGGNNNNNNGNNNNNCCGNNGGNNNNNNAAN

AAAANGGGGNGGNCNNNNCCCCCGGGNNNNNNNNNTTNNNCCCNNNNGGNNNNCCCNNNNNNTNCCC

NGGNGNNGNNNNNNAAAANCNNNNNNNNNTTGGGTTNANNGTTTAAAAAAAAATTTTNAANGNNNNNN

NGGNNNNNNGGNNNNNNNCNNNNNNGNNCNNNNNNAAANNNNNNNNNNNANNNNNNNNNNNAAAAN

NNNNNAAAANNCNNNNNNNNNNNGGNANNTTNNTCCTTTTNNNNNNNNNATTTNNNNNNNNTTNTNNCCA

GNNNANANANANTNNNCCTNCCTTGTTAAAAAAANANCCNTCNNAAANNNAAAAANTATNTTNGNAGGG

NNNNGGGNTTNTTTTTNGGGGTTACCGGCNNNAAACGNCCCTTGGCCCCCCGTTGCATTTAGCAAACCATTT

AGGACAAGCCCCTAGTTTACGGCACATATCCTNNNCNNNNNNNNNNNNNNTTNNANNNNCCNNNNNNNNN

CTTTNNNNNNNNNNNNNNNNNNNNNNCCCGNNNNNNNNNNNNNCNTNNNNNNCNNNNNNNNGNNTGN

AAANNNNNTGNNNNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNN

NNNNNNNN

Read 2:

NCNNNNNANNNNNNNNNNNNCCNNNNAAANANNNNNNNNNNNNNNNNNNNAANNNNNTNNNNNNNN

NNGNGNGTNNNNTNNNNNNNNNNNNNNNNNNNAAANNTTTTNCNNNNNNNNNAANNNNNNNNNAANNN

NNNNNNNNANNNNNNNNAAANNTNCNCNNNNNNNNNGNTTTTTNNAANTTTTTNAANNNNNNNNTNNAA

NNNNGTTTTNNNNGNGNNCCNNNNNGNNNNNNNCNNNNNNCCCNNNNNNNCNNTAAAANTTTTTTTTT

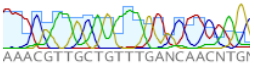
NAACNNTCNNNCNNNNTNNNNNNNNNTTTTTNNNGNNCNNCNCCNGGNAANTTNNNGGNGNNNNCNNN

GGGGNGAAANNNNNNNNCNCNGGGGNNNNGNNCNNNCCNNTNNNNTNNNNNNNNNNCCGGNNNNN

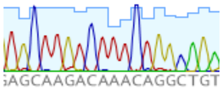
CNNNCANNCNNNNNNCNNNCCNNNGNNNNNNNNNNNNNNNNNNNCNNNGGTTTNNNTTNNNNCNNNC

NNNAAAAAAAAAAAAGNCCAAAATNGNNNNNNNNNNNNNNNNNNNN

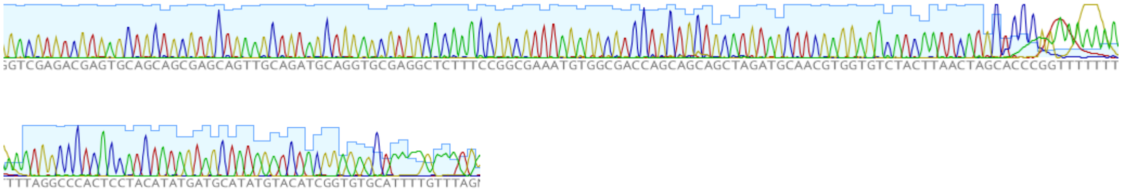
Region 6 (Line 216): Less than 3 N / 100 bp



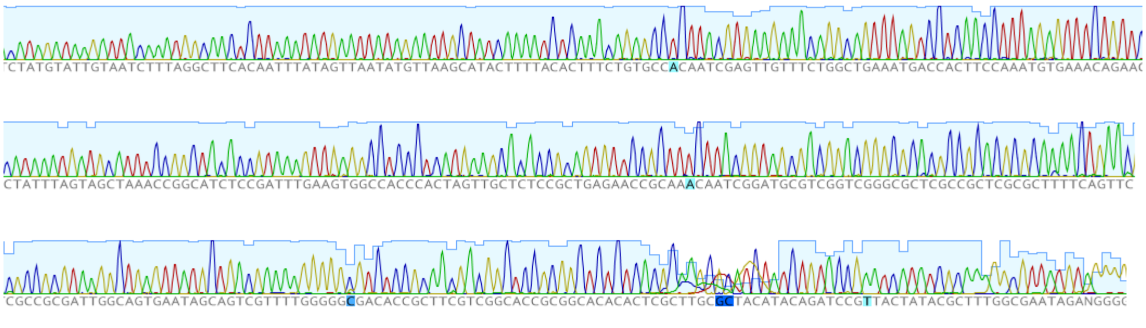
Region 7 (Line 216): Less than 3 N / 100 bp

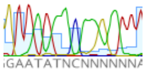


Region 8 (Line 216): Less than 3 N / 100 bp



Region 9 (Line 199): 8 N / 100bp





Region 9 edited sequence:

TATGTATTGTAATCTTTAGGCTTCACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCACAAT

CGAGTTGTTTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGCATCTCC

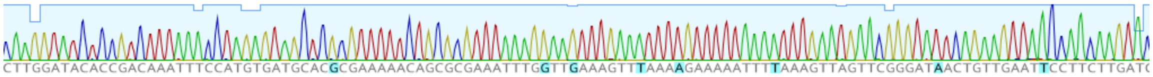
GATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAAACAATCGGATGCGTCGGTCGGGCGCT

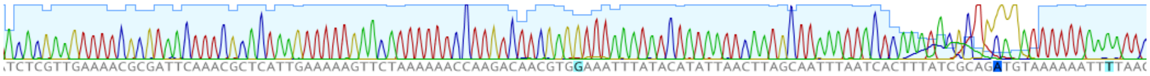
CGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATAGCAGTCGTTTTGGGGGCGACACCGCTTCGTC

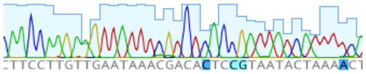
GGCACCGCGGCACACACTCGCTTGCGCTACATACAGATCCGTTACTATACGCTTTGGCGAATAGATGGGGAAT

ATTCTTGTAA

Region 10 (Line 199): Less than 3 N / 100bp







Region 11 (Line 199): Unable to be sequenced. Sequences of reads for that region:

Read 1:

NNNNNNNNNNNNNNNNNNNNNNNATTTTGGGCTTTTTTTTTTTTTGGGNGGGNAAAANNNNNNCCNNN

GNNNNNNNNNNNNNNNNNNNNNGGGGGGNCNNGCGGNNNNNNGNNNNNCCGNNGGNNNNNNAAN

AAAANGGGGNGGNCNNNNCCCCCGGGNNNNNNNNNTTNNNCCCNNNNGGNNNNCCCNNNNNNTNCCC

NGGNGNNGNNNNNNAAAANCNNNNNNNNNTTGGGTTNANNGTTTAAAAAAAAATTTTNAANGNNNNNN

NGGNNNNNNGGNNNNNNNCNNNNNNGNNCNNNNNNAAANNNNNNNNNNANNNNNNNNNNAAAAN

NNNNNAAAANNCNNNNNNNNNNGGNANNTTNNTCCTTTTNNNNNNNNNATTTNNNNNNNTTNTNNCCA

GNNNANANANANTNNNCCTNCCTTGTTAAAAAAANANCCNTCNNAAANNNAAAAANTATNTTNGNAGGG

NNNNGGGNTTNTTTTTNGGGGTTACCGGCNNNAAACGNCCCTTGGCCCCCCGTTGCATTTAGCAAACCATTT

AGGACAAGCCCCTAGTTTACGGCACATATCCTNNNCNNNNNNNNNNNNNNTTNNANNNNCCNNNNNNNNN

CTTTNNNNNNNNNNNNNNNNNNNNNCCCGNNNNNNNNNNNNNCNTNNNNNNCNNNNNNNNGNNTGN

AAANNNNNTGNNNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNN

NNNNNNNN

Read 2:

NNNNNNNNNNNNNNNNNNNGNNNNTTTGGNCTTTTTTTTTTTNNNNNNNGNNNNANNNNAANCCNNNG

NNNNNNNNNNNNNNNNNNCNNNGGNGNGNNNNGNGNNTGNNNGNNNNNCCGGNNGNNNNNNNAN

AANANGGGNNNNNNNNNNNNCCCCNNNNNNCNCNNNTTTCNCCCCNTNGNNNNCNCCNNNNANNTNCN

NGGNGNNGNNCNNNAAAAACNNNNNNNANTTNGNNNGANNNTTNAAAAAAAAANTTTNAANNNNNNN

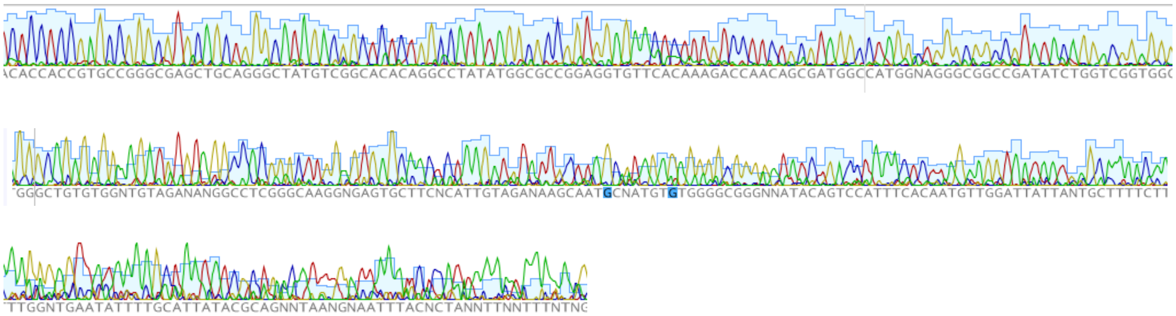
NNGGGNNNNCNGNNNNNNNCCNNNNNGNNCCCNNNNAAAACNNNNTNNAANNNNNNNTTNAAAAAN

TTNNAAAANNCNNNNNNNNNGGGNANNTTNNNNCNNNTNNNNTNNNNNNTTNNNNCNNTTNNNCNNN

NNNNNNAANANTTNNNNNNNNNNNNNNNNNNNANNNNNNNNCCNNNNNNNNNAAANNTTTTTTNNNN

NNNNNNNNGNNNNNTTTNNNNGGNNNNNNNGNNATNANNNNN

Region 12 (Line 199): 8 N / 100bp



Region 12 edited sequence:

CACCACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGTGTTCACAA

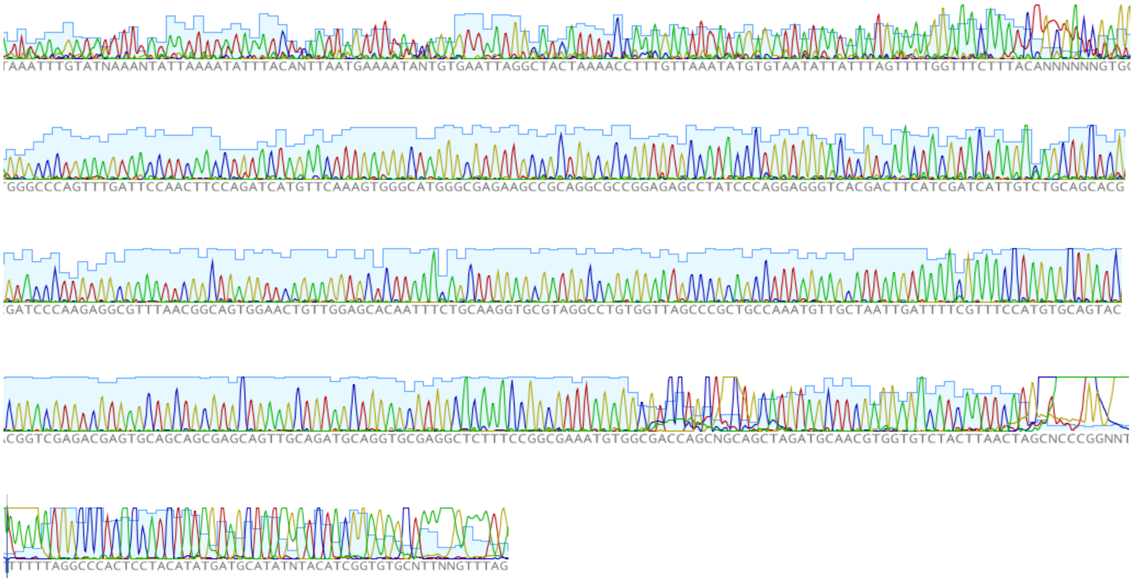
AGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGGTTGTAGAGATGGCC

TCGGGCAAGGTGAGTGCTTCACATTGTAGATAAGCAATGCGATGTGTGGGGCGGGGTATACAGTCCATTTCA

CAATGTTGGATTATTAGTGCTTTTCTTTGGGTGAATATTTTGCATTATACGCAGTCTAAAGTAATTTACACTAAT

TTTATTTTTT

Region 13 (Line 199): Less than 3 N / 100bp



Appendix D: Thesis data compared to unedited independent re-analysis.

The sequences from this thesis are directly compared in the same manner as Appendix B to sequences compiled by an independent re-analysis by a third-party. The independent re-analysis was missing some files (corresponding to the promoter regions), and the re-analysis was completely unedited. Editing bases that the program cannot interpret as the peak in the electropherogram sometimes does not meet the program’s threshold for detection. The independent analysis did not use the same software nor the same process of alignment as was used to generate the data presented throughout this thesis.

```
         1        10        20        30        40        50        60
         |        |         |         |         |         |         |
 FC_CD   CCACCCTGTGTTCTTATCACCAGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA
BJS_CD   CCACGCTGTGTTCTAATCACCAGGGGGATCATCATTTTAGAATCAAGTGCATATGTTATA
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_CD   CCACACTGTGTTCTTATCACCAGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA
199_IR   CCACACTGTGTTCTTATCACCAGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA
216_CD   CCACACTGTGTTCTTATCACCAGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA
216_IR   CCACACTGTGTTCTTATCACCAGGGGGATCATAAGTTCAGAATCAAGCGCATGTGTTATA

 FC_CD   AATATATGTACTATCTATATGTACTTTTATTCATTTTTAATATTTGTTGAT----ATAAT
 FC_IR   AATATATGTACTATCTATATGTACTTTTATTCATTTTNNATATTTGTTGAT----ATAAT
BJS_CD   AATATTTGTACTATCTATATGTACTTTCATTTAGTTTTAATAATTTTTGGTGCTGATAAT
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_CD   AATATATGTACTATCTATATGTACTTTTATTCATTTTTAATATTTGTTGAT----ATAAT
199_IR   AATATATGTACTATCTATATGTACTTTTATTCATTTTTAATATTTGTTGAT----ATAAT
216_CD   AATATATGTACTATCTATATGTACTTTTATTCATTTTTAATATTTGTTGAT----ATAAT
216_IR   AATATATGTACTATCTATATGTACTTTTATTCATTTTTAATATTTGTTGAT----ATAAT

 FC_CD   AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG
 FC_IR   AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG
BJS_CD   AATCATGCGTTCAAAATTAAATATCCCGCCCATCTGTGAATAGTGATGGACGCCTA-ACG
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_CD   AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG
199_IR   AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG
216_CD   AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG
216_IR   AATCATAAGTTCAAAATCAAATATCCCGCCCATCTGTTACTAGTGATGGACGCCTATACG

 FC_CD   ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGACCTATGTTTTTAGCTATTTG
 FC_IR   ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGNCNTATGTTTTTAGCTATTTG
BJS_CD   ATGGTTTACCGCTTTGGAGAGGCAGAAATAATCGGATAGACCAATGTTTTTAGCTATTTT
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_CD   ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGACCTATGTTTTTAGCTATTTG
199_IR   ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGACCTATGTTTTTAGCTATTTG
216_CD   ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGACCTATGTTTTTAGCTATTTG
216_IR   ATGGCTTAACGGTTTGTAGAAGCAGAAATAATCGGATAGACCTATGTTTTTAGCTATTTG

 FC_CD   CAAGAATGTATTTCGAACTTTAATAAACAAAGGTGCAATGTGATGTTTAACTAATATTTT
 FC_IR   CAAGNATGTNTTTCGAACTTTAATAAANAAAGGNNCAANNNGATGNTTAANTAATATTTT
BJS_CD   CAAGAATGTATTTCGAACTTTGTTGAAAAAAGTTGCAATT---TGTTTAATTATTATTTT
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_CD   CAAGAATGTATTTCGAACTTTAATAAAAAAAGGTGCAATGTGATGTTTAACTAATATTTT
199_IR   CAAGAATGTATTTCGAACTTTAATAAAAAAAGGTGCAATGTGATGTTTAACTAATATTTT
216_CD   CAAGAATGTATTTCGAACTTTAATAAAAAAAGGTGCAATGTGATGTTTAACTAATATTTT
216_IR   CAAGAATGTATTTCGAACTTTAATAAAAAAAGGTGCAATGTGATGTTTAACTAATATTTT
```

```
 FC_CD   TAGATGTTC--------------------------------------------TCTGACC
 FC_IR   TAGATGTTC--------------------------------------------TCTGANN
BJS_CD   CAGATGTCCTCAATGTGTTTACCAATTTTGTTAAAAGAAATTTAAGAGCAAAGTCTAAGC
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_CD   TAGATGTTC--------------------------------------------TCTGACG
199_IR   TAGATGTTC--------------------------------------------TCTGACG
216_CD   TAGATGTTC--------------------------------------------TCTGACG
216_IR   TAGATGTTC--------------------------------------------TCTGACG

 FC_CD   CAAAATTTAAATCTTTGACATACATTGTTCAGTTCTAAGCTTTTAACCAATTGTTGAACG
 FC_IR   CAAAATTTAAATCTTTGACATACATTGTTCAGTTCTAAGCTTTTAACCAATTGTTGAACG
BJS_CD   AAAAATTTAACTCTTCGAAATACTCTGCTAAGTT-TAAGCTTTTAACCAATTGTTGAACG
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_CD   CAAAATTTAAATCTTTGACATACATTGCTCAGTTCTAAGCTTTTAACCAATTGTTGAACG
199_IR   CAAAATTTAAATCTTTGACATACATTGCTCAGTTCTAAGCTTTTAACCAATTGTTGAACG
216_CD   CAAAATTTAAATCTTTGACATACATTGCTCAGTTCTAAGCTTTTAACCAATTGTTGAACG
216_IR   CAAAATTTAAATCTTTGACATACATTGCTCAGTTCTAAGCTTTTAACCAATTGTTGAACG

 FC_CD   AAAAGCCTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT
 FC_IR   AAAAGNNTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT
BJS_CD   AAAAGTCCCAGTTTGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTT-TTTCACT
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_CD   AAAAGTCTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT
199_IR   AAAAGTCTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT
216_CD   AAAAGTCTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT
216_IR   AAAAGTCTCAGTTCGATTATCGCTTCTTTGCCACCTCTAAAGCTCTCTCTTCGTTTCGCT

 FC_CD   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG
 FC_IR   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG
BJS_CD   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAGACTTTAGTTAGCTGCTTTTTTTTGG
BJS_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGCTGCTTTTTTTTGG
199_CD   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG
199_IR   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG
216_CD   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG
216_IR   CTTCAAAAACGCCGGGCGACTTGAGTGGCTCGAAAACTTTAGTTAGCTGCTTTTTTTTGG

 FC_CD   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
 FC_IR   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
BJS_CD   CCCGGTAAGACGTACGACAAAGTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
BJS_IR   CCCGGTAAGACGTACGACNAAGTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
199_CD   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
199_IR   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
216_CD   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG
216_IR   CCCGGTAAGACGTAGGACAGGCTGCTTACCTGCGTGCCAAGTTCGGGAGCGCGTTAAACG

 FC_CD   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
 FC_IR   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
BJS_CD   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTGATTATCGCCAAAGGCAA
BJS_IR   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTGATTATCGCCAAAGGCAA
199_CD   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
199_IR   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
216_CD   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
216_IR   TTATGTGTTTGCTCGCCGTGAAAGTTAAATTAAAATAAATCTTATTATCGCCAAAGGCAA
```

```
 FC_CD   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
 FC_IR   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
BJS_CD   AACAAAAAGAAAAATAAAAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
BJS_IR   AACAAAAAGAAAAATAAAAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
199_CD   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
199_IR   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
216_CD   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC
216_IR   AACAAAAAGAAAAA-AACAACGCGGCGCGTGCTTACTAAAGTGCAAGTGTGTGTTGGTGC

 FC_CD   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA
 FC_IR   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA
BJS_CD   TGGATTGCGATAGTGTTTGTGCACTGGCGGCTGGCTGAGCATGTGTGCGTGTGAGCAGCA
BJS_IR   TGGATTGCGATAGTGTTTGTGCACTGGCGGCTGGCTGAGCATGTGTGCGTGTGAGCAGCA
199_CD   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA
199_IR   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA
216_CD   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA
216_IR   TGGATTGCGATAGTGTTTGTGCACAGGCGGCTGGCTGAGCATGTGTGTGTGTGAGCAGCA

 FC_CD   AATTCCAAGAATATAACAAAAGCAGAAAAAACTGAGTTTTGCAAAAATGTCAAACAGGCG
 FC_IR   AATTCCAAGAATATAACAAAAGCAGAAAAAACTGAGTTTTGCAAAAATGTCAAACAGGCG
BJS_CD   AATTCCAAGAATATAACAAAAGCAGAAAAAACAGAGTTTTGCAAAAATGTCAAACAGGCG
BJS_IR   AATTCCAAGAATATAACAAAAGCAGAAAAAACAGAGTTTTGCAAAAATGTCAAACAGGCG
199_CD   AATTCCAAGAATATAACAAAAGCAGAAAAAACTGAGTTTTGCAAAAATGTCAAACAGGCG
199_IR   AATTCCAAGAATATAACAAAAGCAGAAAAAACTGAGTTTTGC-AAAATGTCAAACAGGCG
216_CD   AATTCCAAGAATATAACAAAAGCAGAAAAAACTGAGTTTTGCAAAAATGTCAAACAGGCG
216_IR   AATTCCAAGAATATAACAAAAGCAGAAAAAACTGAGTTTTGCAAAAATGTCAAACAGGCG

 FC_CD   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
 FC_IR   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
BJS_CD   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
BJS_IR   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
199_CD   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
199_IR   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
216_CD   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC
216_IR   AAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGTTTGCAAAAGACACC

 FC_CD   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA
 FC_IR   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA
BJS_CD   GGCAGCAACAACAAATGCAGAGGAGCAGGT------GGCAAGGGAGGAGGAGAATGGCAA
BJS_IR   GGCAGCAACAACAAATGCAGAGGAGCAGGT------GGCAAGGGAGGAGGAGAATGGCAA
199_CD   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA
199_IR   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA
216_CD   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA
216_IR   GGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAGGAGGAGAATGGCAA

 FC_CD   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
 FC_IR   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
BJS_CD   TGGGCATCATAGTACCGTTACAGCTGAAACGCCACCCACACCACCCATCCCGCCCATTCC
BJS_IR   TGGGCATCATAGTACCGTTACAGCTGAAACGCCACCCACACCACCCATCCCGCCCATTCC
199_CD   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
199_IR   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
216_CD   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
216_IR   TGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCCATCCCGCCCATTCC
```

```
 FC_CD   GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
 FC_IR   GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
BJS_CD   GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
BJS_IR   GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
199_CD   GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
199_IR   GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
216_CD   GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA
216_IR   GCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCAAGGTGAGCAGAGCA

 FC_CD   CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
 FC_IR   CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
BJS_CD   CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
BJS_IR   CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
199_CD   CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
199_IR   CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
216_CD   CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC
216_IR   CACCAGCGGTTTTCGAATTGCTGGGCTTTTCCCAACACAAACAGAAACAAGAAAGTCTGC

 FC_CD   ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTCAACATAT
 FC_IR   ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTNTCCCCCTCTTTCAACATAT
BJS_CD   ATTGGCACGTAACTAACAAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTTAACA---
BJS_IR   ATTGGCACGTAACTAACAAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTTAACA---
199_CD   ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTCAACATAT
199_IR   ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTCAACATAT
216_CD   ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTCAACATAT
216_IR   ATTGGCACGTAACTAACTAGCGCTCTCTTTCCCATTTCCTCTCCCCCTCTTTCAACATAT

 FC_CD   GTATCTCTGTCGCGGCTCGGCTCTCAAAGCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT
 FC_IR   GTATCTCTGNCGCGGCTCGGCTCTCNAAGCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT
BJS_CD   -TATCTCTGTCGCGGCTCTGCTCTCAAAGCCATATTTGCGGGGTTTCCCTCTCATTCGCT
BJS_IR   -TATCTCTGTCGCGGCTCTGCTCTCAAAGCCATATTTGCGGGGTTTCCCTCTCATTCGCT
199_CD   GTATCTCTGTCGCGGCTCGGCTCTC-AAGCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT
199_IR   GTATCTCTGTCGCGGCTCGGCTCTCAAAGCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT
216_CD   GTATCTCTGTCGCGGCTCGGCTCTCAAAGCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT
216_IR   GTATCTCTGTCGCGGCTCGGCTCTCAAAGCCCTGTTTGCGGGTTTTCCCTCTCATTTGCT

 FC_CD   CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
 FC_IR   CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
BJS_CD   CTCGCATTTCTTCTCTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
BJS_IR   CTCGCATTTCTTCTCTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
199_CD   CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
199_IR   CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
216_CD   CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA
216_IR   CTCGCTCTTCTTCACTCTCGCTGTGCATTTCACCAGTTTTAATTGTTTAAACGTAAACAA

 FC_CD   CGTG-TTTTGGGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT
 FC_IR   CGTG-TTTTGGGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT
BJS_CD   CGTGTTTTTTGACCACGTTATCTTGTTGTTGGCCTG-GGTGGCTCAACGAACCCCAGAAT
BJS_IR   CGTGTTTTTTGACCACGTTATCTTGTTGTTGGCCTG-GGTGGCTCAACGAACCCCAGAAT
199_CD   CGTG-TTTTGGGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT
199_IR   CGTG-TTTTGGGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT
216_CD   CGTG-TTTTGGGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT
216_IR   CGTG-TTTTGGGCCACGTTATCTTGTTGTTGGCCTGTGGTGGCTCAACGAACCCCAGAAT
```

```
 FC_CD   TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
 FC_IR   TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
BJS_CD   TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
BJS_IR   TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
199_CD   TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
199_IR   TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
216_CD   TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA
216_IR   TTAGCGAACTGCAAGACAGGCACAAAAAATGAACAGCAACAAATTGCGAATTCTGCCTGA

 FC_CD   AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG
 FC_IR   AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG
BJS_CD   AATTAAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCAAAAAATTGG
BJS_IR   AATTAAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCAAAAAATTGG
199_CD   AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG
199_IR   AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG
216_CD   AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG
216_IR   AATTGAACACTTTGTGTGCAGCGCAATTAAAATTCTGCACAGGTGCACTGCGAAAAATCG

 FC_CD   TTAACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC
 FC_IR   TTAACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC
BJS_CD   ATCACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTTAGAAATTTTTGTGGAA
BJS_IR   ATCACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTTAGAAATTTTTGTGGNA
199_CD   TTAACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC
199_IR   TTAACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC
216_CD   TTAACATTTGTGTAACGCATACTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC
216_IR   TTAACATTTGTGTAACGCATANTCGTCGAACACGTAATATATTCAAACATTTTTGTGGAC

 FC_CD   ATTTTATGAGATTGAGATTG-----AAATCTT----------TGCTGCTTTATGTAACAA
 FC_IR   ATTTTATGAGATTGAGATTG-----AAATCTT----------TGCTGCTTTATGTAACAA
BJS_CD   ATTTTAGTAGATTGAGATTGAAACCAAATCTTGAAGTCTAAACGCTCCTTTATGTAACAG
BJS_IR   ATTTTAGTAGATTGAGATTGAAACCAAATCTTGAAGTCTAAACGCTCCTTTATGTAACAG
199_CD   ATTTTATGAGATTGAGATTG-----AAATCTT----------TGCTGCTTTATGTAACAA
199_IR   ATTTTATGAGATTGAGATTG-----AAATCTT----------TGCTGCTTTATGTAACAA
216_CD   ATTTTATGAGATTGAGATTG-----AAATCTT----------TGCTGCTTTATGTAACAA
216_IR   ATTTTATGAGATTGAGATTG-----AAATCTT----------TGCTGCTTTATGTAACAA

 FC_CD   TAAAGATTGTTTTCAGTGATCAACAAGTTTTCGCACTTACAACAGAGCGCCCTTC----A
 FC_IR   TAAAGATTGTTTTCAGTGATCAACAAGTTTTCGCACTTACAACAGAGCGCCCTTC----A
BJS_CD   TAACGATTGTTTTCAGTGATCAACAAGTTTTCTCACTTACAAGAGAGCACCCTTCATAAA
BJS_IR   TAACGATTGTTTTCAGTGATCAACAAGTTTTCTCACTTACAAGAGAGCACCCTTCATAAA
199_CD   TAAAGATTGTTTTCAGTGATCAACAAGTTTTCGCACTTACAACAGAGCGCCCTTC----A
199_IR   TAAAGATTGTTTTCAGTGATCAACAAGTTTTCGCACTTACAACAGAGCGCCCTTC----A
216_CD   TAAAGATTGTTTTCAGTGATCAACAAGTTTTCGCACTTACAACAGAGCGCCCTTC----A
216_IR   TAAAGATTGTTTTCAGTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN----N

 FC_CD   TTTTTTAAACTGATTTTAATGTGTTGAAATATTTTCAATGGAAAATCCCTAATTGCCCGT
 FC_IR   TTTTTTAAACTGATTTTAATGTGTTGAAATATTTTCAATGGAAAATCCCTAATTGCACGT
BJS_CD   TATTTTAAACTGATTTTAATGTGTTCAAATA---------AAAAAGCAATAATTGCTAGG
BJS_IR   TATTTTAAACTGATTTTAATGTGTTCAAATA---------AAAAAGCAATAATTGCTAGG
199_CD   TTTTTTAAACTGATTTTAATGTGTTGAAATATTTTCAATGGAAAATCCCTAATTGCACGT
199_IR   TTTTTTAAACTGATTTTAATGTGTTGAAATATTTTCAATGGAAAATCCCTAATTGCACGT
216_CD   TTTTTTAAACTGATTTTAATGTGTTGAAATATTTTCAATGGAAAATCCCTAATTGCACGT
216_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

```
 FC_CD   TATTAAATCGGGTATACAGTATAAAAACTAAGTTTCTTTATAATGCATGTTATATACGCT
 FC_IR   TATTAAATCGGGTATACAGTATAAAAACTAAGTTTCTTTATAATGCATGTTATATACGCT
BJS_CD   TATTA-----------CATTACAAAAACTAAGTTTCTTTATAATGCATGTTATATTCGCT
BJS_IR   TATTA-----------CATTACAAAAACTAAGTTTCTTTATAATGCATGTTATATTCGCT
199_CD   TATTAAATCGGGTATACAGTATAAAAACTAAGTTTCTTTATAATGCATGTTATATACGCT
199_IR   TATTAAATCGGGTATACAGTATAAAAACTAAGTTTCTTTATAATGCATGTTATATACGCT
216_CD   TATTAAATCGGGTATNCAGNANNAAAANTAAGTTTCTTTATNATNCANGNTANATNNNNT
216_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

 FC_CD   CTGCGTCGTTTTTCGGAGTATTTCGGCATTTATCATAGTCATA--TTTGTGTGGGGCCGC
 FC_IR   CTGCGTCGTTTTTCGGAGTATTTCGGCATTTATCATAGTCANN-NNNNNNNNNNNNNNNN
BJS_CD   CTGCGTCGTTTTTCGTAGTATTTCGGCATTTATCATAGTCATATTTTTGTGTGGGGCCGC
BJS_IR   CTGCGTCGTTTTTCGTAGTATTTCGGCATTTATCATAGTCATATTTTTGTGTGGGGCCGC
199_CD   CTGCGTCGTTTTTCGGAGTATTTCGGCATTTATCATAGTCATA--TTTGTGTGGGGCCGC
199_IR   CTGCGTCGTTTTTCGGAGTATTTCGGCATTTATCATAGTCATA--TTTGTGTGGGGCCGC
216_CD   NNGNGTCGTTTTTNGGNNNNTTTNGGCATTTNTCATAGTCANA--TTTGNGNGGGGCCGC
216_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN--NNNNNNNNNNNNNNN

 FC_CD   CTCAAGTCTTCCAAGATATTTGCTGATAACGCTGACGGAGGACTGATAACACATACACAC
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   CTCAAGTCTTCCAAGATATTTGCTGATAACGCTGACGGTGGACTGATAACACATACACAC
BJS_IR   CTCAAGTCTTCCAAGATATTTGCTGATAACGCTGACGGTGGACTGATAACACATACACAC
199_CD   CTCAAGTCTTCCAAGATATTTGCTGATAACGCTGACGGTGGACTGATAACACATACACAC
199_IR   CTCAAGTCTTCCAAGATATTTGCTGATAACGCTGACGGTGGACTGATAACACATACACAC
216_CD   CNCAAGTCNTNCAAGANATTTGCNGANAACGCNGNCGGTGGNCTGATAACNCATNCACAC
216_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

 FC_CD   GCTCACCTTGAGGGAATTACGGATACGTTTTCCGCTGTTTTTTTTTTGTTGTTGTGTCTT
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTTTTTTTTGTTGTTGTGTCTT
BJS_CD   GCTCACCTTGAGAGAATTACGGATACGTTTTCCGCTGTTTTTTTTT-----TTTTGTCTT
BJS_IR   GCTCACCTTGAGAGAATTACGGATACGTTTTCCGCTGTTTTTTTTT-----TTTTGTCTT
199_CD   GCTCACCTTGAGGGAATTACGGATACGTTTTCCGCTGTTTTTTTTTTGTTGTTGTGTCTT
199_IR   GCTCACCTTGAGGGAATTACGGATACGTTTTCCGCTGTTTTTTTTTTGTTGTTGTGTCTT
216_CD   NCTCNCCTTGAGGGAATTNCGGANNCGTTTTCCGCTGTTTTTTTTTTGTTGTTGTGTCTT
216_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTTTTTTTTTTGTTGTTGTGTCTT

 FC_CD   TTGCGCGGGGGCTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG
 FC_IR   TTGCGCGGGGNCTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG
BJS_CD   ATGCGCAGGGGCTTATTCATAATACACTTCATTTTTTTTGATTTGATTTTTATGGGATAG
BJS_IR   ATGCGCAGGGGCTTATTCATAATACACTTCATTTTTTTTGATTTGATTTTTATGGGATAG
199_CD   TTGCGC-GGGGCTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG
199_IR   TTGCGCNGGGGCTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG
216_CD   TTGCGC-GGGGCTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG
216_IR   TTGCGCNGGGGCTTATTCATAATACAATTAATATTTTTTGATTTGATTTTCATGGGATGG

 FC_CD   GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA
 FC_IR   GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA
BJS_CD   GTGAGTGTTCA--------TCGTTTTGCGGGAAAACATGATTGTCGTATGTTAGCAGGAA
BJS_IR   GTGAGTGTTCA--------TCGTTTTGCGGGAAAACATGATTGTCGTATGTTAGCAGGAA
199_CD   GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA
199_IR   GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA
216_CD   GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA
216_IR   GTGAGTGTTCAAGGGGGTTTCGTTTCGCAGGAAAACACGAATGTCGTATGTTAGCAGGAA
```

```
 FC_CD   ATTTGCGGCAAATATCCATTTTC-TTTGGCACCTGTTTCTATGCCAACTAAGTGAAATTA
 FC_IR   ATTTGCGGCAAATATCCATTTTC-TTTGGCACCTGTTTCTATGCCAACTAAGTGAAATTA
BJS_CD   ATTTGCGACAAACATCCATTTTCTTTTGGCACCTGTTTCTATGGCAACTAAGTGAAATTA
BJS_IR   ATTTGCGACAAACATCCANTTTCTTTTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNATTA
199_CD   ATTTGCGGCAAATATCCATTTTC-TTTGGCACCTGTTTCTATGCCAACTAAGTGAAATTA
199_IR   ATTTGCGGCAAATATCCATTTTC-TTTGGCACCTGTTTCTATGCCAACTAAGTGAAATTA
216_CD   ATTTGCGGCAAATATCCATTTTC-TTTGGCACCTGTTTCTATGCCAACTAAGTGAAATTA
216_IR   ATTTGCGGCAAATATCCATTTTC-TTTGGCACCTGTTTCTATGCCAACTAAGTGAAATTA

 FC_CD   GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA
 FC_IR   GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA
BJS_CD   GCCGAGATTTATGTGTCTGCTGCTCGCTCTTTTTTTCGTTGG--TCTCTATATTTTAATA
BJS_IR   GCCGAGATTTATGTGTCTGCTGCTCGCTCTTTTTTTCGTTGG--TCTCTATATTTTAATA
199_CD   GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA
199_IR   GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA
216_CD   GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA
216_IR   GCCGAGATTTATGTGTCTGCTGCTTGCTCTTTTTTTCGTTGGTCTCTCTATATTTCAATA

 FC_CD   CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATAT-
 FC_IR   CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATAT-
BJS_CD   CCTGTTAATTGATAAGTTGATTGGTTCGCTGGAAAAATGGCAACGTATAGATATAATATT
BJS_IR   CCTGTTAATTGATAAGTTGATTGGTTCGCTGGAAAAATGGCAACGTATAGATATAATATT
199_CD   CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATAT-
199_IR   CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATAT-
216_CD   CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATAT-
216_IR   CCTGTTAATTGATAAGTTGATTAGTTCGCTGGGAAAATGGCAATATATAAATAAAATAT-

 FC_CD   ---TATGTTATTTTT----TTAAAGCATCGTAT------AGATACAAGCAGCGATTCTTT
 FC_IR   ---TATGTTATTTTT----TTAAAGCATCGTAT------AGATACAAGCAGCGATTCTTT
BJS_CD   ACATATGTTTTTTTTTTTATTAAAGCATCGTATATGCTAAGATACAAGCAGTGATTCTTT
BJS_IR   ACATATGTTTTTTTTTTTATTAAAGCATCGTATATGCTAAGATACAAGCAGTGATTCTTT
199_CD   ---TATGTTATTTTT----TTAAAGCATCGTAT------AGATACAAGCAGCGATTCTTT
199_IR   ---TATGTTATTTTT----TTAAAGCATCGTAT------AGATACAAGCAGCGATTCTTT
216_CD   ---TATGTTATTTTT----TTAAAGCATCGTAT------AGATACAAGCAGCGATTCTTT
216_IR   ---TATGTTATTTTT----TTAAAGCATCGTAT------AGATACAAGCAGCGATTCTTT

 FC_CD   AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACAGCATTTATAG
 FC_IR   AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACANCATTTATAG
BJS_CD   AGGAAAGGTTACGAAAAA-------GTTTTTGACTTTAAAGGTTTAAACAGCATTTATAG
BJS_IR   AGGAAAGGTTACGAAAAA-------GTTTTTGACTTTAAAGGTTTAAACAGCATTTATAG
199_CD   AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACAGCATTTATAG
199_IR   AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACAGCATTTATAG
216_CD   AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACAGCATTTATAG
216_IR   AGGAAAGGTACCGAAAAAGTTTTTGGTTTTTGATTTTAAATTTTACAACAGCATTTATAG

 FC_CD   GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTTGACTTTATCTCTCTTTTA
 FC_IR   GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTNNNNNNNNNNNNNNNNNNN
BJS_CD   GGTAATCGGTAGTTCGGACATCTGTAAGCATTGTTCTCATTTG-----------------
BJS_IR   GGTAATCGGTAGTTCGGACATCTGTAAGCATTGTTCTCATTTG-----------------
199_CD   GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTTGACTTTATCTCTCTTTTA
199_IR   GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTTGACTTTATCTCTCTTTTA
216_CD   GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTTGACTTTATCTCTCTTTTA
216_IR   GGTAATCGATAGTTTGGACATGTGTAAGCAGTGTTCTCATTTGACTTTATCTCTCTTTTA
```

```
 FC_CD   ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGCAGGAAAAAAAA
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGGAGGGAAGAAGA
BJS_IR   ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGGAGGGAAGAAGA
199_CD   ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGCAGGGAAGAAGA
199_IR   ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGCAGGGAAGAAGA
216_CD   ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGCAGGGAAGAAGA
216_IR   ATGCCGCCAATCTTGTTCTTCTTCAGCTGGTATACCTGTTTTAGGGGGCAGGGAAGAAGA

 FC_CD   AGAGGAGACGCAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   AGAAGAGACGGAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
BJS_IR   AGAAGAGACGGAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
199_CD   AGAGGAGACGCAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
199_IR   AGAGGAGACGCAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
216_CD   AGAGGAGACGCAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA
216_IR   AGAGGAGACGCAGGGGAGACTATTGGAAGTTAAAACGCGTTTAATACGTGCGTTTGCCCA

 FC_CD   CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCAAA
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCATA
BJS_IR   CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCATA
199_CD   CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCAAA
199_IR   CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCAAA
216_CD   CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCAAA
216_IR   CTCTGCTTTGTAACCTCTTTCGTTATTCTGTTTTCAAATCGAATTCGCTTTCTTCTCAAA

 FC_CD   TCTCCGTCTCATAATGTTTTGCGGTTGTCACCTGGCTAACTGAATGGAATTCCGTTGATT
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   TCTCCGTC-CATAATGTTTTGTGGTTGTCACCTGGCTAACTGAATGGAATTCCGATGATT
BJS_IR   TCTCCGTC-CATAATGTTTTGTGGTTGTCACCTGGCTAACTGAATGGAATTCCGATGATT
199_CD   TCTCCGTCTCATAATGTTTTGCGGTTGTCACCTGGCTAACTGAATGGAATTCCGTTGATT
199_IR   TCTCCGTCTCATAATGTTTTGCGGTTGTCACCTGGCTAACTGAATGGAATTCCGTTGATT
216_CD   TCTCCGTCTCATAATGTTTTGCGGTTGTCACCTGGCTAACTGAATGGAATTCCGTTGATT
216_IR   TCTCCGTCTCATAATGTTTTGCGGTTGTCACCTGGCTAACTGAATGGAATTCCGTTGATT

 FC_CD   GAAGGTCTCTGGCCTAACTGTTGTTGCCTGGGGCTAAGAGAGCGAGAGATCGCGCTCTTG
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   GAAGGTCTCTGGCCTAACTGTTGTTGGCTTTGGTGAAGAGAGCGGGAGCTCGCGCTCTTG
BJS_IR   GAAGGTCTCTGGCCTAACTGTTGTTGGCTTTGGTGAAGAGAGCGGGAGCTCGCGCTCTTG
199_CD   GAAGGTCTCTGGCCTAACTGTTGTTGCCTGGGGCTAAGAGAGCGAGAGATCGCGCTCTTG
199_IR   GAAGGTCTCTGGCCTAACTGTTGTTGCCTNGGGCTAAGAGAGCGAGAGATCGCGCTCTTG
216_CD   GAAGGTCTCTGGCCTAACTGTTGTTGCCT-GGGCTAAGAGAGCGAGAGATCGCGCTCTTG
216_IR   GAAGGTCTCTGGCCTAACTGTTGTTGCCTNGGGCTAAGAGAGCGAGAGATCGCGCTCTTG

 FC_CD   CAACATTCGTATTTGTACTTGAACTTGCTGGTTATCTCTCTCTCTCTCGCTCTCTCTTCT
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATCTCTCTCTCTCNNGCTCTCTCTTCT
BJS_CD   CAACATTCGTATTTGTACTTGAACTTGCAGGTTA----TCTTTCTCTCTCTCTCTCTTTT
BJS_IR   CAACATTCGTATTTGTACTTGAACTTGCAGGTTA----TCTTTCTCTCTCTCTCTCTTTT
199_CD   CAACATTCGTATTTGTACTTGAACTTGCTGGTTA--TCTCTCTCTCTCGCTCTCTCTTCT
199_IR   CAACATTCGTATTTGTACTTGAACTTGCTGGTTA--TCTCTCTCTCTCGCTCTCTCTTCT
216_CD   CAACATTCGTATTTGTACTTGAACTTGCTGGTTA--TCTCTCTCTCTCGCTCTCTCTTCT
216_IR   CAACATTCGTATTTGTACTTGAACTTGCTGGTTA--TCTCTCTCTCTCGCTCTCTCTTCT
```

```
 FC_CD   GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG
 FC_IR   GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG
BJS_CD   GCTCCACGTTCTAGACC-------------------------------TCAGTTCAGTTG
BJS_IR   GCTCCACGTTCTAGACC-------------------------------TCAGTTCAGTTG
199_CD   GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG
199_IR   GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG
216_CD   GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG
216_IR   GCTCCACGTTCTAGCCCTCTCGCTTTGGGCAGCCTGGCATGAGCACAATCAAATCAGTTG

 FC_CD   GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA
 FC_IR   GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA
BJS_CD   GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAAT--------------TGTACA
BJS_IR   GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAAT--------------TGTACA
199_CD   GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA
199_IR   GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA
216_CD   GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA
216_IR   GTTATAATTTTGGGGCTTTCGTGTGTTTGATAAGCATAATTATGGCGTGTACAATGTACA

 FC_CD   CACATATCAGAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC
 FC_IR   CACATATCAGAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC
BJS_CD   CACATATCAGAAAACATTTCCGCATATTACACACGCACGCGCC-----------------
BJS_IR   CACATATCAGAAAACATTTCCGCATATTACACACGCACGCGCC-----------------
199_CD   CACATATCAGAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC
199_IR   CACATATCAGAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC
216_CD   CACATATCAGAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC
216_IR   CACATATCANAAAACATTTCCGCATATTGCACACGCACGCGCCCACACACACACACACAC

 FC_CD   ACACACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAATTTTCTAATGTGTTTT
 FC_IR   ACACACACACGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   ---CACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAAGTTTCTAATGTGTTTT
BJS_IR   ---CACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAAGTTTCTAATGTGTTTT
199_CD   ACACACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAATTTTCTAATGTGTTTT
199_IR   ACACACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAATTTTCTAATGTGTTTT
216_CD   ACACACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAATTTTCTAATGTGTTTT
216_IR   ACACACACACACACGTGCCACTGCCGTTTGCACTAATTTTAGAATTTTCTAATGTGTTTT

 FC_CD   TAAATGCCATGTTTACCGAAGAAAACAGCATAAACATGAGATTCAGAGATTGTGTATGAT
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   TAAATGCCATGTTTACAGAAGAAAACAGCATAAACATGAGATTCACAGATTGTGTATGAT
BJS_IR   TAAATGCCATGTTTACAGAAGAAAACAGCATAAACATGAGATTCACAGATTGTGTATGAT
199_CD   TAAATGCCATGTTTACCGAAGAAAACAGCATAAACATGAGATTCAGAGATTGTGTATGAT
199_IR   TAAATGCCATGTTTACCGAAGAAAACAGCATAAACATGAGATTCAGAGATTGTGTATGAT
216_CD   TAAATGCCATGTTTACCGAAGAAAACAGCATAAACATGAGATTCAGAGATTGTGTATGAT
216_IR   TAAATGCCATGTTTACCGAAGAAAACAGCATAAACATGAGATTCAGAGATTGTGTATGAT

 FC_CD   TTGTTATTTCTATTTATTTAAAA--TACAAACCTCTCAAAACCTATTCGATTACTCTTGT
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNN--NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   TTGTTATTTCCATTTATTTAAAAATTGAAAACCTCTCAAAATCCATTTGATTACTATTTT
BJS_IR   TTGTTATTTCCATTTATTTAAAAATTGAAAACCTCTCAAAATCCATTTGATTACTATTTT
199_CD   TTGTTATTTCTATTTATTTAAAA--TACAAACCTCTCAAAACCTATTCGATTACTCTTGT
199_IR   TTGTTATTTCTATTTATTTAAAA--TACAAACCTCTCAAAACCTATTCGATTACTCTTGT
216_CD   TTGTTATTTCTATTTATTTAAAA--TACAAACCTCTCAAAACCTATTCGATTACTCTTGT
216_IR   TTGTTATTTCTATTTATTTAAAA--TACAAACCTCTCAAAACCTATTCGATTACTCTTGT
```

```
 FC_CD   TCATTTTTGTGCCTTAAATCGGCAAAAAGTAAGCTAAGAAACTTTGAAAAATAAAACTAA
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   ACATTTTTAACTTTTTACTT---------TACTTTAAGAAGCATTGAAAGATAAAACTAA
BJS_IR   ACATTTTTAACTTTTTACTT---------TACTTTAAGAAGCATTGAAAGATAAAACTAA
199_CD   TCATTTTCGTGCCTTAAATTGGCAAAAACTAAGCTAAGAAACATTGAAAAATAAAACTAA
199_IR   TCATTTTCGTGCCTTAAATTGGCAAAAACTAAGCTAAGAAACATTGAAAAATAAAACTAA
216_CD   TCATTTTCGTGCCTTAAATTGGCAAAAACTAAGCTAAGAAACATTGAAAAATAAAACTAA
216_IR   TCATTTTCGTGCCTTAAATTGGCAAAAACTAAGCTAAGAAACATTGAAAAATAAAACTAA

 FC_CD   AAACTATGCTTGT-TTTCCACATTTGTTTACAAGAAATCATTTTGATTGGGTGTCTGATG
 FC_IR   NNNNNNNNNNNNN-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGATG
BJS_CD   AAACTATGGTTGCATATTAACATTTGTTTACATTTAATAATCTTGATTGGGTGTCTGATT
BJS_IR   AAACTATGGTTGCATATTAACATTTGTTTACATTTAATAATCTTGATTGGGTGTCTGATT
199_CD   AAACTATGCTTGT-TTTCCACATTTGTTTACAAGAAATAATTTTGATTGGGTGTCTGATG
199_IR   AAACTATGCTTGT-TTTCCACATTTGTTTACAAGAAATAATTTTGATTGGGTGTCTGATG
216_CD   AAACTATGCTTGT-TTTCCACATTTGTTTACAAGAAATAATTTTGATTGGGTGTCTGATG
216_IR   AAACTATGCTTGT-TTTCCACATTTGTTTACAAGAAATAATTTTGATTGGGTGTCTGATG

 FC_CD   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG
 FC_IR   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG
BJS_CD   GTCTGTTTCTTATTACAACTCACTTTGTTTTGGACGAACTCGTCCTATCAGTCGATCTGA
BJS_IR   GTCTGTTTCTTATTACAACTCACTTTGTTTTGGACGAACTCGTCCTATCAGTCGATCTGA
199_CD   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG
199_IR   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG
216_CD   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG
216_IR   GTCTGATTCTTATTTCAACTCACTTTGTTTTGGACGAAGTCGCCCTATCAGTCAATCTGG

 FC_CD   TAAATTGATCGGTTAGTCATCGAATTTCCAATGCCTCATACATAAATGGGGACCACGGTG
 FC_IR   TAAATTGATCGGTTAGTCATCGAATTTCCAATGCCTCATACATAAATGGGGACCACGGTG
BJS_CD   TAAATTGATCGCTTAGTCATCGAATTTCCAATGCTTCATACATAAATGGGGACCACGGTG
BJS_IR   TAAATTGATCGCTTAGTCATCGAATTTCCAATGCTTCATACATAAATGGGGACCACGGTG
199_CD   TAAATTGATCGGTTAGTCATCGAATTTCAAATGCCTCATACATAAATGGGGACCACGGTG
199_IR   TAAATTGATCGGTTAGTCATCGAATTTCAAATGCCTCATACATAAATGGGGACCACGGTG
216_CD   TAAATTGATCGGTTAGTCATCGAATTTCAAATGCCTCATACATAAATGGGGACCACGGTG
216_IR   TAAATTGATCGGTTAGTCATCGAATTTCAAATGCCTCATACATAAATGGGGACCACGGTG

 FC_CD   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
 FC_IR   CGAANNNNNNNNNNNNNTGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
BJS_CD   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
BJS_IR   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
199_CD   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
199_IR   CGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
216_CD   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG
216_IR   CGAATGTGCAACTTTTATGTGTATCTATAAATATGAAGTATACGCACCGCTGGTCACGTG

 FC_CD   ACAATAGATTTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTCGAA-TCTATTTTT
 FC_IR   ACAATAGATTTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTCGAA-TCTATTTTT
BJS_CD   ACCAAAGA-TTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTTGAA-TCGATTTTC
BJS_IR   ACCAAAGA-TTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTTGAA-TCGATTTTC
199_CD   ACAATAGATTTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTCGAA-TCTATTTTT
199_IR   NNNNNNNNNNTTTCATATAGAGCCAATGAAGTTGATTGACNNGATTNGAANTCTATTTTT
216_CD   ACAATAGATTTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTCGAA-TCTATTTTT
216_IR   ACAATAGATTTTTCATATAGAGCCAATGAAGTTGATTGACTTGATTCGAA-TCTATTTTT
```

```
 FC_CD   TTCAATCTATTTTATGCTGCTCGCTTGTAGAAAAATACATACACTCAATAGCTTAAGTAA
 FC_IR   TNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   AT-----------ATTCTGACCGTTTGTAGAAAAATACCTACACTCAAAAGCTTAAGTAA
BJS_IR   AT-----------ATTCTGACCGTTTGTAGAAAAATACCTACACTCAAAAGCTTAAGTAA
199_CD   TTGAATCTATTTTATTCTGCTCGCTTGTAGAAAAATACATACACTCAATAGCTTAAGTAA
199_IR   TTGAATCTATTNNNNTNTGCTCGCTTGTAGAAAAATACATACACTCAATAGCTTAAGTAA
216_CD   TTGAATCTATTTTATTCTGCTCGCTTGTAGAAAAATACATACACTCAATAGCTTAAGTAA
216_IR   TTGAATCTATTTTATTCTGCTCGCTTGTAGAAAAATACATACACTCAATAGCTTAAGTAA

 FC_CD   ACACATTGCTTATTAGCTAA----CGCGAACATAATAAAATGCCATAAATTAATAATAAA
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   ACACATTGCTTATTAGCCAACAAGCGGGTACATAATAAAATGCCATAAATTAATAATAAA
BJS_IR   ACACATTGCTTATTAGCCAACAAGCGGGTACATAATAAAATGCCATAAATTAATAATAAA
199_CD   ACACATTGCTTATTAGCTAA----CGCGAACATAATAAAATGCCATAAATTAATAATAAA
199_IR   ACACATTGCTTATTAGCTAA----CGCGAACATAATAAAATGCCATAAATTAATAATAAA
216_CD   ACACATTGCTTATTAGCTAA----CGCGAACATAATAAAATGCCATAAATTAATAATAAA
216_IR   ACACATTGCTTATTAGCTAA----CGCGAACATAATAAAATGCCATAAATTAATAATAAA

 FC_CD   TATTCGCGCGGTTAAGACAATTGCAAACCACAAACAAACGCAATACGTTCCACAAGCGAT
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   TATTCTCGCGGCTAAGACAATTGCGAACCACAAACAAACGCAATGCGCTCCAGAAACGAT
BJS_IR   TATTCTCGCGGCTAAGACAATTGCGAACCACAAACAAACGCAATGCGCTCCAGAAACGAT
199_CD   TATTCGCGCGGTTAAGACAATTGCAAACCACAAACAAACGCAATACGTTCCACAAGCGAT
199_IR   TATTCGCGCGGTTAAGACAATTGCAAACCACAAACAAACGCAATACGTTCCACAAGCGAT
216_CD   TATTCGCGCGGTTAAGACAATTGCAAACCACAAACAAACGCAATACGTTCCACAAGCGAT
216_IR   TATTCGCGCGGTTAAGACAATTGCAAACCACAAACAAACGCAATACGTTCCACAAGCGAT

 FC_CD   CGAATACCCCTTTGCGTTCCACAACAAACCGTATCTCGGTGGTGCGTCTCTCTTTCTGGC
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   CGAATACCCCTTTGCGTTCCACAACAAACCGTATCACGGTGGTGTGTCTCTCTTTCTTGT
BJS_IR   CGAATACCCCTTTGCGTTCCACAACAAACCGTATCACGGTGGTGTGTCTCTCTTTCTTGT
199_CD   CGAATACCCCTTTGCGTTCCACAACAAACCGTATCTCGGTGGTGCGTCTCTCTTTCTGGC
199_IR   CGAATACCCCTTTGCGTTCCACAACAAACCGTATCTCGGTGGTGCGTCTCTCTTTCTGGC
216_CD   CGAATACCCCTTTGCGTTCCACAACAAACCGTATCTCGGTGGTGCGTCTCTCTTTCTGGC
216_IR   CGAATACCCCTTTGCGTTCCACAACAAACCGTATCTCGGTGGTGCGTCTCTCTTTCTGGC

 FC_CD   TCTCTTAGCTTCCATACCTCTCTGGCTCATCTTTTTTGCGTCTCCTCTCTCTGAGATGCC
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   GCTCTTAGTTTCCATACCTCTGTTGCTCATC-TTATTGCGTCTCCTCTCTCTGAGATACC
BJS_IR   GCTCTTAGTTTCCATACCTCTGTTGCTCATC-TTATTGCGTCTCCTCTCTCTGAGATACC
199_CD   TCTCTTAGCTTCCATACCTCTCTGGCTCATCTTTTTTGCGTCTCCTCTCTCTGAGATGCC
199_IR   TCTCTTAGCTTCCATACCTCTCTGGCTCATCTTTTTTGCGTCTCCTCTCTCTGAGATGCC
216_CD   TCTCTTAGCTTCCATACCTCTCTGGCTCATCTTTTTTGCGTCTCCTCTCTCTGAGATGCC
216_IR   TCTCTTAGCTTCCATACCTCTCTGGCTCATCTTTTTTGCGTCTCCTCTCTCTGAGATGCC

 FC_CD   ACACAAAAATGGCTATAGTT--TAATGGCAGATATTCTATGTATTGTAATCTTTAGGCTT
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   ATACACAAATGGCTATAGTTTATAGTGGCAGGTATTGTATTTATTGTAAACCGTAGGCTT
BJS_IR   ATACACAAATGGCTATAGTTTATAGTGGCAGGTATTGTATTTATTGTAAACCGTAGGCTT
199_CD   ACACAAAAATGGCTATAGTT--TAATGGCAGATATTCTATGTATTGTAATCTTTAGGCTT
199_IR   ACACAAAAATGGCTATAGTT--TAATGGCAGATATTCTATGTATTGTAATCTTTAGGCTT
216_CD   ACACAAAAATGGCTATAGTT--TAATGGCAGATATTCTATGTATTGTAATCTTTAGGCTT
216_IR   ACACAAAAATGGCTATAGTT--TAATGGCAGATATTCTATGTATTGTAATCTTTAGGCTT
```

```
 FC_CD   CACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCCCAATCGAGTTG
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   CACAATTTATATTTAATATGTTATGCATACTTTTACACTTTCTGTGCCACACTCCAGTAG
BJS_IR   CACAATTTATATTTAATATGTTATGCATACTTTTACACTTTCTGTGCCACACTCCAGTAG
199_CD   CACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCACAATCGAGTTG
199_IR   CACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCACAATCGAGTTG
216_CD   CACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCACAATCGAGTTG
216_IR   CACAATTTATAGTTAATATGTTAAGCATACTTTTACACTTTCTGTGCCACAATCGAGTTG

 FC_CD   TTTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGC
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   TTTCTGACTGAAATGACCACTGGTAAATGTTAAACATAACTATTTAATAGCTAAACCGGC
BJS_IR   TTTCTGACTGAAATGACCACTGGTAAATGTTAAACATAACTATTTAATAGCTAAACCGGC
199_CD   TTTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGC
199_IR   TTTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGC
216_CD   TTTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGC
216_IR   TTTCTGGCTGAAATGACCACTTCCAAATGTGAAACAGAACTATTTAGTAGCTAAACCGGC

 FC_CD   ATCTCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAACCAATCGG
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   ATCCCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGGGCCGTAAACAATCGG
BJS_IR   ATCCCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGGGCCGTAAACAATCGG
199_CD   ATCTCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAAACAATCGG
199_IR   ATCTCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAAACAATCGG
216_CD   ATCTCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAAACAATCGG
216_IR   ATCTCCGATTTGAAGTGGCCACCCACTAGTTGCTCTCCGCTGAGAACCGCAAACAATCGG

 FC_CD   ATGCGTCGGTCGGGCGCTCGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATA
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   ATGCGTCGGTCGGGCGGTCGTCGCTCGCTCTTTTCAGTTCGCCGCGATTGGCAGTGAATA
BJS_IR   ATGCGTCGGTCGGGCGGTCGTCGCTCGCTCTTTTCAGTTCGCCGCGATTGGCAGTGAATA
199_CD   ATGCGTCGGTCGGGCGCTCGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATA
199_IR   ATGCGTCGGTCGGGCGCTCGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATA
216_CD   ATGCGTCGGTCGGGCGCTCGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATA
216_IR   ATGCGTCGGTCGGGCGCTCGCCGCTCGCGCTTTTCAGTTCGCCGCGATTGGCAGTGAATA

 FC_CD   GCAGTCGTTTTGGGGGGGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCTTTA
 FC_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   GCAGTCGGTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGCTA
BJS_IR   GCAGTCGGTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGCTA
199_CD   GCAGTCGTTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGCTA
199_IR   GCAGTCGTTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGCTA
216_CD   GCAGTCGTTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGCTA
216_IR   GCAGTCGTTTTGGGGGCGACACCGCTTCGTCGGCACCGCGGCACACACTCGCTTGCGNNN

 FC_CD   CATACAGATCCGA-TACTATACGCTTTGGCGAATAGATGGGGAATATTCTTGTTAAAAGT
 FC_IR   NNNNNNG-------TACTATACGCTTTGGCGAATAGATGGGGAANNNNNNNNNNNNNNNN
BJS_CD   CATACAAAACCGTATACTGTACGCATTGGCGAATAGATGGGGAATATTCTTGTTAAAAGT
BJS_IR   CATACAAAACCGTATACTGTACGCATTGGCGAATAGATGGGGAATATTCTTGTTAAAAGT
199_CD   CATACAGATCCGT-TACTATACGCTTTGGCGAATAGATGGGGAATATTCTTGT-AAAAGT
199_IR   CATACAGATCCGT-TACTATACGCTTTGGCGAATAGANGGGGAANNNNNNNNNNNNNNNN
216_CD   CATACAGATCCGT-TACTATACGCTTTGGCGAATAGANGGGGAATATNCNNNNNNAAAGT
216_IR   CATACAGATCCGT-TACTATACGCTTTGGCGAATAGANGGGGAATATTCTNGTNAANNNN
```

```
 FC_CD   GCTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGG
 FC_IR   NNNNNNNNNNNNNNNGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGG
BJS_CD   GCTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGG
BJS_IR   GCTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGG
199_CD   GCTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGG
199_IR   NNNNNNNNNNNNNNTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGG
216_CD   GCTAATAAATTGACTGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGG
216_IR   NNNNNNNNNNNNNNNGACTGGCTGCCGCGACAAGTGCTGCGCGTTATGAAATTCGAAAGG

 FC_CD   GTTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG
 FC_IR   GTTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG
BJS_CD   GTTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCAATTCGTCGTGATAAACCAGA
BJS_IR   GTTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCAATTCGTCGTGATAAACCAGA
199_CD   GTTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG
199_IR   GTTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG
216_CD   GTTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG
216_IR   GTTCGGTGCGAGGCGTGCGAACTACTGCCGCCTGTTTGCCATTCGTCGTGATAAACCAGG

 FC_CD   --------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCAC
 FC_IR   --------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCAC
BJS_CD   AAACCAGCCAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCATAACCCAC
BJS_IR   AAACCAGCCAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCATAACCCAC
199_CD   --------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCAC
199_IR   --------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCAC
216_CD   --------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCAC
216_IR   --------CAGAAAACTACACACTACTACGGTGGGAGATAGTGCATAACCCGTAACCCAC

 FC_CD   CCCACCAGT-GAGATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCG
 FC_IR   CCCACCAGT-GAGATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCG
BJS_CD   CCCACCAGTAAAGATTCCATCGATCCGAGAAGGGAATCGGCA-------GTCGCATACCG
BJS_IR   CCCACCAGTAAAGATTCCATCGATCCGAGAAGGGAATCGGCA-------GTCGCATACCG
199_CD   CCCACCAGT-GAGATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCG
199_IR   CCCACCAGT-GAGATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCG
216_CD   CCCACCAGT-GAGATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCG
216_IR   CCCACCAGT-GAGATTCCATCGATCCGAGAGGGGAATCGGCAGCCGGGTGCCGCGTACCG

 FC_CD   GGGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGC
 FC_IR   GGGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGC
BJS_CD   GGGATCGGTCATCCAGCCCCCAGG-------------CCAGTCCAGTCATCGTTCGCTGGC
BJS_IR   GGGATCGGTCATCCAGCCCCCAGG-------------CCAGTCCAGTCATCGTTCGCTGGC
199_CD   GGGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGC
199_IR   GGGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGC
216_CD   GGGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGC
216_IR   GGGATCGGTCATTCCGCCCCCAGGCCAGTCCAGTGCCCAGTCCAGTCATCGTTCGCTGGC

 FC_CD   TCGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTT
 FC_IR   TCGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTT
BJS_CD   TCGCGTCTAAAGTT------CTCTCCAGCTCTACCAGCTTTTAATTGTGAGCTGTCGCTT
BJS_IR   TCGCGTCTAAAGTT------CTCTCCAGCTCTACCAGCTTTTAATTGTGAGCTGTCGCTT
199_CD   TCGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTT
199_IR   TCGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTT
216_CD   TCGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTT
216_IR   TCGCGTCTAAAGTTCTCCAGCTCTCCAGCTCCACCAGCTTTTAATTGTGAGCTGTCAGTT
```

```
 FC_CD   TTTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTT
 FC_IR   TTTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTT
BJS_CD   TTTATTTGCTCTCAGATCGTTCCGCCGCTGTTTGTT-TTTTGTCCCCAGAGTGGGTTGTT
BJS_IR   TTTATTTGCTCTCAGATCGTTCCGCCGCTGTTTGTT-TTTTGTCCCCAGAGTGGGTTGTT
199_CD   TTTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTT
199_IR   TTTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTT
216_CD   TTTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTT
216_IR   TTTATTTGCTCTCGTATCGTTTCGCCGCTATTTTTTGTTTTGTCTCCAGAGTGGGTTGTT

 FC_CD   GTGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGA
 FC_IR   GTGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGA
BJS_CD   GTGGTGTACGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGA
BJS_IR   GTGGTGTACGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGA
199_CD   GTGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGA
199_IR   GTGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGA
216_CD   GTGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGA
216_IR   GTGGTGTATGGTGCGGTGCAATTAGATGTGCATAAATCTGGCTCGGGCACCCACCGGAGA

 FC_CD   CTCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCAT
 FC_IR   CTCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCAT
BJS_CD   CTCTGAAAGTAACACTAGCCCTCGATTGCGGGGA--------AGTTTAAAACAGCGGCAT
BJS_IR   CTCTGAAAGTAACACTAGCCCTCGATTGCGGGGA--------AGTTTAAAACAGCGGCAT
199_CD   CTCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCAT
199_IR   CTCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCAT
216_CD   CTCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCAT
216_IR   CTCTGAAAGTAACACTAGCCCTCGATTGCGGCGATTAAGCACAGTTTAAAACAGCGGCAT

 FC_CD   TATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCA--CTCCAAGGTTCTCTCTCGTC
 FC_IR   TATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCA--CTCCAAGGTTCTCTCTCGTC
BJS_CD   TATCAACCAGCCAGTTGAGAGTAAAAAAGAGGATTGCACTCTCCAAGGTTCTCTCTCGTC
BJS_IR   TATCAACCAGCCAGTTGAGAGTAAAAAAGAGGATTGCACTCTCCAAGGTTCTCTCTCGTC
199_CD   TATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCA--CTCCAAGGTTCTCTCTCGTC
199_IR   TATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCA--CTCCAAGGTTCTCTCTCGTC
216_CD   TATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCA--CTCCAAGGTTCTCTCTCGTC
216_IR   TATCAACCAGGCAGTTGAGAGCACAGGCGAGGATTGCA--CTCCAAGGTTCTCTCTCGTC

 FC_CD   TCGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGA
 FC_IR   TCGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGA
BJS_CD   TCGCTGCGAC-TCGCTGATTAGTGCCGGTTTTGGAGCTATTGGACTGCGCAAACAG-TGA
BJS_IR   TCGCTGCGAC-TCGCTGATTAGTGCCGGTTTTGGAGCTATTGGACTGCGCAAACAG-TGA
199_CD   TCGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGA
199_IR   TCGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGA
216_CD   TCGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGA
216_IR   TCGCTGCAACTTTGCTGATTAGTGCCGGTTTTGGAGCTATTGGATTGGGCAAACAGTTGA

 FC_CD   TTCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGAACCTAGCCAATAGGC
 FC_IR   TTCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGANCCNNNCCAATAGGC
BJS_CD   TTGCCCAGGAGGAGGTCTGCAAAATGCGTCGGAAGAAGTGAGTGATCCAAGCCAATGGGC
BJS_IR   TTGCCCAGGAGGAGGTCTGCAAAATGCGTCGGAAGAAGTGAGTGATCCAAGCCAATGGGC
199_CD   TTCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGAACCTAGCCAATAGGC
199_IR   TTCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGAACCTAGCCAATAGGC
216_CD   TTCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGAACCTAGCCAATAGGC
216_IR   TTCCCCAGGAGGAGGGCTGCAAAATGCGTCGGAAGAAGTGAGTGAACCTAGCCAATAGGC
```

```
 FC_CD   TGATGAAAAATGTGACGCGATTTGCACACAACAATCA-------------CAGCCAATTT
 FC_IR   TGATGAAAAATGTGACGCGATTTGCACACAACAATCA-------------CAGCCAATTT
BJS_CD   TGATGAAAAGTGTGACGCGATTTGCACACAACAATCAAACGAGTAATTGCCAGCCAATTT
BJS_IR   TGATGAAAAGTGTGACGCGATTTGCACACAACAATCAAACGAGTAATTGCCAGCCAATTT
199_CD   TGATGAAAAATGTGACGCGATTTGCACACAACAATCA-------------CAGCCAATTT
199_IR   TGATGAAAAATGTGACGCGATTTGCACACAACAATCA-------------CAGCCAATTT
216_CD   TGATGAAAAATGTGACGCGATTTGCACACAACAATCA-------------CAGCCAATTT
216_IR   TGATGAAAAATGTGACGCGATTTGCACACAACAATCA-------------CAGCCAATTT

 FC_CD   TGGGC---TTTTTTTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAA
 FC_IR   TGGGC---TTTTTTTTTTTGGTAGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAA
BJS_CD   TGGGC--------TTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAA
BJS_IR   TGGGC--------TTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAA
199_CD   TGGGCTTTTTTTTTTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAA
199_IR   TGGGC---TTTTTTTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAA
216_CD   TGGGCTTTTTTTTTTTTTTGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAA
216_IR   TGGGCTTTTTTTTTTTNNNGGTCGGGAATAAATCAAACCTCTGCTCAAATTAGGGCATAA

 FC_CD   TCGCAGGCGTGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGT
 FC_IR   TCGCAGGCGTGACATGCGGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
BJS_CD   TCGCAGGCGTGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGT
BJS_IR   TCGCAGGCGTGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGT
199_CD   TCGCAGGCGTGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGT
199_IR   TCGCAGGCGTGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGT
216_CD   TCGCAGGCGTGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGT
216_IR   TCGCAGGCGTGACATGCGGCTGTCTGCAACACCGGCAGCAGTCAGATAAGATAGGGATGT

 FC_CD   CCTCCTCCCCAGAGTTCGCGGCTTTCACACCCTTGGATACACCGACAAATTTCCATGTGA
 FC_IR   NNNCCTCCCCAGAGTTCGCGGCTTTCACACCNTTGGATACACCGACAAATTTCCATGTGA
BJS_CD   CCTCCTCCCCAGAGTTTGCCGCTTTCACTCCCTTGGATACACCGACAAATTTGTATGTGA
BJS_IR   CCTCCTCCCCAGAGTTTGCCGCTTTCACTCCCTTGGATACACCGACAAATTTGTATGTGA
199_CD   CCTACTCCCCAGCGTTCGCCGCTTTCACACCCTTGGATACACCGACAAATTTCCATGTGA
199_IR   CCTACTCCCCAGCGTTCGCCGCTTTCACACCCTTGGATACACCGACAAATTTCCATGTGA
216_CD   CCTACTCCCCAGCGTTCGCCGCTTTCACACCCTTGGATACACCGACAAATTTCCATGTGA
216_IR   CCTACTCCCCAGCGTTCGCCGCTTTCACACCCTTGGATACACCGACAAATTTCCATGTGA

 FC_CD   TGCACACGAAAAACAGCGCGAAATTTGTTTAAAAGTTGAAAGGAAAAATTTAAAAGTTAG
 FC_IR   TGCACACGAAAAACAGCGCGAAATTTGGTTAAAAGTTGAAAAGAAAAATTTAAAAGTTAG
BJS_CD   TGCACACGGAGAACAGCGTGAAATTTGGTTGAAAGTTCAAATAAAAAATTTTAATGTTAG
BJS_IR   TGCACACGGAGAACAGCGTGAAATTTGGTTGAAAGTTCAAATAAAAAATTTTAATGTTAG
199_CD   TGCACGCGAAAAACAGCGCGAAATTTGGTTGAAAGTTTAAAAGAAAAATTTTAAAGTTAG
199_IR   TGCACGCGAAAAACAGCGCGAAATTTGGTTGAAAGTTTAAAAGAAAAATTTTAAAGTTAG
216_CD   TGCACGCGAAAAACAGCGCGAAATTTGGTTGAAAGTTTAAAAGAAAAATTTTAAAGTTAG
216_IR   TGCACGCGAAAAACAGCGCGAAATTTGGTTGAAAGTTTAAAAGAAAAATTTTAAAGTTAG

 FC_CD   TTCGGGATTACTGTTG------AATACCTTCTTGATCTCGTTGAAAACGCGATTCAAACG
 FC_IR   TTCGGGATTACTGTTG------AATACCTTCTTGATCTCGTTGAAAACGCGATTCAAACG
BJS_CD   TTCGGGATAACTGTTATTTATCAACTGATAATTGATTTCGCTAAAAACTCGATTCAAACA
BJS_IR   TTCGGGATAACTGTTATTTATCAACTGATAATTGATTTCGCTAAAAACTCGATTCAAACA
199_CD   TTCGGGATAACTGTTG------AATTCCTTCTTGATCTCGTTGAAAACGCGATTCAAACG
199_IR   TTCGGGATAACTGTTG------AATTCCTTCTTGATCTCGTTGAAAACGCGATTCAAACG
216_CD   TTCGGGATAACTGTTG------AATTCCTTCTTGATCTCGTTGAAAACGCGATTCAAACG
216_IR   TTCGGGATAACTGTTG------AATTCCTTCTTGATCTCGTTGAAAACGCGATTCAAACG
```

```
 FC_CD   CTCATTGAAAAAGTTCTAAAAAACCAA--------------GACAACGTGA--AAATTTA
 FC_IR   CTCATTGAAAAAGTTCTAAAAAACCAA--------------GACAACGTGA--AAATTTA
BJS_CD   CTCATTGATATAGTCCTAAAAAAAAAGATATATTATATTGCGACAATATGA--AAATTAA
BJS_IR   CTCATTGATATAGTCCTAAAAAAAAAGATATATTATATTGCGACAATATGA--AAATTAA
199_CD   CTCATTGAAAAAGTTCTAAAAAACCAA--------------GACAACGTGG--AAATTTA
199_IR   CTCATTGAAAAAGTTCTAAAAAACCAA--------------GACAACGTGG--AAATTTA
216_CD   CTCATTGAAAAAGTTCTAAAAAACCAA--------------GACAACGTGG--AAATTTA
216_IR   CTCATTGAAAAAGTTCTAAAAAACCAA--------------GACAACGTGGNNAAATTTA

 FC_CD   TACATATTAACTTAGCAATTTAATCACTTTATCGCAGCTGTAAAAAATTAT-------AA
 FC_IR   TACATATTAACTTAGCAATTTAATCACTTTATCGCAGCTGTAAAAAATTAT-------AA
BJS_CD   TATACATTAACTTAGCAATTTAATCACTTTATCCCAGCTGTAAAATATATTGAGTGAAAA
BJS_IR   TATACATTAACTTAGCAATTTAATCACTTTATCCCAGCTGTAAAATATATTGAGTGAAAA
199_CD   TACATATTAACTTAGCAATTTAATCACTTTATCGCAGATGTAAAAAATTTT-------AA
199_IR   TACATATTAACTTAGCAATTTAATCACTTTATCGCAGATGTAAAAAATTTT-------AA
216_CD   TACATATTAACTTAGCAATTTAATCACTTTATCGCAGCTGTAAAAAATTTT-------AA
216_IR   TACATATTAACTTAGCAATTTAATCACTTTATCGCAGCTGTAAAAAATTTT-------AA

 FC_CD   CTTCCTTGTTGAATAAACGACA-ATCATAATACTAAAGCTATTTTCTGAGTGCAGTAGGT
 FC_IR   CTTCCTTGTTGAATAAACGACA-ACCATAATACTAAAACTATTTTCTGAGTGCAGTAGGT
BJS_CD   ATTCCTTGTTCAATAAATTACACACCGAAATACTAAAGCTTTTTTCTGAGTGCAGTAGG-
BJS_IR   ATTCCTTGTTCAATAAATTACACACCGAAATACTAAAGCTTTTTTCTGAGTGCAGTAGG-
199_CD   CTTCCTTGTTGAATAAACGACACTCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
199_IR   CTTCCTTGTTGAATAAACGACACTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-
216_CD   CTTCCTTGTTGAATAAACGACACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN??
216_IR   CTTCCTTGTTGAATAAACGACACTCCGTAATACTAAANNNNNNNNNNNNNNNNNNNNNNN

 FC_CD   ATATTTATCTGGGTTATCGGCGATTAAGCGGCACTTGTCCGCCGTTGCAATTAGCGAAAC
 FC_IR   ATATTTATCTGGGTTATCGGCGATTAAGCGGCACTTGTCCGCCGTTGCAATTAGCGAAAC
BJS_CD   -TATTTATCTGGGTTATCGGCGATTAAGCGGCACTTGTCCGCCGTTGCAATTAGCGAAAC
BJS_IR   -TATTTATCTGGGTTATCGGCGATTAAGCGGCACTTGTCCGCCGTTGCAATTAGCGAAAC
199_CD   NNN??????????????????????????????????????????????NNNNNNGAACC
199_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGAAAC
216_CD   ????????????????????????????????????????????????NNNNNNNNNNNN
216_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

 FC_CD   AATTAGGACAAGCCCCTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAAC
 FC_IR   AATTAGGACAAGCCCCTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAAC
BJS_CD   AATTAGGACAAGCCCCTAGATTAATGCAAATATACTATACTATCAACCATGACATTTAAC
BJS_IR   AATTAGGACAAGCCCCTAGATTAATGCAAATATACTATACTATCAACCATGACATTTAAC
199_CD   AAATAGGACAAGCCCCTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAAC
199_IR   NAATAGGACAAGCCCCTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAAC
216_CD   NNNNNNNNNNNNNCCCTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAAC
216_IR   NNNNNNNNNNNNNNNNTAGATTAATGCAAATATACTATACTATCAACCGTGATATTCAAC

 FC_CD   GTCATCATCCACTC-----TTTTTGGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACG
 FC_IR   GTCATCATCCACTC-----TTTTTGGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACG
BJS_CD   GTCATCATCTACTCTTTTTTTTTTGGATCTTTTAGAGTTGAGTACCGTGTGAAGCAGACG
BJS_IR   GTCATCATCTACTCTTTTTTTTTTGGATCTTTTAGAGTTGAGTACCGTGTGAAGCAGACG
199_CD   GTCATCATCCACTC-----TTTTTGGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACG
199_IR   GTCATCATCCACTC-----TTTTTGGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACG
216_CD   GTCATCATCCACTC-----TTTTTGGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACG
216_IR   GTCATCATCCACTC-----TTTTTGGATCTTACAGAGTTGAGTACCGTGTGAAGCAGACG
```

```
 FC_CD   CCTTCGCGTCCAGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCG
 FC_IR   CCTTCGCGTCCAGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCC
BJS_CD   CCTTCGCGTCCAGTGCAGATGACTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCT
BJS_IR   CCTTCGCGTCCAGTGCAGATGACTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCT
199_CD   CCTTCGCGTCCAGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCC
199_IR   CCTTCGCGTCCAGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCC
216_CD   CCTTCGCGTCCAGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCC
216_IR   CCTTCGCGTCCAGTGCAGATGGCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCC

 FC_CD   CCCGAGGATGAGTTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGG
 FC_IR   CCCGAGGATGAGCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGG
BJS_CD   CCCGAAGATGAGCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGG
BJS_IR   CCCGAAGATGAGCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGG
199_CD   CCCGAGGATGAGCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGG
199_IR   CCCGAGGATGAGCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGG
216_CD   CCCGAGGATGAGCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGG
216_IR   CCCGAGGATGAGCTGGCGGCGCACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGG

 FC_CD   CTCAAAATCAAGAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGA
 FC_IR   CTCAAAATCAAGAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGA
BJS_CD   CTCAAAATCAAGAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGA
BJS_IR   CTCAAAATCAAGAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGA
199_CD   CTCAAAATCAAGAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGA
199_IR   CTCAAAATCAAGAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGA
216_CD   CTCAAAATCAAGAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGA
216_IR   CTCAAAATCAAGAACCGCGACTGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCGTGA

 FC_CD   GTATCCA------CAATCAATACATACGGGGTGCTCCATATGGATAAGGCATCAGGATGG
 FC_IR   GTATCCA------CAATCAATACATACNGGGTGCTCCATATGGATAAGGCATCAGGATGG
BJS_CD   GTATCAGTACAGTCAATCAGTAAATACGGTGTGCTCCATATGGATAAGGCATCAGGAAGG
BJS_IR   GTATCAGTACAGTCAATCAGTAAATACGGTGTGCTCCATATGGATAAGGCATCAGGAAGG
199_CD   GTATCCA------CAATCAATACATACGGGGTGCTCCATATGGATAAGGCATCAGGATGG
199_IR   GTATCCA------CAATCAATACATACGGGGTGCTCCATATGGATAAGGCATCAGGATGG
216_CD   GTATCCA------CAATCAATACATACGGGGTGCTCCATATGGATAAGGCATCAGGATGG
216_IR   GTATCCA------CAATCAATACATACGGGGTGCTCCATATGGATAAGGCATCAGGATGG

 FC_CD   ACGCGATAAGGGTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGAC
 FC_IR   ACGCGATAAGGGTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGAC
BJS_CD   CCGCGATAAGGGTTTGCGTCCGTGCTAGGTACGTGATCGGGAATAGTGACGAACGCAAAC
BJS_IR   CCGCGATAAGGGTTTGCGTCCGTGCTAGGTACGTGATCGGGAATAGTGACGAACGCAAAC
199_CD   ACGCGATAAGGGTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGAC
199_IR   ACGCGATAAGGGTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGAC
216_CD   ACGCGATAAGGGTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGAC
216_IR   ACGCGATAAGGGTTTGCGTCCGGGCTAGGTACGTGATCGGGAATAGTGACGAACGCAGAC

 FC_CD   CCACCAAATGACCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTA
 FC_IR   CCACCAAATGACCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTA
BJS_CD   TCACCAAATGACCATATGCTACCATATGCCGCATCTGAT---------AGTGCCATCTCTA
BJS_IR   TCNNNNNATGACCATATGCTACNATATGCCGCATCTGAT---------ANTGNCATCTCNN
199_CD   CCACCAAATGACCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTA
199_IR   CCACCAAATGACCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTA
216_CD   CCACCAAATGACCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTA
216_IR   CCACCAAATGACCATATGCTACCATATGCCGCATATGATGGTATATGAGTGCCATCTCTA
```

```
 FC_CD   TGGCTGATGTCATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACC
 FC_IR   TGGCTGATGTCATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACC
BJS_CD   TGGCTGATGTCATTTTCGTAATCACTGTTATTA-GAAAATGGAAATAATTTTAAATAACC
BJS_IR   TGGCTGATGTCATTTTCNNNNNNNNNGTTNTTA-GAAAANGGNAATANTTTTAAATANCC
199_CD   TGGCTGATGTCATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACC
199_IR   TGGCTGATGTCATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACC
216_CD   TGGCTGATGTCATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACC
216_IR   TGGCTGATGTCATTTTCGTAATCACTGTTGAAACGAATATGGAAATAATTTTAAATAACC

 FC_CD   TTGAAGTATTTAATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTAGTCT
 FC_IR   TTGAAGTATTTAATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTANTCT
BJS_CD   TTGAAGTACTTAATATCCATTTACAAATAAATATGTTTCAAGAATTGCCCATTTTAGACA
BJS_IR   TTGAAGTANTTAATATCCNTTTNCAAATAAATATGNTTCAAGAATTGCCCNTTTTAGNCA
199_CD   TTGAAGTATTTAATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTAGTCT
199_IR   TTGAAGTATTTAATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTAGTCT
216_CD   TTGAAGTATTTAATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTAGTCT
216_IR   TTGAAGTATTTAATATCCATTTACAATTAAATATGTTTCAAGAATTGCCC--TGTAGTCT

 FC_CD   GTGTAAAAATTGGATTTTAATATAGATTA-AAAAATCTCATTTAATTATTTGGAAATTCA
 FC_IR   GTGTAAAAATTGGANNNNNNNNNNNNNNA-AAAAATCTCATTTAATTATTTGGAAATTCA
BJS_CD   GTTT-AAAATTTGATTTTAATATAAATTACAAAGATCTCAGATAATTATTTGGAAATTAA
BJS_IR   GTTT-AAAATTTGATTTTAATATAAATTACAAAGATNTCAGATAATTATTTGGAAATTAA
199_CD   GTGTAAAAATTGGATTTTAATATAGATTA-AAAAATCTCATTTAATTATTTGGAAATTCA
199_IR   GTGTAAAAATTGGATTTTAATATAGATTA-AAAAATCTCATTTAATTATTTGGAAATTCA
216_CD   GTGTAAAAATTGGATTTTAATATAGATTA-AAAAATCTCATTTAATTATTTGGAAATTCA
216_IR   GTGTAAAAATTGGATTTTAATATAGATTA-AAAAATCTCATTTAATTATTTGGAAATTCA

 FC_CD   TATCTTTAAATTCTAAATATTTAG------AACCGCAAGAATATGCAAATAACAGTCGTC
 FC_IR   TATCTTTAAATTCTAAATATTTAG------ANCNGCAAGAATATGCAAATAACAGTCGTC
BJS_CD   TATCCAGAAATATATAGAATTTAAGGGGGTAACCTCAAGAATATGCAAATAACAGTAGTC
BJS_IR   TATCCAGAAATATATAGAATTTAAGGGGGTAACCTCAAGAATATGCAAATANCAGTAGTC
199_CD   TATCTTTAAATTCTAAATATTTAG------AACCGCAAGAATATGCAAATAACAGTCGTC
199_IR   TATCTTTAAATTCTAAATATTTAG------AACCGCAAGAATATGCAAATAACAGTCGTC
216_CD   TATCTTTAAATTCTAAATATTTAG------AACCGCAAGAATATGCAAATAACAGTCGTC
216_IR   TATCTTTAAATTCTAAATATTTAG------AACCGCAAGAATATGCAAATAACAGTCGTC

 FC_CD   TTTTTAATCTAATAAAAAACATGCAAAGTTTTAGAATTCTGAGAAAAATGCATACT----
 FC_IR   TTTTTAATCTAATAAAAAACATGCAAAGTTTTAGAATTNTGAGAAAAATGCATACT----
BJS_CD   TTTTTATTATAATAAAAAA-ATGCAGATCTCCTGTGTTGTTAATGGTATCCGCTTTTTTT
BJS_IR   TTTTTATTATAATAAAAAA-ATGCAGATNTCCTGTGTTGTTAATGGTATCCGCTTTTTTT
199_CD   TTTTTAATCTAATAAAAAACATGCAAAGTTTTAGAATTCTGAGAAAAATGCATACT----
199_IR   TTTTTAATCTAATAAAAAACATGCAAAGTTTTAGAATTCTGAGAAAAATGCATACT----
216_CD   TTTTTAATCTAATAAAAAACATGCAAAGTTTTAGAATTCTGAGAAAAATGCATACT----
216_IR   TTTTTAATCTAATAAAAAACATGCAAAGTTTTAGAATTCTGAGAAAAATGCATACT----

 FC_CD   ----CATAAGCGACCACTTTCTAAACGTTTTTTCTCAGTTCCGCAGACTTGCCGGCCAGT
 FC_IR   ----CATAAGCGACCACTTTCTAAACGTTTTTTCTCANTTCCGCAGNNTNGCCGGCCAGT
BJS_CD   TTTATATTAGCGACCACTTTCTAAACGCTTTTCCTCAGTTCCGCAGACTTACCGGCCAGT
BJS_IR   TTTATATTAGCGACCACTTTCTAAACGCTTTTCCTCAGTTCCGCAGACTTACCGGCCAGT
199_CD   ----CATAAGCGACCACTTTCTAAACGTTTTTTCTCAGTTCCGCAGACTTGCCGGCCAGT
199_IR   ----CATAAGCGACCACTTTCTAAACGTTTTTTCTCAGTTCCGCAGACTTGCCGGCCAGT
216_CD   ----CATAAGCGACCACTTTCTAAACGTTTTTTCTCAGTTCCGCAGACTTGCCGGCCAGT
216_IR   ----CATAAGCGACCACTTTCTAAACGTTTTTTCTCAGTTCCGCAGACTTGCCGGCCAGT
```

```
 FC_CD   TTGGGTGGCACACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACT
 FC_IR   TTGGGTGGCACACCCAAAAAGACCAGGACGNNNNNNNNNNNNNNNNNNNGGCGCAACNCT
BJS_CD   GTGGGTGGCACCCCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACT
BJS_IR   GTGGGTGGCACCCCCAAAAAGNCCAGGACGGCCAGATCCCGCGTCCTGCGGCGCNNCNNN
199_CD   TTGGGTGGCACACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACT
199_IR   TTGGGTGGCACACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACT
216_CD   TTGGGTGGCACACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACT
216_IR   TTGGGTGGCACACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACT

 FC_CD   ATGGACTGCGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAG
 FC_IR   ATGGACTGCGCCNTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAG
BJS_CD   ATGGGCT-CTCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAG
BJS_IR   NNNNNNN-CGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAG
199_CD   ATGGACTGCGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAG
199_IR   ATGGACTGCGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAG
216_CD   ATGGACTGCGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAG
216_IR   ATGGACTGCGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAG

 FC_CD   CGGCTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACG
 FC_IR   CGGCTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACG
BJS_CD   CGGCTGCAGTCACTCCTTCGCGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACG
BJS_IR   CGGCTGCAGTCACTCCTTCGCGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACG
199_CD   CGGCTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACG
199_IR   CGGCTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACG
216_CD   CGGCTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACG
216_IR   CGGCTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACG

 FC_CD   GCGGTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTG
 FC_IR   GCGGTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTG
BJS_CD   GCGGTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCGCTGGAGTCTCTG
BJS_IR   GCGGTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCGCTGGAGTCTCTG
199_CD   GCGGTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTG
199_IR   GCGGTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTG
216_CD   GCGGTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTG
216_IR   GCGGTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTG

 FC_CD   GACCAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCC
 FC_IR   GACCAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCC
BJS_CD   GACCAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCC
BJS_IR   GACCAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCC
199_CD   GACCAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCC
199_IR   GACCAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCC
216_CD   GACCAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCC
216_IR   GACCAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCC

 FC_CD   AAGGTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGC
 FC_IR   AAGGTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGC
BJS_CD   AAGGTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGC
BJS_IR   AAGGTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGC
199_CD   AAGGTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGC
199_IR   AAGGTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGC
216_CD   AAGGTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGC
216_IR   AAGGTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGC
```

```
 FC_CD   TCCTCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGAT
 FC_IR   TCCTCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGAT
BJS_CD   TCCTCGGCGCCGCCGTTGGCTTTCGCTGCCCAACTTCCCGCTGGAATGCTAAGGAGTGAT
BJS_IR   TCCTCGGCGCCGCCGTTGGCTTTCGCTGCCCAACTTCCCGCTGGAATGCTAAGGAGTGAT
199_CD   TCCTCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGAT
199_IR   TCCTCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGAT
216_CD   TCCTCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGAT
216_IR   TCCTCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGAT

 FC_CD   GGAAGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTA
 FC_IR   GGAAGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTA
BJS_CD   GGAAGAGCAACACCTGGGTTGGGGAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTA
BJS_IR   GGAAGAGCAACACCTGGGTTGGGGAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTA
199_CD   GGAAGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTA
199_IR   GGAAGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTA
216_CD   GGAAGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTA
216_IR   GGAAGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTA

 FC_CD   ATTAAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGG
 FC_IR   ATTAAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGG
BJS_CD   ATTAAGCTCGGTAGCGTAGATCGACAGGACGCCAAACTATCACAGGAGGAGCACACCTGG
BJS_IR   ATTAAGCTCGGTAGCGTAGATCGACAGGACGCCAAACTATCACAGGAGGAGCACACCTGG
199_CD   ATTAAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGG
199_IR   ATTAAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGG
216_CD   ATTAAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGG
216_IR   ATTAAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGG

 FC_CD   CAGACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACG
 FC_IR   CAGACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACG
BJS_CD   CAGACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACT
BJS_IR   CAGACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACT
199_CD   CAGACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACG
199_IR   CAGACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACG
216_CD   CAGACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACG
216_IR   CAGACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACG

 FC_CD   GTGGAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACC
 FC_IR   GTGGAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACC
BJS_CD   GTGGAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACC
BJS_IR   GTGGAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACC
199_CD   GTGGAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACC
199_IR   GTGGAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACC
216_CD   GTGGAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACC
216_IR   GTGGAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACC

 FC_CD   CACATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTG
 FC_IR   CACATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTG
BJS_CD   CACATTATAAACTACAAGTTCCAGCCGCGTTATCGCCGCGAGCCGAGTCTGATAAGCTTG
BJS_IR   CACATTATAAACTACAAGTTCCAGCCGCGTTATCGCCGCGAGCCGAGTCTGATAAGCTTG
199_CD   CACATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTG
199_IR   CACATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTG
216_CD   CACATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTG
216_IR   CACATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTG
```

```
 FC_CD   GATAGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCA
 FC_IR   GATAGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCA
BJS_CD   GATAGTGGCACTCATTCCGATAGCAATTCTAATGCCAGCTGTATGTGAAACCATTCTTCA
BJS_IR   GATAGTGGCACTCATTCCGATAGCAATTCTAATGCCAGCTGTATGTGAAACCATTCTTCA
199_CD   GATAGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCA
199_IR   GATAGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCA
216_CD   GATAGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCA
216_IR   GATAGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTGTATGTCAAACCATCATTCA

 FC_CD   AAAGTAAACCCCTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAG
 FC_IR   AAAGTAAACCCCTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAG
BJS_CD   AAAGTAAACCCCTTATTATTTATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAG
BJS_IR   AAAGTAAACCCCTTATTATTTATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAG
199_CD   AAAGTAAACCCCTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAG
199_IR   AAAGTAAACCCCTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAG
216_CD   AAAGTAAACCCCTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAG
216_IR   AAAGTAAACCCCTTGTTATTCATTATTTAAATCTATTCTCCAGCTCCACTGCCCAGCAAG

 FC_CD   ATGTGCCAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCG
 FC_IR   ATGTGCCAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCG
BJS_CD   ATGTGCCAAGGATGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCG
BJS_IR   ATGTGCCAAGGATGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCG
199_CD   ATGTGCCAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCG
199_IR   ATGTGCCAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCG
216_CD   ATGTGCCAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCG
216_IR   ATGTGCCAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCG

 FC_CD   CTGCGAGAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCC
 FC_IR   CTGCGAGAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCC
BJS_CD   CTGCGAGAAGTGGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCC
BJS_IR   CTGCGAGAAGTGGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCC
199_CD   CTGCGAGAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCC
199_IR   CTGCGAGAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCC
216_CD   CTGCGAGAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCC
216_IR   CTGCGAGAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCC

 FC_CD   TCGCAGGCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATC
 FC_IR   TCGCAGGCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATC
BJS_CD   TCACAGGCCATGGGCGCCCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATC
BJS_IR   TCACAGGCCATGGGCGCCCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATC
199_CD   TCGCAGGCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATC
199_IR   TCGCAGGCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATC
216_CD   TCGCAGGCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATC
216_IR   TCGCAGGCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATC

 FC_CD   AAGTCCATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTA
 FC_IR   AAGTCCATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTA
BJS_CD   AAGTCCATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTTTA
BJS_IR   AAGTCCATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTTTA
199_CD   AAGTCCATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTA
199_IR   AAGTCCATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTA
216_CD   AAGTCCATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTA
216_IR   AAGTCCATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTA
```

```
 FC_CD   GGAAAAATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCC
 FC_IR   GGAAAAATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCC
BJS_CD   GGAAAAATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCC
BJS_IR   GGAAAAATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGNGCACTTCCTCC
199_CD   GGAAAAATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCC
199_IR   GGAAAAATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCC
216_CD   GGAAAAATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCC
216_IR   GGAAAAATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCC

 FC_CD   TACATCGCCTCCGACAGCGGT-AGTTCATCAGCATC-GGGTGTAGAGAACGATGATTCTG
 FC_IR   TACATCGCCTCCGACAGCGGTNAGTTCATCAGCATC-GGGTGTAGAGAACGATGATTCTG
BJS_CD   TACATCGCCACCGACAGCGGT-AGTTCATCAGCATCGGGGTGTAGAGAACGATGATTCGG
BJS_IR   TACATCGCCACCGACAGCGGT-AGTTCATCAGCATCNGGGTGTAGAGAACGATGATTCGG
199_CD   TACATCGCCTCCGACAGCGGT-AGTTCATCAGCATC-GGGTGTAGAGAACGATGATTCTG
199_IR   TACATCGCCTCCGACAGCGGT-AGTTCATCAGCATCGGGGTGTAGAGAACGATGATTCTG
216_CD   TACATCGCCTCCGACAGCGGT-AGTTCATCAGCATC-GGGTGTAGAGAACGATGATTCTG
216_IR   TACATCGCCTCCGACAGCGGT-AGTTCATCAGCATC-GGGTGTAGAGAACGATGATTCTG

 FC_CD   CTGTTAACTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGG
 FC_IR   CTGTTAACTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGG
BJS_CD   CTGTTAACTCAATGGATTCATCAAAGCCGCCCAGTATGGCCGGATCAGCCTCGCGAAAGG
BJS_IR   CTGTTAACTCAATGGATTCATCAAAGCCGCCCAGTATGGCCGGATCAGCCTCGCGAAAGG
199_CD   CTGTTAACTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGG
199_IR   CTGTTAACTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGG
216_CD   CTGTTAACTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGG
216_IR   CTGTTAACTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGG

 FC_CD   GTGTGACACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCG
 FC_IR   GTGTGACACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCG
BJS_CD   GTGTGACACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCG
BJS_IR   GTGTGACACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCG
199_CD   GTGTGACACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCG
199_IR   GTGTGACACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCG
216_CD   GTGTGACACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCG
216_IR   GTGTGACACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCG

 FC_CD   AGACATCAAGTAGCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATAT
 FC_IR   AGACATCAAGTAGCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATAT
BJS_CD   AGACATCAAGTAGCAATGAGTAAGTATATACGAAGTAATACGAAATTGAAGAAAAGAAAT
BJS_IR   AGACATCAAGTAGCAATGAGTAAGTATATACGAAGTAATACGAAATTGAAGAAAAGAAAT
199_CD   AGACATCAAGTAGCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATAT
199_IR   AGACATCAAGTAGCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATAT
216_CD   AGACATCAAGTAGCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATAT
216_IR   AGACATCAAGTAGCAATGAGTATGTATATACGAAGTAGTACGAAATTGAAGAAAACATAT

 FC_CD   AACATACATTTTCCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCT
 FC_IR   AACATACATTTTCCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCT
BJS_CD   AATATACATTTTCCTTCAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCT
BJS_IR   AATATACATTTTCCTTCAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCT
199_CD   AACATACATTTTCCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCT
199_IR   AACATACATTTTCCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCT
216_CD   AACATACATTTTCCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCT
216_IR   AACATACATTTTCCTTTAGATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCT
```

```
 FC_CD   CTCACAGCCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCAT
 FC_IR   CTCACAGCCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCAT
BJS_CD   CTCACAGCCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCAT
BJS_IR   CTCACAGCCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCAT
199_CD   CTCACAGCCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCAT
199_IR   CTCACAGCCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCAT
216_CD   CTCACAGCCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCAT
216_IR   CTCACAGCCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCAT

 FC_CD   TGGGTACGTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCA
 FC_IR   TGGGTACGTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCA
BJS_CD   TGGGTACGTGCAGCAGCAATGGGACAGGAGAGCAATCCACTGGGCTGTATCGCAAGTTCA
BJS_IR   TGGGTACGTGCAGCAGCAATGGGACAGGAGAGCAATCCACTGGGCTGTATCGCAAGTTCA
199_CD   TGGGTACGTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCA
199_IR   TGGGTACGTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCA
216_CD   TGGGTACGTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCA
216_IR   TGGGTACGTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCA

 FC_CD   TAGAAAATGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGC
 FC_IR   TAGAAAATGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGC
BJS_CD   TTGAAAATGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCCTTTCTACACAAGCTGC
BJS_IR   TTGAAAATGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCCTTTCTACACAAGCTGC
199_CD   TAGAAAATGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGC
199_IR   TAGAAAATGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGC
216_CD   TAGAAAATGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGC
216_IR   TAGAAAATGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGC

 FC_CD   ACAACGTGGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAG
 FC_IR   ACAACGTGGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAG
BJS_CD   ACAACGTGGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTCG
BJS_IR   ACAACGTGGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTCG
199_CD   ACAACGTGGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAG
199_IR   ACAACGTGGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAG
216_CD   ACAACGTGGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAG
216_IR   ACAACGTGGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAG

 FC_CD   ATTACGAGTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCA
 FC_IR   ATTACGAGTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCA
BJS_CD   ATTACGAGTCGGATGCAGAGTCCATTGAAGAAGACGTACCAAGGCTAGATCCTGAAATCA
BJS_IR   ATTACGAGTCGGATGCAGAGTCCATTGAAGAAGACGTACCAAGGCTAGATCCTGAAATCA
199_CD   ATTACGAGTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCA
199_IR   ATTACGAGTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCA
216_CD   ATTACGAGTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCA
216_IR   ATTACGAGTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCA

 FC_CD   GCCGCGAACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCA
 FC_IR   GCCGCGAACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCA
BJS_CD   GCCGCGAACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCA
BJS_IR   GCCGCGAACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCA
199_CD   GCCGCGAACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCA
199_IR   GCCGCGAACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCA
216_CD   GCCGCGAACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCA
216_IR   GCCGCGAACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCA
```

```
 FC_CD   TAAATCTGCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCA
 FC_IR   TAAATCTGCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCA
BJS_CD   TAAACCTGCCGTCGTACATTCCCACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCA
BJS_IR   TAAACCTGCCGTCGTACATTCCCACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCA
199_CD   TAAATCTGCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCA
199_IR   TAAATCTGCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCA
216_CD   TAAATCTGCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCA
216_IR   TAAATCTGCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCA

 FC_CD   TGCATGAGTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCC
 FC_IR   TGCATGAGTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCC
BJS_CD   TGCATGAGTTCTTGCGCATGCGTCTCGAAACCAGACCCGTTCGACCCAATCCACTGAGCC
BJS_IR   TGCATGAGTTCTTGCGCATGCGTCTCGAAACCAGACCCGTTCGACCCAATCCACTGAGCC
199_CD   TGCATGAGTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCC
199_IR   TGCATGAGTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCC
216_CD   TGCATGAGTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCC
216_IR   TGCATGAGTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCC

 FC_CD   TGGAACAGTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGC
 FC_IR   TGGAACAGTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGC
BJS_CD   TGGAACAGTTGATGAAGGAGCTTAGGGAGGGCTTGACCCTAGCCTTGACCCATCGGGAGC
BJS_IR   TGGAACAGTTGATGAAGGAGCTTAGGGAGGGCTTGACCCTAGCCTTGACCCATCGGGAGC
199_CD   TGGAACAGTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGC
199_IR   TGGAACAGTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGC
216_CD   TGGAACAGTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGC
216_IR   TGGAACAGTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGC

 FC_CD   GATACCAGCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATA
 FC_IR   GATACCAGCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATA
BJS_CD   GATACCAGCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTGGGCAGTTATA
BJS_IR   GATACCAGCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTGGGCAGTTATA
199_CD   GATACCAGCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATA
199_IR   GATACCAGCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATA
216_CD   GATACCAGCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATA
216_IR   GATACCAGCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATA

 FC_CD   TTAGCATTTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGT
 FC_IR   TTAGCATTTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGT
BJS_CD   TTAGCATTTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTATTTGGAGT
BJS_IR   TTAGCATTTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTATTTGGAGT
199_CD   TTAGCATTTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGT
199_IR   TTAGCATTTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGT
216_CD   TTAGCATTTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGT
216_IR   TTAGCATTTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGT

 FC_CD   ACATCGATCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGA
 FC_IR   ACATCGATCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGA
BJS_CD   ACATCGATCAACTAGTGCTGGTTGCCGTTCCGGAAGGCAACCAGAAGTCCGTCCTTGAGA
BJS_IR   ACATCGATCAACTAGTGCTGGTTGCCGTTCCGGAAGGCAACCAGAAGTCCGTCCTTGAGA
199_CD   ACATCGATCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGA
199_IR   ACATCGATCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGA
216_CD   ACATCGATCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGA
216_IR   ACATCGATCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGA
```

```
 FC_CD   AGGAATGGATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCAT
 FC_IR   AGGAATGGATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCAT
BJS_CD   AGGAATGGATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCAT
BJS_IR   AGGAATGGATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCAT
199_CD   AGGAATGGATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCAT
199_IR   AGGAATGGATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCAT
216_CD   AGGAATGGATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCAT
216_IR   AGGAATGGATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCAT

 FC_CD   CTCAAAAATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGA
 FC_IR   CTCAAAAATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGA
BJS_CD   CTCAAAAATTCTGTGGCATAATCAGCAAACTTTTGCGGAGCATTTCCGAACGCCTGGTGA
BJS_IR   CTCAAAAATTCTGTGGCATAATCAGCAAACTTTTGCGGAGCATTTCCGAACGCCTGGTGA
199_CD   CTCAAAAATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGA
199_IR   CTCAAAAATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGA
216_CD   CTCAAAAATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGA
216_IR   CTCAAAAATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGA

 FC_CD   AGCGTACCGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGG
 FC_IR   AGCGTACCGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGG
BJS_CD   AGCGCACCGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGG
BJS_IR   AGCGCACCGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGG
199_CD   AGCGTACCGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGG
199_IR   AGCGTACCGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGG
216_CD   AGCGTACCGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGG
216_IR   AGCGTACCGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGG

 FC_CD   AGGTTAAATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAAC
 FC_IR   AGGTTAAATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAAC
BJS_CD   AGGTTAAATGGCAGCTGCTGACCATTTGCAGAGAGACACAATCGCTGCTCACAGTCGAAC
BJS_IR   AGGTTAAATGGCAGCTGCTGACCATTTGCAGAGAGACACAATCGCTGCTCACAGTCGAAC
199_CD   AGGTTAAATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAAC
199_IR   AGGTTAAATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAAC
216_CD   AGGTTAAATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAAC
216_IR   AGGTTAAATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAAC

 FC_CD   GGGAGCGCTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGA
 FC_IR   GGGAGCGCTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGA
BJS_CD   GGGAGCGCTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTCGAGACGA
BJS_IR   GGGAGCGCTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTCGAGACGA
199_CD   GGGAGCGCTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGA
199_IR   GGGAGCGCTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGA
216_CD   GGGAGCGCTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGA
216_IR   GGGAGCGCTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGA

 FC_CD   CGGATTTTCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCT
 FC_IR   CGGATTTTCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCT
BJS_CD   CGGATTTTCATCGCGAGCACTATGAACATGACGTGGCCAACCAACAGCATGACTTCATCT
BJS_IR   CGGATTTTCATCGCGAGCACTATGAACATGACGTGGCCAACCAACAGCATGACTTCATCT
199_CD   CGGATTTTCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCT
199_IR   CGGATTTTCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCT
216_CD   CGGATTTTCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCT
216_IR   CGGATTTTCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCT
```

```
 FC_CD   GTTCGGACGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACA
 FC_IR   GTTCGGACGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACA
BJS_CD   GTTCGGACGTAAAGGCGGCGTTTAAACTCTTGCAGCAGGACGTTCTGCAAGTGCGCAACA
BJS_IR   GTTCGGACGTAAAGGCGGCGTTTAAACTCTTGCAGCAGGACGTTCTGCAAGTGCGCAACA
199_CD   GTTCGGACGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACA
199_IR   GTTCGGACGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACA
216_CD   GTTCGGACGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACA
216_IR   GTTCGGACGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACA

 FC_CD   AGTTGACGGCAATAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATT
 FC_IR   AGNTGACGGCAATAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATT
BJS_CD   AGTTGACGGCAATAATTGAAGGAGTACAAAAACGGTGCTGTTTGAGCAACATGCGAGATT
BJS_IR   AGTTGACGGCAATAATTGAAGGAGTACAAAAACGGTGCTGTTTGAGCAACATGCGAGATT
199_CD   AGTTGACGGCAATAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATT
199_IR   AGTTGACGGCAATAATGGAAGNAGTGCAAAAACGTTGCTGTTTGANNNNNNNNNNNNNNN
216_CD   AGTTGACGGCAATAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATT
216_IR   AGTTGACGGCAATAATTGAAGGAGTGCAAAAACGTTGCTGTTTNNNNNNNNNNNNNNNNN

 FC_CD   TGGATGAGCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGT
 FC_IR   TGGATGAGCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGT
BJS_CD   TGGATGAGCAAGACAAACAGGCTGTGCTGTCGCGCACTCGAGAGATTTTGCATCAGGGGT
BJS_IR   TGGATGAGCAAGACAAACAGGCTGTGCTGTCGCGCACTCGAGAGATTTTGCATCAGGGGT
199_CD   TTTATGAGCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGT
199_IR   NNNNNNNNCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGT
216_CD   TGGATGAGCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGT
216_IR   NNNNNNNNCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGT

 FC_CD   ACAAATTTGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGG
 FC_IR   ACAAATTTGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGG
BJS_CD   ACAAATTTGGATTTGAATATCACAAGGACGTCATTCGGTTGTTCGAGCAGAAGATTATGG
BJS_IR   ACAAATTTGGATTTGAATATCACAAGGACGTCATTCGGTTGTTCGAGCAGAAGATTATGG
199_CD   ACAAATTTGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGG
199_IR   ACAAATTTGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGG
216_CD   ACAAATTTGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGG
216_IR   ACAAATTTGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGG

 FC_CD   ACCAAAAGGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTA
 FC_IR   ACCAAAAGGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTA
BJS_CD   ACCAAAAGGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTA
BJS_IR   ACCAAAAGGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTA
199_CD   ACCAAAAGGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTA
199_IR   ACCAAAAGGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTA
216_CD   ACCAAAAGGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTA
216_IR   ACCAAAAGGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTA

 FC_CD   AGATGTGGATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTT
 FC_IR   AGATGTGGATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTT
BJS_CD   AGATGTGGATGCATTTCGTAATGGAGCGTTGCGAACGTGGACGAGGAATGCGTCCGCGTT
BJS_IR   AGATGTGGATGCATTTCGTAATGGAGCGTTGCGAACGTGGACGAGGAATGCGTCCGCGTT
199_CD   AGATGTGGATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTT
199_IR   AGATGTGGATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTT
216_CD   AGATGTGGATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTT
216_IR   AGATGTGGATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTT
```

```
 FC_CD   GGGCTTCCCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACT
 FC_IR   GGGCTTCCCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACT
BJS_CD   GGGCCTCCCAGGGTCTGGAGTTTCTGATTCTTGCCTGTGATCCACAAATCACCCAGCACT
BJS_IR   GGGCCTCCCAGGGTCTGGAGTTTCTGATTCTTGCCTGTGATCCACAAATCACCCAGCACT
199_CD   GGGCTTCCCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACT
199_IR   GGGCTTCCCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACT
216_CD   GGGCTTCCCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACT
216_IR   GGGCTTCCCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACT

 FC_CD   TGGACGACGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGA
 FC_IR   TGGACGACGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGA
BJS_CD   TGGATGATGACGAGTTTGAGGCGCTAAAGCAGCAAATGGATCGCTGTATTTCGCACGTGA
BJS_IR   TGGATGATGACGAGTTTGAGGCGCTAAAGCAGCAAATGGATCGCTGTATTTCGCACGTGA
199_CD   TGGACGACGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGA
199_IR   TGGACGACGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGA
216_CD   TGGACGACGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGA
216_IR   TGGACGACGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGA

 FC_CD   TTGGCATCACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAA
 FC_IR   TTGGCATCACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAA
BJS_CD   TTGGCATCACCTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGCA
BJS_IR   TTGGCATCACCTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGCA
199_CD   TTGGCATCACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAA
199_IR   TTGGCATCACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAA
216_CD   TTGGCATCACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAA
216_IR   TTGGCATCACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAA

 FC_CD   AGACTTCATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTG
 FC_IR   AGACTTCATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTG
BJS_CD   AGACTTCATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTG
BJS_IR   AGACTTCATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTG
199_CD   AGACTTCATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTG
199_IR   AGACTTCATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTG
216_CD   AGACTTCATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTG
216_IR   AGACTTCATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTG

 FC_CD   GCATGGTTCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGC
 FC_IR   GCATGGTTCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGC
BJS_CD   GCATGGTCCTTAATCCTAATACGCCGCCACTGCAGTCGCCACCGTACAACAAGTTACTGC
BJS_IR   GCATGGTCCTTAATCCTAATACGCCGCCACTGCAGTCGCCACCGTACAACAAGTTACTGC
199_CD   GCATGGTTCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGC
199_IR   GCATGGTTCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGC
216_CD   GCATGGTTCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGC
216_IR   GCATGGTTCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGC

 FC_CD   ATCCGCAGTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTT
 FC_IR   ATCCGCAGTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTT
BJS_CD   ATCCGCAGTTCAGCTTAAAGGAGGATGTGTCAGGAACTTCGTACAGTCCCGTTGACAGTT
BJS_IR   ATCCGCAGTTCAGCTTAAAGGAGGATGTGTCAGGAACTTCGTACAGTCCCGTTGACAGTT
199_CD   ATCCGCAGTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTT
199_IR   ATCCGCAGTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTT
216_CD   ATCCGCAGTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTT
216_IR   ATCCGCAGTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTT
```

```
 FC_CD   CAGACTATGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGC
 FC_IR   CAGACTATGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGC
BJS_CD   CAGACTATGTTGACACTCCGTGCCAAAGGAGTGCCAACGGCGAGCTGCGTCTGCTGGTGC
BJS_IR   CAGACTATGTTGACACTCCGTGCCAAAGGAGTGCCAACGGCGAGCTGCGTCTGCTGGTGC
199_CD   CAGACTATGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGC
199_IR   CAGACTATGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGC
216_CD   CAGACTATGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGC
216_IR   CAGACTATGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGC

 FC_CD   CCCAGACGCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGG
 FC_IR   CCCAGACGCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGG
BJS_CD   CCCAGACGCCGCCGACTCCAGCATCTCCCGGAAAGAGCAGCCTCGAAAGTACACCGCTGG
BJS_IR   CCCAGACGCCGCCGACTCCAGCATCTCCCGGAAAGAGCAGCCTCGAAAGTACACCGCTGG
199_CD   CCCAGACGCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGG
199_IR   CCCAGACGCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGG
216_CD   CCCAGACGCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGG
216_IR   CCCAGACGCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGG

 FC_CD   CCTTGCGACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACG
 FC_IR   CCTTGCGACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACG
BJS_CD   CCTTGCGGCAGGAACGTGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACG
BJS_IR   CCTTGCGGCAGGAACGTGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACG
199_CD   CCTTGCGACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACG
199_IR   CCTTGCGACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACG
216_CD   CCTTGCGACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACG
216_IR   CCTTGCGACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACG

 FC_CD   GGCTGCGCGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGC
 FC_IR   GGCTGCGCGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGC
BJS_CD   GGCTACGCGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTTAATTCCAGCGACAAGGTGC
BJS_IR   GGCTACGCGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTTAATTCCAGCGACAAGGTGC
199_CD   GGCTGCGCGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGC
199_IR   GGCTGCGCGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGC
216_CD   GGCTGCGCGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGC
216_IR   GGCTGCGCGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGC

 FC_CD   ATATAAGGGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGAC
 FC_IR   ATATAAGGGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGAC
BJS_CD   ATATAAGGGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGAC
BJS_IR   ATATAAGGGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGAC
199_CD   ATATAAGGGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGAC
199_IR   ATATAAGGGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGAC
216_CD   ATATAAGGGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGAC
216_IR   ATATAAGGGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGAC

 FC_CD   GATTCGGCAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAG
 FC_IR   GATTCGGCAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAG
BJS_CD   GATTCGGCAAGGTATACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAG
BJS_IR   GATTCGGCAAGGTATACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAG
199_CD   GATTCGGCAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAG
199_IR   GATTCGGCAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAG
216_CD   GATTCGGCAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAG
216_IR   GATTCGGCAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAG
```

```
 FC_CD   AGATCGCAATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGA
 FC_IR   AGATCGCAATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGA
BJS_CD   AGATCGCAATCCAGCCGGGAGAGACGCGAGCACTCAAGAATGTGGCCGAAGAGCTAAAGA
BJS_IR   AGATCGCAATCCAGCCGGGAGAGACGCGAGCACTCAAGAATGTGGCCGAAGAGCTAAAGA
199_CD   AGATCGCAATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGA
199_IR   AGATCGCAATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGA
216_CD   AGATCGCAATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGA
216_IR   AGATCGCAATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGA

 FC_CD   TCCTGGAAGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCG
 FC_IR   TCCTGGAAGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCG
BJS_CD   TCTTGGAAGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCG
BJS_IR   TCTTGGAAGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCG
199_CD   TCCTGGAAGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCG
199_IR   TCCTGGAAGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCG
216_CD   TCCTGGAAGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCG
216_IR   TCCTGGAAGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCG

 FC_CD   AAGAGCTGCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGC
 FC_IR   AAGAGCTGCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGC
BJS_CD   AGGAGCTGCTTATTTTCATGGAGCTGTGCTCTGAGGGCACCCTTGAGTCACTGGTGGAGC
BJS_IR   AGGAGCTGCTTATTTTCATGGAGCTGTGCTCTGAGGGCACCCTTGAGTCACTGGTGGAGC
199_CD   AAGAGCTGCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGC
199_IR   AAGAGCTGCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGC
216_CD   AAGAGCTGCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGC
216_IR   AAGAGCTGCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGC

 FC_CD   TGACTGGTAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCG
 FC_IR   TGACTGGTAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCG
BJS_CD   TGACTGGTAATCTGCCGGAGGCGCTTACGCGACGTTTCACCGCCCAGCTGTTGTCGGGCG
BJS_IR   TGACTGGTAATCTGCCGGAGGCGCTTACGCGACGTTTCACCGCCCAGCTGTTGTCGGGCG
199_CD   TGACTGGTAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCG
199_IR   TGACTGGTAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCG
216_CD   TGACTGGTAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCG
216_IR   TGACTGGTAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCG

 FC_CD   TGTCAGAGCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCC
 FC_IR   TGTCAGAGCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCC
BJS_CD   TGTCTGAGCTGCACAAGCATGGCATTGTGCATCGCGACATCAAGACTGCTAACATATTCC
BJS_IR   TGTCTGAGCTGCACAAGCATGGCATTGTGCATCGCGACATCAAGACTGCTAACATATTCC
199_CD   TGTCAGAGCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCC
199_IR   TGTCAGAGCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCC
216_CD   TGTCAGAGCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCC
216_IR   TGTCAGAGCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCC

 FC_CD   TCGTGGACGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGG
 FC_IR   TCGTGGACGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGG
BJS_CD   TCGTGGACGGCAGCAACAGCCTAAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAAG
BJS_IR   TCGTGGACGGCAGCAACAGCCTAAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAAG
199_CD   TCGTGGACGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGG
199_IR   TCGTGGACGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGG
216_CD   TCGTGGACGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGG
216_IR   TCGTGGACGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGG
```

```
 FC_CD   CGCACACCACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGC
 FC_IR   CGCACACCACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGC
BJS_CD   CGCACACCACTGTGCCCGGCGAGCTGCAGGGCTATGTAGGCACGCAGGCCTATATGGCGC
BJS_IR   CGCACACCACTGTGCCCGGCGAGCTGCAGGGCTATGTAGGCACGCAGGCCTATATGGCGC
199_CD   CGCACACCACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGC
199_IR   CGCACACCACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGC
216_CD   CGCACACCACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGC
216_IR   CGCACACCACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGC

 FC_CD   CGGAGGTGTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGG
 FC_IR   CGGAGGTGTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGG
BJS_CD   CTGAGGTTTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGG
BJS_IR   CTGAGGTTTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGG
199_CD   CGGAGGTGTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGG
199_IR   CGGAGGTGTTCACAAAGACCAACAGCGATGGCCATGGNAGGGCGGCCGATATCTGGTCGG
216_CD   CGGAGGTGTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGG
216_IR   CGGAGGTGTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGG

 FC_CD   TGGGCTGTGTGGTTGTAGAGATGGCCTCGGGCAAGGTGAGTGCTTCACATTGTAGATAAG
 FC_IR   TGGGCTGTGTGGTTGTAGAGATGGCCTCGGGCAAGGTGAGTGCTTCACATTGTAGATAAG
BJS_CD   TGGGCTGTGTGGTTGTCGAGATGGCCTCGGGCAAGGTGAGTGCTTCGCATTGTAGATAAG
BJS_IR   TGGGCTGTGTGGTGTCCGAGATGGCCTCGGGCNAGGTGAGTGCTTCGCATTGTANATAAG
199_CD   TGGGCTGTGTGGTTGTAGAGANGGCCTCGGGCAAGNTGAGTGCTTCNCATTGTAGANAAG
199_IR   TGGGCTGTGTGGNTGTAGANANGGCCTCGGGCAAGGNGAGTGCTTCNCATTGTAGANAAG
216_CD   TGGGCTGTGTGGTTGTAGAGATGGCCTCGGGCAAGGTGAGTGCTTCACATTGTAGATAAG
216_IR   TGGGCTGTGTGGTTGTAGAGATGGCCTCGGGCAAGGTGAGTGCTTCACATTGTAGATAAG

 FC_CD   CAATTCGATGTATGGGGCGGGGTATACAGTCCATTTCACAATGTTGGATTATTAGTGCTT
 FC_IR   CAATTCGATGTATGGGGCGGGGTATACAGTCCATTTCNCAATGTTGGATTATTAGTGCTT
BJS_CD   CAATTCGATGTATGGGGCGGGGTATACAGTCCATTTCACAATGTTGGATTATTAGTGCTT
BJS_IR   CAATTCGATGTATNGGGNNNNNNNNNNNNNNNNNNNNNCAATGTTGGATTATTAGTGCTT
199_CD   CAATGCNATGTGTGGGGCGGGNNATACAGTCCATTTCACAATGTTGGATTATTANTGCTT
199_IR   CAATGCNATGTGTGGGGCGGGNNATACAGTCCATTTCACAATGTTGGATTATTANTGCTT
216_CD   CAATGCGATGTGTGGGGCGGGGTATACAGTCCATTTCACAATGTTGGATTATTAGTGCTT
216_IR   CAATGCGATGTGTGGGGCGGGGTATACAGTCCATTTCACAATGTTGGATTATTAGTGCTT

 FC_CD   TTCTTTGGGTGAATATTTTGCATTATACGCAGTCTAAAGTAATTTACA-CTAATTTTATT
 FC_IR   TTCTTTGGGTGAATATTTTGCATTATACGCAGTCTAAAGTAATTTACA-CTAATTTTATT
BJS_CD   TTCTTTGGGTGAA-----------------------------------------------
BJS_IR   TTCTTTGGGTGAA-----------------------------------------------
199_CD   TTCTTTGGNTGAATATTTTGCATTATACGCAGNNTAANGNAATTTACANCTANNTTNATT
199_IR   TTCTTTGGNTGAATATTTTGCATTATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
216_CD   TTCTTTGGGTGAATATTTTGCATTATACGCAGTCTAAAGTAATTTACA-CTAATTTTATT
216_IR   TTCTTTGGGTGAATATTTTGCATTATACGCAGTCTAAAGTAATTTACA-CTAATTTTATT

 FC_CD   TTTTGAATGAATGCAGTACTCTTTTACTCTTTTCCTTTTAATACTAAATTTGTATAAAAA
 FC_IR   TTTTGAATGAATGCAGTACTCTTTTACTCTTTTCCTTTTAATACTAAATTTGTATAAAAA
BJS_CD   -----------------------------------------------------------A
BJS_IR   -----------------------------------------------------------A
199_CD   TTTTGAATGAATGCAGTACTCTTTTACTCTTTTCCTTTTAATACTAAATTTGTATAAAAA
199_IR   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
216_CD   TTTTGAATGAATGCAGTACTCTTTTACTCTTTTCCTTTTAATACTAAATTTGTATAAAAA
216_IR   TTTTGAATGAATGCAGTACTCTTTTACTCTTTTCCTTTTAATACTAAATTTGTATAAAAA
```

```
 FC_CD   TATTAAAATATTT-ACACTTAATGAAAATACTGTGAATTAGGCTACTAAAACCTTTGTTA
 FC_IR   TATTAAAATATTT-ACACTTAATGAAAATACTGTGAATTAGGCTACTAAAACCTTTGTTA
BJS_CD   TTTTGAAATGTTTAACATTTAA--------------------------------------
BJS_IR   TTTTGAAATGTTTAACATTTAA--------------------------------------
199_CD   TATTAAAATATTT-ACACTTAATGAAAATACTGTGAATTAGGCTACTAAAACCTTTGTTA
199_IR   TATTAAAATATTT-ACANTTAATGAAAATANTGTGAATTAGGCTACTAAAACCTTTGTTA
216_CD   TATTAAAATATTT-ACACTTAATGAAAATACTGTGAATTAGGCTACTAAAACCTTTGTTA
216_IR   TATTAAAATATTT-ACACTTAATGAAAATACTGTGAATTAGGCTACTAAAACCTTTGTTA

 FC_CD   AATATGTGTAATATTATTTAGTTTTGGTTTCTTTACAGCGTCCGTGGGCCCAGTTTGATT
 FC_IR   AATATGTGTAATATTATTTAGTTTTGGTTTCTTTACAGCGTCCGTGGGCCCAGTTTGATT
BJS_CD   --------------TATATATTTATGGTTTCTTTACAGCGTCCTTGGGCCCAGTTTGATT
BJS_IR   --------------TATATATTTATGGTTTCTTTACAGCGTCCTTGGGCCCAGTTTGATT
199_CD   AATATGTGTAATATTATTTAGTTTTGGTTTCTTTACAGCGTCCGTGGGCCCAGTTTGATT
199_IR   AATATGTGTAATATTATTTAGTTTTGGTTTCTTTACANNNNNNGTGGGCCCAGTTTGATT
216_CD   AATATGTGTAATATTATTTAGTTTTGGTTTCTTTACAGCGTCCGTGGGCCCAGTTTGATT
216_IR   AATATGTGTAATATTATTTAGTTTTGGTTTCTTTACAGCGTCCGTGGGCCCAGTTTGATT

 FC_CD   CCAACTTCCAGATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCC
 FC_IR   CCAACTTCCAGATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCC
BJS_CD   CCAACTTCCAGATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCACCGGAGAGCC
BJS_IR   CCAACTTCCAGATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCACCGGAGAGCC
199_CD   CCAACTTCCAGATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCC
199_IR   CCAACTTCCAGATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCC
216_CD   CCAACTTCCAGATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCC
216_IR   CCAACTTCCAGATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCC

 FC_CD   TATCCCAGGAGGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTT
 FC_IR   TATCCCAGGAGGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTT
BJS_CD   TATCCCAGGAGGGTCACGACTTCATCGATCATTGTCTGCAGCATGATCCCAAGAGGCGTT
BJS_IR   TATCCCAGGAGGGTCACGACTTCATCGATCATTGTCTGCAGCATGATCCCAAGAGGCGTT
199_CD   TATCCCAGGAGGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTT
199_IR   TATCCCAGGAGGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTT
216_CD   TATCCCAGGAGGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTT
216_IR   TATCCCAGGAGGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTT

 FC_CD   TAACGGCAGTGGAACTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCC
 FC_IR   TAACGGCAGTGGAACTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCC
BJS_CD   TAACGGCAGTGGAACTGTTGGAGCACAATTTCTGCAAGGTGCGTATGCCTGTGGTTAGCT
BJS_IR   TAACGGCAGTGGAACTGTTGGAGCACAATTTCTGCAAGGTGCGTATGCCTGTGGTTAGCT
199_CD   TAACGGCAGTGGAACTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCC
199_IR   TAACGGCAGTGGAACTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCC
216_CD   TAACGGCAGTGGAACTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCC
216_IR   TAACGGCAGTGGAACTGTTGGAGCACAATTTCTGCAAGGTGCGTAGGCCTGTGGTTAGCC

 FC_CD   CGCTGCCAAATGTTGCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGC
 FC_IR   CGCTGCCAAATGTTGCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGC
BJS_CD   CGCTGCCAAATGTTGCTAATTGATTTTCGTTTACATGTGCAGTACGGTCGAGACGAGTGC
BJS_IR   CGCTGCCAAATGTTGCTAATTGATTTTCGTTTACATGTGCAGTACGGTCGAGACGAGTGC
199_CD   CGCTGCCAAATGTTGCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGC
199_IR   CGCTGCCAAATGTTGCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGC
216_CD   CGCTGCCAAATGTTGCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGC
216_IR   CGCTGCCAAATGTTGCTAATTGATTTTCGTTTCCATGTGCAGTACGGTCGAGACGAGTGC
```

```
 FC_CD   AGCAGCGAGCAGTTGCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGC
 FC_IR   AGCAGCGAGCAGTTGCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGC
BJS_CD   AGCAGTGAGCAGTTGCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGC
BJS_IR   AGCAGTGAGCAGTTGCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGC
199_CD   AGCAGCGAGCAGTTGCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGC
199_IR   AGCAGCGAGCAGTTGCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGC
216_CD   AGCAGCGAGCAGTTGCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGC
216_IR   AGCAGCGAGCAGTTGCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGC

 FC_CD   AGCAGCTAGATGCAACGTGGTGTCTACTTAACTAGCACCCGG-TTTTTTTTAGGCCCACT
 FC_IR   AGCAGCTAGATGCAACGTGGTGTCTACTTAACTAGCACCCGG-TTTTTTTTAGGCCCACT
BJS_CD   AGCAGCTAGGTGCTACGTGGTGTCTACTTAACTAGCTTGCGGTTTTTTTTTAGGCGCACC
BJS_IR   AGCAGCTAGGTGCTACGTGGTGTCTACTTAACTAGCTTGCGGTTTTTTTTTAGGCGCACC
199_CD   AGCAGCTAGATGCAACGTGGTGTCTACTTAACTAGCACCCGG-TTTTTTTTAGGCCCACT
199_IR   NGCAGCTAGATGCAACGTGGTGTCTACTTAACTAGCNCCCGG-NNTTTTTTAGGCCCACT
216_CD   AGCAGCTAGATGCAACGTGGTGTCTACTTAACTAGCACCCGG-TTTTTTTTAGGCCCACT
216_IR   AGCAGCTAGATGCAACGTGGTGTCTACTTAACTAGCACCCGG-TTTTTTTTAGGCCCACT

 FC_CD   CCTACATATGATGCATATGTACATCGGTGTGCATTTTGTTTAG
 FC_IR   CCTACATATGATGCATATGTACATCGGTGTNNNNNNNNNNNNN
BJS_CD   CCTACATATGATGCATATGTACATCGGTGTGGATTTTGTTTAG
BJS_IR   CCTACATATGATGCATATGTACATCGGTGTGGATTTTGTTTA-
199_CD   CCTACATATGATGCATATGTACATCGGTGTGCATTTTGTTTAG
199_IR   CCTACATATGATGCATATNTACATCGGTGTGCNNNNNNNNNNN
216_CD   CCTACATATGATGCATATGTACATCGGTGTGCATTTTGTTTAG
216_IR   CCTACATATGATGCATATGTACATCGGTGTGCATTTTGTTTAG
```

Appendix E: Sequence comparison for transcript RB of *Mekk1* between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216).

```
                 1        10        20        30        40        50        60
                 |        |         |         |         |         |         |
  FC - RB   ATGTCAAACAGGCGAAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGT
 BJS - RB   ATGTCAAACAGGCGAAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGT
 199 - RB   ATGTCAAACAGGCGAAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGT
 216 - RB   ATGTCAAACAGGCGAAGAGTGCGAACAATTGATTATCTGGCGCTGCAGCAGAGTTTGCGT

  FC - RB   TTGCAAAAGACACCGGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAG
 BJS - RB   TTGCAAAAGACACCGGCAGCAACAACAAATGCAGAGGAGCAGGT------GGCAAGGGAG
 199 - RB   TTGCAAAAGACACCGGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAG
 216 - RB   TTGCAAAAGACACCGGCAGCAACAACAAATGCAGAGGACCAAGTGGGAAAGGCAGGGGAG

  FC - RB   GAGGAGAATGGCAATGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCC
 BJS - RB   GAGGAGAATGGCAATGGGCATCATAGTACCGTTACAGCTGAAACGCCACCCACACCACCC
 199 - RB   GAGGAGAATGGCAATGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCC
 216 - RB   GAGGAGAATGGCAATGGGCATCATAGTGCCGTTACAGCTGAAACGCCACCCCCACCACCC

  FC - RB   ATCCCGCCCATTCCGCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCA
 BJS - RB   ATCCCGCCCATTCCGCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCA
 199 - RB   ATCCCGCCCATTCCGCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCA
 216 - RB   ATCCCGCCCATTCCGCCCATACGATTACGACGCGAGCAGAGCGTCGAGGAGGATGTTGCA

  FC - RB   AGAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCCAGTGCAGATGGCTCGAAATCGG
 BJS - RB   AGAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCCAGTGCAGATGACTCGAAATCGG
 199 - RB   AGAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCCAGTGCAGATGGCTCGAAATCGG
 216 - RB   AGAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCCAGTGCAGATGGCTCGAAATCGG

  FC - RB   ATTGGTGCTCTGGAGGAGGACATGCCGCCCGAGGATGAGTTGGCGGCGCACTATGAGGCC
 BJS - RB   ATTGGTGCTCTGGAGGAGGACATGCCTCCCGAAGATGAGCTGGCGGCGCACTATGAGGCC
 199 - RB   ATTGGTGCTCTGGAGGAGGACATGCCCCCCGAGGATGAGCTGGCGGCGCACTATGAGGCC
 216 - RB   ATTGGTGCTCTGGAGGAGGACATGCCCCCCGAGGATGAGCTGGCGGCGCACTATGAGGCC

  FC - RB   TTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAAGAACCGCGACTGGGAACGCAAA
 BJS - RB   TTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAAGAACCGCGACTGGGAACGCAAA
 199 - RB   TTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAAGAACCGCGACTGGGAACGCAAA
 216 - RB   TTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAAGAACCGCGACTGGGAACGCAAA

  FC - RB   CAGAAGGTGGTCAATGTGACAGCTTCCGCAGACTTGCCGGCCAGTTTGGGTGGCACACCC
 BJS - RB   CAGAAGGTGGTCAATGTGACAGCTTCCGCAGACTTACCGGCCAGTGTGGGTGGCACCCCC
 199 - RB   CAGAAGGTGGTCAATGTGACAGCTTCCGCAGACTTGCCGGCCAGTTTGGGTGGCACACCC
 216 - RB   CAGAAGGTGGTCAATGTGACAGCTTCCGCAGACTTGCCGGCCAGTTTGGGTGGCACACCC

  FC - RB   AAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATGGACTGCGCCCTG
 BJS - RB   AAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATGGGCTGCTCCCTG
 199 - RB   AAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATGGACTGCGCCCTG
 216 - RB   AAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATGGACTGCGCCCTG

  FC - RB   CTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGGCTGCAGTCACTC
 BJS - RB   CTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGGCTGCAGTCACTC
 199 - RB   CTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGGCTGCAGTCACTC
 216 - RB   CTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGGCTGCAGTCACTC
```

```
 FC - RB   CTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCGGTGGCTGCTCCA
BJS - RB   CTTCGCGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCGGTGGCTGCTCCA
199 - RB   CTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCGGTGGCTGCTCCA
216 - RB   CTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCGGTGGCTGCTCCA

 FC - RB   CGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGACCAGATCATGCCG
BJS - RB   CGAGGTAACAGCTTCCACGAGACAGCTCATCCGCTGGAGTCTCTGGACCAGATCATGCCG
199 - RB   CGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGACCAGATCATGCCG
216 - RB   CGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGACCAGATCATGCCG

 FC - RB   CTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAGGTGGTGGAGAGC
BJS - RB   CTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAGGTGGTGGAGAGC
199 - RB   CTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAGGTGGTGGAGAGC
216 - RB   CTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAGGTGGTGGAGAGC

 FC - RB   TGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCCTCGGCGCCGCCG
BJS - RB   TGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCCTCGGCGCCGCCG
199 - RB   TGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCCTCGGCGCCGCCG
216 - RB   TGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCCTCGGCGCCGCCG

 FC - RB   TTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGAAGAGCAACGCCT
BJS - RB   TTGGCTTTCGCTGCCCAACTTCCCGCTGGAATGCTAAGGAGTGATGGAAGAGCAACACCT
199 - RB   TTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGAAGAGCAACGCCT
216 - RB   TTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGAAGAGCAACGCCT

 FC - RB   GGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATTAAGCTCGGTAGC
BJS - RB   GGGTTGGGGAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATTAAGCTCGGTAGC
199 - RB   GGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATTAAGCTCGGTAGC
216 - RB   GGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATTAAGCTCGGTAGC

 FC - RB   GTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAGACTGAGCTGAAG
BJS - RB   GTAGATCGACAGGACGCCAAACTATCACAGGAGGAGCACACCTGGCAGACTGAGCTGAAG
199 - RB   GTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAGACTGAGCTGAAG
216 - RB   GTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAGACTGAGCTGAAG

 FC - RB   GATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTGGAGCAGCAGGAC
BJS - RB   GATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACTGTGGAGCAGCAGGAC
199 - RB   GATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTGGAGCAGCAGGAC
216 - RB   GATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTGGAGCAGCAGGAC

 FC - RB   AAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCACATTATAAACTAC
BJS - RB   AAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCACATTATAAACTAC
199 - RB   AAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCACATTATAAACTAC
216 - RB   AAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCACATTATAAACTAC

 FC - RB   AAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGATAGTGGCATTCAT
BJS - RB   AAGTTCCAGCCGCGTTATCGCCGCGAGCCGAGTCTGATAAGCTTGGATAGTGGCACTCAT
199 - RB   AAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGATAGTGGCATTCAT
216 - RB   AAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGATAGTGGCATTCAT

 FC - RB   TCCGATAGCAATTCTAATGCTAGCTCTCCACTGCCCAGCAAGATGTGCCAAGGCTGCATG
BJS - RB   TCCGATAGCAATTCTAATGCCAGCTCTCCACTGCCCAGCAAGATGTGCCAAGGATGCATG
199 - RB   TCCGATAGCAATTCTAATGCTAGCTCTCCACTGCCCAGCAAGATGTGCCAAGGCTGCATG
216 - RB   TCCGATAGCAATTCTAATGCTAGCTCTCCACTGCCCAGCAAGATGTGCCAAGGCTGCATG
```

```
 FC - RB   TCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGAGAGGTAGAGGGT
BJS - RB   TCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGAGAAGTGGAGGGT
199 - RB   TCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGAGAGGTAGAGGGT
216 - RB   TCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGAGAGGTAGAGGGT

 FC - RB   ATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAGGCCATGGGCGCT
BJS - RB   ATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCACAGGCCATGGGCGCC
199 - RB   ATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAGGCCATGGGCGCT
216 - RB   ATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAGGCCATGGGCGCT

 FC - RB   CTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCCATGTGCTTATGG
BJS - RB   CTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCCATGTGCTTATGG
199 - RB   CTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCCATGTGCTTATGG
216 - RB   CTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCCATGTGCTTATGG

 FC - RB   TATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAAATTCTGGCAAGA
BJS - RB   TATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTTTAGGAAAAATTCTGGCAAGA
199 - RB   TATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAAATTCTGGCAAGA
216 - RB   TATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAAATTCTGGCAAGA

 FC - RB   CTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATCGCCTCCGACAGC
BJS - RB   CTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATCGCCACCGACAGC
199 - RB   CTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATCGCCTCCGACAGC
216 - RB   CTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATCGCCTCCGACAGC

 FC - RB   GGTAGTTCATCAGCATCGGGTGTAGAGAACGATGATTCTGCTGTTAACTCAATGGACTCA
BJS - RB   GGTAGTTCATCAGCATCGGGTGTAGAGAACGATGATTCGGCTGTTAACTCAATGGATTCA
199 - RB   GGTAGTTCATCAGCATCGGGTGTAGAGAACGATGATTCTGCTGTTAACTCAATGGACTCA
216 - RB   GGTAGTTCATCAGCATCGGGTGTAGAGAACGATGATTCTGCTGTTAACTCAATGGACTCA

 FC - RB   TCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGACACCGTGCCACAAA
BJS - RB   TCAAAGCCGCCCAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGACACCGTGCCACAAA
199 - RB   TCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGACACCGTGCCACAAA
216 - RB   TCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGACACCGTGCCACAAA

 FC - RB   GTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATCAAGTAGCAATGAA
BJS - RB   GTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATCAAGTAGCAATGAA
199 - RB   GTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATCAAGTAGCAATGAA
216 - RB   GTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATCAAGTAGCAATGAA

 FC - RB   TCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAGCCATATGCGCAAG
BJS - RB   TCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAGCCATATGCGCAAG
199 - RB   TCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAGCCATATGCGCAAG
216 - RB   TCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAGCCATATGCGCAAG

 FC - RB   GGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTACGTGCAGCAGCAAT
BJS - RB   GGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTACGTGCAGCAGCAAT
199 - RB   GGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTACGTGCAGCAGCAAT
216 - RB   GGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTACGTGCAGCAGCAAT

 FC - RB   GGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAATGTGCTCAAAAGT
BJS - RB   GGGACAGGAGAGCAATCCACTGGGCTGTATCGCAAGTTCATTGAAAATGTGCTCAAAAGT
199 - RB   GGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAATGTGCTCAAAAGT
216 - RB   GGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAATGTGCTCAAAAGT
```

```
 FC - RB   CGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGTGGCATTGTACAAG
BJS - RB   CGAGGACTGGCCAAGTCGCTGGCCTTTCTACACAAGCTGCACAACGTGGCATTGTACAAG
199 - RB   CGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGTGGCATTGTACAAG
216 - RB   CGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGTGGCATTGTACAAG

 FC - RB   GCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGAGTCGGATGCGGAG
BJS - RB   GCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTCGATTACGAGTCGGATGCAGAG
199 - RB   GCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGAGTCGGATGCGGAG
216 - RB   GCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGAGTCGGATGCGGAG

 FC - RB   TCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGAACAGGTTGTGGAG
BJS - RB   TCCATTGAAGAAGACGTACCAAGGCTAGATCCTGAAATCAGCCGCGAACAGGTTGTGGAG
199 - RB   TCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGAACAGGTTGTGGAG
216 - RB   TCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGAACAGGTTGTGGAG

 FC - RB   CTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCTGCCGTCGTACATT
BJS - RB   CTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAACCTGCCGTCGTACATT
199 - RB   CTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCTGCCGTCGTACATT
216 - RB   CTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCTGCCGTCGTACATT

 FC - RB   CCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGAGTTCTTGCGCATG
BJS - RB   CCCACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGAGTTCTTGCGCATG
199 - RB   CCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGAGTTCTTGCGCATG
216 - RB   CCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGAGTTCTTGCGCATG

 FC - RB   CGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACAGTTGATGAAGGAG
BJS - RB   CGTCTCGAAACCAGACCCGTTCGACCCAATCCACTGAGCCTGGAACAGTTGATGAAGGAG
199 - RB   CGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACAGTTGATGAAGGAG
216 - RB   CGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACAGTTGATGAAGGAG

 FC - RB   CTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCAGCGGCACATAACC
BJS - RB   CTTAGGGAGGGCTTGACCCTAGCCTTGACCCATCGGGAGCGATACCAGCGGCACATAACC
199 - RB   CTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCAGCGGCACATAACC
216 - RB   CTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCAGCGGCACATAACC

 FC - RB   ACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCATTTTAAACCACTAC
BJS - RB   ACGGCGCTGGTGGAGAATGAGGCGGAACTGGGCAGTTATATTAGCATTTTAAACCACTAC
199 - RB   ACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCATTTTAAACCACTAC
216 - RB   ACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCATTTTAAACCACTAC

 FC - RB   GATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGATCAACTAGTGCTG
BJS - RB   GATGCTACGGTGCGAAAGACTTTCGAGCTGTATTTGGAGTACATCGATCAACTAGTGCTG
199 - RB   GATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGATCAACTAGTGCTG
216 - RB   GATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGATCAACTAGTGCTG

 FC - RB   GTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATGGATGTTTACCAAG
BJS - RB   GTTGCCGTTCCGGAAGGCAACCAGAAGTCCGTCCTTGAGAAGGAATGGATGTTTACCAAG
199 - RB   GTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATGGATGTTTACCAAG
216 - RB   GTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATGGATGTTTACCAAG

 FC - RB   CTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAAATTCTGTGGCATA
BJS - RB   CTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAAATTCTGTGGCATA
199 - RB   CTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAAATTCTGTGGCATA
216 - RB   CTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAAATTCTGTGGCATA
```

```
 FC - RB   ATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTACCGTTGAGCTAGAT
BJS - RB   ATCAGCAAACTTTTGCGGAGCATTTCCGAACGCCTGGTGAAGCGCACCGTTGAGCTAGAT
199 - RB   ATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTACCGTTGAGCTAGAT
216 - RB   ATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTACCGTTGAGCTAGAT

 FC - RB   CAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAAATGGCAACTGCTG
BJS - RB   CAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAAATGGCAGCTGCTG
199 - RB   CAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAAATGGCAACTGCTG
216 - RB   CAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAAATGGCAACTGCTG

 FC - RB   ACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCGCTCCATTAAGGTG
BJS - RB   ACCATTTGCAGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCGCTCCATTAAGGTG
199 - RB   ACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCGCTCCATTAAGGTG
216 - RB   ACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCGCTCCATTAAGGTG

 FC - RB   CTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTTTCATCGCGAGCAT
BJS - RB   CTCTTCTTTGCCAAAACCTTTTGTCGCGACGTCGAGACGACGGATTTTCATCGCGAGCAC
199 - RB   CTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTTTCATCGCGAGCAT
216 - RB   CTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTTTCATCGCGAGCAT

 FC - RB   TACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGACGTCAAGGCGGCG
BJS - RB   TATGAACATGACGTGGCCAACCAACAGCATGACTTCATCTGTTCGGACGTAAAGGCGGCG
199 - RB   TACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGACGTCAAGGCGGCG
216 - RB   TACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGACGTCAAGGCGGCG

 FC - RB   TTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGACGGCAATAATTGAA
BJS - RB   TTTAAACTCTTGCAGCAGGACGTTCTGCAAGTGCGCAACAAGTTGACGGCAATAATTGAA
199 - RB   TTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGACGGCAATAATTGAA
216 - RB   TTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGACGGCAATAATTGAA

 FC - RB   GGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTGGATGAGCAAGACAAACAG
BJS - RB   GGAGTACAAAAACGGTGCTGTTTGAGCAACATGCGAGATTTGGATGAGCAAGACAAACAG
199 - RB   GGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTTTATGAGCAAGACAAACAG
216 - RB   GGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTGGATGAGCAAGACAAACAG

 FC - RB   GCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATTTGGATTCGAGTAT
BJS - RB   GCTGTGCTGTCGCGCACTCGAGAGATTTTGCATCAGGGGTACAAATTTGGATTTGAATAT
199 - RB   GCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATTTGGATTCGAGTAT
216 - RB   GCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATTTGGATTCGAGTAT

 FC - RB   CACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAAGGACAGCGGTGCT
BJS - RB   CACAAGGACGTCATTCGGTTGTTCGAGCAGAAGATTATGGACCAAAAGGACAGCGGTGCT
199 - RB   CACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAAGGACAGCGGTGCT
216 - RB   CACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAAGGACAGCGGTGCT

 FC - RB   CACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTGGATGCATTTCGTG
BJS - RB   CACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTGGATGCATTTCGTA
199 - RB   CACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTGGATGCATTTCGTG
216 - RB   CACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTGGATGCATTTCGTG

 FC - RB   ATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTCCCAGGGTCTGGAG
BJS - RB   ATGGAGCGTTGCGAACGTGGACGAGGAATGCGTCCGCGTTGGGCCTCCCAGGGTCTGGAG
199 - RB   ATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTCCCAGGGTCTGGAG
216 - RB   ATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTCCCAGGGTCTGGAG
```

```
 FC - RB   TTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGACGACGAGTTTGAG
BJS - RB   TTTCTGATTCTTGCCTGTGATCCACAAATCACCCAGCACTTGGATGATGACGAGTTTGAG
199 - RB   TTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGACGACGAGTTTGAG
216 - RB   TTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGACGACGAGTTTGAG

 FC - RB   GCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCATCACTTCGGAACCC
BJS - RB   GCGCTAAAGCAGCAAATGGATCGCTGTATTTCGCACGTGATTGGCATCACCTCGGAACCC
199 - RB   GCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCATCACTTCGGAACCC
216 - RB   GCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCATCACTTCGGAACCC

 FC - RB   GAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTCATCGCCGGCCACC
BJS - RB   GAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGCAAGACTTCATCGCCGGCCACC
199 - RB   GAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTCATCGCCGGCCACC
216 - RB   GAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTCATCGCCGGCCACC

 FC - RB   TCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGTTCTTAATCCGAAT
BJS - RB   TCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGTCCTTAATCCTAAT
199 - RB   TCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGTTCTTAATCCGAAT
216 - RB   TCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGTTCTTAATCCGAAT

 FC - RB   ACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCAGTTCAGTTTAAAG
BJS - RB   ACGCCGCCACTGCAGTCGCCACCGTACAACAAGTTACTGCATCCGCAGTTCAGCTTAAAG
199 - RB   ACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCAGTTCAGTTTAAAG
216 - RB   ACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCAGTTCAGTTTAAAG

 FC - RB   GAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTATGTCGACACTCCG
BJS - RB   GAGGATGTGTCAGGAACTTCGTACAGTCCCGTTGACAGTTCAGACTATGTTGACACTCCG
199 - RB   GAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTATGTCGACACTCCG
216 - RB   GAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTATGTCGACACTCCG

 FC - RB   TGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGACGCCTCCGACTCCA
BJS - RB   TGCCAAAGGAGTGCCAACGGCGAGCTGCGTCTGCTGGTGCCCCAGACGCCGCCGACTCCA
199 - RB   TGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGACGCCTCCGACTCCA
216 - RB   TGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGACGCCTCCGACTCCA

 FC - RB   GCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCGACAGGAACGCGTT
BJS - RB   GCATCTCCCGGAAAGAGCAGCCTCGAAAGTACACCGCTGGCCTTGCGGCAGGAACGTGTT
199 - RB   GCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCGACAGGAACGCGTT
216 - RB   GCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCGACAGGAACGCGTT

 FC - RB   AGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCGCGAACGAAGGTTG
BJS - RB   AGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTACGCGAACGAAGGTTG
199 - RB   AGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCGCGAACGAAGGTTG
216 - RB   AGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCGCGAACGAAGGTTG

 FC - RB   ATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAGGGCGCGTAGCGTT
BJS - RB   ATTGGGCAGGTCAAGTCTCTTAATTCCAGCGACAAGGTGCATATAAGGGCGCGTAGCGTT
199 - RB   ATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAGGGCGCGTAGCGTT
216 - RB   ATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAGGGCGCGTAGCGTT

 FC - RB   CATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGGCAAGGTGTACACG
BJS - RB   CATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGGCAAGGTATACACG
199 - RB   CATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGGCAAGGTGTACACG
216 - RB   CATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGGCAAGGTGTACACG
```

```
 FC - RB   GCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGCAATCCAGCCGGGA
BJS - RB   GCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGCAATCCAGCCGGGA
199 - RB   GCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGCAATCCAGCCGGGA
216 - RB   GCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGCAATCCAGCCGGGA

 FC - RB   GAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGAAGGAATAAAGCAC
BJS - RB   GAGACGCGAGCACTCAAGAATGTGGCCGAAGAGCTAAAGATCTTGGAAGGAATAAAGCAC
199 - RB   GAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGAAGGAATAAAGCAC
216 - RB   GAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGAAGGAATAAAGCAC

 FC - RB   AAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCTGCTCATTTTCATG
BJS - RB   AAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAGGAGCTGCTTATTTTCATG
199 - RB   AAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCTGCTCATTTTCATG
216 - RB   AAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCTGCTCATTTTCATG

 FC - RB   GAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGGTAATCTGCCGGAG
BJS - RB   GAGCTGTGCTCTGAGGGCACCCTTGAGTCACTGGTGGAGCTGACTGGTAATCTGCCGGAG
199 - RB   GAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGGTAATCTGCCGGAG
216 - RB   GAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGGTAATCTGCCGGAG

 FC - RB   GCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGAGCTGCACAAGCAT
BJS - RB   GCGCTTACGCGACGTTTCACCGCCCAGCTGTTGTCGGGCGTGTCTGAGCTGCACAAGCAT
199 - RB   GCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGAGCTGCACAAGCAT
216 - RB   GCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGAGCTGCACAAGCAT

 FC - RB   GGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGACGGCAGCAACAGC
BJS - RB   GGCATTGTGCATCGCGACATCAAGACTGCTAACATATTCCTCGTGGACGGCAGCAACAGC
199 - RB   GGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGACGGCAGCAACAGC
216 - RB   GGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGACGGCAGCAACAGC

 FC - RB   CTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACACCACCGTGCCGGGC
BJS - RB   CTAAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAAGCGCACACCACTGTGCCCGGC
199 - RB   CTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACACCACCGTGCCGGGC
216 - RB   CTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACACCACCGTGCCGGGC

 FC - RB   GAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGTGTTCACAAAGACC
BJS - RB   GAGCTGCAGGGCTATGTAGGCACGCAGGCCTATATGGCGCCTGAGGTTTTCACAAAGACC
199 - RB   GAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGTGTTCACAAAGACC
216 - RB   GAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGTGTTCACAAAGACC

 FC - RB   AACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGGTTGTAGAG
BJS - RB   AACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGGTTGTCGAG
199 - RB   AACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGGTTGTAGAG
216 - RB   AACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTGTGTGGTTGTAGAG

 FC - RB   ATGGCCTCGGGCAAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCAGATCATGTTCAAA
BJS - RB   ATGGCCTCGGGCAAGCGTCCTTGGGCCCAGTTTGATTCCAACTTCCAGATCATGTTCAAA
199 - RB   ATGGCCTCGGGCAAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCAGATCATGTTCAAA
216 - RB   ATGGCCTCGGGCAAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCAGATCATGTTCAAA

 FC - RB   GTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGAGGGTCACGACTTC
BJS - RB   GTGGGCATGGGCGAGAAGCCGCAGGCACCGGAGAGCCTATCCCAGGAGGGTCACGACTTC
199 - RB   GTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGAGGGTCACGACTTC
216 - RB   GTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGAGGGTCACGACTTC
```

```
 FC - RB   ATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGTGGAACTGTTGGAG
BJS - RB   ATCGATCATTGTCTGCAGCATGATCCCAAGAGGCGTTTAACGGCAGTGGAACTGTTGGAG
199 - RB   ATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGTGGAACTGTTGGAG
216 - RB   ATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGTGGAACTGTTGGAG

 FC - RB   CACAATTTCTGCAAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTTGCAGATGCAGGTG
BJS - RB   CACAATTTCTGCAAGTACGGTCGAGACGAGTGCAGCAGTGAGCAGTTGCAGATGCAGGTG
199 - RB   CACAATTTCTGCAAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTTGCAGATGCAGGTG
216 - RB   CACAATTTCTGCAAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTTGCAGATGCAGGTG

 FC - RB   CGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAG
BJS - RB   CGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAG
199 - RB   CGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAG
216 - RB   CGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAG
```

Appendix F: Sequence comparison for transcript RD/RC of *Mekk1* between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216).

```
           1        10        20        30        40        50        60
           |        |         |         |         |         |         |
 FC - RC   ATGCGTCGGAAGAAAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCCAGTGCAGATG
BJS - RD   ATGCGTCGGAAGAAAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCCAGTGCAGATG
199 - RC   ATGCGTCGGAAGAAAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCCAGTGCAGATG
216 - RC   ATGCGTCGGAAGAAAGTTGAGTACCGTGTGAAGCAGACGCCTTCGCGTCCAGTGCAGATG

 FC - RC   ACTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCGCCCGAGGATGAGTTGGCGGCG
BJS - RD   ACTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCTCCCGAAGATGAGCTGGCGGCG
199 - RC   GCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCCCCCGAGGATGAGCTGGCGGCG
216 - RC   GCTCGAAATCGGATTGGTGCTCTGGAGGAGGACATGCCCCCCGAGGATGAGCTGGCGGCG

 FC - RC   CACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAAGAACCGCGAC
BJS - RD   CACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAAGAACCGCGAC
199 - RC   CACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAAGAACCGCGAC
216 - RC   CACTATGAGGCCTTCGGCACTACTCCGCCACGTACCAGGCTCAAAATCAAGAACCGCGAC

 FC - RC   TGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCTTCCGCAGACTTGCCGGCCAGTTTG
BJS - RD   TGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCTTCCGCAGACTTACCGGCCAGTGTG
199 - RC   TGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCTTCCGCAGACTTGCCGGCCAGTTTG
216 - RC   TGGGAACGCAAACAGAAGGTGGTCAATGTGACAGCTTCCGCAGACTTGCCGGCCAGTTTG

 FC - RC   GGTGGCACACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATG
BJS - RD   GGTGGCACCCCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATG
199 - RC   GGTGGCACACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATG
216 - RC   GGTGGCACACCCAAAAAGACCAGGACGGCCAGATCCCGCGTCCTGCGGCGCAACACTATG

 FC - RC   GACTGCGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGG
BJS - RD   GGCT-CTCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGG
199 - RC   GACTGCGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGG
216 - RC   GACTGCGCCCTGCTCAACGAGATGTTTGTCAACGATGAGAGCAAGCGGACGGACAAGCGG

 FC - RC   CTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCG
BJS - RD   CTGCAGTCACTCCTTCGCGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCG
199 - RC   CTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCG
216 - RC   CTGCAGTCACTCCTTCGTGACTCGGAACGTGAGATGAAGAACTCGTTGGCCACAACGGCG

 FC - RC   GTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGAC
BJS - RD   GTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCGCTGGAGTCTCTGGAC
199 - RC   GTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGAC
216 - RC   GTGGCTGCTCCACGAGGTAACAGCTTCCACGAGACAGCTCATCCACTGGAGTCTCTGGAC

 FC - RC   CAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAG
BJS - RD   CAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAG
199 - RC   CAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAG
216 - RC   CAGATCATGCCGCTCAATTCGGAGGCAATGGCGCCGATTCCACTGCGAATTGCTTCCAAG

 FC - RC   GTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCC
BJS - RD   GTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCC
199 - RC   GTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCC
216 - RC   GTGGTGGAGAGCTGCAATCGCTATCGCTGCGTGTCCTCACGTCCCATTGGGTATCGCTCC
```

```
 FC - RC   TCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGA
BJS - RD   TCGGCGCCGCCGTTGGCTTTCGCTGCCCAACTTCCCGCTGGAATGCTAAGGAGTGATGGA
199 - RC   TCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGA
216 - RC   TCGGCGCCGCCGTTGGCTTTCGCTGCTCAACTTCCCGCTGGAATGTTAAGGAGTGATGGA

 FC - RC   AGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATT
BJS - RD   AGAGCAACACCTGGGTTGGGGAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATT
199 - RC   AGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATT
216 - RC   AGAGCAACGCCTGGGTTGGGAAAAAGAAAGGACTTCCACGAAACGTTTGCCAATCTAATT

 FC - RC   AAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAG
BJS - RD   AAGCTCGGTAGCGTAGATCGACAGGACGCCAAACTATCACAGGAGGAGCACACCTGGCAG
199 - RC   AAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAG
216 - RC   AAGCTCGGTAGCGTAGACCGACAGGATGCCAAACTATCACAAGAGGAGCACACCTGGCAG

 FC - RC   ACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTG
BJS - RD   ACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACTGTG
199 - RC   ACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTG
216 - RC   ACTGAGCTGAAGGATCTCATCTGGCTAGAGCTACAGGCCTGGCAAGCAGACCGAACGGTG

 FC - RC   GAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCAC
BJS - RD   GAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCAC
199 - RC   GAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCAC
216 - RC   GAGCAGCAGGACAAGTACCTTTTCGAAGCGCGTCAGGGAGTCTCCGATCTGCTCACCCAC

 FC - RC   ATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGAT
BJS - RD   ATTATAAACTACAAGTTCCAGCCGCGTTATCGCCGCGAGCCGAGTCTGATAAGCTTGGAT
199 - RC   ATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGAT
216 - RC   ATTATAAACTACAAGTTCCAGCCGCGCTATCGCCGCGAGCCGAGTCTGATCAGCCTGGAT

 FC - RC   AGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTCTCCACTGCCCAGCAAGATGTGC
BJS - RD   AGTGGCACTCATTCCGATAGCAATTCTAATGCCAGCTCTCCACTGCCCAGCAAGATGTGC
199 - RC   AGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTCTCCACTGCCCAGCAAGATGTGC
216 - RC   AGTGGCATTCATTCCGATAGCAATTCTAATGCTAGCTCTCCACTGCCCAGCAAGATGTGC

 FC - RC   CAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGA
BJS - RD   CAAGGATGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGA
199 - RC   CAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGA
216 - RC   CAAGGCTGCATGTCGCTGTACTGCAAGGATTGCATGGATCATCAGGAGTTGGCGCTGCGA

 FC - RC   GAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAG
BJS - RD   GAAGTGGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCACAG
199 - RC   GAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAG
216 - RC   GAGGTAGAGGGTATGCTGACGCGCCTGGAAGCTGCAGAGGCACTCTACCCTTCCTCGCAG

 FC - RC   GCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCC
BJS - RD   GCCATGGGCGCCCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCC
199 - RC   GCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCC
216 - RC   GCCATGGGCGCTCTGCATCCCATTTACAAATCACAGAGCTTTGTAGGGCGCATCAAGTCC

 FC - RC   ATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAA
BJS - RD   ATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTTTAGGAAAA
199 - RC   ATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAA
216 - RC   ATGTGCTTATGGTATAACATTACCAAGCAGAACAAGTTGAAGCTCAGTATTCTAGGAAAA
```

```
 FC - RC   ATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATC
BJS - RD   ATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATC
199 - RC   ATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATC
216 - RC   ATTCTGGCAAGACTGCAGGATGAGAAGTTCTCTTGGCCCGTGTGCACTTCCTCCTACATC

 FC - RC   GCCTCCGACAGCGGTAGTTCATCAGCATC-GGGTGTAGAGAACGATGATTCTGCTGTTAA
BJS - RD   GCCACCGACAGCGGTAGTTCATCAGCATCGGGGTGTAGAGAACGATGATTCGGCTGTTAA
199 - RC   GCCTCCGACAGCGGTAGTTCATCAGCATC-GGGTGTAGAGAACGATGATTCTGCTGTTAA
216 - RC   GCCTCCGACAGCGGTAGTTCATCAGCATC-GGGTGTAGAGAACGATGATTCTGCTGTTAA

 FC - RC   CTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGAC
BJS - RD   CTCAATGGATTCATCAAAGCCGCCCAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGAC
199 - RC   CTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGAC
216 - RC   CTCAATGGACTCATCAAAGCCACCAAGTATGGCCGGATCAGCCTCGCGAAAGGGTGTGAC

 FC - RC   ACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATC
BJS - RD   ACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATC
199 - RC   ACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATC
216 - RC   ACCGTGCCACAAAGTGCAATTCATGTTAAACGACGCCACCCACGTGCCAGGCGAGACATC

 FC - RC   AAGTAGCAATGAATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAG
BJS - RD   AAGTAGCAATGAATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAG
199 - RC   AAGTAGCAATGAATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAG
216 - RC   AAGTAGCAATGAATCCACCTCCACGGAGGTAAGCCAGTGGTCCAGCGAGTGCTCTCACAG

 FC - RC   CCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTAC
BJS - RD   CCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTAC
199 - RC   CCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTAC
216 - RC   CCATATGCGCAAGGGATCCATGCACGACATCAACATTTTCAGTGTCGAGCCATTGGGTAC

 FC - RC   GTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAA
BJS - RD   GTGCAGCAGCAATGGGACAGGAGAGCAATCCACTGGGCTGTATCGCAAGTTCATTGAAAA
199 - RC   GTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAA
216 - RC   GTGCAGCAGCAATGGGACAGGAGAGCAGTCCACTGGGCTGTATCGCAAGTTCATAGAAAA

 FC - RC   TGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGT
BJS - RD   TGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCCTTTCTACACAAGCTGCACAACGT
199 - RC   TGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGT
216 - RC   TGTGCTCAAAAGTCGAGGACTGGCCAAGTCGCTGGCTTTTCTACACAAGCTGCACAACGT

 FC - RC   GGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGA
BJS - RD   GGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTCGATTACGA
199 - RC   GGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGA
216 - RC   GGCATTGTACAAGGCGCACATAGCTCTGGAAAAACCCGGAGCTGAAGACTTAGATTACGA

 FC - RC   GTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGA
BJS - RD   GTCGGATGCAGAGTCCATTGAAGAAGACGTACCAAGGCTAGATCCTGAAATCAGCCGCGA
199 - RC   GTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGA
216 - RC   GTCGGATGCGGAGTCCATTGAGGAAGACGTACCAAGGCTAGATCCTCAAATCAGCCGCGA

 FC - RC   ACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCT
BJS - RD   ACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAACCT
199 - RC   ACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCT
216 - RC   ACAGGTTGTGGAGCTGCGCACCTACGGCTACTGGAGCGAGGAGGCACAATCCATAAATCT
```

```
 FC - RC   GCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGA
BJS - RD   GCCGTCGTACATTCCCACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGA
199 - RC   GCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGA
216 - RC   GCCGTCGTACATTCCTACCTTTGTATTTCTGAGCGGAATTCCGTTGCAGTTCATGCATGA

 FC - RC   GTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACA
BJS - RD   GTTCTTGCGCATGCGTCTCGAAACCAGACCCGTTCGACCCAATCCACTGAGCCTGGAACA
199 - RC   GTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACA
216 - RC   GTTCTTGCGCATGCGTCTCGAGACGAGACCCGTTCGCCCCAATCCACTGAGCCTGGAACA

 FC - RC   GTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCA
BJS - RD   GTTGATGAAGGAGCTTAGGGAGGGCTTGACCCTAGCCTTGACCCATCGGGAGCGATACCA
199 - RC   GTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCA
216 - RC   GTTGATGAAGGAGCTCAGGGAGGGCTTGACCCTAGCCTTAACCCATCGGGAGCGATACCA

 FC - RC   GCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCAT
BJS - RD   GCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTGGGCAGTTATATTAGCAT
199 - RC   GCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCAT
216 - RC   GCGGCACATAACCACGGCGCTGGTGGAGAATGAGGCGGAACTAGGCAGTTATATTAGCAT

 FC - RC   TTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGA
BJS - RD   TTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTATTTGGAGTACATCGA
199 - RC   TTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGA
216 - RC   TTTAAACCACTACGATGCTACGGTGCGAAAGACTTTCGAGCTGTACTTGGAGTACATCGA

 FC - RC   TCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATG
BJS - RD   TCAACTAGTGCTGGTTGCCGTTCCGGAAGGCAACCAGAAGTCCGTCCTTGAGAAGGAATG
199 - RC   TCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATG
216 - RC   TCAACTAGTGCTGGTTGCCGTTCCGGAGGGCAACCAGAAGTCCGTCCTTGAGAAGGAATG

 FC - RC   GATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAA
BJS - RD   GATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAA
199 - RC   GATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAA
216 - RC   GATGTTTACCAAGCTTATCAGTCCCATGATAAAGGGCATGCACACGTTAGCATCTCAAAA

 FC - RC   ATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTAC
BJS - RD   ATTCTGTGGCATAATCAGCAAACTTTTGCGGAGCATTTCCGAACGCCTGGTGAAGCGCAC
199 - RC   ATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTAC
216 - RC   ATTCTGTGGCATAATCAGCAAGCTTTTGCGGAGCATCTCCGAACGCCTGGTGAAGCGTAC

 FC - RC   CGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAA
BJS - RD   CGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAA
199 - RC   CGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAA
216 - RC   CGTTGAGCTAGATCAACAGATCGATGGCACCGCCGATACCGACGACAACGAGGAGGTTAA

 FC - RC   ATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCG
BJS - RD   ATGGCAGCTGCTGACCATTTGCAGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCG
199 - RC   ATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCG
216 - RC   ATGGCAACTGCTGACCATTTGCCGAGAGACACAATCGCTGCTCACAGTCGAACGGGAGCG

 FC - RC   CTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTT
BJS - RD   CTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTCGAGACGACGGATTT
199 - RC   CTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTT
216 - RC   CTCCATTAAGGTGCTCTTCTTTGCCAAAACCTTTTGTCGCGACGTAGAGACGACGGATTT
```

```
 FC - RC   TCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGA
BJS - RD   TCATCGCGAGCACTATGAACATGACGTGGCCAACCAACAGCATGACTTCATCTGTTCGGA
199 - RC   TCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGA
216 - RC   TCATCGCGAGCATTACGAACATGACGTAGCCAACCAGCAGCATGACTTCATCTGTTCGGA

 FC - RC   CGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGAC
BJS - RD   CGTAAAGGCGGCGTTTAAACTCTTGCAGCAGGACGTTCTGCAAGTGCGCAACAAGTTGAC
199 - RC   CGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGAC
216 - RC   CGTCAAGGCGGCGTTCAAGCTACTGCAGCAGGACGTTTTGCAAGTGCGCAACAAGTTGAC

 FC - RC   GGCAATAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTGGATGA
BJS - RD   GGCAATAATTGAAGGAGTACAAAAACGGTGCTGTTTGAGCAACATGCGAGATTTGGATGA
199 - RC   GGCAATAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTTTATGA
216 - RC   GGCAATAATTGAAGGAGTGCAAAAACGTTGCTGTTTGAGCAACATGCGAGATTTGGATGA

 FC - RC   GCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATT
BJS - RD   GCAAGACAAACAGGCTGTGCTGTCGCGCACTCGAGAGATTTTGCATCAGGGGTACAAATT
199 - RC   GCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATT
216 - RC   GCAAGACAAACAGGCTGTGCTGTCACGCACTCGAGAGATTTTGCATCAGGGGTACAAATT

 FC - RC   TGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAA
BJS - RD   TGGATTTGAATATCACAAGGACGTCATTCGGTTGTTCGAGCAGAAGATTATGGACCAAAA
199 - RC   TGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAA
216 - RC   TGGATTCGAGTATCACAAGGACGTCATTCGGCTGTTCGAGCAGAAGATTATGGACCAAAA

 FC - RC   GGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTG
BJS - RD   GGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTG
199 - RC   GGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTG
216 - RC   GGACAGCGGTGCTCACACAGTAGATCTTGCCTTAGGAATTATCGCTTACGCTAAGATGTG

 FC - RC   GATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTC
BJS - RD   GATGCATTTCGTAATGGAGCGTTGCGAACGTGGACGAGGAATGCGTCCGCGTTGGGCCTC
199 - RC   GATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTC
216 - RC   GATGCATTTCGTGATGGAGCGTTGCGAACGTGGGCGAGGAATGCGTCCGCGTTGGGCTTC

 FC - RC   CCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGA
BJS - RD   CCAGGGTCTGGAGTTTCTGATTCTTGCCTGTGATCCACAAATCACCCAGCACTTGGATGA
199 - RC   CCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGA
216 - RC   CCAGGGTCTGGAGTTTTTGATTCTTGCCTGTGATCCACAAATTACCCAGCACTTGGACGA

 FC - RC   CGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCAT
BJS - RD   TGACGAGTTTGAGGCGCTAAAGCAGCAAATGGATCGCTGTATTTCGCACGTGATTGGCAT
199 - RC   CGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCAT
216 - RC   CGACGAGTTTGAGGCGCTAAAGCAGCAAATGGACCGCTGTATTTCGCACGTGATTGGCAT

 FC - RC   CACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTC
BJS - RD   CACCTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGCAAGACTTC
199 - RC   CACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTC
216 - RC   CACTTCGGAACCCGAAAAGGTCGCCAAGAAAAAGGCTTCGCCGCGCACTCGAAAGACTTC

 FC - RC   ATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGT
BJS - RD   ATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGT
199 - RC   ATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGT
216 - RC   ATCGCCGGCCACCTCGCGTTCCCGGACACCAACACGGACTCCAATGTCCGCTGGCATGGT
```

```
 FC - RC   TCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCA
BJS - RD   CCTTAATCCTAATACGCCGCCACTGCAGTCGCCACCGTACAACAAGTTACTGCATCCGCA
199 - RC   TCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCA
216 - RC   TCTTAATCCGAATACGCCGCCACTGCAGTCGCCACCATACAACAAGTTACTGCATCCGCA

 FC - RC   GTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTA
BJS - RD   GTTCAGCTTAAAGGAGGATGTGTCAGGAACTTCGTACAGTCCCGTTGACAGTTCAGACTA
199 - RC   GTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTA
216 - RC   GTTCAGTTTAAAGGAGGATGTGTCAGTAAACTCGTACAGTCCCGTTGACAGTTCAGACTA

 FC - RC   TGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGAC
BJS - RD   TGTTGACACTCCGTGCCAAAGGAGTGCCAACGGCGAGCTGCGTCTGCTGGTGCCCCAGAC
199 - RC   TGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGAC
216 - RC   TGTCGACACTCCGTGCCAGAGGAGTGCCAATGGCGAGCTGCGCCTGCTGGTGCCCCAGAC

 FC - RC   GCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCG
BJS - RD   GCCGCCGACTCCAGCATCTCCCGGAAAGAGCAGCCTCGAAAGTACACCGCTGGCCTTGCG
199 - RC   GCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCG
216 - RC   GCCTCCGACTCCAGCATCTCCTGGAAAGAGCAGCCTCGAAGGCACACCGCTGGCCTTGCG

 FC - RC   ACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCG
BJS - RD   GCAGGAACGTGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTACG
199 - RC   ACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCG
216 - RC   ACAGGAACGCGTTAGAGATGCCGTCAACCGTTTGGATATGGATCTAGAGGACGGGCTGCG

 FC - RC   CGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAG
BJS - RD   CGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTTAATTCCAGCGACAAGGTGCATATAAG
199 - RC   CGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAG
216 - RC   CGAACGAAGGTTGATTGGGCAGGTCAAGTCTCTCAATTCTAGCGACAAGGTGCATATAAG

 FC - RC   GGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGG
BJS - RD   GGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGG
199 - RC   GGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGG
216 - RC   GGCGCGTAGCGTTCATTTCCGCTGGCATCGTGGCATTAAGATCGGCCAGGGACGATTCGG

 FC - RC   CAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGC
BJS - RD   CAAGGTATACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGC
199 - RC   CAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGC
216 - RC   CAAGGTGTACACGGCGGTGAACAATAACACGGGAGAGCTGATGGCCATGAAAGAGATCGC

 FC - RC   AATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGA
BJS - RD   AATCCAGCCGGGAGAGACGCGAGCACTCAAGAATGTGGCCGAAGAGCTAAAGATCTTGGA
199 - RC   AATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGA
216 - RC   AATCCAGCCGGGAGAGACGCGAGCGCTCAAGAATGTGGCCGAAGAGCTAAAGATCCTGGA

 FC - RC   AGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCT
BJS - RD   AGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAGGAGCT
199 - RC   AGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCT
216 - RC   AGGAATAAAGCACAAAAACCTGGTGCGCTACTATGGCATCGAGGTGCACCGCGAAGAGCT

 FC - RC   GCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGG
BJS - RD   GCTTATTTTCATGGAGCTGTGCTCTGAGGGCACCCTTGAGTCACTGGTGGAGCTGACTGG
199 - RC   GCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGG
216 - RC   GCTCATTTTCATGGAGCTGTGCTCTGAGGGCACCCTCGAGTCGCTGGTGGAGCTGACTGG
```

```
 FC - RC   TAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGA
BJS - RD   TAATCTGCCGGAGGCGCTTACGCGACGTTTCACCGCCCAGCTGTTGTCGGGCGTGTCTGA
199 - RC   TAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGA
216 - RC   TAATCTGCCGGAGGCGCTTACGAGACGATTCACCGCCCAGCTGTTGTCGGGCGTGTCAGA

 FC - RC   GCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGA
BJS - RD   GCTGCACAAGCATGGCATTGTGCATCGCGACATCAAGACTGCTAACATATTCCTCGTGGA
199 - RC   GCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGA
216 - RC   GCTGCACAAGCATGGCATTGTGCACCGCGACATCAAGACTGCTAACATATTCCTCGTGGA

 FC - RC   CGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACAC
BJS - RD   CGGCAGCAACAGCCTAAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAAGCGCACAC
199 - RC   CGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACAC
216 - RC   CGGCAGCAACAGCCTGAAACTGGGCGATTTCGGATCAGCAGTGAAGATCCAGGCGCACAC

 FC - RC   CACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGT
BJS - RD   CACTGTGCCCGGCGAGCTGCAGGGCTATGTAGGCACGCAGGCCTATATGGCGCCTGAGGT
199 - RC   CACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGT
216 - RC   CACCGTGCCGGGCGAGCTGCAGGGCTATGTCGGCACACAGGCCTATATGGCGCCGGAGGT

 FC - RC   GTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTG
BJS - RD   TTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTG
199 - RC   GTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTG
216 - RC   GTTCACAAAGACCAACAGCGATGGCCATGGCAGGGCGGCCGATATCTGGTCGGTGGGCTG

 FC - RC   TGTGGTTGTAGAGATGGCCTCGGGCAAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCA
BJS - RD   TGTGGTTGTCGAGATGGCCTCGGGCAAGCGTCCTTGGGCCCAGTTTGATTCCAACTTCCA
199 - RC   TGTGGTTGTAGAGATGGCCTCGGGCAAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCA
216 - RC   TGTGGTTGTAGAGATGGCCTCGGGCAAGCGTCCGTGGGCCCAGTTTGATTCCAACTTCCA

 FC - RC   GATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGA
BJS - RD   GATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCACCGGAGAGCCTATCCCAGGA
199 - RC   GATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGA
216 - RC   GATCATGTTCAAAGTGGGCATGGGCGAGAAGCCGCAGGCGCCGGAGAGCCTATCCCAGGA

 FC - RC   GGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGT
BJS - RD   GGGTCACGACTTCATCGATCATTGTCTGCAGCATGATCCCAAGAGGCGTTTAACGGCAGT
199 - RC   GGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGT
216 - RC   GGGTCACGACTTCATCGATCATTGTCTGCAGCACGATCCCAAGAGGCGTTTAACGGCAGT

 FC - RC   GGAACTGTTGGAGCACAATTTCTGCAAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTT
BJS - RD   GGAACTGTTGGAGCACAATTTCTGCAAGTACGGTCGAGACGAGTGCAGCAGTGAGCAGTT
199 - RC   GGAACTGTTGGAGCACAATTTCTGCAAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTT
216 - RC   GGAACTGTTGGAGCACAATTTCTGCAAGTACGGTCGAGACGAGTGCAGCAGCGAGCAGTT

 FC - RC   GCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAG
BJS - RD   GCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAG
199 - RC   GCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAG
216 - RC   GCAGATGCAGGTGCGAGGCTCTTTCCGGCGAAATGTGGCGACCAGCAGCAGCTAG
```

Appendix G: Sequence comparison for primary structure of RB protein of *Mekk1* between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216).

```
Key: X = 5' and 3' nucleotides of kinase domain
     X = variant (not unique to FC)
```

```
       1        10        20        30        40        50        60
       |        |         |         |         |         |         |
  FC   MSNRRRVRTIDYLALQQSLRLQKTPAATTNAEDQVGKAGEEENGNGHHSAVTAETPPPPP
 BJS   MSNRRRVRTIDYLALQQSLRLQKTPAATTNAEEQV--AREEENGNGHHSTVTAETPPTPP
 199   MSNRRRVRTIDYLALQQSLRLQKTPAATTNAEDQVGKAGEEENGNGHHSAVTAETPPPPP
 216   MSNRRRVRTIDYLALQQSLRLQKTPAATTNAEDQVGKAGEEENGNGHHSAVTAETPPPPP

  FC   IPPIPPIRLRREQSVEEDVARVEYRVKQTPSRPVQMARNRIGALEEDMPPEDELAAHYEA
 BJS   IPPIPPIRLRREQSVEEDVARVEYRVKQTPSRPVQMTRNRIGALEEDMPPEDELAAHYEA
 199   IPPIPPIRLRREQSVEEDVARVEYRVKQTPSRPVQMARNRIGALEEDMPPEDELAAHYEA
 216   IPPIPPIRLRREQSVEEDVARVEYRVKQTPSRPVQMARNRIGALEEDMPPEDELAAHYEA

  FC   FGTTPPRTRLKIKNRDWERKQKVVNVTASADLPASLGGTPKKTRTARSRVLRRNTMDCAL
 BJS   FGTTPPRTRLKIKNRDWERKQKVVNVTASADLPASVGGTPKKTRTARSRVLRRNTMGCSL
 199   FGTTPPRTRLKIKNRDWERKQKVVNVTASADLPASLGGTPKKTRTARSRVLRRNTMDCAL
 216   FGTTPPRTRLKIKNRDWERKQKVVNVTASADLPASLGGTPKKTRTARSRVLRRNTMDCAL

  FC   LNEMFVNDESKRTDKRLQSLLRDSEREMKNSLATTAVAAPRGNSFHETAHPLESLDQIMP
 BJS   LNEMFVNDESKRTDKRLQSLLRDSEREMKNSLATTAVAAPRGNSFHETAHPLESLDQIMP
 199   LNEMFVNDESKRTDKRLQSLLRDSEREMKNSLATTAVAAPRGNSFHETAHPLESLDQIMP
 216   LNEMFVNDESKRTDKRLQSLLRDSEREMKNSLATTAVAAPRGNSFHETAHPLESLDQIMP

  FC   LNSEAMAPIPLRIASKVVESCNRYRCVSSRPIGYRSSAPPLAFAAQLPAGMLRSDGRATP
 BJS   LNSEAMAPIPLRIASKVVESCNRYRCVSSRPIGYRSSAPPLAFAAQLPAGMLRSDGRATP
 199   LNSEAMAPIPLRIASKVVESCNRYRCVSSRPIGYRSSAPPLAFAAQLPAGMLRSDGRATP
 216   LNSEAMAPIPLRIASKVVESCNRYRCVSSRPIGYRSSAPPLAFAAQLPAGMLRSDGRATP

  FC   GLGKRKDFHETFANLIKLGSVDRQDAKLSQEEHTWQTELKDLIWLELQAWQADRTVEQQD
 BJS   GLGKRKDFHETFANLIKLGSVDRQDAKLSQEEHTWQTELKDLIWLELQAWQADRTVEQQD
 199   GLGKRKDFHETFANLIKLGSVDRQDAKLSQEEHTWQTELKDLIWLELQAWQADRTVEQQD
 216   GLGKRKDFHETFANLIKLGSVDRQDAKLSQEEHTWQTELKDLIWLELQAWQADRTVEQQD

  FC   KYLFEARQGVSDLLTHIINYKFQPRYRREPSLISLDSGIHSDSNSNASSPLPSKMCQGCM
 BJS   KYLFEARQGVSDLLTHIINYKFQPRYRREPSLISLDSGTHSDSNSNASSPLPSKMCQGCM
 199   KYLFEARQGVSDLLTHIINYKFQPRYRREPSLISLDSGIHSDSNSNASSPLPSKMCQGCM
 216   KYLFEARQGVSDLLTHIINYKFQPRYRREPSLISLDSGIHSDSNSNASSPLPSKMCQGCM

  FC   SLYCKDCMDHQELALREVEGMLTRLEAAEALYPSSQAMGALHPIYKSQSFVGRIKSMCLW
 BJS   SLYCKDCMDHQELALREVEGMLTRLEAAEALYPSSQAMGALHPIYKSQSFVGRIKSMCLW
 199   SLYCKDCMDHQELALREVEGMLTRLEAAEALYPSSQAMGALHPIYKSQSFVGRIKSMCLW
 216   SLYCKDCMDHQELALREVEGMLTRLEAAEALYPSSQAMGALHPIYKSQSFVGRIKSMCLW

  FC   YNITKQNKLKLSILGKILARLQDEKFSWPVCTSSYIASDSGSSSASGVENDDSAVNSMDS
 BJS   YNITKQNKLKLSILGKILARLQDEKFSWPVCTSSYIATDSGSSSASGVENDDSAVNSMDS
 199   YNITKQNKLKLSILGKILARLQDEKFSWPVCTSSYIASDSGSSSASGVENDDSAVNSMDS
 216   YNITKQNKLKLSILGKILARLQDEKFSWPVCTSSYIASDSGSSSASGVENDDSAVNSMDS
```

```
 FC   SKPPSMAGSASRKGVTPCHKVQFMLNDATHVPGETSSSNESTSTEVSQWSSECSHSHMRK
BJS   SKPPSMAGSASRKGVTPCHKVQFMLNDATHVPGETSSSNESTSTEVSQWSSECSHSHMRK
199   SKPPSMAGSASRKGVTPCHKVQFMLNDATHVPGETSSSNESTSTEVSQWSSECSHSHMRK
216   SKPPSMAGSASRKGVTPCHKVQFMLNDATHVPGETSSSNESTSTEVSQWSSECSHSHMRK

 FC   GSMHDINIFSVEPLGTCSSNGTGEQSTGLYRKFIENVLKSRGLAKSLAFLHKLHNVALYK
BJS   GSMHDINIFSVEPLGTCSSNGTGEQSTGLYRKFIENVLKSRGLAKSLAFLHKLHNVALYK
199   GSMHDINIFSVEPLGTCSSNGTGEQSTGLYRKFIENVLKSRGLAKSLAFLHKLHNVALYK
216   GSMHDINIFSVEPLGTCSSNGTGEQSTGLYRKFIENVLKSRGLAKSLAFLHKLHNVALYK

 FC   AHIALEKPGAEDLDYESDAESIEEDVPRLDPQISREQVVELRTYGYWSEEAQSINLPSYI
BJS   AHIALEKPGAEDFDYESDAESIEEDVPRLDPEISREQVVELRTYGYWSEEAQSINLPSYI
199   AHIALEKPGAEDLDYESDAESIEEDVPRLDPQISREQVVELRTYGYWSEEAQSINLPSYI
216   AHIALEKPGAEDLDYESDAESIEEDVPRLDPQISREQVVELRTYGYWSEEAQSINLPSYI

 FC   PTFVFLSGIPLQFMHEFLRMRLETRPVRPNPLSLEQLMKELREGLTLALTHRERYQRHIT
BJS   PTFVFLSGIPLQFMHEFLRMRLETRPVRPNPLSLEQLMKELREGLTLALTHRERYQRHIT
199   PTFVFLSGIPLQFMHEFLRMRLETRPVRPNPLSLEQLMKELREGLTLALTHRERYQRHIT
216   PTFVFLSGIPLQFMHEFLRMRLETRPVRPNPLSLEQLMKELREGLTLALTHRERYQRHIT

 FC   TALVENEAELGSYISILNHYDATVRKTFELYLEYIDQLVLVAVPEGNQKSVLEKEWMFTK
BJS   TALVENEAELGSYISILNHYDATVRKTFELYLEYIDQLVLVAVPEGNQKSVLEKEWMFTK
199   TALVENEAELGSYISILNHYDATVRKTFELYLEYIDQLVLVAVPEGNQKSVLEKEWMFTK
216   TALVENEAELGSYISILNHYDATVRKTFELYLEYIDQLVLVAVPEGNQKSVLEKEWMFTK

 FC   LISPMIKGMHTLASQKFCGIISKLLRSISERLVKRTVELDQQIDGTADTDDNEEVKWQLL
BJS   LISPMIKGMHTLASQKFCGIISKLLRSISERLVKRTVELDQQIDGTADTDDNEEVKWQLL
199   LISPMIKGMHTLASQKFCGIISKLLRSISERLVKRTVELDQQIDGTADTDDNEEVKWQLL
216   LISPMIKGMHTLASQKFCGIISKLLRSISERLVKRTVELDQQIDGTADTDDNEEVKWQLL

 FC   TICRETQSLLTVERERSIKVLFFAKTFCRDVETTDFHREHYEHDVANQQHDFICSDVKAA
BJS   TICRETQSLLTVERERSIKVLFFAKTFCRDVETTDFHREHYEHDVANQQHDFICSDVKAA
199   TICRETQSLLTVERERSIKVLFFAKTFCRDVETTDFHREHYEHDVANQQHDFICSDVKAA
216   TICRETQSLLTVERERSIKVLFFAKTFCRDVETTDFHREHYEHDVANQQHDFICSDVKAA

 FC   FKLLQQDVLQVRNKLTAIIEGVQKRCCLSNMRDLDEQDKQAVLSRTREILHQGYKFGFEY
BJS   FKLLQQDVLQVRNKLTAIIEGVQKRCCLSNMRDLDEQDKQAVLSRTREILHQGYKFGFEY
199   FKLLQQDVLQVRNKLTAIIEGVQKRCCLSNMRDFYEQDKQAVLSRTREILHQGYKFGFEY
216   FKLLQQDVLQVRNKLTAIIEGVQKRCCLSNMRDLDEQDKQAVLSRTREILHQGYKFGFEY

 FC   HKDVIRLFEQKIMDQKDSGAHTVDLALGIIAYAKMWMHFVMERCERGRGMRPRWASQGLE
BJS   HKDVIRLFEQKIMDQKDSGAHTVDLALGIIAYAKMWMHFVMERCERGRGMRPRWASQGLE
199   HKDVIRLFEQKIMDQKDSGAHTVDLALGIIAYAKMWMHFVMERCERGRGMRPRWASQGLE
216   HKDVIRLFEQKIMDQKDSGAHTVDLALGIIAYAKMWMHFVMERCERGRGMRPRWASQGLE

 FC   FLILACDPQITQHLDDDEFEALKQQMDRCISHVIGITSEPEKVAKKKASPRTRKTSSPAT
BJS   FLILACDPQITQHLDDDEFEALKQQMDRCISHVIGITSEPEKVAKKKASPRTRKTSSPAT
199   FLILACDPQITQHLDDDEFEALKQQMDRCISHVIGITSEPEKVAKKKASPRTRKTSSPAT
216   FLILACDPQITQHLDDDEFEALKQQMDRCISHVIGITSEPEKVAKKKASPRTRKTSSPAT

 FC   SRSRTPTRTPMSAGMVLNPNTPPLQSPPYNKLLHPQFSLKEDVSVNSYSPVDSSDYVDTP
BJS   SRSRTPTRTPMSAGMVLNPNTPPLQSPPYNKLLHPQFSLKEDVSGTSYSPVDSSDYVDTP
199   SRSRTPTRTPMSAGMVLNPNTPPLQSPPYNKLLHPQFSLKEDVSVNSYSPVDSSDYVDTP
216   SRSRTPTRTPMSAGMVLNPNTPPLQSPPYNKLLHPQFSLKEDVSVNSYSPVDSSDYVDTP
```

```
 FC    CQRSANGELRLLVPQTPPTPASPGKSSLEGTPLALRQERVRDAVNRLDMDLEDGLRERRL
BJS    CQRSANGELRLLVPQTPPTPASPGKSSLESTPLALRQERVRDAVNRLDMDLEDGLRERRL
199    CQRSANGELRLLVPQTPPTPASPGKSSLEGTPLALRQERVRDAVNRLDMDLEDGLRERRL
216    CQRSANGELRLLVPQTPPTPASPGKSSLEGTPLALRQERVRDAVNRLDMDLEDGLRERRL

 FC    IGQVKSLNSSDKVHIRARSVHFRWHRGIKIGQGRFGKVYTAVNNNTGELMAMKEIAIQPG
BJS    IGQVKSLNSSDKVHIRARSVHFRWHRGIKIGQGRFGKVYTAVNNNTGELMAMKEIAIQPG
199    IGQVKSLNSSDKVHIRARSVHFRWHRGIKIGQGRFGKVYTAVNNNTGELMAMKEIAIQPG
216    IGQVKSLNSSDKVHIRARSVHFRWHRGIKIGQGRFGKVYTAVNNNTGELMAMKEIAIQPG

 FC    ETRALKNVAEELKILEGIKHKNLVRYYGIEVHREELLIFMELCSEGTLESLVELTGNLPE
BJS    ETRALKNVAEELKILEGIKHKNLVRYYGIEVHREELLIFMELCSEGTLESLVELTGNLPE
199    ETRALKNVAEELKILEGIKHKNLVRYYGIEVHREELLIFMELCSEGTLESLVELTGNLPE
216    ETRALKNVAEELKILEGIKHKNLVRYYGIEVHREELLIFMELCSEGTLESLVELTGNLPE

 FC    ALTRRFTAQLLSGVSELHKHGIVHRDIKTANIFLVDGSNSLKLGDFGSAVKIQAHTTVPG
BJS    ALTRRFTAQLLSGVSELHKHGIVHRDIKTANIFLVDGSNSLKLGDFGSAVKIQAHTTVPG
199    ALTRRFTAQLLSGVSELHKHGIVHRDIKTANIFLVDGSNSLKLGDFGSAVKIQAHTTVPG
216    ALTRRFTAQLLSGVSELHKHGIVHRDIKTANIFLVDGSNSLKLGDFGSAVKIQAHTTVPG

 FC    ELQGYVGTQAYMAPEVFTKTNSDGHGRAADIWSVGCVVVEMASGKRPWAQFDSNFQIMFK
BJS    ELQGYVGTQAYMAPEVFTKTNSDGHGRAADIWSVGCVVVEMASGKRPWAQFDSNFQIMFK
199    ELQGYVGTQAYMAPEVFTKTNSDGHGRAADIWSVGCVVVEMASGKRPWAQFDSNFQIMFK
216    ELQGYVGTQAYMAPEVFTKTNSDGHGRAADIWSVGCVVVEMASGKRPWAQFDSNFQIMFK

 FC    VGMGEKPQAPESLSQEGHDFIDHCLQHDPKRRLTAVELLEHNFCKYGRDECSSEQLQMQV
BJS    VGMGEKPQAPESLSQEGHDFIDHCLQHDPKRRLTAVELLEHNFCKYGRDECSSEQLQMQV
199    VGMGEKPQAPESLSQEGHDFIDHCLQHDPKRRLTAVELLEHNFCKYGRDECSSEQLQMQV
216    VGMGEKPQAPESLSQEGHDFIDHCLQHDPKRRLTAVELLEHNFCKYGRDECSSEQLQMQV

 FC    RGSFRRNVATSSS
BJS    RGSFRRNVATSSS
199    RGSFRRNVATSSS
216    RGSFRRNVATSSS
```

Appendix H: Sequence comparison for primary structure of RD/RC protein of *Mekk1* between *D. melanogaster* (BJS) and three strains of *D. simulans* (FC, 199 and 216).

```
Key: X = 5' and 3' nucleotides of kinase domain
     X = variant (not unique to FC)

      1        10        20        30        40        50        60
      |        |         |         |         |         |         |
 FC   MRRKKVEYRVKQTPSRPVQMTRNRIGALEEDMPPEDELAAHYEAFGTTPPRTRLKIKNRD
BJS   MRRKKVEYRVKQTPSRPVQMTRNRIGALEEDMPPEDELAAHYEAFGTTPPRTRLKIKNRD
199   MRRKKVEYRVKQTPSRPVQMARNRIGALEEDMPPEDELAAHYEAFGTTPPRTRLKIKNRD
216   MRRKKVEYRVKQTPSRPVQMARNRIGALEEDMPPEDELAAHYEAFGTTPPRTRLKIKNRD

 FC   WERKQKVVNVTASADLPASLGGTPKKTRTARSRVLRRNTMDCALLNEMFVNDESKRTDKR
BJS   WERKQKVVNVTASADLPASVGGTPKKTRTARSRVLRRNTMGCSLLNEMFVNDESKRTDKR
199   WERKQKVVNVTASADLPASLGGTPKKTRTARSRVLRRNTMDCALLNEMFVNDESKRTDKR
216   WERKQKVVNVTASADLPASLGGTPKKTRTARSRVLRRNTMDCALLNEMFVNDESKRTDKR

 FC   LQSLLRDSEREMKNSLATTAVAAPRGNSFHETAHPLESLDQIMPLNSEAMAPIPLRIASK
BJS   LQSLLRDSEREMKNSLATTAVAAPRGNSFHETAHPLESLDQIMPLNSEAMAPIPLRIASK
199   LQSLLRDSEREMKNSLATTAVAAPRGNSFHETAHPLESLDQIMPLNSEAMAPIPLRIASK
216   LQSLLRDSEREMKNSLATTAVAAPRGNSFHETAHPLESLDQIMPLNSEAMAPIPLRIASK

 FC   VVESCNRYRCVSSRPIGYRSSAPPLAFAAQLPAGMLRSDGRATPGLGKRKDFHETFANLI
BJS   VVESCNRYRCVSSRPIGYRSSAPPLAFAAQLPAGMLRSDGRATPGLGKRKDFHETFANLI
199   VVESCNRYRCVSSRPIGYRSSAPPLAFAAQLPAGMLRSDGRATPGLGKRKDFHETFANLI
216   VVESCNRYRCVSSRPIGYRSSAPPLAFAAQLPAGMLRSDGRATPGLGKRKDFHETFANLI

 FC   KLGSVDRQDAKLSQEEHTWQTELKDLIWLELQAWQADRTVEQQDKYLFEARQGVSDLLTH
BJS   KLGSVDRQDAKLSQEEHTWQTELKDLIWLELQAWQADRTVEQQDKYLFEARQGVSDLLTH
199   KLGSVDRQDAKLSQEEHTWQTELKDLIWLELQAWQADRTVEQQDKYLFEARQGVSDLLTH
216   KLGSVDRQDAKLSQEEHTWQTELKDLIWLELQAWQADRTVEQQDKYLFEARQGVSDLLTH

 FC   IINYKFQPRYRREPSLISLDSGIHSDSNSNASSPLPSKMCQGCMSLYCKDCMDHQELALR
BJS   IINYKFQPRYRREPSLISLDSGTHSDSNSNASSPLPSKMCQGCMSLYCKDCMDHQELALR
199   IINYKFQPRYRREPSLISLDSGIHSDSNSNASSPLPSKMCQGCMSLYCKDCMDHQELALR
216   IINYKFQPRYRREPSLISLDSGIHSDSNSNASSPLPSKMCQGCMSLYCKDCMDHQELALR

 FC   EVEGMLTRLEAAEALYPSSQAMGALHPIYKSQSFVGRIKSMCLWYNITKQNKLKLSILGK
BJS   EVEGMLTRLEAAEALYPSSQAMGALHPIYKSQSFVGRIKSMCLWYNITKQNKLKLSILGK
199   EVEGMLTRLEAAEALYPSSQAMGALHPIYKSQSFVGRIKSMCLWYNITKQNKLKLSILGK
216   EVEGMLTRLEAAEALYPSSQAMGALHPIYKSQSFVGRIKSMCLWYNITKQNKLKLSILGK

 FC   ILARLQDEKFSWPVCTSSYIASDSGSSSASGVENDDSAVNSMDSSKPPSMAGSASRKGVT
BJS   ILARLQDEKFSWPVCTSSYIATDSGSSSASGVENDDSAVNSMDSSKPPSMAGSASRKGVT
199   ILARLQDEKFSWPVCTSSYIASDSGSSSASGVENDDSAVNSMDSSKPPSMAGSASRKGVT
216   ILARLQDEKFSWPVCTSSYIASDSGSSSASGVENDDSAVNSMDSSKPPSMAGSASRKGVT

 FC   PCHKVQFMLNDATHVPGETSSSNESTSTEVSQWSSECSHSHMRKGSMHDINIFSVEPLGT
BJS   PCHKVQFMLNDATHVPGETSSSNESTSTEVSQWSSECSHSHMRKGSMHDINIFSVEPLGT
199   PCHKVQFMLNDATHVPGETSSSNESTSTEVSQWSSECSHSHMRKGSMHDINIFSVEPLGT
216   PCHKVQFMLNDATHVPGETSSSNESTSTEVSQWSSECSHSHMRKGSMHDINIFSVEPLGT

 FC   CSSNGTGEQSTGLYRKFIENVLKSRGLAKSLAFLHKLHNVALYKAHIALEKPGAEDLDYE
BJS   CSSNGTGEQSTGLYRKFIENVLKSRGLAKSLAFLHKLHNVALYKAHIALEKPGAEDFDYE
199   CSSNGTGEQSTGLYRKFIENVLKSRGLAKSLAFLHKLHNVALYKAHIALEKPGAEDLDYE
216   CSSNGTGEQSTGLYRKFIENVLKSRGLAKSLAFLHKLHNVALYKAHIALEKPGAEDLDYE
```

```
 FC   SDAESIEEDVPRLDPQISREQVVELRTYGYWSEEAQSINLPSYIPTFVFLSGIPLQFMHE
BJS   SDAESIEEDVPRLDPEISREQVVELRTYGYWSEEAQSINLPSYIPTFVFLSGIPLQFMHE
199   SDAESIEEDVPRLDPQISREQVVELRTYGYWSEEAQSINLPSYIPTFVFLSGIPLQFMHE
216   SDAESIEEDVPRLDPQISREQVVELRTYGYWSEEAQSINLPSYIPTFVFLSGIPLQFMHE

 FC   FLRMRLETRPVRPNPLSLEQLMKELREGLTLALTHRERYQRHITTALVENEAELGSYISI
BJS   FLRMRLETRPVRPNPLSLEQLMKELREGLTLALTHRERYQRHITTALVENEAELGSYISI
199   FLRMRLETRPVRPNPLSLEQLMKELREGLTLALTHRERYQRHITTALVENEAELGSYISI
216   FLRMRLETRPVRPNPLSLEQLMKELREGLTLALTHRERYQRHITTALVENEAELGSYISI

 FC   LNHYDATVRKTFELYLEYIDQLVLVAVPEGNQKSVLEKEWMFTKLISPMIKGMHTLASQK
BJS   LNHYDATVRKTFELYLEYIDQLVLVAVPEGNQKSVLEKEWMFTKLISPMIKGMHTLASQK
199   LNHYDATVRKTFELYLEYIDQLVLVAVPEGNQKSVLEKEWMFTKLISPMIKGMHTLASQK
216   LNHYDATVRKTFELYLEYIDQLVLVAVPEGNQKSVLEKEWMFTKLISPMIKGMHTLASQK

 FC   FCGIISKLLRSISERLVKRTVELDQQIDGTADTDDNEEVKWQLLTICRETQSLLTVERER
BJS   FCGIISKLLRSISERLVKRTVELDQQIDGTADTDDNEEVKWQLLTICRETQSLLTVERER
199   FCGIISKLLRSISERLVKRTVELDQQIDGTADTDDNEEVKWQLLTICRETQSLLTVERER
216   FCGIISKLLRSISERLVKRTVELDQQIDGTADTDDNEEVKWQLLTICRETQSLLTVERER

 FC   SIKVLFFAKTFCRDVETTDFHREHYEHDVANQQHDFICSDVKAAFKLLQQDVLQVRNKLT
BJS   SIKVLFFAKTFCRDVETTDFHREHYEHDVANQQHDFICSDVKAAFKLLQQDVLQVRNKLT
199   SIKVLFFAKTFCRDVETTDFHREHYEHDVANQQHDFICSDVKAAFKLLQQDVLQVRNKLT
216   SIKVLFFAKTFCRDVETTDFHREHYEHDVANQQHDFICSDVKAAFKLLQQDVLQVRNKLT

 FC   AIIEGVQKRCCLSNMRDLDEQDKQAVLSRTREILHQGYKFGFEYHKDVIRLFEQKIMDQK
BJS   AIIEGVQKRCCLSNMRDLDEQDKQAVLSRTREILHQGYKFGFEYHKDVIRLFEQKIMDQK
199   AIIEGVQKRCCLSNMRDFYEQDKQAVLSRTREILHQGYKFGFEYHKDVIRLFEQKIMDQK
216   AIIEGVQKRCCLSNMRDLDEQDKQAVLSRTREILHQGYKFGFEYHKDVIRLFEQKIMDQK

 FC   DSGAHTVDLALGIIAYAKMWMHFVMERCERGRGMRPRWASQGLEFLILACDPQITQHLDD
BJS   DSGAHTVDLALGIIAYAKMWMHFVMERCERGRGMRPRWASQGLEFLILACDPQITQHLDD
199   DSGAHTVDLALGIIAYAKMWMHFVMERCERGRGMRPRWASQGLEFLILACDPQITQHLDD
216   DSGAHTVDLALGIIAYAKMWMHFVMERCERGRGMRPRWASQGLEFLILACDPQITQHLDD

 FC   DEFEALKQQMDRCISHVIGITSEPEKVAKKKASPRTRKTSSPATSRSRTPTRTPMSAGMV
BJS   DEFEALKQQMDRCISHVIGITSEPEKVAKKKASPRTRKTSSPATSRSRTPTRTPMSAGMV
199   DEFEALKQQMDRCISHVIGITSEPEKVAKKKASPRTRKTSSPATSRSRTPTRTPMSAGMV
216   DEFEALKQQMDRCISHVIGITSEPEKVAKKKASPRTRKTSSPATSRSRTPTRTPMSAGMV

 FC   LNPNTPPLQSPPYNKLLHPQFSLKEDVSVNSYSPVDSSDYVDTPCQRSANGELRLLVPQT
BJS   LNPNTPPLQSPPYNKLLHPQFSLKEDVSGTSYSPVDSSDYVDTPCQRSANGELRLLVPQT
199   LNPNTPPLQSPPYNKLLHPQFSLKEDVSVNSYSPVDSSDYVDTPCQRSANGELRLLVPQT
216   LNPNTPPLQSPPYNKLLHPQFSLKEDVSVNSYSPVDSSDYVDTPCQRSANGELRLLVPQT

 FC   PPTPASPGKSSLEGTPLALRQERVRDAVNRLDMDLEDGLRERRLIGQVKSLNSSDKVHIR
BJS   PPTPASPGKSSLESTPLALRQERVRDAVNRLDMDLEDGLRERRLIGQVKSLNSSDKVHIR
199   PPTPASPGKSSLEGTPLALRQERVRDAVNRLDMDLEDGLRERRLIGQVKSLNSSDKVHIR
216   PPTPASPGKSSLEGTPLALRQERVRDAVNRLDMDLEDGLRERRLIGQVKSLNSSDKVHIR

 FC   ARSVHFRWHRGIKIGQGRFGKVYTAVNNNTGELMAMKEIAIQPGETRALKNVAEELKILE
BJS   ARSVHFRWHRGIKIGQGRFGKVYTAVNNNTGELMAMKEIAIQPGETRALKNVAEELKILE
199   ARSVHFRWHRGIKIGQGRFGKVYTAVNNNTGELMAMKEIAIQPGETRALKNVAEELKILE
216   ARSVHFRWHRGIKIGQGRFGKVYTAVNNNTGELMAMKEIAIQPGETRALKNVAEELKILE
```

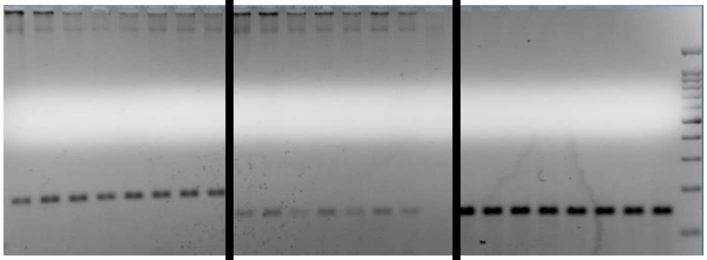
```
 FC     GIKHKNLVRYYGIEVHREELLIFMELCSEGTLESLVELTGNLPEALTRRFTAQLLSGVSE
BJS     GIKHKNLVRYYGIEVHREELLIFMELCSEGTLESLVELTGNLPEALTRRFTAQLLSGVSE
199     GIKHKNLVRYYGIEVHREELLIFMELCSEGTLESLVELTGNLPEALTRRFTAQLLSGVSE
216     GIKHKNLVRYYGIEVHREELLIFMELCSEGTLESLVELTGNLPEALTRRFTAQLLSGVSE

 FC     LHKHGIVHRDIKTANIFLVDGSNSLKLGDFGSAVKIQAHTTVPGELQGYVGTQAYMAPEV
BJS     LHKHGIVHRDIKTANIFLVDGSNSLKLGDFGSAVKIQAHTTVPGELQGYVGTQAYMAPEV
199     LHKHGIVHRDIKTANIFLVDGSNSLKLGDFGSAVKIQAHTTVPGELQGYVGTQAYMAPEV
216     LHKHGIVHRDIKTANIFLVDGSNSLKLGDFGSAVKIQAHTTVPGELQGYVGTQAYMAPEV

 FC     FTKTNSDGHGRAADIWSVGCVVVEMASGKRPWAQFDSNFQIMFKVGMGEKPQAPESLSQE
BJS     FTKTNSDGHGRAADIWSVGCVVVEMASGKRPWAQFDSNFQIMFKVGMGEKPQAPESLSQE
199     FTKTNSDGHGRAADIWSVGCVVVEMASGKRPWAQFDSNFQIMFKVGMGEKPQAPESLSQE
216     FTKTNSDGHGRAADIWSVGCVVVEMASGKRPWAQFDSNFQIMFKVGMGEKPQAPESLSQE

 FC     GHDFIDHCLQHDPKRRLTAVELLEHNFCKYGRDECSSEQLQMQVRGSFRRNVATSSS
BJS     GHDFIDHCLQHDPKRRLTAVELLEHNFCKYGRDECSSEQLQMQVRGSFRRNVATSSS
199     GHDFIDHCLQHDPKRRLTAVELLEHNFCKYGRDECSSEQLQMQVRGSFRRNVATSSS
216     GHDFIDHCLQHDPKRRLTAVELLEHNFCKYGRDECSSEQLQMQVRGSFRRNVATSSS
```

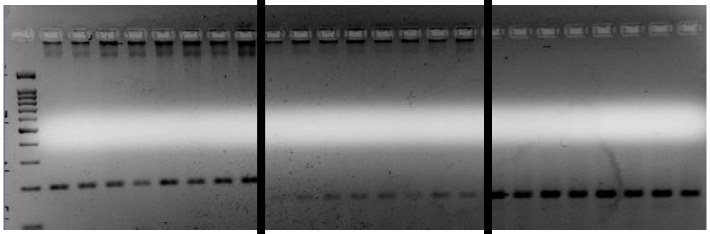
Appendix I: Gel images from RT-PCR comparing relative levels of *Mekk1* expression in females between *D. simulans* (FC) and *D. melanogaster* (BJS).

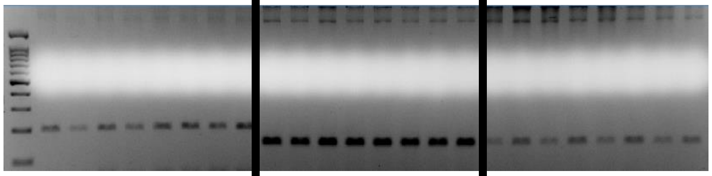
Labels RB and RD/RC correspond to *Mekk1* transcripts of interest, while control gene is labelled as *Rpl32*. “m” and “s” correspond to samples from *D. melanogaster* (BJS) and *D. simulans* (FC), respectively. Samples on the same gel are technical replicates while different gels indicate biological replicates. Ladder (farthest left or right lane depending on gel) represents 100bp increments.

<u>Adult Heads</u>

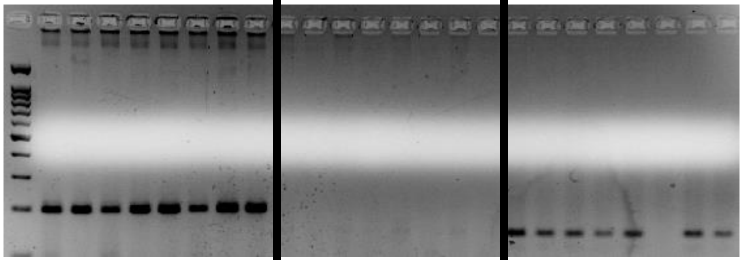
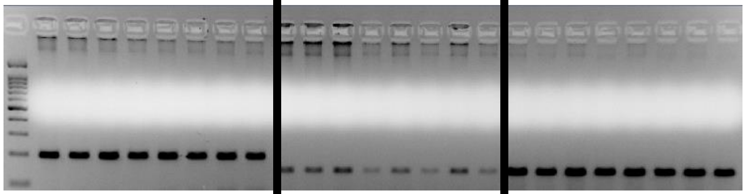
RB	RD	Rpl
m/s/m/s/m/s/m/s	m/s/m/s/m/s/m/s	m/s/m/s/m/s/m/s



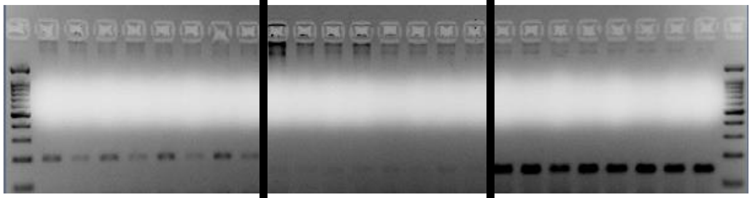
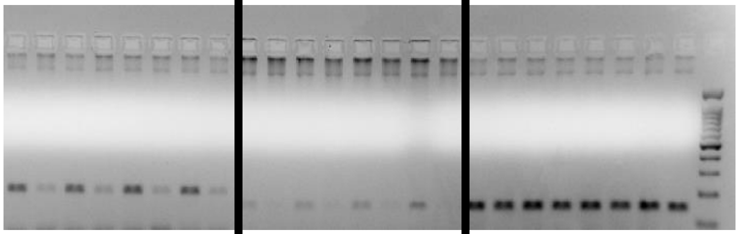




Adult Bodies

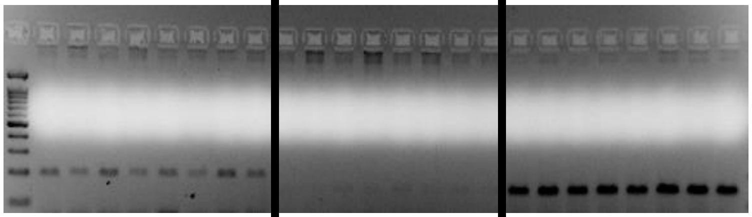
<table>
<tr><td>RB</td><td>RD</td><td>Rpl</td></tr>
<tr><td>m/s/m/s/m/s/m/s</td><td>m/s/m/s/m/s/m/s</td><td>m/s/m/s/m/s/m/s</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3"></td></tr>
</table>

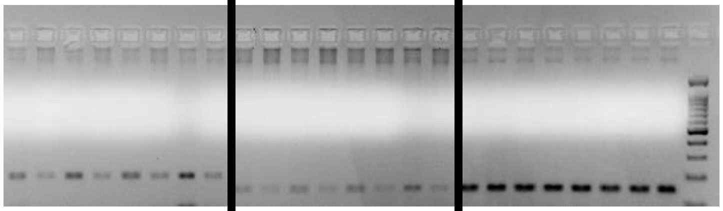
Pupae Heads

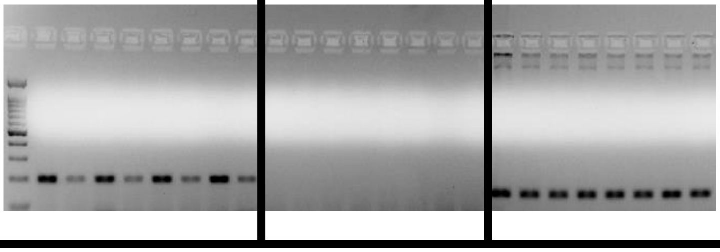
RB	RD	Rpl
m/s/m/s/m/s/m/s	m/s/m/s/m/s/m/s	m/s/m/s/m/s/m/s
		
		
		

Pupae Bodies

RB	RD	Rpl
m/s/m/s/m/s/m/s	m/s/m/s/m/s/m/s	m/s/m/s/m/s/m/s

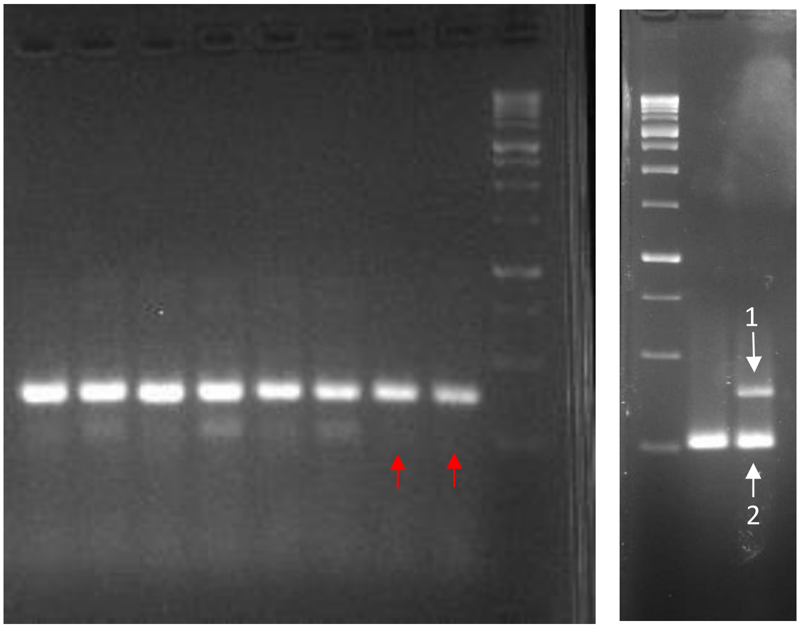






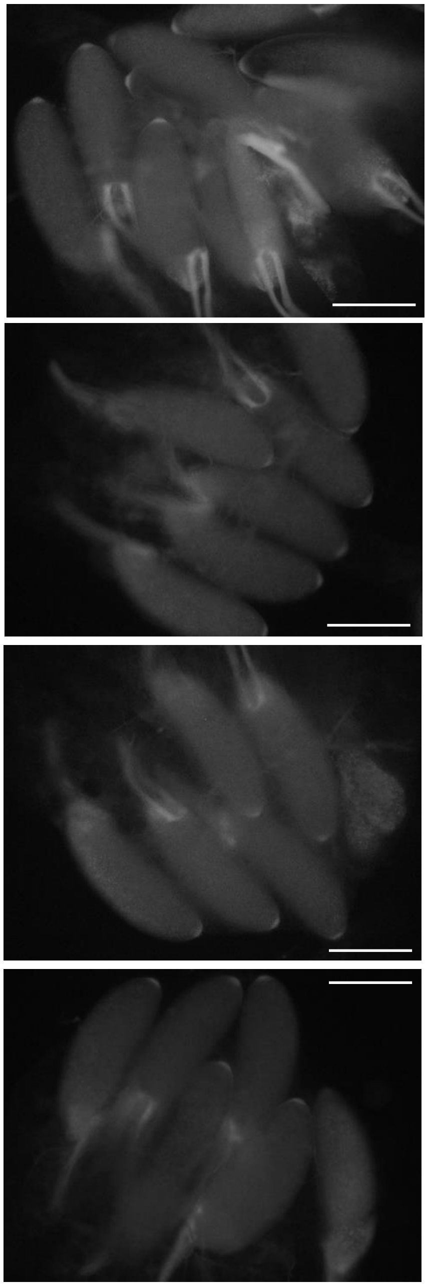
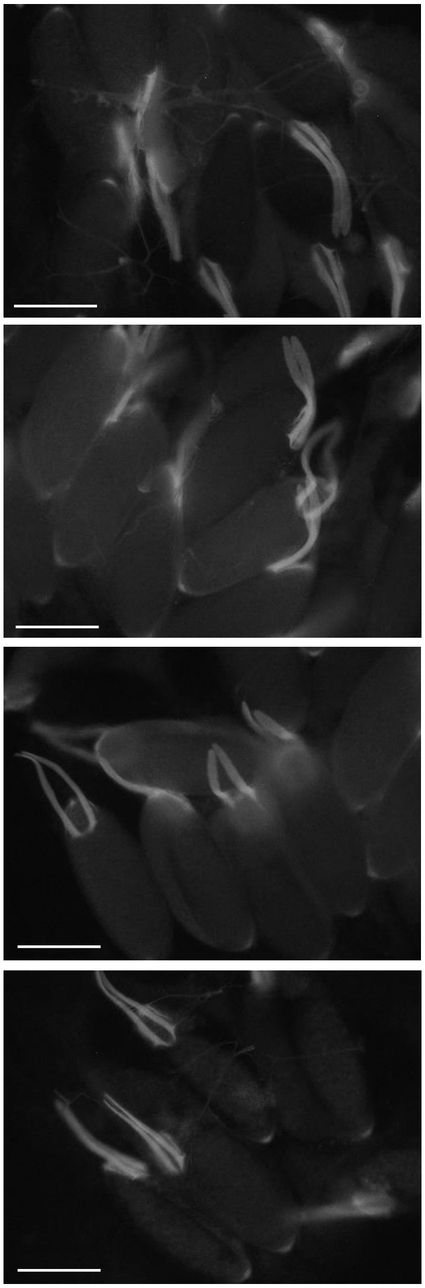
Appendix J: Genotypes of flies from behavioural assays examining tissue-specific rescue of *Mekk1* and its effect on behaviour. .

Left gel indicates flies (n=8) used in behavioural assays. White arrows indicate expected band size for *P*-element insertion (1) and for the UAS element (2). The right gel contains a positive control while the left individuals that were assayed for behaviour and were expected to have both elements. All reactions were run from the same master mix at the same time. Red arrows indicate where *P*-element insertion is not present (likely due to recombination).



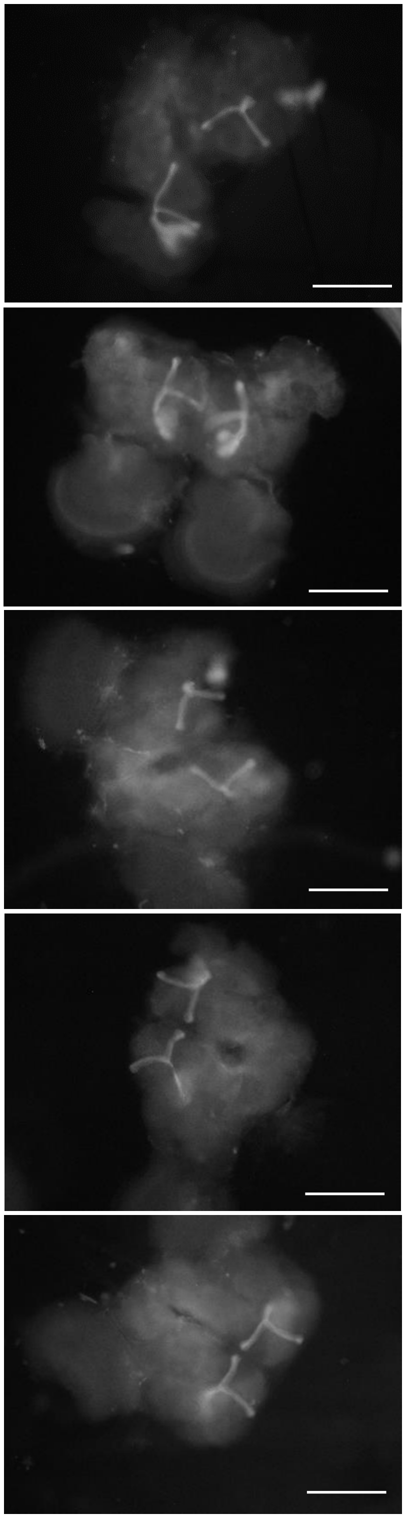
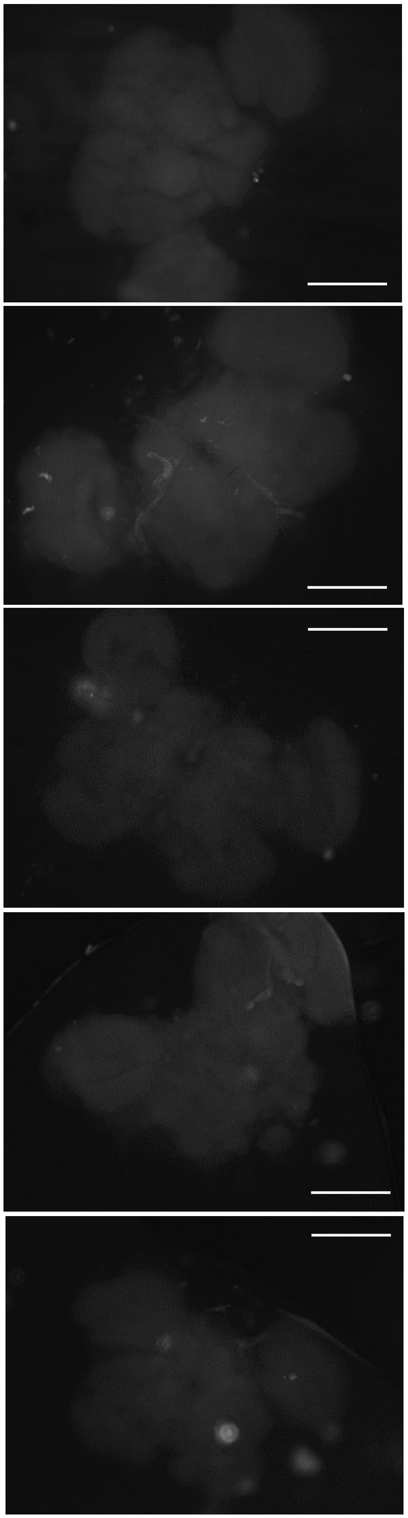
Appendix K: Fluorescent microscopy examining female adult ovaries of the strain with an enhancer trap within *Mekk1* in *D. melanogaster*.

Images taken under 10X microscopy, scale bars are 50µm.

Enhancer Trap	Control
	

Appendix L: Fluorescent microscopy examining female adult brains of the strain with an enhancer trap within *Mekk1* in *D. melanogaster*.

Images taken under 10X microscopy, scale bars are 50µm.

Enhancer trap	Control
	

CV

CURRENT POSITION

M.Sc. candidate (*Cell and Molecular Biology)* **2014 - Present**
Western University, London, Ontario

- Project title: "Characterizing the genetic basis of behavioural isolation between *Drosophila* species"

EDUCATION

Certificate in University Teaching and Learning *(Teaching Support Center)* **2014 - 2016**
Western University, London, Ontario
Bachelor of Science *(Honours Specialization in Biology)* **2010 - 2014**
Western University, London, Ontario

- Dean's Honour List, Thesis student

SCHOLARSHIPS AND AWARDS

Boys and Girls Club of Canada Scholarship (2000 CAD) **2017**
Boys and Girls Club of Canada Central Region (Ontario)
Ontario Volunteer Service Award **2016**
Ontario Ministry of Citizenship, Immigration and International Trade
Graduate Student Travel Award (250 CAD) **2016**
Western University
Queen Elizabeth II Graduate Scholarship in Science and Technology (15,000 CAD)
Province of Ontario and Western University **2015-2016**
Western Graduate Research Scholarship (4150 CAD per annum) **2016**
Western University
Best Ecology & Evolution presentation (50 CAD) **2014**
Honours Biology Day, Western University
Dean's Honour List **2012-2014**
Western University
University of Western Ontario Scholarship of Distinction (1000 CAD) **2010**
Western University

CERTIFICATIONS

Standard First Aid/CPR-C **2015**
High Five – Principles of Healthy Child Development (PHCD) **2014**
Supervisor Health and Safety Awareness **2014**
Biosafety Training **2014**
Comprehensive WHMIS **2013**
Occupational Health and Safety **2013**
General Laboratory Safety and Hazardous Waste Management Training **2013**
Accessibility for Ontarians with Disabilities Act Training **2013**

TEACHING

Guest Lecturer **2015**
Behavioural Genetics (BIO3598)
Department of Biology, Western University, London, Ontario
Teaching Assistant **2014-2016**

Cell Biology Laboratory (BIO3326); Behavioural Genetics (BIO3598); Genomics and Beyond: A Laboratory Course (BIO3596); Biology for Science/General Biology (BIO1001/1201)
Department of Biology, Western University, London, Ontario

PUBLICATIONS

Dooner, C. (2017). Preventing Lab Mistakes: The Importance of Including Rationale in Laboratory Protocols. Teaching Innovation Projects, 7(1), 3.

Dooner, J. **Dooner, C.** and Lee, S. (2013). Comparison of cerebral oximetry and transcranial doppler in decision to shunt for carotid endarterectomy. Canadian Journal of Surgery **56** (6 Suppl 1): S154-S155.

CONFERENCE PRESENTATIONS

Dooner, C., Kohalmi, S.E., and Moehring, A.J. and Characterizing the genetic basis of behavioural isolation between *Drosophila* species. *Canadian Society of Zoology*, London, ON, Canada **2016**.

Dooner, C., Kohalmi, S.E., and Moehring, A.J. Characterizing the genetic basis of behavioural isolation between *Drosophila* species. *Fallona Family Interdisciplinary Showcase*, London, ON, Canada **2015**.

Dooner, C., Kohalmi, S.E., and Moehring, A.J. Characterizing the genetic basis of behavioural isolation between *Drosophila* species. *Biology Graduate Research Forum*, London, ON, Canada **2015**.

Dooner, C., and Moehring, A.J. The effect of accessory proteins on sperm survivability in *Drosophila pseudoobscura*. *Ontario Ecology, Ethology, and Evolution Colloquium*, Guelph, ON, Canada **2014**.

Dooner, C., and Moehring, A. J. The effect of accessory proteins on sperm survivability in *Drosophila pseudoobscura*. *Ontario Biology Day*, Mississauga, ON, Canada **2014**.

Dooner, C., Lee, S. and Dooner, J. Comparison of cerebral oximetry and transcranial doppler in decision to shunt for carotid endararterectomy. *Canadian Society for Vascular Surgery Annual Meeting on Vascular Surgery*, Edmonton, Alberta, Canada **2013**.