To Boys, Girls Have Different Tastes: Can Differences in Gustatory Receptor Genes Account for Premating Isolation in *Drosophila virilis*?

By Chengyuan Bo

Supervisor: Dr. Alberto Civetta

Committee: Dr. J. Franck and Ms. D. Rittberg

A thesis submitted in partial fulfillment of the Honours Thesis in Biology (05.4111/6)

Department of Biology

The University of Winnipeg

2008

ABSTRACT

Sexual isolation between closely-related species can be premating and/or postmating. *Drosophila virilis* males show strong premating isolation towards closely-related heterospecific females of *Drosophila virilis* group. A previous study involving courtship behaviour examination and genetic mapping suggested that X-linked gustatory receptor genes may be responsible for this isolation. Using information from the *D. melanogaster* genome and currently available data on the *D. virilis* genome, I have identified five gustatory receptor genes: Gr2a, Gr5a, Gr8a, Gr9a and Gr10a in *D. virilis*. Four of them (except Gr5a) were successfully isolated from *D. americana, D. novamexicana, D. texana* and *D. virilis* using PCR; and Gr2a and Gr9a were successfully sequenced. Molecular evolution analysis for Gr2a and Gr9a show no evidence of positive selection at either of the genes. There is evidence of significant changes in rates of evolution among species, with no indication of *D. virilis* branch-specific acceleration. It is possible that Gr genes work together to account for *D. virilis* male pre-mating isolation towards heterospecific females.

ACKNOWLEDGEMENTS

I wish to place on record my deep sense of gratitude, profound thanks to my supervisor, Dr. Alberto Civetta for his valuable guidance, encouragement and support towards this project. Barb Brouwers, Lisa Levesque and Desiree Nickel also deserve credit for their help in every aspect in this research. I would like to thank my committee members, Dr. Franck and Ms. Rittberg, for being willing to deal with a trouble-maker who is of course, called Chengyuan Bo. I am grateful to Dr. Ric Moodie, the coordinator of this thesis course, for answering my questions and providing me advice patiently, it is with extra “honour” considering this year is the last year before Dr. Moodie’s retirement.

TABLE OF CONTENTS

ABSTRACT……... ii
ACKNOWLEDGEMENTS……... iii
TABLE OF CONTENTS…… iv
LIST OF FIGURES…… v
LIST OF TABLES……. vi
LIST OF ABBREVIATIONS…… vii

1. INTRODUCTION…….. **Page 1-6**

1.1 Speciation……. Page 1-2
1.2 *Drosophila Virilis* Group…….. Page 2-3
1.3 Chemosensory Receptors (Gustatory receptors)…….. Page 3-5
1.4 X chromosome effect on sexual isolation between *D.virilis* and sister species Page 5-6

2. MATERIALS & METHODS…….. **Page 7-17**

2.1 *Drosophila* Fly stock…… Page 7
2.2 Candidate gustatory receptor genes in *D. virilis*……..…. Page 7-9
2.3 Primers design…….…. Page 9-10
2.4 DNA extraction……. Page 10-11
2.5 PCR …….. Page 11-12
2.6 Cleaning and quantifying PCR products…….. Page 12-14
2.7 DNA sequencing……... Page 14-16
2.8 Data analysis……. Page 16-17

3. RESULTS…… **Page 18-24**

3.1 In silico identification of candidate gustatory genes in *D. virilis*…….…. Page 18-19
3.2.1 Gr2a…….. Page 19-22
3.2.2 Gr9a…….. Page 22-24
3.2.3 Gr5a, Gr8a and Gr10a…….. Page 24

4. DISCUSSION…… **Page 25-28**

5. CONCLUSIONS..…….. **Page 29**

6. FUTURE STUDY……... **Page 29**

7. REFERENCES…..…….. **Page 30-32**

APPENDIX…… **Page 33-40**

LIST OF FIGURES

Figure 1-1. Expression sites of gustatory receptors in *Drosophila* species. Gustatory receptors are found in taste neurons located on the labellum, pharynx, the legs and anterior wing margins — Page 4

Figure 2-1. Alignment between the Gr2a gene of *D.melanogaster* and a similar sequence (candidate Gr2a gene) of *D. virilis* — Page 8

Figure 3-1a. The sequence of the *D.melanogaster* Gr10a gene aligned with homologous sequence in *D. virilis*, with a small portion of Gr10a gene of *D.melanogaster* showing no homologous sequence in *D. virilis* genome — Page 18

Figure 3-1b. The Gr10b gene of *D. melanogaster* has no homologous sequence in *D. virilis* — Page 18

Figure 3-2. MotifScan result using the amino acid sequence of candidate Gr9a gene found in *D. virilis* — Page 19

Figure 3-3. Multiple sequence alignments of candidate Gr2a genes from *D. americana, D. novamexicana, D. texana* and *D. virilis* — Page 32-35

Figure 3-4. The alignment of amino acid products of candidate Gr2a genes for *D. americana, D. novamexicana, D. texana* and *D. virilis* — Page 36

Figure 3-5.The proportion of nonsynonymous to synonymous substitutions (d_N/d_S) accumulated along the branch leading to *D. novamexicana, D. texana* and *D. virilis* from the common ancestor, based on the sequence data of Gr2a — Page 22

Figure 3-6. Multiple sequence alignments of candidate Gr9a genes from *D. americana, D. novamexicana, D. texana* and *D. virilis* — Page 37-39

Figure 3-7. The alignment of amino acid products of candidate Gr9a genes for *D. americana, D. novamexicana, D. texana* and *D. virilis* — Page 40

Figure 3-8. The proportion of nonsynonymous to synonymous substitutions (d_N/d_S) accumulated along the branch leading to *D. novamexicana, D. texana* and *D. virilis* from the common ancestor, based on the sequence data of Gr9a — Page 24

LIST OF TABLES

Table 2-1. Primer description and its corresponding sequence (5' to 3') Page 9-10

Table 2-2. The chemical composition and concentration used in PCR reaction Page 11

Table 2-3. Temperature and cycling condition used in PCR reaction to amplify Gr genes Page 12

Table 2-4. Table for estimating the required amount (in ng) of double stranded DNA for sequencing, provided by Bechman Coulter, Inc Page 13

Table 2-5. The chemical composition and concentrations used in the sequencing reaction Page 15

Table 2-6. Temperature and cycling conditions used for the sequencing reaction Page 15

Table 3-1. Summary of sequence changes in the candidate Gr2a genes between *D. americana, D. novamexicana, D. texana* and *D. virilis* Page 21

Table 3-2. Results from Phylogenetic Analysis by Maximum Likelihood (PAML) analysis of the Gr2a gene in comparison between *D. novamexicana, D. texana* and *D. virilis* Page 21

Table 3-3. Summary of sequence changes in the candidate Gr9a genes between *D. americana, D. novamexicana, D. texana* and *D. virilis* Page 23

Table 3-4. Results from Phylogenetic Analysis by Maximum Likelihood (PAML) analysis of the Gr9a gene in comparison between *D. novamexicana, D. texana* and *D. virilis* Page 23

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide-triphosphate
EDTA	Ethylenediaminetetraacetate
Gr	Gustatory receptor
$MgCl_2$	Magnesium Chloride
mL	Millilitre
µL	Microlitre
PCR	Polymerase Chain Reaction
SDDW	Sterile distilled deionized water
SDS	Sodium dodecyl sulfate
fmol	Femtomole (One quadrillionth mole)
d_N/d_S	Ratio of nonsynonymous to synonymous substitutions

1. INTRODUCTION

1.1 Speciation

Speciation can be defined as the evolutionary process by which new biological species arise. The biological species concept states that "a species is a group of actually or potentially interbreeding individuals who are reproductively isolated from other such groups." (Coyne and Orr, 2004). Therefore, the divergence from a common ancestor into two or more species is dependent upon the establishment of barriers to gene flow. In allopatric speciation, the initial barrier is geographic, and over time the populations may acquire unique genetic changes through a balance between mutations and natural selection, since different geographic regions are likely to exert different selective pressures. When enough change has accumulated the two populations may no longer be able to interbreed, even if they were brought together geographically. In this mode, the geographic isolation may lead to reproductive isolation and the formation of new species. Although allopatric speciation is thought to be the main mechanism of speciation, occasionally two populations can live in the same geographic location and become reproductively isolated (i.e. sympatric and parapatric speciation). For example, new plant species can form in sympatry by simply duplicating the number of chromosomes. Interbreeding between plants that differ in chromosome number usually results in the production of sterile offspring.

Reproductive isolation between closely related species is commonly the result of premating and/or postmating incompatibilities. Premating isolation can be caused by ecological isolation (species live in different habitats), temporal isolation (species breed at different time), behavioral isolation (species copulation requires distinct courtship and mating rituals) or mechanical isolation (sexual organ incompatibility) which prevent the production of unfit zygotes (Coyne and Orr 2004). Alternatively, postmating isolation results from failure of

hybrid zygotes developing normally with ability to reproduce. The postzygotic isolation mechanisms include hybrid inviability, hybrid sterility or more commonly, the hybrid is healthy and fertile, but less fit. Enhanced premating isolation in areas of species overlap, for instance, is often attributed to the reinforcement of mating discrimination by natural selection against maladaptive hybridization (Dobzhansky, 1937; Ortiz *et al.*, 2004).

1.2 *Drosophila virilis* Group:

Drosophila is a genus that belongs to the family *Drosophilidae*, known to us as "fruit flies". The best known member of the genus is *Drosophila melanogaster*, which has been used in research since 1906 and whose genome was fully sequenced in 2000 (Adams *et al*., 2000). The vast amount of genetic research conducted on *D. melanogaster* means that many of its genes are well annotated in terms of function. Another group of species within the genus is the *virilis* group. This group is divided into two phylads: *virilis* and *montana* (phylad is a term referring to the phylogenetic subgroup within the *Drosophila virilis* group). The gross morphological differences of members of the *virilis* phylad are minimal, with the exception of differences in the structure of the male genitalia (Kulikov *et al*., 2004). Based on chromosomal rearrangement, the phylogenetic tree of the *virilis* phylad was constructed (Throckmorton, 1982): *D. virilis* and *D. lummei* have the same chromosomal arrangement: one sex chromosome pair plus four separate autosomal chromosome pairs. *D. novamexicana* and *D. texana* form a separate group with chromosomes 2 and 3 fused to form one giant chromosome. *D. americana* shows, in addition to the fusion of chromosomes 2 and 3, a polymorphism involving X chromosome fusion with chromosome 4. This model places *D. americana* outside of *D. novamexicana* and *D. texana*, but phylogenetic reconstruction based

upon nuclear DNA sequence data place *D. americana* as a sister species to *D. novamexicana* (Spicer and Bell, 2002; McAllister *et al*., 2004).

Previous work analyzed the male courtship behaviour of closely related species of the *virilis* phylad when paired with conspecific or heterospecific females and found that *Drosophila virilis* males prefer not to mate with closely-related heterospecific females (Nickel, 2007). During the early stage of the courtship repertoire, *Drosophila* males tap both conspecific and heterospecific females similarly, with their fore tarsi, the very distal portion of the front legs, and sample female cuticular hydrocarbons (Liimatainen, Jallon, 2007; Kaori Ishii *et al.*, 2001). Behavioral assays demonstrated that *D. virilis* males walk away from heterospecific females after tapping, and thus terminate the courtship ritual, suggesting the possible role for gustation in conspecific female identification (Nickel, 2007).

1.3 Chemosensory Receptors (Gustatory receptors):

Animals use their chemosensory systems to detect and discriminate among chemical signals in the environment such as different food sources and chemical cues needed to recognize possible mates (Cobb and Jallon, 1990; Coyne et al. 1994; Jallon, 1984). The observation that *Drosophila* males have significantly more taste bristles on their forelegs (about 50) than the females (about 37) (Nayak and Singh, 1983) along with the fact that female pheromones are non-volatile but concentrated in the abdomen region (Coyne et al., 1994; Ferveur et al., 1996), suggest that chemosensation through contact mediates mate recognition.

In *Drosophila*, the receptions of gustatory (taste) and olfactory (smell) signals are key elements in species recognition. Unlike humans, in which taste perception is restricted to one

organ, fruit flies have taste bristles not only on the labellum, which is equivalent to the human tongue, but also on the wings, legs, genitalia and in the pharynx (Natasha *et al*., 2005; Stocker, 1994) (Figure 1-1).

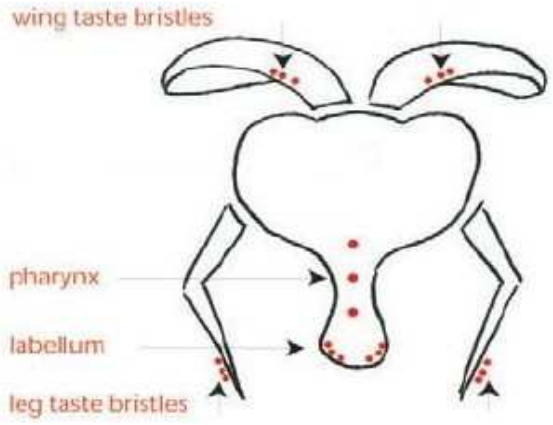

Figure 1-1. Expression sites of gustatory receptors in *Drosophila* species. Gustatory receptors are found in taste neurons located on the labellum, pharynx, the legs and anterior wing margins.

The taste bristles are associated with taste neurons expressing a class of G-Protein-coupled receptors, called gustatory receptors (Clyne, 2000; Scott, 2001). Current research has predicted that the insect chemoreceptor superfamily in *Drosophila melanogaster* consists of 62 odorant receptor (Or) and 68 gustatory receptor (Gr) proteins, encoded by families of 60 Or and 60 Gr genes, respectively, through alternative splicing (Clyne et al., 1999; Robertson *et al*., 2003).

In members of the *Drosophila melanogaster* group, cuticular hydrocarbons have been shown to contribute to mate choice within the species (Jallon, 1984) and sexual isolation between species (Cobb and Jallon, 1990; Coyne et al. 1994). In members of the *Drosophila virilis* group, variations in the female cuticular hydrocarbon profiles may allow and help the males

of the *virilis* phylad in species recognition. Bartelt *et al*. (1986) has shown that the female hydrocarbon profile of *D.virilis* is composed of a twenty-five carbon chain (11-pentacosene), *D. novamexicana* is composed of a twenty-seven carbon chain, and *D. texana* is composed of a twenty-nine carbon chain.

1.4 X chromosome effect on sexual isolation between *D. virilis* and sister species

As mentioned before, sexual isolation in *Drosophila* genus is frequently observed. The studies of sexual isolation within this genus have been usually conducted by examining the inviability and sterility of resulting interspecies hybrid generations (Patterson and Stone, 1952). Using a backcross approach, Coyne and Orr (1989) created *D. novamexicana* males with their X chromosome derived from *D.virilis*, used these males to cross with "regular" *D.novamexicana* females and tested the sterility and inviability of their offspings. This research found that the genes with the largest effect on hybrid sterility and inviability are on the X chromosome and they suggested that the large effect of the X chromosome on postzygotic isolation results from the rapid divergence of X-linked loci by natural selection. In addition, by studying the mating behaviours between species of another *Drosophila* group (*D. pseudoobscura* and *D. persimilis*), it was observed that during courtship, males and females from different species can stop mating due to a behavioral barrier. Therefore, premating isolation was suggested as a reinforcement mechanism to the sexual isolation (Ortiz *et al.*, 2004). The reinforcement occurs when natural selection strengthens behavioral discrimination to prevent costly interspecies matings, which would produce sterile or inviable hybrids. Premating isolation has evolutionary advantages since it stops a bad combination before combination happens, and

from an evolutionary view, this increases the fitness of both male and female flies. Nickel (2007) used a similar cross approach to test which chromosome was most likely responsible for premating isolation among species of the *D. virilis* group. Their results also suggested that genes with the largest effect on premating isolation are most likely located on X chromosome, although X- autosomal interaction may also affect pre-mating isolation.

So, to summarize what is known:

1. Sexual isolation is observed between *D.virilis* males and females of closely-related species. Such isolation is caused by mating discrimination after tapping.
2. The ability of *D.virilis* males to recognize females as conspecfic or heterospecific is most likely dependent on gustatory receptors located in leg bristles.
3. Genetic mapping research suggests a major effect of genes on the X chromosome affecting premating isolation.
4. Gustatory receptor genes, which encode gustatory receptors, are located on autosomal chromosomes and the X chromosome.

The purpose of this research is to attempt to answer the question: “do X-linked gustatory receptor genes account for premating isolation in the *Drosophila virilis* group?”

To answer this question, I identified five candidate gustatory receptor genes on the *D. virilis* X chromosome and then proceeded to analyze patterns of sequence divergence between *D. virilis* and its close relatives for signatures of adaptive diversification between species and/or acceleration/deceleration of rates of evolution in the *D. virilis* lineage.

2. MATERIALS & METHODS

2.1 *Drosophila* fly stock:

The fly stocks used in this research were obtained from the *Drosophila* species stock center in Tuscon, Arizona. The four species of the *Drosophila virilis* group are wild type *Drosophila virilis* (stock # 15010 - 1051.00), *D. americana* (stock # 15010 – 0951.00), *D. novamexicana* (stock # 15010 - 1031.04) and *D. texana* (stock # 15010 – 1041.16).

Stocks were maintained in round (237 mL) polypropylene bottles containing standard cornmeal-molasses (CM) medium. All fly stocks were maintained in a controlled environment with a twelve-hour light/dark cycle at 20^{o}C.

2.2 Candidate gustatory receptor genes in *D.virilis*:

The genome of *Drosophila melanogaster* is well understood in terms of gene location and function. Recently, the entire genome of *D.virilis* was sequenced, but gene anotation has not been finished, so identifying the candidate X-linked gustatory receptor genes of *D.virilis* was the first task. I used data on *D. melanogaster* X-linked gustatory receptors (FLYBASE) to search, using the UCSC database (http://genome.ucsc.edu/), for *Drosophila virilis* genome sequences that aligned with the *Drosophila melanogaster* genes (Figure 2-1).

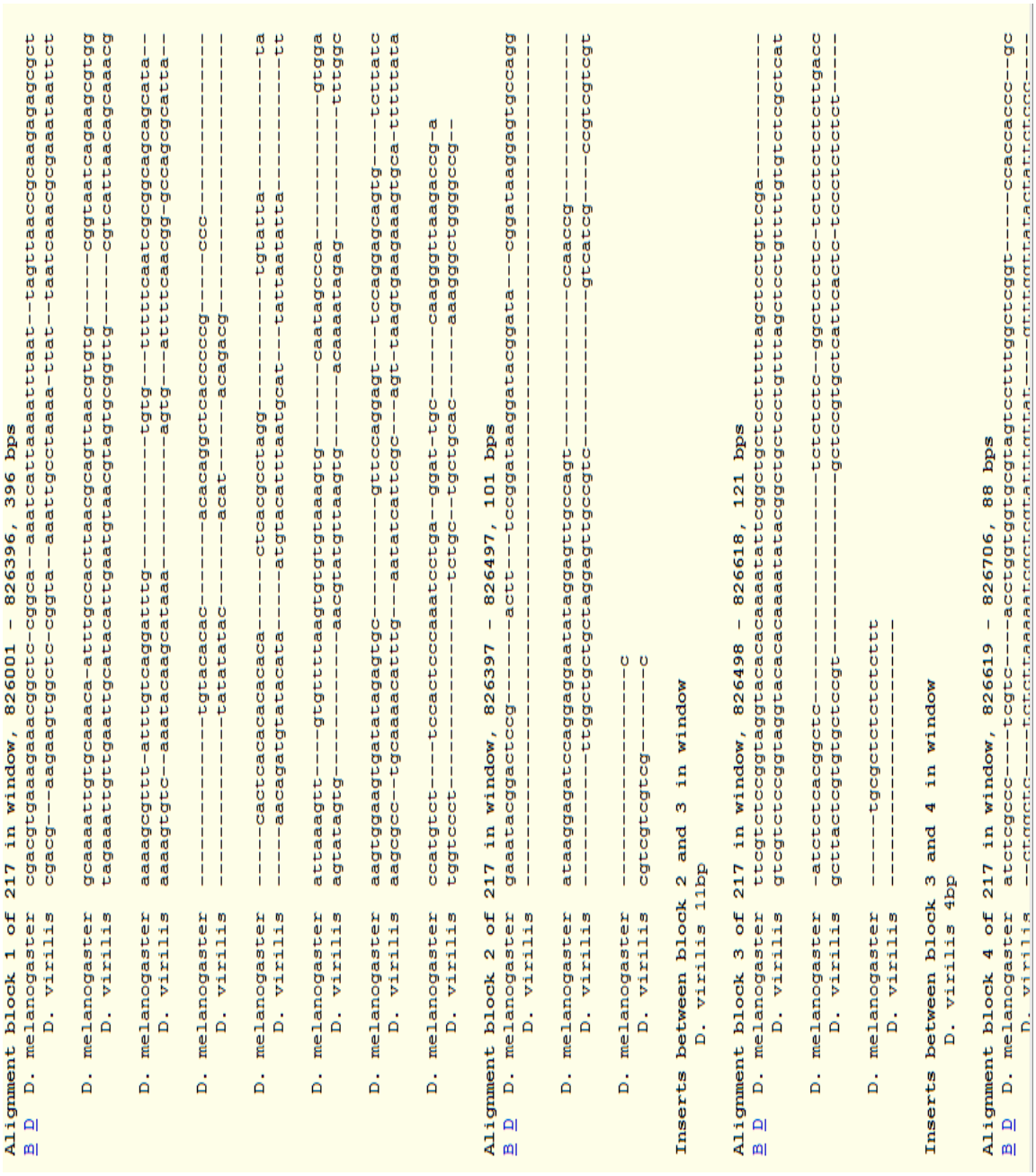

Figure 2-1. Alignment between the Gr2a gene of *D.melanogaster* and a similar sequence (candidate Gr2a gene) of *D. virilis* (http://genome.ucsc.edu/cgi-bin/hgc?hgsid_=104448765&o=826000&t=851000&g=multiz15way&i=multiz15way&c=chr2L&l=826000&r=851000&db=dm3&pix=620).

Note: the gaps (-) that are shared by both *D. melanogaster* and *D. virilis* are due to some other Drosophila species in the UCSC database having nucleotide insertions at these alignment positions.

2.3 Primer design:

The primers were designed using a computer program called Primer 3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). For each fly species there were two pairs of primers: one primer pair (named sequencing primers) was designed in such a way that they could cover as much sequence as possible of the gene of interest. Targets of the other primer pair (named internal primers) were located inside the gene of interest, the purpose of this primer pair was to guarantee we could obtain a complete sequence of the gene considering the fact that sequencing from one single primer can render about 600 base pairs of readable sequence. Information about the primers is summarized in Table 2-1. The primers were purchased from INVITROGEN.

Table 2-1. Primer description and its corresponding sequence (5' to 3')

Primer Description	**Sequence (5' to 3')**
Sequence Primer (SQ) *D.virilis* Gr2a_Forward primer	ATGGACATTTTGGACGCACT
SQ *D.virilis* Gr2a_Reverse primer	CCTCGCTCATGTGGAATTGT
SQ *D.virilis* Gr5a_Forward primer	TGCCTCCAATCGATTCTTCT
SQ *D.virilis* Gr5a_Reverse primer	CTTTTCATCCTCGTGGAATTG
SQ *D.virilis* Gr8a_Forward primer	GGCTTCATCTGCGCTTCTAT
SQ *D.virilis* Gr8a_Reverse primer	ATCCACTCAACTTGGGCATC
SQ *D.virilis* Gr9a_Forward primer	CGTCTACTTTCGGCTTTTGG
SQ *D.virilis* Gr9a_Reverse primer	GAAAGTGACCAGCGTGTTGA
SQ *D.virilis* Gr10a_Forward primer	GAGCGGCACGAGTACAAGTT

SQ *D.virilis* Gr10a_Reverse primer	TCGAATTGGACGAGCGTTAT
Internal *D.virilis* Gr2a_Forward primer	CGCCCTGCTGAATCGCTGCT
Internal *D.virilis* Gr2a_Reverse primer	GCCCGCTGCGCTGAAGTGTA
Internal *D.virilis* Gr5a_Forward primer	GACACGCTCTTCACCATCAA
Internal *D.virilis* Gr5a_Reverse primer	GATGTAGTTCCAGCCGAAGG
Internal *D.virilis* Gr8a_Forward primer	TTGCCAGCGTTGCAGCCAGT
Internal *D.virilis* Gr8a_Reverse primer	AGCTGGTAGGCGATCCGGGG
Internal *D.virilis* Gr9a_Forward primer	CGCCAGCTCTCCCGCCTCTA
Internal *D.virilis* Gr9a_Reverse primer	TGTATGCGTGTGAGTCCGGGT
Internal *D.virilis* Gr10a_Forward primer	CGGCCTCTATGCGTTCGGCT
Internal *D.virilis* Gr10a_Reverse primer	CGGTCCAGCTGAGCAGCCC

2.4 DNA extraction:

For each *Drosophila virilis* group species (*Drosophila virilis, D. americana*, *D.novamexicana* and *D.texana*), 7 flies were collected into a 1.5 mL micro-centrifuge sample tube. The flies were killed by being placed at - 20 ^{o}C for 20 minutes. 100 µL of homogenizing buffer (0.1 M TrisHCl, 0.1 M EDTA and 1% SDS) was added to each micro-centrifuge tube, and, the fly bodies were macerated using sterilized micro-pestles. Homogenized flies were incubated at 70 ^{o}C for 30 minutes. After this incubation, 14 µL of 8M potassium acetate was added, followed by incubation for 30 minute on ice. The sample was centrifuged at 18,300 *g* (RCF) and at 4 ^{o}C for 20 minutes. The supernatant was transferred to a new tube and 50 µL of 100% isopropanol was added followed by a 10 minute incubation at room temperature. DNA was pelleted by

centrifugation at 18,300 *g* (RCF) for 10 minutes at room temperature. The DNA pellet was washed twice with cold (previously stored at -20 ^{o}C) 70% ethanol. After drying, DNA was resuspended in 40 µL SDDW and stored at -20 ^{o}C.

2.5 Polymerase Chain Reaction (PCR):

Primers designed using *D.virilis* candidate Gr genes sequence information were used to isolate and amplify the same candidate X-linked gustatory receptor gene from *D. americana*, *D.novamexicana* and *D.texana*. First, PCR was conducted on *D. virilis* under different temperatures and $MgCl_2$ concentrations to test the primers and to identify the best conditions for PCR. After the PCR reaction was completed, gel electrophoresis (using 2% agarose gel) was used to verify that the expected gene had been amplified. The same PCR conditions used with *D. virilis* were used to isolate the candidate gustatory receptor genes from the other 3 species. The reagent aliquots used are shown in Table 2-2. The best thermal condition is summarized in Table 2-3.

Table 2-2. The chemical composition and concentration used in PCR reaction.

Reagent	**Volume (µL)**	**Final Concentration (or Amount)**
DNA	2	1 µg
Primer	0.6 + 0.6 (two primers)	0.1 to 1µM
10X PCR Buffer	2.5	1 to 4 mM of magnesium chloride
10X Taq polymerase	0.2	2 U/100µL
dNTPs	0.6	0.2 mM
SDDW	18.5	

Table 2-3. Temperature and cycling condition used in PCR reaction to amplify Gr genes.

Steps	Temperature (^{o}C)	Time length
1. Denature	95	5 minutes
2. Denature	95	45 seconds
3. Annealing	50	50 seconds
4. Extension	60	4 minutes
5. Repeat step 2 to 4 for 35 more cycles		~ 180 minutes
6. Final extension	60	10 minutes

2.6 Cleaning and quantifying PCR products:

After PCR, the PCR products were cleaned using "Wizard™ SV System" DNA clean kit (Promega cat # A9281) manufactured by Wizard™ Company. The purpose of cleaning PCR products is to eliminate the inorganic impurities, which could have bad effects on sequence quality as well as to remove primers, primer dimers and excess nucleotides from the PCR product. Each PCR product required a total volume of 1.2 mL of membrane wash solution (prepared by using one sixth volume of membrane wash solution from clean kit plus five sixths of 95% ethanol) and equal volume of membrane binding solution. The detailed procedures can be found in the company website (http://www.promega.com/applications/pcr/featuresandbenefits/Wizard_SV_Gel_PCR_Clean-Up_System.htm).

First, for clarification, the term "sequencing reaction" refers to the process in which the previously isolated and cleaned PCR products were used as templates for sequencing primers

binding and amplification. The process was conducted using a PCR cycler, not a DNA sequencer.

It is necessary to calculate the template aliquot required for the sequencing reaction based upon the following criteria:

1. the size of the PCR product;

2. the quantity of DNA in the cleaned product.

When designing primers, the sizes of all five gene products were estimated to be around 1250 base pairs (Gr2a-1362 bp, Gr5a-1338 bp. Gr8a-1280 bp, Gr9a-1093 bp, Gr10a-1186 bp). According to the template preparation sheet for the CEQ 2000 Dye Terminator Cycle Sequencing with Quick Start Kit BCI P/N 60818-AA, provided by Beckman Coulter, Inc, the recommended amount of double-stranded DNA required for the sequencing reaction is 25-100 fmol. In order to have 25-100 fmol of DNA, it was determined that no less than 20.25 ng and no more than 81.25 ng of ds DNA (approximately 1250 base pairs in size) should be used (see **bolded** in Table 2-4).

Table 2-4. Table for estimating the required amount (in ng) of double stranded DNA for sequencing, provided by Bechman Coulter, Inc.

Size (kilobase pairs)	ng for 25 fmol	ng for 100 fmol
0.5	8.1	33
1.0	16	65
1.25	**20.25**	**81.25**
2.0	33	130

To determine the volume of cleaned PCR product to be used in the sequencing reaction, the concentration of the cleaned PCR products must be known. For each cleaned PCR product, a 4 µL sample was loaded onto a gel and electrophoresed along with a standard low mass ladder. The band intensities of PCR products on the gel image were compared to the intensity of the standard low mass ladder bands, allowing the amount of DNA (in ng) in each PCR product to be estimated. Knowing the theoretical amount of DNA that should be used for the sequencing reaction and the quantity of cleaned PCR products available, I determined the proper volume of each sample to be added to the sequencing reaction mix. In this experiment, 4 µL samples were used in the sequencing reaction. For example, 4 µL of the PCR product of Gr2a gene isolated from *D. texana* was estimated to contain 57 ng DNA which is within the desirable quantity range of 20.25 to 81.25 ng.

2.7 DNA sequencing:

The sequencing reaction mix was prepared by running PCR using single sequencing primers. The reagent aliquots used for preparing the sequencing reaction mix are shown in Table 2-5. The thermal conditions are summarized in Table 2-6.

Table 2-5. The chemical composition and final concentrations used in the sequencing reaction.

Reagent	Volume (µL)	Final Concentration (or Amount)
DNA	4	25-100 fmol
Forward or Reverse Primer	1.0 (only one primer)	10µM (10pmol/µl)
Sequencing Buffer	1.2	
DCTS Quick start Master mix	2	
Pellet paint	0.25	
dNTPs	0.6	0.2 mM
SDDW	11.55	

Table 2-6. Temperature and cycling conditions used for the sequencing reaction.

Steps	Temperature (^{o}C)	Time length
1. Denature	95	5 minutes
2. Denature	95	45 seconds
3. Annealing	50	45 seconds
4. Extension	60	4 minutes
5. Repeat step 2 to 4 for 35 more cycles		~ 180 minutes
6. Final extension	60	8 minutes

Note: the thermal conditions of DNA PCR and DNA sequence reaction have certain similarities. However, the similarities between Table 2-3 and Table 2-6 do not imply that thermal condition of sequence reaction should be the same as PCR thermal condition used for isolation and amplification.

After Thermal Cycling, 5µl of freshly made Stop Solution [1.5M sodium acetate, 50mM EDTA and 1µl of 20 mg/ml glycogen], (1:1 mix from 3M sodium acetate and 100mM EDTA) was added to each sample to stop further cycling. The sample was then precipitated by using 95% cold ethanol and centrifuging at 18,300 *g* (RCF) for 15 minutes at 4 ^{o}C, and washed twice with 70% cold ethanol.

Samples were loaded into the CEQ 2000 XL sequencer to run sequencing. The detailed procedure for manipulating the CEQ 2000XL sequencer can be found in the manufacturer's website (http://innovation.swmed.edu/research/genomics/res_gen_beck.html).

2.8 Data analysis:

After retrieving the sequence data from the sequencer, the candidate gene sequences were assembled after proof reading the sequence trace files. The full sequences of each gene were aligned between the four species using the program Clustal X (ftp://ftp-igbmc.ustrasbg.fr/pub/ClustalX/) (Thompson *et al*., 1997). The nucleotide sequences were translated into amino acid sequences using online translate tool (http://www.expasy.ch/tools/dna.html). A program called "MotifScan" (http://myhits.isb- sib.ch/cgi-bin/motif_scan) (Thakallapally *et al*., 2000) was used to search for the presence of protein domains found in known gustatory receptor proteins. Sequence analyses in a phylogenetic context were performed using MEGA4 (www.megasoftware.net) (Tamura *et al*., 2007) and the CODEML software package in Phylogenetic analysis by maximum likelihood (PAML) (http://abacus.gene.ucl.ac.uk/software/paml.html#bugs) (Yang *et al*., 2000). The analysis was used to detect whether there is positive selection acting on a gene and whether the effect is specific to a given lineage. Comparison of likelihood estimates of a tree topology under model 7 (M7 = no positive

selection, $0 < d_N/d_S < 1$) with an alternative model 8 (M8 = positive selection, $d_N/d_S > 1$) can be used to detect whether there is positive selection acting on codon sites. Comparison of likelihood estimates of a tree topology under model 0, which assumes there is constant d_N/d_S ratio across all compared lineages, with model 1, which assumes that d_N/d_S ratio freely varies across lineages, were also conducted to determine whether the evolutionary rates in different species are constant (Wong *et al*., 2004).

3. RESULTS

3.1. *In silico* identification of candidate gustatory genes in *D. virilis*

Based on the alignment of known X-linked gustatory genes in *D. melanogaster* to similar sequences in the recently posted genome of *D. virilis*, I identified only Gr2a, Gr5a, Gr8a, Gr9a and Gr10a. As an example I show the hit of Gr10a *D.melanogaster* sequence to the *D. virilis* genome (Fig 3-1a). No homologous sequence was found to the *D. melanogaster* Gr10b gene (Figure 3-1b).

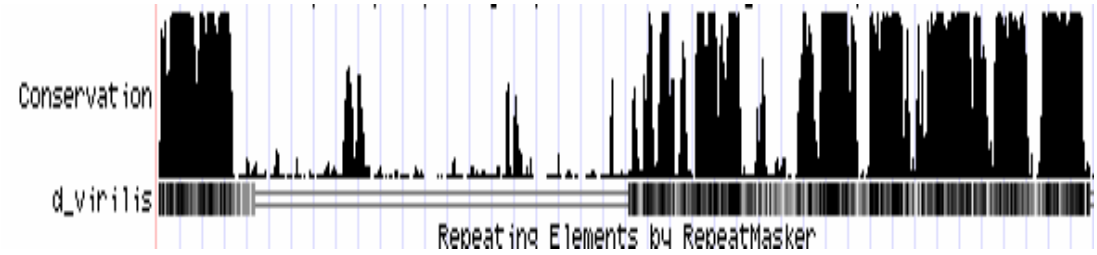

Figure 3-1a. The sequence of the *D.melanogaster* Gr10a gene aligned with homologous sequence in *D. virilis*, with a small portion of Gr10a gene of *D.melanogaster* showing no homologous sequence in *D. virilis* genome (two parallel lines). Conservation refers to the amount of sequence identity across all species aligned in the UCSC database.

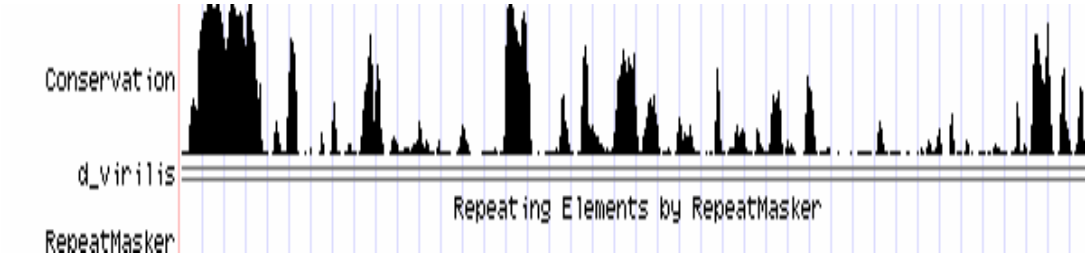

Figure 3-1b. The Gr10b gene of *D. melanogaster* has no homologous sequence in *D. virilis* (two parallel lines). Conservation refers to the amount of sequence identity across all species aligned in the UCSC database.

For the 5 identified candidate Gr genes in *D. virilis*, I found protein domains that are expected in chemosensory (gustatory) receptors (Figure 3-2).

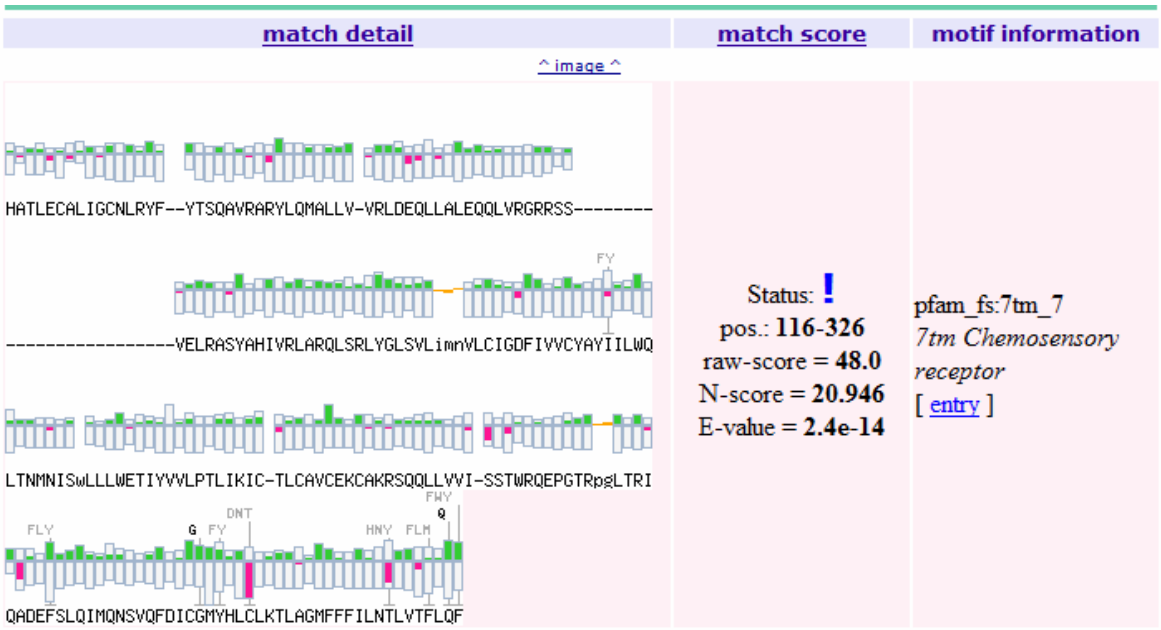

Figure 3-2. MotifScan result using the amino acid sequence of candidate Gr9a gene found in *D. virilis* (gustatory receptors belong to the superfamily called chemosensory receptors).

Using the same approaches, the candidate gustatory receptor genes (both Gr2a and Gr9a) isolated from *D. americana, D. novamexicana* and *D. texana* were found to show gustatory receptor protein domains.

3.2 DNA sequence analysis of candidate Gr genes among species of the *virilis* group

3.2.1 Gr2a

The candidate Gr2a gene was sequenced in *D. novamexicana, D. texana* and *D. virilis*, with only a partial sequence obtained for *D. americana*. The alignments of Gr2a DNA and amino

acid sequences are shown in Figures 3-3 and 3-4, respectively (in Appendix). The nucleotide changes in the candidate Gr2a genes between *D. americana, D. novamexicana, D. texana* and *D. virilis* are summarized in Table 3-1.

I conducted phylogenetic analysis by maximum likelihood (PAML) for comparisons among the three species (*D. novamexicana, D. texana* and *D. virilis*). *D. americana* was not included in the PAML analysis due to the fact that only a partial sequence was obtained, and the lack of sufficient sequence data would affect the accuracy of PAML tests (Yang, 1997). The CODEML software packaged in PAML was used to detect whether there is positive selection acting on a specific lineage or codon sites. Comparison of likelihood estimates of a tree topology using sequence data of Gr2a gene under model 7 (M7 = no positive selection, $0 < d_N/d_S < 1$) with an alternative model 8 (M8 = positive selection, $d_N/d_S > 1$) shows that M8 does not fit the data better than M7 (Table 3-2). That is, there is no evidence of positive selection at codon sites within the gene. I also compared of likelihood estimates of a tree topology under model 0, which assumes there is constant d_N/d_S ratio across lineages, with model 1, which assumes d_N/d_S ratio varies across lineages (Wong *et al*., 2004). For Gr2a, it shows there is significant variation in terms of evolutionary rate (Table 3-2). Figure 3-5 shows a three-species tree topology with branches leading from a common ancestor to each species representing the lineage specific rate of evolution (d_N/d_S).

Table 3-1. Summary of sequence changes in the candidate Gr2a genes between *D. novamexicana, D. texana* and *D. virilis*.

	d_N	d_S	d_N/d_S	**Interpretation (regarding positive selection)**
D. virilis –D. nova	0.015	0.101	0.099	No
D. virilis - D. texana	0.025	0.104	0.240	No
D. nova- D. texana	0.011	0.011	1.000	No (neutrality)

Table 3-2. Results from Phylogenetic analysis by maximum likelihood (PAML) Analysis of the Gr2a gene in comparison between *D. novamexicana, D. texana* and *D. virilis*.

Model	**Degree of freedom**	**Log likelihood**	**2 Δl^c**	**P-value (df)**	**Interpretation**
7	7	-1749.54	0	0.998 (2)	No positive selection
8	9	-1749.54			
0	6	-1749.53	10.94	0.012 (3)	Variable evolutionary rate
1	9	-1744.06			

2 Δl^c = likelihood ratio test, is two times the difference in log likelihood between two models that are compared. Significance is tested by using a Chi-Square test with degrees of freedom given by the difference between the numbers of independent parameter in each model (same rule applies to Table 3-4) (Yang Z. 1997).

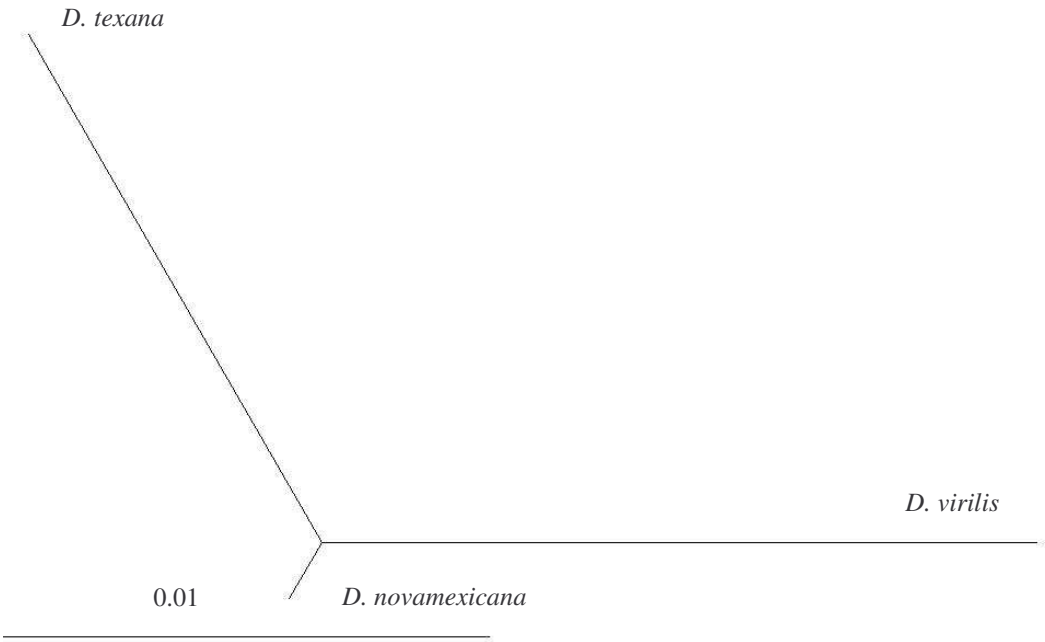

Figure 3-5. The proportion of nonsynonymous to synonymous substitutions (d_N/d_S) accumulated along the branch leading to *D. novamexicana, D. texana* and *D. virilis* from the common ancestor (shown in figure as the intersection of three species), based on the sequence data of Gr2a. It can be seen that for Gr2a, *D. texana* and *D. virilis* diverged more from the common ancestor relative to *D. novamexicana*. The distance stands for the amount of nonsynonymous to synonymous substitutions accumulated along each branch.

3.2.2 Gr9a

The candidate Gr9a gene was sequenced in *D. americana, D. novamexicana, D. texana* and *D. virilis*, with only a partial sequence obtained for *D. americana*. The alignments of Gr9a DNA and amino acid sequences are shown in figures 3-6 and 3-7, respectively (in Appendix). Nucleotide changes in the candidate Gr9a genes between *D. americana, D. novamexicana, D. texana* and *D. virilis* are summarized in Table 3-3.

PAML test was conducted for the three species (*D. novamexicana, D. texana* and *D. virilis*). *D. americana* was not included into the PAML analysis due to similar reason as in Gr2a.

Comparison of likelihood estimates of a tree topology using sequence data of Gr9a gene under model 7 with an alternative model 8 shows no evidence of positive selection at codon sites within the gene. For Gr9a, comparison of likelihood estimates of a tree topology under model 0 with model 1 shows the existence of significant variation in evolutionary rates (Table 3-4). A three-species tree topology with branches leading from a common ancestor to each species representing the lineage specific rate of evolution (d_N/d_S), based on the sequence data of Gr9a, is shown in Figure 3-8.

Table 3-3. Summary of sequence changes in the candidate Gr9a genes between *D. nov amexicana, D. texana* and *D. virilis*.

	d_N	**d_S**	**d_N/d_S**	**Interpretation (regarding positive selection)**
D. virilis –D. nova	0.045	0.060	0.750	No
D. virilis - D. texana	0.022	0.071	0.310	No
D. nova- D. texana	0.037	0.050	0.740	No

Table 3-4. Results from Phylogenetic analysis by maximum likelihood (PAML) Analysis of the Gr9a gene in comparison between *D. novamexicana, D. texana* and *D. virilis*.

Model	**Degree of freedom**	**Log likelihood**	**2 Δl^c**	**P-value (df)**	**Interpretation**
7	7	-977.91	0.06	0.970 (2)	No positive selection
8	9	-977.88			
0	6	-978.35	5.08	0.166 (3)	constant evolutionary rate
1	9	-975.81			

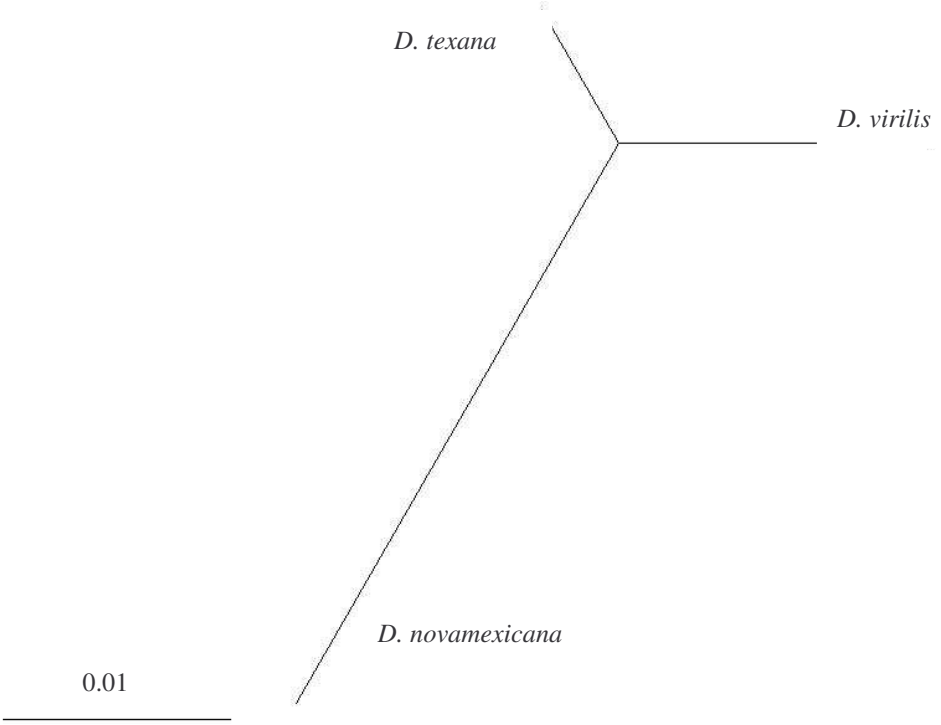

Figure 3-8. The proportion of nonsynonymous to synonymous substitutions (d_N/d_S) accumulated along the branch leading to *D. novamexicana, D. texana* and *D. virilis* from the common ancestor (shown in figure as the intersection of three species), based on the sequence data of Gr9a: it can be seen that *D. texana* and *D. virilis* diverged less from the common ancestor relative to *D. novamexicana*. The distance stands for the amount of nonsynonymous to synonymous substitutions accumulated along each branch.

3.2.3 Gr5a, Gr8a and Gr10a

Candidate gustatory receptor genes of Gr8a and Gr10a for *D. americana, D. novamexicana, D. texana* and *D. virilis* were isolated and amplified using PCR, and were sequenced. But the quality of Gr8a and Gr10a sequences were too poor to be analyzed. The candidate Gr5a genes could not be isolated from any of these four species, and DNA sequencing was not applied to Gr5a.

5. DISCUSSION

Functions of specific proteins are primarily determined by their structures, which are usually secondary, tertiary and quaternary structures formed by the primary structures. Primary structures, which are the amino acid sequences, are determined by the genetic information, the nucleotide sequences in the protein coding region of the gene(Voet *et al.*, 2004). By assessing the possible functions of the protein product using MotifScan program, we can interpret whether the changes at the nucleotide level are significant enough to cause phenotypic differences (gustatory receptors with different functions). The MotifScan results indicate that all isolated candidate X-linked Gr genes (Gr2a and Gr9a) from *D. americana, D. novamexicana, D. texana* and *D. virilis* still probably function as Gr genes as in *D. melanogaster*. Therefore, based on the data of Gr2a and Gr9a, we cannot conclude that the nucleotide changes cause significant changes at the protein level to eventually change their functions.

The ratio of non-synonymous to synonymous mutations (d_N/d_S) is a method that can be used to assess of natural selection acting on the protein (Yang *et al.*, 2000). The rationale of this test is that if synonymous mutations are essentially neutral because they do not result in amino acid replacement, then the rate of synonymous site evolution (d_S) will equal the mutation rate. Nonsynonymous mutations that result in change of amino acid sequences are more likely to be under the effect of selection. Thus, in situations where such mutations are deleterious the ratio of d_N/d_S is predicted to be significantly lower than 1, and to reflect the effect of purifying selection. In contrast, if nonsynonymous mutations are maintained in

populations by positive selection, then the d_N/d_S ratio is predicted to be significantly larger than 1. Simplistically, values of this ratio smaller than 1 means negative purifying selection, the ratio equal to 1 implies neutral evolution, and a value larger than 1 implies positive selection (Nielsen and Yang, 1998; Yang *et al.*, 2000). The calculated average value of d_N/d_S suggested that there was no positive selection among Gr2a and Gr9a genes of *D. americana*, *D. novamexicana* and *D. virilis*. (Table 3-1 and Table 3-3, respectively).

However, the ratio averaged over all sites and all lineages is almost never larger than 1, since positive selection is unlikely to have affected all sites over a long time. Thus, interest has been focused on detecting positive selection that affects only some lineages (in this case, only *D. virilis*) or some sites (Yang, 1997; Yang, 2007). In this research, the CODEML software package in phylogenetic analysis by maximum likelihood (PAML) was also used to detect whether there was positive selection acting on a specific lineage or codon sites. Because premating isolation was only observed from *D.virilis* males towards other heterospecific females, and not between males and females of other species, positive selection was expected to act only on *D. virilis*. Instead of directly calculating the average value of d_N/d_S, PAML estimates the likelihood under different models of evolution that make alternative assumptions about the d_N/d_S ratio over codon sites and lineages. For both Gr2a and Gr9a, I detected no evidence of positive selection at specific codon sites. Results from M7/M8 further depicted that no significant positive selection has happened on codon sites within Gr9a gene. Generally, this rejects the possibility of Gr9a being responsible for the observed premating isolation. Only Gr2a showed evidence of significant variation in rates of evolution across lineages, but the rate of evolution was not the highest or slowest in the *virilis* branch. It is possible that the

lack of any signal for Gr9a is linked to the fact that the last portion of this gene sequence (see figure 3-5, 3-7 written in red) were completely different between *D. novamexicana* and *D. virilis*, but this portion of the sequence was not include in the PAML analysis (only 2 species sequences available). These facts create difficulty in explaining the result. To have a clear conclusion, more data are required.

The phylogenic trees built based on the data of Gr2a and Gr9a (Figure 3-5 and Figure 3-8, respectively) show that the Gr2a gene in *D.virilis* and *D. texana* have accumulated a higher proportion of d_N/d_S change from the common ancestor compared to *D. novamexicana*; while Gr9a gene in *D. virilis* and *D. texana* have accumulated a lower proportion of d_N/d_S change from the common ancestor compared to *D. novamexicana*. Based on the data of Gr2a and Gr9a, there is no evidence showing that either of these genes can explain the premating isolation between *D. virilis* and close-related heterospecific females. While it is possible that none of these X-linked gustatory receptor genes (Gr2a, Gr9a) account for the premating isolation, it is also possible that the premating isolation is the combined result of divergence at two or more X-linked gustatory receptor genes.

To understand this second interpretation, let us examine Figure 3-5 and 3-8 again. In Figure 3-5, the result can be interpreted as the Gr2a gene contributing to the sexual isolation between *D. virilis* and *D. texana*, but not between *D. virilis* and *D. novamexicana*. Because *D. virilis* and *D. texana* diverged a lot relative to each other, but *D. virilis* and *D. novamexicana* did not diverge as much relative to each other. Similarly figure 3-8 can be interpreted as the Gr9a gene contributes to the sexual isolation between *D. virilis* and *D. novamexicana*, but not between *D. virilis* and *D. texana*. The combined effect of divergence at both Gr2a and Gr9a

genes can explain *D. virilis* sexual isolation. Obviously, more data is needed in order to derive a clear conclusion. It is possible that none of the 2 Gr genes analyzed have any role in premating isolation between *D. virilis* males and heterospecific females.

Why was Gr5a gene not isolated and amplified in any species?

This may due to the thermal and chemical conditions of the PCR reaction. We know that for a primer to work properly, the concentration of Mg^{2+}can affect the annealing of the primer to the template DNA by stabilizing the primer-template interaction; it also stabilizes the replication complex of polymerase with template-primer. It can therefore also increase non-specific annealing and produce undesirable PCR products (multiple bands on the gel) if the concentration of Mg^{2+} is too high. Temperature (annealing temperature, extension temperature), is another important variable affecting primer binding. Usually, the thermal profiles of primers are provided by the manufacturer. But in this research, the primers were designed from the *D. virilis* genome, and were allowed to work on the extracted DNAs of *D. americana, D. novamexicana, D. texana*. The optimal temperature condition may change for these heterospecific DNA's(Griffiths et al., 2005). However, due to the fact that survey experiments were extensively conducted by changing the temperature profile and Mg^{2+} concentrations, it is not likely that the failure to get a PCR product of Gr5a was caused by technical problems. Gr5a codes for gustatory receptors that are responsible for detecting sugars and appears to be activated specifically by trehalose (Fischler *et al*., 2007). It is also known that among the superfamily of gustatory genes, Gr5a is one of the most polymorphic Gr genes at least in *D. melanogaster* (Ueno *et al.*, 2001), meaning that there are usually many

nucleotide variants in Gr5a between individuals of the same species. This provides the most likely explanation for the failure to amplify Gr5a genes from all four species: if the primers I designed target sequences that can be allelic between individuals, then it is possible that the primers would not bind to the template DNA of a specific strain leading to failure of the PCR reaction.

5. CONCLUSIONS

- There are nucleotide differences (insertion/deletion, substitution) between *D. americana, D. novamexicana, D. texana* and *D. virilis* for Gr2a and Gr9a;
- Candidate genes (Gr2a, Gr9a) of all four *virilis* group species are likely to function as gustatory receptor genes, since their protein products all have the functional domain of typical gustatory receptors.
- Phylogenetic analysis by maximum likelihood (PAML) suggests there is no evidence of positive selection specific to *D. virilis*.
- It is possible that two or more gustatory receptor genes work together to account for *D. virlis* male premating isolation towards closely-related heterospecific females.

6. FUTURE STUDY

- Complete Gr8a, Gr10a sequencing and assembly.
- Sequence Gr5a in *virilis* group species (redesign primers for Gr5a).
- Gene expression assays

7. REFERENCES

- Adams, M. D., Celniker S, E., and Holt R, A., *et al*. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287 (5461): 2185–95.
- Bartelt, N. C., and T. L. Einstein. 1986. Comment on ‘‘Reliability of low-energy electron diffraction for studies of surface order-disorder phenomena’’. *Phys. Rev. Lett*. 56: 2881 – 2881.
- Bryan C. Caletka and Bryant F. McAllister. 2004. A genealogical view of chromosomal evolution and species delimitation in the *Drosophila virilis* species subgroup. *Molecular Phylogenetics and Evolution.* 3*:* Issue 3, 664-670.
- Buckley, S. H., T. Tregenza, and R. K. Butlin. 1887. Speciation and sinmal train genetics. *Trends in ecology & evolution*. 12: 299-301.
- Clyne, P. J., C. G. Warr., and J. R. Carlson., 2000. Candidate taste receptors in Drosophila. *Science* 287: 1830-1834.
- Coyne, J. A., A. P. Crittenden., K. Mah., 1994. Genetics of a pheromonal difference contributing to reproductive isolation in Drosophila. *Science* 265: 1461–1464.
- Coyne, J.A. and H.A. Orr. 1989. Patterns of speciation in *Drosophila*. *Evolution*. 43: 362-391.
- Coyne, J.A. and H.A. Orr. 2004. *Speciation.* Sinauer Associates, Sunderland, MA
- Cobb. M., B. Burnet., R. Blizard and J. M. Jallon. 1990. Altered mating behaviour in a Carsonian population of Drosophila sechellia. *Evolution*. 44:2057-2068.
- Dobzhansky, Th. 1937. Genetics and the Origin of Species. Columbia University Press, New York.

- Ferveur, J. F., M. Cobb., H. Boukella., and J.M. Jallon., 1996. World-wide variation in Drosophila melanogaster sex pheromone: behavioural effects, genetic bases and potential evolutionary consequences. *Genetic*, 97: 73–80.
- Fischler, W., P. Kong, S. Marella and K. Scott. 2007. The detection of carbonation by the *Drosophila* gustatory system. *Nature* 448: 1054-1057.
- Futuyma, D.J. 1986. *Evolutionary Biology*. Sunderland, Mass: Sinauer Associates, Inc.
- Griffiths et al., 2005. *Introduction to Genetic Analysis*. 8th Edition. W. H. Freeman and Company. Chapter 11, page 360-364.
- Harrison D. Stalker. 1942. The Inheritance of a Subspecific Character in the Virilis Complex of Drosophila. *The American Naturalist*, 76: 426-431.
- Jallon, J. M. 1984. A few chemical words exchanged by Drosophila during courtship and mating. *Behav. Genet*. 14.441-476.
- Kaori, Ishii., H. Yoshiyuki., C. Katagiri., M. T. Kimura., 2001. Sexual isolation and cuticular hydrocarbons in *Drosophila elegans. Nature*, 98: 392-399.
- Kulikov, A. M., A. I. Melnikov., N. G. Gornostaev., O. E. Lazebny., V. G. Mitrofanov., 2004. Morphometriv analysisof males genitalia in sibling species of *Drosophila virilis.* Sturt. Russ. *J. Genet*. 40: 125-138.
- Liimatainen. J., and J. M. Jallon., 2007. Genetic Analysis of Cuticular Hydrocarbons and Their Effect on Courtship in Drosophila virilis and D. lummei. *Behav Genet*: 17557202.
- Markow, T. A. and P. M. O'Grady (2006). *Drosophila*: A guide to species identification and use. London, UK, Elsevier Inc.

- Natasha T., B. Steve and H. Amrein. 2005. Function and expression of the Drosophila Gr genes in the perception of sweet, bitter and pheromone compounds. *Chem. Senses*: 30: i270-i272.
- Nayak, S.V. and R.N. Singh., 1983. Sensilla on the tarsal segments and mouthparts of adult a
- Nickel. D., 2007. You're not my kind of girl: The genetics of premating isolation in the *Drosophila virilis* group. B.Sc. Honours Thesis. Department of Biology, Universty of Winnipeg.
- Ortiz-Barrientos. D., B. A. Counterman., M. A. F. Noor,. 2004. The Genetics of Speciation by Reinforcement. *PLoS Biol* 2(12): e416.
- Patterson, J., W. Stone. 1952. Evolution in the Genus Drosophila. New York: Macmillan Co.
- Robertson, H.M., C.G. Warr., and J.R. Carlson., 2003. Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. Proceedings of the National Academy of Sciences USA 100, pp. 14537–14542.
- Rosenzweig, M.L. 1995. Species Diversity in Space and Time. Cambridge: *Cambridge University Press*.
- Scott, K., et al. 2001. A chemosensory gene family encoding candidate gustatory and olfactory receptors in Drosophila. *Cell* 104: 661-673.
- Spicer, G. S., C. D. Bell., 2002. Molecular phylogeny of the Drosophila virilis species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. *Ann Entomol Soc Am*. 95: 156-161.
- Stalker. H. D., 1942. Sexual isolation studies in the species complex Drosophila virilis. *Genetics*. 27 (2):238-57.

- Stocker. R. F., 1994. The organization of the chemosensory system of Drosophila melanogaster: a review. *Cell Tissue Res*. 175: 3-26.
- Thakallapally, R., W. Kibbe., D. Lang., and B. Korber.2000. Motifscan: A Web-based Tool to Find HLA Anchor Residues in Proteins or Peptides. Northwestern University, Chicago. http://www.hiv.lanl.gov/content /immunol- og/pdf/2000/1/Thakallapally.pdf.
- Thompson, J. D., Gibson, R. J., Plewniak, F., Jeanmougin, F., and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 24:4876-4882.
- Throckmorton, L. H., 1982. The virilis species group. In: Ashburner, M., Carson, H. L., Thompson, J. N. Jr. Jr. (Eds.), The genetics and biology of Drosophila. *Academic press*, New York. 3: 227-296.
- Ueno, K., M. Ohta, H. Morita, Y. Mikuni, S. Nakajima, K. Yamamoto, K. Isono.2001. Trehalose sensitivity in Drosophila correlates with mutations in and expression of the gustatory receptor gene Gr5a. *Current Biology* 11: 1451-1455.
- Voet. D., F. G. Voet., and C. W. Pratt. 2004. Fundamentals of Biochemistry, Second Edition. John Wiley & Sons, Inc. NJ. Page: 40-197.
- Wessells, N.K. and J. L. Hopson. 1988. *Biology*. New York: Random House. Chapter 43.
- Wong, W. S,. Z. Yang, N. Goldman, and R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*. 168: 1041-1051.
- Yang, Z., and R. Nielsen. 1998. Synonymous and nonsymonymous rate variation in nuclear genes of mammals. *J. Mol. Evol*. 46: 409-418.

- Yang, Z., R. Nielsen., N. Goldman, and A-M. K. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid *sites. Genetics. 155: 431-449.*
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* Oct; 13(5) 555-6. pmid:9367129.
- Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*; 24: 1586-1591.

APPENDIX

```
D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      --------------------------------ACTGCANCGTATCGGCCAGCTGTGCAAC
D.tex-Gr2a_nuc      -------------------------------------GTATCGGCCAGCTGTGCAAC
D.vir-Gr2a_nuc      TTTGGCATGGACATTTTGGACGCACTGGGGCCACTGCAGCGTATCGGCCAGCTGTGCAAC

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      CTGTGGCAGTGGCGCTTGGATACTGCCACACGGCAGCTGCAACGTTCCCGCTGGCTGGAG
D.tex-Gr2a_nuc      CTGTGGCAGTGGCGCTTGGATACTGCCACACGGCAGCTGCAACGTTCCCGCTGGCTGGAG
D.vir-Gr2a_nuc      CTGTGGCAGTGGCGCTTGGATACTGCCACACGGCAGCTGCAACGTTCCCGCTGGCTGGAG

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      GCCTACAGCTGTGTGGTGCTGCTGCTGTCCACGGGCTATCTTCTATACGGCTTGTTTAAT
D.tex-Gr2a_nuc      GCCTACAGCTGTGTGGTGCTGCTGCTGTCCACGGGCTATCTTCTATACGGCTTGTTTAAC
D.vir-Gr2a_nuc      GCCTACGGCTGTGTGGTGCTGCTACTGTCCACGGGTTACCTTCTGTACGGCTTGTTTGAC

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      GAGAGTCGCACTGCCGATGGCGAGTCCAATGAAACCACCAACATTGGCCACACGGTAGAC
D.tex-Gr2a_nuc      GAGAGTCGCACTGCCGATGGCGAGTCCAATGAAACCACCAACATTGGCCACACGGTAGAC
D.vir-Gr2a_nuc      GAGAGTCGCACCGCCGATGGCGAGTCCAATGAAACCACCAACATTGGCCACACGGTAGAC

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      TCCATACAGTTGGTGGGCATACGTGTGGCTCACGCAACGGCGCTTCTGGAGGCGCTGTTG
D.tex-Gr2a_nuc      TCCATACAGTTGGTGGGCATACGTGTGGCTCACGCAACGGCGCTTCTGGAGGCGCTGTTG
D.vir-Gr2a_nuc      TCCATACAGCTGGTGGGCATACGTGTGGCTCACGCAACGGCGCTTCTGGAGGCGCTATTG

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      CAGCGTCAAGCGCAGGGCCGCTTCTTCGCGCAGCTGCGCGAGATCGACCGTGAATTCGCG
D.tex-Gr2a_nuc      CAGCGTCAAGCGCAGGGCCGCTTCTTCGCGCAGCTGCGCGAGATCGACCGTGAATCCGCG
D.vir-Gr2a_nuc      CAGCGTCAAGCGCAGGGCCGCTTCTTTGCGCAGCTGCGCGAGATCGACCGCGAATTCGCG

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      AATGTGCTGCATCTGGATCTGGAGAATGGCCAATTGCGAAGTCAAATGAACCGCCGAGCA
D.tex-Gr2a_nuc      AATGTGCTGCATCTGGATCTGGAGAATGGCCAATTGCGAAGTCAAATGAACCGCCGAGCG
D.vir-Gr2a_nuc      AATGTGCTGCATCTGGACCTGGAGAATGGCCAATTCCGAAGTCAAATGAACCGCCGAGCA

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      CTGTGGATGCTGTCCGGCTATTTGGTTAGCCAGCTCNTCATATTGTTTACCAAAATGCTN
D.tex-Gr2a_nuc      CTGTGGATGCTGTCCGGCTATATGGTTAGCCAGCTCTTCATATTGTCTACCAAAATGCTT
D.vir-Gr2a_nuc      CTGTGGATGCTATCCGGCTATTTGGTTAGCCAGATCTTCATACTGTTTACCAAAATGCTT
```

```
D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      GCGCCGGATAGCCACNTTATGACCTATGGGCTGTGCTATTTACTGCCGCTGCTGGTCTGT
D.tex-Gr2a_nuc      GCGCCGGATAGCCACTTTATGACCTATGGGCTGTGCTATGTACTGCCGCTGCTGGTCTGT
D.vir-Gr2a_nuc      GCGCCGGATAGCCACTTTATGATCTATTGGCTGTGCTATTTGCTGCCGCTGCTGGTCTGT

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      GGCGTGCGCTATTTTCAGATCTACACGGCCGTNCAGATTGTGCACCACCGTCTGGAGATG
D.tex-Gr2a_nuc      GGCGTGCGCTATTTTCAGATCTACACGGCCGTNCAGATTGTGCACCACCGTCTGGAGATG
D.vir-Gr2a_nuc      GGCGTGCGCTATTTTCAGATCTACACGGCCGTCCAGATTGTGCACCACCGTCTGGAGATG

D.ame-Gr2a_nuc      ------------------------------------------------------------
D.nov-Gr2a_nuc      ATGCTGCGCTTGCTGGAGTNCCTGCAGCTGAACGAGGCGGAGCAGCCGCAGCCATTTGTG
D.tex-Gr2a_nuc      ATGCTGCGCTTGCTGGAGTCCCTGCAGCTGAACGAGGCGGAGCAGCGGCAGCCATTTGTG
D.vir-Gr2a_nuc      CTGCTGCGCTTGCTGGAGTCCCTGCAGT---------CGGAGCAGCCGCAGCCATTTGTC

D.ame-Gr2a_nuc      ----------------------------------------------TTGCATGCAGCGT
D.nov-Gr2a_nuc      TGGNAGTCCATAAAGGAGGAGCTCGAAATGGAGGCGCTGGCGGCGGCTTGCATGCAGCGT
D.tex-Gr2a_nuc      TGCAAGTCCATAAAGGAGGAGCTCGAAATGGAGGCGCTGGCGGCGGCTTGCATGCAGCGT
D.vir-Gr2a_nuc      TGCAAGTCCATAAAGGAGGAGCTCGAAATGGAACGCTGGCGGCGGCTTGCATGCAGCGT

D.ame-Gr2a_nuc      CTGTCGNCCGCCAGAAAATTGTATCAGCGCCTCTGGCTATNGCTCGCCCTGCTGAATCGC
D.nov-Gr2a_nuc      CTGTTGGCCGCCAGAAAATTGTATCAGCGCCTCTGGCTATTGGTCGCCCTGCTGAATCGC
D.tex-Gr2a_nuc      CTGTTGGCCGCCAGAANATTGTATCAGCGCCTCTGGCTATTGGTCGCCCTGCTGAATCGC
D.vir-Gr2a_nuc      CTGTTGGCTGCCAGAAAATTGTATCAGCGCCTCTGGCTGTTGGTCGCCCTGCTGAATCGC

D.ame-Gr2a_nuc      TGTTATGGCCTCTCCATNCTGGTCAACTTGGGCAACGACTTTCTGGCCATTACCTCCAAC
D.nov-Gr2a_nuc      TGTTATGGCCTCTCCATACTGGTCAACTTGGGCAACGACTTTCTGGCCATCACCTCCAAC
D.tex-Gr2a_nuc      TGTTATGGCCTCTCCATACTGGTCTACTTGGGCAACGACTTTCTGGCCATCACCTCCAAC
D.vir-Gr2a_nuc      TGCTATGGCCTCTCCATGCTGGTCAACTTGGGCAACGACTTTCTGGCCATCACCTCCAAC

D.ame-Gr2a_nuc      TGCTATTGGATCTTCTTGAACTTTCGCCAGTTTGCCGCCTCGCCCTACGATTTCCTGCAA
D.nov-Gr2a_nuc      TGCTATTGGATCTTCTTGAACTTTCGCCAGTTTGCCGCCTCGCCCTACGATTTCCTGCAA
D.tex-Gr2a_nuc      TGCTATTGGATCTTCTTGAACTTTCGCCAGTTTGCCGCCTCGCCCTACGATTTCCTGCAA
D.vir-Gr2a_nuc      TGTTATTGGATCTTCTTGAACTTTCGCCGGTTTGCCGCCTCGCCCTACGATTTCCTGCAA

D.ame-Gr2a_nuc      ATTGTAGGCAGCACCCTTTGGTCTGCACCGCATTTGGGCAACGTTCTTTTTCTGGCTCTT
D.nov-Gr2a_nuc      ATTGTAGGCAGCACCCTTTGGTCTGCACCGCATTTGGGCAACGTTCTTTTTCTGGCTCTT
D.tex-Gr2a_nuc      ATTGTAGGCAGCACCCTTTGGTCTGCACCGCATTTGGGCAACGTTCTTTTTCTGGCTCTT
D.vir-Gr2a_nuc      ATTGTGGGCAGCACCCTTTGGTCTGCACCGCATTTGGGCAACGTTCTTTTTCTGGCTCTC

D.ame-Gr2a_nuc      ATATGCGAACGAGCCGCTTATTTTACAACGCGGCTTGCCTTGAGTCTGCATCACATCCGC
D.nov-Gr2a_nuc      ATATGCGAACGAGCCGCTTATTTTACAACGCGGCTTGCCTTGAGTCTGCATCACATCCGC
D.tex-Gr2a_nuc      ATATGCGAACGAGCCGCTTATTTTACAACGCGGCTTGCCTTGAGTCTGCATCATATCCGC
D.vir-Gr2a_nuc      ATATGCGAACGAGCCGCTTATTTTACAACGCGGCTTGCCTTGAGTCTGCATCACATCCGC
```

```
D.ame-Gr2a_nuc      GTGGATCTCCAGAATGACAGCCACAATGCACTGATCACCCAGTTCTCCCTGCAGCTGCTG
D.nov-Gr2a_nuc      GTGGATCTCCAGAATGACAGCCACAATGCACTGATCACCCAGTTCTCCCTGCAGCTGCTG
D.tex-Gr2a_nuc      GTGGATCTCCAGAATGACAGCCACAATGCACTGATCACCCAGTTCTCCCTGCAGCTGCTG
D.vir-Gr2a_nuc      GTCGATCTCCAGAATGACAGCCACAATGCACTGATCACCCAGTTCTCCCTGCAGCTGCTG

D.ame-Gr2a_nuc      CATCAGCGACTACACTTCAGCGCAGCGGGCTTTTTTGATGTGGACTGCACGTTGCTTTAC
D.nov-Gr2a_nuc      CATCAGCGACTACACTTCAGCGCAGCGGGCTTTTTTGATGTGGACTGCACGTTGCTTTAC
D.tex-Gr2a_nuc      CATCAGCGACTACACTTCAGCGCAGCGGGCTTNTTTGATGTGGACTGCACGTCGCTTTAC
D.vir-Gr2a_nuc      CATCAGAGACTACACTTCAGCGCAGCGGGCTTTTTCGATGTGGACTGCACGTTGCTTTAC

D.ame-Gr2a_nuc      ACGATTGTTGGTGCCACAACCACGTATCTCATCATTTTGATACAATTTCACATGA-----
D.nov-Gr2a_nuc      ACTATTGTTGGTGCCACAACCACGTATCTCATCATTTNGATACAATTCCACATGAGC---
D.tex-Gr2a_nuc      ACTATTGTTGGTGCCACAACCACGTATCTCATCATTTTGATACAATCCCACATGAGGGAG
D.vir-Gr2a_nuc      ACAATTGTTGGTGCCACAACCACGTATCTCATCATTTTGATACAATTCCACATGAGCGAG

D.ame-Gr2a_nuc      ---------------------------------
D.nov-Gr2a_nuc      ---------------------------------
D.tex-Gr2a_nuc      GA-------------------------------
D.vir-Gr2a_nuc      GGCAGCTTGAGCAGCGGTACGGCCGGGGGATAG
```

Figure 3-3. Multiple sequence alignments of candidate Gr2a genes from *D. americana, D. novamexicana, D. texana* and *D. virilis* with nucleotide substitutions highlighted with yellow and nucleotide insertion/deletion with green.

Note: the symbol “N” stands for “non-readable”, due to multiply signals with similar strength for one nucleotide. The same rule applies to Figure 4-5.

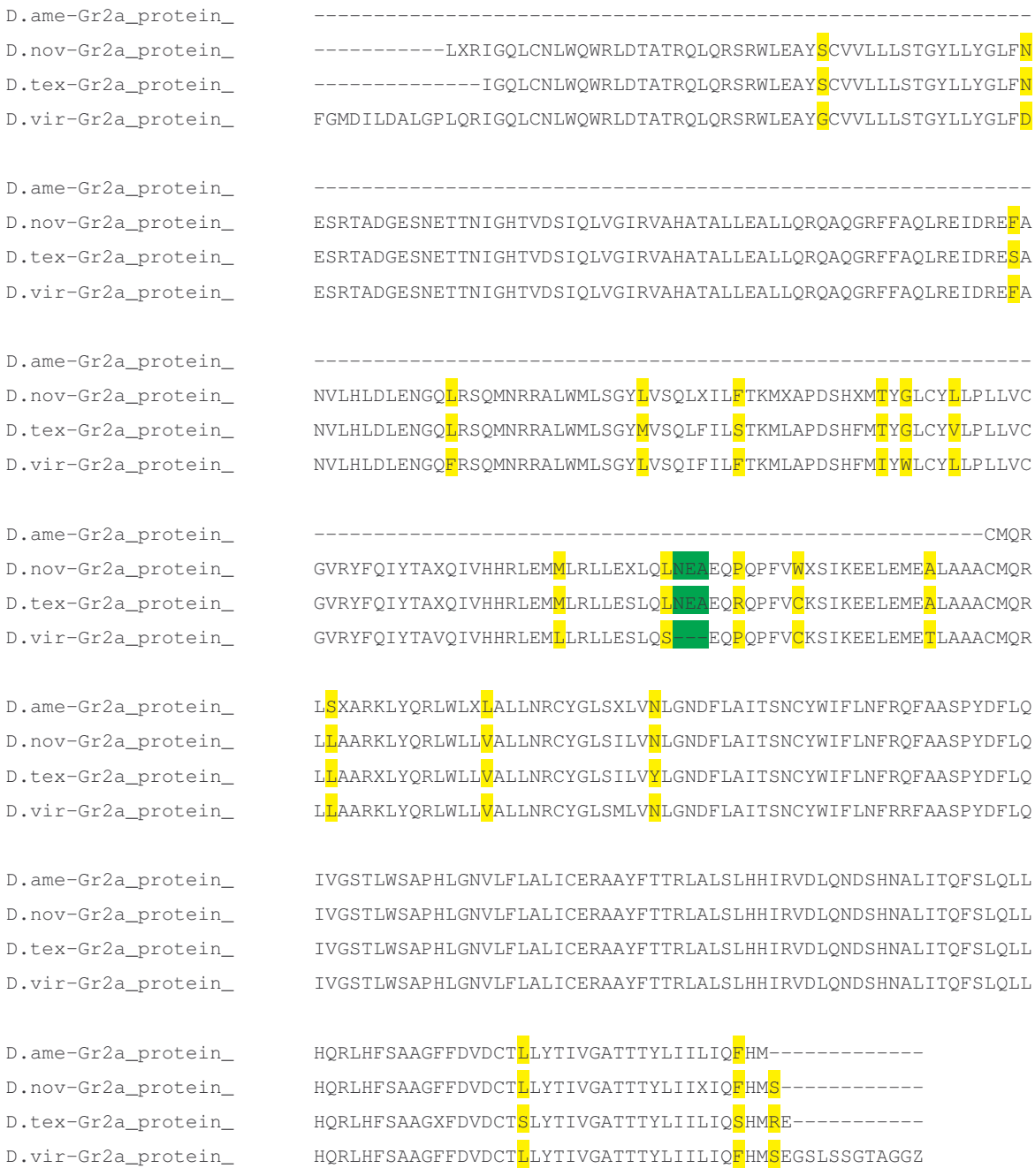

Figure 3-4. The alignment of amino acid products of candidate Gr2a genes for *D. americana, D. novamexicana, D. texana* and *D. virilis* with amino acid substitutions highlighted with yellow paint and amino acid insertion/deletion with green paint.

Note: the symbol “X” means amino acid which is unidentifiable due to the presence of “N” in its codon. The same rule applies to Figure 4-6.

D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
GGAACTCTTTCTGGCCGTCTACTTTCGGCTTTTGGCGCTCACGTGCAGCTCACGGAAGCG
D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
GGCGGTCCAGCGGGTGCTGAGCGGCTACCTGGCGCTGATCCTGGGCGATATGGTAATACA
GACGGTCCAGCGGGTGCTGAGCGGCTATCTAGCGCTAATCCTGGGCGATATGGTAATACA
D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
TGCAGGAGCCGCAGGGGCA
GTTNNGGAATTACTTCGGCAACGGCTGGCATATGACCAACTGGCAGCAGCCGCAGCGGCA
GTTTTGGCATTACTTCGGCAACGGCTGGCAGATGACCAACCTGCAGCAGCCGCAGCGGCA
D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
TCAGTTCAATGCCTATGACAAGATCATCGATGCCATGAGTGNGNCNNATNNGTGGGCCCA
GCAGTTCAATGCCTATGACAAGATCATCGATGCCATGAGTGTGGCCCATCTGCTGGCCCA
ACAGTTCAACGCCTATGCAAAGATCATTGATGCCATGAGTGTGGCCCATCTGCTGGCCCA
D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
TATGCTGATGGTGCTGCACGCGCTTGTGTGTCGGTATCGGGAGCGCAGGCTGTTGGCCCA
TATGCTGATGGTGCTGCAGGCGCTTGTGTTGCGGAATCGTGAGCGCAGGCTGTTGGCCCA
TATGCTGATGGTGCTGCAGGCGCTTGTGTTGCGGAATCGGGAGCGCAGGCTGTTGGCCCA
D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
GCTGCCGACAGTGGGACCGCCGTCAACATTGATCAACGTGCGTCTCCATGCGACGCTAGA
GCTGCCGACAGTGGGACCGCCGTCCACATGGATCAACGTGCGTCTCCATGCGACGCTAGA
GCTGCCGACAATGGGACCCCCGTCCACATTGATCAACGTGCGCCTCCATGCGACGCTGGA
D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
GTGCGCCCTAATCGGCTGCGCTCTCACAATGGGCACATACCAGTCGCTGCAAAGCCATCA
GTGCGCCCTAATCGGCTGCGCTTTCACAATGGGCACCTACCAGTCGCTGCAAAGCCATCA
GTGCGCCCTAATCGGCTGC
D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
CAACTTGCGCTACTTCTACACCTCGCAGGCTGTTCGGGCGCCTAATCTGCA
GCTGATGCCCGACTTGCGCTACTTCTACACCTCGCAGGCTGTTCGGGCGCGTTATCTGCA
GCTGATGCCCAACTTGCGTNACTTCTACACCTCGCAGGCTGTTCGGGCGCGTTATCTGCA
AACTTGCGCTACTTCTACACCTCGCAGGCTGTTCGGGCGCGTTATCTGCA
D.amer-Gr9a_nuc
D.nov-Gr9a_nuc
D.tex-Gr9a_nuc
D.vir-Gr9a_nuc
GATGACACTGCTTGTTGTCCGGCTGGATGGGCAGCTGCTGGCACTGGAGCAACAGCTGGT
GATGACACTGCTTGTTGTCCGGCTGGATGGGCAGCTGCTGGCACTGGAGCAACAGCTGGT
GATGACACTGCTTGTTGTCCGGCTAGATGGGCAGCTGCTGGCACTGGAGCAACAGCTGGT
GATGGCACTGCTTGTTGTCCGGCTGGATGAGCAGCTGCTGGCACTGGAGCAACAGCTGGT

```
D.amer-Gr9a_nuc     GAGGGGACGACGCTCCAATGTGGAGCTGCGTGCCAGCTATGCGCATATTGTGCGCCTGGC
D.nov-Gr9a_nuc      GAGGGGACGACGCTCCAATGTGGAGCTGCGTGCCAGCTATGCGCATATTGTGCGCCTGGC
D.tex-Gr9a_nuc      GAGGGGACGACGCTCCAATGTGGAGCTGCGTGCCAGCTATGCGCATATTGTGCGCCTGGC
D.vir-Gr9a_nuc      GAGGGGACGACGCTCCAGTGTGGAGCTGCGTGCCAGCTATGCGCATATTGTGCGCCTGGC

D.amer-Gr9a_nuc     GCGCCAGCTCTCCCGCCTCTATGGGCTATCCGTGCTGATGA---ACGTGCTTTGCATGGG
D.nov-Gr9a_nuc      GCGCCAGCTCTCCCGCCTCTATGGGCTATCCGTGCTGATGATGAACGTGCTTTGCATTGG
D.tex-Gr9a_nuc      GCGCCAGCTCTCCCGCCTCTATGGGCTATCCGTGCTGATGATGAACGTGCTTTGCATTGG
D.vir-Gr9a_nuc      GCGCCAGCTCTCCCGCCTCTATGGGCTATCCGTGCTGATAATGAACGTGCTTTGCATTGG

D.amer-Gr9a_nuc     CGACTTGATTGTGGTGTGCTATGCCTATATTATACTCTGGCAGCTGACCAACATGAATAT
D.nov-Gr9a_nuc      CGACTTTATTGTGGTGTGCTATGCCTATATTATACTCAGGCAGCTGACCAACATGAATAT
D.tex-Gr9a_nuc      CGACTTTATTGGTGTGTGCAATGCCTATATTATACTCTGGCAGCTGACCAACATGAATAT
D.vir-Gr9a_nuc      CGACTTTATTGTGGTGTGCTATGCCTATATTATACTCTGGCAGCTGACCAACATGAATAT

D.amer-Gr9a_nuc     GAGCTGGCTGCTGCTCTGGCAGACCATATACGTGGTGCTGCCCACGCTGATCAAGATCTG
D.nov-Gr9a_nuc      AAGCTGGCTGCTGCTCTGGCAGACCATATACGTGGTGCTGCCCACGCTGATCAAGATCTG
D.tex-Gr9a_nuc      TAGCTGGCTGCTGCTCTGGGAGACCATATACGTGGTGCTGCCCACGCTGATCAAGATCTG
D.vir-Gr9a_nuc      TAGCTGGCTGCTGCTCTGGGAGACCATATACGTGGTGCTGCCCACGCTGATCAAGATCTG

D.amer-Gr9a_nuc     TACGCTGTGCGCCGTCTGCGGAGGGAAGTGTGCCAAGCGGTCGCAGCAGCTTCTGGTCGT
D.nov-Gr9a_nuc      TACGCTGTGCGCCGTCTGCGAT---AAGTGTGCCAAGCGGTCGCAGCAGCTTCTGGTCGT
D.tex-Gr9a_nuc      TACGCTGTGCGCCGTCTGCGAG---AAGTGTGCCAAGCGC--------------------
D.vir-Gr9a_nuc      TACGCTGTGCGCCGTCTGCGAG---AAGTGTGCCAAGCGCTCGCAGCAGCTTCTGGTCGT

D.amer-Gr9a_nuc     GATCGGCGGCACTTGGGAGCCAGGGGACGGGAACAGGACCCGGACTCACACGCATACAAG
D.nov-Gr9a_nuc      GATCGGCGGCACTTGGGAGCCAGGAGAGGGGAACTAGACCCAAACTCACACGCATACAAG
D.tex-Gr9a_nuc      ------------------------------------------------------------
D.vir-Gr9a_nuc      GATCAGCAGCACTTGGCGCCAGGAGCCGGGAACGAGACCCGGACTCACACGCATACAAGC

D.amer-Gr9a_nuc     CCGATGCACTCACACGCATACAAGCCGATGAATTCTCGCCACAGGTAATGCAGAA-TTTC
D.nov-Gr9a_nuc      CCGATG----------------------AATTCTCGCTACAGATTATGCAGNAATTCC
D.tex-Gr9a_nuc      ------------------------------------------------------------
D.vir-Gr9a_nuc      CGATGAATTCTCTCTACAGATAATGCAGAATTCCGTGCAGTTCGATATCTGCGGCATGTA

D.amer-Gr9a_nuc     GTGCAGTCGGATGTCTGCGGCATGTACCATCTATGTCTGAAGACCATGGCTGGG------
D.nov-Gr9a_nuc      GTGCAGTTCGATGTCTGCGGCATGTACCATCTATGTCTGAAGACCATGGCTGGGATGTTT
D.tex-Gr9a_nuc      ------------------------------------------------------------
D.vir-Gr9a_nuc      CCATCTATGTCTGAAGACCTTGGCCGGGATGTTTTTCTTCATACTCAACACGCTGGTCAC

D.amer-Gr9a_nuc     ------------------------------
D.nov-Gr9a_nuc      TTCTTCATACTCAACACGCTGGTCACTTTCC
D.tex-Gr9a_nuc      ------------------------------
D.vir-Gr9a_nuc      TTTCCTTCAGTTCGTG---------------
```

Figure 3-6. Multiple sequence alignments of candidate Gr9a genes from *D. americana, D. novamexicana, D. texana* and *D. virilis* with nucleotide substitutions highlighted with yellow and nucleotide insertion/deletion with green. At the end of alignments, sequences written with red ink indicating non-similar sequences between species.

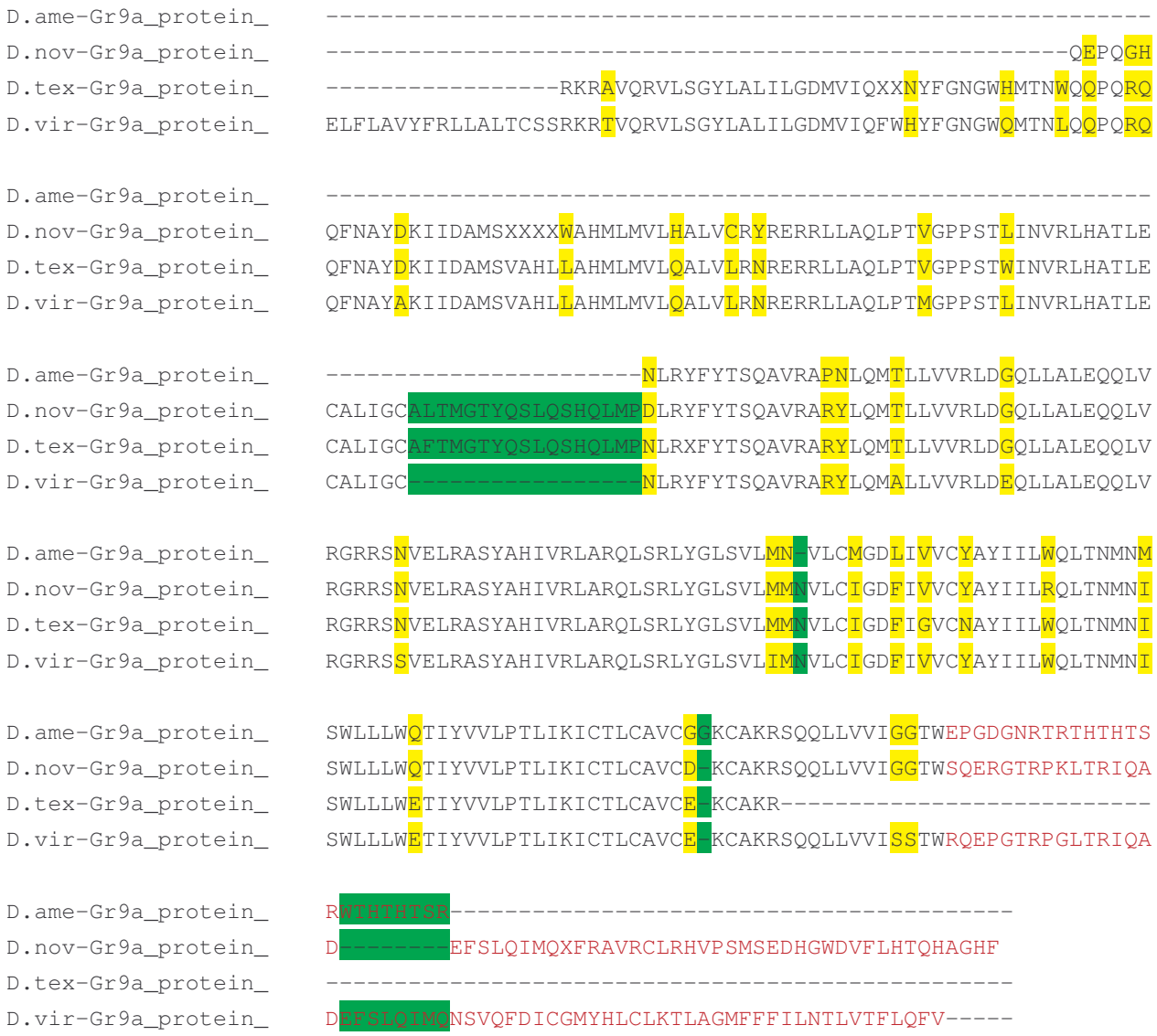

Figure 3-7. The alignment of amino acid products of candidate Gr9a genes for *D. americana, D. novamexicana, D. texana* and *D. virilis* with amino acid substitutions highlighted with yellow and amino acid insertion/deletion with green. Amino acids written with red ink indicate non-similar amino acids between species.