DISSERTATION

AEDES AEGYPTI VECTOR COMPETENCE AND GENE FLOW IN MEXICO. ASSOCIATION MAPPING SOFTWARE FOR TESTING CANDIDATES GENES ASSOCIATED WITH A PHENOTYPE

Submitted by

Saul Lozano-Fuentes

Department of Microbiology, Immunology and Pathology

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, CO

Summer 2005

COLORADO STATE UNIVERSITY

April 29, 2005

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY SAUL LOZANO-FUENTES ENTITLED '*AEDES AEGYPTI* VECTOR COMPETENCE AND GENE FLOW IN THE STATE OF VERACRUZ, MEXICO, AND ASSOCIATION MAPPING SOFTWARE FOR TESTING CANDIDATES GENES ASSOCIATED WITH A PHENOTYPE' BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Mo Salman

Advisor

Co-Advisor

Department Head

ABSTRACT OF DISSERTATION

AEDES AEGYPTI VECTOR COMPETENCE AND GENE FLOW IN THE STATE OF VERACRUZ, MEXICO, AND ASSOCIATION MAPPING SOFTWARE FOR TESTING CANDIDATES GENES ASSOCIATED WITH A PHENOTYPE

I have discovered an abrupt barrier to gene flow in *Ae. aegypti* associated with the Neovolcanic axis (NVA) along the Gulf Coast of Mexico. The mountains associated with the NVA limit the distribution of *Ae. aegypti* to ~15 km between the mountains and the Atlantic Ocean. Most importantly the change in gene flow appears to have also caused an abrupt increase in the frequency of mosquitoes with a midgut infection barrier (MIB) to the south of the NVA and a gradual increase in mosquitoes with a midgut escape barrier (MEB) south of the axis. The results of this study demonstrate how a barrier in gene flow can also cause a shift in the frequencies of alleles at MIB and MEB loci and thereby cause changes in susceptibility.

Early trypsin (ET) is considered a candidate gene controlling MIB and MEB. It has been shown that *Ae. aegypti* females fed an infectious blood meal, containing a soybean trypsin inhibitor, presented a decrease in midgut DEN2 RNA copies. DEN2 proteolytic processing, mediated by midgut trypsins, probably influenced the infection (Molina-Cruz *et al.* 2004). Six criteria were applied to minimize the detection of false positive QTNs when testing the association between segregating sites in ET and DEN2 susceptibility in *Ae. aegypti*. First, *a priori* hypothesis that ET is a candidate gene was established. Second, large sample sizes were used to maximize the analysis power.

Third, the independence of segregation of alternate nucleotides at segregating sites was tested. Fourth, I used mtDNA (neutral marker) to identify false QTNs arising from analysis of genetically distinct populations. Fifth, populations were subdivided (panmictic units) to independently validated QTNs. Sixth, patterns of association between alleles and genotypes were examined when presumptive QTNs arise.

No consistent associations between segregating sites in ET and susceptibility were found. However, little systematic disequilibrium was detected in the ET gene allowing the possibility that *cis* elements in close proximity to the ET gene were undetected. Additionally, mutations in enhancers or promoters residing outside the analyzed region, which could be associated with the phenotype, remain a *de facto* qualification of QTN studies.

Saul Lozano-Fuentes
Department of Microbiology,
Immunology and Pathology
Colorado State University
Fort Collins, CO 80523
Summer 2005

ACKNOWLEDGMENTS

Stalin said *gratitude is a sickness suffered by dogs*. I do not fill like a dog but I do feel a great deal of gratitude.

I will like to thank the members of my doctoral committee for putting time aside to help me complete my formation as scientist. I whish I had spent more time to learn from Dr. Freier, Dr. Moore and Dr. Salman. Time and objectives limited these interactions. Still, I will always be in indebted to you.

To Dr. Barry Beaty and Dr. William C. Black IV, I offer my interminable gratitude. I thank Dr. Beaty, for his kind support and his rigorous interest in my research. I will try to follow his example of success through hard work and his apparent innate abilities to navigate the biopolitical waters. I thank Dr. Black, for his continuous support and orientation. I consider him part of my family. His way of seeing life has leaved a profound mark in me as well as the trips to Pingree Park.

I also would like to acknowledge Dr. Ildefonso Fernandez-Salas for opening to me "doors and windows of opportunity". I thank Dr. Adriana Flores-Suarez for allowing me to used her laboratory space and tolerate my music. And finally I will like to thank Dr. Roberto Mercado-Hernandez for his words of support and for teaching me how to approach problems logically.

I will like to specially thank Dr. Norma Gorrochotegui-Escalante, Dr. Kristine Bennett, Dr. Alvaro Molina and Dr. Francisco Diaz for enduring my constant questions and giving me their friendship.

To my fellow students at AIDL and at the Black Lab, thank you for making the long days of work bearable and amusing. I offer my continuing friendship to Conny, Carlos, Debra, Erick, Isabel, Ian, Jason, Maria de Alba and Tiffany. Alejandro "La güera" Espinoza, Jorge "Bambam" Gomez, Luis "Li" Peña and Carlos "3b" Velazco for your borderless friendship, thank you.

This work could not have completed without the help of Don Flick, Karen Fleming, Allison Fredenberg, Jenn Holmes, and John Montgomery, and several work-study students, who when through the grueling task of ordering, stocking, preparing and making aliquots of reagents and supplies, among many other things.

I apologize to my sister Janet and my new brother Juanjo for missing their wedding. Thank you for understanding the reasons. I thank my Sister Xochilt for the long passionate talks in how to solve the world's problems. I thank my father and mother for 29 years of love and support, even though they asked me when I was going to get "a real job".

I am grateful of my wife Karla and my son Isaac for their patience (especially in the last months before the completion of this document) and love. Thank you for following me into this journey. Regardless of the uncertainty and the difficulties, you kept walking next to me.

DEDICATION

To Karla and Isaac.

Table of Contents

List of tables

List of figures

I Literature Review

Introduction:

It has been close to 120 years since the discovery by Ross that a mosquito could transmit a disease, and the world is still burdened by vector borne disease. Wars have been halted and many economical enterprises have not been completed as a result of the human, vector, and pathogen interaction. The list of pathogens transmitted by a vector is large, especially arthropod the number of borne viruses (arboviruses). However, the most important vector borne diseases are: African trypanosomiasis, dengue, Leishmaniasis, Malaria, Chagas disease and lymphatic filariasis (Table 1).

These diseases are no longer a public health problem in developed countries, only reappearing in traveling tourists or immigrants; for this reason they have been named "diseases of the poor" (World Health Report, 2002). This expression can be considered prejudiced, but suggests that improving the economic welfare of families could help better control vector borne diseases and considerably reduce the number of deaths worldwide.

The damage vector borne diseases do on humanity is a consequence of insects that must take a blood meal for development of their eggs, flight or as food source. During blood feeding pathogens are transmitted. The fact that mosquito females from many species usually have multiple gonadotropic cycles, and that some species, such as *Ae. aegypti* and *An. gambiae*, even feed several times within a single cycle, makes them ideal vectors of pathogens like arboviruses, protozoan and filarial parasites. Arboviruses in particular have seen a global resurgence in the past 2 decades (Gubler, 1996).

It is not completely understood why there was an increase in these diseases. Factors such as alterations in global climate (Patz *et al.*, 1995), disturbance of ecosystems (Epstein, 1995), increased urbanization and poverty (Gubler, 1996; Monath 1994), insecticide resistance (Brown and Brown, 1974; Goldman *et al.*, 1986; Mouches *et al.* 1987; Thomson *et al.,* 1993; Hemingway *et al.,* 2002; Ranson *et al.,* 2002a, b), and the collapse of public health infrastructure and medical entomological control have been hypothesized to contributed to this resurgence. Mexico can serve as an example to demonstrate the speed of the recovery of a mosquito population capable and its ability to generate and sustain an epidemic. In 1962 Mexico was declared *Ae aegypti* free; thirteen years later in 1975 Mexico was recolonized at both borders and only five years later, in 1980, there was a major epidemic with more than 50,000 dengue cases (SSA, 2001).

In my opinion, there are many other issues that factor into the increase of arboviral diseases, that given areas have unique socio-political problems that must be overcome. This is further potentiated by the economic gap between countries that supply insecticides and equipment for vector control and those in dire need of it (Goodman and Mills, 1999). All these factors, in addition to the facility with which people and cargo

now travel globally, are contributing to the distribution and increased prevalence of viruses and their vectors. The West Nile Virus USA epidemic serves as an example, because it clearly showed how fast dispersal can occur and how many states in the USA had lost their public health capabilities (CDC, 2004). The future for vector borne disease control in the world seems bleak.

Vector control, when well implemented, is an effective tool to control vector borne diseases. The logic behind vector control is straightforward. The vector is ultimately necessary for the continuous transmission of the disease. Unfortunately, as mentioned previously, the increasing prevalence of insecticide resistance - often to multiple groups of insecticides (Mouches *et al.,* 1987; Thomson *et al.*, 1993; Goldman *et al.,* 1986; Brown and Brown, 1974) makes traditional forms of vector control difficult in many areas. In addition, there is little interest from the insecticide companies to create new molecules for vector control; agricultural pest control is more profitable and the public health sector provides small revenues. Vector-borne disease control will require new strategies.

The symbiosis between pathogen and vector is considered circumscribed, thus vector control is typically targeted to particular species. This symbiosis can be described within the concept of vector competence. Vector competence is defined as the intrinsic permissiveness of a vector to infection, replication, and transmission of a pathogen (Hardy, 1988). Vector competence does not explain whether transmission actually could occur or could be maintained in the environment. It simply defines whether a potential vector is capable of supporting, replicating, and delivering a pathogen. Vectorial capacity (MacDonald, 1973) accounts vector competence but also includes such aspects

as vector abundance, host preference, longevity, and the extrinsic incubation period of the given pathogen in the given vector. Nevertheless, virus transmission through an arthropod vector is not a simple matter. Vector competence of *Ae aegypti* is dependent on its susceptibility to midgut infection with a virus following an infectious blood meal and its permissiveness to virus dissemination to the salivary glands for transmission to a subsequent host. Variation in the mosquito susceptibility is well documented. Variation in midgut infection (MI) and disseminated infection (DI) are due primarily to midgut infection barriers (MIB) and midgut escape barriers (MEB).

Several countries went through eradication campaigns of disease vectors (PAHO, 1997) and most attempts failed. Eradication for all countries is not a realistic option, as it would require careful planning, global cooperation and persistence all driven by an unrestricted budget. The new approach, determined by an ever smaller budget, is that control efforts should forget eradication and instead concentrate on finding ways to control vector species in ways that minimize exposure to bites reducing the number of cases and as much as possible the incidence of disease.

Another route for disease control is the modification of the vector-pathogen relationship by introducing genes that confer resistance to diseases. To accomplish this, a better understanding of the interactions between the virus and its vector is required. Recent advances in molecular biology techniques give researchers more tools to investigate disease vectors (Besansky and Collins, 1992; Collins, 1994; Crampton *et al.*, 1994; Carlson *et al.*, 1995; Beaty, 2000; Black *et al.*, 2000; Black *et al.*, 2001; Xiong *et al.*, 1989; Olson *et al.*, 2000; Powers *et al.*, 1996; Beard *et al.*, 1993; Afanasiev *et al.*,

1994; Yee *et al.,* 1994). This information and the completion of the *Aedes aegypti* genome project may create effective techniques for disease control.

The importance of biological factors is sometimes undermined by non-biological factors (Meade *et al.*, 1988). An approach that involves assessing the interactions and associations between elements of the physical and cultural environments, that could explain the outcome of an outbreak, is contemplated inside landscape epidemiology. Landscape epidemiology involves the identification of elements linked to geographical areas where disease is transmitted. That is, by knowing the vegetation and geologic conditions necessary for the maintenance of specific pathogens in nature, one can use the landscape to identify the spatial and temporal distribution of disease risk (Pavlovsky, 1966). Under this concept environmental elements, including elevation, temperature, rainfall, and humidity, influence the presence, development, activity, and longevity of pathogens, vectors, zoonotic reservoirs of infection, and their interactions with humans (Meade *et al.*, 1988).

The use of computers has been integrated in the overall strategy of disease control for the data storing, data querying, meta analysis and in the decision making process. Drawing maps and studying spatial pattern is a human activity done for centuries. In comparison, the need to reduce information to numbers is rather recent. The human eye and brain are capable of analyzing complex patterns but not without bias.

Geographic Information System (GIS) and Global Positioning Systems (GPS) technologies are currently a fast and accurate method for the acquisition of data for control of vector borne diseases (Wagner *et al.,* 1979; Washino and Wood, 1994; Hayes *et al.,* 1985; Hugh-Jones, 1989; Kitron, *et al.,* 1994; Malone *et al.,* 1994; Kitron and

Kazmietcsak, 1997; Dister *et al.,* 1997; Kitron, 1998; Morrison *et al.,* 1998). In addition, GIS is a powerful tool in the decision-making process.

The Internet is further enhancing the GIS technology. The WHO initiative is called "The Global Atlas of Infectious Diseases" (http: //www.who.int/ GlobalAtlas/ home.asp). The key components of the site include but are not limited to: 1) Data Query allowing users to browse, view, query, search the contents of the WHO's Communicable Disease global database and output data in reports, charts and maps. 2) Interactive Mapping that provides a user-friendly mapping interface that allows users to select geographic areas of interest and create maps of diseases, the location of health facilities, schools, roads and geographic features. In the near future, direct data dumping to the internationally accessible web servers will be a reality. However, socio-economic concerns will surely limit this resource. The travel industry will not be happy with a Yellow Fever report in Cancun, Acapulco, Aruba or any other tourist city. The impact of computer-based technologies on vector borne diseases will be discussed later.

Aedes(Stegomyia) *aegypti* (Linnaeus)

Aedes aegypti per se

The history of the classification of *Ae aegypti* is confusing and interesting. Historically *Ae aegypti* was first described by Linnaeus as *Culex aegypti.* However Linnaeus' original description did not match with *Ae aegypti per se,* but to *Ochleorotatus caspius.* The oldest description that matches *Ae aegypti* is from 1805; Fabricius named the mosquito *Culex fasciatus.* Reed published that the intermediary mosquito and responsible of transmitting yellow fever was *Culex fasciatus* (Reed, 1900). The name was later changed by Theobald to *Stegomyia fasciata* in 1901. The name remained

unchanged until 1962, when it was renamed as *Stegomyia aegypti* by Mattingly (Mattingly *et al.,* 1962). The main reason the name *aegypti* was reintroduced was the historical link to the name *aegypti.*

Ae aegypti is a domestic mosquito. Its adult and larval habitats are closely related to man (anthropophilic), it oviposits preferentially in containers found in domestic and peridomestic areas (Gordon, 1988). *Ae. aegypti* females feed mainly from human (antropophagic) (Tabachnick, 1991; Scott *et al.*, 2000a). Its distribution is pantropical between the 30° N and 20° S (Trend *et al.,* 1983); however altitude clearly limits its distribution.

There are two subspecies, "black and light", inside *Ae aegypti. Ae aegypti formosus*, the "black" subspecies, occurs in sylvatic habitats in sub-Saharan Africa it has been shown to be quite resistant to infection with dengue and yellow fever, and is probably not a significant vector. *Ae ae formosus* preferentially oviposits in tree holes. On the other hand, *Ae aegypti aegypti*, the "light" subspecies is primarily an urban mosquito. It has been shown to be highly susceptible to infection with dengue and is the vector most associated with dengue epidemics. *Ae ae aegypti* it preferentially oviposits in artificial water containers such as cisterns and tanks; however, tires, refrigerators, washers, empty soda bottles and other objects that can hold water, can also serve as breeding sites. It feeds almost exclusively on human blood. Natural selection may favor females that feed on human blood without feeding on sugar (Scott *et al.,* 1997; Harrington *et al.,* 2001). This adaptation is advantageous, because it provides increased energy reserves and a fitness advantage (Harrington *et al.,* 2001; Costero *et al.,* 1998), in

addition to the possibility of an increased flight range (Wells, 2002, personal communication).

Theoretically, anywhere that *Ae aegypti* occurs, there is potential for dengue transmission. However, in practice, this is not the case. *Ae aegypti* vectorial capacity differs throughout the world and vector competence also changes. Vector competence for flaviviruses is genetically determined in part (Aitken *et al.,* 1977; Wallis *et al.,* 1983; Miller and Mitchell, 1991; Rosen *et al.,* 1985; Lorenz *et al.,* 1984; Vazeille *et al.,* 2001).

Vector Competence:

How dengue virus infects the mosquito midgut is not completely understood. Research with other arboviruses (Bunyaviruses and Orbiviruses) suggests the need for proteolytic processing of virus proteins involved with host cell receptor attachment to midgut epithelial cells for efficient vector infection (Ludwig *et al.,* 1991; Ludwig *et al.,* 1989; Mertens *et al.,* 1996; Xu *et al.,* 1997). No virus receptors have been found in *Ae aegypti.*

However, it has been shown that when *Ae. aegypti* females are fed a dengue infectious blood meal that also contains soybean trypsin inhibitor (STI) there is a 91-97% decrease in midgut DEN2 RNA copies (Molina-Cruz *et al.,* 2004). STI treatment also resulted in slower DEN2 replication in the midgut, less DEN2 E protein expression, and decreased dissemination to the thorax. A second uninfected blood meal, 7 days after the STI-treated infectious meal, significantly increased DEN2 replication in the midgut and recovered oogenesis, suggesting that the lower viral infection caused by STI was in part due to a nutritional effect. Mosquitoes fed DEN2 digested *in vitro* with bovine trypsin (before STI addition) exhibited a transient increase in midgut DEN2 4 days post-

infection. Blood digestion and possibly DEN2 proteolytic processing, mediated by midgut trypsins, influence the rate of DEN2 infection, replication and dissemination (Molina-Cruz *et al.*, 2004).

Despite the lack of detailed understanding of the mechanism of dengue infection of midgut epithelial cells, its genome must be transcribed, translated and replicated for production of mature virions. If the virus cannot infect the midgut, the mosquito is said to have a midgut infection barrier (MIB). MIB is associated with attachment, penetration, and uncoating of the virus in the mosquito midgut. Studies on vector competence of *Ae aegypti* have consistently shown that MIB has a major effect on the potential for flavivirus transmission (Schoepp *et al.,* 1990; Bosio *et al.,* 1998; Gubler *et al.,* 1979; Tabachnick *et al.,* 1985).

Once an infection has been established in the midgut, the virus must be able to disseminate from the midgut into the hemocoele in order to infect the salivary glands. The mosquito is said to have a midgut escape barrier (MEB) if the virus is unable to disseminate from the midgut or if the virus disseminates but is unable to infect secondary organs. MEB has been shown to be partially responsible for the resistance of *Ae aegypti* mosquitoes to yellow fever (Miller and Mitchell, 1991).

Virus must infect the salivary glands before it is transmitted during mosquito blood feeding. Salivary gland barriers have been shown in several systems (Whitfield *et al.,* 1973; Whitfield *et al.,* 1971; Takahashi and Suzuki, 1979; Sriurairatna and Bhamarapravati, 1977; Takahashi, 1980; Grimstad *et al.,* 1985; Jupp, 1985; Kramer *et al.,* 1981; Beaty and Bishop, 1988; Paulson *et al.,* 1989). However, salivary gland

Infection barriers (SIB) or salivary gland escape barriers (SEB) have not been described in *Ae aegypti* transmission of flaviviruses.

Variation in *Ae aegypti* vector competence has been well documented. Many studies have shown that the *Ae aegypti* range of susceptibility to infection with flaviviruses is, at least in part, genetically determined (Aitken *et al.,* 1977; Wallis *et al.*, 1983; Tabachnick *et al.,* 1985; Rosen *et al.,* 1985; Black *et al.,* 2002). There is geographic variation in oral infection rates with yellow fever virus in populations of *Ae aegypti* from around the world (Tabachnick *et al.,* 1985). This susceptibility corresponded with the 8 genetic groupings found using allozyme analysis (Tabachnick and Powell, 1979). The populations found to be least susceptible to infection (7-34% disseminated infection) were from the *Ae aegypti formosus* subspecies, while populations from the *Ae aegypti aegypti* subspecies from the Caribbean (34-53% DI) and east Africa (29-57% DI) were much more susceptible (Tabachnick *et al.,* 1985).

Gubler *et al.* (1979), among others (Rosen *et al.,* 1985; Tardieux *et al.*, 1990; Vazeille-Falcoz *et al.,* 1999), have also shown that *Ae aegypti* is extremely variable in its susceptibility to infection with dengue viruses. Analysis of 13 geographic populations from around the globe showed that populations from the South Pacific and Southeast Asia had a higher susceptibility to dengue (9-62% DI) than did populations from East and West Africa (0-12%) (Gubler *et al.,* 1979).

The variation in vector competence for flavivirus is not limited to differences among populations, but has been documented even within *Ae. aegypti aegypti* and *Ae. aegypti formosus* families (Miller and Mitchell, 1991). There were significant differences in DI (0-70%) and transmission (0-33%) of yellow fever virus among individual families

within a population of *Ae. aegypti formosus* from Nigeria. Similar results were obtained within families *Ae. aegypti aegypti* from Puerto Rico (27-100% DI and 0-44% transmission) (Miller and Mitchell, 1991).

Ae. albopictus, an implicated vector of dengue virus, also exhibits considerable variation in its susceptibility to dengue viruses (Boromisa *et al.,* 1987; Bosio *et al.,* 1992; Gubler and Rosen, 1976). *Ae. triseriatus* varies in its susceptibility to LaCrosse (LAC) virus, and in its rate of transovarial transmission of LAC virus and is also genetically controlled (Grimstad *et al.,* 1977; Miller *et al.,* 1982; Graham *et al.* 1999).

Non biological factors involved in dengue transmission:

Ae aegypti is closely linked to the human environment since its origin in the sub-Saharan region. The dramatic increases in the densities of *Ae aegypti*, due in part to increased larval and adult habitats, is a result of fast growth of cities, lack of adequate public services (water, trash collection sewer, etc) and human life styles (Gubler, 1993). Additionally, commerce has created mechanisms for rapid movement of mosquitoes among cities (Gubler and Trent, 1994).

Vector competence is only a small portion of the vectorial capacity equation. There are other factors that influence whether a virus will actually be transmitted. Non biological factors, like rainfall and temperature, play a central role in vector borne diseases. When biological and non biological factors are taken into account, the ability to predict disease transmission becomes complex. These extrinsic environmental factors are an integral part of determining whether a vector will transmit a given virus at a given time (Hardy, 1988).

Among the events that are influenced by environmental factors is the amount of time it takes a vector to be able to transmit a virus once it has ingested an infected blood meal; this is the extrinsic incubation period (EIP). The length of the EIP is considered to be inversely proportional to the temperature the vector is exposed once it has an infected blood meal. In early studies the EIP of *Haemagogus capricornii* for yellow fever virus went from 28 days when incubated at 25 °C to just 12 days when incubated at 30 °C (Bates and Roca-Garcia, 1946). Researchers have found similar results in other systems (Chamberlain and Sudia, 1955; Kramer *et al.,* 1983; Hurlbut, 1973). The EIP for DEN-2 in *Ae aegypti* went from 12 days at 30 °C to 7 days at 32-35 °C (Watts *et al.,* 1987). Temperature also produces effects in the maturation of mosquitoes; high temperatures produce small mosquito females, and this forces the females to constantly blood feed to obtain sufficient protein for egg production (Focks *et al.,* 1995). The transmission of dengue mainly occurs in the hottest months of the year (Gubler and Trent, 1994) and mosquito size may be one of the reasons. Larval nutrition also has an impact in the size of the adult. Smaller females of *Ae triseriatus*, nutritionally deprived as larvae, were shown to transmit LAC more efficiently than larger females (Grimstad and Haramis, 1984; Grimstad and Walker, 1991). Smaller *Culex tritaeniorhynchus* females were more susceptible to West Nile virus (Baqar *et al.,* 1980).

The working hypothesis is that smaller females have an increase in vector competence as a result of ingesting more frequent blood meals to compensate for their nutritional deprivation. However, Sumanochitrapon *et al.* (1998) found that larger *Ae aegypti* females were more susceptible to infection with DEN-2. Additionally, researchers have also shown that larger body size translates to longer life span, e.g., *Oc.*

punctor (Packer and Corbet, 1989), *Ae. aegypti* (Briegel, 1990) in the laboratory, and *An. arabiensis* (Ameneshewa and Service, 1996) in field populations.

Daily survivorship is more important than susceptibility of the vector population (MacDonald, 1973), and in consequence, an increase in susceptibility to infection from smaller size may be outweighed by an increase in longevity from larger size.

Ae aegypti behavior may also play an important role in dengue virus transmission. *Ae aegypti* females can take several blood meals per gonotrophic cycle (Scott *et al.,* 1993, 2000b). Additionally, feeding exclusively from human blood without taking sugar meals may promote virus transmission increasing human contact and giving a fitness advantage to the female mosquito (Harrington *et al.,* 2001). This behavior may explain transmission of dengue in areas were *Ae aegypti* densities are low (Kuno, 1997; Focks *et al.,* 2000).

There are other environmental factors that influence vector competence like rainfall and human settlements. Increased precipitation has the potential to increase the number and quality of resting places sites for vectors (Githeko *et al.,* 2000). In South America more than 70% of the population lives in rural areas and thus only a small proportion is exposed to rural infections. In contrast, in Africa, more than 70% of the population lives in rural areas making vector control (e.g. removal of larval breading sites) difficult (Githeko *et al.,* 2000).

Aedes aegypti in Mexico

The dengue vector has a long history in Mexico; the registers of the Spaniard "conquistadors" suggested the transmission of YF among soldiers and the native population (Nathan, 1991). The first reports of epidemics came from the cities of Campeche and Merida in the Mexican Caribbean. Even though, the transmission of YF

stopped in Mexico, *Ae aegypti* was abundat until 1957 when the eradication campaigns were launched (Soper, 1967). In 1962 Mexico was officially declared *Ae. aegypti* free, thirteen year later (1975) the mosquito recolonized both borders and only five years later (1980), there was a major epidemic with more than 50,000 Dengue cases (SSA, 2001).

Eradication campaigns in the 1950's and 1960's nearly eradicated *Ae aegypti* from many American countries. These campaigns stopped in the 1970's in part because of reduction in vector borne diseases in many areas. The mosquito was able to return to its pre-campaign prevalence in recent years, exceeding its previous boundaries.

In recent years, the Mexican authorities changed their objectives from eradication to control. However, not reporting cases to give the appearance that dengue cases are under control in Mexico, may have hurt the credibility of the sanitary authorities. In 2000 in the city of Monterrey alone there were more than 10,000 cases, the following year the city reported only 13 cases (SSA, 2001).

Dengue

The term dengue was introduced into medical literature as a translation from the Swahili *dinga*, *dyenga* or *ki denga pepo*. The term used in Indonesia was knokkel-koorts and the term used in Philadelphia in 1789 was Break bone Fever or Dandy Fever to describe the unpleasant sensation of deep muscular pain (Halstead and Porterfiel, 1980).

Vectors capable of maintaining dengue in urban cycles are: *Ae aegypti*, *Ae albopictus*, *Ae (Gymnometopa) mediovitatus* and *Ae scutellaris, Ae luteocephalus, Ae polynesiensis* (Soper, 1967; Gould *et al.,* 1968; Metselaar *et al.,* 1980; Gubler, 1988; Fan *et al.,* 1989), but the primary vector for urban transmission is *Ae aegypti aegypti.*

Sylvatic cycles have been documented in Southeast Asia between susceptible monkey species and mosquito vectors of the *Ae (Finlaya) niveus* complex (Rudnick, 1965, 1978). In West Africa sylvatic cycles have been documented involving a number vector species: *Ae (Diceromyia) furcifer, Ae (D) taylori,* and the *Ae (Stegomyia) africanus* complex (Cordellier *et al.,* 1983; Cornet *et al.* 1984). Although a sylvatic cycle has not been discovered in South and Central America, it is presumed to exist (Rosen, 1958). Provocatively, Platt *et al.* (2000) argue they have detected dengue infection using neutralizing antibodies in sera from bats collected in urban areas of Costa Rica and Ecuador.

Dengue, virus isolates from sylvan cycles in West Africa have been shown to be genetically distinct from isolates obtained in urban cases adding another piece to the puzzle. This suggests that the two cycles may be unrelated (Kerschner *et al.,* 1986).

Dengue Disease Manifestations:

A typical case of dengue fever (DF) begins 2-7 days after an infectious bite. The initial symptoms include high fever, headache, retroorbital pain, lumbosacral aching pain and conjunctival congestion. The fever may last for 6-7 days, followed by excruciating myalgia, anorexia, nausea, vomiting that leads to weakness and prostration. Petechiae may appear on the first or second day. Convalescence can be prolonged and may include general weakness, depression, bradycardia, and ventricular extrasystoles. This presentation of dengue is not consider life threatening, but could severely impact the economy of a household.

Dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) are severe forms of dengue virus infection. In most cases, the disease initially presents as DF, but

after 2-5 days progresses to prostration, restlessness, irritability, shock with cold clammy extremities, diaphoresis, rapid respiration, rapid pulse, and hypotension. Spontaneous hemorrhages can occur, both internally and externally. DHF is distinguished from DF by the presence of thrombocytopenia and hemoconcentration. Hypotension or profound shock characterizes DSS (WHO, 1998).

Treatment for DF and DHF/DSS is supportive only. When there are signs of dehydration, treatment includes careful monitoring of electrolytes and fluids by hydration therapy and maintenance of the circulating fluid volume. This type of supportive therapy has succeeded in decreasing the case fatality rate from nearly 40% to less than 5%.

Perhaps the most alarming factor in the recent dengue pandemic has been the rise in the number of individuals progressing to the more severe forms of the disease: DHF and DSS. Reasons for this increase may include antibody dependent immune enhancement (Halstead, 1982; Halstead 1988) and/or increased virus virulence (Rico-Hesse *et al.,* 1997; Holmes and Burch, 2000).

Infection with one serotype confers life-long immunity to exposure to homologous, but not heterologous serotypes. DHF is seen most commonly in patients suffering a secondary infection with a dengue serotype heterologous to the primary infection, suggesting a potential for antibody dependent immune enhancement (ADE) (Halstead, 1982, 1988). In ADE, antibody binds to dengue particles but does not neutralize it. The virus-antibody complexes more efficiently infect macrophages, thereby changing the pathogenesis of the infection.

Differences in virus virulence have been suggested as another determinant of dengue disease outcome. The first occurrence of large numbers of DHF cases in the

Americas coincided with the introduction of a new strain of DEN-2 that was similar to virus circulating in Southeast Asia (Rico-Hesse, 1990). Similar results were seen when DEN-3 was introduced into the Americas (Gubler and Clark, 1995). However, it is difficult, to distinguish between effects from the introduction of new virus strains from other factors, such as ADE, that may play a role in disease severity.

Dengue Epidemiology:

Dengue is currently considered (along with several other important illnesses) to be an emerging disease (Gubler, 1996; WHO, 2002). It is sometimes considered to be the most important arboviruses disease of human beings (Gubler, 1998; Monath 1994). Two and a half billion people (2/5 of the global population) are at risk for dengue infection. There are currently approximately 50-100 million cases of classical dengue fever (DF) annually and an estimated 500,000 cases per year of DHF/DSS (Monath, 1994; Gubler and Clark, 1995). The number of dengue cases is increasing globally justifying its classification as an emerging disease. (Dove, 1998; Gratz, 1999; Chareonsook *et al.,* 1999; Aviles *et al.,* 1999; Yamada *et al.,* 1999; Clarke, 2002)

Dengue is currently endemic in over 100 countries in the world. Epidemics occur regularly in Southeast Asia, India, the western Pacific, Mexico and Central and South America (Clarke, 2002). Prior to 1970 only 9 countries had reported DHF epidemics - as of 1995, that number had increased more than 4 fold (WHO Fact Sheet No 117 Revised April 2002).

Dengue was considered a disease of the tropics and/or of developing countries. This view, however, is changing. Dengue has been diagnosed in Southern United States

(Pena *et al.,* 1999). There is clear disparity between the number and severity of cases seen in the southern United States and Northern Mexico.

DF/DHF/DSS has been an economic burden not easily supported on the developing world governments. Large sums of money have to be spent on vector control, medical care, and hospitalization costs. There is also a decreased productivity and disruption of tourism (VonAllmen *et al.,* 1979; Gubler, 1988; Meltzer *et al.,* 1998), as well as a reduction of school attendance. In Puerto Rico and Thailand the highest infection rates have been found in small children and women than older children and men. (Halstead *et al.,* 1969; Morens *et al.,* 1986). In Mexico, school aged children showed the highest attack rates from mosquitoes than any other aged group. (Rodriguez-Figueroa *et al.,* 1995).

Dengue Virus:

Dengue viruses belong to the genus Flavivirus in the family Flaviviridae. This family is made up of single stranded RNA viruses of positive polarity (Monath, 1996; Rice, 1996; Heinz *et al.,* 2000; Burke and Monath, 2001). Although the mode of transmission has been understood since the early 1900's (Kruif, 1926; Graham, 1903), the etiology of dengue fever was not determined until the mid 1940s (Sabin, 1952).

Dengue viruses have a genome length of approximately 11,000 bases. The genome codes for three structural and seven nonstructural proteins. The genome is translated as a single polyprotein that is cleaved cotranslationally and posttranslationally by host and viral proteases into the functional viral proteins. (Henschal and Putnak, 1990, Monath and Heinz, 1996; Burke and Monath, 2001).

Quantitative Genetics

Quantitative genetics is the study of the genetics underlying quantitative traits. A Quantitative trait is a phenotypic characteristic that, in a population, has a distribution approximates that normality. There are three categories of quantitative traits:

Continuous traits are those for which there is a continuum of possible phenotypes. An example of a continuous trait is height or length.

Meristic traits are those for which there are discrete integer values, where the phenotype is within a given range and always represented by an integral value. An example of this is the number of bristles on the abdominal sternites in *Drosophila* (Breese and Mather, 1957), or the number of pectin teeth and comb scales on mosquito larvae (Wood, 1976).

Discrete traits are characteristics that are usually present or absent in an individual. A threshold must be met by a number of genotypic and environmental factors in order for a given phenotype to occur or not. An example would be disease susceptibility.

Quantitative genetics provides a means by which to determine how much of the variation in a phenotypic trait (phenotypic variance, V_P) is influenced by genetics (genotypic variance, V_G) and the environmental (environmental variance, V_E). It also provides insight into how individual alleles contribute to a phenotype of interest. V_G is further subdivided into additive (V_A), dominance (V_D), and epistatic (V_I) variance. Specific genotypes may even perform differently in different environments (genotype-environment interaction, V_{GxE}). Statistical methods have been developed to analyze all of these factors in the context of a defined system (Becker, 1992; Falconer and Mackay, 1996; Mather and Jinks, 1982). In general, the formula for genetic variance is:

$$V_G = V_A + V_D + V_I. \quad \text{(i)}$$

and the formula for phenotypic variance is:

$$V_P = V_A + V_D + V_I + V_E + V_{GxE}. \quad \text{(ii)}$$

An understanding of these variance components is critical to determining heritability. Determining heritability is essential for implementing any form of selection. If variation in a character is due primarily to environmental, dominance, or epistatic effects, it is not a good character upon which to base selection (Houle 1992). Broad sense heritability can be calculated, but takes into account all the genetic variance components (V_G), which can not make a very accurate prediction of actual heritability. A better estimate is the narrow-sense heritability, which takes into account only V_A and is calculated as:

$$h^2 = \frac{V_A}{V_P} \quad \text{(iii)}$$

Heritability is also a function of allele frequencies, and can be expressed as:

$$h^2 = \frac{\sum 2pq[a + (q - p)d]^2}{\sigma^2} \quad \text{(iv)}$$

where p and q are the frequencies of 2 alleles at a given locus, a and d are measures of the additive and dominance effects of the locus, and σ^2 is the variance of the trait in the population (Hartl and Clark, 1989). These values are summed over all the loci that affect the character, with each locus having potential for different values of p, q, a, and d. Allele frequencies can vary over distance and time, which makes heritability a population specific and not a species specific characteristic. The experimental unit should always be the population and not the species when trying to determine h^2.

Population Genetics of *Aedes aegypti:*

Ae aegypti populations differ genetically as shown by biochemical and molecular genetic markers and on vector competence for dengue viruses (Gubler *et al.,* 1979; Wallis *et al.,* 1983; Rosen *et al.,* 1985; Tabachnick *et al.,* 1985; Tardieux *et al.*, 1990; Vazeille-Falcoz *et al.,* 1999; Vazeille *et al.,* 2001; Bennet *et al.,* 2002). Genetic studies of allozymes in *Ae aegypti* populations from around the world identified 8 genetic groups (Tabachnick, 1991; Tabachnick and Powell, 1979). The study included two sylvan groups (*Ae aegypti formosus*) from West and East Africa and six domestic groups (*Ae aegypti aegypti*) from East Africa, Southeast USA, Southwest USA-Mexico, Central-South America, Caribbean, and Southeast Asia-Pacific.

The genetic variation found among various groups of *Ae aegypti* is relatively small when compared to other insect populations (Tabachnick and Powell, 1978 and 1979; Tabachnick *et al.,* 1979; Wallis *et al.,* 1983 and 1984). The subspecies that forms *Ae aegypti* may have separated when the Sahara began to dry about 4,000 years ago (McIntosh and McIntosh, 1981). This finding suggests that the division into sylvatic and domestic subspecies occurred recently in evolutionary history. Populations of *Ae aegypti* would have been stranded in North Africa and adapted to live in water storage containers and feed on humans (Tabachnick, 1991). Humans have been implicated as the primary factor in facilitating the spread and diversification of *Ae. aegypti* via commerce and exploration (Black *et al.,* 2002).

Recent studies have investigated gene flow among *Ae aegypti* populations at a more regional level (Apostle *et al.,* 1996; Vazeille *et al.,* 2001, Gorrochotegui-Escalante *et al.,* 2000; Gorrochotegui-Escalante *et al.,* 2002; Bennet *et al.*, 2002). A study in Puerto Rico (Apostle *et al.,* 1996) using RAPD-PCR markers examined gene flow in 16

mosquito groups from 6 cities. This study was able to find more than two times the heterozygosity (0.345) detected in previous studies using allozymes.

Studies on mosquitoes collected from Mexico (Gorrochotegui-Escalante *et al.*, 2000 and 2002) have suggested that gene flow among *Ae aegypti* populations varies among regions depending on the amount of human commerce and barriers to natural mosquito migration. This study demonstrated three genetically homogeneous *Ae aegypti* groups in Mexico corresponding to the geographic locations of Northeastern Mexico, the Pacific coast and the Yucatan. Northeastern collections were not strongly isolated by distance and showed moderate gene flow and low genetic diversity. Collections from the Yucatan had wide variation in genetic diversity, low levels of gene flow and were isolated by distance. Collections along the Pacific coast showed high genetic diversity and a large amount of gene flow. Gorrochotegui-Escalante *et al.* (2002) suggested that populations of mosquitoes across Mexico can be expected to remain genetically homogeneous within distances of 150 km, assuming there is no local selection or natural barriers to migration. These studies also suggest that genes released into a given population should spread rapidly into populations within a 150 km radius.

Quantitative Trait Loci (QTL):

Often, the actual gene(s) that define a phenotype are not known. Each of the gene(s) that condition the phenotype of interest is considered as a locus, and is known as a quantitative trait locus (QTL).

QTL mapping calculates the statistical probability that a give phenotypic trait is associated with a given marker on the linkage map. In this way, the majority of the genome can be discounted as not being associated with the phenotype of interest, limiting

the area in which a candidate gene may be found. This identifies genome areas important to the trait of interest when the actual genes are not known. In practice, QTL mapping requires a linkage map with polymorphic markers that cover the entire genome and a quantitative trait that shows variation within or between populations or strains (Falconer and Mackay, 1996; Mackay, 2001). Ideally, the marker loci are highly polymorphic, abundant, codominant, and neutral with respect to fitness and the quantitative trait. The concept of linking marker loci to phenotypic characters was first shown by Sax (1923) through crosses between inbred lines of bean plants (*Phaseolus vulgaris*).

Susceptibility to dengue viruses seems to be under the control of more than one genetic locus, and is influenced by environmental factors. Bosio *et al.* (1998) performed a quantitative genetic study *of Ae aegypti* susceptibility to midgut and disseminated infection with DEN-2 to determine how useful quantitative genetics would be in segregating genetic and environmental components of vector competence. Wallis *et al.* (1984) found that several genes of major effect appeared to control vector competence for flaviviruses in *Ae aegypti*, and Miller and Mitchell (1991) concluded that there were likely to be two loci of major effect involved in vector competence of *Ae aegypti* for flaviviruses. Recent studies using advance intercrossed families have detected up to 9 QTLs associated with MIB (Bosio *et al.,* 2000; Bennet *et al.,* 2004; Machorro-Gomez, 2004)

Association Mapping:

Finding a QTL is only a starting point in determining the genes that actually condition a quantitative trait. Association mapping is a method that is being used to evaluate phenotypic associations with genotypes at candidate loci (Long *et al.*, 1999;

Cardon *et al.*, 2001; Tabor *et al.,* 2002) and has been proposed as a general method for detecting loci for susceptibility to complex human diseases (Risch, 2000). Evaluating an association between markers of a candidate gene and a phenotype requires a sample of individuals from the field, each of whom has been genotyped at the marker loci and evaluated for the trait phenotype. For discrete traits, such as vector competence, the population sample is stratified according to infection phenotype, and an association between a marker and a phenotypic trait is revealed by a significant difference in marker allele or genotype frequencies among mosquitoes with alternative phenotypes (Cardon *et al.*, 2001).

Taken to a finer level, if the genotypes at all segregating sites in the candidate gene are determined, one or a few of them should correspond to the site causing the phenotypic effect. These sites have come to be known as quantitative trait nucleotides (QTNs) (Long *et al.*, 1998). The power to detect an association between the QTN and the trait phenotype is much higher than with QTL mapping studies. Genetic epidemiologists have rapidly embraced QTN mapping as a powerful tool for identifying heritable genetic predisposition to human disease. *Drosophila* researchers have mapped QTNs in ectodermally expressed genes that control bristle number in *Drosophila* (Mackay, 2001b). Domestic animal and plant breeders have mapped QTNs associated with increased yield or other desirable characters and then used these as selectable markers for more rapid crop and animal improvement (e.g. (Winter *et al.*, 2002)).

Four principal factors confound Association Mapping of QTNs, leading to both false positive and negative associations (Risch, 2000; Black *et al.*, 2001). These include 1) lack of independence among segregating sites due to linkage disequilibrium, 2)

inadvertent analysis admixtures of populations, 3) unadjusted significance thresholds when inference testing and, 4) loss of statistical power to detect a true QTN because of small sample size. In this dissertation I have developed software to assess these factors.

In theory, the amount of disequilibrium among proximal segregating sites is a function of the physical distance corresponding to $4Nc$ (Kruglyak, 1999; Long *et al.*, 1999), where N is the effective population size and c is the centiMorgan distance between a pair of segregating sites. Linkage disequilibrium can be identified from regression of all possible pairwise disequilibria between markers in a genomic region against their physical distances (Hudson, 1987). Disequilibrium analysis must be completed for every candidate gene and segregating site because $4Nc$ varies by at least an order of magnitude among different gene regions and in different species (Long *et al.*, 1999; Mackay 2001a, b).

Admixtures of populations with different frequencies of alleles in the candidate gene and different values of the phenotypic trait can generate false QTNs. False associations arising from admixtures of racially or socioeconomically stratified human populations are legendary in sociology and medical epidemiology (Gould, 1995; Risch, 2000; Cardon *et al.*, 2001). This problem can be alleviated by experimental designs that control for population structure (Risch, 2000; Black *et al.*, 2001). Subdividing sample populations into panmictic units can provide independent tests of a QTN and of genotype-phenotype associations. Assuming no epistasis and no QTN by environment interactions (but see Long *et al.* 1996, Lukens *et al.* 1999), a valid QTN should be repeatedly detected in all populations and the patterns of genotype-phenotype associations should be consistent.

Testing for trait associations of multiple segregating sites while simultaneously avoiding false positives requires a downward adjustment of the significance threshold when inference testing. Permutation tests developed for QTL mapping can be applied (Churchill *et al.*, 1994) as can traditional adjustments of experiment-wise error rates (e.g. Bonferroni corrections). The likelihood of avoiding false negatives (i.e. the power to detect a true QTN) depends not only on the numbers of individuals sampled, but also on the nature of genotype-phenotype associations (e.g. additive vs. dominant) and on the density of polymorphic markers (Long *et al.*, 1999; Luo *et al.*, 2000)

Landscape epidemiology and genetics

Landscape epidemiology involves the identification of geographical areas where disease is transmitted. It is an approach that involves the interactions and associations between elements of the physical and cultural environments. This multifactorial idea was first proposed by the Russian epidemiologist Pavlovsky (1966); his idea proposes that by knowing the vegetation and geologic conditions necessary for the maintenance of specific pathogens in nature, one can use the landscape to identify the spatial and temporal distribution of disease risk. Under this concept environmental elements, including elevation, temperature, rainfall, and humidity, influence the presence, development, activity, and longevity of pathogens, vectors, zoonotic reservoirs of infection, and their interactions with humans (Meade *et al.,* 1988). Vegetation type and distribution are also influenced by the environmental variables mentioned above, and can be expressed as landscape elements.

Examples of studies of disease that focused on identifying and mapping vector habitats, or assessing environmental factors related to vector habitat quality began in the

late 1970's (Barnes and Cibula, 1979; Wagner *et al.*, 1979; Cross *et al.*, 1984; Hayes *et al.*, 1985; Linthicum *et al.*, 1987, 1999; Welch *et al.*, 1989; Rogers and Randolph, 1991; Hugh-Jones *et al.*, 1992; Pope *et al.*, 1992). There are many studies that investigated the application of GIS and spatial analysis techniques to identify and map landscape elements that collectively define vector and human population dynamics related to disease transmission risk (Daniel and Kolar, 1990; Wood *et al.*, 1991; Glass *et al.*, 1992; Wood *et al.*, 1992b, 1992c; Dister *et al.*, 1993, 1997; Beck *et al.*, 1994, 1997, 2000). Hay *et al.* (1996) reviewed how data from meteorological satellites can be used to study arthropod vectors of disease.

The use of GIS has many implications for health scientists because it provides the ability to store, integrate, query, display, and analyze data from a molecular to continent wide resolution (Piglucci and Barbujani, 1991; Sokal and Thomson, 1998; Piertney *et al.*, 1998; Keyghobadi *et al.*, 1999; Sumner *et al.*, 2001; Castric *et al.*, 2001; Roach *et al.*, 2001). Field observations on environmental conditions (locally sensed data), including vegetation, water, and topography, can be combined in a GIS to direct interpretation of RS data and facilitate characterization of the landscape in terms of vector and pathogen prevalence.

The associations between disease risk variables (e.g., vector, pathogen, and reservoir host abundance and distribution) and environmental variables can be quantified using the spatial analysis capabilities of GIS. Landscape pattern analysis, combined with statistical analysis, allows us to define landscape predictors of disease risk that can be applied in larger regions where field data are unavailable. This makes GIS a powerful set

of tools for disease surveillance, predicting potential disease outbreaks, and targeting intervention programs.

All this computational power comes with a price. Ecological fallacy happens when individual-level causal processes are inferred from group level data. The likelihood of generating a fallacy greatly increases when uncritically analyzing large amounts of data with "point-and-click" software. A simple example of an ecological fallacy is: Betty goes to a school where the average math grade is A. Simon's school only averages a B in math. The ecological fallacy would lead us to conclude that Betty is a better mathematician than Simon, which is not necessarily the case.

"Prosecutor's fallacy" is an example of hypothesis testing clouded by ecological fallacy. Consider for instance the case of Sally Clark, who was accused in 1998 of having killed her first child at 11 weeks of age, then conceived another child and killed it at 8 weeks of age. The defense claimed that these were two cases of sudden infant death syndrome; neither prosecution nor defense offered any other explanations for the deaths. The prosecution had expert witness Sir Roy Meadow testify that the probability of two children in the same family dying from sudden infant death syndrome is about 1 in 73 million. To provide proper context for this number, the probability of a mother killing one child, conceiving another and killing that one too, should have been estimated and compared to the 1 in 73 million figure. Ms. Clark was convicted in 1999, resulting in a press release by the Royal Statistical Society which pointed out the mistake (Royal Statistical Society, 2001, news release).

Finding whether a person is innocent or guilty in mathematical terms is a form of binary classification. Let's start with a thought experiment. In a big bowl with one

thousand balls, some of them are made of wood and some of made of plastic. Wooden balls are white (100%), and only 1% of the plastic balls are white, the others are red. A ball is pulled out at random, and it is white. Given the information, how likely is it that the selected ball is made of wood? Is it 99%? No! Maybe the bowl contains only 10 wooden and 990 plastic balls. Without that information (the *a priori* probability), we cannot make any statement. In this thought experiment, we can think of the wooden balls as "accused is guilty", the plastic balls as "accused is innocent", and the white balls as "the observed evidence."

The fallacy can be analyzed using conditional probability: Suppose E is the observed evidence, and I stands for "accused is innocent". We know that $P(E|I)$ (the probability that the evidence would be observed if the accused were innocent) is tiny. The prosecutor wrongly concludes that $P(I|E)$ (the probability that the accused is innocent, given the evidence E) is comparatively tiny. However, $P(E|I)$ and $P(I|E)$ are quite different; using Bayes' theorem we see $P(I|E) = P(E|I) \cdot P(I) / P(E)$. Thus the *a priori* probability of innocence $P(I)$ and the overall probability of the observed evidence $P(E)$ need to be taken into account. If $P(I)$ is much larger than $P(E)$, then $P(I|E)$ can be large as well. The fallacy lies in the fact that the *a priori* probability of guilt is not taken into account. If this probability is small, then the only effect of the presented evidence is to increase that probability somewhat, but not necessarily dramatically.

The prosecutor's is not fallacious if the *a priori* odds of guilt are assumed to be 1:1. In a Bayesian approach to personal probabilities, where probabilities represent degrees of belief of reasonable persons, this assumption can be justified as follows: a

completely unbiased person, without having been shown any evidence and without any prior knowledge, will estimate the *a priori* odds of guilt as 1:1.

An ecological fallacy would occur in GIS analysis if groups are geographical areas and the results of the analysis of the aggregated data are assumed to reflect individual level relationships (Steel and Holt, 1997). The ecological fallacy is caused by several factors including the non-random composition of the groups and the correlations between individuals within the same group, among this correlation spatial autocorrelation (SA). This lack of independence can be identified through the autocorrelation tests described below.

It is important to emphasize that SA does not refer to the distribution of samples on a map, but the variables observed at these locations. Thus, there may be randomly distributed sampling locations supporting an SA variable, as well as spatially clustered samples supporting spatially random values of a variable (Sokal, 2004). Spatial autocorrelation affects the common assumption of most conventional statistical tests of independence among experimental units. Positive SA will increase the apparent significance of the *t*-tests, ANOVA's, G^2 or χ^2, and correlation or regression coefficients (Sokal, 2004). Testing for SA is relatively new and its use is not widespread when compared to time series analysis (temporal autocorrelation testing).

By proposing an *a priori* hypothesis it is possible to minimize the effect ecological fallacies could have in a study. For example, in the present study I collected mosquitoes based in observations made by Gorrochotegui *et al.* (2000, 2002) by analyzing *Ae aegypti* mitochondrial DNA data, which revealed three genetically homogenous populations with a break in gene flow somewhere in the state of Veracruz.

Data was contrasted among and between the three genetically homogenous areas (Gorrochotegui *et al.*, 2000, 2002). I decided to divide the Gulf of Mexico collection into North and South Atlantic groups (figure 20) using as a division line the Neovolcanic Axis (NVA). This constituted an *a priori* hypothesis since this geographic feature separates Mexico into Neartic and Neotropical regions (Cox and Moore, 1993). It is believed that around the NVA there is a biogeographical transition (Halffter, 1983; Escalante *et al.*, 2004). Based in the distribution of insects the Isthmus of Tehuantepec, just south of the NVA is the transition border Hallfter (1983). Escalante *et al.* (2004), based in mammal distribution, believe the border is dynamic and placed the border further north around the Tropic of Cancer.

SPATIAL POINT PATTERNS

A dot map is a representation of objects as points on a map to reveal spatial distributions. The dot map has been the most popular representation tools for centuries. Its widespread use is based on the simplicity and clarity with which it displays spatial distributions. Examples include the position of individual trees within a forest, the location of schools, towns, hospital, etc.

The search for specific spatial patterns in a dot map should be directed by *a priori* hypotheses arising from evident geographical barriers or corridors on maps and from previous studies, particularly studies that have developed theories and models. In this way, these techniques can be employed to search for specific types of patterns. Patterns are the result of some process at a given point in time and space, and the analysis of point processes ought to incorporate *a priori* hypotheses how the pattern evolved. Patterns can be viewed as a snapshot at a fixed time point. It is sometimes difficult to grasp the idea

that a continuous process can produce a pattern. Accordingly, when studying a specific pattern at time point x reflects the history of the processes that led to the pattern at x.

Most biological populations are distributed in a non-random pattern. Their spatial distribution is determined by the environment, including habitat availability, climatic conditions, food and breeding resources. It is far easier to study relatively sessile organisms that are completely dependent on their immediate environment. That is why plant ecologists were the first to recognize three distinct spatial distributions: random (Poisson), aggregated (negative binomial), and regular (uniform e.g. row crops, houses in a city block). This distinction among patterns is made for convenience, since in a realistic situation a continuum occurs among the three patterns.

Random Spatial Pattern

Objects are said to be distributed at random if the relative location of an individual object is independent of the location of any other object within an infinitely large area. Cottam *et al.* (1953) describe random spatial patterns as "an absence of fixed distances from one another, nor is there a marked gradient from regions where the population is dense to regions where it is rather sparse". Furthermore, "Any major portion of the population will have approximately the same number of objects as any other equal portion, but within each portion, there will be areas of greater than average density and there will be distinct bare areas". For a random spatial pattern the average distance between individual points is given by:

$$\frac{1}{2\sqrt{\rho}}, \qquad \text{(v)}$$

where ρ = density of the number of points per unit area.

For such process, the probability that a region Δ with area $|\Delta|$ contains x points is given by the Poisson distribution:

$$p(x) = \frac{e^{-\lambda|\Delta|}(\lambda|\Delta|)^x}{x!}, x = 0,1,2,3 \qquad \text{(vi)}$$

Aggregated Spatial Pattern

An aggregated population is one in which the objects occur in clumps of varying densities and sizes. Aggregated populations are very common in nature (Pielou, 1960). Many processes are potentially involved in the formation of spatially aggregated biological populations. Cole (1946) tried to explain aggregation in terms of a *contagious* process rather than by an *under*-dispersed process, which has been commonly mistaken to mean a " uniform" or "regular" spatial pattern (Ashby, 1935). Cole (1946) defines a "contagious" distribution as one in which "the presence of one or more organisms within a sample unit influences the probability of other organisms occurring in the same sample." Usually, contagiousness is positive, resulting in an excess of densely occupied samples, with few samples yielding single individuals or a smaller number of individuals than would be expected by chance. Pielou (1960) explained aggregation in terms of the reproductive and dispersal mechanisms of the species, and the site conditions of the study area. Some plants reproduce vegetatively (e.g. aspen, bunch-grass, etc.), such that offspring tend to cluster around parent plants.

For modeling aggregated populations we use the negative binomial distribution. We assume that the number of clusters follows a Poisson distribution, while the numbers of individuals are assumed to be independently distributed logarithmically. The density function of the negative binomial distribution is given by:

$$p(x) = \frac{(k+x-1)!}{x!(k-1)!} p^k (1-p)^k (1-p), x = 0,1,2,... \quad \text{(vii)}$$

Were p is the mean to the variance ratio and k is a measure of spatial heterogeneity in the population. For an aggregated spatial pattern we would expect the variance to be larger than the mean, thus $p < 1$. The parameter p is estimated as:

$$\hat{p} = \frac{\bar{x}}{s^2} \quad \text{(viii)}$$

Ae. aegypti exhibits a clear dependency on humans for food and breeding sites. Thus the aggregation of humans in and around villages and cities could be the underlying process that creates an aggregated distribution of the mosquito and ultimately dengue cases. Despite the apparent clustering of dengue human cases within households previously reported (Halstead *et al.*, 1969; Waterman *et al.*, 1982) there is no formal spatial research on the distribution pattern of *Ae. aegypti* and dengue with three exceptions.

Morrison *et al.* (1998) found that dengue cases in Florida, Puerto Rico clustered temporally over short periods of times. Within seven weeks of the first reported case, 9,000 people were affected. This presumably indicates that only a few, or the same, mosquitoes were causing the household cases while infected humans traveling within the town may have facilitated the rapid spread of infections. Lozano-Fuentes (2000) found spatial clustering within the city of Monterrey, Mexico. In the 1998 dengue epidemic, one municipality had close to 60% of the 967 cases. Within a household the father and the mother simultaneously exhibited dengue symptoms and a few days later children in the same household also showed symptoms.

Getis *et al.* (2003) examined the numbers of adult, pupae, larvae or pupae positive containers and all water-holding containers. Adults clustered strongly within houses and weakly to a distance of 30 meters beyond the household; clustering was not detected beyond 10 meters of a positive household. Larvae and pupae were not spatially structure beyond the household. Water containers were found ubiquitously, meaning all households were at equal risk at all times. However, Getis *et al.* (2003) found that as the number of water containers increased in a household so did the risk of *Ae. aegypti* pupae and adult infestations. This is not a surprising since larval distribution is independent among households. Households clusters in may arise from mosquito egg lying behavior and variation among households in cleaning practices (Apostol *et al.*, 1994).

Getis *et al.* (2003) found a small, but significant spatial structure of adult mosquitoes. The adult mosquitoes clustered within distances of ~10 meters and to a smaller extent out to 30 meters. This included neighboring houses, which agrees with mark-release-recapture experiments on the dispersal of adult *Ae. aegypti.* Most experiments concluded that the typical dispersal distance was less than 100 meters (Trpis and Hausermann, 1986; Edman and Scott, 1987; Anderson *et al.*, 1990; Rodhain and Rosen, 1997; Ordoñes *et al.*, 1997; Edman *et al.*, 1998; Morrison *et al.*, 1999; Thomas *et al.*, 2000a, b).

One the most interest results from Getis *et al.* (2003) is that mosquitoes tended to remain in proximity to the container in which they developed. Mosquito life stages were also significantly correlated. Larval clusters were correlated with pupal clusters and pupal clusters were correlated with adult clusters, however larval clusters were not

correlated with adult clusters. All this could arise from having water containers in all searched households, thus reducing the need for adult dispersal.

Regular Spatial Pattern

A regular, or uniform, spatial distribution is one in which the objects are evenly distributed. In biological populations it is extremely rare to find a uniformly distributed population. It has been thought that a uniform spatial pattern could occur in nature when there is a high degree of competition for space between individuals (Pielou, 1960). Cole (1946) describes this as a "*negative contagious*" process. An example of a possible cause of this process would be allelopathy (antagonism) in which some plants like the creosote bush (*Larrea tridentata*) and black walnut (*Juglans nigra*) produce toxic substances. Such toxic chemicals inhibit the growth of susceptible plants within the immediate vicinity increasing their density and zone of influence within a given area. The result is a uniform spatial pattern due to inhibition of individuals from growing close to one another.

The double Poisson model is an accurate representation of a uniformly distributed variable. This model is a modified Poisson distribution that combines a purely random process with a specified competitive process. One way of thinking about this model is in the context in which it was developed. Let us say that we divide a state into counties of different sizes. Next we want to assign towns (county seats and non-county seats) to the individual counties. The process is made "competitive" by limiting the number of county seats to one per county. Once a county has been allocated a county seat, it cannot be allocated another. The non-county seats are allocated randomly without this restriction

and, thus, conforming to the Poisson law. These two processes are combined to form a modified Poisson model that produces a pattern more regular than random.

The probability function is given by:

$$p(x) = \begin{cases} qe^{-\lambda} x = 0 \\ \dfrac{q\lambda^x e^{-\lambda}}{x!} + \dfrac{q\lambda^{x-1} e^{-\lambda}}{(x-1)!} x = 1,2,3,... \end{cases} \quad \text{(ix)}$$

where $0 > p > 1$ and $q = 1 - p$. The distribution depends on two parameter, λ and p. These can be estimated form the mean and the variance as follows:

$$\hat{p} = \sqrt{\bar{x} - s^2} \quad \text{(x)}$$

and

$$\hat{\lambda} = \bar{x} - \hat{p} \quad \text{(xi)}$$

where $\bar{x}$ and s^2 are the sample mean and the variance respectively. Both parameters vary from 0 to 1. The parameter p is interpreted as indicating the degree of regularity. If $p = 0$, $\lambda = \bar{x}$ the process is purely Poison, and when $p = 1$, there is a county in every county.

SPATIAL AUTOCORRELATION

Spatial autocorrelation occurs when the value (nominal, ordinal or interval) at one locality is dependent on the values of this variable at neighboring localities (Sokal and Oden, 1978). That is, the arrangement of values is not random (this randomness does not refer to the spatial distribution of the points). If similar values tend to be near each other it is said that there is positive spatial correlation. On the other hand, if different values tend to be near each other it is said that there is negative spatial correlation. For example, in a chess board you can randomly assign black or white to each grid box. The possibilities of obtaining a "checkerboard" pattern are very low; a checkerboard pattern

represents a negative correlation. Would you expect all boxes on one side to be black, and all boxes on the other to be white? No! That would be positive correlation.

Spatial autocorrelation is important when performing regression analysis and other tests. Suppose that we have multivariate spatial data and we are looking for relationships between the variables. We could use regression analysis to find relationships. However, if there is spatial autocorrelation in the residuals, it is possible to get unrealistic values for the significance and confidence limits for the coefficients. The reason is that ordinary least squares assumes that samples are all independent but if there is some kind of correlation between them, they are not independent. Thus, our estimate of the number of degrees of freedom would be too high, and our estimate of the standard errors would be too low. This could lead us to believe that some coefficients are significant when in fact they are not. Depending on the nature of the autocorrelation, and the spatial structure of the data, the autocorrelation can introduce slowly changing (i.e. low frequency) variability into the data and overestimate the slope.

For example, imagine we have entomological data from 10 neighborhoods spread all over the northeast of a city. The data includes variables such as median income, numbers of pupae positive containers, and dengue cases. We regress dengue cases against the other two variables only to realize that n=10 is not powerful enough to detect significant results. The data from 10 additional neighborhoods is added, but the chosen neighborhoods are immediately adjacent to the original 10. If incomes and cleaning practices hardly change from one neighborhood to the next, no additional information data is being added. If spatial autocorrelation is present in the variables, one would be doing the calculations as if there were n=20 cases, but in reality we only have 10

independent cases. Accordingly, one would be overestimating the number of degrees of freedom, getting unrealistic p values, and underestimating the standard errors of the coefficients.

One of the most common formats to display and infer SA is by using correlograms (Cliff and Ord, 1973, 1981; Sokal and Oden, 1978a). A correlogram consists of a series of estimated autocorrelation coefficients, calculated for different spatial relationships. The two common autocorrelation coefficients are Moran's I and Geary's c.

Moran's I

The test proposed by Moran (1950) is most commonly used in correlogram analysis. It is calculated as:

$$I = n\frac{\sum_{ij} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{W\sum_{i=1}^{n}(y_i - \bar{y})^2}, \qquad \text{(xii)}$$

where y_i is the value of the variable at the *ith* location, n is the number of points, w_{ij} indicates the spatial relationship of points i and j (see below), $\sum_{ij}$ indicates the double sum over all i and all j, $W = \sum_{ij} w_{ij}$, the sum of the values in the weight matrix. Moran's I is akin to the Pearson's correlation coefficient, as it is based on a product moment formulation. I will usually range from 1 to –1, with an expected value of $\frac{-1}{n-1}$; for large n, this is approximately zero. Positive values of I indicate positive spatial autocorrelation, negative values negative spatial autocorrelation.

Geary's c

A second metric of SA (a squared-difference coefficient) from Geary (1954) is:

$$c = (n-1)\frac{\sum_{ij} w_{ij}(y_i - y_j)^2}{2W\sum_{i=1}^{n}(y_i - \bar{y})^2}. \qquad \text{(xiii)}$$

Geary's c ranges from zero to infinity, with an expected value of 1 under no autocorrelation. Values from zero to one indicate positive spatial autocorrelation, values above 1 negative spatial autocorrelation. Geary's correlogram is essentially a standardized semivariogram. Moran's I and Geary's c usually yield similar interpretations; differences between them are described in Sokal (1979).

Correlograms

To compute the autocorrelation coefficients, one must choose a system to assign weights w_{ij} to connect the localities. Weights assignment is a crucial part of the testing. These weights may be based on a potentially elaborate connection matrix or simply on distance. Usually the weight matrix is binary, such that w_{ij} equals 1 if there is a connection between points i and j, and 0 if not. However, the weights do not have to be binary and other weighting schemes can be used; for example, the inverse of the distance (or squared distance) between points i and j. Rather than using a single set of weights to calculate an overall measure of spatial autocorrelation, one normally uses an ordered series of weights that depicts different spatial relationships among the localities. A common method is to use a series of classes representing successively larger distances (*e.g.,* 0 to 10 km, 10 to 20 km, etc.), in which a pair of points is given a weight of 1 if the distance between them is within the range of the class and estimate zero otherwise. The autocorrelation coefficients are then calculated separately for each distance class. The plot of the autocorrelation coefficients against distance is the correlogram.

When the raw data are expressed as distances (*e.g.* Paiwise F_{ST} or any other measure of genetic distance) rather than values at individual points, a correlogram can also be constructed using the Mantel coefficient (Sokal *et al.,* 1987). For this analysis, the Mantel correlation is calculated between the data distance matrix and the weight matrix. In this case, w_{ij} contains a 0 if points i and j are connected, and a 1 if they are not. The reversed weighting scheme is used because the data are expressed as distances; therefore high values (*i.e.,* 1) indicate dissimilarity, while low values (*i.e.,* 0) indicate similarity. The Mantel correlation is then plotted versus distance as in any other correlogram.

The significance of individual autocorrelation coefficients can be determined from their moments (Cliff and Ord, 1973, 1981; Sokal and Oden 1978a, b), while that of an entire correlogram is usually calculated using a Bonferroni procedure (Oden, 1984). A Bonferroni procedure is necessary to deal with the lack of independence of the distance classes; the same points are included within each class, although they are paired differently.

In summary, when using point data both Moran's I and Geary's c can be calculated, and in the case of distance data, Mantel correlation can be calculated. After choosing a data type, one must choose the connection scheme. Connections can be based either on distances and distance classes or on a single connections matrix.

Semivariograms

The semivariogram (often referred to as just as variogram) is the primary form of spatial analysis used in geostatistics (Isaaks and Srivastava, 1989). It is very similar to a correlogram analysis using Geary's c.

The semivariogram is calculated as:

$$g = \frac{\sum_{ij} w_{ij}(y_i - y_j)^2}{2W}, \qquad \text{(xiv)}$$

where y_i is the value of the variable at the *ith* location, n is the number of points, w_{ij} indicates the spatial relationship of points i and j, $\sum_{ij}$ indicates the double sum over all i and all j, $i \neq j$, and $W = \sum_{ij} w_{ij}$, the sum of the values in the weight matrix. Comparing this to the formula for Geary's c equation one sees that

$$c = g\frac{(n-1)}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{g}{Var(y)}. \qquad \text{(xv)}$$

Geary's c is simply a semivariogram standardized by the variance.

Anisotropy Analyses

Anisotropy analyses are specifically designed for two-dimensional data and are used to determine directional spatial patterns. Anisotropy refers to data that are heterogeneous in different directions. Rosenberg (2000) compares and contrasts some of these methods.

Bearing analysis. The bearing analysis (Falsetti and Sokal, 1993) is a method of determining the direction of greatest correlation between data distance and geographic distance. The procedure starts with two matrices: a data distance matrix V and a geographic distance matrix D. D is transformed into a new matrix, G_θ, by multiplying each entry of D by the squared cosine of the angle between a fixed bearing (θ) and that of each pair of points

$$G_{ij} = D_{ij} \cos^2(\theta - \alpha_{ij}), \qquad \text{(xvi)}$$

where G_{ij} is the ij^{th} element of matrix G, D_{ij} is the ij^{th} element of matrix D, and α_{ij} is the angular bearing of points I and j. If the two bearings point in the same direction ($\theta - \alpha_{ij} = 0$), the $\cos^2$ will equal one; if the bearings are at right angles to each other, the $\cos^2$ will equal zero and if the bearings are at opossite (180^o) angles to each other, the $\cos^2$ will equal -1. This transformation essentially weights each geographic distance by its alignment with a test direction.

The correlation between V and G_θ is calculated via a Mantel test and repeated for a set of θ. The maximal correlation occurs when large distances in the data are correlated with large geographic distances between points that lie roughly in the same direction as the fixed bearing. The direction of maximal correlation indicates the most likely direction of a gradient.

Mantel Test

The mantel test is based in the test proposed by Nathan Mantel in 1967 (Mantel, 1967). The test was created with the purpose of detecting temporal-spatial clusters in diseases. Since the test is very general and can use any kind distance matrices (expressed as dissimilarities) it provides a simple, but conservative, test of clustering. The test is used to estimate the association between two independent dissimilarity matrices describing the same set of entities and to test whether the association is stronger than one would expect form chance.

For whatever distance scales we may employ, let *Xij* be the spatial measure between *i* and *j*, and *Yij* the genetic distance measure. Our test is "Z":

$$Z = \sum_{i<j} \sum Xij Yij \qquad \text{(xvii)}$$

This can be compared with its null expectation; the null distribution of Z can be obtained by a finite population approach. We have *n* measures of genetic distance and *n* in geographic dimensions. The hypothesis of no clustering is the same as the genetic distances are matched at random with geographic distances, there being a total of *n!* equally probable set of matches.

Modified *t* test for correlation (CRH correction)

Clifford, Richardson and Hémon (1989) proposed a modified test of association based on the correlation coefficient or the covariance between two spatially autocorrelated processes (CRH modified *t* test for correlation). The test they proposed could be used for both lattice and non-lattice data and is based on the evaluation of an "effective sample size" that takes into account the spatial structure. When processes are positively autocorrelated the effective sample size is reduce under this method. The method used for evaluating the reduction is via and approximation of the variance of the correlation coefficient.

Clifford *et al.* (1989), with the use of theoretical and empirical data, showed that uncritical usage of the correlation coefficient between positively autocorrelated processes leads to an inflated proportion of Type I errors. They additionally showed that theoretically the magnitude of the inflation is consistent with a reduction in the effective sample size. The modified *t* test they proposed show that when "adjustments are made, the type I error rate is much closer to the nominal level". Additionally, the proposed method does not require a particular parametric model for the type of spatial autocorrelation. It manages equally well with data points spaced regularly or irregularly. The method is robust in the presence of anisotropy, thus isotropy need not be assumed.

Biplots

Principal component analysis is a multivariate technique for examining relationships among several quantitative variables. For the purposes of this dissertation, these variables are the frequencies of genetic markers in populations. Assume that we have ***j*** populations with ***i*** alleles measured in each population. These are entered into a ***j*** x ***i*** matrix (**Y**). The correlation in allele frequencies is calculated between each pair of populations to generate a symmetrical ***i*** x ***i*** correlation matrix (ρ).

A principal component is a linear combination of a standardized measure of allele frequencies multiplied by a coefficient called the eigenvector. There are ***i*** different principal components. The sum of the products of the eigenvectors and the standardized allele frequencies for each principal component is called the eigenvalue. The eigenvalue is proportional to the magnitude of the general correlation in allele frequencies among populations. If allele frequencies at different loci covary among populations then the eigenvalue will be large. However, if allele frequencies vary independently among populations then the eigenvalue will be small.

The principal component with the largest eigenvector is defined as the "first" principal component (PC1) while the principal component with the next largest eigenvector is defined as the "second" principal component (PC2) and so on. While there are ***i*** principal components, usually only the PC1 and PC2 and occasionally the third (PC3) components are necessary to explain most of the relationships among genetic markers.

Bivariate plots of the first two principal components are useful in summarizing relationships among populations especially when they account for >50% of the covariance among allele frequencies. In this chapter, we explore facets of principal

components known as biplot analysis (Gabriel, 1971). Biplot analysis is a multivariate statistical technique that graphically displays patterns of variation within and among individual loci.

Biplot analysis requires little additional computation and identifies the magnitude of variances and covariances among individual loci. There are three useful properties of biplots. The length of a biplot vector is proportional to the variance in an allele frequency among populations and indicates how much one locus varies among populations with respect to other loci in the dataset. Additionally the cosine of the angle between two biplot vectors equals the correlation in allele frequencies among populations at the two loci. If both vectors point in the same direction, then the angle between them approaches zero. The cosine of 0° is one, which is also the correlation in allele frequencies between those two loci. If both vectors form a right angle, cosine 90° = 0 indicating that allele frequencies at different loci vary independently among populations. If vectors point in opposite directions, cosine 180° = -1 indicating that alleles at different loci occur in exactly opposite frequencies in populations.

Summary

The purposes of this dissertation are two-fold: create novel tools to test the relationship between segregating sites in a nucleotide sequence and a given a phenotype (association mapping) and incorporate spatial analysis tools into the assessment of population genetics and vector competence data. The specific objectives of this dissertation are multiple and varied. They are represented in each of the chapters.

The next (second) chapter focuses in the mathematical concepts behind the association mapping computer programs. The chapter also contains a short example

clarifying the use and results from the software. The third chapter is an extensive association mapping study in which the association of segregating sites in the Early Trypsin gene with susceptibility to Dengue-2 virus in the mosquito *Aedes aegypti* is assessed. This chapter also illustrates the niche association mapping computer programs, presented in chapter 2. The 4th chapter contains a study of vector competence and gene flow of *Aedes aegypti* in the state of Veracruz which is an extension of the work started by Gorrochotegui-Escalante *et al.* (2000, 2001) and Bennett *et al.* (2002). This chapter serves as a preamble to the fifth chapter. Chapter 5 incorporates spatial analysis elements to assess population genetics, vector competence and dengue case distribution.

II PGenome: Association mapping software

Introduction

Association mapping evaluates the relationship between segregating sites at candidate genes with phenotypes (Tabor *et al.*, 2002) and has been proposed as a general method for detecting genetic markers associated with susceptibility to complex human diseases (Risch, 2000). Association mapping requires a sample of individuals from the field, each of whom has been genotyped at the candidate locus and evaluated for the trait phenotype. When a phenotype is found to be consistently statistically associated with particular segregating sites, the site is called a quantitative trait nucleotide (QTN) (Long and Langley, 1999).

Genetic epidemiologists have rapidly embraced QTN mapping as a powerful tool for identifying heritable genetic predisposition to human disease (Risch, 2000). *Drosophila* researchers have mapped QTNs in ectodermally expressed genes that control bristle number in *Drosophila* (Mackay, 2001b). Domestic animal and plant breeders have mapped QTNs associated with increased yield or other desirable characters and then used

these as selectable markers for more rapid crop and animal improvement (e.g. Winter *et al.*, (2002)).

Four principal factors confound association mapping of QTNs, leading to both false positive and negative associations (Risch, 2000; Black *et al.*, 2001). These include 1) lack of independence among segregating sites due to linkage disequilibrium, 2) inadvertent analysis of admixtures of populations, 3) unadjusted significance thresholds when inference testing and, 4) loss of statistical power to detect a true QTN because of small sample size.

Lack of independence among segregating sites

An analysis of linkage disequilibrium is necessary to determine the extent to which polymorphic sites in a candidate gene segregate independently in the populations under study. If segregating sites are in disequilibrium then it may not be possible to associate a single nucleotide with a phenotype, rather the phenotype will be associated with all sites that are in disequilibrium. In theory, the amount of disequilibrium among proximal segregating sites is a function of the physical distance corresponding to $4Nc$, where N is the effective population size and c is the CentiMorgan distance between a pair of segregating sites. Linkage disequilibrium can be identified from regression of all possible pairwise disequilibria between markers in a genomic region against their physical distances (Hudson, 1987). Disequilibrium analysis must be completed for every candidate gene and segregating site because $4Nc$ varies by at least an order of magnitude among different gene regions and in different species (Long *et al.*, 1999; Mackay, 2001a,b).

Analysis of admixtures of populations

Admixtures of populations with different frequencies of alleles in the candidate gene and different values of the phenotypic trait can generate false QTNs. False associations arising from admixtures of racially or socioeconomically stratified human populations are legendary in sociology and medical epidemiology (Gould, 1995). This problem can be alleviated by experimental designs that control for population structure (Risch, 2000; Black *et al.*, 2001). Subdividing sample populations into panmictic units can provide independent tests of a QTN and of genotype-phenotype associations. Assuming no epistasis and no QTN by environment interactions, a valid QTN should be repeatedly detected in all populations and the patterns of genotype-phenotype associations should be consistent.

Significance thresholds when inference testing

Testing for trait associations of multiple segregating sites while simultaneously avoiding false positives requires a downward adjustment of the significance threshold when inference testing. Permutation tests developed for QTL mapping can also be applied (Doerge and Churchill, 1996) as can traditional adjustments of experiment-wise error rates (e.g. Bonferroni corrections).

Loss of statistical power to detect a true QTN

The likelihood of avoiding false negatives (i.e. the power to detect a true QTN) depends not only on the numbers of individuals sampled, but also on the nature of genotype-phenotype associations (e.g. additive vs. dominant) and on the density of polymorphic markers (Long *et al.*, 1999; Luo *et al.*, 2000).

We have written a series of 8 programs: PGMerge, PGSort, PGFis, PGLD, PGMan, PGTheta, PGPW, PGCon (PGenome) that, respectively, assemble, sort, perform

an analysis of genotype frequencies at segregating sites, do two types of linkage disequilibrium analysis among segregating sites, perform association mapping based upon allele frequencies, and based upon genotypes. We illustrate the use of these programs using a phenotype/genotype dataset from the mosquito *Aedes aegypti*

Association mapping of vector competence for Dengue virus with trypsin genes in the mosquito Aedes aegypti

On a worldwide basis *Ae. aegypti* is the most common vector of yellow fever and dengue fever flaviviruses. Vector competence refers to the intrinsic permissiveness of an arthropod vector to infection, replication, and transmission of a virus. When a mosquito takes a viremic bloodmeal, the virus must first establish an infection in the mosquito midgut by overcoming a midgut infection barrier (MIB). After replication in the midgut epithelium, the virus must pass through a midgut escape barrier (MEB) and replicate in other tissues. Finally, the virus must infect the salivary glands and be shed in the saliva to be transmitted to the next vertebrate host. Throughout its worldwide range, *Ae. aegypti* exhibits continuous variation in vector competence for flaviviruses (Bennett *et al.*, 2002). We have shown that when *Ae. aegypti* females are fed a dengue infectious blood meal that also contains soybean trypsin inhibitor (STI) there is a 91-97% decrease in midgut DEN2 RNA copies (Molina-Cruz *et al.*, 2004). In the present study we test for associations between naturally occurring genotypes at individual segregating sites in the *Abundant Trypsin* gene and predisposition to midgut infection with DEN2 virus.

System and Methods

Dataset construction

Mosquitoes were collected from 13 locations from along the Pacific coast of Mexico. A total of 684 mosquitoes were articially bloodfed a meal containing DEN virus

(Bennett *et al.*, 2002). After 2 weeks, the DEN infection phenotype was scored (11=SUSC, 10=MEB, 00=MIB) in each mosquito. A PHENOTYPE datafile was created that contained for each mosquito a 5 letter label indicating the collection site, a two integer field containing an unique identification number followed by a two character field containing the DEN infection phenotype.

The entire *Abundant Trypsin* gene was amplified with the Polymerase Chain Reaction. Single Strand Conformation Polymorphisms (SSCP) analysis (Black and DuTeau 1997) was used to identify novel genotypes and each novel SSCP genotype was assigned a numerical label. Genotype number was entered into a three-integer field for each mosquito in the PHENOTYPE file.

All unique genotype sequences were assembled into a single GENOTYPE dataset and machine aligned using CLUSTAL W (Thompson *et al.*, 1994) such that each line contained a unique diploid, phase unknown, genotype. An 'A' indicated an A/A homozygote, 'C'=C/C, 'G'=G/G, 'T'=T/T. For heterozygotes, 'M' = A/C, 'R'=A/G, 'W'=A/T, 'S'=C/G, 'Y'=C/T, and 'K'=G/T. Gaps representing insertions or deletions (indels) were encoded as '-'=-/-, '1'= -/A, '2'= -/C, '3'= -/G, '4'=-/T.

PGMerge read the SSCP genotype number from the PHENOTYPE datafile and replaced it with the corresponding sequence from the GENOTYPE file. The final PHENOTYPE-GENOTYPE file contained 684 rows corresponding to each phenotyped mosquito and 795 columns of nucleotides. In total, 38 segregating sites were found.

PGFis - Analysis of Genotype Frequencies

Wright's F_{IS} summarizes the relationship between observed and expected heterozygotes at each segregating site (Wright, 1965).

$$F_{IS} = 1 - \left(\frac{H_{o(i)}}{H_{e(i)}} \right) \quad (1)$$

and $H_{o(i)}$ and $H_{e(i)}$ are respectively the observed and expected frequencies of heterozygotes containing nucleotide *i* at a segregating site. Weir and Cockerham's *f* is an estimator of F_{IS} that is unbiased by small or unequal sample sizes(Weir and Cockerham, 1984) and is calculated as:

$$f = \frac{b}{b + c} \quad (2)$$

where:

$$b = \frac{1}{2(\bar{n} - 1)} \left(\sum_y n_y H_{e(iy)} - \frac{(2\bar{n} - 1)}{2\bar{n}} \sum_y n_y H_{o(iy)} \right) \quad (3)$$

$$c = \frac{\sum_y n_y H_{o(iy)}}{2\bar{n}} \quad (4)$$

where $\bar{n} = \sum_y n_y / 31$, $H_{e(iy)} = 1 - \sum_i p_i^2$ and p_i is the frequency of nucleotide *i* at a segregating site and n_y is the size of collection *y*. *f* was calculated at each of the 38 segregating sites in each of the 13 geographic collections (Figure 1). There is a clear excess of homozygotes at most segregating sites. This pattern is consistent with an hypothesis of strong purifying selection acting in the functional region of *Abundant Trypsin* such that heterozygotes and homozygotes of the alternative nucleotide are selected against.

PGLD -Linkage disequilibrium analysis

Most traditional measures of linkage disequilibrium assume random union of gametes and knowledge of the gametic phase of alternative nucleotides at segregating sites. When working with genotypic data from natural populations, random mating

cannot be assumed and, when analyzing sequence data directly from the PCR products of a diploid individual, the phase of nucleotides at segregating sites is unknown. Cockerham and Weir (1977) suggested instead that a composite linkage disequilibrium coefficient, $\Delta_{ij,}$ be used.

A gamete with nucleotide *i* at segregating site A and nucleotide *j* at segregating site B is denoted as A_iB_j. The frequency with which an individual is formed by the union of gametes A_iB_j and A_kB_h is denoted $P_{ij}^{kh} (= P_{kh}^{ij})$ where *i* and *k* are alternative nucleotides at segregating site A and *j* and *h* are nucleotides at segregating site B. Dots denote sums of genotypic frequencies at both segregating sites. Therefore, the frequency with which A_iB_j gametes join with all other types of gametes to form zygotes is denoted by:

$$P_{ij}^{\cdot\cdot} = \sum_k \sum_h P_{ij}^{kh} \tag{5}$$

The frequency with which gametes bearing *i* and gametes bearing *j* join to form zygotes is denoted by:

$$P_{\cdot j}^{i\cdot} = \sum_k \sum_h P_{kj}^{ih} \tag{6}$$

The frequencies of nucleotides *i* and *j* are denoted p_i and p_j. If we assume that gametes can be sampled from a population and that all zygotes are formed by random union of gametes then linkage equilibrium exists if the frequency of individuals bearing nucleotides *i* and *j* acquired from a single parent ($P_{ij}^{\cdot\cdot}$) is equal to the frequency of individuals bearing *i* and *j* from different parents ($P_{\cdot j}^{i\cdot}$). Alternatively, if *i* and *j* appear together more often in parents than is predicted by their independent frequencies, then equilibrium does not occur. A statistical dependency D_{ij}^{w} (within individual

disequilibrium component) (Cockerham and Weir 1977) exists between the nucleotides and:

$$D_{ij}^{w} = P_{ij}^{\cdot\cdot} - P_{j}^{i\cdot} \tag{7}$$

If we assume random union of gametes in the population:

$$P_{ij}^{\cdot\cdot} = P_{j}^{i\cdot} = p_i p_j \tag{8}$$

If linkage equilibrium does not exist then:

$$P_{ij}^{\cdot\cdot} \neq P_{j}^{i\cdot} = p_i p_j \tag{9}$$

If union of gametes is not random:

$$P_{j}^{i\cdot} \neq p_i p_j \tag{10}$$

and a second statistical dependency D_{ij}^{b} (between individual disequilibrium component) exists between i and j where (Cockerham and Weir, 1977):

$$D_{ij}^{b} = P_{j}^{i\cdot} - p_i p_j \tag{11}$$

D_{ij}^{w} and D_{ij}^{b} sum to the coefficient of linkage disequilibrium D_{ij}. Therefore:

$$\begin{aligned} D_{ij} &= (P_{ij}^{\cdot\cdot} - P_{j}^{i\cdot}) + (P_{j}^{i\cdot} - p_i p) \\ &= P_{ij}^{\cdot\cdot} - p_i p_j \end{aligned} \tag{12}$$

Cockerham and Weir (1977) observed that, for D_{ij} to be an accurate measure of linkage disequilibrium there must be random union of gametes. Furthermore, estimation of D_{ij} requires that the frequency of A_iB_j gametes in the population be known. When working with genotypic data from natural populations, it is not reasonable to assume that random mating occurs. Furthermore when analyzing sequence data directly from the PCR products of a diploid individual the phase of nucleotides at segregating sites is

unknown. Cockerham and Weir suggested that a composite linkage disequilibrium coefficient, Δ_{ij}, be used instead of D_{ij}. Their formulation is:

$$\begin{aligned}\Delta_{ij} &= D_{ij}^{w} + 2D_{ij}^{b} \\ &= P_{ij}^{\cdot\cdot} + P_{j}^{i\cdot} - 2p_i p_j\end{aligned} \tag{13}$$

Δ_{ij} contains $P_{ij}^{\cdot\cdot}$ and $P_{j}^{i\cdot}$ so that an estimate of the frequency of either is not required. Furthermore, Δ_{ij} is not biased by departures from random mating in the population. Note that, if $D_{ij}^{b} = 0$, $\Delta_{ij} = D_{ij}$.

PGLD sums the number of times, T_{ij}, that a nucleotide *i* at one segregating site appears with another nucleotide *j* at a second segregating site in the same individual and therefore:

$$T_{ij}/N_y = P_{ij}^{\cdot\cdot} + P_{j}^{i\cdot} \tag{14}$$

where N_y is the number of collections sampled. So that an unbiased estimate of Δ_{ij} is:

$$\Delta_{ij} = (N_y/(N_y - 1))((T_{ij}/N_y) - 2p_i p_j) \tag{15}$$

We assumed that the 13 collections made in this study represent a series of subdivided populations and therefore that linkage disequilibrium could result from both random (nonsystematic) and epistatic (systematic) causes. When migration among subdivided populations is limited, novel mutations and random drift may generate different unique nucleotide combinations and create disequilibrium. Alternatively, systematic disequilibrium that is consistent in all subdivided populations is assumed to arise through epistatic natural selection (Lewontin and Kojima 1960).

To differentiate between random and epistatic causes of disequilibrium, (Ohta, 1982a;b) devised a method for partitioning the variance of dilocus disequilibrium

coefficients into within- and between-subpopulation components. Ohta's theory is based on the assumption that when selection produces specific nucleotide combinations, they should appear consistently among subpopulations. The variance in the observed frequency of nucleotide combinations should therefore be small with respect to the variance expected on the hypothesis of random drift. Ohta's five "D-statistics" describe the variance in observed and expected frequencies of nucleotide combinations. Black and Krafsur (1985) integrated Δ_{ij} into Ohta's D-statistics. PGLD computes Ohta's five D-statistics for all possible pairs of segregating sites and reports whether patterns are consistent with 1) nonsystematic disequilibrium, 2) systematic disequilibrium, or 3) unequal systematic disequilibrium.

We follow Ohta's notation and label the statistics with subscripts I, S, and T (I-Individuals, S-Subpopulations, T-Total population). The variance of disequilibrium in the total population D_{IT}^2, is:

$$D_{IT}^2 = \sum_s \left(\sum_i \sum_j \left(T_{ij,s} - 2\bar{p}_i \bar{p}_j \right)^2 \right) / x \qquad (16)$$

where $T_{ij,s}$ is the frequency that nucleotides i and j appear together in individuals in subpopulation s, $\bar{p}_i$ and $\bar{p}_j$ are the weighted mean nucleotide frequencies and x is the number of subpopulations. D_{IT}^2 indicates when i and j appear together in individuals more often than is predicted by their independent frequencies in the overall population.

The variance of disequilibrium in subpopulations, D_{IS}^2, is calculated as:

$$D_{IS}^2 = \sum_s \left(\sum_i \sum_j \left(T_{ij,s} - 2p_{is} p_{js} \right)^2 \right) / x \qquad (17)$$

where p_{is} and p_{js} are the frequencies of i and j in subpopulation s. D_{IS}^2 is $\Delta_{ij,s}$ squared and summed over all nucleotide pairs. It is therefore a measure of the average squared disequilibrium in subpopulations. The variance in the expected frequency of individuals bearing i and j, D_{ST}^2 is:

$$D_{ST}^2 = \sum_s \left(\sum_i \sum_j \left(p_{is} p_{js} - 2\bar{p}_i \bar{p}_j \right)^2 \right) / x \qquad (18)$$

D_{ST}^2 is a predictor of the variance of disequilibrium expected under the hypothesis of genetic drift. If subpopulations are differentiated, nucleotide combinations will have been established independently among subpopulations and D_{ST}^2 will be greater than D_{IS}^2. Alternatively, systematic epistatic selection among subpopulation will cause the frequency of nucleotide combinations to converge so that D_{ST}^2 is deflated and less than D_{IS}^2. $D_{ST}^{'2}$ is the variance of disequilibrium in the total population:

$$D_{ST}^{'2} = \sum_i \sum_j \left(\bar{T}_{ij} - 2\bar{p}_i \bar{p}_j \right)^2 \qquad (19)$$

where $\bar{T}_{ij}$ is the mean frequency with which i and j appear together in individuals in the overall population. Note that $D_{ST}^{'2}$ is Δ_{ij} for the whole population squared and summed over all nucleotide pairs. $D_{ST}^{'2}$ is therefore an index of disequilibrium in the total population. $D_{IS}^{'2}$ is the variance in the observed frequency that i and j appear together in individuals in subpopulations.

$$D_{IS}^{'2} = \sum_s \left(\sum_i \sum_j \left(T_{ij,s} - \bar{T}_{ij} \right)^2 \right) / x \qquad (20)$$

If random drift caused the variance in the frequency of nucleotide combinations, D'^2_{IS} will exceed D'^2_{ST}. Uniform selection among subpopulations will increase D'^2_{ST} and deflate the variance D'^2_{IS}. Ohta (1982a) showed that,

$$D^2_{IT} = D'^2_{IS} + D'^2_{ST} \quad (21)$$

Equation (21) therefore allows an evaluation of the relative contributions of systematic and nonsystematic processes to disequilibrium.

Disequilibrium is nonsystematic when caused by random drift such that:

$$D^2_{IS} < D^2_{ST}, \text{ and } D'^2_{IS} > D'^2_{ST} \quad (22).$$

Alternatively, if disequilibrium arises through systematic processes for specific nucleotide pairs in subpopulations then,

$$D^2_{IS} > D^2_{ST}, \text{ and } D'^2_{IS} < D'^2_{ST} \quad (23)$$

Unequal systematic disequilibrium arises when selection for specific nucleotide pairs occurs but only in a few subpopulations:

$$D^2_{IS} > D^2_{ST}, \text{ and } D'^2_{IS} > D'^2_{ST} \quad (24)$$

D^2_{IS} is greater than D^2_{ST} because disequilibrium in subpopulations is greater than that expected by random drift. But D'^2_{IS} is greater than D'^2_{ST} because the variance in frequencies with which nucleotides appear together in individuals will be greater than disequilibrium in the total population.

PGLD output outcomes can be easily displayed in a half matrix (Figure 2). Among the 13 collections, a transition at site 234 was in systematic disequilibrium with a transition at site 249. Transitions at segregating sites 285 and 297 were in systematic disequilibrium with one another and with a transversion at site 300 and a transversion at

site 401. A transversion at site 468 and a transversion at site 619 were also in disequilibrium.

An additional test of disequilibrium is performed in PGLD by regressing D'^2_{ST} on the physical distance among segregating sites (Figure 3). Under a hypothesis of disequilibrium, sites in proximity to one another should have greater disequilibrium than sites further away and there should be a negative slope. The significance of this regression analysis was evaluated (Mantel, 1967). The slope is negative, and some adjacent sites had large disequilibrium coefficients. This is consistent with a general pattern of disequilibrium even though the Mantel probability was 1.000. Taken together, the Ohta and regression analyses suggest that the most disequilibrium in *Abundant trypsin* is nonsystematic but that systematic disequilibrium occurs in four gene regions.

PGTheta - Association mapping with allele frequencies

A series of statistical tests of association between allele frequencies and phenotypes are performed in PGTheta. For all analyses, the PHENOTYPE-GENOTYPE file was partitioned according to phenotype using PGSort. For association mapping of MIB, pi in SUSC and MEB mosquitoes was compared to pi in MIB mosquitoes. For each alternative nucleotide at a segregating site, Wright's FST (Wright, 1965) was estimated among mosquitoes with different phenotypes as:

$$F_{STi} = \frac{s_{pi}^2}{\bar{p}_i(1-\bar{p}_i)} \tag{25}$$

where: $s_{pi}^2 = \frac{\sum_d n_d SS_{id}}{\bar{n}}$, $SS_{id} = (p_{id} - \bar{p}_i)^2$ and $\bar{p}_i = \frac{\sum_d n_d p_{id}}{2\bar{n}}$ where n_d is the number of mosquitoes in a phenotypic group (*SUSC +MEB*, *MIB*, or *SUSC*) and

$$\bar{n} = \sum_d n_d / 2 \tag{26}$$

A contingency χ^2 was then computed for each allele at a segregating site as:

$$\chi_i^2 = 2\mathrm{N}\sum_i (s_{pi}^2 / \bar{p}_i) \tag{27}$$

with degrees of freedom = (number of alternative nucleotides -1) and $\mathrm{N} = \sum_d n_d$. An estimator of F_{STi} that is unbiased by small or unequal sample sizes is θ_i (Weir and Cockerham, 1984) computed by PGTheta as:

$$\theta_i = \frac{a_i}{a_i + b_i + c_i} \tag{28}$$

where:

$$a_i = \frac{1}{n_c}\left(\sum_d n_d SS_{id} - \frac{1}{2(\bar{n}-1)}\right)\left(\sum_d n_d \mathrm{H}_{e(id)} - \frac{1}{2}\sum_d n_d \mathrm{H}_{o(id)}\right) \tag{29}$$

$$b_i = \frac{1}{2(\bar{n}-1)}\left(\sum_d n_d \mathrm{H}_{e(id)} - \frac{(2\bar{n}-1)}{2\bar{n}}\sum_d n_d \mathrm{H}_{o(id)}\right) \tag{30}$$

$$c_i = \frac{\sum_d n_d \mathrm{H}_{o(id)}}{2\bar{n}} \tag{31}$$

where $n_c = 2\bar{n} - \left(\sum_d n_d^2 / 2\bar{n}\right)$ and $\mathrm{H}_{e(id)} = 2\,p_{id}\,(1\text{-}\,p_{id})$

which is the expected proportion of mosquitoes heterozygous for nucleotide *i* in each phenotypic group *d* and $\mathrm{H}_{o(id)}$ is the observed proportion of heterozygous mosquitoes for nucleotide *i*.

The consistency of θ_i is evaluated following the protocol of Doerge and Churchill (1996). PGTheta permutes the original dataset by randomly assigning the genotype of one mosquito to another mosquito. After all genotypes are shuffled, θ_i is estimated between phenotypic groups and stored in memory. After 10,000 permutations, all θ_i are sorted. The 9,500th and 9,900th largest values respectively define the 95% and 99%

thresholds at each segregating site. Experiment-wise thresholds were computed by storing the θ_l among all segregating sites in one permutation and the 95% and 99% experiment-wise thresholds were the 9,500th and 9,900th largest values among all permutations (Figure 4). Two putative QTNs at sites 285 and 300 were detected in the Pacific.

PGCon- Association mapping based upon genotypes

PGCon performs a series of contingency χ^2 tests of the association between specific genotypes at segregating sites and phenotypes. There were 15 potential genotypes at each segregating site. Genotype association mapping of MIB compares the frequency of a genotype in SUSC + MEB mosquitoes to that in MIB mosquitoes. At each segregating site a contingency χ^2 value was calculated as:

$$\chi^2 = \sum_g \sum_d \frac{(g_{gd} - \hat{g}_{gd})^2}{\hat{g}_{gd}} \tag{32}$$

where g_{gd} is the observed frequency of a genotype g in phenotypic group d and $\hat{g}_{gd}$ is the expected frequency calculated from the independent frequencies of g_g and the proportion of mosquitoes of each phenotype. There are (number of genotypes-1) degrees of freedom. Because of the large number of segregating sites tested simultaneously, Bonferroni adjusted significance thresholds were used:

$$\alpha` = \alpha/r \tag{33}$$

where r is the number of segregating sites. Figure 5 indicates that there were significant associations between genotypes at segregating site 681 using the 95% Bonferroni adjusted significance threshold. Thus there was no additional support for the putative QTNs at sites 285 and 300 that were detected by PGTheta (Figure 4).

PGCon also estimates the minimal effect (m) that a nucleic acid substitution or a difference in genotype could exert on a phenotype and still be statistically significant in association mapping. This is especially important given the variable sample sizes among 1) geographic regions, 2) phenotypic groups 3) nucleotides and 4) genotypes at each segregating site. m is iteratively estimated. For allelic effects we considered a general model where the frequency of susceptible (+) or refractory (-) mosquitoes in a population were respectively (f_+) or (f_-). Nucleotide frequencies in a *MIB* or *MEB* comparison are p_i. We adopted a general model where the number of mosquitoes of a phenotype d and having an nucleotide i is:

$$n_{id} = 2N(f_d p_i + m) \tag{34}$$

where N is the total sample size in a comparison and m is the minimal effect of a nucleic acid substitution or an alternative genotype. Values of n_{id} were calculated for all observed combinations of i and d in a geographic region such that:

		Nucleotide		
		I	*K*	
Phenotype	*Susceptible*	$2N(f_+p_i + m)$	$2N(f_+p_k - m)$	$n_+/2N = f_+$
	Refractory	$2N(f_-p_i - m)$	$2N(f_-p_k + m)$	$n_-/2N = f_-$
		$n_i/2N = p_i$	$n_k/2N = p_k$	1.0

(35)

We initially assumed that f_+ and f_- are unaffected by i such that $m = 0$ and $n_{id} = 2Nf_d p_i$. For each segregating site we used the estimates of f_+, f_-, p_i, p_k, and N observed in a *MIB* or *MEB* comparison. By adding or subtracting m in each cell, the marginal probabilities remain constant. We then iteratively permuted m from 0.00001 to 1.0000. With each increment a χ^2 with 1 degree of freedom was estimated where :

$$\chi^2 = \sum_d \sum_i \frac{((2N(f_d p_i \pm m)) - 2Nf_d p_i)^2}{2Nf_d p_i} \quad (36)$$

the value of m was recorded when $\chi^2 = 3.84$ ($\alpha = 0.05$) and $\chi^2 = 6.64$ ($\alpha = 0.01$).

A different procedure was applied with 3 nucleotides:

		Nucleotide			
		I	*K*	*L*	
Phenotype	*Susceptible*	2N(*f*+*pi* +*m*)	2N*f*+*pk*	2N(*f*+*pl* –*m*)	N+/2N = *f*+
	Refractory	2N(*f*-*pi* – *m*)	2N*f*-*pk*	2N(*f*-*pl*+*m*)	n-/2N = *f*-
		ni/2N = *pi*	*nk*/2N = *pk*	*Nl*/2N = *pl*	1.0

(37)

where i was the nucleotide with the highest frequency, k was the nucleotide with the intermediate frequency and l was the nucleotide with the lowest frequency and m was iteratively estimated as above.

The same procedure as outlined above for estimating m with 2 alternative nucleotides was applied for estimating m with 2 alternative genotypes (usually when an alternative nucleotide was rare) except that N rather than 2N was used in all calculations. For three genotypes the same procedure as outlined above for estimating m with 3 alternative nucleotides was used except that i was the most common homozygous genotype, k the most common heterozygous genotype, and l was the homozygous genotype for the less common nucleotide and N was used in all calculations. n_{id} became small in some cells and artificially inflated $\chi 2$ when there were more than 2 nucleotides at a segregating site. Under these circumstances, the three most common genotypes were identified and i became the most common genotype, k the next most frequent genotype and l the third most frequent genotype. The remaining genotypes were discarded but i, k, and l genotypes never were less than 97% of N. Sample sizes were large in the Pacific

regions, and *m* never exceeded 2.5% (Figure 6). Thus in general our experimental design had the power to detect effects as small as 2.5%. Our experimental design and sample sizes would not have detected allelic and genotypic substitution effects that caused a smaller percent change in phenotype.

We conclude that there is little consistent evidence for an association between nucleotide substitutions in the *Abundant Trypsin* gene and MIB in *Ae. aegypti* populations along the Pacific coast of Mexico.

III Association mapping of segregating sites in the Early Trypsin gene and susceptibility to Dengue-2 virus in the mosquito Aedes aegypti

Introduction

On a worldwide basis the mosquito *Aedes aegypti* is the most common vector of yellow fever and dengue fever flaviviruses (Gubler, 2002). Vector competence refers to the intrinsic permissiveness of an arthropod vector to infection, replication, and transmission of a virus (Hardy, 1988; Woodring *et al.*, 1996). When a mosquito takes a viremic bloodmeal, the virus encounters several barriers to infection. First, the virus must establish an infection in the mosquito midgut by overcoming a midgut infection barrier (MIB). After replication in the midgut epithelium, the virus must pass through a midgut escape barrier (MEB) and replicate in other tissues. Finally, the virus must infect the salivary glands and be shed in the saliva to be transmitted to the next vertebrate host.

Throughout its worldwide range, *Ae. aegypti* exhibits continuous variation in vector competence for flaviviruses (Gubler *et al.*, 1979; Tabachnick *et al.*, 1985). Most

recently we estimated MIB and MEB rates in 1,470 *Ae. aegypti aegypti* from 24 collections throughout Mexico. The MIB rate among collections ranged from ~10-60% and from ~5-45% in MEB (Bennett *et al.*, 2002). Variation in midgut susceptibility to infection with flaviviruses between individual *Ae. aegypti* is quantitative, with population variation in the amount of virus in the midgut and head approximating a normal distribution (Bosio *et al.*, 1998). We have verified the quantitative nature of vector competence by mapping, to date, 9 QTL that appear to condition transmission of DEN2 in *Ae. aegypti* (Bosio *et al.*, 2000; Bennett *et al.*, 2004; Gomez-Machorro *et al.*, 2004). However, if we are to determine the molecular genetic basis of susceptibility to flaviviruses among individual *Ae. aegypti* we need to go beyond statistical descriptions to identify and determine the properties of the individual genes underlying variation in vector competence.

Little is known of the early events of flavivirus infection in insect midgut epithelial cells but for other arboviruses, (e.g. *Bunyaviruses* and *Orbiviruses*), there is a prerequisite for proteolytic processing of virion surface proteins for efficient vector-midgut cell interaction (Mellor *et al.*, 1981; Ludwig *et al.*, 1989; Ludwig *et al.*, 1991). In *Aedes* spp. a small amount of a proteolytic enzyme called Early Trypsin (ET) is secreted into the lumen of the midgut immediately following ingestion of blood or any other pool of free amino acids. ET is part of a unique signal transduction system (Noriega *et al.*, 2001) in which a large pool of transcribed ET mRNA resides in the midgut of newly eclosed adults and translation is induced in the presence of free amino acids (Noriega *et al.*, 2001). The function of ET may be to "taste" the incoming meal to determine if there is sufficient protein to support a gonadotrophic cycle. If so, the signal transduction

pathway activates the transcription and translation of other proteins secreted in the midgut to commence blood meal digestion.

It has been shown that when *Ae. aegypti* females are fed a dengue infectious blood meal that also contains soybean trypsin inhibitor (STI) there is a 91-97% decrease in midgut DEN2 RNA copies (Molina-Cruz *et al.*, 2004). STI treatment also resulted in slower DEN2 replication in the midgut, less DEN2 E protein expression, and decreased dissemination to the thorax. A second uninfected blood meal, 7 days after the STI-treated infectious meal, significantly increased DEN2 replication in the midgut and recovered oogenesis, suggesting that the lower viral infection caused by STI was in part due to a nutritional effect. Mosquitoes fed DEN2 digested *in vitro* with bovine trypsin (before STI addition) exhibited a transient increase in midgut DEN2 4 days post-infection. Blood digestion and possibly DEN2 proteolytic processing, mediated by midgut trypsins, influence the rate of DEN2 infection, replication and dissemination (Molina-Cruz *et al.*, 2004).

We wished to test the hypothesis that segregating sites in *Early Trypsin* are associated with natural variation in midgut susceptibility to DEN2 virus. Cosegregation of intragenic recombinant genotypes in a candidate gene with the phenotype provides direct genetic proof that a candidate gene constitutes a QTL (e.g. the tomato apoplastic invertase gene and fruit sugar content QTL (Fridman *et al.*, 2000). However, there is very low resolution of base pairs/cM in the *Ae. aegypti* genome (1.0 – 3.4 Mbp/cM (Brown *et al.*, 2001) and, not surprisingly, no *Early Trypsin* recombinants have been detected in F_7 or F_8 advanced intercross lines (Bennett *et al.*, 2004). Functional complementation, in which a phenotype is rescued in transgenic organisms, is another

tool for gene confirmation (Cormier *et al.*, 1997; Frary *et al.*, 2000; Mackay, 2001b). However defined deletion stocks do not exist for *Ae. aegypti* and transgenesis in *Ae. aegypti* is not currently a routine procedure. Association mapping is an alternative method that is being used to evaluate phenotypic associations with genotypes at candidate loci (Long and Langley, 1999; Cardon and Bell, 2001; Tabor *et al.*, 2002) and has been proposed as a general method for detecting loci for susceptibility to complex human diseases (Risch, 2000).

Evaluating association between markers at a candidate gene and a phenotype requires a sample of individuals from the field, each of whom has been genotyped at the marker loci and evaluated for the trait phenotype. For discrete traits, such as vector competence, the population sample is stratified according to infection phenotype, and an association between a marker and a phenotypic trait is revealed as a significant difference in marker allele or genotype frequencies among mosquitoes with alternative phenotypes (Cardon and Bell, 2001).

Taken to a finer level, if the genotypes at all segregating sites in the candidate gene are determined, one or a few of them should correspond to the site causing the phenotypic effect. These sites have come to be known as quantitative trait nucleotides (QTNs) (Long *et al.*, 1998). The power to detect an association between the QTN and the trait phenotype is much higher than with QTL mapping studies. Genetic epidemiologists have rapidly embraced QTN mapping as a powerful tool for identifying heritable genetic predisposition to human disease. *Drosophila* researchers have mapped QTNs in ectodermally expressed genes that control bristle number in *Drosophila* (Mackay, 2001b). Domestic animal and plant breeders have mapped QTNs associated

with increased yield or other desirable characters and then used these as selectable markers for more rapid crop and animal improvement (e.g. (Winter *et al.*, 2002).

Four principal factors confound Association Mapping of QTNs, leading to both false positive and negative associations (Risch, 2000; Black *et al.*, 2001). These include 1) lack of independence among segregating sites due to linkage disequilibrium, 2) inadvertent analysis of admixtures of populations, 3) unadjusted significance thresholds when inference testing and, 4) loss of statistical power to detect a true QTN because of small sample size.

An analysis of linkage disequilibrium is necessary to determine the extent to which polymorphic sites in a candidate gene segregate independently in the populations under study. If segregating sites are in disequilibrium then it may not be possible to associate a single nucleotide with a phenotype, rather the phenotype will be associated with all sites that are in disequilibrium. In theory, the amount of disequilibrium among proximal segregating sites is a function of the effective recombination rate ($4Nc$) (Kruglyak, 1999; Long *et al.*, 1999), where N is the effective population size and c is the centiMorgan distance between a pair of segregating sites. Disequilibrium analysis must be completed for every candidate gene and segregating site because $4Nc$ varies by at least an order of magnitude among different gene regions and in different species (Long *et al.*, 1999; Mackay, 2001b; Mackay, 2001a).

Admixtures of populations with different frequencies of alleles in the candidate gene and different values of the phenotypic trait can generate false QTNs. False associations arising from admixtures of racially or socioeconomically stratified human populations are legendary in sociology and Medical Epidemiology (Gould, 1995; Risch,

2000; Cardon *et al.*, 2001). For example, we might wish to identify the genes that cause differences in cuticle pigmentation between *Ae. a. aegypti* and *Ae. a. formosus*. There are many genes at which alleles are differentiated between the two subspecies because of geographic distance alone. Therefore, if dark and light mosquitoes were compared without respect to geographic origin, almost any gene would be associated with cuticle pigmentation. This problem can be alleviated by experimental designs that control for population structure (Risch, 2000; Black *et al.*, 2001). Subdividing sample populations into panmictic units can also provide independent tests of a QTN and of genotype-phenotype associations. Assuming no epistasis and no QTN by environment interactions (but see(Long *et al.*, 1996; Lukens and Doebley, 1999), a valid QTN should be repeatedly detected in all populations and the patterns of genotype-phenotype associations should be consistent.

Testing for trait associations of multiple segregating sites while simultaneously avoiding false positives requires a downward adjustment of the significance threshold when inference testing. Permutation tests developed for QTL mapping can be applied (Churchill and Doerge, 1994) as can traditional adjustments of experiment-wise error rates (e.g. Bonferroni corrections). The likelihood of avoiding false negatives (i.e. the power to detect a true QTN) depends not only on the numbers of individuals sampled, but also on the nature of genotype-phenotype associations (e.g. additive vs. dominant) and on the density of polymorphic markers (Long *et al.*, 1999; Luo *et al.*, 2000).

In this chapter we test for associations between naturally occurring genotypes at individual segregating sites in the *Early Trypsin* gene and predisposition to midgut infection with DEN2 virus. We begin by outlining the distribution of segregating sites

and the distribution of polymorphisms along the *Early Trypsin* gene and among collections. Based upon a previous studies of gene flow in Mexico (Gorrochotegui-Escalante *et al.*, 2000) we know that *Ae. aegypti* in Mexico exists as panmictic populations in four geographic regions. We demonstrate how ignoring this breeding structure can lead to false QTNs. Next we perform an analysis of linkage disequilibrium and show that while disequilibrium occurs in all four regions there is very little systematic disequilibrium arising through epistasis. Association mapping is then performed based upon allele frequency differences among phenotypes using a permutation procedure (Churchill *et al.*, 1994) and based upon genotype frequency differences among phenotypes using a Bonferroni correction. Finally, we perform a series of power calculations for each segregating site to define the minimal effects that were detectable in our study. This allowed us to place quantitative limits on the ability of the current study to detect QTN in *Ae. aegypti Early Trypsin*.

Materials And Methods

Mosquito Collection and Susceptibility Phenotype Determination

Mosquitoes were collected as eggs from the sites in Mexico and the United States (Table 2). Eggs were hatched in tap water and the emerged larvae were supplied with dried liver powder and ground Tetramin fish food.

Dengue serotype 2 JAM-1409 strain of dengue was used in all experiments. Feeding, and DEN2 infection protocols follow (Bennett *et al.*, 2002). When viral antigen was not detected in head tissues, the respective abdomens, which had been stored at -70°C, were assayed by immunofluoresence for virus antigen. Detection of virus antigen in midgut tissue revealed a midgut infection (MI) and detection of virus antigen in head

tissue revealed a disseminated infection (DI). Mosquitoes were not assayed for salivary gland infection or for actual virus transmission. Completely refractory mosquitoes had no midgut or head infection and were scored as *MIB*. Midgut Escape Barrier (*MEB*) mosquitoes had a midgut infection but no head infection. Susceptible (*SUSC)* mosquitoes had a head infection. The numbers of adults analyzed in each collection site are listed in Table 1 as are the rates of *SUSC*, *MIB*, and *MEB* mosquitoes.

DNA extraction, PCR, SSCP and sequencing

DNA was extracted from the thoraces of individual mosquitoes (Black and DuTeau, 1997) and the DNA pellet was resuspended in 500 ml TE buffer (50 mm Tris-HCl, 5 mm EDTA, pH 8.0). A 50-ml aliquot of this DNA was overlaid with sterile mineral oil and stored at 4°C for daily use in the polymerase chain reaction (PCR). The remainder was stored in plastic screw-top vials at -70°C.

Early Trypsin was amplified in two overlapping pieces using the four primers shown in bold italics in Figure 7 with an annealing temperature of 60°C and using the conditions described in (Black *et al.*, 1997). SSCP and silver staining procedures also follow (Black *et al.*, 1997). Each novel genotype detected by SSCP was sequenced and at least two sequence determinations from different mosquitoes with the same genotype were completed. Most duplicate sequences were obtained from mosquitoes from different collections. PCR products were sequenced directly. Products were purified using a QiaQuick PCR purification kit (Qiagen) with the final product suspended in 50 μl of ddH_20 and then sequenced using the Taq Fluorescent Sequencing Dye Deoxy Terminator cycle-sequencing reactions on an ABI sequencer (Davis Sequencing, Davis, CA). Sequences were obtained in both directions using the four primers (Figure 7).

Initial analyses of the 3' fragment revealed length polymorphisms associated with indels in the 13 nucleotides 5' to the start site and within the intron (Figure 7). Individual bands were picked from the SSCP gel and sequenced in 70 heterozygous mosquitoes.

Dataset construction

For both the 5' and 3' fragments, each novel SSCP genotype was assigned a numerical label. For each mosquito the initial PHENOTYPE datafile contained a 5 letter label indicating the collection site (Table 1), a two integer field containing an unique identification number for the mosquito followed by a two character field containing the DEN infection phenotype (11=*SUSC*, 10=*MEB*, 00=*MIB*). Two three-integer fields containing the SSCP genotype number for the 5' and 3' fragments followed these 3 strings.

All 5' sequences were assembled into a single GENOTYPE dataset and machine aligned using CLUSTAL W (Thompson *et al.*, 1994). The same was done for all 3' sequences. Each line of data contained a diploid, phase unknown, genotype. An 'A' indicated an A/A homozygote, 'C'=C/C, 'G'=G/G, 'T'=T/T. For heterozygotes, 'M' = A/C, 'R'=A/G, 'W'=A/T, 'S'=C/G, 'Y'=C/T, and 'K'=G/T. Gaps representing insertions or deletions (indels) were encoded as '-'=-/-, '1'= -/A, '2'= -/C, '3'= -/G, '4'=-/T. All unique substitutions were checked against the original sequence trace files and corrected where necessary. Next a sequence identity matrix was derived in BioEdit (Hall, 1999.) to identify identical sequences that had been incorrectly assigned unique numerical codes and to identify different sequences that had been incorrectly assigned the same numerical code. When this was completed all unique SSCP genotypes were assigned a new numerical code and these codes were corrected in the PHENOTYPE datafile. The

program PGMerge (Lozano-Fuentes *et al.*, 2004) read the SSCP genotype number for the 5' and 3' fragments from the PHENOTYPE datafile and replaced these with the corresponding sequences from the GENOTYPE dataset. The overlapping regions of the combined sequence dataset were then identified. Portions corresponding to the primers were deleted and the 102 nucleotide sequences that overlapped in the 5' and 3' fragments were compared. No discrepancies were detected and one section was deleted. The final PHENOTYPE-GENOTYPE file therefore contained 1,642 rows corresponding to each phenotyped mosquito and 869 columns of nucleotides.

Analysis of sequence variability

In each of the 31 collections, DnaSP (Rozas and Rozas, 1999) was used to determine the number of haplotypes, the haplotype diversity, the number of segregating sites, π (Nei, 1987), equation 10.5) and the standard deviation of π ((Nei, 1987) equation 10.7), θ per segregating site (Watterson, 1975) and F* (Fu and Li, 1993) and *D* (Tajima, 1983) tests for neutrality. These same statistics were estimated for each of the four regions and for the entire study. π was also estimated for each segregating site among all mosquitoes and plotted to display levels of variation across the entire gene.

Analysis of genotype frequencies

To summarize the relationship between observed and expected heterozygotes at each segregating site, Wright's F_{IS} (= 1 – (Observed number of heterozygotes / expected number of heterozygotes) (Wright, 1965) was estimated among all 31 collections. Weir and Cockerham's *f* (Weir and Cockerham, 1984) is an estimator of F_{IS} that is unbiased by small or unequal sample sizes. Detailed formulae appear in chapter 2.

Partitioning of the dataset

Admixtures of populations with different frequencies of alleles in the candidate gene and different values of the trait can cause the inference of false QTNs. Based on previous studies of gene flow in Mexico (Gorrochotegui-Escalante *et al.*, 2000) we know that *Ae. aegypti* in Mexico exists as four panmictic populations: Northeast, Central Atlantic, Yucatan, and Pacific. DEN susceptibility varies widely among these (Bennett *et al.*, 2002).

The following steps were taken to test the likelihood of detecting a false QTN in our analyses. We examined variation between *SUSC* + *MEB* mosquitoes and *MIB* mosquitoes and variation between *SUSC* and *MEB* mosquitoes with respect to a mitochondrial (mtDNA) marker, a 387 bp region of the NADH dehydrogenase subunit 4 (ND4) gene. In so doing we assume that 1) variation in ND4 has no influence on DEN2 midgut susceptibility and 2) variation in ND4 broadly reflects effects that act on the nuclear genome (e.g. inbreeding, migration, genetic drift). With these assumptions, if admixtures of populations cause the inference of false QTNs then we would expect our methods to identify putative QTN's in the mtDNA and we could expect false QTNs to be detected in the *Early Trypsin* gene. Primers used to amplify the ND4 gene were ND4+ (5'-GTD YAT TTA TGA TTR CCT AA-3') and ND4- (5'-CTT CGD CTT CCW ADW CGT TC-3'). PCR, SSCP and sequencing are described in (Gorrochotegui-Escalante *et al.*, 2000). PHENOTYPE-mtDNA GENOTYPE files were generated as described above for the *Early Trypsin* gene. Association mapping of mtDNA QTNs was initially done in all mosquitoes and then repeated amongst mosquitoes in each of the four regions to determine if false associations arise in genetically uniform groups.

Linkage disequilibrium analysis

We assumed that the 31 collections represent a series of subdivided populations and therefore that linkage disequilibrium could result from both random (nonsystematic) and epistatic (systematic) causes. When migration among subdivided populations is limited, novel mutations and random drift may generate different unique nucleotide combinations and create disequilibrium (Ohta and Kimura, 1969). Alternatively, systematic disequilibrium that is consistent in all subdivided populations is assumed to arise through epistatic natural selection (Lewontin and Kojima, 1960).

PGLD calculates Ohta's five D-statistics uses (Ohta, 1982a; Ohta, 1982b) for all possible pairs of segregating sites and reports whether patterns are consistent with 1) nonsystematic disequilibrium, 2) systematic disequilibrium, or 3) unequal systematic disequilibrium (Lozano-Fuentes *et al.*, 2004). These outcomes are then displayed in a half matrix in which a white square indicates nonsystematic disequilibrium, gray squares indicate unequal systematic disequilibrium and black squares indicate systematic disequilibrium. An additional test of disequilibrium was performed by regressing D'^{2}_{ST} on the physical distance among segregating sites. Under an hypothesis of disequilibrium, sites in proximity to one another should have greater disequilibrium than sites further away and there should be a negative slope. The significance of this regression analysis was evaluated (Mantel, 1967).

Association mapping based upon allele frequencies

PGTheta performed a series of statistical tests of association between specific nucleotides and *MIB*, *MEB*, or *SUSC* phenotypes -chapter 2-. For all analyses, the PHENOTYPE-GENOTYPE file was partitioned according to phenotype. The frequency

of a nucleotide i at a segregating site is p_i. For association mapping of *MIB*, p_i in *SUSC* and *MEB* mosquitoes was compared to p_i in *MIB* mosquitoes. For association mapping of *MEB*, p_i was compared between *SUSC* and *MEB* mosquitoes. For each alternative nucleotide at a segregating site, θ (Weir *et al.*, 1984) was estimated and the consistency of θ was evaluated with a permutation test (Doerge and Churchill, 1996). Detailed formulae for these procedures appear in chapter 2.

Association mapping based upon genotype frequencies

PGCon performs a series of contingency χ^2 tests of the association between specific genotypes at segregating sites and *MIB*, *MEB*, or *SUSC* phenotypes (Lozano-Fuentes *et al.*, 2004). Genotype association mapping of *MIB* compared the frequency of a genotype in *SUSC* and *MEB* mosquitoes to that in *MIB* mosquitoes. For genotype association mapping of *MEB*, genotype frequencies in *SUSC* mosquitoes were compared to those in *MEB* mosquitoes. Because of the large number of segregating sites simultaneously tested, Bonferroni adjusted significance thresholds were used. Detailed formulae for these procedures appear in chapter 2.

Power calculations

We estimated the minimal effect (m) that a nucleic acid substitution or a difference in genotype could exert on a phenotype and still be statistically significant in association mapping of *MIB* or *MEB*. This was especially important given the variable sample sizes among 1) geographic regions, 2) phenotypic groups 3) nucleotides and 4) genotypes at each segregating site. m was iteratively estimated in the program PGCon (Lozano-Fuentes *et al.*, 2004).

Results

Comparison of SSCP and sequence analyses for genotype detection

Table 3 is a comparison of presumed genotype numbers identified by SSCP and actual numbers as determined by sequencing. In the 3' fragment we sequenced PCR products from 52 mosquitoes representing 25 unique SSCP patterns. Sequencing identified 26 unique genotypes. One substitution was undetected by SSCP. This comparison was complicated in the 5' fragment by length polymorphisms associated with indels in the 13 nucleotides 5' to the start site and within the intron (Figure 7). The number of genotypes identified in deletion homozygotes by SSCP and sequencing were the same. In contrast, in insertion homozygotes SSCP identified 5 more genotypes than sequencing. In 36 indel heterozygotes, the insertion bearing allele was picked from the SSCP gel and sequenced. SSCP analyses suggested the presence of 18 genotypes, which was confirmed by sequence analysis. In an additional 34 indel heterozygotes, the insertion bearing allele was picked and sequenced. SSCP predicted 17 genotypes while sequence analysis identified only 4.

Segregating sites in the Early Trypsin gene

There were 869 sites in the aligned *Early trypsin* gene dataset (Figure 7). There were 744 sites in coding regions and the three noncoding regions contained the remaining 125 sites. There were a total of 90 segregating sites (Table 3). Coding regions contained 44 segregating sites and 46 sites were in noncoding regions (Table 4). In the noncoding region, 25 segregating sites contained gaps. In the coding region, 13 of the segregating sites encoded synonymous substitutions while the remaining 31 were nonsynonymous. Among the 3,286 alleles sampled, counting only those sites without gaps, there were 282

genotypes. Gene diversity was 0.9203 (std. dev. = 0.0038). The nucleotide diversity (π) was 0.0117 and θ/per site was 0.0097. The average number of nucleotide differences (k) was 9.841. π was also estimated for each segregating site among all mosquitoes and plotted to indicate levels of variation across the entire gene (Figure 8). The highest level of diversity occurred in the intron.

Fu and Li's F* was estimated in all 31 collections and across the gene within the 4 geographic regions (Figure 9). F* is a normalized comparison of the number of all mutations (η) to the number of singletons (η_s). This analysis assumes that F*>0 ($\eta>\eta_s$) under balancing selection, F* $\simeq$ 0 ($\eta \simeq \eta_s$) with neutral substitutions and F*< 0 ($\eta< \eta_s$) under purifying selection. A significant positive F* was detected in 16 collections and a significant negative F* occurred in 2 of the 31 collections (analyses not shown) suggesting that balancing selection may be acting on *Early trypsin* in over half of our collections. F* was strongly correlated among all four geographic regions (see Pearson's correlation coefficients in the box in Figure 9). However, analysis of F* across the gene within the 4 geographic regions indicated that significant values of F* are not equally distributed. The majority of segregating sites appear to be neutral however, in all four regions F* was large and positive in sites 100-200, the regions that contain the intron. In two of the four regions F* was large and positive in sites 325-375 suggesting some form of balancing selection in the 5' end of the active trypsin. A significant negative F* (purifying selection) was only detected in the Central Atlantic region in position 260-270.

Analysis of genotype frequencies

Weir and Cockerham's *f* was calculated at each segregating site among all 31 geographic collections (Figure 10). *f* ranges from –1 to +1, and a positive *f* indicates an

excess of homozygotes while a negative f indicates an excess of heterozygotes. Figure 10A is a frequency histogram of f among all 90 segregating sites. There is a clear excess of homozygotes at most segregating sites. However, Figure 10B indicates that this homozygote excess is not equally distributed across the gene. Of the 90 segregating sites, 49 are located in the first 176 nucleotides, which include the 5' noncoding region, the signal peptide, and the intron. In this region only segregating site 160 demonstrated a large homozygote excess. In contrast, there are many segregating sites with excess homozygotes in the remainder of the gene. Specifically notice the clustering of large f values between the first group of conserved trypsin residues and the cutting residues, the same area that has the lowest gene diversity (Figure 8). These patterns are consistent with an hypothesis of strong purifying selection acting in the functional region of *Early Trypsin* such that heterozygotes and homozygotes of the alternative nucleotide are selected against. Two sites (160 and 240) in the 5' noncoding region-5'signal peptide-intron region have large f values. As described below, these sites also exhibit systematic disequilibrium. These patterns are consistent with an hypothesis of epistatic selection acting to maintain homozygotes at these two sites. However, the nature of this epistasis is unclear; segregating site 160 is an A ←→ C transversion in the intron and site 240 is a nonsynonymous A ←→ G transition in the functional portion of *Early Trypsin* (Table 9)

Partitioning of the dataset

We tested the likelihood of detecting false MIB and MEB QTN by examining variation between *SUSC* , *MIB,* and *MEB* mosquitoes with respect to the ND4 mtDNA gene. Association mapping of MIB QTNs was initially done in all mosquitoes (Figure 11A). Remembering that a putative QTN is identified at a segregating site when the

original θ exceeded the 95% threshold, there were 4 putative QTNs identified in the mtDNA when *SUSC* + *MEB* mosquitoes and *MIB* mosquitoes were compared (Sites 83, 171, 297, 303 in Figure 11A). This suggests that similar false QTNs could arise in analyses of the *Early Trypsin* gene if association tests are performed on all mosquitoes irrespective of their geographic origin. This analysis was repeated among mosquitoes in the Northeast, Central Atlantic, Yucatan and Pacific regions. These collections were genetically homogeneous (Gorrochotegui-Escalante *et al.*, 2000) and no QTNs were detected in any of the four regions. The results for the Pacific region are shown in Figure 11B. All subsequent analyses of MIB were performed separately in the four regions.

There were no QTNs identified between all *SUSC* and *MEB* mosquitoes (Figure 11C). This analysis was repeated among mosquitoes in the four regions. No QTNs were detected in Northeast, Central Atlantic, Yucatan (results for the Yucatan region are shown in Figure 11D). However in the Pacific, there were 14 putative MEB QTNs identified (Figure 11E). These associations disappeared when Coyuca and Culiacan mosquitoes were removed from the Pacific analysis (Figure 11F). All subsequent analyses were performed separately in the four regions but with Coyuca and Culiacan collections removed.

Linkage disequilibrium analysis

Results of the analysis of Ohta's D-statistics among all pairs of segregating sites are shown in Figure 12. Systematic disequilibrium is evident in the Central Atlantic collections and occurred among alleles at most nucleotides in the 5' end of the gene and substitutions at site 176 in the 3' end of the intron, a synonymous C ←→ T transition at

segregating site 242 and a synonymous A ←→ G transition at segregating site 344 (Table 9). In contrast, very little systematic disequilibrium occurred in the Northeast, Yucatan and Pacific collections. Sites 160 in the intron and the A ←→ G nonsynonymous transitions at 240 were in systematic disequilibrium in Northeast, Central Atlantic and Pacific. Sites 160, 161 and 171 in the intron were in systematic disequilibrium in the Northeast and Central Atlantic collections. Alleles at site 242 were in systematic disequilibrium with alleles at site 344 in Yucatan and Pacific collections. The G ←→ T synonymous transition at 560 was in systematic disequilibrium with a C ←→ T nonsynonymous transition in a second codon position at position 598 in Central Atlantic and Yucatan and with a synonymous transversion in a third codon position at 737 in the Northeast and Central Atlantic collections. We have no explanation as to why any of these sites are in systematic disequilibrium.

An additional test of disequilibrium was performed by regressing D'^{2}_{ST} on the physical distance among segregating sites (Figure 13). In all four regions, the slope was negative, and some adjacent sites had large disequilibrium coefficients. This is consistent with a general pattern of disequilibrium even though the Mantel probability was 1.00. Taken together, the Ohta and regression analyses suggest that systematic disequilibrium occurs in only a few segregating sites.

Association mapping based upon allele frequencies

For association mapping of *MIB*, the PHENOTYPE-GENOTYPE file was partitioned into mosquitoes with an MI (*SUSC* + *MEB*) and compared to mosquitoes without an MI *(MIB*). For each alternative nucleotide at a segregating site, θ was estimated and the consistency of θ was evaluated following (Doerge *et al.*, 1996) (Figure

14). Five putative QTNs were identified among all mosquitoes. Three putative QTNs were identified in the Northeast, four in the Central Atlantic, none in the Yucatan and 14 in the Pacific. Three putative QTNs at sites 240, 305 and 344 were detected in both the Central Atlantic and Pacific. There is a A ←→ G nonsynonymous (Val ←→ Iso) transition at segregating site 240, a synonymous A ←→ G transition at 305, and synonymous A ←→ G transition at 344. While 240 and 344 were involved in systematic disequilibrium, they were not in disequilibrium with one another.

The relationships between alleles and midgut infection rates (MIR) at sites 240, 305, and 344 are plotted in Figure 15. A guanine at site 240 (encoding a Valine) slightly increased MIR in the Northeast, Yucatan and the Pacific but the opposite pattern was seen in the Central Atlantic. An adenine at site 305 increased MIR in the Central Atlantic and Yucatan but the opposite pattern was seen in the Northeast and Pacific collections. A guanine at site 344 increased MIR in the Northeast and Central Atlantic but the opposite pattern was seen in the Yucatan and Pacific collections. Central Atlantic. No consistent association between alleles at sites 240, 305, and 344 with MIR was evident.

A series of contingency χ^2 tests of association between alternate genotypes and the number of mosquitoes with and without an MI were performed for association mapping of MIB (Figure 16). Only in the Pacific did any tests exceed the 95% adjusted α value. Using instead the unadjusted 95% α value, seven putative QTNs were identified in the Northeast, two in the Central Atlantic, none in the Yucatan and 15 in the Pacific. Genotypes at 240 were detected in both the Central Atlantic and Pacific. Genotypes at 305, 560, and 615 were detected in both the Northeast and Pacific.

The relationships between genotypes at sites 240, 305, and 344 and MIR are plotted in Figure 17. A greater proportion of guanine homozygotes at site 240 were infected in the Northeast, and Central Atlantic but in the Yucatan and the Pacific the opposite pattern occurred. A greater proportion of adenine homozygotes at site 305 were infected in the Northeast and the Pacific, but in the Yucatan and Central Atlantic the opposite pattern was seen. A lower proportion of adenine homozygotes at site 344 were infected in the Northeast, Central Atlantic, and Yucatan, but the opposite pattern was seen in the Pacific. No consistent association between genotypes at sites 240, 305, and 344 and MIR was evident.

For association mapping of *MEB*, mosquitoes with a DI (*SUSC*) were compared to mosquitoes without a DI (*MEB*) (Figure 18). Putative QTNs were only detected in the Pacific. No analyses of consistency could be performed.

Power calculations

The minimal threshold of detection (m) varied from 0 - 4.5% in the Northeast and Central Atlantic, the regions with the smallest sample sizes (Figure 19). Sample sizes were large in both Yucatan and Pacific regions, and m never exceeded 2.5%. Thus in general our experimental design had the power to detect effects as small as 2.5% in the Yucatan and Pacific regions, and 4.5% in the Northeast and Central Atlantic. Our experimental design would not have detected allelic and genotypic substitution effects that caused a smaller percent change in phenotype.

Discussion

We have applied six criteria to test for association between segregating sites in *Early Trypsin* and DEN2 susceptibility in *Ae. aegypti*. Most of these criteria were

intended to minimize the likelihood of detecting false positive QTNs. First, we established *a priori* hypotheses regarding *Early Trypsin* as a candidate gene. Second, we used large sample sizes to maximize the power of our tests to detect QTNs. Third, we tested for independent segregation of alternate nucleotides at segregating sites and in general found very little evidence of systematic disequilibrium in the *Early Trypsin* gene. Fourth, we used mtDNA as a neutral marker to identify the potential for false QTNs arising from analysis of genetically distinct populations. Fifth, subdividing populations into panmictic units allowed independent tests of the validity and repeatability of a QTN. Sixth, we examined the pattern of association between specific alleles and genotypes with MIR when PGTheta (Figure 14) or PGCon (Figure 16) indicated a presumptive QTN.

Applying these 6 criteria, no consistent associations between segregating sites in *Early Trypsin* and susceptibility to DEN2 were found in *Ae. aegypti* in North America. However, it is important to remember that ET remains the only proteolytic enzyme found in the midgut of *Ae. aegypti* immediately following a bloodmeal and that ET is ultimately necessary to activate the transcription and translation of other secreted proteins in the midgut (Noriega *et al.*, 2001). The first observation may not be relevant because viable DEN virus remains in the gut for days after blood feeding, sufficient time for other trypsins to appear in the gut (Molina-Cruz *et al.*, 2004).

The apparent lack of association with segregating sites in the structural gene does not preclude the possibility that *cis* or *trans* acting elements influence transcription and possibly translation of ET and are therefore associated with susceptibility to DEN virus. The fact that we detected little systematic disequilibrium allows the possibility that *cis* elements in close proximity to the *Early Trypsin* gene were undetected. The possibility

that mutations in enhancers or promoters residing outside the analyzed region remains a *de facto* qualification for all QTN studies. To test for variation in transcription it would probably be faster (but no less expensive) to repeat the sampling scheme adopted in this paper and instead test mosquitoes for differences in the amounts of *Early Trypsin* mRNA in the midgut prior to bloodfeeding. To test for variation in translation, mosquito midguts could be tested serologically for differences in the amounts of ET. The results of this study do not therefore preclude transcriptional or translational differences in Early Trypsin among *MIB*, *MEB*, and *SUSC* mosquitoes. Up to 11 additional trypsins, chymotrypsins, elastases or trypsin like enzymes have been discovered in the *Ae. aegypti* genome (Fernando Noriega, personal communication). Any factors that influence ET activity also therefore affect transcription and translation of these other genes (Noriega *et al.*, 2001).

Large sample sizes maximize the power of genotype-phenotype association tests. Our power calculations (Figure 19) suggest that sensitivity was correlated with sample size and varied from 2.5-4.5%. In practice, being able to detect QTNs with small effects in less polymorphic sites is of little value because alternate alleles and genotypes are rare and thus not very predictive in natural populations (Risch, 2000).

We tested for independent segregation of alternate nucleotides at segregating sites and in general found very little evidence of systematic disequilibrium in the *Early Trypsin* gene. This suggests that $4Nc$ is sufficiently large in *Ae. aegypti* populations to disrupt disequilibrium. The fact that Ohta's analysis identified very little systematic disequilibrium (Figure 12) suggests that the general pattern of disequilibrium found in each region (Figure 13) is caused by recent mutations and founder's effects. The only

exception to this pattern involved two sites (160 and 240) in the region containing the 5' noncoding region and the signal peptide-intron region. Both sites also had large *f* values. These patterns are consistent with an hypothesis of epistatic selection acting to maintaining homozygotes at these two sites, although the nature of this epistasis is unclear.

Sampling of panmictic units was an important step taken to avoid the detection of false QTNs. Note that if we had not subdivided our analyses among geographic regions that we would have inferred the existence of 5 MIB QTNs (Figure 14) and 4 MEB QTNs (Figure 18). Subdividing populations into panmictic units simultaneously provided a fifth criterion to reduce the likelihood of false QTN detection. Independent analyses in these four geographic regions allowed independent tests of the validity and repeatability of a QTN. A valid QTN should be repeatedly detected in all regions and, as importantly, the patterns of genotype-phenotype associations should be consistent. These latter statements assume no epistasis and no QTN by environment interactions, yet there are examples of both epistasis and QTN by environment interactions (Long *et al.*, 1996; Lukens *et al.*, 1999)). The same may be true among geographic regions in *Ae. aegypti*. However, if this is true, the application of QTNs towards predicting vector competence in field populations may be strictly regional and possibly have a seasonal or temporal component.

IV Aedes aegypti Vector Competence and Gene Flow in the state of Veracruz, Mexico

Introduction

Aedes aegypti is the main vector of the four serotypes of the Dengue virus. There are 50 to 100 million cases with thousands of deaths cases every (Gubler, 2002). Most of the dengue cases are asymptomatic or have mild symptoms. During epidemics severe infections with profuse bleeding and shock that can be observed, 20% of these cases can have a fatal ending (WHO, 2002).

Ae. aegypti populations exhibit a large amount of genetic variability in vector competence for flaviviruses (Aitken *et al.,* 1977; Tabachnick *et al.,* 1985; Rosen *et al.* 1985), including dengue 2 viruses (Gubler *et al.,* 1979; Rosen *et al.,* 1985; Tardieux *et al.,* 1990; Vazeille-Falcoz *et al.,* 1999). Vector competence for flaviviruses in *Ae. aegypti* is thought to be controlled by at least two mechanisms, the midgut infection barrier (MIB) and the midgut escape barrier (MEB) (Miller and Mitchell, 1991; Bosio *et al.,* 1998).

Black *et al.* (2002) suggested that the variation in dengue infection rates among natural populations of *Ae. aegypti* may be due to segregation of alleles at each of nine previously detected QTL. Additionally, the alleles of the nine QTL should vary independently in frequency among populations. This idea would predict that varying proportions of incompetent and competent mosquitoes would be found in *Ae. aegypti* populations, in addition, environmental factor would also contribute to the variation in susceptibility. Bennett and collaborators (2002) assessed the validity of this model by estimating the rates of MIB and MEB in mosquito collections throughout Mexico. They discovered as much variation in vector competence for dengue virus serotype 2 (DEN-2) as was described earlier in surveys of vector competence for Yellow Fever (YF) virus (Tabachnick *et al.*, 1985). The MIB and MEB rate were found to be weakly correlated, supporting the prediction that independent sets of genes control MIB and MEB.

Gorrochotegui-Escalante *et al.* (2000, 2001) detected a barrier to gene flow along the Gulf of Mexico between northern collections from Houston, Texas to Tuxpan, Veracruz and southern collections from Minatitlan to Cancun. In an attempt to locate more precisely the geographic barrier to gene flow between Tuxpan and Minatitlan, we concentrated collections in the state of Veracruz.

Veracruz is a state that plays an important role in dengue epidemics in Mexico. The state has cases throughout the year (DGE, 2001-2004). The state climate is a factor that could explain the number of dengue cases. Close to 90% of the state is warm (24 - 26°C annual average), humid (annual rainfall from 600 to 4500 mm) and most of the rains occur in summer. Veracruz, located along the Golf Mexico (22°28', 17°09' N; 93°36', 98°39' W), is divided into two (north and south gulf costal plains) by the

Neovolcanic axis (Figure 20). This axis is formed by volcanoes located in an almost east-west line over the 19th parallel.

Materials and methods

Mosquito collection. The locations of *Ae. aegypti* collections are listed in Table 4, as are the number of adults analyzed in each collection for vector competence and population structure. The geographic locations of all sampling sites appear in Figure 20. Mosquito were collected as larvae, hatched and reared to adults in the laboratory. Adults were individually examined to confirm that they were *Ae. aegypti*. The 2003 collection was processed for analysis of vector competence and mtDNA markers. The 2004 collection was used only for analysis of mtDNA markers.

All experiments were performed on mosquitoes that had spent ≤ 2 generations in the laboratory to avoid effects of colonization and inbreeding (Lorenz *et al.,* 1984). Laboratory colonies of *Aedes aegypti aegypti* D2S3 (Bennett *et al.*, 2003) served as highly susceptible controls.

Vector competence

Virus: The virus used was dengue 2 JAM1409 which was isolated in 1983 in Jamaica and belongs to toptotype Jamaica and subtype IIIb (Deubel *et al.,* 1986). To prepare stock virus, dengue 2 JAM1409 was amplified in C6/36 cells. A confluent T150 flask of C6/36 cells was infected at an MOI of 1.5. Virus was incubated for 14 days at 28 °C in L15 medium supplemented with 2% heat inactivated FBS, penicillin (1%), streptomycin (1%) and L-glutamine (1%). Virus was harvested and placed in aliquots of 0.5 ml, which were stored at -70C. For the *per os* challenges, 0.5ml aliquots were used to infect confluent T75 flasks of C6/36 cells at an MOI of 1.5. Virus was incubated for 14

days at 28 °C in L15 medium as above, and the medium was changed on day 7. Virus and cells were harvested by scraping with a cell scraper and collecting medium, virus and cells in a 50ml conical centrifuge tube. The virus suspension was mixed 1:1 with defibrinated sheep blood and placed in membrane feeders covered with sterile hog gut membranes (Higgs and Beaty, 1996). Pre- and post-blood meal virus titers were determined by inoculating serial 10-fold dilutions of the respective meal into C6/36 cells in a 96 well plate. Cells were assayed by immunofluorescence and $TCID_{50}$ titers were calculated using the Karber formula (Schoepp *et al.* 1990). Undiluted virus titers ranged from 7.5 to 8.5 logs/ml in most oral challenges.

***Per os* Challenges**: 30-40 mosquitoes/1pint carton were sucrose starved and water deprived 24 hours prior to blood feeding. Blood meals were maintained at a constant 37 °C. Mosquitoes were allowed 45 min to 1 h to feed. Fully engorged mosquitoes were selected and held for 14 days at a constant temperature of 27 °C and 80% relative humidity in an insectary with a 12-hour photoperiod. Heads and abdomen were assayed by immunofluorescense to determine DI and MI respectively.

Immunofluorescence: 14 days post oral challenge, mosquito heads were severed, squashed onto acid washed slides, acetone fixed, and assayed for dengue virus by indirect immunoflourescence assay (IFA) to determine disseminated infection rates. Slides were incubated for 40 minutes at 37 °C with a mouse derived primary monoclonal antibody directed against a Flavivirus E gene epitope (Gould *et al.* 1985) diluted 1/200 in PBS. Slides were then washed 2 times in PBS. A mixture of biotinylated sheep anti mouse antibody (Amersham) and Evans blue (1%) as a counterstain was made at a dilution of 1/200 in PBS. Slides were incubated for 40 minutes at 37 C followed by 2

washes in PBS. Slides were then incubated 10 minutes with streptavidin fluorescein (Amersham) diluted 1/200 in PBS to detect the primary-secondary antibody complexes. Slides were washed twice in PBS and once in water. Glycerin containing Diazobicyclo(2,2,2)Octane (DABCO Sigma D-2522) was applied, and a coverslip placed on each slide. Heads were examined at 200x using an Olympus BH2-RFCA microscope. Thoraces and abdomen were frozen at -70C for DNA extractions or IFA, if necessary. Abdomens of mosquitoes lacking DI were subsequently assayed by immunofluorescence to determine midgut infection rates (MI); thoraces were kept for DNA extractions.

Population structure

DNA was obtained from whole or parts of individual mosquitoes previously stored at -70°C. Primers used to amplify the Nicotinamide Adenine Dinucleotide Dehydrogenase subunit 4 mitochondrial gene (*ND4*) and all the polymerase chain reaction and electrophoresis conditions are as reported earlier by Gorrochotegui-Escalante *et al.* (2000). Haplotypes were assigned numbers according to Gorrochotegui-Escalante *et al.* (2000, 2001). The *ND4* polymerase chain reaction products from 20 individuals representing each of the 9 haplotypes were sequenced at least once along both strands on an ABI sequencer (Davis Sequencing, Davis, California). These were compared to the sequences reported previously by Gorrochotegui-Escalante *et al.* (2001) and assigned the same numeric labels of Gorrochotegui-Escalante *et al.* (2001). No new haplotypes were discovered. Phylogenetic relationships among haplotypes were described by Gorrochotegui-Escalante *et al.* (2000).

Statistical analysis of mitochondrial haplotype frequencies. Variation in haplotype frequencies within and among collection sites and regions was examined

analysis of molecular variance (AMOVA) (Excoffier *et al.*, 1992). Arlequin 2.000 estimated pairwise Slatkin's "linearized F_{ST}" [= $F_{ST}/(1 - F_{ST})$] (Slatkin, 1993) among populations and computed the significance of the variance components associated with each level of genetic structure by a nonparametric permutation test with 100,000 pseudoreplicates (Excoffier *et al.*, 1992). For each collection, the nucleotide sequence and the frequency of each haplotype were also entered into DnaSP (Rozas and Rozas, 1999). The number of polymorphic sites was estimated as well the average number of nucleotide differences (k) among individuals in a collection (Tajima 1983). $F_{ST}/(1-F_{ST})$ were used to construct a dendrogram among all collections by means of unweighted pair-group method with arithmetic averaging analysis (Sokal and Sneath, PHA 1963) in the NEIGHBOR procedure in PHYLIP3.5C (Felsenstein, 1993).

Geographic distances were obtained by Geographic Information Systems (Environmental Systems Research Institute, Redlands, CA). Linearized F_{ST} values were regressed on the pairwise geographic distances among populations to determine if geographic distance among populations serves as a barrier to gene flow (Slatkin 1993). The Mantel test was performed by PGMan (Lozano-Fuentes *et al.*, 2004). The reciprocal of the slope estimated by this regression provides an estimate of the average effective population size (Rousset, 1997).

Results

Vector competence

The disseminated infection rates (DIR) among mosquitoes from the 10 Veracruz collections are shown in Table 5. To determine anatomic barriers that condition transmission potential, mosquitoes from the 10 experimental *Ae. aegypti* collections and

the D2S3 control were phenotyped for the presence of MIB and MEB. The DIR for D2S3 was 89% (n = 60). Figure 21 indicates that the lowest MIB rate occurs to the north of the Neovolcanic axis associated with the Poza Rica collection, while the highest MIB rate occurs to the south of the axis associated primarily with the Alvarado collection. There is a gradual decrease in MIB (green bar graph) from Panuco south to Poza Rica and then a gradual increase in MIB to Alvarado that is maintained in all collection to the south of Alvarado. The MIB rates are significantly different between collections to the north and south of the Neovolcanic axis (t, $p < 0.0001$).

Figure 21 shows that MEB slowly increased from north to south. In Panuco, Tantoyuca and Poza Rica MEB range between 2-6%. MEB jumped to 24% in Martinez de la Torre to the north of the Neovolcanic axis and was 27% to the south of the Neovolcanic axis. The MEB rate ranged from 6-28% in the remaining collections south of the Neovolcanic axis. MEB was found significantly different between mosquitoes collected north and south of the neovolcanic axis (t, $p = 0.0011$). The MIB and MEB rates were not correlated in the state ($R^2 = 0.0097$, $p > .5$).

Gene Flow

The ND4 gene was amplified and surveyed for variation by SSCP among 654 mosquitoes among 19 collections (Cosoleacaque was not collected in 2004). Nine different *ND4* haplotypes were detected with SSCP. The *ND4* gene was sequenced in 20 individual mosquitoes. Sequences of mosquitoes with identical SSCP patterns were identical within each haplotype, and SSCP patterns differed among individuals with even small sequence differences. This supports previous observations (Gorrochotegui-Escalante *et al.* 2000, 2001) that SSCP is specific and reproducible in identifying single

nucleotide substitutions in the *ND4* gene in *Ae. aegypti.* All the sequenced haplotypes were compared to previously reported sequences (GenBank accession numbers AF334841–AF334865), and no new sequences were detected.

A cluster analysis of pairwise linearized F_{ST} values among all 19 collections collections and including the 17 Gulf Coast collections of (Gorrochotegui-Escalante *et al.,* 2001) is shown in Figure 22. All collections north of the Neovolcanic axis, with the exception of Nuevo Laredo, fall within a single cluster. The majority of south collections fall within a separate cluster, with the exceptions of Minatitlan 1996 and Moloacan.

Gorrochotegui-Escalante (2002) reported similar results for Moloacan, Minatitlan 1996 and Nuevo Laredo. They are distantly related to one another and to the other collections. Moloacan and Minatitlan are in close geographic proximity to one another but are also in proximity to Acayucan, Coatzacoalcos, Cosoleacaque and Moloacan. The extreme genetic isolation of Nuevo Laredo was reported for both RAPD and mtDNA markers in Gorrochotegui-Escalante (2000).

Figure 23 shows the cluster analysis of pairwise linearized F_{ST} values among the all Veracruz collections in the present study. There are again two distinct clusters, one containing primarily collections north of the Neovolcanic axis and a cluster containing only collections south of the Neovolcanic axis. The 2003 Acayucan and Zempola collections fell in the Northern cluster but were in the southern cluster in 2004.

Nested analysis of haplotype frequencies. Haplotype frequencies were compared among collections within cities and among cities by AMOVA (Excoffier *et al.,* 1992). This partitioning was done among 1) all Gulf Coast Collections north vs. south of

the Neovolcanic axis 2) 2003 and 2004 Vera Cruz collections north vs. south of the Neovolcanic axis, 3) Collection done in 2003 vs. 2004.

A total of 16% of the variation in haplotype frequencies arose between Gulf Coast collections north and south of the Neovolcanic axis (Table 6). These observations agree with the clustering patterns shown in Figure 22. In the second AMOVA, 24% of the variation was accounted by differences between collections from north and south of the neovolcanic axis within the state of Veracruz. Ten percent of the variation arose from within the groups and 66% of the variation came from the collection sites. These observations agree with the clustering patterns shown in Figure 23. Between years (AMOVA 3), the variation only accounted for -2.7% of the variation, 26% of the variation arose from within the years and 77% of the variation arose within collections. The pattern of variation found among north and south collections somewhat agrees with the diversity analysis (Figure 25).

Genetic isolation by distance. Pairwise linearized F_{ST} values were regressed against the pairwise geographic distances among collections (Figure 24, Table 7) to determine if gene flow among collections was correlated with geographic distance (*i.e.*, to test for genetic isolation by distance). This analysis indicated a significant correlation between genetic and geographic distances among all collections, within collection made south of the neovolcanic axis and within the state of Veracruz. However, no correlation was detected among collections made north of the Neovolcanic axis or collections made within the state of Veracruz.

Haplotype diversity. Collections from the North consistently had a lower average number of haplotype differences among individuals (k) and fewer polymorphic

sites than collections from the South, (Figure 25). Thus, mosquitoes in the South collections not only had the greatest number of mtDNA haplotypes, but the sequences within these haplotypes were more diverse than in the north collections.

Discussion

We have discovered an abrupt barrier to gene flow in *Ae. aegypti* associated with the Neovolcanic axis along the Gulf Coast of Mexico. For various reasons, *Ae. aegypti* is limited in its distribution in Mexico to altitudes below 2000 feet. The high mountains and uplifts associated with the Neovolcanic axis therefore limit the distribution of *Ae. aegypti* to ~15 km between the mountains and the Atlantic Ocean. In itself this narrow corridor doesn't serve as a barrier to natural aerial dispersal of *Ae. aegypti.* However, it is well known that *Ae. aegypti* and other container breeding mosquitoes can be transported over long distances through human commerce. Eggs are laid in protective containers such as tires which can then be transported over long distances. There are no large cities (> 10,000 inhabitants) close to the Neovolcanic axis below 3000 feet in our study area (Figure 26). A single, narrow paved road (interstate 180) crosses this ~15km corridor from North to South (Figure 26). With regards to trade, most of the commerce in Veracruz occurs between Veracruz City and Mexico City, no significant commerce occurs between northern and southern parts of our study area (North American Transportation in Figures, 2000)

The coast in our study area is lined with Alvarado Mangrove swamps. Furthermore, there is an abrupt change in vegetative coverage at the Neovolcanic axis. Veracruz moist forests occur to the north of the axis (1500 - 2000 mm rain/yr) but extensively. Veracruz dry forest occurs to the south (1000-1200 mm rain/yr). Human

modifications are also considerable north and south of the Neovolcanic axis (figure 42). Extensive areas of grassland are found north and south of the neovolcanic axis (*Paspalum vaginatum* (turfgrass), *Cynodon plectostachyum* (star grass), *Digitaria decumbens* (*Pangola grass*), *Pennisetum clandestinum* (kikuyu grass), *Panicum maximum* (Guinea grass) used in cattle feeding have disrupted the local vegetation. The grasslands are intermixed with agriculture lands (*Zea mays* (corn), *Saccharum officinarum* (sugar cane), *Carica papaya* (papaya), *Mangifera indica* (mango), and *Citrus sinensis* (Orange). However, south of the neovolcanic axis areas of rainforest can be found intermixed with agriculture lands and commercial grasslands. In contrast, immediately north of the neovolcanix axis grasslands are observed. The grasslands are intermixed with agriculture lands and rainforest are almost scarce.

To the north of the axis isolation by distance was not detected (Gorrochotegui-Escalante 2000). This along with the lower variability in ND4 gene from mosquitoes collected in this area (Figure 25) suggests that variation among local populations may arise through founder's effect and genetic drift. In contrast, isolation by distance was detected south of the axis (table 7) and these populations generally had higher variability. In this area, gene flow may occur more through aerial dispersal.

Regardless of the cause, the abrupt change in gene flow between *Ae. aegypti* populations north and south of the Neovolcanic axis appears to have be also associated with an abrupt increase in the frequency of mosquitoes with an MIB and a gradual increase in mosquitoes with an MEB south of the axis. Basically this has caused *Ae. aegypti* to the north of the axis to be susceptible to DEN transmission while those to the south are refractory.

A quantitative genetic analysis of DEN2 MIB and MEB was made in a population of *Ae. a. aegypti* from San Juan, Puerto Rico collected in 1995 and of an *Ae. a. formosus* population collected from Ibo village, Nigeria in 1995 (Bosio *et al.*, 1998). A standard half-sib breeding design was used to estimate the heritability for MI titers. In both populations, the heritability for MIB was 0.41 and was 0.39 for DI titers in the *Ae. a. formosus* population. In the Puerto Rican *Ae. a. aegypti*, DI titers appeared to be controlled by dominant alleles (Bosio *et al.*, 1998). These observations indicate that DEN susceptibility is a quantitive trait and predict that the frequencies of susceptibility alleles should vary continuously among populations.

A great deal of information on the locations of QTL that control MIB and MEB in *Ae. aegypti* has been accumulated. To date we have identified 9 QTL that appear to condition transmission of DEN2 in *Ae. aegypti* (Bosio *et al.*, 2000; Gomez-Machorro *et al.*, 2003 and Bennet *et al.* in review). Six of these (1 on chromosome I, 3 on II, and 2 on III) condition an MIB and three (one on each chromosome) condition a MEB.

Taken together, the results to date suggest that variation in dengue infection rates among natural populations of *Ae. aegypti* may be due to the segregation of alleles at least 9 QTL. Differences in dengue susceptibility among populations reflect differences in the frequency of alleles at the MIB and MEB loci. Our mapping results suggest that alleles at the 9 MIB or MEB loci should vary independently in frequency among populations. Independent segregation along with a myriad of environmental factors suggest that an *Ae. aegypti* population will probably be composed of varying proportions of incompetent and competent mosquitoes. Mosquitoes will range from being completely refractory to oral infection, to being susceptible to midgut infection but unable to transmit virus, to being

fully competent to acquire and transmit DEN. Bennett *et al.* (2002) tested this model by estimating MIB and MEB rates in *Ae. aegypti* populations from throughout Mexico and showed a great deal of variation in the frequency of alleles that control DEN-2 vector competence. Bennett *et al.* (2002) showed that alleles do not appear to operate independently as predicted. There was a weak but significant correlation between MIB and MEB rates. This suggests that MIB and MEB may be controlled in part by similar genes. In other words, a gene that confers a MIB may also confer a MEB. However, the correlation was not strong, supporting the prediction of independent sets of genes controlling MIB and MEB.

A pattern that Bennett *et al.* (2002) did not detect was an abrupt shift in vector competence among proximal populations. There are large areas of Mexico (e.g. Quintana Roo) that are highly susceptible to infection. Similarly, there were large areas (e.g. Pacific coast) that were moderately susceptible. The results of this study demonstrate how a barrier in gene flow over a short geographic distances can also cause an abrupt shift in the frequencies of alleles at MIB and MEB QTL and thereby cause abrupt changes in the susceptibility of populations.

V Explicit spatial analysis of Aedes aegypti vector competence, ND4 gene variation and dengue cases

Introduction

Population genetic theory predicts that populations will exhibit internal spatial autocorrelation if they have restricted gene flow (Wright 1943, 1978). Computer simulated studies have confirmed that a positive spatial autocorrelation that declines with distance will usually develop under restricted gene flow and without genetic drift (Sokal and Wartenberg, 1983; Sokal *et al.*, 1989; Epperson, 1990, 1995a, b; Sokal and Jacquez, 1991)

There have been several attempts to analyze the spatial autocorrelation implicit in population genetic structure, by incorporating geographic dimensions. Slatkin (1989, 1991) and Slatkin and Maddison (1989, 1990) proposed to estimate the minimum numbers of migration event from gene genealogies, looked at relationships linking such parameters with gene flow rates (m), inbreeding coefficients and effective population sizes (N_e). Excoffier *et al.* (1992) developed a method for the

analysis of molecular variance, AMOVA. Excoffier *et al.* (1992) used Wright's (1965) *F* statistics, with values depending on the correlation between haplotypes at various nested geographic levels. In this chapter, Moran's *I* and Geary's *c* spatial autocorrelation indices are presented as another alternative.

The calculation of correlation coefficients is a common practice. However, the calculation of correlation coefficients for variables observed at a variety of different spatial locations can cause false inferences. This is because Pearson's product moment correlation coefficient, can be influenced by spatial and temporal autocorrelation (Student, 1914; Richardson and Hémon, 1981; Clifford *et al.*, 1989). For this reason, there is need to assess spatial autocorrelation in data obtained from collection sites arranged in a spatial pattern. The late development of tools for assessing spatial autocorrelation has left this branch of statistics lagging behind. As a result, there are only a few developed tools and their usage is not common among researchers.

There are numerous articles that report significant correlation between environmental factors and vector competence, not to mention between disease cases and environmental factors. Sokal (2004) presents an excellent review of the effect spatial autocorrelation has on correlations and other quantitative statistics.

In this chapter I apply Spatial statistics to the data presented in chapter 4. The data obtain from the state of Veracruz were added to previous data obtain by Gorrochothegui-Escalante *et al.* (2000 and 2002) and Bennett *et al.* (2002). Thus, a country wide sample area will be analyzed for vector competence and ND4 haplotype frequencies.

The following spatial explicit hypotheses will be addressed in this chapter:

- *Aedes aegypti* ND4 gene frequencies are independent of geographic distance.
- *Aedes aegypti* vector competence to dengue virus is independent of the distance.
- Dengue cases, classical and hemorrhagic, are randomly distributed in the state of Veracruz.
- Dengue cases are independent of: altitude, inhabitants density, precipitation, solar radiation, mosquito susceptibility to dengue 2 and temperature.

Results

Spatial analysis of the ND4 gene in Mexico

In the 55 collections, 26 unique ND4 haplotypes were identified. The most common haplotypes were 3, 2, 14, 1 and 19. These respectively occurred in 79%, 73%, 38%, 36% and 34% of the 55 collections. Haplotypes 7 (Nuevo Laredo), 20 (Tucson) and 26 (Oaxaca) appeared in only single collections (Table 5). The minimum and maximum distance found between collection sites were, respectively, 2.27 and 2727 km.

Data were divided into 15 distance classes with ~equal number of points (~99) in each class. Figure 27 is the distribution of the pairwise comparisons in the 15 classes. The last distance class (1533-2728 km) lacks points at all angles showing most of the collections are oriented West and North of one another. Most collections were located along a North-South orientation along coasts, while distant collections tended to be on different coasts.

I and *c* correlograms

When performing correlograms for all data, 18 of the 26 possible *I*-correlograms and 12 of the 26 *c*-correlograms were significant (Bonferroni $p < 0.05$). Some haplotypes were removed because of low occurrence (found in < 3 collection sites). Removing rare haplotypes resulted in completely removing some collection sites Houston (n=0), Miguel Aleman (n=0), Houston (n =1) and Tucson (n = 2). The removal of these collection sites modified somewhat the distances and the distribution of the points (figure 28) and new distance classes were created.

The remaining 10 haplotypes were 1, 2, 5, 10, 11, 13, 14, 16, 19 and 24 and they showed significant correlograms. Table 10 shows that 41 of all 130 *I* values were statistically significant. Two of the 10 variables were significantly autocorrelated in distance class 0-194 km. In classes 194-290 km and 1410-2139 km classes, 5 variables were significantly autocorrelated. Significant autocorrelation coefficients in the second distance class range from 0.26 to 0.77 for *I* and 0.17 to 0.28 for *c*. Thus, there is substantial positive autocorrelation between collection sites up to 98 km apart.

Figure 29 indicates a classical spatial autocorrelation with proximate collections being positively autocorrelated but beyond 607 km for *I* and 418 km for *c*, the autocorrelation becomes largely negative. Average correlograms may better indicate the decline of genetic similarity with distance than single haplotypes frequencies, because random variation in single haplotypes will tend to cancel one another out (Sokal *et al.*, 1986).

Bearing Correlograms were performed to assess the influence of direction over the spatial autocorrelation of haplotypes. A bearing correlogram is calculated in the same way as a normal correlogram, except that the weights are scaled to indicate direction and

distance. In the first and second distance classes, there is positive autocorrelation up to 290 km in all directions (figure 30). Positive autocorrelations become focused in north-south direction between 607-800 km. This is again because most proximate collections were located along a North-South orientation along coasts.

The spatial analysis discussed so far assessed individual and average haplotype frequencies. Sokal (1983) has termed this a "multivariable" study as distinguished from a "multivariate" approach, in which the spatially distributed variables are considered jointly. Several multivariate approaches can be employed for spatial analysis. I will present two such approaches in this dissertation, a Mantel correlogram analysis and a Principal Components analysis.

Mantel Correlograms

The mantel correlogram analysis was performed comparing genetic distances between collection sites against binary distance matrices representing the geographic connections between collection sites at 13 distance classes. Slatkin's linearized F_{ST} (Slatkin, 1995) was used as metric in the genetic distance matrix based on all 26 haplotype frequencies. Two mantel correlograms were performed with different geographic distance matrices, one of which was transformed by applying natural logarithm to the geographic distances.

The strength of association between both matrices, genetic and geographic, was tested using a Mantel test. The association was greater in the natural logarithm transformed geographic distance ($r = 0.7804$, $t = 17.8315$, $p = 0.000$; simulation (10,000 iterations), $p = 0.00010$). Linear geographic distances showed a weaker non-significant association ($r = 0.113$, $t = 1.602$, $p = 0.054$; simulation (10,000 iterations), $p = 0.077$).

Both Mantel correlograms for the linearized F_{ST} are presented in table 11. Only 4 of the 13 distance classes were statistically significant. Each distance class autocorrelation was tested independently by a Mantel test. Both correlograms present similar results. Both show significant positive spatial autocorrelation at short distances and significant negative spatial autocorrelation at large distances. Thus it can be said that, when assessing the combined effects of all haplotype frequencies, there is still positive spatial autocorrelation (Mantel $r = 0.067$) up to 607 kilometers. However there is larger positive autocorrelation (Mantel $r = 0.104$) at distances up to 194 km. At larger distances the autocorrelation is negative. Significant negative autocorrelation can be detected starting at 1045 km. Nevertheless, the change from positive to negative autocorrelation is around 709 and 884 km (Figure 31). With the exception of distance classes 2, 3, 4, 6 and 13 the correlation decrease monotonically. The results suggest that as geographic distance increases, "overall" genetic dissimilarity increases and that the rate of decrease can be quantitatively calculated.

To determine the direction of greatest correlation between data distance and geographic distance a bearing analysis was performed on the F_{ST} genetic distance matrix. The bearing analysis (figure 32) gave similar result to the bearing correlogram (figure 30) showing a maximum correlation (15.8%) between 75-80 degrees. The test shows the angle of maximum correlation.

Principal component analysis

All haplotypes frequencies were employed as independent variables in a principal component analysis. To assess the contribution of each haplotype a biplot analysis was performed. The principal component analysis shows that principal component 1 (PC1)

and principal component 2 (PC2) account for < 50% of the overall variation (table 12). This could arise because of the large number of independent variables (haplotypes) with small occurrence. Figure 33 shows that haplotypes 3, 8, 9, 21 and 23 are separated from the main group (group 1). Collection sites Houston and Miguel Aleman were separated by these haplotypes. Haplotypes 10, 11 24 (group 2) are distributed in the second quadrant. Haplotypes 1, 2, 5 and 6 (group 3) are found at ~270° between the third and fourth quadrant. A small group of haplotypes (group 4) can be observed in the third quadrant formede by haplotypes 4, 7, 16, 17, 18 and 20.

The angle between groups 1, 2 and 3 is ~90°. The angle between groups 2 and 3 is ~180°. Thus, group 1 varies independently of group 2 and 3. The angle between group 2 an 3 is ~180° indicating that haplotypes variation occur in exactly opposite frequencies. Group 4 accounts for a small amount of the overall variation but can be separated from the other three groups. Group 4 is in between groups 2 and 3 and opposite to group 1. Only twelve pairs of haplotypes had a correlation greater than 0.50 (3-8, 4-7, 5-6, 8-9, 9-21, 9-23, 13-19, 15-24, 16-18, 17-20 and 21-23). Haplotype pairs 4-7, 5-6, 8-9 and 17-20 had a correlation > 0.85. Most of the Pacific and south Atlantic collectionswere associated with the second group. All the North Atlantic and most of the Oaxaca collections are associated with the third group. Group 4 is associated with Acayucan, Alvarado, and Minatitlan.

A second principal component analysis was performed after removing uncommon haplotypes. PC1 and PC2 accounted for 41% of the total variation. In the resulting biplot (figure 34) the North Atlantic, South Atlantic and the Pacific collections fall into

different groups. The Oaxaca and the North Atlantic collections are very close to one another.

Given the small amount of variation accounted for by the principal components, a spatial analysis using PCA results, is not recommended (Sokal *et al.* 1986) and thus no inference of the spatial genetic variation using principal components could be made.

Spatial analysis of the vector competence of *Aedes aegypti* to dengue-2 in Mexico.

The possibility of spatial autocorrelation in midgut infection barrier (MIB), midgut escape barrier (MEB) and susceptibility (SUC) rates (table 15) was assessed by correlograms. Data were linearized using $arcsin(x)^{1/2}$ transformation before performing the correlograms. The *I* correlograms shows that MIB and SUC are autocorrelated thus invalidating previous analysis of correlation between these three variables (MIB, MEB and SUC).

The *I*-correlograms show that MIB is positively autocorrelated at distances up to 279 km ($p < 0.001$) and negatively autocorrelated at a distance of 598 km ($p < 0.01$). However, there are two directional autocorrelation changes; the first change is from negative to positive at a distance of 728 km. The second change, from positive to negative, is at distance class 1471km. The SUC *I*-correlogram showed that susceptibility is highly positively autocorrelated up to a distance of 445 km ($p < 0.001$). At larger distances, starting at distance class 728 km, the autocorrelation is negative. The *c*-correlograms only showed SUC to be negatively autocorrelated at large distances starting at distance class 728km ($p < 0.001$). However, the overall correlograms is significant and it shows that there is high positive autocorrelation at small distances.

A second correlation analysis was performed using CRH correction, for data with spatial autocorrelation. This showed that the previous significant association, previously reported between MIB and MEB arose from spatial autocorrelation (table 16).

The spatially autocorrelated variables, MIB (figure 35) and SUC (figure 36), were interpolated based in neighboring points and a search sphere of 600km. The resulting models are not especially good (EMRS = 9.9 for MIB and (EMRS = 14.4 for SUC). Nevertheless, a spatial signature can be detected. The distribution of SUC is relatively patchy. This agrees with our hypothesis that independently varying proportions of QTLs could control *Ae. aegypti* susceptibility to dengue virus.

Dengue cases in the state of Veracruz Mexico during 2004 and their association to altitude, inhabitant density, precipitation, solar radiation, mosquito susceptibility to dengue 2 and temperature.

Dengue cases were obtained from the State of Veracruz Health Services Department data repository (http://www.ssaver.gob.mx/). The cases were organized by municipalities (figure 37). Generally the city with the county seat is the largest city in the municipality, thus instead of using the municipality polygon centroid, the county seat geographic position was used to summarize 2004 dengue cases in the state of Veracruz. Figure 37 shows that dengue cases appear to have a contagious distribution.

Average accumulated rainfall, temperature (average, maximum and minimum), vegetation types and solar radiation (average, maximum and minimum) were obtained from the Consejo Nacional para el conocimiento y uso de la Biodiversidad data repository website (CONABIO, http://www.conabio.gob.mx/). Altitudes and total number of inhabitants in each municipality were obtained from the Instituto Nacional de Estadistica, Geografia e Informatica data repository (INEGI, http://www.inegi.gob.mx/).

Mosquito susceptibility for each municipality was obtained from an Inversed Distance Weighting (IDW) susceptibility prediction map made from mosquito collections throughout Mexico (0.599 x SUC + 13.064, Error Root Mean Square= 14.37). For ranged data, the class midpoint was used for each county seat.

Altitude, inhabitant density and mosquito susceptibility are used for the remainder of analyzes. The other response variables (Average accumulated rainfall, temperature (average, maximum and minimum), and solar radiation (average, maximum and minimum) were not correlated with total dengue cases ($p > 0.25$). Figures 38-42 show altitudes, isotherms, isohyets, annual solar radiation and vegetation for the state of Veracruz region. Dengue cases were normalized by total number of inhabitants in each municipality. Variables were transformed to approach normality. Inhabitant density and altitude were lognormal ($p < 0.05$). Mosquito SUC followed a normal distribution ($p < 0.05$) and was not further transformed. Only municipalities with dengue fever (DF) and dengue hemorrhagic fever (DHF) cases were considered. Only 117 of 210 of the municipalities presented at least one DF case. Only 43 municipalities presented at least one DHF case. After removing municipalities without dengue cases, DF and DHF were lognormal ($p < 0.05$).

All variables show significant *I*-correlograms (450 pairs/class in 15 distance classes for DF and 100 pairs/class in 9 classes for DHF). Interestingly, DF and DHF presented a high positive significant autocorrelation at small distances up to 36 km. DF and DHF remain negatively autocorrelated up to distances of 189 and 219 km. Significant positive autocorrelation was recovered for DF and DHF at 278 and 358 km respectively (figure 43). Given the fact that all variables, explanatory and response

variables, showed some degree of spatial autocorrelation the usage of conventional statistics is not valid. Each variable correlation was measured and tested using CRH corrected correlations (table 17). The DF vs. DHF correlation is not shown because they are mutually exclusive by definition. As explained previously, significant correlations from conventional testing can become non-significant once spatial autocorrelation is accounted for. The degrees of freedom are reduced because some observations are not independent.

If only conventional statistic were used to analyze the information, the results would point to significant negative correlation of dengue fever cases and altitude. Additionally we would say that mosquito susceptibility increases with altitude. Once we account for spatial autocorrelation, none of the tested variables correlated with the DF cases. The complementary table, containing the correlation results, of DHF vs. altitude, inhabitant density and mosquito susceptibility are not shown because there were no significant results even with conventional statistics. The lack of significant results, when analyzing DHF, can be due the smaller number of observations and thus a reduction in the analytical power.

Both response variables (DF and DHF) show similar trends with respect to inhabitant density, altitude and susceptibility (figure 44). The regressions are also not valid due to spatial autocorrelation in the residuals, and consequently lack independence (Durbin-Watson < 1.0). However, two discernable trends can be observed from figure 44, the number of dengue cases increases as the density of inhabitants increases and decreases as the altitude decreases.

Discussion

It is clear from the spatial analysis of the ND4 haplotypes that genomes are spatially autocorrelated. There are many examples in plants in which some allozyme alleles were spatially autocorrelated (Schnabel *et al.*, 1991; Perry and Knowles, 1991; Loiselle *et al.*, 1995). Similar patterns have been found in animal mitochondrial DNA haplotypes (Nevo *et al.*, 1993; Zink, 1994). The two approaches, multivariate and multivariable revealed, significant spatial structure to the ND4 in *Ae. aegypti* in Mexico. Both approaches present a clear cutoff at which geographic and genetic distances became independent. There is detectable migration up to distances of 194 km when the totality of Mexico is taken into account.

In figures 33 and 34 the Oaxaca city collections and the North Atlantic collections are found to be very similar to one another, suggesting similar underlying events generated the ND4 variation among Oaxaca and North Atlantic *Ae. aegypti*. Oaxaca City is 1550 meters above sea level. The highest recorded dengue outbreack reported in Mexico is at 1770 meter above sea level in 1988 (Herrera-Basto *et al.*, 1992), nonetheless 80% of the cases are concentrated < 600 meters above sea level (Escobar-Mesa and Gomez-Dantes, 2003). Mosquitoes collected in Oaxaca City also had smaller molecular diversity suggesting that mosquitoes collected at this site went through a diversity reduction.

Principal components analysis of the ND4 haplotypes indicated spatial autocorrelation. The prediction maps created from the principal components were used to examine the spatial distribution of vector competence. PC2 was significantly correlated with MIB. Even though the prediction maps present a patchy landscape for *Ae. aegypti* vector competence the correlograms suggest that the underlying process that

generates the variation has a spatial component. It follows that since vector competence is genetically controlled (Bennett *et al.*, 2003, 2004) the associated "genes" will also be spatially autocorrelated.

Dengue cases were found to be spatially autocorrelated at small distances (figure 36). This finding suggests that proximate municipalities are more likely to exchange dengue cases either through movement of infected people or mosquitoes. This observation also suggest that municipalities with large of number inhabitants are more likely to maintain dengue transmission. Similar results were reported for Veracruz state in the 1995-1998 dengue epidemics (Escobar-Mesa 2003). This is a surprising result since infected humans are consider to move considerable distances, additionally, *Ae. aegypti* is found ubiquitously in most cities at less than 600 meters above sea level.

The failed attempt of modeling dengue cases based in environmental variables can be a result of two main reasons: time and spatial autocorrelation. Modelling dengue cases requires analysis of an event that fluctuates considerably over space and time. This kind of modeling is referred as nonstationary modeling. The lack of correlation of dengue cases with what would seem to be obvious variables (e.g. human population density) could be the result of clumping and averaging data over time.

Cummings *et al.*, (2004) represent the only serious attempt to simultaneously analyze temporal and spatial variables in DHF cases. With the help of a massive database (850,000 infections occurring in 72 provinces of Thailand during the period 1983-1997), Cummings *et al.* (2004) examined the spatial–temporal dynamics of DHF incidence. They used a sophisticated, yet elegant, empirical mode decomposition method to show the existence of a spatial–temporal traveling wave in the incidence of DHF. The

"wave" of DHF cases emanates from Bangkok and moves radially at a speed of 148 km per month. Put in other words, dengue is spatial-temporally correlated. This finding provides an important starting point for detecting and characterizing the key processes that contribute to the spatial–temporal dynamics of DHF. In the near future we probably will be able assess the role other dynamic variables as temperature, rainfall or solar radiation have on dengue epidemics.

It appears that dengue cases in the state of Veracruz also move in a wave. Veracruz municipality is the first city to present a dengue case in the first week of 2004, followed by additional cases in the adjacent municipalities. In the 10^{th} week of the same year Martinez de la Torre presented dengue cases, this municipality is not adjacent and has a large number of inhabitants. In contrast, a closer municipality (Xalapa) has a larger number of inhabitants, but did not present cases until the 30^{th} week. There is a great deal of trade between these two municipalities (North American Transport, 2000) and dengue cases appear earlier in Martinez de la Torre than in Xalapa. By the 10^{th} week, the average temperature, as an effect of altitude, in Xalapa is 16.2 °C and in Martinez de la Torre is 21.1°C. Altitude could thus be responsible for the lag in dengue cases at Xalapa.

Regardless of the lack of a model to make inferences about the role the environment has on dengue epidemic, it is clear from the spatial analysis made in this chapter that dengue cases have a contagious distribution even at a statewide scale. Additionally, efforts aimed at reduce dengue cases in a city could have a larger impact than expected.

VI Conclusions

Chapter 2 and 3 focused upon association mapping. Not much can be concluded from chapter 2 in regards to results since it is a theoretical chapter. However, Mantel in 1967, was only able visualize an analysis of two 10 by 10 matrices with 10,000 permutations. Technology limited the methods for analyzing data. Most of the population genetics seminal work can be dated back to the 1960's. However, the first PC was not available until 1981. The first IBM PC ran on a 4.77 MHz Intel 8088 microprocessor. The PC came equipped with 16 kilobytes of memory, expandable to 256k. PCs came with one or two 160k floppy disk drives and an optional color monitor. The price tag started at $1,565, which would be nearly $4,000 today. The PC using the software presented in chapter 2 is more than 200 times faster than the first IBM and has 63 times more RAM memory. The output created by PGLD, for example, would require close to twenty-nine 160k disks. It easy to forget, due to our familiarity with computers, that many analyses were not possible 20 years ago. The mathematical

methods proposed in chapter 2 and executed in chapter 3, clearly demonstrate that analysis of large data sets can be achieved in a reasonable amount of time.

The principal rationale for association mapping is to understand phenotype, albeit indirectly, rather than genomes per se. Understanding this relationship between genotype (e.g., genomic data) and phenotype is a necessary step in successfully modeling complex traits. I believe that genotypes should be considered the class to which an organism belongs based upon its genes. However, more commonly we analyze partial genotypes. Now it is possible to obtain genotypic data from a myriad of sources such as DNA sequences, Single nucleotide polymorphisms (SNPs), Microsatellites and Protein sequences. On the other hand, phenotype is the class to which an organism belongs based upon its observable physical characteristics. Again, we also analyze partial phenotypes. It is clear that phenotype can be affected by factors such as heredity, epitasis, epigenesis, environment and experimental errors.

Association mapping should answer should answer what differences in phenotype can be explained by a genotype. It should determine particular genotypic features important in developing a particular phenotype. Additionally, it should measure how accurately a phenotype can be predicted based on genotype data. The software presented in this dissertation can answer these questions. However, there are many analytical challenges that need answering. Such as numerous levels of variables, mixture of variable types and potential non-additive interactions between variables.

In chapter 3 the role of ET in the susceptibility to dengue virus was assessed. Despite the fact, that no association was found, the apparent lack of association in a structural gene segregating sites does not preclude the possibility that other elements

influence transcription and possibly translation. Mutations in enhancers or promoters residing outside the analyzed region could also explain this lack of association. Our reductionist view of the "phenotype" mosquito-susceptibility could also be blamed. For example, it can be argued that the "phenotype" traffic jam contains patterns of behavior that cannot be reduced to the behavior of an individual car. Similarly, ET could have an important but not determinant role in the mosquito susceptibility phenotype. The observations made in chapter 3 suggest that the application of QTNs towards predicting vector competence in field populations may be strictly regional and, very possibly, have a seasonal or temporal component. Despite the lack of positive results, I believe modeling is required to quantify the effects variables have over complex traits. As with most models, in time the major variables and interactions that control the trait are revealed. Association mapping, as proposed in chapter 3 using the six criteria is not infallible; it is an "intelligent guess" approach to solve complex traits.

Chapter 4 and 5 deal mainly with gene flow, population structure and vector competence. In chapter 4, I used conventional population genetic approaches to assess them. I introduced spatial statistic tools and used them to address these problems in chapter 5. Darwin emphasized that population isolation was one important factor promoting population changes. However, there is a balance between gene flow and genetic drift. Genetic drift alone in completely isolated populations, will tend to fix alleles in the differentiated populations. Gene flow among populations will prevent fixation and it will avert substantial genetic differentiation due to drift.

The observations made in chapter 4, based in mtDNA data, appear to support genetic drift and/or founder's effect as responsible for the observed patterns. The results

further confirm what has been known for many years, *Aedes aegypti* shows a great deal of variation in VC for dengue viruses and that this variation is controlled by genetic factors. However, this dissertation demonstrated that VC can drastically change at small distances and that VC could be modified by breaks in gene flow.

Population genetics predicts that population genotypic variation will be modified by distance, thus phenotypes controlled by genes should also be affected by geography. Could the results in chapter 3 (the lack of association between early trypsin (ET) and vector competence (VC), and the association between ND4 and VC) be a result of spatial autocorrelation? Could ET and abundant trypsin (AT) be autocorrelated? Could autocorrelation set off Type I error rate in association mapping as the results suggested? I showed that MIB, SUC and ND4 are autocorrelated and most likely ET and AT are spatially autocorrelated.

Now this dissertation presents advancements and tools towards the application of landscape genetics. Manel *et al.* (2003) proposed the term landscape genetics in the frame of biological conservation, ecological understanding and evolution. This concept is open to other fields of the biological sciences and has direct application in disease transmission studies.

When Dr. Robert Sokal addressed the current state of statistical and other quantitative approaches to human biology, he pointed to five common statistical blunders (Sokal, 2004). Dr. Sokal considered as blunders the use of canned computer programs while ignoring the underlying assumptions, failure to employ multiple comparisons methods, ignoring spatial, temporal and phylogenetic autocorrelation, presenting distorted ordination plots and claims based on insufficient evidence. In chapter 5, I

delineated tools to test for spatially autocorrelation. Additionally, multivariable and multivariate approaches to summarize genetic data and test for autocorrelation were presented, all this as an effort to avoid such "blunders".

References

Afanasiev BN, Kuzlov YV, Carlson JO, Beaty BJ. 1994. Densovirus of *Aedes aegypti* as an expression vector in mosquito cells. *Exp Parasitol.* **79**:322-339.

Aitken THG, Downs WG, Shope RE, 1977. *Aedes aegypti* strain fitness for yellow fever virus transmission. *Am J Trop Med Hyg* **26**:985-990.

Anderson RA, Edman JD and Scott TW. 1990. Rubidium and cesium as host blood-markers to study multiple blood feeding by mosquitoes (Diptera: Culicidae). *J. Med. Entomol.* **27**:999-1001.

Apostol BL, Black WCIV, Reiter P, Miller BR. 1996. Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. *Heredity* **76**:325-334.

Aviles G, Rangeon G, Vorndam V, Briones A, Baroni P, Enria D, Sabattini MS. 1999. Dengue reemergence in Argentina. *Emerg Infect Dis.* **5**:575.

Baqar S, Hayes CG, Ahmed T. 1980. The effect of larval rearing conditions and adult age on the susceptibility of *Culex tritaeniorhynchus* to infection with west nile virus. *Mosquito News.* **40**:165.

Barnes C. and Cibula W. 1979. Some implications of remote sensing technology in insect control programs including mosquitoes. *Mosq. News* **39**:271-282.

Bates M and Roca-Garcia M. 1946. The development of the virus of yellow fever in *Haemagogus* mosquitoes. *Am J Trop Med Hyg.* **26**:585.

Beard CB, O'Neill SL, Tesh RB, Richards FF, Aksoy S. 1993. Modification of arthropod vector competence via symbiotic bacteria. *Parasitol Today.* **9**:179-183.

Beaty BJ and Bishop DH. 1988. Bunyavirus-vector interactions. *Virus Res.* **10**:289.

Beaty BJ. 2000. Genetic manipulation of vectors: a potential novel approach for control of vector-borne diseases. *PNAS.* **97**:10295-10297

Beck LR, Lobitz BM, Wood BL. 2000. Remote sensing and human health: New sensors and new opportunities. *Emerging Infectious Diseases* **6**:217-227.

Beck LR, Rodríguez MH, Dister SW, Rodríguez AD, Rejmankova E, Ulloa A, Meza RA, Roberts DR, Paris JF, Spanner MA, Washino RK, Hacker C, Legters LJ. 1994. Remote sensing as a landscape epidemiological tool to identify villages at high risk for malaria transmission. *Am. J. Trop. Med. Hyg.* **51**:271-280.

Beck LR, Rodríguez MH, Dister SW, Rodríguez AD, Washino RK, Roberts DR, Spanner MA. 1997. Assessment of a remote sensing based model for predicting malaria transmission risk in villages of Chiapas, Mexico. *Am. J. Trop. Med. Hyg.* **56**:99-106.

Becker WA. 1992. *Manual of Quantitative Genetics*, 5th ed. Academic Enterprises, Pullman, WA.

Bennett KE, Olson KE, Munoz MDL, Fernandez-Salas I, Farfan JA, Higgs S, Black WC IV, Beaty BJ. 2002. Variation in vector competence for dengue 2 virus among 24 collections of *Aedes aegypti* form Mexico and the United States. *Am J Trop Med Hyg.* **67**:85-92.

Bennett, K.E., B.J. Beaty, and W.C. Black IV. 2003. Selection of *D2S3,* an *Aedes aegypti* strain with high oral susceptibility to Dengue 2 virus and *D2MEB,* a strain with a midgut barrier to Dengue 2 escape. *Journal of Medical Entomology* (in press)

Bennett, K.E., Flick, D., Fleming, K.H., Beaty, B.J., Black, W.C., 2004. Quantitative Trait Loci that Control Dengue-2 Virus Dissemination in the Mosquito Aedes aegypti. Genetics (submitted).

Bertorelle G and Barbujan G. 1995. Analysis of DNA diversity by spatial autocorrelation. *Genetics.* **140**: 811-819

Besansky NJ, Collins FH. 1992. The mosquito genome: organization, evolution, and manipulation. *Parasitol Today.* **8**:186-192.

Black WC and Duteau NM. 1997. RAPD-PCR and SSCP Analysis for insect population genetic studies. In: Crampton, J., Beard, C.B., Louis, C. (eds), The Molecular Biology of Insect Disease Vectors: A Methods Manual, Vol. Chapman and Hall., New York., pp. 361-373

Black WC and Krafsur ES. 1985. A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. Theoretical and Applied Genetics, **70:**491-496.

Black WC IV and Munstermann LM. 1996. Molecular taxonomy and systematics of arthropod vectors, pp. 438-470 in *Biology of Disease Vectors,* edited by W. C. Marquart and B. Beaty. University Press of Colorado, Boulder.

Black WC IV, Baer CF, Antolin MF, DuTeau NM. 2001. Populations genomics: genome-wide sampling of insect populations. *Annual Rev Entomol.* **46**:441.

Black WC IV, Bennett KE, Gorrochotegui-Escalante N, Barillas-Mury C, Fernandez-Salas I, De Lourdes Munoz M, Farfan-Ale JA, Olson KE and Beaty BJ. 2002. Flavivirus susceptibility in *Aedes aegypti. Archives of Medical Research.* **33**:379-388.

Boromisa RD, Grimstad PR, Rai KS. 1987. Variation in the vector competence of geographic strains of *Aedes albopictus* for dengue 1 virus. *J Am Mosq Control Assoc.* **3**:378-386.

Bosio CF, Beaty BJ, Black WC, 1998. Quantitative genetics of vector competence for dengue-2 virus in *Aedes aegypti. Am J Trop Med Hyg* **59**:965-970.

Bosio CF, Fulton RE, Salasek MJ, Beaty BJ, Black WC, 2000. Quantitative trait loci that control vector competence for dengue-2 virus in the mosquito *Aedes aegypti. Genetics* **156**:687-698.

Bosio CF, Thomas RE, Grimstad PR, Rai KS. 1992. Variation in the efficiency of vertical transmission of dengue-1 virus by strains of *Aedes albopictus* (Diptera:Culicidae). *J Med Entomol.* **29**:985-989.

Breese EL and Mather K. 1957. The organization of polygenic activity within a chromosome in Drosophila. I. Hair characters. *Heredity.* **11**:337-395.

Briegel H. 1990. Metabolic relationship between body size, reserves, and fecundity of *Aedes aegypti. J Insect Physiol.* **36**:165-172.

Brown SE, Severson DW, Smith LA, Knudson DL. 2001. Integration of the Aedes aegypti mosquito genetic linkage and physical maps. *Genetics* **157:**1299-1305.

Brown TM and Brown AWA. 1974. Experimental induction of resistance to a juvenile hormone mimic. *J Econom Entomol.* **67**:799-801.

Burke DS and Monath TP. 2001. Flaviviruses, p. 1043-1126, In D.M. Knipe, and P.M. Howely (ed), Fields Virology 4th ed. Lippincott Williams and Wilkins, Philadelphia, PA.

Carbone DL and Black WC IV. 2000. Biplot analysis: a graphical technique to display relationships among genetic markers in large datasets. *Insect Molecular Biology* (in revision)

Cardon LR and Bell JI. 2001. Association study designs for complex diseases. *Nat Rev Genet* **2:**91-99.

Carlson JO, Olson KE, Higgs S, Beaty BJ. 1995. Molecular genetic manipulation of mosquito vectors. *Ann Rev Entomol.* **40**:359-388.

Castric V. *et al.* 2001 Landscape structure and hierarchical genetic diversity in the brook charr. Salvelinus fontinalus. *Evolution* **55:**1016–1028

Centers for Disease Control and Prevention (CDC). 1994. *Addressing emerging infectious disease threats: A strategy for the United States.* Atlanta, GA: U.S. Dept. of Health and Human Services, CDC.

Centers for Disease Control and Prevention (CDC). 1998. *Preventing emerging infectious diseases: A strategy for the 21st Century.* Atlanta, GA: U.S. Department of Health and Human Services.

Centers for Disease Control and Prevention (CDC). 2004. *West Nile Hopepage* Available from http://www.cdc.gov/ncidod/dvbid/westnile/index.htm

Chamberlain RW, Sudia WD. 1955. The effects of temperature upon the extrinsic incubation of eastern equine encephalitis in mosquitoes. *Am J Hyg.* **62**:295.

Chareonsook O, Foy HM, Teeraratkul A, Silarug N. 1999. Changing epidemiology of dengue hemorrhagic fever in Thailand. *Epidemiol Infect.* **122**:161.

Churchill GA, Doerge RW. 1994. Empirical threshold values for quantitative trait mapping. Genetics **138:**963-971.

Clarke T. 2002. Dengue virus: break-bone fever. *Nature.* **416**:672-674.

Clifford P, Richardson S and Hémon D. 1989. Assessing the significance of the correlation between two spatial processes. *Biometrics.* **45**:123-134.

Cockerham CC and Weir BS. 1977 Digenic descent measures for finite populations. Genetical Research. **30:**121- 147.

Collins FH. 1994. Prospects for malaria control through the genetic manipulation of its vectors. *Parasitol Today.* **10**:370-371.

Cordellier R, Bouchite, Rouche J-C, Monteny N, Diaco B, Akoliba P. 1983. Circulation selvatique du virus dengue 2 en 1980, dans les savanes sub-soudaniennes de Cote D'Ivoire. Cah ORSTOM, Ser *Entomo Med Parasitol.* **21**:165-179.

Cormier, R.T., Hong, K.H., Halberg, R.B., Hawkins, T.L., Richardson, P., Mulherkar, R., Dove, W.F., Lander, E.S., 1997. Secretory phospholipase Pla2g2a confers resistance to intestinal tumorigenesis. *Nat Genet* **17**:88-91.

Cornet M, Saluzzo J, Hervy H, Digoutte J, Germain M, Chauvancy M, Eryaud M, Ferrara L, Heme G, Legros F. 1984. Dengue 2 au Senegal oriental: une poussee epizootique en milieu selvatique; isolements du virus a partir de moustiques d'un singe et considerations epidemiologiques. Cah ORSTOM. *Ser Entomol Med Parasitol.* **22**:313-323.

Costero A, Edman JD, Clark GG, Scott TW. 1998. Life table study of *Aedes aegypti* (Diptera: Culicidae) in Puerto Rico fed only human blood versus blood plus sugar. *J Med Entomol.* **35**:809-813.

Cox CB and Moore PD. 1993. Biogeography: And ecological and evolutionary approach. 5th ed. Black Well Science, Inc., Cambridge, US

Crampton JM, Warren A, Lycett GJ, Hughs MA, Comley IP, and Eggleston P. 1994. Genetic manipulation of insect vectors as a strategy for the control of vector-borne disease. *Ann Trop Med Parasitol.* **88**:3-12.

Cross, E., R. Perrine, C. Sheffield, and G. Passaglia. 1984. Predicting areas endemic for schistosomiasis using weather variables and a Landsat data base. *Mil. Med.* **149**:542-544.

Cummings DAT, Irizarry RA, Huang NE, Endy TP, Nisalak A, Ungchusak K, BurkeDS. 2004. Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand *NATURE.* **427**:344-347

Daniel M and Kolar J. 1990. Using satellite data to forecast occurrence of the common tick *Ixodes ricinus* (L.). *J. Hygiene Epidem. Microbiol. Immun.* **34**:243-252.

de Kruif P. 1926. *Microbe Hunters*. Harcourt, Brace, NY.

DeCock KM, Monath TP, Nasidi A, Tukei PM, Enriquez J, Lichfield P, Craven RB, Fabiyi A, Okafor BC, Ravaonjanahary C, *et al.*. 1988. Epidemic yellow fever in eastern Nigeria, 1986. *Lancet.* **1**:630.

Deubel V, Kinney RM and Trent DW. 1986. Nucleotide sequences and duduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. *Virology.* **155**:365-377.

DGE 2001. Direccion General de Epidemiologia/Secretaria de Salud. Boletin Epidemiologico

DGE 2002. Direccion General de Epidemiologia/Secretaria de Salud. Boletin Epidemiologico

DGE 2003. Direccion General de Epidemiologia/Secretaria de Salud. Boletin Epidemiologia ISSN 1405-2636

DGE 2004. Direccion General de Epidemiologia/Secretaria de Salud. Boletin Epidemiologia. ISSN 1405-2636

Dister SW, Beck LR, Wood BL, Falco R and Fish D. 1993. The use of GIS and remote sensing technologies in a landscape approach to the study of Lyme disease transmission risk. *Proc., Seventh Annual Symposium on GIS in Forestry, Environment and Natural Resources Management*, 15-18 February 1993, Vancouver, British Columbia, Canada.

Dister SW, Fish D, Bros S, Frank DH and Wood BL. 1997. Landscape characterization of peridomestic risk for Lyme disease using satellite imagery. *Am. J. Trop. Med. Hyg.* **57**(6):687-692.

Doerge RW and Churchill GA. 1996 Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**:285-294.

Dove A. 1998. Dengue fever on the increase. *Nat. Med.* **4**:543.

Edman JD and Scott TW. 1987. Host defensive behavior and the feeding success of mosquitoes. *Insect Science and its Application.* **8**:617-622.

Edman JD, Scott TW, Costero A, Morrison AC, Harrington LC, Clark GG, 1998. *Aedes aegypti* (L.) (Diptera: Culicidae) movement influenced by availability of oviposition sites. *J Med Entomol* **35**:578–583.

Epperson BK. 1990. Spatial autocorrelation of genotypes under directional selection. *Genetics.* **124**:757-771

Epperson BK. 1995a. Spatial distribution of genotypes under isolation by distance. *Genetics.* **140**: 1431-1440

Epperson BK. 1995b. Spatial structure of two-locus genotypes under isolation by distance. *Genetics.* **140**:365-375

Epstein PR. 1995. Emerging diseases and ecosystem instability: new threats to public health. *Am J Public Health.* **85**:168-172.

Escalante T, Rodriguez G and Morrone JJ. 2004. The diversification of Nearctic mammals in the Mexican transition zone. *Biological Journal of the Linnean Society.* **83**:327–339

Escobar-Mesa J, Gómez-Dantés H. 2003. Determinantes de la transmisión de dengue en Veracruz: un abordaje ecológico para su control. *Salud Publica Mex.* **45**:43-53.

Excoffier L, Smouse PE, Quattro JM, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* **131**:479–491.

Falconer DS ane Mackay TFC. 1996. *Introduction to Quantitative Genetics*, 4th ed. Longman Group Ltd., Essex UK.

Fan W, Yu S, Cosgriff T. 1989. The reemergence of dengue in China. Rev Infect Dis 2 (Supplement 4): S847-S853.

Felsenstein J. 1993. *Phylogeny Inference Package.* Version 3.5C. Seattle, WA: University of Washington.

Focks DA, Brenner RJ, Hays J, Daniels E. 2000. Transmission thresholds for dengue in terms of *Aedes aegypti* pupae per person with discussion of their utility in source reduction efforts. *Am J Trop Med Hyg.* **62**:11-18.

Focks DA, Daniels E; Haile DG and Keesling JE. 1995. A simulation model of the epidemiology of urban dengue fever: Literature analysis, model development, preliminary validation and examples of simulation results. *Am J. of Trop. Med. And Hyg.* **53(3)**: 489-506.

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**:85-88.

Fridman E, Pleban T, Zamir D. 2000. A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc *Natl Acad Sci USA* **97:**4718-4723.

Fu YX and Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* **133:**693-709.

Gabriel, K. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika.* **58**:453-467.

Githeko AK, Lindsay SW, Confalonieri UE, Patz JA. 2000. Climate change and vector-borne diseases: a regional analysis. *Bulletin of the World Health Organization.* **78**:1136-1147.

Glass GE, Morgan III JM, Johnson DT, Noy PM, Israel E, and Schwartz BS. 1992. Infectious disease epidemiology and GIS: a case study of Lyme disease. *GeoInfo Systems* **2**:65-69.

Goldman IF, Arnold J, Calrton BC. 1986. Selection for a resistance *to Bacillus thuringiensis* subspecies *israelensis* in field and laboratory populations of the mosquito *Aedes aegypti. J Invert Path.* **47**:317-324.

Gomez-Machorro C, Bennett KE, Munoz ML, Black IV WC. 2004. Quantitative Trait Loci affecting Dengue midgut infection barriers in an advanced intercross line of *Aedes aegypti. Insect Molecular Biology* (In Press).

Goodman CA and Mills AJ. 1999. Review article: the evidence base on the cost-effectiveness of malaria control measures in Africa. *Health Policy and Planning.* **14**: 301-312.

Gordon AJ. 1988. Mixed strategies in the health education community participation. An evaluation of Dengue control in the Dominican Republic. *Health and Education Research.* **3** 399-419.

Gorrochotegui-Escalante N, Lourdes Munoz M, Fernandez-Salas I, Beaty BJ, Black WC, 2000. Genetic isolation by distance among *Aedes aegypti* populations along the northeastern coast of Mexico. *Am J Trop Med Hyg* **62**:200-209.

Gorrochotegui-Escalante N., Gomez-Machorro C., Lozano-Fuentes S., Fernandez-Salas I, Lourdes Munoz M, Farfan-Ale JA, Garcia-Rejon J, Beaty BJ, and Black WC, 2002. Breeding structure of Aedes aegypti populations in Mexico varies by region. *Am J. Trop. Med. Hyg.* **66:**213-222.

Gould DJ, Yuill TM, Moussa MA, Simasathien P, Rutledge LC. 1968. An insular outbreak of dengue hemorrhagic fever. III. Identification of vectors and observation on vector ecology. *Am J Trop Med Hyg.* **17**:609-618.

Gould EA, Buckley A, Cammack N, Barrett ADT, Clegg JCS, Ishak R, and Varma MGR. 1985. Examination of the immunological relationships between flaviviruses using yellow fever virus monoclonal antibodies. *J. Gen. Virol.* **66**:1369-1382.

Gould SJ. 1995 Ghosts of bell curves past: the mismeasure of man continues, as a current bestseller revives academic racism's old arguments. *Nat. Hist.* **104**:12-19.

Graham DH, Holmes JL, Higgs S, Beaty BJ, Black WC IV. 1999. Selection of refractory and permissive strains of *Aedes triseriatus* (Diptera: Culicidae) for transovarial transmission of La Crosse virus. *Journal of Medical Entomology.* **36**. 671-678.

Graham H. 1903. The Dengue. A study of its pathology and mode of propagation. *J Trop Med Hyg London.* **6**:209-214.

Gratz NG. 1999. Emerging and resurging vector-borne diseases. *Annual Rev Entomol.* **44**:51.

Grimstad PR and Haramis LD. 1984. *Aedes triseriatus* (Diptera:Culicidae) and La Crosse virus. III. Enhanced oral transmission by nutrition-deprived mosquitoes. *J Med Entomol.* **21**:249-256.

Grimstad PR, Craig GB Jr, Ross QE, Yuill TM. 1977. *Aedes triseriatus* and La Crosse virus: geographic variation in vector susceptibility and ability to transmit. *Am J Trop Med Hyg.* **26**:990-996.

Grimstad PR, Paulson SL, Craig GB Jr. 1985. Vector competence of *Aedes hendersoni* (Diptera:Culicidae) for La Crosse virus and evidence of a salivary-gland escape barrier. *J Med Entomol.* **22**:447.

Grimstad PR, Walker ED. 1991. Aedes triseriatus (Diptera:Culicidae) and La Crosse virus IV. Nutritional deprivation of larvae affects adult barriers to infection and transmission. *J Med Entomol.* **28**:378-386.

Gubler DJ and Clark GG. 1995. Dengue/dengue hemorrhagic fever: the emergence of a global health problem. *Emerg Inf Dis.* **1**:55-57.

Gubler DJ and Rosen L. 1976. Variation among geographic strains of *Aedes albopictus* in susceptibility to infection with dengue viruses. *Am J Trop Med Hyg.* **25**:318.

Gubler DJ and Trent DW. 1994. Emergence of Epidemic Dengue/Dengue Hemorhagic Fever as a Public Health Problem in the Americas. Infectious Agents and Disease. **2**:383-393.

Gubler DJ, 1998. Resurgent vector-borne diseases as a global health problem. *Emerging Infectious Diseases* **4**:442-450.

Gubler DJ, Nalim S, Tan R, Saipan H, Saroso JS. 1979. Variation in susceptibility to oral infection with dengue viruses among geographic strains of *Aedes aegypti. Am J Trop Med Hyg.* **28**:1045-1052.

Gubler DJ, Novak RJ, Vergne E, Colon NA, Velez M, Fowler J. 1985. *Aedes (Gymnometopa) mediovittatus* (Diptera:Culuicidae), a potential maintenance vector of dengue viruses in Puerto Rico. *J Med Entomol.* **5**:469-475.

Gubler DJ. 1988. Dengue. In: *The Arboviruses: Epidemiology and Ecology, Volume II.* TP Monath (ed). CRC Press, Boca Raton, FL. pp 223-260.

Gubler DJ. 1993. Emergent and Resurgent Arboviral disease as Public Health Problems U. S. Department of Health and human Dervices, Public health Service CDC: 257–273.

Gubler DJ. 1996. The global resurgence of arboviral diseases. *Trans Roy Soc Trop Med Hyg.* **90**:449-451.

Gubler DJ. 2002. Epidemic Dengue/Hemorragic fever as a public health, social and economic problem of the 21st century. *Trends Microbiol.* **10:**100-103.

Gubler DJ. 2002a. The global emergence/resurgence of arboviral diseases as public health problems. *Arch Med Res* **33:**330-342.

Halffter G. 1983. Biogeography of the montane entomofauna of Mexico and Central America. *Ann. Rev. Entomol.* **32**:95-114

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**:95-98.

Halstead SB and Porterfield J. 1980. Enhancement of dengue Virus Infection in monocytes by Flavivirus antisera. *Am. J. Prop. Med. Hyg.* **29**:638-642.

Halstead SB, Scanlon JE, Umpaivit P, Udonsakdi A. 1969. Dengue and Chickungunya virus infection in man in Thailand, 1962–1964. IV. Epidemiologic studies in the Bangkok metropolitan area. *Am J Trop Med Hyg* **18**:997–1033.

Halstead SB. 1982. Immune enhancement of viral infection. *Prog Allergy.* **31**:301-364.

Halstead SB. 1988. Pathogenesis of dengue: challenges to molecular biology. *Science* **239**:476-481.

Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege EO. 2002. African Pastoralism: Genetic Imprint of origins and Migrations. *Nature.* **296**: 336-339

Hardy JL. 1988. Susceptibility and resistance of vector mosquitoes. Monath TP, ed. *Arboviruses. Epidemiology and Ecology, Volume I.* Boca Raton, FL: CRC Press., pp 87-126.

Harrington LC, Edman JD, Scott TW. 2001. Why do female *Aedes aegypti* (Diptera: Culicidae) feed preferentially and frequently on human blood? *J Med Entomol.* **38(3)**:411-422.

Hartl DL, Clark AG. 1989. *Principles of Population Genetics* 2nd ed. Sinauer Associates, Inc., Sunderland, MA.

Hay S, Tucker C, Rogers D, Packer M. 1996. Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease. *Ann. Trop. Med. Parasitol.* **90**(1):1-19.

Hayes RO, Maxwell EL, Mitchell CJ, Woodzick TL. 1985. Detection, identification and classification of mosquito larval habitats using remote sensing scanners in earth-orbiting satellites. *Bull. World Health Org.* **63**:361-374.

Heinz FX, Collett MS, Purcell RH, Gould EA, Howard CR, Houghton M, Moormann RJM, Rice CM, Theil HJ. 2000. Family Flaviviridae, pp 859-878. In MH Van Regenmortel, CM Fauquet, CM Bishop, DHL Carstens, E Estes, MK Lemon, S Maniloff, J Mayo, MA McGeoch, D Pringle, CR and RB Wickner (ed). Virus Taxonomy. 7th Report of the International Committee for the Taxonomy of Viruses. Academic Press. San Diego, CA.

Hemingway J, Field L and Vontas J. 2002. An overview of insecticide resistance. *Science.* **298**:96-97.

Henschal EA, Putnak JR. 1990. The dengue viruses. *Clin Microbiol Rev.* **3**:376-396.

Herrara-Basto E, Prevots DR, Zarate ML, SILVA JL, Sepulveda-Amor J. 1992. First reported outbreak of classical dengue fever at 1,700 meters above sea level in Guerrero State, Mexico, June 1988. *Am. J. Trop. Med. Hyg.* **46**:649-653

Higgs S and Beaty BJ. 1996. Rearing and containment of mosquito vectors. The Biology of Disease Vectors. 595-605.

Holmes, Eddie C. and Burch, Susanna S. 2000. The causes and consequences of genetic variaion in dengue virus. *Trends in Microbiology*. **8(2)**:74-77.

Houk EJ, Kramer LD, Hardy JL, Presser SB. 1986. An interspecific mosquito model for the mesenteronal infection barrier to western equine encephalymyelitis virus (*Culex tarsalis* and *Culex pipiens*). *Am J Trop Med Hyg*. **35**:632.

Houle D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics*. **130**:195-204.

Hudson RR. 1987. Estimating the recombination parameter of a finite population model without selection. *Genet. Res.* **50:**245-250.

Hugh-Jones M, Barre N, Nelson G, Wehnes K, Warner J, Garvin J, and Garris G. 1992. Landsat-TM identification of *Amblyomma variegatum* (Acari: Ixodidae) habitats in Guadeloupe. *Remote Sens. Environ.* **40**:43-55.

Hugh-Jones M. 1989. Applications of remote sensing to the identification of the habitats of parasites an disease vectors. *Parasitology Today*. **5(8)**:244-251.

Hugh-Jones M. 1991. The remote recognition of tick habitats *Journal of Agricultural Entomology*. **8(4)**:309-316.

Hurlbut HS. 1973. The effect of environmental temperature upon the transmission of St. Louis encephalitis virus by *Culex pipiens quinquefasciatus*. *J Med Entomol*. **10**:1-12.

Jupp PG. 1985. *Culex theileri* and Sindbis virus; salivary gland infection in relation to transmission. *J Am Mosq Control Assoc.* **1**:374.

Kerschner JH, Vorndam AV, Monath TP, Trent DW. 1986. Genetic and epidemiologic studies of dengue type 2 viruses by hybridization using synthetic deoxyoligonucleotides as probes. *J Gen Virol*. **67**:2645-2666.

Keyghobadi N. 1999. Influence of landscape on the population genetic structure of the alpine butterfly Parnassius smintheus (Papilionidae). *Mol. Ecol.* **8**:1481–1495

Kitron U, Pener HP, Costin C, Orshan L, Greenber Z and Shalom U. 1994. Geographic Information System in Malaria Surveillance; Mosquito Breedeng and Imported cases in Israel, 1992. *Am. J. Trop. Med. and Hyg.* **50**:550-556.

Kitron U. 1998. Landscape ecology and epidemiology of vector-born disease: tools for spatial analysis. *Journal of medical Entomology*. **35**: 435-445.

Kitron U. and Kazmierczak JJ. 1997. Spatial analysis of the distribution of Lym disease in Wisconsin. *American Journal of Epidemiology*. **145**:558-566.

Kramer LD, Hardy JL, Presser SB, Houk EJ. 1981. Dissemination barriers for western equine encephalomyelitis virus in *Culex tarsalis* infected after ingestion of low viral doses. *Am J Trop Med Hyg*. **30**:190-197.

Kramer LD, Hardy JL, Presser SB. 1983. Effect of temperature of extrinsic incubation on the vector competence of Culex tarsalis for western equine encephalomyelitis virus. *Am J Trop Med Hyg*. **32**:1130-1139.

Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22:**139-144.

Kuno G. 1997. Factors influencing the transmission of dengue viruses. DJ Gubler and G Kuno (eds). Dengue and Dengue Hemorrhagic Fever. CAB New York., pp 61-88.

Lepiniec L, Dalgarno L, Huong VT, Monath TP, Digoutte JP, Deubel V. 1994. Geographic distribution and evolution of yellow fever viruses based on direct sequencing of genomic cDNA fragments. *J. Gen. Virol.* **75**:417.

Lewontin RC and Kojima K. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution* **14**:458-472.

Linthicum KJ, Anyamba A, Tucker CJ, Kelley PW, Myers MF and Peters CJ. 1999. Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science* **285**:397-4000.

Linthicum KJ, Bailey CL, Davies FG and Tucker CJ. 1987. Detection of Rift Valley fever viral activity in Kenya by satellite remote sensing imagery. *Science* **235**:1656-1659.

Loiselle R, Sork V, Nason J. 1995. Spatial genetic structure of tropical understory shrub, *Psychotria officinalis* (Rudiaceae). *Am. J. Bot.* **82**:1420-1425

Long AD and Langley CH. 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res.* **9**:720-731.

Long AD, Lyman RF, Langley CH, Mackay TF. 1998. Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in Drosophila melanogaster. *Genetics* **149**:999-1017.

Long AD, Mullaney SL, Mackay TF, Langley CH. 1996. Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in Drosophila melanogaster. *Genetics* **144**:1497-1510.

Lorenz L, Beaty BJ, Aitken THG, Wallis GP, and Tabachnick WJ, 1984. The effect of colonization upon *Aedes aegypti* susceptibility to oral infection with yellow fever virus. *Am J Trop Med Hyg* **33**:690-694.

Lozano-Fuentes S. 2000. Análisis espacial y temporal de los casos de dengue en el area metropolitana d Monterrey N.L., México, durante 1998. Universidad Autonoma de Nuevo Leon, Mexico. Master of Science Thesis.

Lozano-Fuentes, S., Gorrochotegui-Escalante, N., Black, W.C., 2004. PGenome: Association mapping software. *Bioinformatics* (Submitted).

Ludwig GV, Christensen BM, Yuill TM, Schultz KT. 1989. Enzyme processing of La Crosse virus glycoprotein G1: a bunyavirus-vector infection model. *Virology* **171**:108-113.

Ludwig GV, Israel BA, Christensen BM, Yuill TM, Schultz KT. 1991. Role of LaCrosse virus glycoproteins in attachment of virus to host cells. *Virology.* **181**:564.

Ludwig GV, Israel BA, Christensen BM, Yuill TM, Schultz KT. 1991. Monoclonal antibodies directed against the envelope glycoproteins of La Crosse virus. *Microb Pathog* **11**:411-421.

Lukens LN, Doebley J. 1999. Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genetical Research* **74**:291-302.

Luo ZW, Tao SH, Zeng ZB. 2000. Inferring linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Genetics* **156**:457-467.

MacDonald G. 1973. *Dynamics of Tropical Disease.* Bruce-Chwatt LJ and Glanville VJ (eds.), Oxford University Press.

Mackay TF. 2001a. The genetic architecture of quantitative traits. *Annu Rev Genet* **35:**303-339.

Mackay TF. 2001b. Quantitative trait loci in Drosophila. *Nat Rev Genet* **2:**11-20.

Malone, J, Hub O, Fehler D, Wilson P, Wilensky D, Holmes R and Elmagdoub A. 1994. Temperature Data From Satellite Imagery and the Distribution of Schistosomiasis in Egypt. *Am. J. Trop. Med. and Hyg.* **50(6)**:714-722.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* **27:**209-220.

Mather K and Jinks JL. 1982. *Biometrical Genetics*, 3rd ed. Chapman and Hall, London.

Mattingly PF, Stone A and Knight KL. 1962. *Culex aegypti* Linnaeus, 1972 (INSECTA, DIPTERA); Proposed validation and interpretation under the plenary powers of the species named. *Bull. Zool. Nomencl.* **19** part 4.

McBratney AB and Webster R. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science* **37:**617-639.

Meade MS, Florin JW, Gesler WM. 1988. *Medical Geography.* The Guilford Press, New York.

Mellor PS, Jennings DM, Braverman Y, Boorman J. 1981. Infection of Israeli culicoides with African horse sickness, blue tongue and akabane viruses. *Acta Virol* **25**:401-407.

Meltzer MI, Rigau-Perez JG, Clark GG, Reiter P, and Gubler G. 1998. Using disability-adjusted life years to asses the economic impact of dengue in Puerto Rico: 1984-1994. *Am J Trop Med Hyg.* **59**:265-271.

Mertens PP, Burroughs JN, Walton A, Welby MP, Fu H, O'Hara RS, Brookes SM, Mellor PS. 1996. Enhanced infectivity of modified bluetongue virus particles for two insect cell lines and for two *Culicoides* vector species. *Virology.* **217**:582.

Metselaar D, Grainger CR, Olei KG, Reynolds DG, Pudney M, Leake CJ, Tukei PM, D'Offay RM, Simpsin DIH. 1980. An outbreak of type 2 dengue fever in the Seychelles, probably transmitted by *Aedes albopictus* (Skuse). *Bull WHO.* **58**:937-943.

Miller BR and CJ Mitchell. 1991. Genetic selection of a flavivirus-refractory strain of the Yellow Fever mosquito *Aedes aegypti. Am. J. Trop Med Hyg.* **45**:399-407.

Miller BR, Beaty BJ, Lorenz LH. 1982. Variation of La-Crosse virus filial infection rates in geographic strains of *Aedes triseriatus* diptera Culicidae. *Journal of medical entomology* **19**:213-214

Miller BR, Monath TP, Tabachnick WJ, Ezike VI. 1989. Epidemic yellow fever caused by an incompetent mosquito vector. *Trop Med Parasitol.* **40**:396-399.

Molina-Cruz A, Gupta L, Richardson JR, Bennett KE, Black IV WC and Barillas-Mury C. 2004. Inhibition Of Midgut Trypsin Activity During Blood Feeding Decreases Dengue-2 Virus Infection And Dissemination In Aedes aegypti. *Am. J. Trop. Med. Hyg.*, (in press)

Monath TP and Heinz FX. 1996. Flaviviruses. *Fields Virology 3rd ed.* Philadelphia: Lippincott-Raven Publishers. pp 961-1034.

Monath TP. 1994. Dengue: the risk to developed and developing countries. *Proc. National Acad Sci* **91**:2395-2400.

Monath TP. 1994. Vector-borne emergent disease. *An NY Acad Sci.* **740**:126.

Morens DM, Rigau-Perez JG, Lopez-Correa RH, Moore CG, Ruiz-Tiben EE, Salther GE, Chiriboga J, Eliason DA, Casta-Velez A, Woodall JP and Dengue Outbreak Investigation Group. 1986. Dengue in Puerto Rico, 1977: Public Health response to characterize and control an epidemic of multiple serotypes. *Am. J. of Trop. Med. and Hyg.* **35**:197-211.

Morrison AC, Costero A, Edman JD, and Scott TW. 1999. Increased fecundity of female *Aedes aegypti* (Diptera: Culicidae) fed human blood before release in a mark-recapture study in Puerto Rico. *J. Amer. Mosq. Control Assoc.* (Barr Memorial). **15**:98-104.

Morrison AC, Getis A, Santiago M, Rigau-Perez J, Reiter P. 1998. Exploratory space-time analysis of reported dengue cases during an outbreak in Florida, Puerto Rico, 1991–1992. *Am J Trop Med Hyg* **58**:287–298.

Mouches C, Magnin M, Berge JB, De Silvestri M, Beyssat V, Pasteur N, Georghiou GP. 1987. Over production of detoxifying esterases in organophosphate-resistant *Culex* mosquitoes and their presence in other insects. *Proc Nat Acad Sci USA.* **82**:2113-2116.

Muttitanon W, Kongthong P, Kongkanon C, Yoksan S, Gonzalez JP and Babazan P. Spatial and temporal dynamics of Dengue Hemorrhagic Fever Epidemics (Nakhon Pathom province, Thailand, 1997-2001). www.GISdevelopment.net Newslatter. http://www.gisdevelopment.net/application/health/planning/healthp0010.htm

Nasidi A, Monath TP, DeCock K, Tomori O, Cordellier R, Olaleye OD, Harry TO, Adeniyi JA, Sorungbe AO, Ajose-Coker AO, *et al.*. 1989. Urban Yellow fever epidemic in western Nigeria. *Trans R Soc Trop Med Hyg.* **83**:401.

Nathan MB. 1991. The Status of Aedes aegypti Control Programs in the Caribbean. Annual Reunion of the American Mosquito Control Association, New Orleans. pp. 12.

Nei M, 1987. *Molecular Evolutionary Genetics.* New York: Columbia University Press.

Nevo E, Honeycutt RL, Yonekawa H, Nelson K and Hanzawa N. 1993. Mitochondrial DNA polymorphism in subterranean mole-rats of the *Spalax ehrenbergi* superspecies in Israel and it peripheral isolates. *Mol. Biol. Evol.* **10**:590-604

Noriega FG, Edgar KA, Goodman WG, Shah DK, Wells MA. 2001. Neuroendocrine factors affecting the steady-state levels of early trypsin mRNA in *Aedes aegypti. J Insect Physiol* **47**:515-522.

North American Transportation in Figures. 2000. U.S. Department of Transportation, Bureau of Transportation Statistics, U.S. Department of Commerce, Census Bureau; Statistics Canada; Transport Canada; Instituto Mexicano del Transporte; Instituto Nacional de Estadística, Geografía e Informática; and Secretaría de Comunicaciones y Transportes, *North American Transportation in Figures*, BTS00-05, Washington, DC.

Ohta T and Kimura M. 1969. Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* **63**:229-238.

Ohta T. 1982a. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc Natl Acad Sci USA* **79**:1940-1944.

Ohta T. 1982b. Linkage disequilibrium with the island model. *Genetics* **101**:139-155.

Olson KE, Myles KM, Seabaugh RC, Higgs S, Carlson JO, Beaty BJ. 2000. Development of a Sindbis virus expression system that efficiently expresses green fluorescent protein in the midguts of *Aedes aegypti* following per os infection. *Insect Mol Biol.* **9**:57-65.

Ordonez JG, Fernandez Salas I, Flores-Leal A. 1997. Monitoring dispersal of marked Aedes aegypti females under field conditions using sticky ovitraps in Monterrey, northeastern Mexico. *J Am Mosq Control Assoc* **13**:121.

Packer MJ and Corbet PD. 1989. Size variation and reproductive success of females of *Aedes punctor* (Diptera: Culicidae). *Ecological Entomol.* **14**:297-309.

PAHO. 1997. The feasibility of eradicating Aedes aegypti in the Americas. Rev Panam Salud Publica. [online]. Jan. 1997, vol.1, no.1 [cited 18 April 2005], p.68-72. Available from World Wide Web: <http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S1020-49891997000100023&lng=en&nrm=iso>.

Patz JA, Epstein PR, Burke TA, Balbus JM. 1995. Global climate change and emerging infectious diseases. *JAMA.* **175**:217-223.

Paulson SL, Grimstad PR, Craig GB Jr. 1989. Midgut and salivary gland barriers to LaCrosse virus dissemination in mosquitoes of the *Aedes triseriatus* group. *Med Vet Entomol.* **3**:113.

Pavlovsky EN. 1966. *The natural nidality of transmissible disease* (N.D. Levine, ed.). University of Illinois Press, Urbana.

Peña G, Svenkerud E, Liszka B, Hendricks K, Rawlings J, 1999. Underdiagnosis of Dengue - Laredo, Texas, 1999. *MMWR* **50**:57-9.

Perry DJ and Knowles P. 1991. Evolutionary implications of allozyme and RAPD variation in the diploid population of buffalograss (*Buchloe dactylodes* (Nutt. Engelm.). *Mol. Ecol.* **4**:135-147

Piertney S. *et al.* 1998. Local genetic structure in red grouse (*Lagopus lagopus scoticus*): evidence from microsatellite DNA markers. *Mol. Ecol.* **7**:1645–1654

Piglucci M and Barbujani G. 1991. Geographical patterns of gene frequencies in Italian populations of *Ornithogalum montanum* (Liliaceae). *Genet. Res.* **58**:95–104

Platt KB, Mangiafico JA, Rocha OJ, Zaldivar ME, Mora J, Trueba G and Rowley WA, Detection of Dengue Virus Neutralizing Antibodies in Bats from Costa Rica and Ecuador. 2000. *J. Med. Entomol.* **37(6)**:965-967

Pope KO, Sheffner EJ, Linthicum KJ, Bailey CL, Logan TM, Kasischke ES, Birney K, Njogu AR and Roberts CR. 1992. Identification of central Kenyan Rift Valley fever virus vector habitats with Landsat TM and evaluation of their flooding status with airborne imaging radar. *Remote Sens. Environ.* **40**:185-196.

Powers AM, Kamrud KI, Olson KE, Higgs S, Carlson JO, Beaty BJ. 1996. Molecularly engineered resistance to California serogroup virus replication in mosquito cells and mosquitoes. *Proc Nat Acad Sci.* **93**:4187-4191.

Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH, Feyereisen R. 2002a. Evolution of supergene families associated with insecticide resistance. *Science.* **298**:179-181.

Ranson H, Nikou D, Hutchinson M, Wang X, Roth CW, Hemingway J, and Collins FH. 2002b. Molecular analysis of multiple cytochrome P450 genes from the malaria vector, *Anopheles gambiae. Insect molecular Biology.* **11**:409-418.

Reed W, Caroll J, Agramonte A, Leazar JW. 1900. The etiology of yellow fever – a preliminary note. *Proceedings of the 28th Annual Meeting American Public Health Association.* Indianapolis, IN:3-19

Rice C, 1996 *Flaviviridae*: The viruses and their replication . *Fields Virology 3rd ed.* Philadelphia: Lippincott-Raven Publishers. pp 931-960.

Richardson S and Hémon D. 1981. On the variance of the sample correlation between two independent lattice processes. Journal of applied Probability **18**:943-948

Rico-Hesse R, Harrison LM, Alba-Salas R, Tovar D, Nisalak A, Ramos C, Boshell J, R de Mesa MT, Nogueira RMR, and Travassos da Rosa A. 1997. Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* **230**:244-251.

Rico-Hesse R. 1990. Molecular evolution and distribution of dengue viruses type-1 and type 2 in nature. *Virology.* **174**:479-493.

Risch NJ. 2000. Searching for genetic determinants in the new millennium. *Nature* **405**, 847-856.

Roach JL *et al.* 2001. Genetic structure of ametapopulation of blacktailed prairie dogs. *J. Mammal.* **82**:946–959

Rodhain F, Rosen L, 1997. Mosquito vectors and dengue virusvector relationships. Gubler DJ, Kuno G, eds. *Dengue and Dengue Hemorrhagic Fever.* New York: CAB International, 61–88.

Rodriguez-Figueroa L, Rigau-Perez JG, Suarez EL and Reiter P. 1995. Risk factors for dengue infection during an outbreak in Yanes, Puerto Rico in 1991. *Am. J. of Trop. Med. and Hyg.* **52**:496-505.

Rogers DJ and Randolph SE. 1991. Mortality rates and population density of tsetse flies correlated with satellite imagery. *Nature* **351**:739-741.

Rogers DJ, and Randolph SE. 1993. Distribution of tsetse and ticks in Africa, past, present and future. *Parasitol. Today* **9**:266-271.

Rogers DJ, Hay SI and Packer MJ. 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Ann. Trop. Med. Parasitol.* **90**:225-241.

Rosen L. 1958. Experimental infection of New World monkeys with dengue and yellow fever viruses. *Am J Trop Med Hyg.* **7**:406-410.

Rosen, L, Roseboom LE, Gubler DJ, Lein JC, and Chaniotis BN, 1985. Comparative susceptibility of mosquito species and strains to oral and parenteral infection with dengue and Japanese encephalitis viruses. *Am J Trop Med Hyg* **34**:603-615.

Rosenberg MS. 2001. PASSAGE. Pattern Analysis, Spatial Statistics, and Geographic Exegesis. Version 1.1. Department of Biology, Arizona State University, Tempe, AZ.

Rousset F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics 145:* 1219–1228.

Royal Statistical Society. 2001. Royal Statistical Society concerned by issues raised in Sally Clark case. Embargoed until 00.00hrs Tuesday 23 October 2001. News release. http://www.rss.org.uk/docs/Royal%20Statistical%20Society.doc

Rozas J, Rozas R, 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics 15:* 174–175.

Rudnick A. 1965. Studies on the ecology of dengue in Malaysia: a preliminary report. *J Med Entomol.* **2**:203-208.

Rudnick A. 1978. Ecology of dengue virus. *Asian J Inf Dis.* **2**:156-160.

Sabin AB. 1952. Research on dengue during World War II. *Am J Trop Med Hyg.* **1**:30-50.

Sax K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phasiolus vulgaris* (Leguminosae). *Genetics.* **8**:552-556.

Schnabel A, Laushamn RH and Hamrick JL. 1991. Comparative genetic structure of two co-occurring tree species, *Maclura pomifera* (Moraceae) and *Gleditsi triacanthos* (Leguminosae). *Heredity.* **67**:357-364

Schoepp RJ, Beaty BJ, Eckels KH. 1990. Dengue 3 virus infection of *Aedes albopictus* and *Aedes aegypti*: comparison of parent and progeny candidate vaccine viruses. *Am. J. Trop. Med. Hyg.* **42(1)**:89-96.

Scott TW, Amerasinghe PH, Morrison AC, Lorenz LH, Clark GG, Strickman D, Kittayapong P, Edman JD. 2000b. Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: Blood feeding frequency. *J Med Entomol.* **37**:89-101.

Scott TW, Amerasinghe PH, Morrison AC, Lorenz LH, Clark GG, StrickmanD, Kittayapong P, Edman JD. 2000a. Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: Blood feeding frequency *Journal of Medical Entomology* **37**:89-101.

Scott TW, Clark GG, Lorenz LH, Amerasinghe PH, Reiter P, Edman JD. 1993. Detection of multiple bloodfeeding in *Aedes aegypti* (Diptera:Culicidae) during a single gonotrophic cycle using a histologic technique. *J Med Entomol.* **30**:94-99.

Scott TW, Morrison AC, Lorenz LH, Clark GG, Strickman D, Kittayapong P, Zhou h, Edman JD. 2000a.. Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: population dynamics. *J Med Entomol.* **37**:77-88.

Scott TW, Morrison AC, Lorenz LH, Clark GG, Strickman D, Kittayapong P, Zhou H, Edman JD. 2000b. Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: Population dynamics *Journal of Medical Entomology.* **37**:77-88.

Scott TW, Naksathit A, Day JF, Kittayapong P, Edman JD. 1997. A fitness advantage for *Ae aegytpi* and the viruses it transmits when females feed only on human blood. *Am J Trop Med Hyg.* **57**:235-239.

Slatkin M and Maddison WP. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics.* **123**:603-613

Slatkin M and Maddison WP. 1990. Detecting isolation by distance using phylogenies of genes. *Genetics.* **126**:249-260

Slatkin M. 1989. Detecting a small amount of gene flow from phylogenies of alleles. *Genetics.* **121**:609-612

Slatkin M. 1991. Inbreeding coefficients and coalescence times. *Genet. Res.* **58**:167-175

Slatkin M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**:264–279.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics.* **139**:457-462

Sokal RR and Jacquez GM. 1991. Testing inferences about microevolutionary processes by means of spatial autocorrelation analysis. *Evolution.* **45**:152-168

Sokal RR and Thomson BA. 1998. Spatial genetic structure of human populations in Japan. *Hum. Biol.* **70**:1–22

Sokal RR and Wartenberg DE. 1983. A test of spatial autocorrelation using an isolation-by-distance model. *Genetics.* **105**:219-237

Sokal RR, Jacquez GM and Wooten MC. 1989. Spatial autocorrelation analysis of the migration and selection. *Genetics* **121**:845-855

Sokal RR, Oden NL and Thompson BA. 1999. A problem with synthetic maps. *Hum. Biol.* **71**:1-13

Sokal RR, Smouse PE and Neel JV. 1986. The genetic structure of the tribal population, the yanomama Indians. XV. Patterns inferred by autocorrelation analysis. *Genetics.* **114**:259-287

Sokal RR, Sneath PHA, 1963. *Principles of Numerical Taxonomy.* San Francisco: Freeman.

Sokal RR. 1983. Analyzing character variation in geographic space. Pp 384-403. In: numerical taxonomy, edited by J. Felsenstein. Springer-Verlag, New York

Sokal RR. 2004. Raymond Pearl's Legacy: The proper measure of Man. *American Journal of Human Biology.* **16**:113-124

Soper FL. 1967. Dinamics of Aedes aegypti Distribution and Density, Seasonal Fluctuations in the Americans. *WHO, Bull.* **36**:536-538.

Sriurairatna S, Bhamarapravati N. 1977. Replication of dengue-2 virus is *Aedes albopictus* mosquitoes. An electron microscopic study. *Am J Trop Med Hyg.* **26**:119.

SSA, 2001. Secretaria de Salud publica. Programa Nacional de Salud 2001-2006. Programa de accion: Enfermedades transmitidas por Vector. 29-43

SSAVER-SSA. Servicios de Salud Veracruz-Secretaria de Salud y Asistencia. 2004. http://www.ssaver.gob.mx/

statistics with special regard to systems of mating. *Evolution* **19**:395-420.

Student (Gosset WS). 1914. The elimination of spurious correlation due to position in time or space. *Biometrika.* **10**:179-181

Sumanochitrapon W, Strickman D, Sithiprasasna R, Kittayapong P, Innis BL. 1998. Effect of size and geographic origin of *Aedes aegypti* on oral infection with dengue-2 virus. *Am J Trop Med Hyg.* **58**:283-286.

Sumner J. et al. (2001) Neighbourhood size, dispersal and density estimates in the prickly forest skink (Gnypetoscincus queenslandiae) using individual genetic and demographic methods. *Mol. Ecol.* **10:**1917–1927

Tabachnick WJ and Powell JR. 1978. Genetic structure of the East African domestic population of *Aedes aegypti*. *Nature*. **272**:535-537.

Tabachnick WJ and Powell JR. 1979. A world-wide survey of genetic variation in the yellow fever mosquito, *Aedes aegypti*. *Genetical Research* **34(3)**:215-229.

Tabachnick WJ, Munstermann LE, Powell JR. 1979. Genetic distinctiveness of sympatric forms of *Aedes aegypti* in East Africa. *Evolution.* **33**:287-295.

Tabachnick WJ, Wallis GP, Aitken THG, Miller BR, Amato GD, Lorenz L, Powell JR, Beaty BJ. 1985. Oral infection of *Aedes aegypti* with Yellow Fever virus: geographic variation and genetic considerations. *Am J Trop Med Hyg*. **34**:1219-1224.

Tabor HK, Risch NJ, Myers RM. 2002. Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* **3**:391-397.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105:**437-460.

Takahashi M, Suzuki K. 1979. Japanese encephalitis virus in mosquito salivary glands. *Am J Trop Med Hyg*. **28**:122.

Takahashi M. 1980.Variation in susceptibility among colony strains *of Culex tritaeniorhynchus* to Japanese encephalitis virus infection. *Jpn J Med Sci Biol.* **33**:321-333.

Tardieux I, Poupel O, Lapchin L, and Rodhain F. 1990. Variation among strains of *Aedes aegypti* in susceptibility to oral infection with dengue virus type 2. *Am J Trop Med Hyg* **43(3)**:308-313.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**:4673-4680.

Thomson M, Shotkoski F, French-Constant RH. 1993. Cloning and sequencing of the cyclodiene insecticide resistance gene from the yellow fever mosquito *Aedes aegypti*. *FEBS Lett.* **325**:187-190.

Trpis M, Hausermann W, 1986. Dispersal and other population parameters of *Aedes aegypti* in an African village and their possible significance in epidemiology of vector-borne diseases. *Am J Trop Med Hyg* **35**:1263–1279.

Vazeille M, Mousson L, Rakatoarivony I, Villeret R, Rodhain F, Duchemin JB, Failloux AB. 2001. Population genetic structure and competence as a vector for dengue type 2 virus of *Aedes aegypti* and *Aedes albopictus* from Madagascar. *Am J Trop Med Hyg*. **65(5)**:491-497.

Vazeille-Falcoz M, Mousson L, Rodhain F, Chungue E, and Failloux A, 1999. Variation in oral susceptibility to dengue type 2 virus of populations of *Aedes aegypti* from the islands of Tahiti and Moorea, French Polynesia. *Am J Trop Med Hyg* **60(2)**:292-299.

von Allmen SD, Lopez-Correa RH, Woodall JP, Morens DM, Chiriboga J, Casta-Velez A. 1979. Epidemic dengue fever in Puerto Rico, 1977: a cost analysis. *Am J Trop Med Hyg*. **28**:1040-1044.

Wagner VE, Hill-Rowley R, Narlok SA and Newson HD. 1979. Remote Sensing: A Rapid and Accurate Method of Data Adquisition for a Newly Formed Mosquito Control District. Mosquito News. **39(2)**:2283-287.

Wallis GP, Tabachnick WJ, Powell JR. 1983. Macrogeographic genetic variation in a human commensal: *Aedes aegypti*, the yellow fever mosquito. *Genet Res.* **41**:241-258.

Wallis GP, Tabachnick WJ, Powell JR. 1984. Genetic heterogeneity among Caribbean populations of *Aedes aegypti. Am J Trop Med Hyg.* **33**:492-498.

Washino RK and Wood BL. 1994. Application of remote sensing to arthropod vector surveillance and control. *Am. J. of Trop. Med. and Hyg.* **50(6)**:134-144.

Waterman SH, Novak RJ, Sather GE, Bailey RE, Rios L, Gubler DJ. 1985. Dengue transmission in two Puerto Rican communities in 1982. *Am J Trop Med Hyg* **34**: 625–632.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7:**256-276.

Weir BS and Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38:**1358-1370.

Welch JB, Olson JK, Hart WG, Ingle SG, and Davis MR. 1989. Use of aerial color-infrared photography as a survey technique for Psorophora columbiae oviposition habitats in Texas ricelands. *J. Am. Mosq. Control Assoc.* **5**:147-160.

Whitfield SG, Murphy FA, Sudia WD. 1971. Eastern equine encephalomyelitis virus: an electron microscopic study of *Aedes triseriatus* (Say) salivary gland infection. *Virology.* **43**:110.

Whitfield SG, Murphy FA, Sudia WD. 1973. St Louis encephalitis virus: an ultrastructural study of infection in a mosquito vector. *Virology.* **56**:70.

WHO. 1998. Dengue and Dengue Haemorrhagic Fever. WHO Fact Sheet No 117, WHO, Geneva. Revised April 2002. http://www.who.int/inf-fs/en/fact117.html.

WHO. 2002. Fact Sheet no. 1212313

Winter A, Kramer W, Werner FA, Kollers S, Kata S, Durstewitz G, Buitkamp J, Womack JE, Thaller G, Fries R. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc Natl Acad Sci USA* **99:**9300-9305.

Wood BL, Beck LR, Lawless JG, and Vesecky JF. 1992a. Preliminary considerations for a small satellite to monitor environmental change associated with vector-borne. *J. Imaging Science and Tech.* **36(5)**:431-439.

Wood BL, Beck LR, Washino RK, Hibbard K, and Salute JS. 1992b. Estimating high mosquito-producing rice fields using spectral and spatial data. *Int. J. Remote Sens.* **13**:2813-2826.

Wood BL, Beck LR, Washino RK, Palchick S, and Sebesta P. 1991. Spectral and spatial characterization of rice field mosquito habitat. *Int. J. Remote Sens.* **12**:621-626.

Wood BL, Washino RK, Beck LR, Hibbard K, Pitcairn M, Roberts D, Rejmankova E, Paris J, Hacker C , Salute JS, Sebesta P, and Legters L. 1992c. Distinguishing high and

low anopheline-producing rice fields using remote sensing and GIS technologies. *Prev. Vet. Med.* **11**:277-288.

Wood RJ. 1976. Spine number on the pecten and comb of fourth instar larvae of *Aedes aegypti* L. differences between two populations. *Genetica.* **46**:33-44.

Woodring JL, Higgs S, Beaty BJ, 1996. Natural cycles of vector borne pathogens. In: Marquardt, W.C., Beaty, B.J. (Eds.), Biology of Disease Vectors, vol. University Press of Colorado, Boulder, Colorado, pp. 51–72

Wright S. 1943. Isolation by distance. *Genetics.* **28**:114-138

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19:**395-420.

Wright S. 1978. Evolution and the genetics of populations, Vol. 4, Variability within and among natural populations. University of Chicago Press, Chicago.

Xiong C, Levis R, Shen P, Schlesinger S Rice CM, Huang HV. 1989. Sindbis virus: an efficient broad host range vector for gene expression in animal cells. *Science.* **243**:118-119.

Xu G, Wilson W, Mecham J, Murphy K, Zhou EM, Tabachnick W. 1997. VP7: an attachment protein of bluetongue virus for cellular receptors in *Culicoides variipennis. J Gen Virol.* **78**:1617.

Yamada KI, Takasaki T, Nawa M, Nakayama M, Arai YT, Yabe S, Kurane I. 1999. The features of imported dengue fever cases from 1996 to 1999. *Jpn J Infect Dis.* **52**:257.

Yee J, Friedmann T, Burns JC. 1994. Generation of high-titer pseudotyped retroviral vectors with very broad host range. *Methods Cell Biol.* **43**:99-112.

Zink RM. 1994. The geography of mitochondrial DNA variation, population structure, hybridization, and species limits in the fox sparrow (*Passerrella iliaca*). *Evolution.* **48**:96-111

Table 1. List of priority diseases for WHO excluding non-vector borne diseases. DALYs* Disability Adjusted Life Years (the number of healthy years of life lost due to premature death and disability). Source: World Health Report, 2002.

	TDR disease category	Disease burden DALYs* (thousands)			Deaths (thousands)		
		Total	Male	Female	Total	Male	Female
African trypanosomiasis	1	1,598	1,029	568	50	32	18
Dengue	1	653	287	366	21	10	11
Leishmaniasis	1	2,357	1410	946	59	35	24
Malaria	2	42,280	20,024	22,256	1,124	532	592
Chagas disease	3	649	333	316	13	7	6
Lymphatic filariasis	3	5,644	4,317	1,327	0	0	0
Onchocerciasis	3	987	571	416	0	0	0

Table 2. The locations, sample sizes of *Aedes aegypti* collections in Mexico and the United States.

Regions **State**	**Cities**	**Locations in cities**	**Date**	**Number of Individuals Analyzed**	**MIB**	**MEB**	**SUSC**
Northeast region							
Texas	Houston		10 / 05 / 98	49	21 (43%)	17 (35%)	11 (22%)
Nuevo Leon	Monterrey		7 / 09 / 96	51	17 (33%)	3 (6%)	31 (61%)
Tamaulipas	Miguel Aleman		6 / 16 / 98	51	7 (14%)	7 (14%)	37 (73%)
	Nuevo Laredo		8 / 10 / 97	57	20 (35%)	7 (12%)	30 (53%)
TOTAL				208			
Central Atlantic region							
Veracruz	Moloacan		9 / 21 / 98	44	7 (16%)	5 (11%)	32 (73%)
Tabasco	Villahermosa		9 / 10 / 98	58	35 (60%)	10 (17%)	13 (22%)
Campeche	Campeche		8 / 10 / 98	56	17 (30%)	15 (27%)	24 (43%)
	Ciudad del Carmen		8 / 15 / 98	58	19 (33%)	14 (24%)	25 (43%)
TOTAL				216			
Yucatan region							
Yucatan	Merida		7 / 01 / 99	58	3 (5%)	3 (5%)	52 (90%)
		North	7 / 10 / 99	60	4 (7%)	3 (5%)	53 (88%)
		South	7 / 10 / 99	37	3 (8%)	3 (8%)	31 (84%)
		Central	7 / 10 / 99	59	12 (20%)	5 (8%)	42 (71%)
		East	7 / 10 / 99	60	3 (5%)	4 (7%)	53 (88%)
		West	7 / 10 / 99	60	11 (18%)	5 (8%)	44 (73%)
Quintana Roo	Cancun	Central	6 / 11 / 99	32	8 (25%)	0 (0%)	24 (75%)
		North	6 / 11 / 99	58	4 (7%)	1 (2%)	53 (91%)
Chetumal		Central	6 / 12 / 99	54	10 (19%)	0 (0%)	44 (81%)
		North	6 / 12 / 99	58	9 (16%)	1 (2%)	48 (83%)
TOTAL				536			

Table 2 (continued)

Pacific Coastal region						
Chiapas	Tapachula I	8 / 15 / 98	60	3 (5%)	3 (5%)	54 (90%)
	Tapachula II	8 / 16 / 98	37	7 (19%)	4 (11%)	26 (70%)
Oaxaca	Puerto Escondido	6 / 12 / 99	58	17 (29%)	5 (9%)	36 (62%)
Guerrero	Coyuca de Benitez	11 / 13 / 99	60	12 (20%)	8 (13%)	40 (67%)
	Ixtapa-Zihuatanejo	11 / 14 / 99	60	39 (65%)	11 (18%)	10 (17%)
Michoacan	Lazaro Cardenas	11 / 14 / 99	59	43 (73%)	7 (12%)	9 (15%)
Colima	Manzanillo	11 / 15 / 99	59	20 (34%)	9 (15%)	30 (51%)
Guadalajara	Puerto Vallarta	11 / 15 / 99	60	29 (48%)	12 (20%)	19 (32%)
Sinaloa	Mazatlan	11 / 16 / 99	59	21 (36%)	4 (7%)	34 (58%)
	Culiacan	7 / 27 / 98	17	5 (29%)	0 (0%)	12 (71%)
Sonora	Guaymas	7 / 18 / 00	38	6 (16%)	9 (24%)	23 (61%)
	Hermosillo	7 / 18 / 00	55	24 (44%)	6 (11%)	25 (45%)
Arizona	Tucson	9 / 10 / 98	60	17 (28%)	2 (3%)	41 (68%)
TOTAL			682			
Total	31 collections		1642			

Table 3: Comparison of genotype numbers in SSCP and sequence analyses in the early trypsin gene.

	Sequences	Presumed Genotypes SSCP	Actual Genotypes sequencing	Diff
5' fragment				
Deletion homozygotes	5	3	3	0
Insertion homozygotes	40	25	20	+5
Indel heterozygotes				
Delition	34	17	4	+13
Insertion	36	18	18	0
3' Fragment	52	25	26	-1
TOTAL	167	88	71	19

Table 3. Nucleotide or gap states at each of the 90 segregating sites in the early trypsin gene. Substitutions are listed from left to right according to relative frequencies, the first nucleotide being the most common. The frequency of the most common nucleotide is listed as is the expected heterozygosity at the segregating site.

Seg. site	Nucleo (Fig 1)	Freq.Most common nuc	Hexp	Hobs	Substitution	Codon	a.a.
1	28	0.860	0.226	0.215	T > C		
2	39	0.865, 0.133	0.220	0.208	C > A > T		
3	48	0.860	0.226	0.215	C > -		
4	49	0.860	0.226	0.215	T > -		
5	50	0.860	0.226	0.215	G > -		
6	51	0.860	0.226	0.215	A > -		
7	52	0.860	0.226	0.215	A > -		
8	53	0.860	0.226	0.215	C > -		
9	54	0.860	0.226	0.215	C > -		
10	55	0.860	0.226	0.215	A > -		
11	56	0.860	0.226	0.215	C > -		
12	57	0.860	0.226	0.215	A > -		
13	58	0.860	0.226	0.215	G > -		
14	59	0.860	0.226	0.215	C > -		
15	60	0.860	0.226	0.215	C > -		
16	70	0.866	0.218	0.206	T > C	TTT > CTT	F <=> L
17	72	0.862,0.133	0.229	0.214	T > C > G	TTT > TTC >TTG	F <=> L
18	88	0.999	0.002	0.002	T > A	TGT > ATG	C <=> M
19	118	0.753	0.273	0.330	G > T	g	
20	119	0.860	0.226	0.215	G > T	g	
21	127	0.862	0.228	0.212	G > T	g	
22	133	0.860	0.226	0.215	C > T	c	
23	134	0.998	0.002	0.003	G > T	g	
24	136	0.862	0.228	0.212	G > A	g	
25	137	0.860,0.138	0.228	0.215	G > T > A	g	
26	141	0.998	0.002	0.003	A > -	a	
27	142	0.998	0.002	0.003	C > -	c	
28	143	0.860,0.138	0.228	0.215	A > C > -	a	
29	144	0.860	0.226	0.215	A > G	a	
30	145	0.998	0.002	0.003	C > T	c	
31	147	0.860,0.138	0.228	0.215	- > T > A	-	
32	148	0.998	0.002	0.003	C > A	c	
33	150	0.998	0.002	0.003	A > T	a	

Table 3 (Continued)

Seg. site	Nuc. (Fig. 1)	Freq. most common nuc.	H_{exp}	H_{obs}	Substitution	codon	a.a.
34	151	0.998	0.002	0.003	G > T	g	
35	153	0.860	0.226	0.215	T > C	t	
36	154	0.993,0.005	0.013	0.014	G > A > C	g	
37	155	0.860	0.226	0.215	G > A	g	
38	156	0.860,0.138	0.228	0.215	T > G > -	t	
39	157	0.998	0.002	0.003	T > -	t	
40	160	0.747	0.067	0.314	A > C	a	
41	161	0.843,0.155	0.164	0.237	T > C > -	t	
42	162	0.998	0.002	0.003	T > -	t	
43	163	0.998	0.002	0.003	T > -	t	
44	164	0.998	0.002	0.003	C > -	c	
45	165	0.860,0.138	0.228	0.215	T > C > -	t	
46	166	0.860,0.138	0.228	0.215	A > G > -	a	
47	170	0.998	0.002	0.003	T > A	t	
48	171	0.848	0.200	0.225	C > G	c	
49	176	0.691,0.162	0.410	0.430	A > C > T	a	
50	214	0.996	0.008	0.008	G > A	GGC > GAC	G <=> D
51	240	0.723	0.069	0.312	G > A	GTC > ATC	V <=> I
52	242	0.666	0.394	0.409	C > T	GTC > GTT	V
53	284	1.000	0.000	0.000	C > A	TGC > TGA	C <=> >
54	305	0.893	0.165	0.163	G > A	CCG > CCA	P
55	338	0.860	0.226	0.215	G > T	ACG > ACT	T
56	344	0.654	0.400	0.417	A > G	ACA > ACG	T
57	354	0.862	0.226	0.210	G > A	GCC > ACC	A <=> T
58	355	0.854	0.226	0.220	C > T	GCC > GTC	A <=> V
59	366	0.942	0.056	0.095	C > A	CGG > AGG	R
60	374	0.862	0.228	0.212	T > G	GGT > GGG	G
61	398	0.862	0.228	0.212	C > A	GGC > GGA	G
62	402	0.851	0.226	0.222	A > G	ATA > GTA	I <=> V
63	434	0.992	0.002	0.014	G > A	CCG > CCA	P
64	454	0.995	0.011	0.011	A > T	TAC > TTC	Y <=> F
65	470	0.995	0.010	0.009	A > G	TCA > TCG	S
66	486	0.967	0.067	0.060	G > A	GAA > AAA	E <=> K
67	531	0.998	0.004	0.004	G > T	GAG > TAG	E <=> >
68	534	0.982	0.013	0.028	G > T	GCT > TCG	A <=> S
69	540	0.998	0.004	0.004	G > A	GAA > TAA	E <=> >

Table 3 (Continued)

Seg. site	Nuc. (Fig. 1)	Freq.most common nuc.	H_{exp}	H_{obs}	Substitution	codon	a.a.
70	558	0.999	0.003	0.003	G > T	GCT > TCT	A <=> S
71	560	0.737	0.107	0.324	T > G	GCT > GCG	A
72	567	0.948	0.000	0.046	A > G	ACC > GCC	T <=> A
73	593	0.960	0.013	0.066	G > C	AAG > AAC	K <=> N
74	598	0.522	0.069	0.382	T > C	GTT > GCT	V <=> A
75	602	0.990	0.004	0.017	T > G	TTT > TTG	F <=> L
76	608	0.978	0.020	0.038	G > T	ATG > ATT	M <=> I
77	615	0.954	0.000	0.073	T > C	TTG > CTG	L
78	621	0.998	0.004	0.004	C > A	CGC > AGC	R <=> S
79	640	0.992	0.004	0.014	C > A	TCG > TAG	S <=> >
80	651	0.997	0.000	0.005	G > T	GCG > TCG	A <=> S
81	653	0.988	0.000	0.019	G > A	GCG > GCA	A
82	659	0.939	0.122	0.097	T > C	TGT > TGC	C
83	679	0.969	0.063	0.056	T > C	GTT > GCT	V <=> A
84	695	0.969	0.063	0.056	G > T	GAG > GAT	E <=> D
85	715	0.992	0.000	0.015	A > T	TAT > TTT	Y <=> F
86	721	0.941	0.117	0.092	C > A	GCT > GAT	A <=> D
87	737	0.677,0.320	0.0840	0.316	T > C > A	TCT > TCC > TCA	S
88	739	0.966	0.068	0.061	G > A	TGC > TAC	C <=> Y
89	769	0.992	0.016	0.014	C > A	TCT > TAT	S <=> Y
90	779	0.959	0.082	0.066	A > C	AAA > AAC	K <=> N

Table 4. Locations, dates of collections, coordinates, and sample sizes of *Aedes aegypti* collections in Mexico.

State	City/Region	Date(s) m/y	Latitude	Longitude	n
Nuevo Leon	Monterrey North	7/96	N25°40'00.12"	W100°18'00.00"	57
	South	7/96	N25°28'00.12"	W100°10'01.20"	58
	West	7/96	N25°30'00.00"	W100°04'58.80"	58
	East	7/96	N25°40'59.88"	W100°22'01.20"	58
Tamaulipas	Ciudad Victoria	8/96	N23°40'00.12"	W099°15'00.00"	59
	Miguel Aleman	6/98	N26°23'30.00"	W099°03'39.00"	50
	Matamoros	7/96	N26°15'00.00"	W097°28'00.12"	59
	Nuevo Laredo	8/97	N27°30'00.00"	W099°28'00.12"	48
	Reynosa	7/97	N26°10'00.12"	W098°10'00.12"	59
	Tampico	8/96	N23°40'00.12"	W097°49'59.88"	59
Veracruz	Acayucan	08/03, 08/04	N17°57'43.07"	W094°24'45.17"	138
	Alvarado	08/03, 08/04	N18°46'27.19"	W095°45'48.80"	116
	Coatzacoalcos	08/03, 08/04	N18°08'26.91"	W094°24'47.15"	120
	Cololeacaque	08/03	N17°57'43.07"	W094°32'09.79"	63
	Martinez*	08/03, 08/04	N20°02'59.97"	W097°02'19.77"	102
	Minatitlan	9/96, 08/03, 08/04	N17°58'47.00"	W094°32'27.00"	161
	Moloacan	9/98	N17°59'09.00"	W094°20'46.00"	55
	Panuco	08/03, 08/04	N22°03'12.47"	W098°11'11.78"	141
	Poza Rica	08/03, 08/04	N20°32'37.18"	W097°28'14.83"	105
	Tantoyuca	08/03, 08/04	N21°20'30.33"	W098°13'39.88"	118
	Tuxpan	8/96	N21°10'00.12"	W097°25'00.12"	59
	Zempoala	08/03, 08/04	N19°26'41.60"	W096°24'23.31"	105
Tabasco	Villahermosa	9/98	N17°59'59.99"	W092°54'00.00"	58
Campeche	Campeche	8/98	N19°53'59.99"	W090°36'00.01"	53
	Cd. del Carmen	8/98	N18°35'59.99"	W091°47'59.99"	52
Yucatan	Merida	7/99	N20°57'00.01"	W089°38'23.99"	57
	North	7/99	N21°00'44.64"	W089°37'51.60"	49
	South	7/99	N20°57'06.84"	W089°38'26.88"	35
	Central	7/99	N20°57'58.68"	W089°39'57.24"	46
	East	7/99	N20°59'28.32"	W089°35'00.60"	53
	West	7/99	N20°58'39.00"	W089°39'28.80"	60
Quintana Roo	Cancun Central	6/99	N21°08'24.01"	W086°52'47.99"	32
	North	6/99	N21°09'03.61"	W086°52'38.97"	53
	Chetumal Central	6/99	N18°29'59.99"	W088°18'00.00"	38
	North	6/99	N18°30'29.31"	W088°17'49.97"	54
Total					2488

Martinez* = Martinez de la Torre

Table 5. Vector competence to dengue 2 (Jam 1409) of *Aedes aegypti* from the 2003 Veracruz collection.

Region	Collection	n	[1]H+	[2]M+	MIR	DIR	MIB%	MEB%
North	Panuco	60	23	39	0.6500	0.5897	0.3500	0.0603
	Tantoyuca	77	45	60	0.7792	0.7500	0.2208	0.0292
	Martinez*	72	37	61	0.8472	0.6066	0.1528	0.2407
	Poza Rica	47	35	41	0.8723	0.8537	0.1277	0.0187
Total		*256*	*140*	*201*	*0.7852*	*0.6965*	*0.2148*	*0.0886*
South	Alvarado	56	6	20	0.3571	0.3000	0.6429	0.0571
	Acayucacan	73	8	35	0.4795	0.2286	0.5205	0.2509
	Minatitlan	60	12	29	0.4833	0.4138	0.5167	0.0695
	Cosoleacaque	63	8	33	0.5238	0.2424	0.4762	0.2814
	Coatzasacoalcos	71	22	46	0.6479	0.4783	0.3521	0.1696
	Zempoala	75	25	55	0.7333	0.4545	0.2667	0.2788
Total		*398*	*81*	*218*	*0.5477*	*0.3716*	*0.4523*	*0.1762*

Martinez*=Martinez de la Torre
1 Head Positive
2 Midgut Positive

Table 6. Partitioning of variation in the frequency of genetic markers among *Aedes aegypti* collections.

Source of variation	d.f.	S.S	Var. comp.	Var (%)
Among north-south groups	1	103.8	0.0790	16.3
Within north-south groups	44	249.87	0.0990	20.4
Within cities inside north-south groups	2435	756.33	0.3083	63.4
Within the state of Veracruz collection				
Among north-south groups	1	505.09	1.0167	24.1
within north-south groups	17	449.33	0.4217	10.0
within cities in the north-south groups	1055	2928.83	2.7761	65.9
2003 vs. 2004 collections				
Among years	1	7.14	-0.0959	-2.7
Within year	17	947.28	0.9406	26.0
Within collections of 2003 and 2004	1055	2928.83	2.7761	76.7

Table 7. Regression of $F_{ST}/(1-F_{ST})$ from ND4 mitochondrial markers on geographic distances. The *p* refers to Mantel probability (10,000 iterations).

Collection	Slope $F_{ST}/(1-F_{ST})$ =	Intercept	R^2	*p*	$4D\pi\sigma^2$ probability
All collections	0.0001 x geo. distance	0.3398	0.2901	0.0001	
	0.0892 ln(geo. distance)	0.1123	0.1113	0.0001	11
North	0.0003 geo. distance	0.2127	0.0388	0.0899	
	0.0901 ln(geo. distance)	-0.1638	0.0409	0.0648	11
South	0.0002 geo. distance	0.2464	0.0939	0.0016	
	0.0508 ln(geo. distance)	0.0451	0.1611	0.0001	20
Within Veracruz	0.0011 geo. distance	0.1270	0.4305	0.0023	
	0.1925 ln(geo. distance)	-0.5866	0.3891	0.0009	5
North	0.0003 geo. distance	0.0855	0.0374	0.3291	
	0.0371 ln(geo. distance)	-0.0453	0.0232	0.2900	27
South	0.0020 geo. distance	0.0657	0.4187	0.0590	
	0.1525 ln(geo. distance)	-0.3623	0.3061	0.09	7

Table 8. Fixation indexes between and among *Aedes aegypti* gulf collections. North-south collection groups were separated by the their relative location to the Neovolcanic Axis.

Collection	*F*	Index
Among north-south groups	*F*(among)	0.2432
Within north-south groups	*F*(within)	0.3662
Within cities inside north-south groups	*F*(all collections)	0.1626
Within the state of Veracruz collection		
Martinez* in the north group		
Among north-south groups	*F*(among)	0.1323
Within north-south groups	*F*(within)	0.3248
Within cities in the north-south groups	*F*(all collections)	0.2219
Martinez* in the south group		
Among north-south groups	*F*(among)	0.1319
within north-south groups	*F*(within)	0.3413
within cities in the north-south groups	*F*(all collections)	0.2412
2003 vs. 2004 collections		
Among years	*F*(among)	0.2531
Within year	*F*(within)	0.2330
Within collections of 2003 and 2004	*F*(all collections)	0.0265

Martinez* = Martinez de la Torres
1023 permutations

Table 9. List of the of 26 ND4 haplotypes frequencies found in *Aedes aegypti* mosquitoes collections made throughout Mexico.

id	Longitude	Latitude	n	1	2	3	4	5	6	7	8	9	10
acayu	-94.62970734	18.00087929	133	0.068	0.030	0	0	0	0	0	0	0	0
alvar	-95.76355556	18.77421944	113	0.027	0.027	0.027	0	0.009	0	0	0	0	0
campe	-90.52552795	19.85453987	53	0	0.151	0.283	0	0	0	0	0	0	0
cancn	-86.87749167	21.15100278	53	0	0	0.189	0	0	0	0	0	0	0
cancu	-86.87999722	21.14000278	32	0	0	0.031	0	0	0	0	0	0	0
cdcar	-91.79999722	18.59999722	52	0.019	0.096	0.058	0	0	0	0	0	0	0.115
cdvic	-99.13921356	23.73859978	59	0.475	0.407	0.051	0	0.051	0.017	0	0	0	0
chetn	-88.29721389	18.50814167	54	0	0.037	0.333	0	0	0	0	0	0	0
chetu	-88.30000000	18.49999722	38	0	0.053	0	0	0	0	0	0	0	0
coatz	-94.41309722	18.14080833	117	0.034	0	0.026	0	0	0	0	0	0	0
cosol	-94.53605278	17.96196389	36	0	0	0.048	0	0	0	0	0	0	0
coyuc	-100.08679962	17.00638008	50	0	0.200	0	0	0	0	0	0	0	0.020
culia	-107.38549805	24.80484962	36	0	0	0	0.083	0	0	0	0	0	0
houst	-95.38672638	29.76869965	50	0	0	0.660	0	0	0	0	0.180	0.160	0
ixtap	-101.56680298	17.64336967	59	0	0	0.068	0	0	0	0	0	0	0.085
juchi	-98.63184357	16.61890030	61	0	1.000	0	0	0	0	0	0	0	0
lcard	-102.19249725	17.92494011	42	0	0.190	0.048	0	0	0	0	0	0	0.095
manza	-104.03935988	18.95876052	49	0	0.061	0.286	0	0	0	0	0	0	0
marti	-97.03882500	20.04999167	99	0.071	0.141	0.040	0	0.394	0.303	0	0	0	0
mata	-97.50218964	25.86116982	59	0.220	0.458	0.136	0.169	0	0.017	0	0	0	0
mazat	-106.41609955	23.20383072	49	0	0.082	0.306	0	0	0	0	0	0	0
merda	-89.63999722	20.95000278	57	0	0.018	0.053	0	0	0	0	0	0	0.088
merdc	-89.66590000	20.96630000	46	0	0.022	0.022	0	0	0	0	0	0	0.044
merde	-89.58350000	20.99120000	53	0	0.019	0.019	0	0	0	0	0	0	0.038
merdn	-89.63100000	21.01240000	49	0	0	0.020	0	0	0	0	0	0	0
merds	-89.64080000	20.95190000	35	0	0	0.057	0	0	0	0	0	0	0.086
merdw	-89.65800000	20.97750000	60	0	0.017	0.083	0	0	0	0	0	0	0.100

Table 9. continue

id	Longitude	Latitude	n	1	2	3	4	5	6	7	8	9	10
migu	-99.06083333	26.39166667	50	0	0	0.400	0	0	0	0	0.060	0.180	0
minat	-94.54083333	17.97972222	155	0	0.026	0.032	0	0.026	0	0	0	0	0
moloa	-94.34611111	17.98583333	55	0.018	0.236	0	0	0	0	0	0	0	0
mtye	-100.36700000	25.68330000	58	0.793	0.121	0.086	0	0	0	0	0	0	0
mtyn	-100.30000000	25.66670000	57	0.456	0.246	0.246	0	0	0.053	0	0	0	0
mtys	-100.16700000	25.46670000	58	0.534	0.224	0.241	0	0	0	0	0	0	0
mtyw	-100.08300000	25.50000000	58	0.603	0.293	0.103	0	0	0	0	0	0	0
nvola	-99.46670000	27.50000000	48	0.021	0	0	0.896	0	0	0.083	0	0	0
oaxc	-96.71365356	17.06192017	60	0	0.950	0.050	0	0	0	0	0	0	0
oaxn	-96.77132288	17.12943372	60	0	0.583	0.350	0	0	0	0	0	0	0
oaxs	-96.70136474	16.99573593	59	0	0.559	0.220	0	0	0	0	0	0	0
oaxse	-96.78375988	16.97863505	60	0	0.367	0.633	0	0	0	0	0	0	0
panuc	-98.18660556	22.05346389	132	0.402	0.288	0.106	0.076	0.053	0.076	0	0	0	0
pochu	-96.46259308	15.73468971	59	0	1.000	0	0	0	0	0	0	0	0
pozri	-97.47078611	20.54366111	98	0.337	0.286	0.235	0.031	0.020	0.092	0	0	0	0
ptoes	-97.06862640	15.85956001	117	0	0.427	0.162	0	0	0	0	0	0	0
ptova	-105.23930359	20.61449051	52	0	0.058	0.077	0	0	0	0	0	0	0.115
reyno	-98.16670000	26.16670000	59	0.559	0.051	0.356	0	0	0.034	0	0	0	0
tampi	-97.84262848	22.24323082	59	0.271	0.576	0.102	0	0.051	0	0	0	0	0
tanto	-98.22774444	21.34175833	116	0.664	0.276	0.060	0	0	0	0	0	0	0
tapaI	-92.26068115	14.90674973	57	0	0.053	0.158	0	0	0	0	0	0	0.193
tapan	-94.21723175	16.37220955	70	0	0.900	0.100	0	0	0	0	0	0	0
tapII	-92.24115900	14.91367800	42	0	0	0.190	0	0	0	0	0	0	0.119
tucso	-110.89170074	32.19581985	59	0	0	0	0.254	0	0	0	0	0	0
tuxpa	-97.40339661	20.95841026	59	0.458	0.136	0.288	0	0.085	0.034	0	0	0	0
vilah	-92.90000000	17.99999722	58	0	0.259	0	0	0	0	0	0	0	0
zanat	-94.35711670	16.47145081	60	0	0.600	0.400	0	0	0	0	0	0	0
zempo	-96.40647500	19.44488889	104	0.135	0	0	0	0.308	0.279	0	0	0	0

Table 9. continue

id	n	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
acayu	133	0	0	0	0.406	0	0.414	0	0	0.083	0	0	0	0	0	0	0
alvar	113	0	0	0	0.566	0	0.310	0	0	0.035	0	0	0	0	0	0	0
campe	53	0.057	0.151	0	0.113	0	0.151	0	0	0	0	0	0.094	0	0	0	0
cancn	53	0	0	0	0.321	0	0	0	0	0.019	0	0	0	0	0.472	0	0
cancu	32	0.312	0.406	0	0.219	0	0	0	0	0	0	0	0	0	0.031	0	0
cdcar	52	0.058	0.019	0	0.058	0.039	0	0	0	0	0	0	0.077	0	0.212	0.250	0
cdvic	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chetn	54	0	0.019	0	0	0.222	0	0	0	0.019	0	0	0	0	0.370	0	0
chetu	38	0.368	0.053	0	0.026	0.158	0	0	0	0	0	0	0	0.053	0.289	0	0
coatz	117	0	0	0	0.573	0	0.145	0	0	0.222	0	0	0	0	0	0	0
cosol	36	0	0	0	0.825	0	0.127	0	0	0	0	0	0	0	0	0	0
coyuc	50	0.040	0.160	0	0	0	0	0	0	0.420	0	0	0	0	0.140	0.020	0
culia	36	0	0	0.194	0.028	0	0	0.056	0	0.639	0	0	0	0	0	0	0
houst	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ixtap	59	0.441	0	0	0.102	0	0	0	0	0.288	0	0	0	0	0.017	0	0
juchi	61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lcard	42	0.286	0	0.024	0	0	0	0	0	0.071	0	0	0	0	0.071	0.214	0
manza	49	0	0.020	0.306	0	0	0	0	0	0.286	0	0	0	0	0.041	0	0
marti	99	0	0	0	0	0	0	0	0	0.051	0	0	0	0	0	0	0
mata	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mazat	49	0.020	0.061	0.102	0	0	0	0	0	0.327	0	0	0	0	0.102	0	0
merda	57	0.193	0.018	0	0.175	0	0	0	0	0	0	0	0	0	0.175	0.281	0
merdc	46	0.500	0	0	0.217	0	0	0	0	0	0	0	0	0	0	0.196	0
merde	53	0.660	0	0	0.264	0	0	0	0	0	0	0	0	0	0	0	0
merdn	49	0.286	0	0	0.571	0	0	0	0	0	0	0	0	0	0	0.122	0
merds	35	0.486	0	0	0.314	0	0	0	0	0	0	0.029	0	0	0	0.029	0
merdw	60	0.400	0	0	0.367	0	0	0	0	0	0	0	0	0	0	0.033	0
migu	50	0	0	0	0	0	0	0	0	0	0	0.040	0	0.320	0	0	0
minat	155	0	0	0.006	0.374	0	0.458	0	0.019	0.058	0	0	0	0	0	0	0

Table 9. continue

id	n	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
moloa	55	0	0	0	0	0	0.436	0	0.273	0.018	0	0	0	0.018	0	0	0
mtye	58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mtyn	57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mtys	58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mtyw	58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
nvola	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
oaxc	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
oaxn	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.067
oaxs	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.220
oaxse	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
panuc	132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pochu	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pozri	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ptoes	117	0.017	0.188	0	0	0	0	0	0	0.145	0	0	0	0	0.060	0	0
ptova	52	0.077	0	0.096	0	0	0	0	0	0.346	0	0	0	0	0.231	0	0
reyno	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tampi	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tanto	116	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tapaI	57	0.053	0.140	0	0.018	0	0	0	0	0.246	0	0	0	0	0.140	0	0
tapan	70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tapII	42	0.262	0.119	0	0.048	0	0	0	0	0.119	0	0	0	0	0.143	0	0
tucso	59	0	0	0.034	0	0	0	0.271	0	0	0.441	0	0	0	0	0	0
tuxpa	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
vilah	58	0	0.345	0	0	0.035	0	0	0	0.345	0	0	0	0	0.017	0	0
zanat	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
zempo	104	0	0	0	0.221	0	0.058	0	0	0	0	0	0	0	0	0	0

Geographical coordinates NAD 1927. The complete names for the *id* can be found at the end of table 13.

Table 10. Spatial autocorrelation coefficients *I of* 10 selected ND4 haplotype frequencies. All *I*-correlograms are significant (Bonferroni, $p < 0.05$, expected value -0.02)

UL	194	290	418	530	607	709	800	884	979	1045	1183	1410	2139
pairs	98	98	98	98	98	98	99	98	98	98	98	98	98
hap/bin	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.33	0.77**	0.21*	0.31**	0.09	0.20*	-0.20	-0.24*	-0.32**	-0.77**	-0.53**	-0.39**	-0.13
2	0.69	0.30**	-0.08	-0.08	-0.03	0.020	0.15	-0.44**	-0.36**	-0.01	-0.34**	-0.14	0.20*
5	0.07	0.04	-0.11	-0.09	-0.06	-0.10	-0.25**	-0.07	0.00	0.00	0.00	0.03	0.07
10	1.77*	0.05	0.12	-0.06	-0.39**	-0.16	0.14	-0.24*	-0.18	0.06	0.16*	-0.05	-0.01
11	1.81*	0.38**	0.33**	0.04	-0.15	-0.39**	-0.03	-0.53**	-0.38**	0.02	-0.16	0.07	-0.23*
13	0.07	0.26**	0.07	0.08	0.19**	-0.01	0.05	-0.15	0.01	-0.11	-0.04	-0.12	-0.43**
14	0.73	0.03	-0.37**	-0.04	0.08	0.57**	-0.01	-0.24*	-0.15	-0.11	-0.2	-0.26**	-0.23*
16	0.16	-0.15	-0.22*	0.11	-0.17	-0.27**	-0.04	-0.11	-0.01	-0.12	-0.01	0.03	0.04
19	0.28	0.28**	0.12	-0.01	-0.04	-0.03	-0.12	0.02	-0.06	-0.01	0.06	-0.19	-0.31**
24	-0.01	0.08	0.23**	0.04	0.15*	-0.18	-0.05	-0.05	-0.17	-0.01	-0.29**	-0.14	-0.1
Avg.	0.59	0.20	0.03	0.03	-0.03	-0.04	-0.03	-0.20	-0.16	-0.11	-0.14	-0.11	-0.11

Avg = Average
Hap = Haplotype
UL= Uper Limit distance in kilometers
* $p < 0.05$
** $p < 0.01$

Table 11. Correlogram for linearized F_{ST} (F_{ST}/1- F_{ST}) for all 26 Nd4 haplotype frequencies. Each interval was tested using Mantel permutation test.

Linear geographic distance (km)				*ln* geographic distance (km)				
UL	# Pairs	Mantel **r**	**p**	UL	# Pairs	Mantel **r**	**p**	
194	105	0.104	0.000	5.3	103	0.101	0.001	***(**)
290	102	0.017	0.675	5.7	102	0.017	0.675	
418	99	0.018	0.681	6.0	105	0.019	0.691	
530	106	0.014	0.646	6.3	97	0.010	0.607	
607	101	0.067	0.038	6.4	105	0.064	0.045	*
709	104	0.014	0.662	6.6	101	0.025	0.775	
800	105	0.058	0.946	6.7	103	0.047	0.908	
884	105	-0.006	0.433	6.8	106	0.001	0.510	
979	103	-0.021	0.256	6.9	110	-0.024	0.221	
1045	106	-0.012	0.379	7.0	101	-0.008	0.411	
1183	135	-0.100	0.002	7.1	149	-0.101	0.002	**
1410	144	-0.137	0.000	7.3	130	-0.139	0.000	***
2728	170	0.016	0.602	8.0	171	0.016	0.599	

UL = upper limit in kilometers
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$
(**) p for LN transformed geographic distance in first distance class. The probabilities for the remaining distance classes are similar.

Table 12. Eigenvalues for the principal components analysis of 26 ND4 haplotype frequencies. Listed is the variance accounted for by each of the principal components, the eigenvectors calculated from a correlation matrix. The two higher Eigenvalues are denoted as principal component 1 and principal component 2

Eigenvalues	Variance Proportion	Cumulative
3.196	0.123	0.123
2.843	0.109	0.232
2.372	0.091	0.323
2.106	0.081	0.404
1.882	0.072	0.477
1.838	0.071	0.548
1.675	0.064	0.612
1.482	0.057	0.669
1.209	0.047	0.716
1.150	0.044	0.760
1.082	0.042	0.801
1.077	0.041	0.843
0.915	0.035	0.878
0.730	0.028	0.906
0.592	0.023	0.929
0.542	0.021	0.950
0.405	0.016	0.965
0.283	0.011	0.976
0.210	0.008	0.984
0.185	0.007	0.991
0.152	0.006	0.997
0.046	0.002	0.999
0.019	0.001	1

Table 13. The first and second principal component coordinates of each population, and the biplot coordinates for each genetic locus showed graphically in figure 33. The list is sorted by the variation (Var.) in each collection. The table also presents the spatial autocorrelation (SA) from each haplotype (X significant, O non-significant spatial autocorrelation). Output corresponds with Figure 33. The Group (G) each collection sites corresponds (□) refer to North Atlantic collection, (○) refers to South Atlantic collection, (◇) refers to Pacific and Oaxaca (black filled) collection. The abbreviations of the collection sites are described in at the end of table 9.

Col.	PC1	PC2	Var.	G	Hap.	PC1	PC2	Var.	Ocurrace	SA
migu	9.826	2.725	10.197	□	9	4.997	0.458	5.018	3.6%	X
houst	6.863	0.851	6.916	□	23	4.001	0.523	4.035	5.5%	O
cdcar	-1.495	3.277	3.602	○	8	4.022	0.292	4.033	3.6%	X
marti	-0.269	-3.517	3.527	□	21	3.476	0.661	3.538	3.6%	O
merda	-1.398	2.794	3.124	○	3	3.518	-0.004	3.518	80.0%	O
zempo	-0.484	-2.919	2.959	○	10	-1.385	1.357	1.939	21.8%	X
merds	0.563	2.651	2.710	○	11	-1.276	1.280	1.807	32.7%	X
tapaI	-1.023	2.365	2.577	◇	25	-1.256	1.038	1.629	14.5%	O
lcard	-1.198	2.198	2.503	◇	24	-0.951	1.090	1.447	29.1%	X
merdc	-1.238	2.165	2.494	○	14	-1.312	0.554	1.424	38.2%	X
nvola	-0.707	-2.234	2.343	□	19	-1.207	0.544	1.324	34.5%	X
chetu	-0.438	2.250	2.292	○	1	0.413	-1.055	1.133	36.4%	X
tucso	-0.994	-1.998	2.232	◇	12	-0.864	0.551	1.025	23.6%	O
tapII	-0.763	2.084	2.219	◇	2	0.224	-0.980	1.005	74.5%	X
ptova	-1.077	1.938	2.217	◇	6	-0.093	-0.993	0.997	16.4%	O
ixtap	-1.019	1.806	2.074	◇	5	-0.173	-0.944	0.960	16.4%	X
merdw	-0.975	1.823	2.067	○	16	-0.712	-0.157	0.729	14.5%	X
merde	-0.990	1.618	1.897	○	4	-0.410	-0.539	0.677	10.9%	O
panuc	0.114	-1.853	1.857	□	22	-0.456	0.457	0.646	3.6%	X
chetn	0.053	1.738	1.739	○	15	-0.224	0.568	0.611	7.3%	O
pozri	0.423	-1.659	1.712	□	17	-0.497	-0.330	0.597	3.6%	X
tuxpa	0.562	-1.612	1.707	□	20	-0.413	-0.348	0.540	1.8%	X
cancu	-1.000	1.366	1.693	○	13	-0.459	0.182	0.494	12.7%	X
cdvic	0.046	-1.688	1.689	□	7	-0.294	-0.389	0.488	1.8%	X
merdn	-1.027	1.288	1.647	○	26	0.343	-0.311	0.463	3.6%	O
oaxs	0.627	-1.518	1.642	◆	18	-0.216	-0.195	0.291	3.6%	O
mtye	0.185	-1.566	1.577	□						
mtyn	0.492	-1.481	1.561	□						
tanto	0.111	-1.541	1.545	□						
reyno	0.756	-1.340	1.539	□						
oaxse	1.323	-0.716	1.504	◆						
tampi	0.141	-1.491	1.498	□						
mtyw	0.201	-1.474	1.488	□						
cancn	-0.455	1.358	1.432	○						
coyuc	-1.006	0.994	1.414	◇						
mtys	0.503	-1.311	1.404	□						

Table 13 continue.

Collection	PC1	PC2	Var.	G
mata	0.155	-1.388	1.397	□
oaxn	0.757	-1.115	1.348	◆
juchi	-0.102	-1.337	1.341	◇
pochu	-0.102	-1.337	1.341	◇
oaxc	0.010	-1.288	1.288	◆
tapan	0.123	-1.239	1.245	◇
zanat	0.798	-0.945	1.237	◇
moloa	-0.466	-1.101	1.196	○
culia	-1.075	0.336	1.126	◇
vilah	-0.898	0.626	1.095	○
coatz	-0.827	0.255	0.865	○
acayu	-0.813	-0.217	0.841	○
minat	-0.760	-0.312	0.822	○
cosol	-0.747	0.266	0.793	○
alvar	-0.761	-0.075	0.765	○
campe	-0.179	0.673	0.696	○
mazat	-0.042	0.497	0.499	◇
manza	-0.120	0.368	0.387	◇
ptoes	-0.184	-0.040	0.188	◇

Acayucan=acayu, Alvarado=alvar, Campeche=campe, Cancun Central=cancn, Cancun North=cancu, Chetumal Central=chetu, Chetumal North=chetn, Ciudad del Carmen=cdcar, Ciudad Victoria=cdvic, Coatzacoalcos=coatz, Cosoleacaque=cosol, Coyuca de Benites=coyuc, Culiacan=culia, Guaymas=guaym, Hermosillo=hermo, Houston=houst, Ixtapa- Zihuatanejo=ixtap, Juchitan=juchi, Lazaro Cardenas=lcard, Manzanillo=manza, Martinez de la Torre=marti, Matamoros=mata, Mazatlan=mazat, Merida=merda, Merida Central=merdc, Merida East=merde, Merida North=merdn, Merida South=merds, Merida Weast=merdw, Miguel leman=migue, Minatitlan=minat, Moloacan=moloa, Monterrey East=mtye, Monterrey North=mtyn, Monterrey South=mtys, Monterrey West=mtyw, Nuevo Laredo=nvola, Oaxaca Central=oaxc, Oaxaca North=oaxn, Oaxaca South=oaxs, Oaxaca South-East=oaxse, Panuco=panuc, Pochutla=pochu, Poza Rica=pozri, Puerto Escondido=ptoes, Puerto Vallarta=ptova, Reynosa=reyno, Tampico=tampi, Tantoyuca=tanto, Tapachula I=tapaI, Tapachula II=tapan, Tapanatepec=tapII, Tucson=tucso, Tuxpam=tuxpa, Villa Hermosa=villah, Zanatepec=zanat, Zempoala=zempo.

Table 14. List of vector competence components for mosquitoes collected throughout Mexico.

Collection Site	Longitude	Latitude	MIB	MEB	SUC
Acayucan	-94.62970734	18.00087929	0.521	0.251	0.123
Alvarado	-95.76001740	18.77404976	0.643	0.057	0.143
Campeche	-90.52552795	19.85453987	0.367	0.316	0.433
Ciudad del Carmen	-91.98669283	18.61944986	0.417	0.257	0.433
Coatzacoalcos	-94.41271210	18.14044952	0.352	0.170	0.296
Cosoleacaque	-94.90099335	17.94359016	0.476	0.281	0.127
Coyuca de Benitez	-100.08679962	17.00638008	0.160	0.130	0.733
Culiacan	-107.38549805	24.80484962	0.172	0.048	0.788
Guaymas	-110.89209747	27.92354965	0.154	0.273	0.615
Hermosillo	-110.96219635	29.09417915	0.436	0.194	0.455
Houston	-95.38672638	29.76869965	0.400	0.350	0.390
Ixtapa	-101.56680298	17.64336967	0.588	0.429	0.235
Lazaro Cardenas	-102.19249725	17.92494011	0.470	0.140	0.457
Manzanillo	-104.03935988	18.95876052	0.290	0.200	0.564
Martinez de la Torre	-97.04261780	20.05847931	0.153	0.240	0.514
Mazatlan	-106.41609955	23.20383072	0.286	0.100	0.643
Merida	-89.62757874	20.98339081	0.215	0.137	0.678
Miguel Aleman	-99.01666670	26.38333330	0.311	0.133	0.597
Minatitlan	-94.53204346	17.96666908	0.520	0.070	0.200
Moloacan	-94.34361112	17.98583330	0.252	0.239	0.569
Monterrey	-100.31710052	25.67734909	0.314	0.169	0.570
Nuevo Laredo	-99.51055145	27.48100090	0.430	0.160	0.480
Panuco	-98.18283081	22.05258942	0.350	0.060	0.601
Poza Rica	-97.47090912	20.54340935	0.128	0.019	0.766
Puerto Vallarta	-105.23930359	20.61449051	0.500	0.400	0.300
Tantoyuca	-98.22939301	21.34853935	0.221	0.029	0.558
Tapachula	-92.26068115	14.90674973	0.222	0.071	0.722
Tapachula	-92.24115900	14.91367800	0.450	0.091	0.500
Tucson	-110.89170074	32.19581985	0.303	0.073	0.646
Villahermosa	-92.93678284	17.97628975	0.411	0.175	0.486
Zempola	-96.40296936	19.44329071	0.267	0.279	0.329

Longitude and Latitude in geographical coordinates (NAD 1927). MIB =Midgut infection Barrier, MEB=Midgut Escape Barrier, SUC=Suceptibility

Table 15. Spatial autocorrelation coefficients *I* and *c* for MIB', MEB' and SUC'. The rates where $arcsin(x)^{1/2}$ transformed. Each data class is tested independently. Each data class was 82 pairs. Expected values *I*=-0.03 Expected value for *c* = 1.0.

UL	279	445	598	728	886	1057	1219	1471	1868	2723	
I											
MIB'	0.34***	0.17	-0.30**	-0.22	0.33***	-0.12	0.04	-0.37***	-0.13	0.02	X
MEB'	0.06	0.20*	-0.05	-0.2	-0.09	-0.19	0.10	-0.17	-0.03	0.13	
SUC'	0.52***	0.40***	-0.13	-0.44***	0.08	-0.34**	-0.01	-0.26*	-0.14	0.07	X
c											
MIB'	0.78	0.80	1.39*	1.32	0.75	1.13	0.69	1.50**	1.00	0.64	
MEB'	0.91	0.86	0.96	1.07	0.93	1.09	0.92	1.33	1.12	0.81	
SUC'	0.65	0.84	1.07	1.61***	0.83	1.35*	0.77	1.26	0.91	0.7	X

I=Moran's *I* correalogram. ***c***= Geary's *c* correlograms. UL = upper limit in km. X significant correlogram ($p < 0.05$).

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

Table 16. Vector competence components pair wise correlations. All components were $arcsin(x)^{1/2}$ transformed.

					+	++
		cov	*p*(cov)	***r***	*p*(***r***)	*p*(***r***)
MIB'	MEB'	1.6562	0.0977	0.3386	0.0304	0.0984
MIB'	SUC'	-3.1678	0.0015	-0.8928	0.0000	0.0000
MEB'	SUC'	-2.6717	0.0075	-0.6276	0.0000	0.0039

+ conventional probability calculation. ++ CRH corrected probability calculation (Clifford, Richardson, and Hémon, 1989). cov= covariance. ***r*** Pearson's product-moment correlation between two variables.

Table 17. Pairwise correlations of dengue cases, altitude, mosquito susceptibility and inhabitant density.

		cov	*P*(cov)	**r**	p(**r**)+	p(**r**)++	Effective *n*	+/-
logDF	logALT	-1.4013	0.1611	-0.2797	0.0023	0.1655	26.0939	-
logDF	SUC'	-0.567	0.5707	-0.0738	0.4289	0.5752	59.9699	
logDF	logDEN	0.3347	0.7378	0.043	0.6451	0.7409	61.5359	
logALT	SUC'	0.7973	0.4253	0.2247	0.0149	0.4482	13.5949	-
logALT	logDEN	1.4813	0.1385	0.2927	0.0014	0.1416	26.6169	-
SUC'	logDEN	0.5277	0.5977	0.0951	0.3079	0.6059	31.8108	

logDEN = $\log_{10}$(total inhabitant in the municipalty/square degrees), logALT= $\log_{10}$(altitude), SUC'= *arcsin*(susceptibility)$^{1/2}$, logDF=$\log_{10}$(dengue fever cases * 100,000), logDHF=$\log_{10}$(dengue hemorrhagic fever cases * 100,000). + conventional probability calculation. ++ CRH corrected probability cov= covariance. ***r*** Pearson's correlation between two variables.

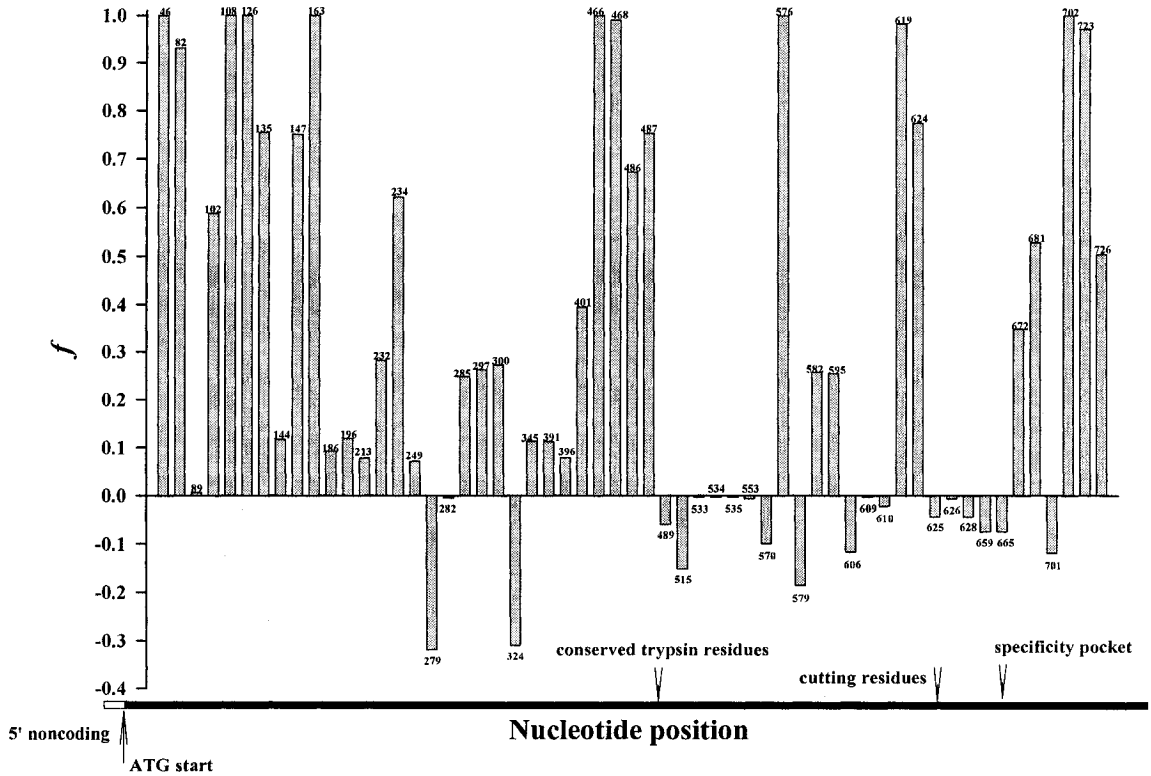


Figure 1: Weir and Cockerham's (1984) f at each segregating site among all 31 geographic collections of *Aedes aegypti*. A positive f indicates an excess of homozygotes and a negative f indicates an excess of heterozygotes.

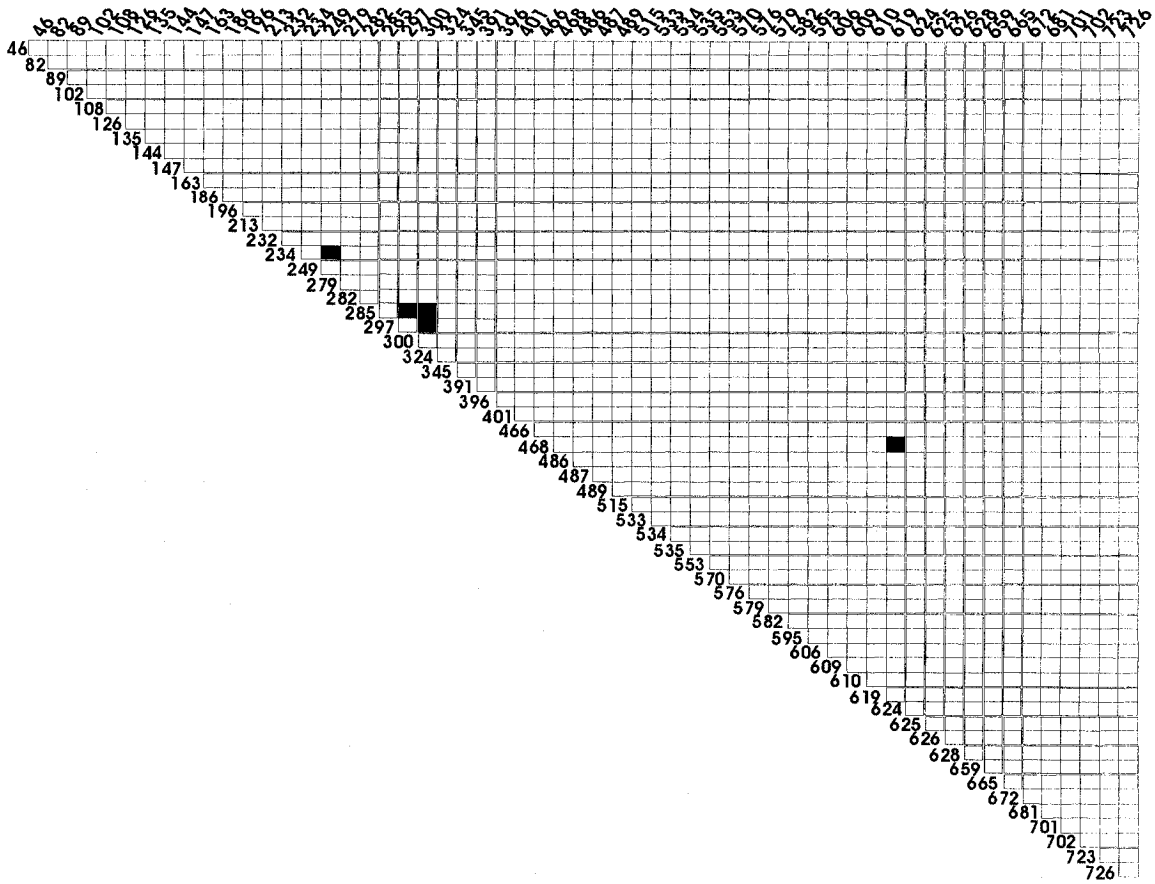


Figure 2: Analysis of linkage disequilibrium among all pairs of segregating sites among Pacific collections. A black square corresponds to a pattern of systematic disequilibrium consistent with epistasis. A clear square indicates disequilibrium due to genetic drift.

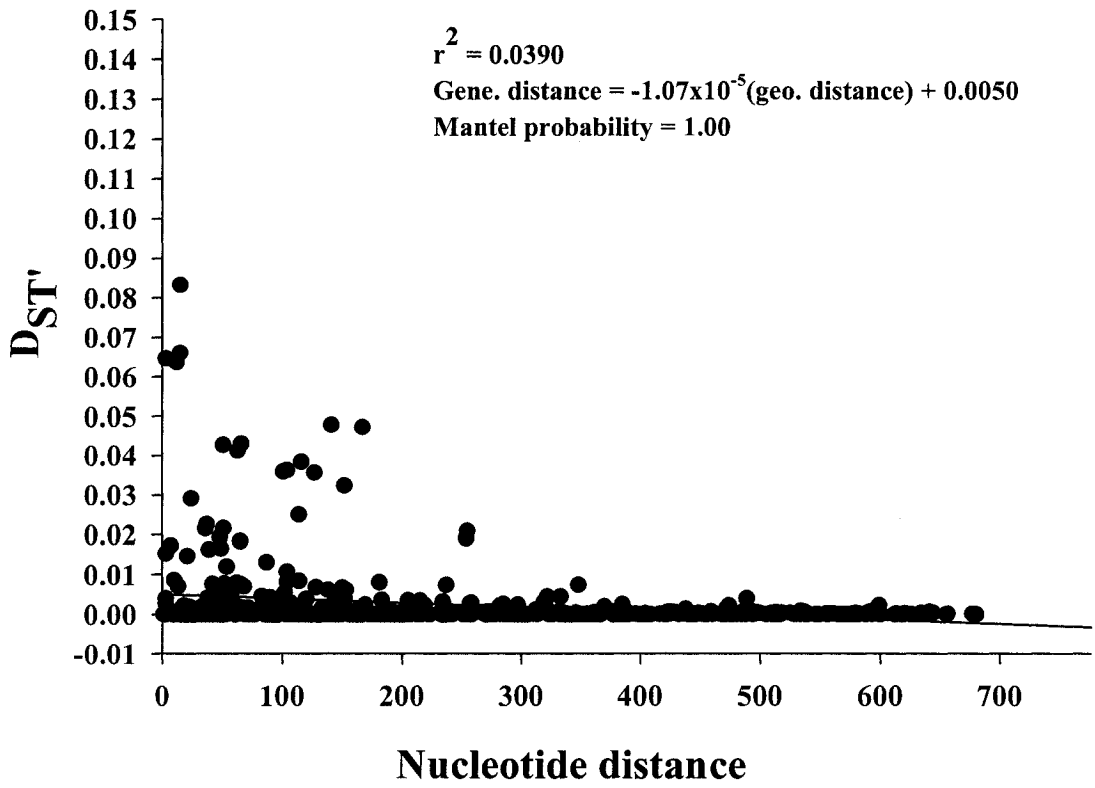


Figure 3: Regression analysis of $D_{ST}^{'2}$ on the physical distance among segregating sites. The significance of this regression analysis was evaluated with a permutation test (Mantel, 1967). The regression model is listed for each site along with the Mantel probability.

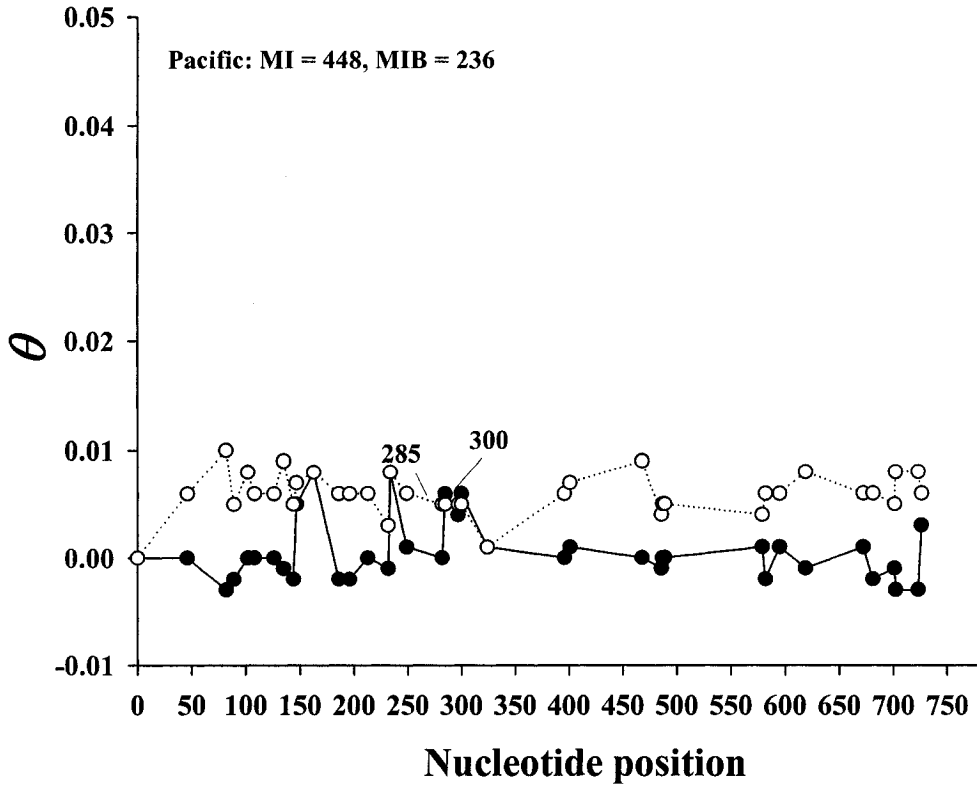


Figure 4: Association mapping of MIB among collections along the Pacific based upon nucleotide frequencies. A putative QTN, in the abundant trypsin gene, is identified at a segregating site when the original θ (solid line and circles) exceeded the 95% threshold (dotted line, empty circle). For each geographic region, putative QTNs are indicated with the number of their segregating site.

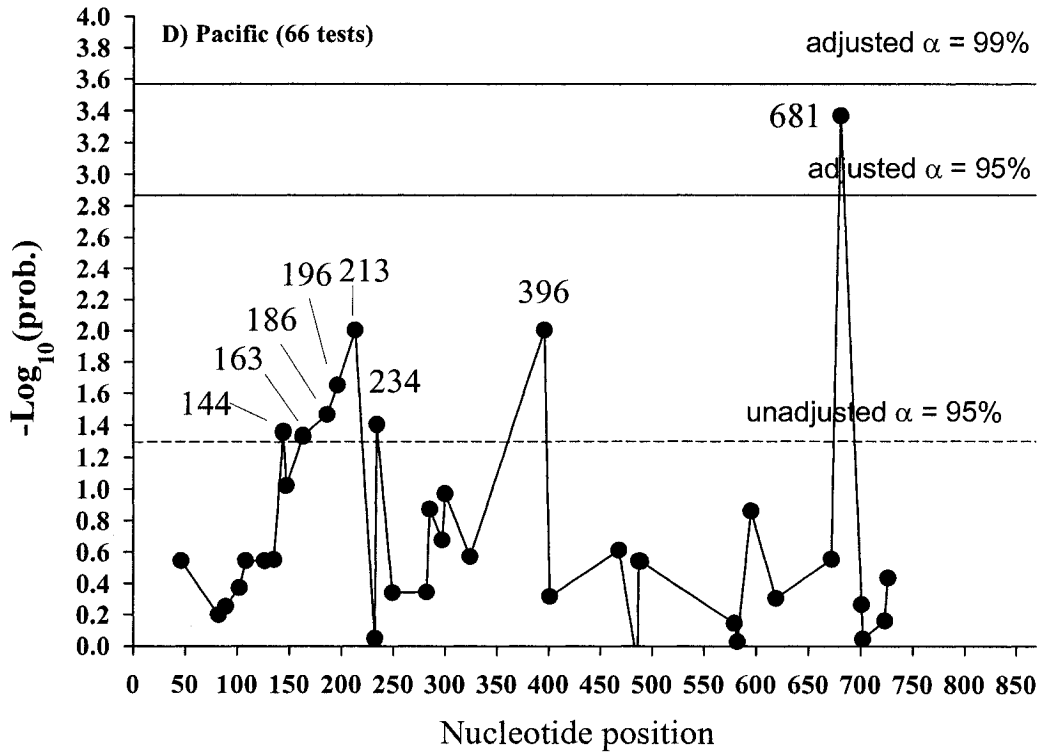


Figure 5: Association mapping of MIB based upon abundant trypsin genotype frequencies. Contingency $\chi 2$ tests of association between alternate genotypes and the number of mosquitoes with and without an MI were performed. The $-\log_{10}(p)$ score (p = 1- the cumulative probability associated with the test $\chi 2$ statistic) was then plotted against nucleotide position at each segregating site.

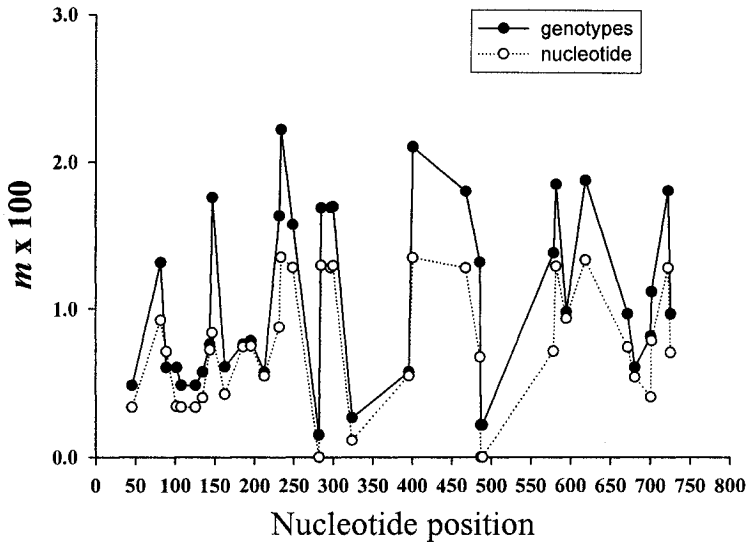


Figure 6: Minimal allelic and genotypic effects (m) that a difference in genotype (solid line and circles) or nucleic acid substitution (dotted line, empty circle) could exert on a phenotype and still be statistically significant.

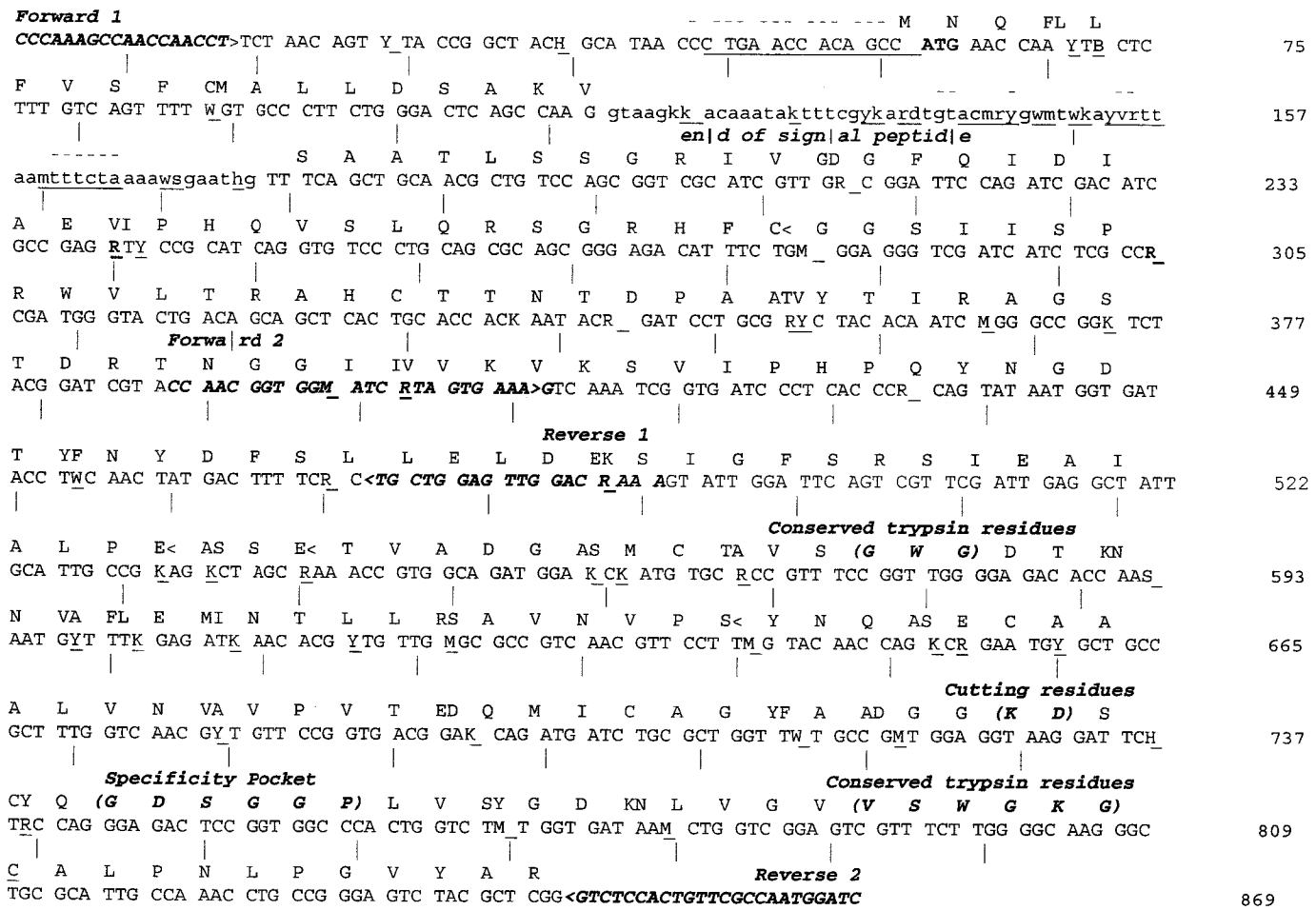


Figure 7. Map of segregating sites in *Early Trypsin*. A '|' appears once every 10 nucleotides. Segregating sites are underlined. Amino acid(s) appear above each codon. The gene was amplified in two overlapping pieces using the four primers shown in bold italics. Segregating sites discussed in the text (240, 305, and 344) are bolded and underlined. Landmarks (e.g. cutting residues) are from Noriega et al (1996).

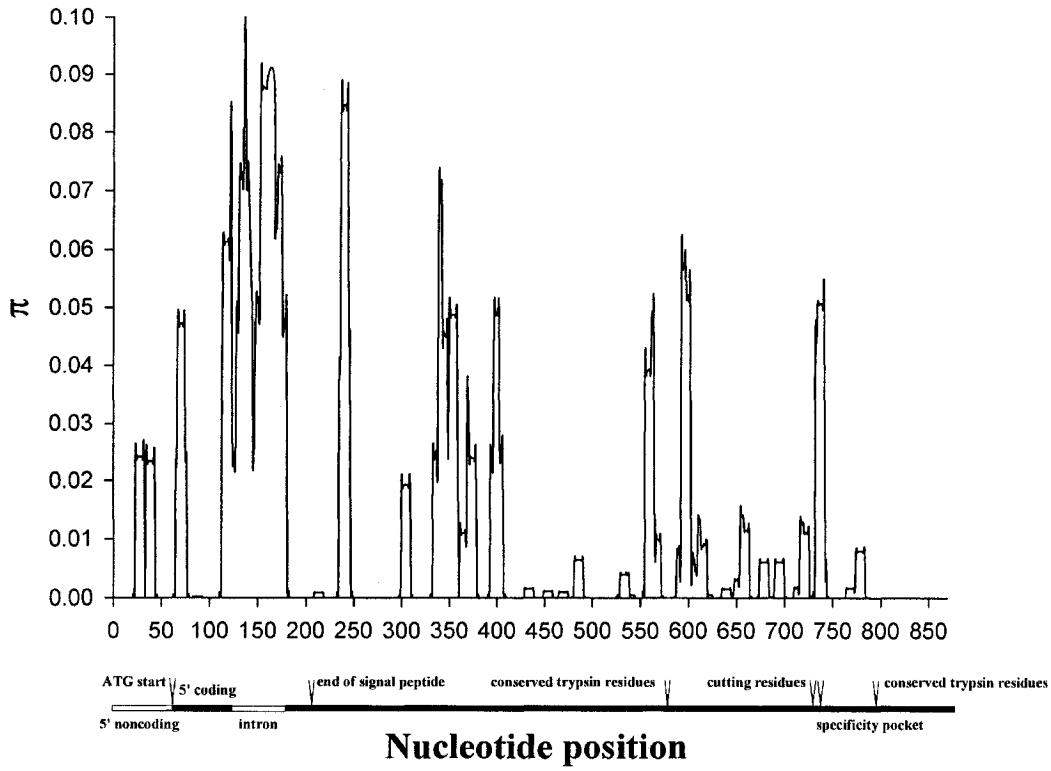


Figure 8. Plot of nucleotide diversity (π) across the sequenced regions of *Early Trypsin*. Landmarks at the base of the figure are from Figure 1.

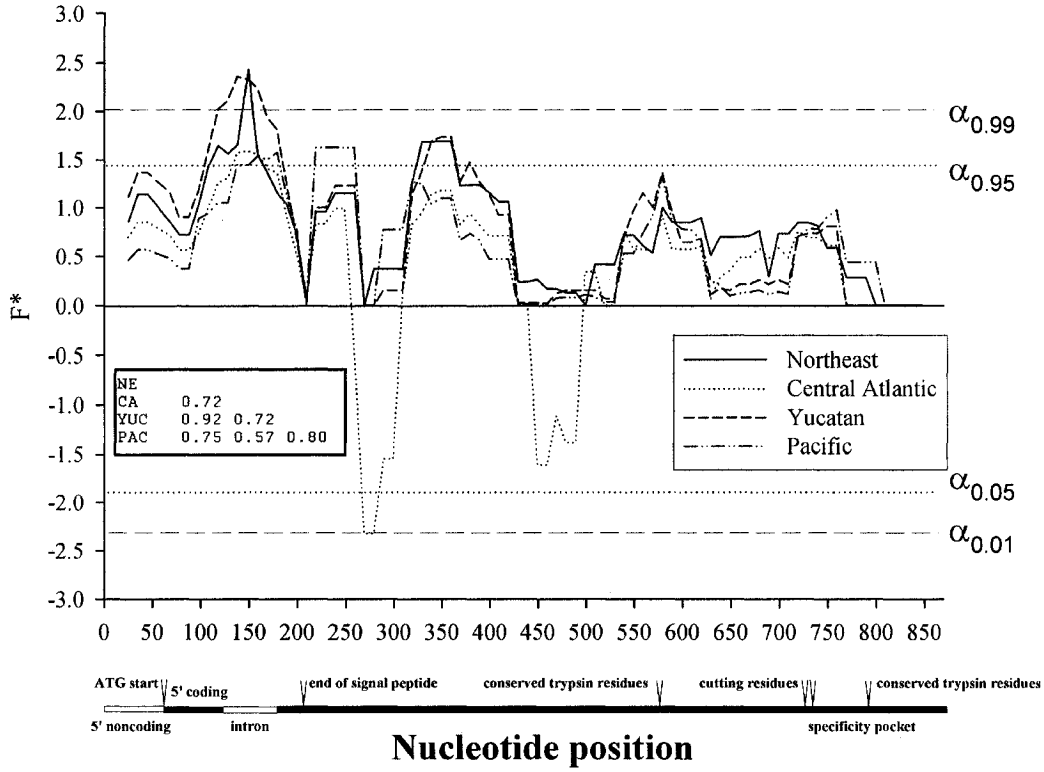


Figure 9. Plot of Fu and Li's F* across *Early Trypsin*. The α threshold values are from Table 4 for n=100 in FU and LI (1993). F* is a normalized comparison of the number of all mutations (η) to the number of singletons (η_s). This analysis assumes that F*>0 ($\eta > \eta_s$) under balancing selection, F* $\simeq$ 0 ($\eta \simeq \eta_s$) with neutral substitutions and F*< 0 ($\eta < \eta_s$) under purifying selection. Pearson correlation coefficients of F* among populations appear in the box. Landmarks at the base of the figure are from Figure 1.

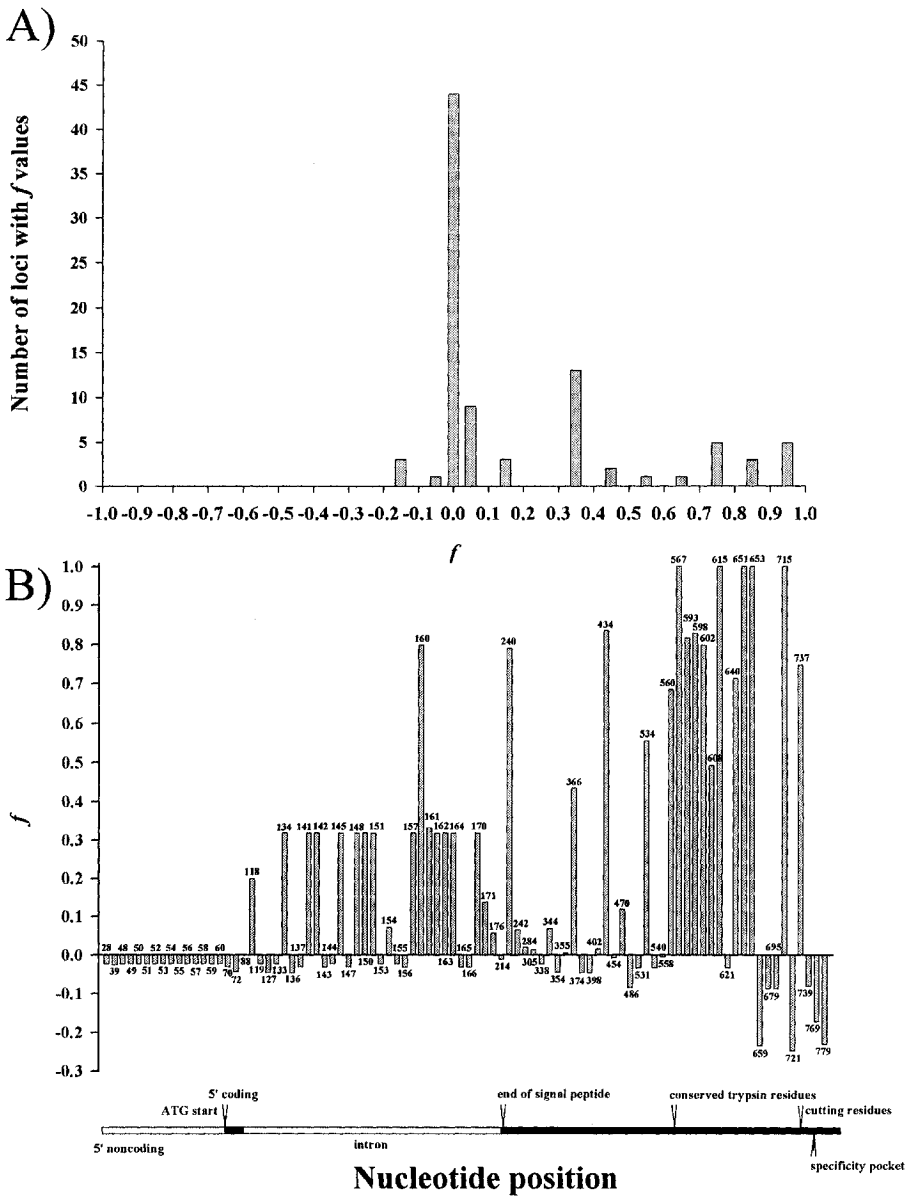


Figure 10. Weir and Cockerham's (1984) f at each segregating site among all 31 geographic collections. A positive f indicates an excess of homozygotes and a negative f indicates an excess of heterozygotes. A) A frequency histogram of f among all 90 segregating sites. B) f at each segregating site. Landmarks at the base of the figure are from Figure 1 but note that distances among sites are not on a linear scale.

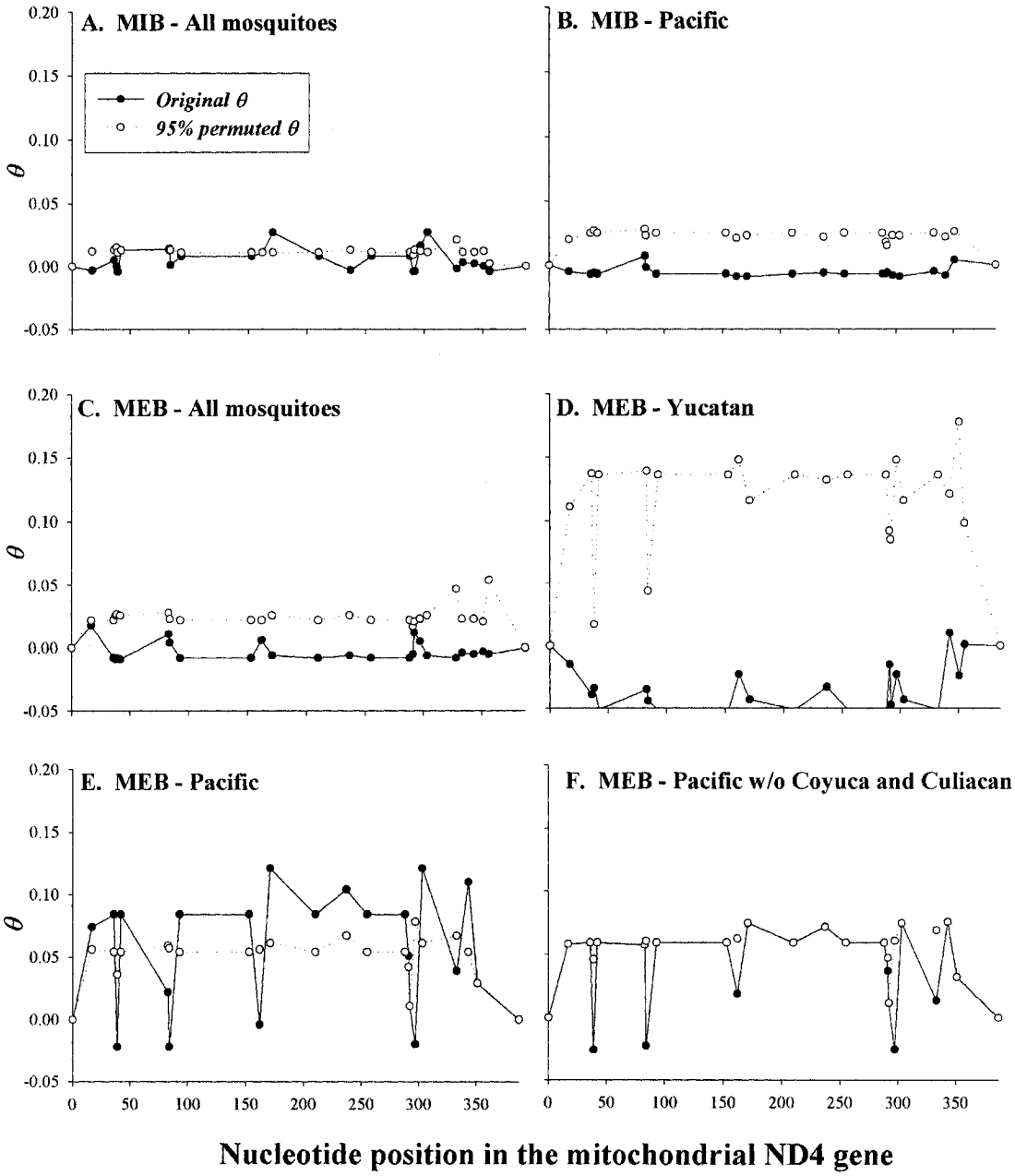


Figure 11. Association mapping of MIB and MEB with polymorphisms in the ND4 gene of the mtDNA. For MIB (Figs. A and B), nucleotide frequencies were compared between SUSC + MEB and MIB mosquitoes. For MEB (Figs. C-F), nucleotide frequencies were compared between SUSC and MEB mosquitoes.

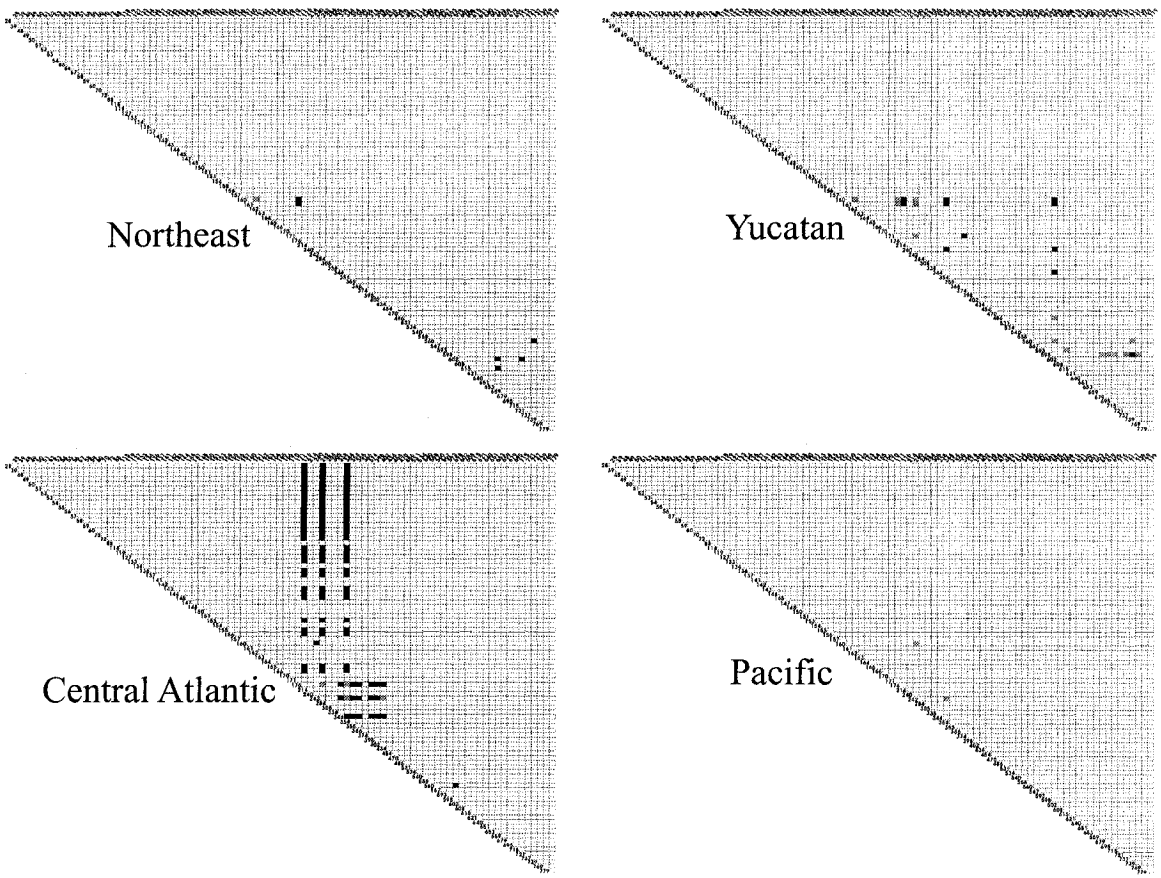


Figure 12. Analysis of linkage disequilibrium among all pairs of segregating sites at the early trypsin gene. A black square corresponds to a pattern of systematic disequilibrium consistent with epistasis ($D^2_{IS} > D^2_{ST}$, and $D'^2_{IS} < D'^2_{ST}$). A gray square corresponds to a pattern of nonsystematic disequilibrium consistent with epistasis in only some collections ($D^2_{IS} > D^2_{ST}$, and $D'^2_{IS} > D'^2_{ST}$). A clear square indicates disequilibrium due to genetic drift ($D^2_{IS} < D^2_{ST}$, and $D'^2_{IS} > D'^2_{ST}$).

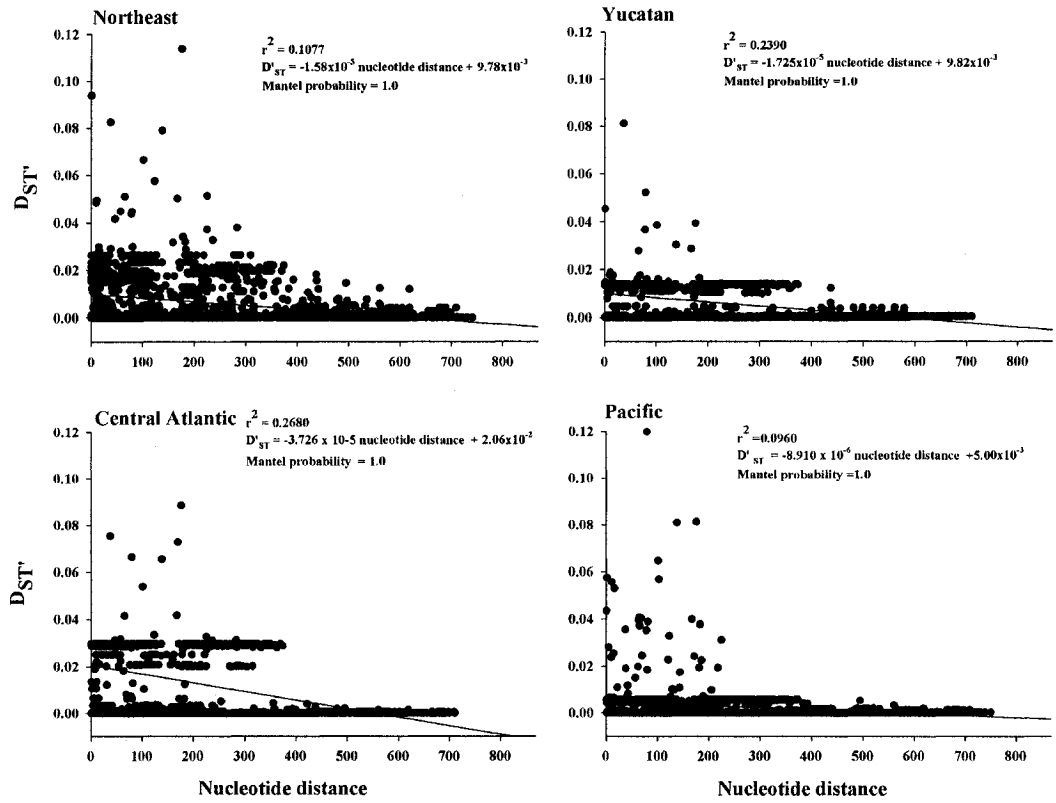


Figure 13. Regression analysis of D'^2_{ST} on the physical distance among segregating sites. The significance of this regression analysis was evaluated with a permutation test (Mantel, 1967). The regression model is listed for each site along with the Mantel probability.

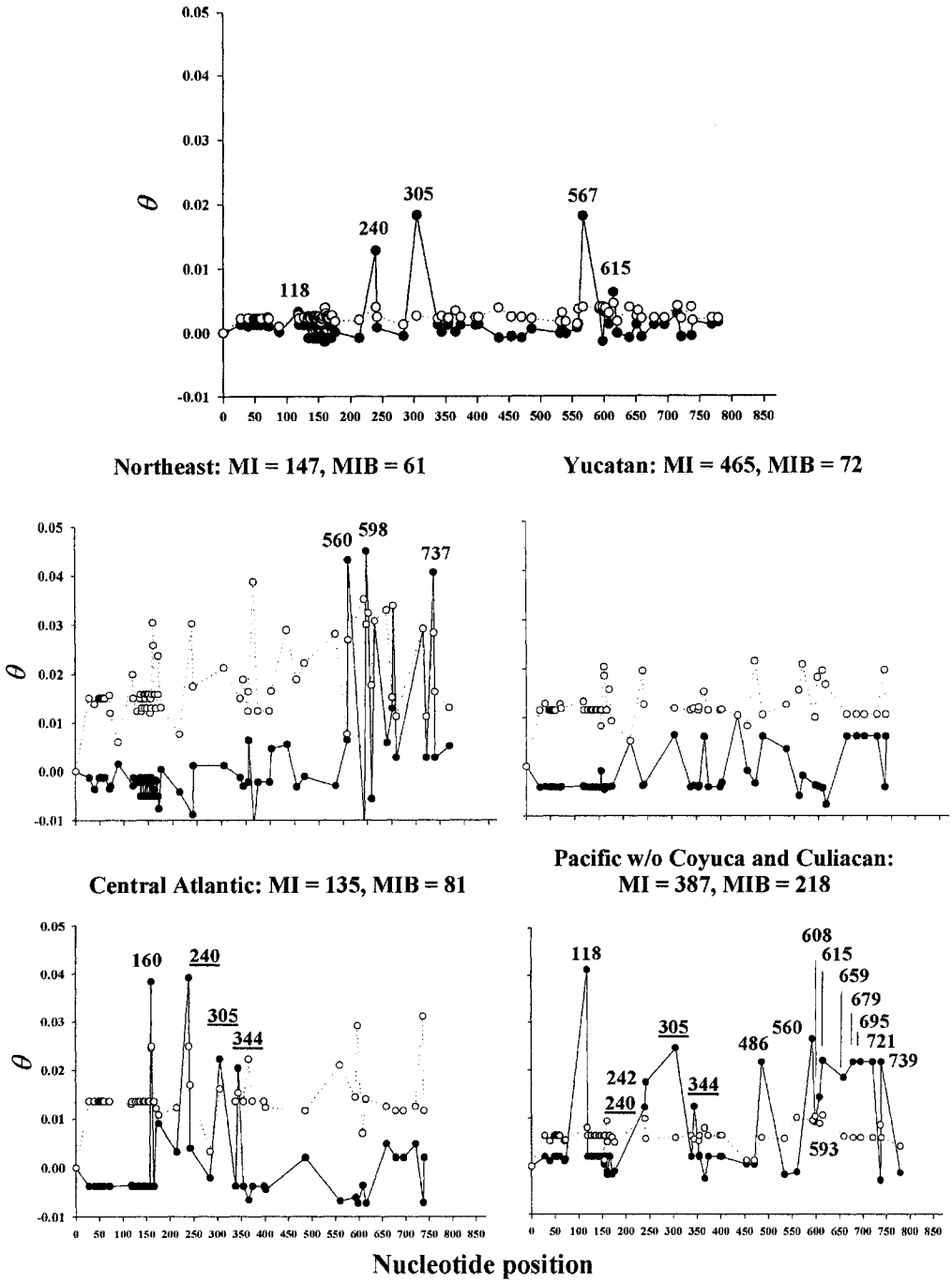


Figure 14. Association mapping of *MIB* based upon early trypsin nucleotide frequencies. At each segregating site, nucleotide frequencies in *SUSC* + *MEB* mosquitoes were compared to nucleotide frequencies in *MIB* mosquitoes. A putative QTN is identified at a segregating site when the original θ (solid line and circles) exceeded the 95% threshold (dotted line, empty circle). For each geographic region, putative QTNs are indicated with the number of their segregating site.

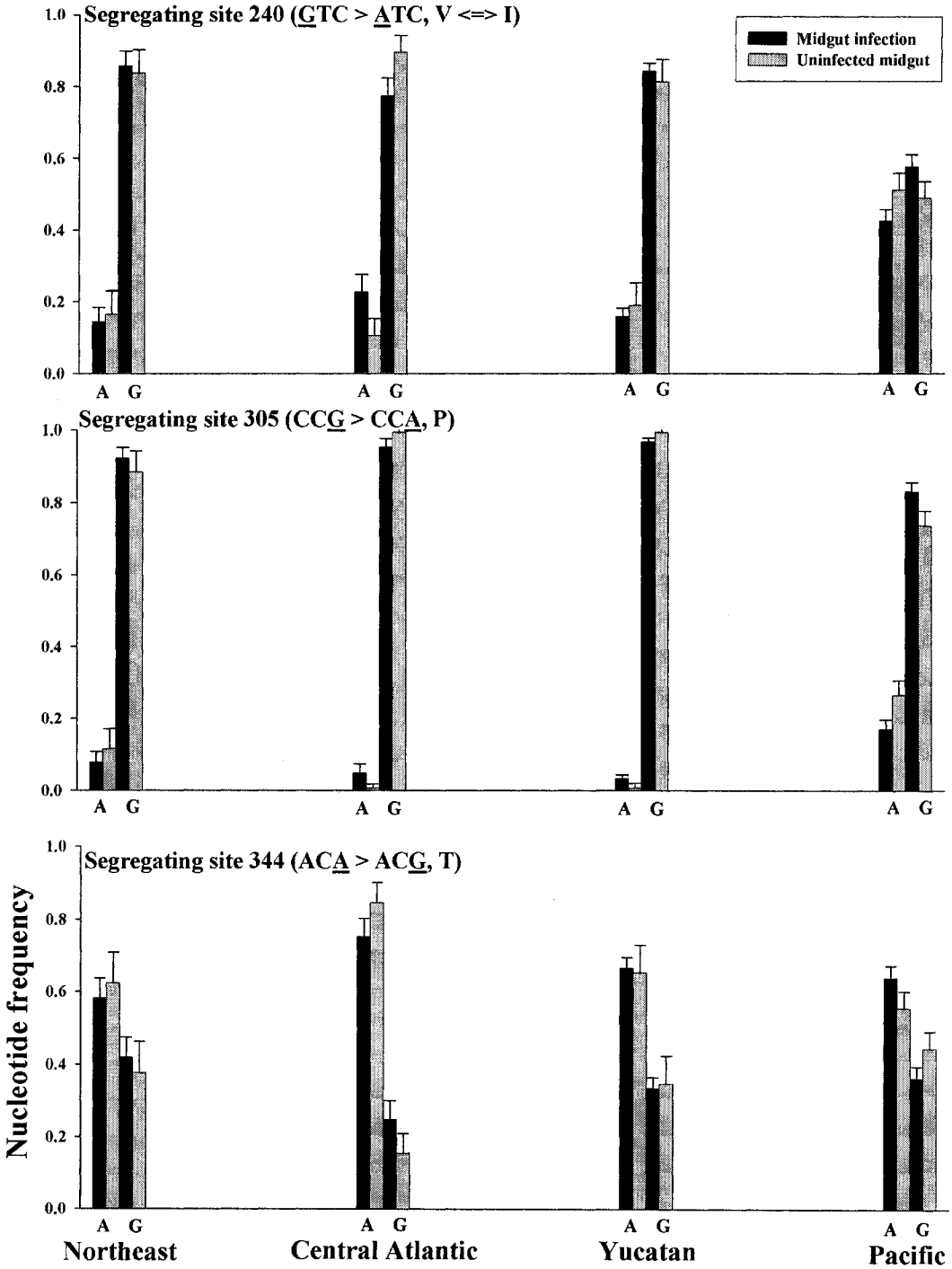


Figure 15. The relationships between nucleotide frequencies and MIR at segregating sites 240, 305, and 344 in the early trypsin gene. Nucleotide frequencies in mosquitoes *with* a infected midgut are indicated with a black bar. Nucleotide frequencies in mosquitoes *without* an infected midgut are indicated with a gray bar.

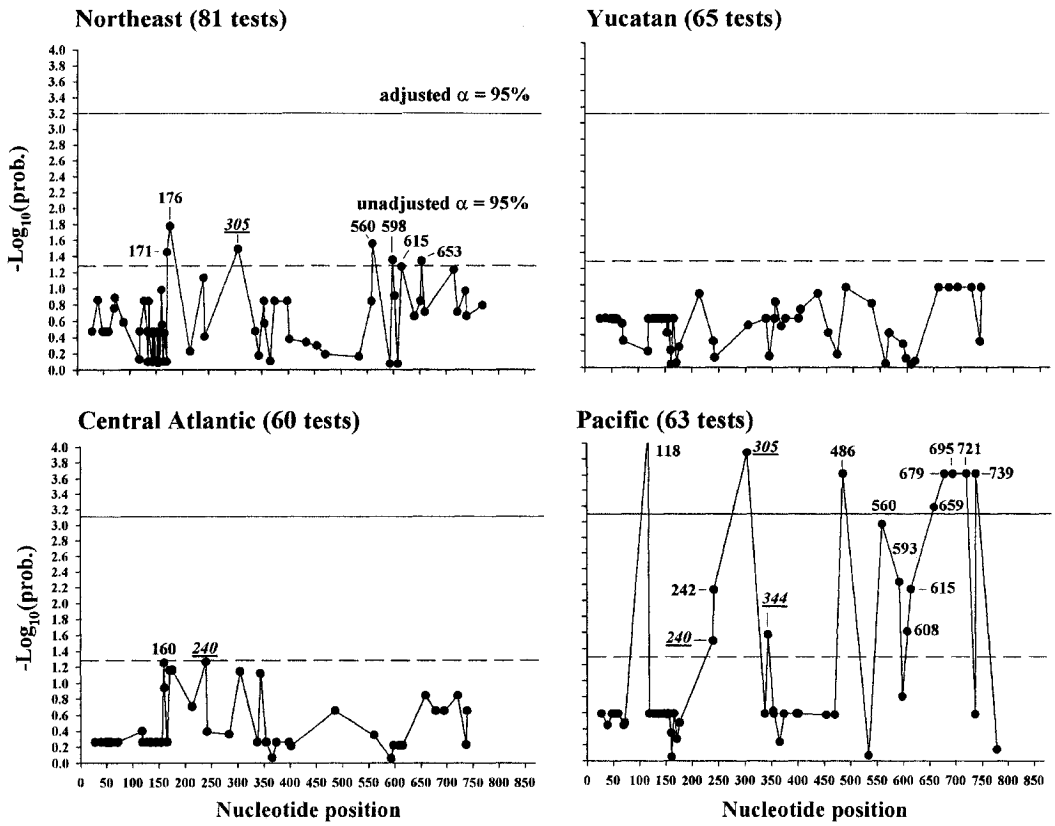


Figure 16. Association mapping of *MIB* based upon genotype frequencies. Contingency χ^2 tests of association between alternate genotypes and the number of mosquitoes with and without an MI were performed. The $-\log_{10}(p)$ score (p = 1- the cumulative probability associated with the test χ^2 statistic) was then plotted against nucleotide position at each segregating site in each of the 4 geographic regions.

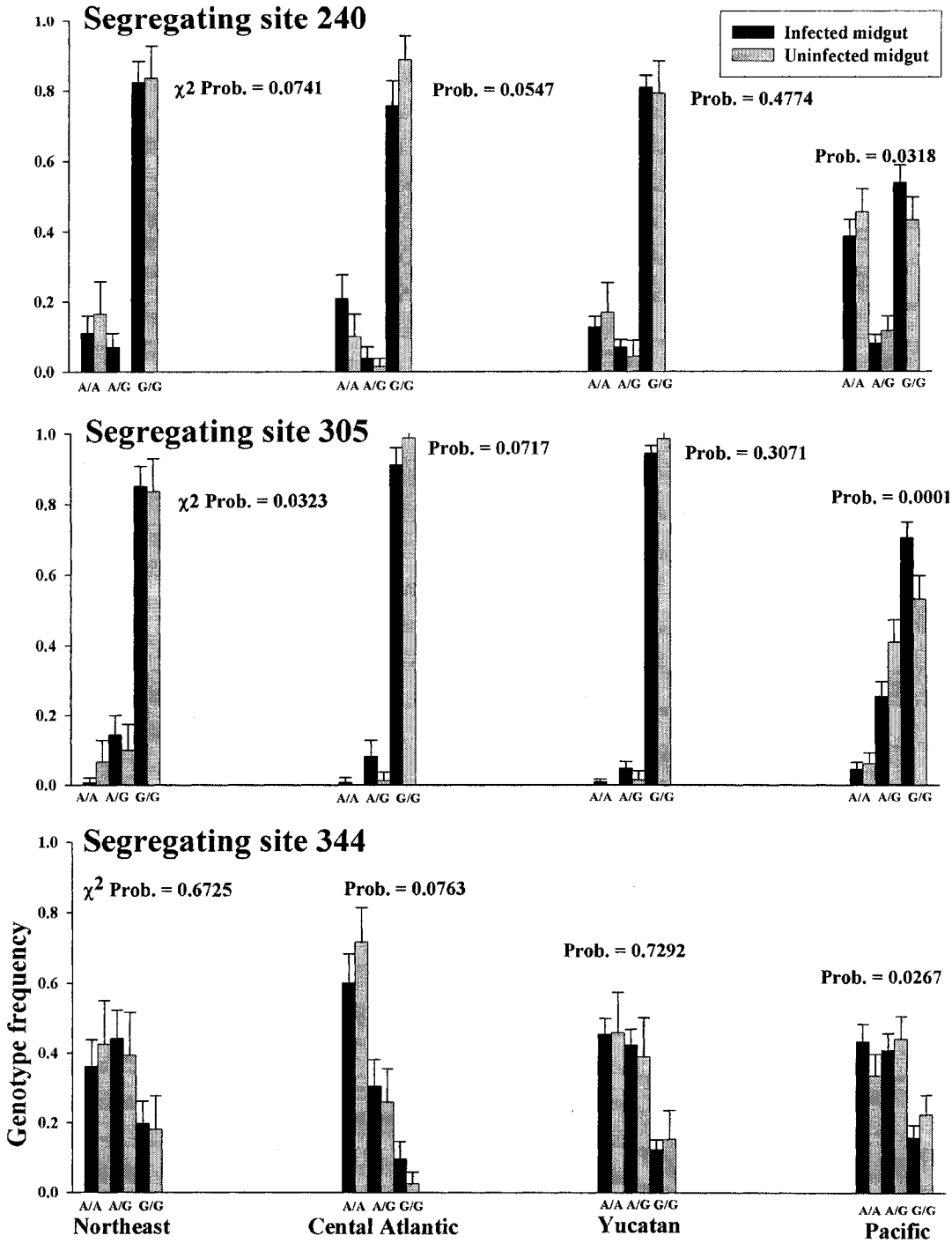


Figure 17. The relationships between genotype frequencies and midgut infection rates (MIR) at segregating sites 240, 305, and 344. Genotype frequencies in mosquitoes *with* a infected midgut are indicated with a black bar. Genotype frequencies in mosquitoes *without* an infected midgut are indicated with a gray bar.

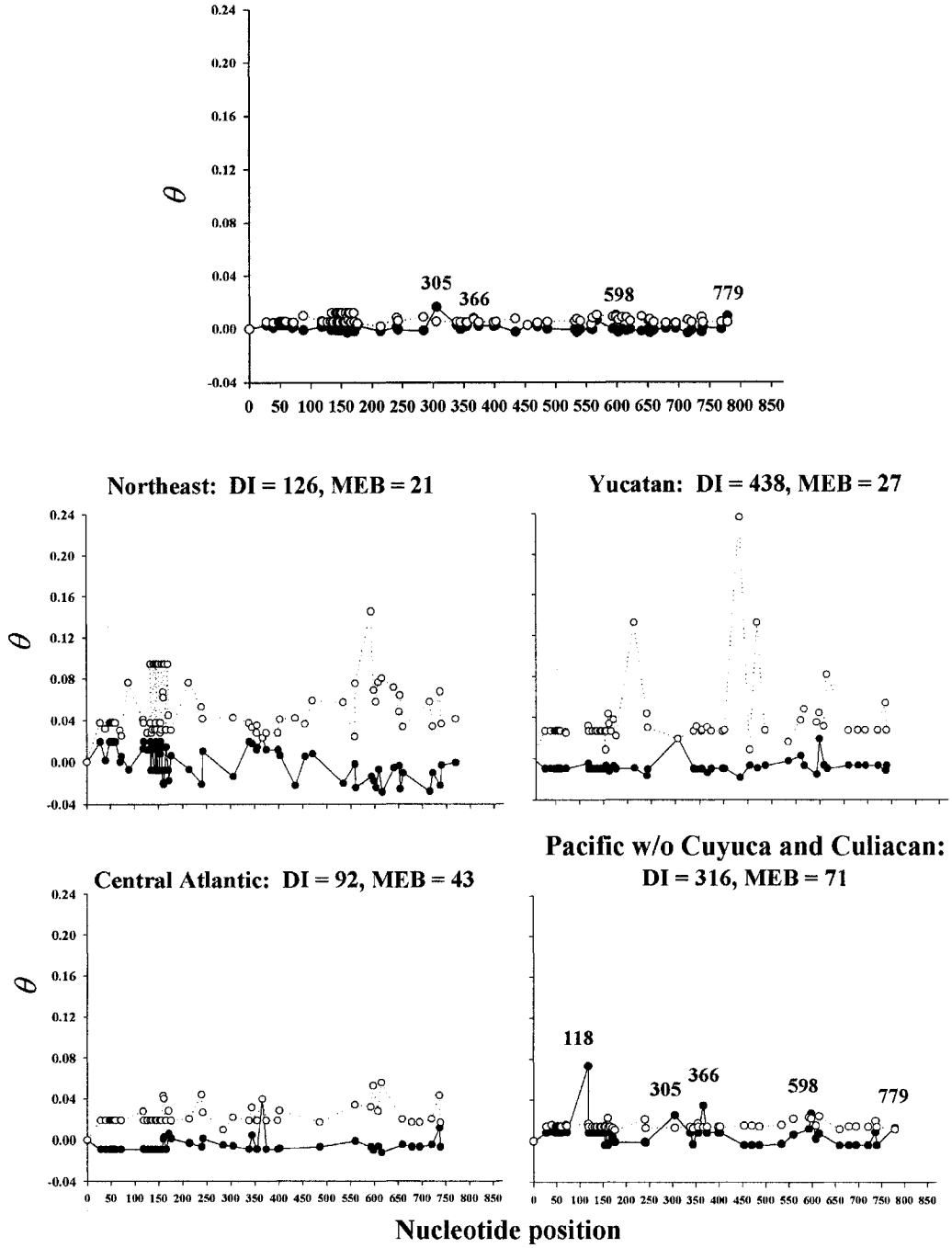


Figure 18. Association mapping of *MEB* based upon nucleotide frequencies. At each segregating site, nucleotide frequencies in *SUSC* mosquitoes were compared to nucleotide frequencies in *MEB* mosquitoes. Putative QTNs are indicated with the number of their segregating site.

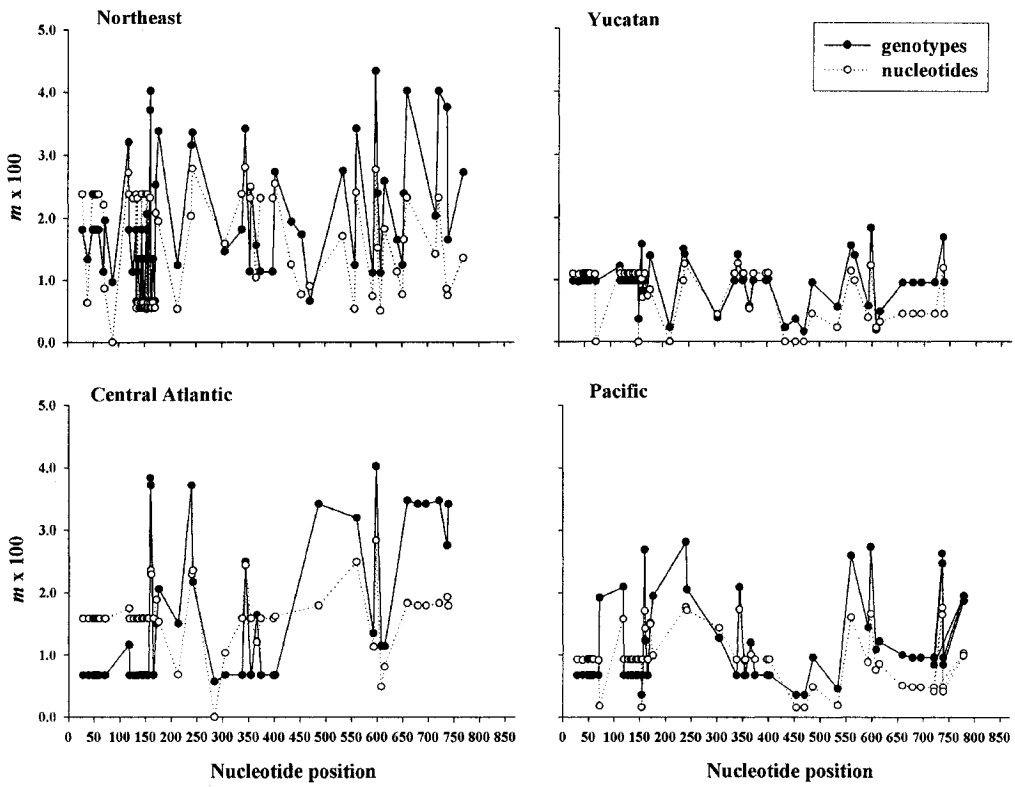


Figure 19. Minimal allelic and genotypic effects (m) that a difference in genotype (solid line and circles) or nucleic acid substitution (dotted line, empty circle) could exert on a phenotype and still be statistically significant.

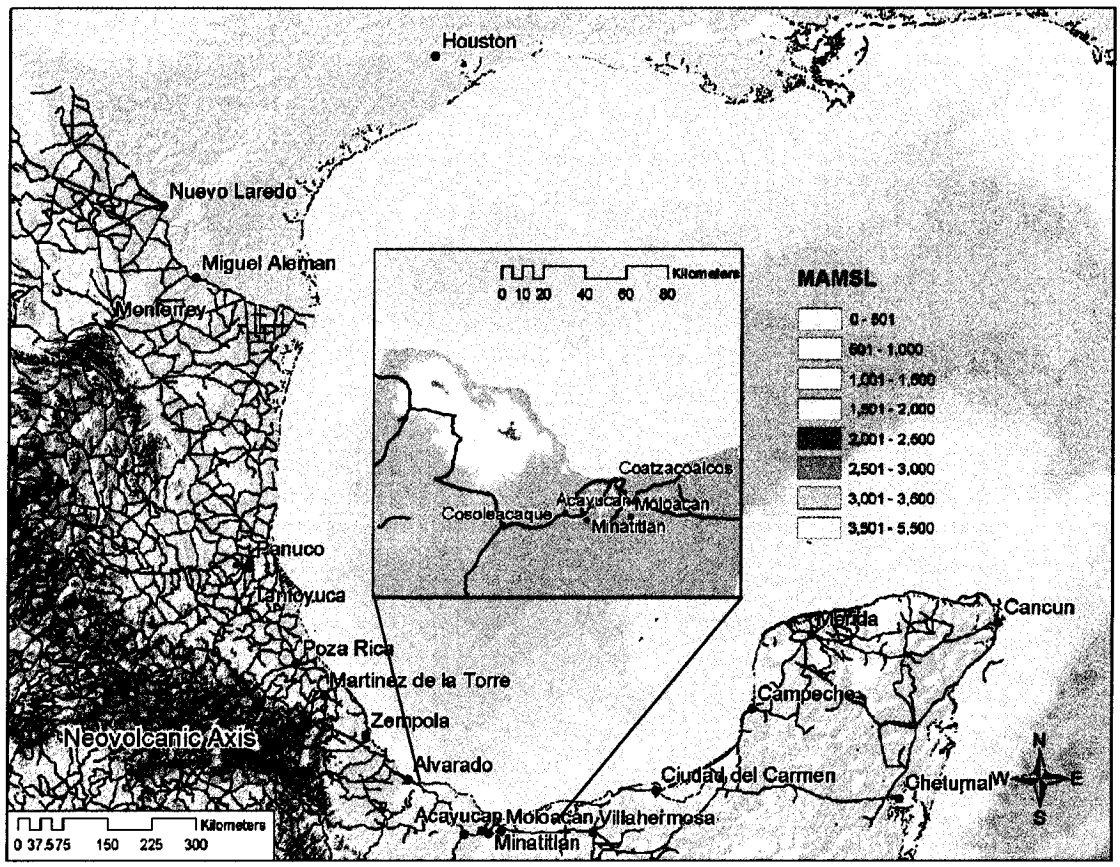
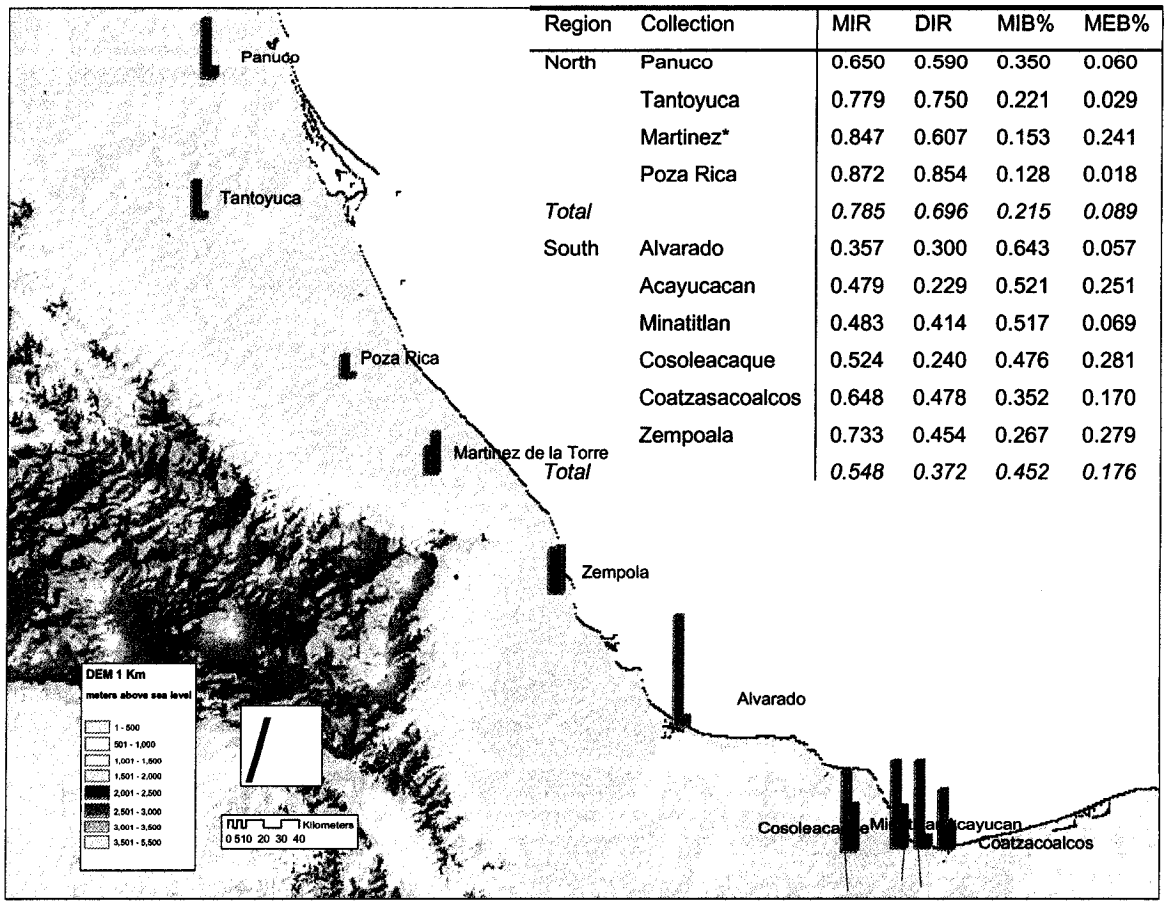


Figure 20. Map of the collections sites and position of the Neovolcanic Axis. The black lines represent main Mexican roads. MAMSL = Meters Above Mean Sea Level.



Region	Collection	MIR	DIR	MIB%	MEB%
North	Panuco	0.650	0.590	0.350	0.060
	Tantoyuca	0.779	0.750	0.221	0.029
	Martinez*	0.847	0.607	0.153	0.241
	Poza Rica	0.872	0.854	0.128	0.018
Total		*0.785*	*0.696*	*0.215*	*0.089*
South	Alvarado	0.357	0.300	0.643	0.057
	Acayucacan	0.479	0.229	0.521	0.251
	Minatitlan	0.483	0.414	0.517	0.069
	Cosoleacaque	0.524	0.240	0.476	0.281
	Coatzasacoalcos	0.648	0.478	0.352	0.170
	Zempoala	0.733	0.454	0.267	0.279
Total		*0.548*	*0.372*	*0.452*	*0.176*

Figure 21. Vector competence of *Aedes aegypti* to Dengue 2 (Jam 1409) in the state of Veracruz. Green bars represent MIB. Red bars represent MEB.

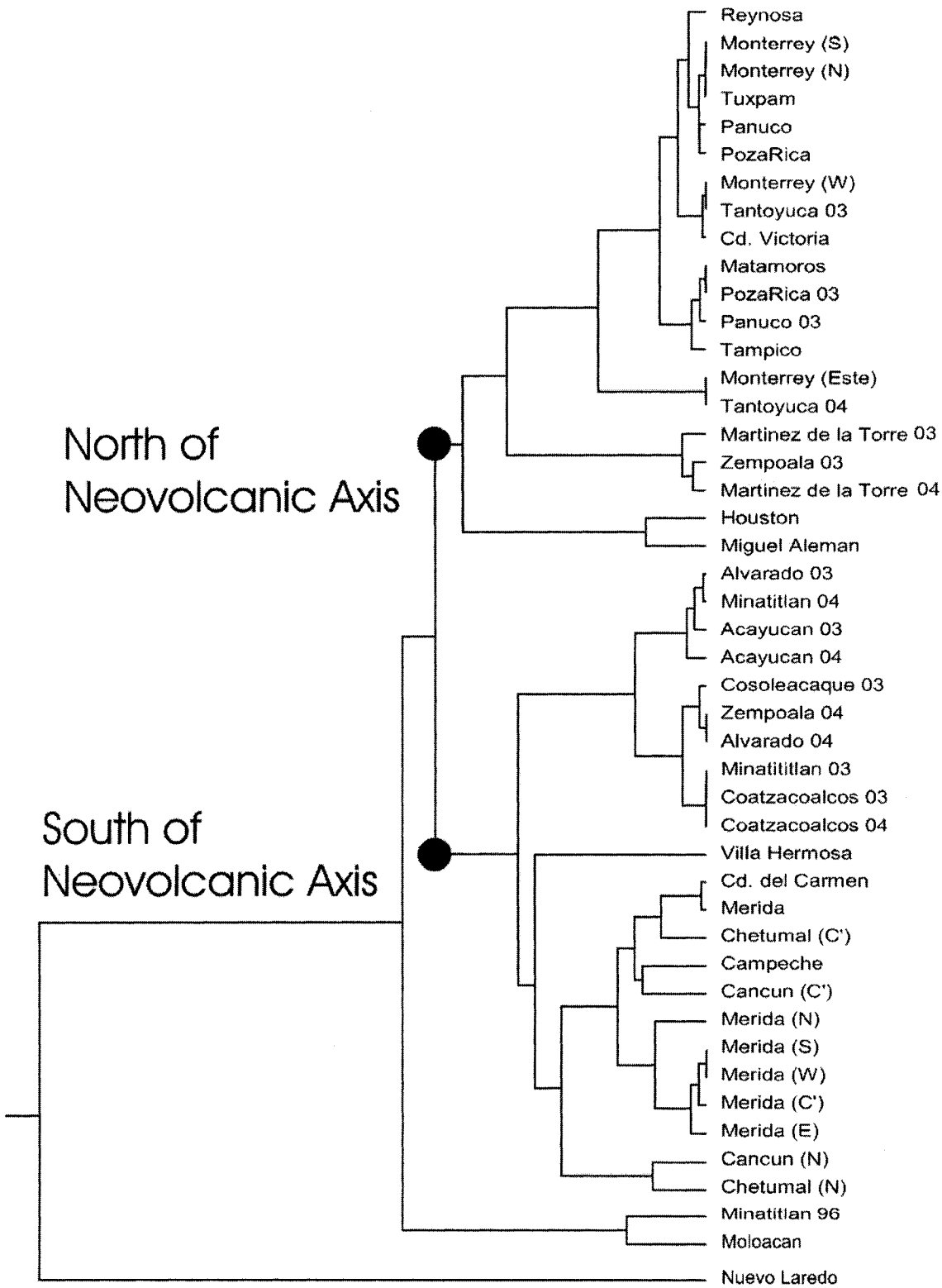


Figure 22. Unweighted Pair Group Method using arithmetic average cluster analysis of pairwise $F_{ST}/(1- F_{ST})$ relationship between collections.

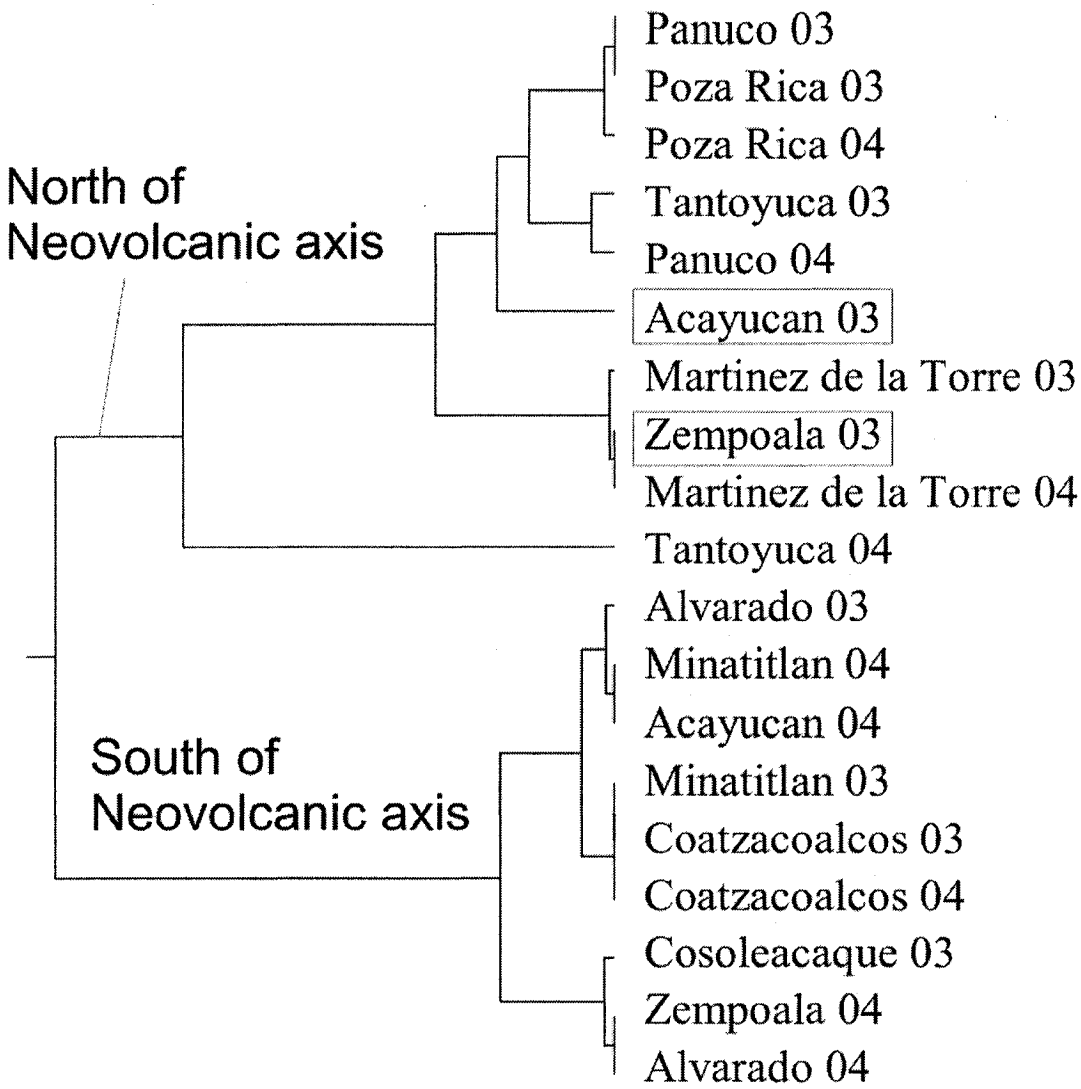


Figure 23. Cluster analysis for the Veracruz collections. Boxes indicated collection placed in an unexpected cluster.

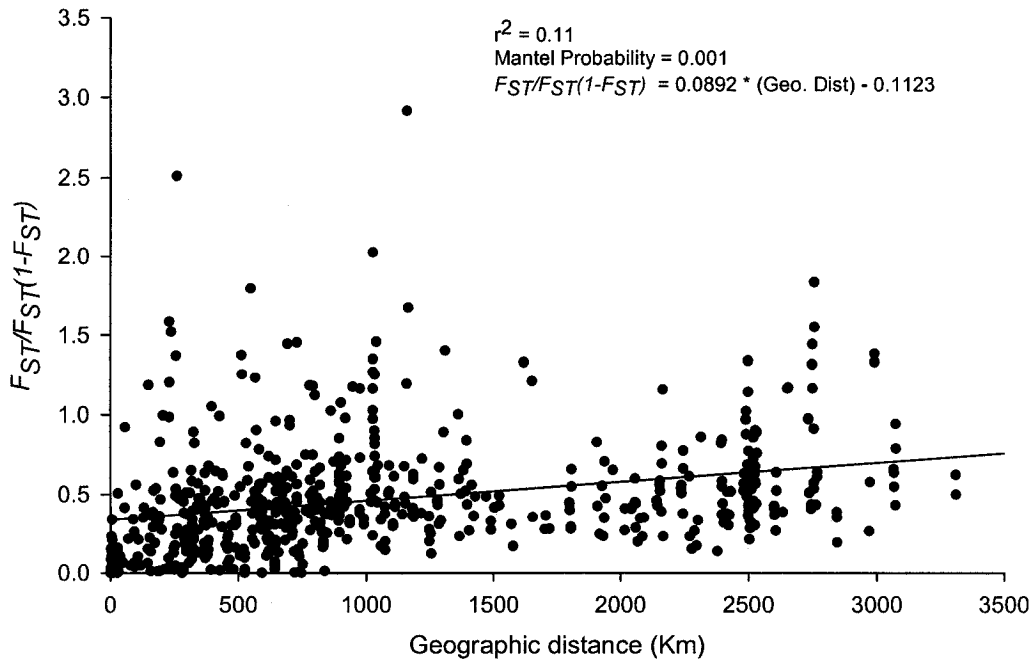


Figure 24. Regression of linearized F_{ST} on geographic distance. Mantel probability obtained from 10,000 permutations. Each dot represents a pair wise comparison between the genetic distance matrix and the geographic distance matrix.

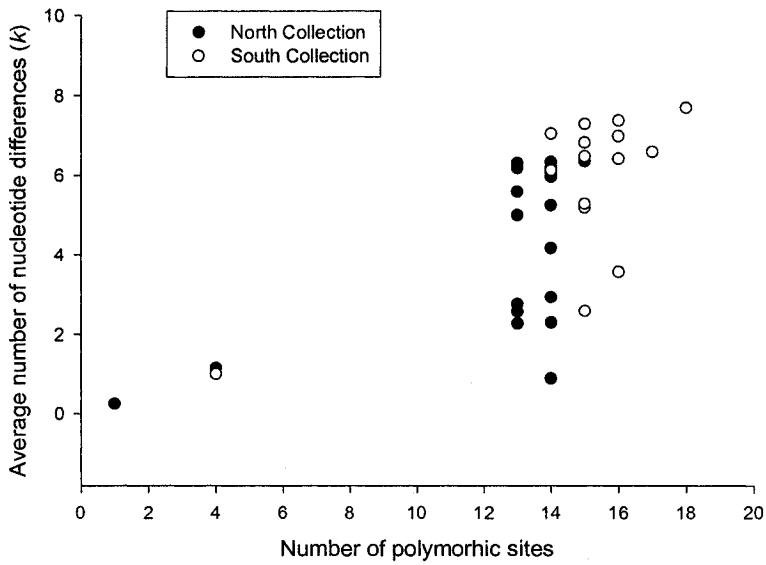


Figure 25. Nucleotide differences versus average number of nucleotides differences. The north and south differentiation is in reference to the Neovolcanic Axis. Data from the same collection sites realized in different years was not pooled.

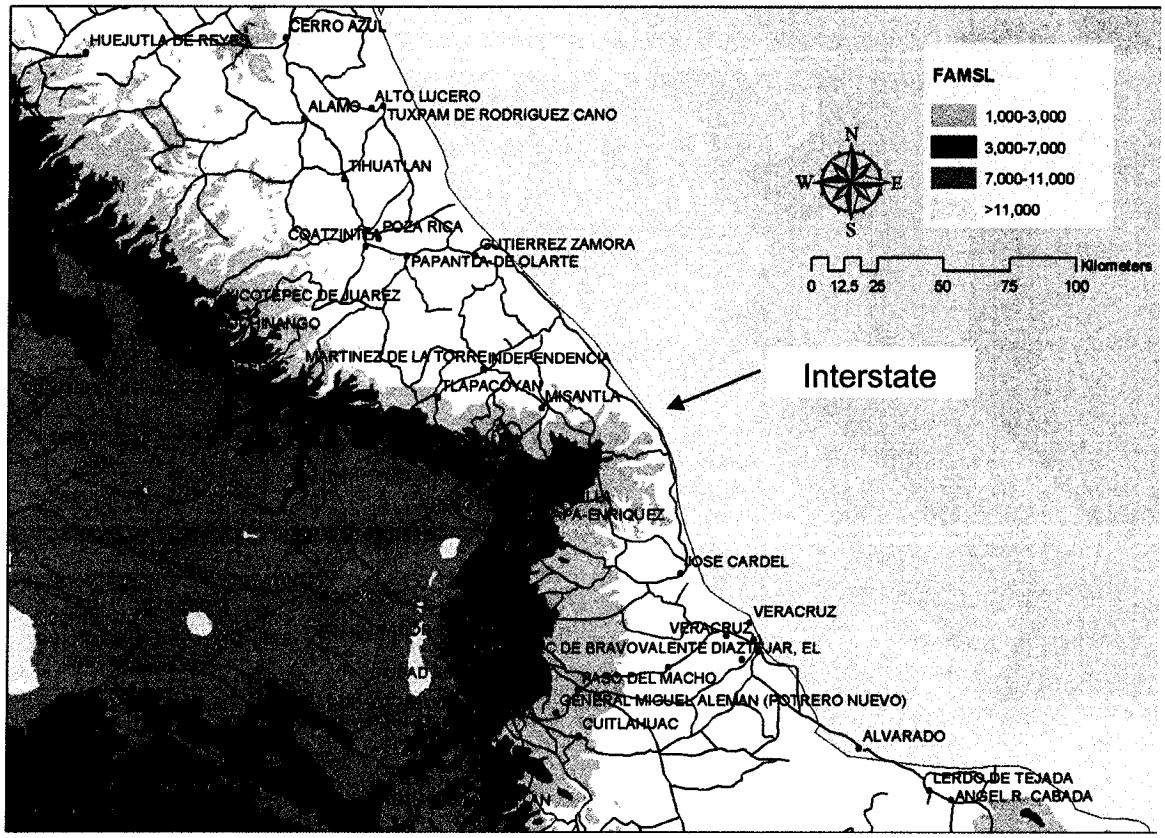


Figure 26. Regional distribution of cities with more than 10,000 inhabitants (Black dots). Black continuous lines represent main Mexican roads. FAMSL=Feet Above Mean Sea Level.

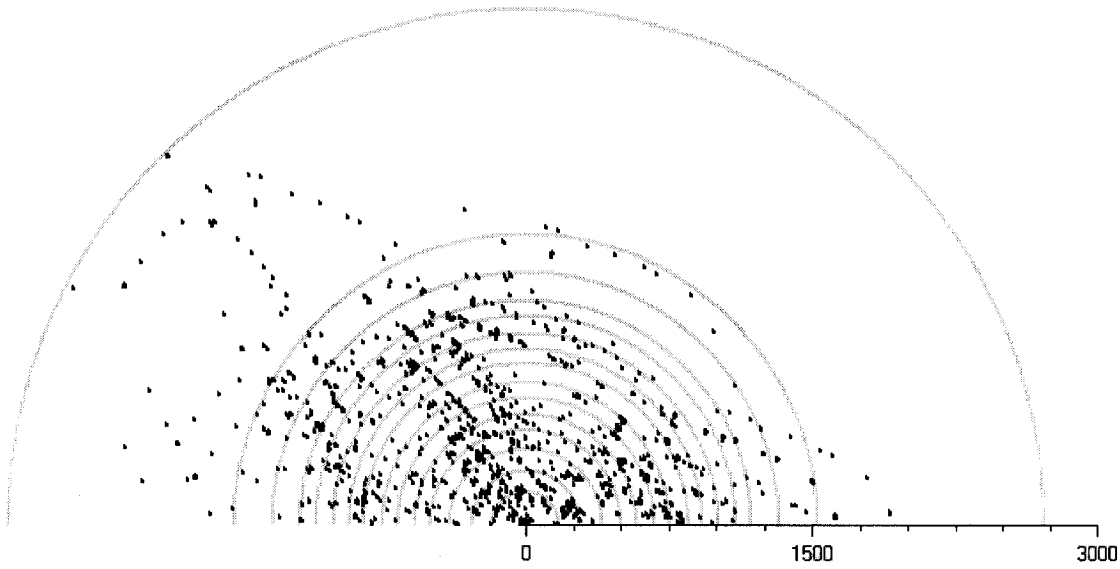


Figure 27. Distance/Direction distribution of the distances and directions among the 55 collection sites. The 1,485 (55*54/2) pairwise distances are shown is kilometers. The gray lines represent the 15 distance classes with n= ~99 (184, 286, 409, 510, 594, 683, 768, 864, 939, 1016, 1112, 1191, 1336, 1533, 2728 km)

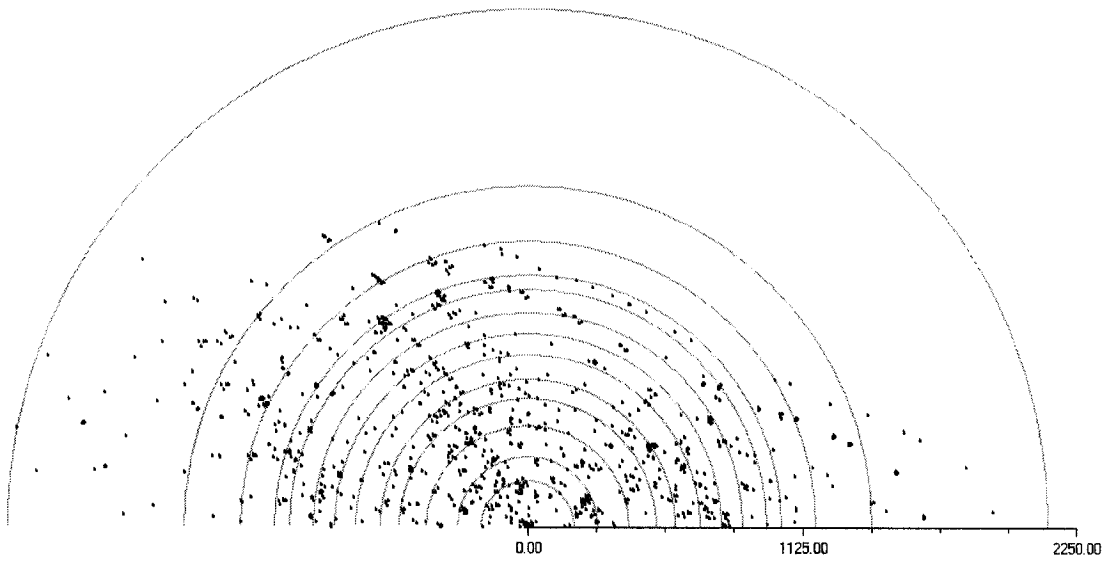


Figure 28. Distance/Direction distribution of the pair wise comparison points after removing non significant and low occurrence haplotypes. The distances are shown is kilometers. The gray lines represent de 13 distance classes with *n* = ~98 (184, 286, 409, 510, 594, 683, 768, 864, 939, 1016, 1112, 1191, 1336 km)

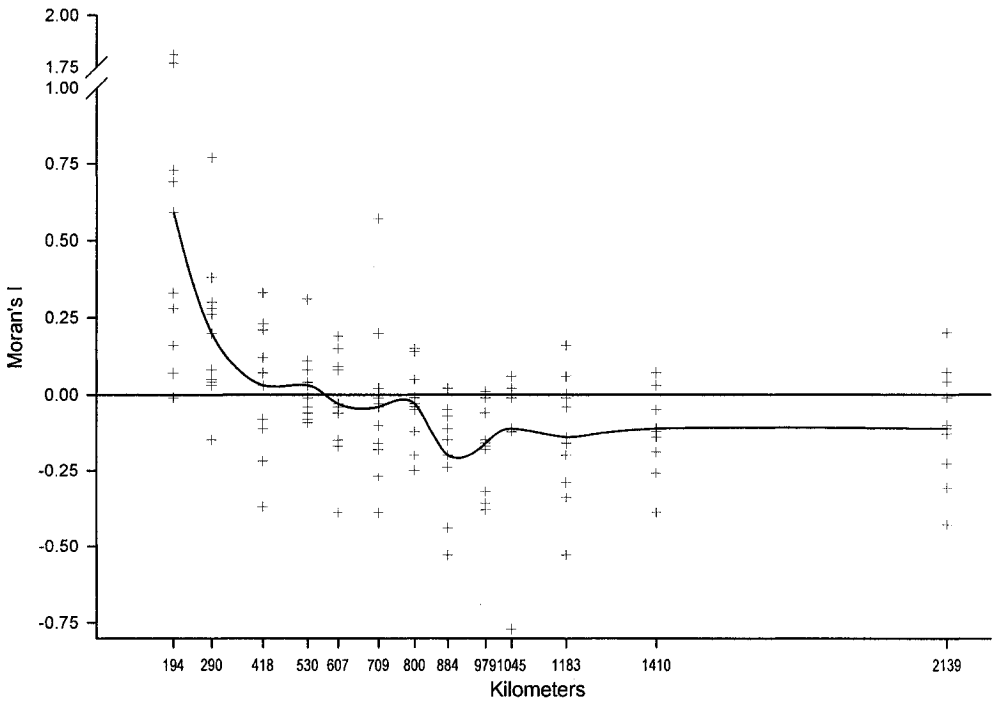


Figure 29. Overall spatial *I*-correlogram for 10 haplotype frequencies. Ordinate: spatial autocorrelation coefficient (Moran's *I*). Abscissa: distance classes in kilometers. Distances indicate the upper limit of the distance class. The average correlogram per distance class are connected by a continuous line.

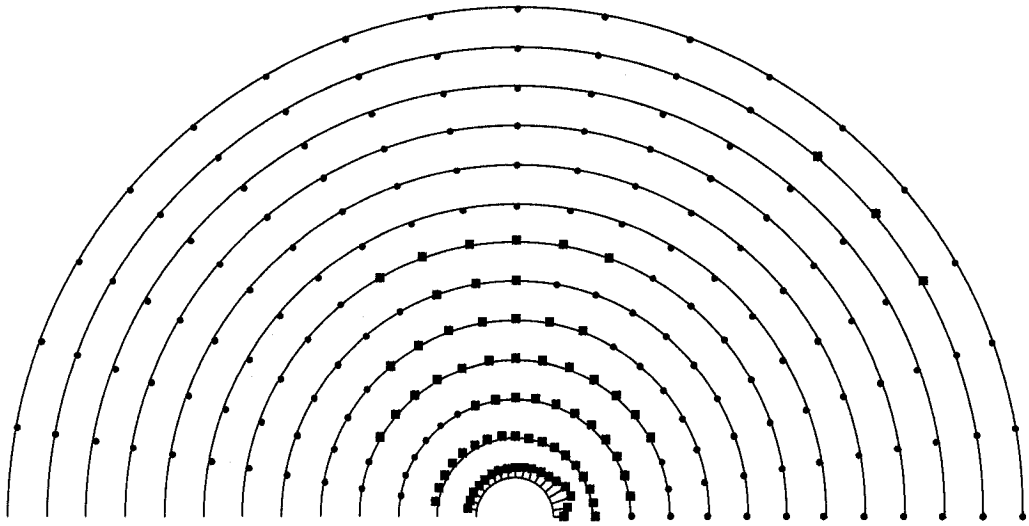


Figure 30. Average bearing correlogram for haplotypes 2, 5, 10, 11, 13, 14, 16, 19 and 24. Each line represents a distance class (194, 290, 418, 530, 607, 709, 800, 884, 979, 1045, 1183, 1410, 2139 km). Each angular intervals has an angle increment of 5°. Red squares represent positive spatial autocorrelation. Blue circles represent negative autocorrelation. The length of the line connecting the data point to the distance class line represent the amount of correlation.

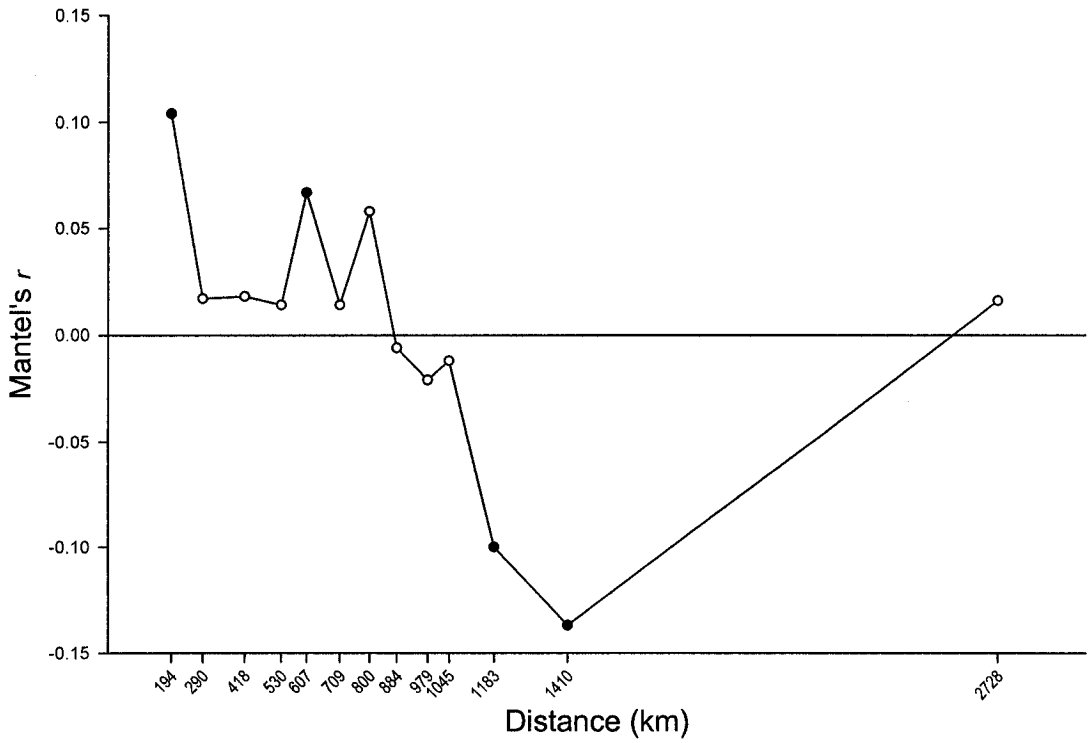


Figure 31. Mantel correlogram for 26 ND4 haplotypes frequencies. Ordinate: Mantel autocorrelation coefficient. Abscissa: distance classes in kilometers. Distances indicate the upper limit of the distance class. Dark circles denote significant (Mantel permutation, $p < 0.05$). The plot is a graphic representation of table 11.

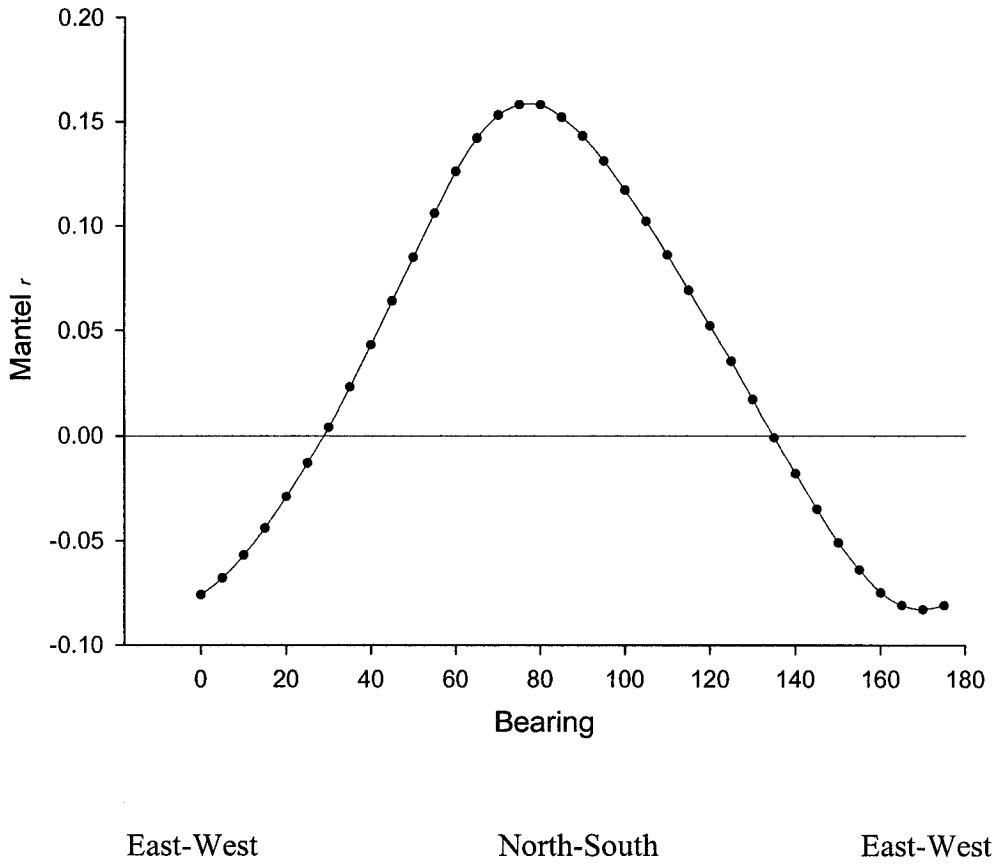


Figure 32. Plot showing the direction of maximum spatial autocorrelation for $F_{ST}/(1-F_{ST})$. Bearing units represent degrees from 0° (pure east) through 90° (pure north) to (but not including) 180°.

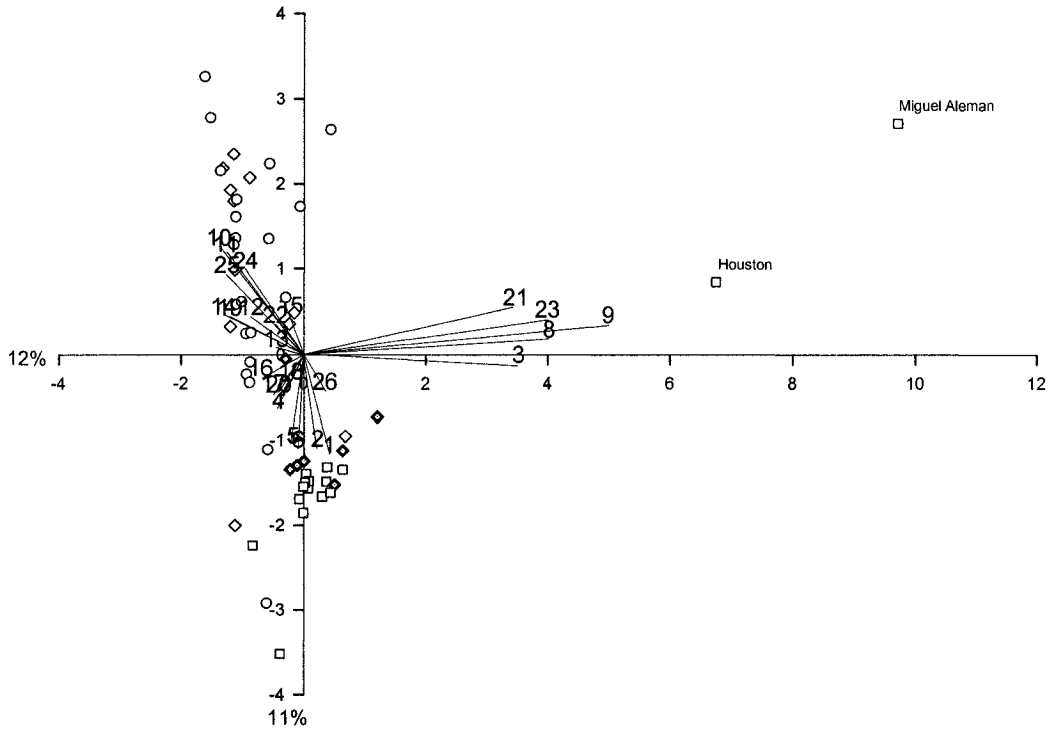


Figure 33. Graphic display of population principal components for 55 collections of *Aedes aegypti* and 26 biplot vectors from ND4 haplotypes. Each symbol represents a population, while each number represents a biplot vector. (PC1 = 12%, PC2 = 11%). The coordinates of the first and second principal components correspond to the values in table 13. (□) refer to North Atlantic collection, (○) refers to South Atlantic collection, (◇) refers to Pacific and Oaxaca (white dot in center) collection.

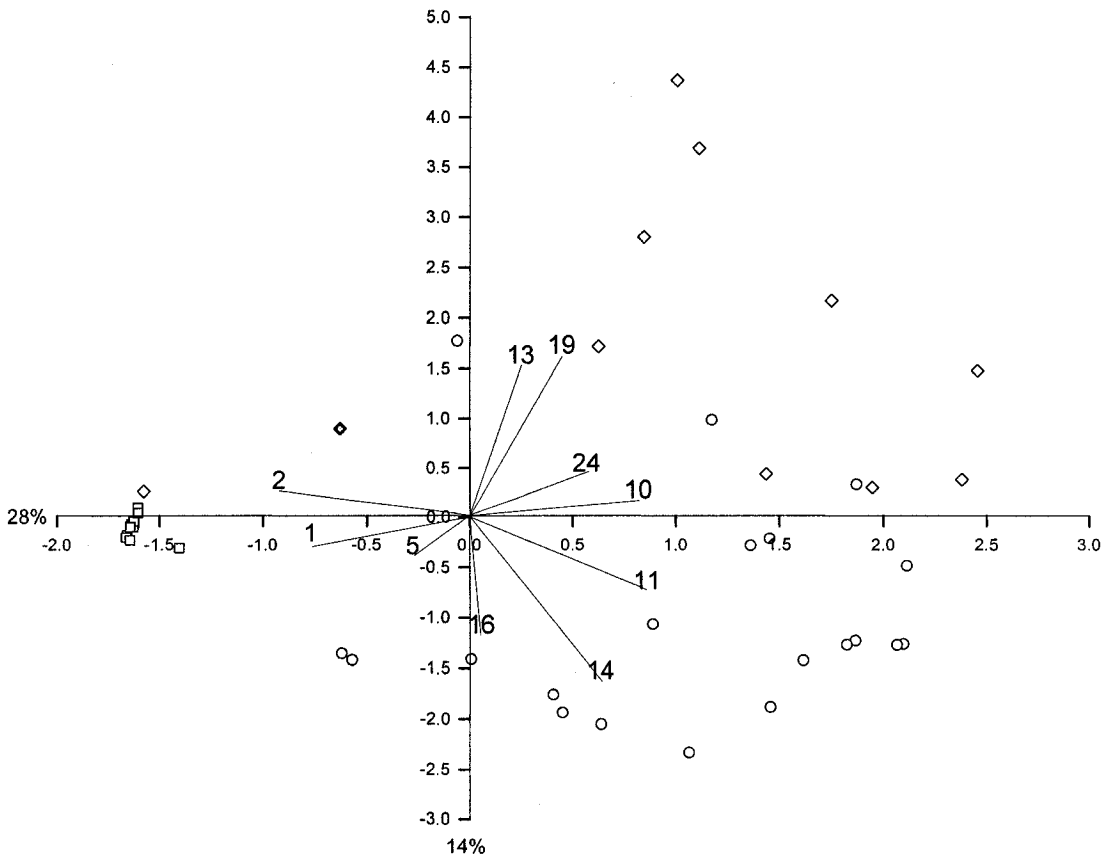


Figure 34. Graphic display of population principal components for 51 collections of *Aedes aegypti* and 10 biplot vectors for individual loci. Each symbol represents a population, while each number represents a biplot vector. (PC1 = 28%, PC2 = 14%). (□) refer to North Atlantic collection, (○) refers to South Atlantic collection, (◇) refers to Pacific and Oaxaca (black fill white center) collection. This graph was obtained from BIPLOT (Black *et al*).

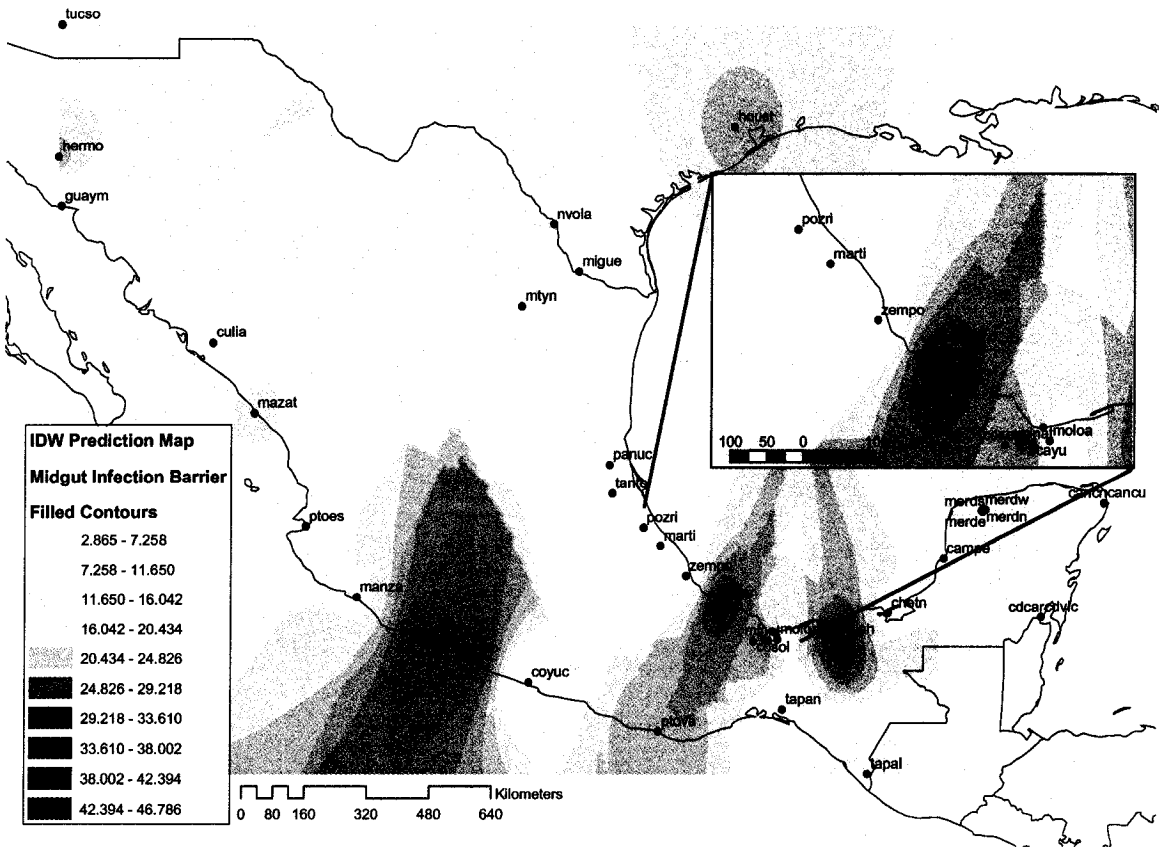


Figure 35. Inversed distance weighting MIB prediction map obtained from the Mexican *Aedes aegypti* vector competence study . ERMS = 9.9. Power: 2.2956. Neighbors to include: 5 (include at least 3). 700 km search sphere.

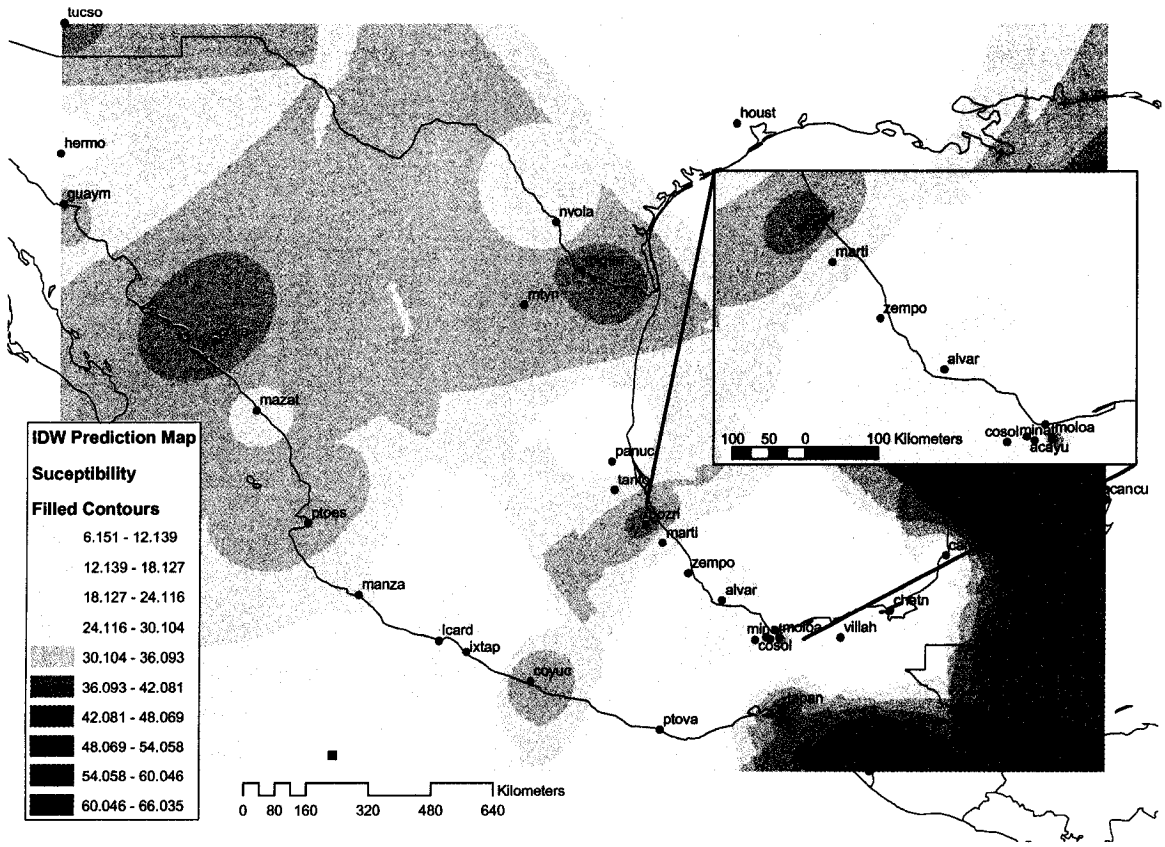


Figure 36. Inversed distance weighting SUC prediction map obtained from the Mexican *Aedes aegypti* vector competence study. ERMS=14.37. Power: 2.322. Neighbors to Include: 5 (include at least 3). 700 km search sphere.

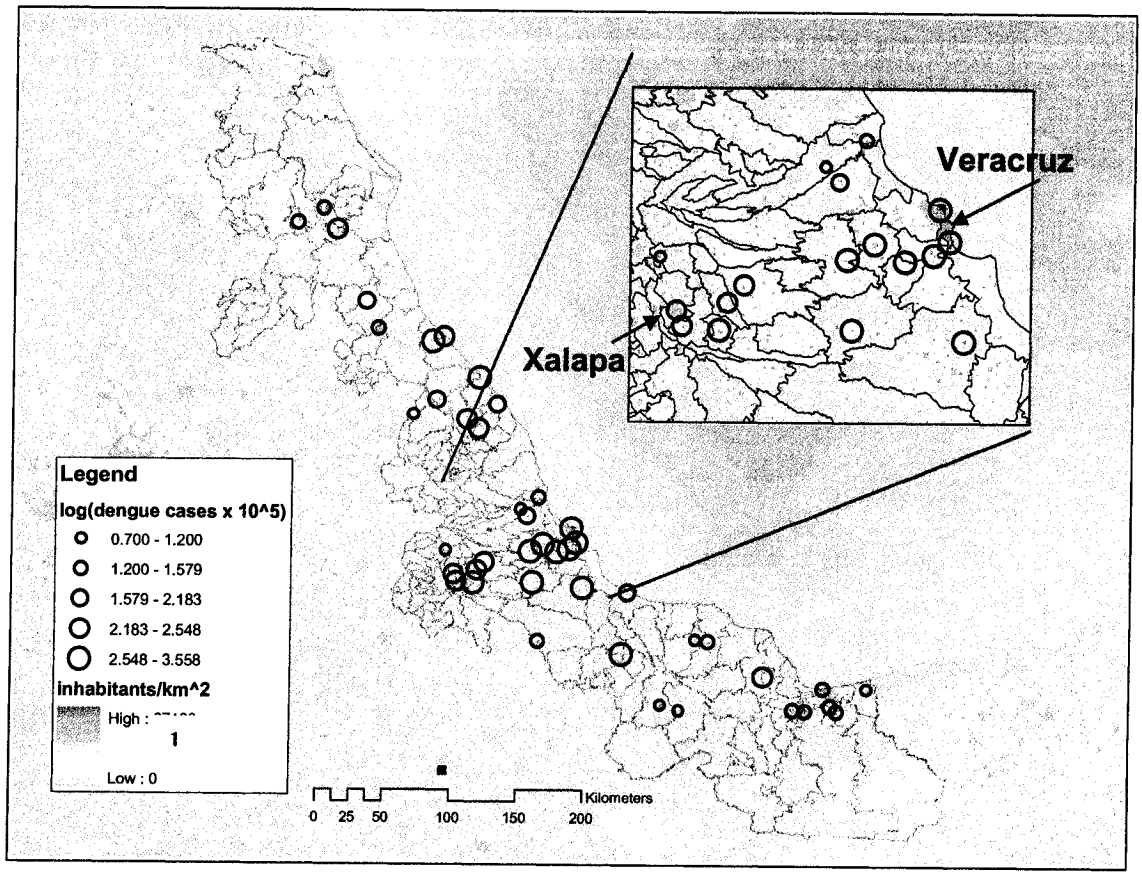


Figure 37. Distribution of total dengue cases (classic + hemorrhagic x 100,000) normalized by total number of inhabitant per municipality and Log_{10} linearized. Areas without people are represented in white (LandScan 2003 Global Population Database).

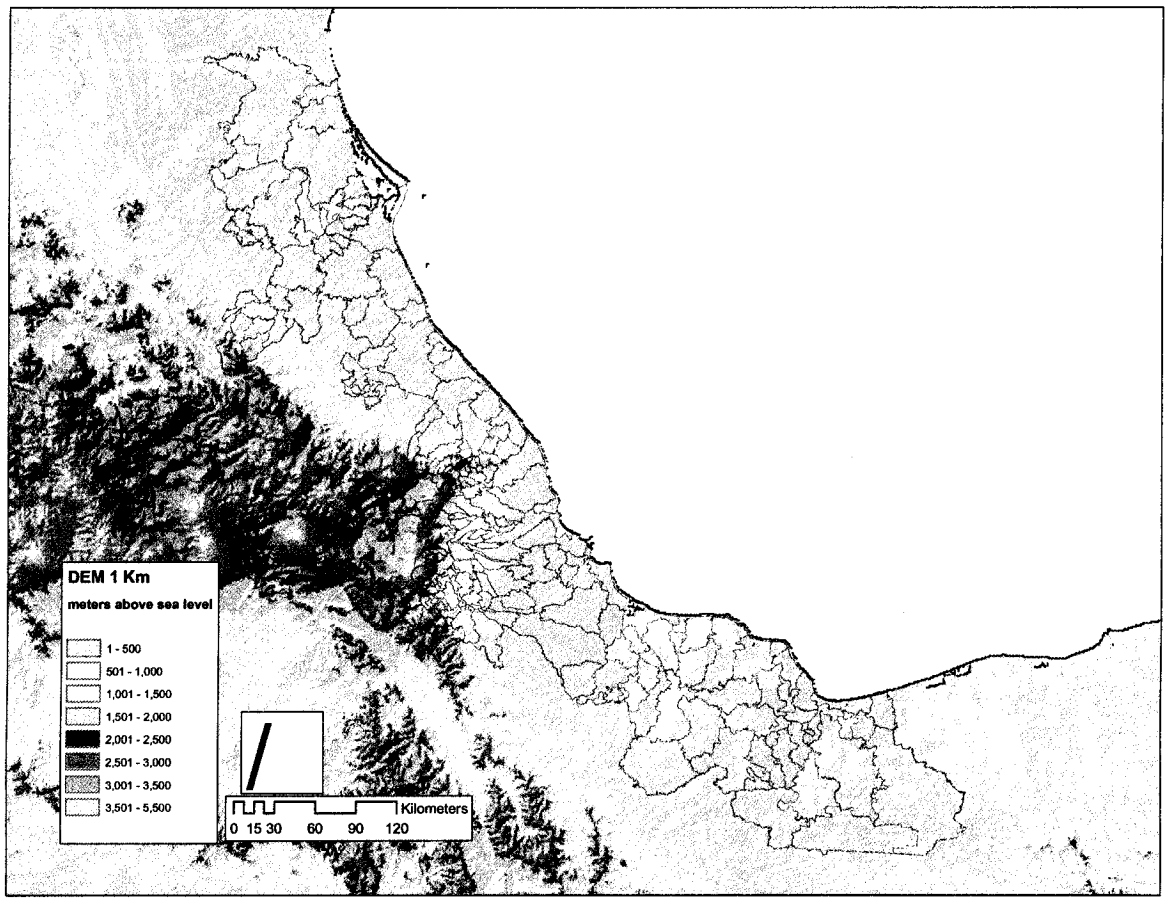


Figure 38. Veracruz state topography based on 1 km digital elevation model. Altitudes used for calculations were obtained from INEGI tables.

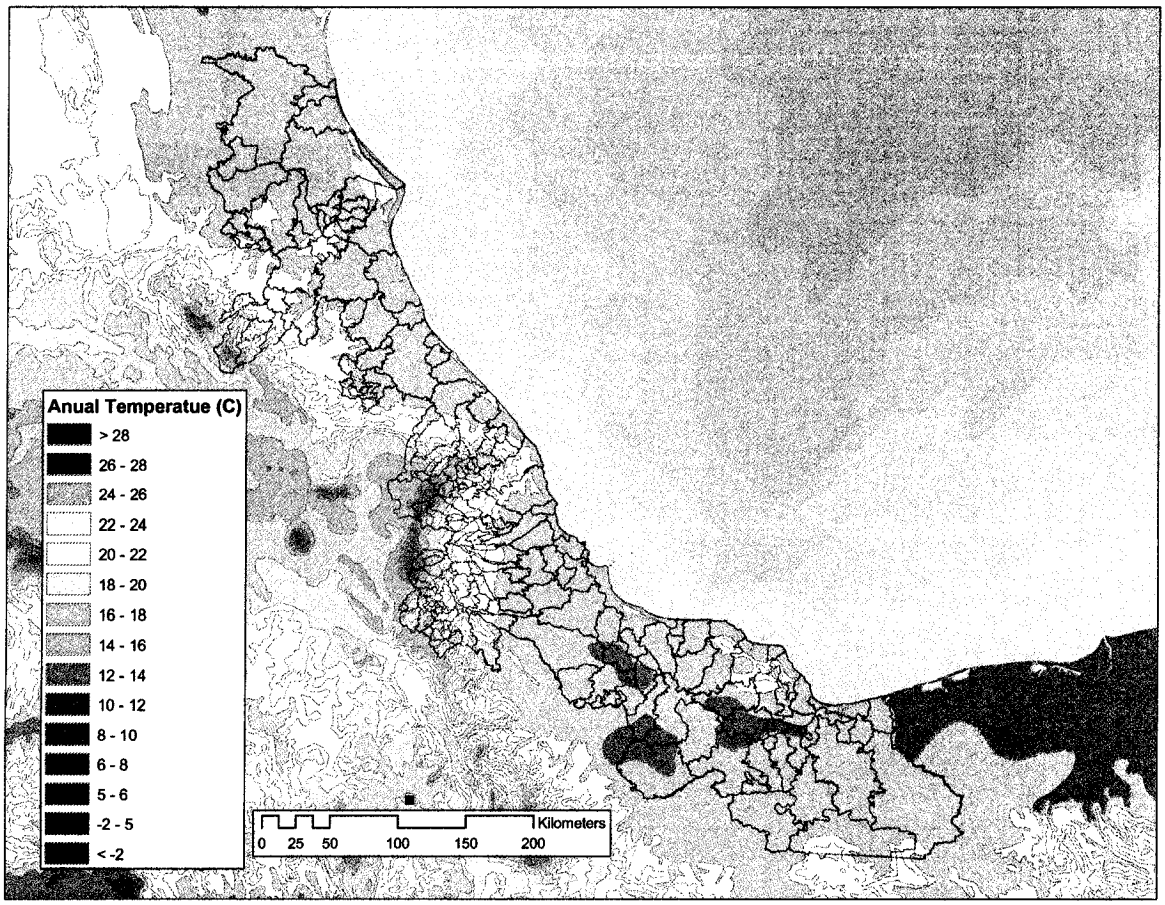


Figure 39. State of Veracruz isotherms (filled contours). The most variation in temperature can be observed as a result of the Neovolcanic axis.

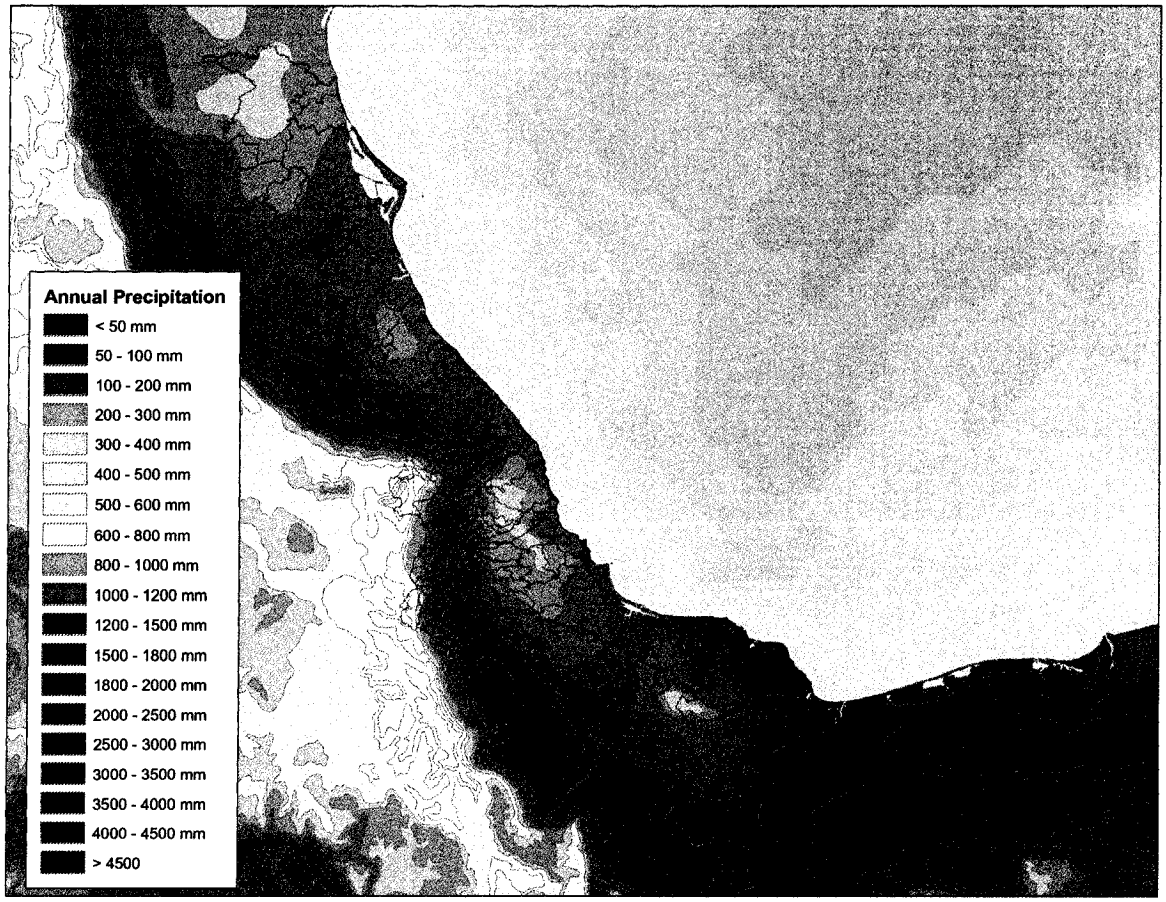


Figure 40. State of Veracruz isohyets (filled contours). The Neovolcanic Axis and the Sierra Madre Oriental generate orographic lifting for the humid Gulf of Mexico winds.

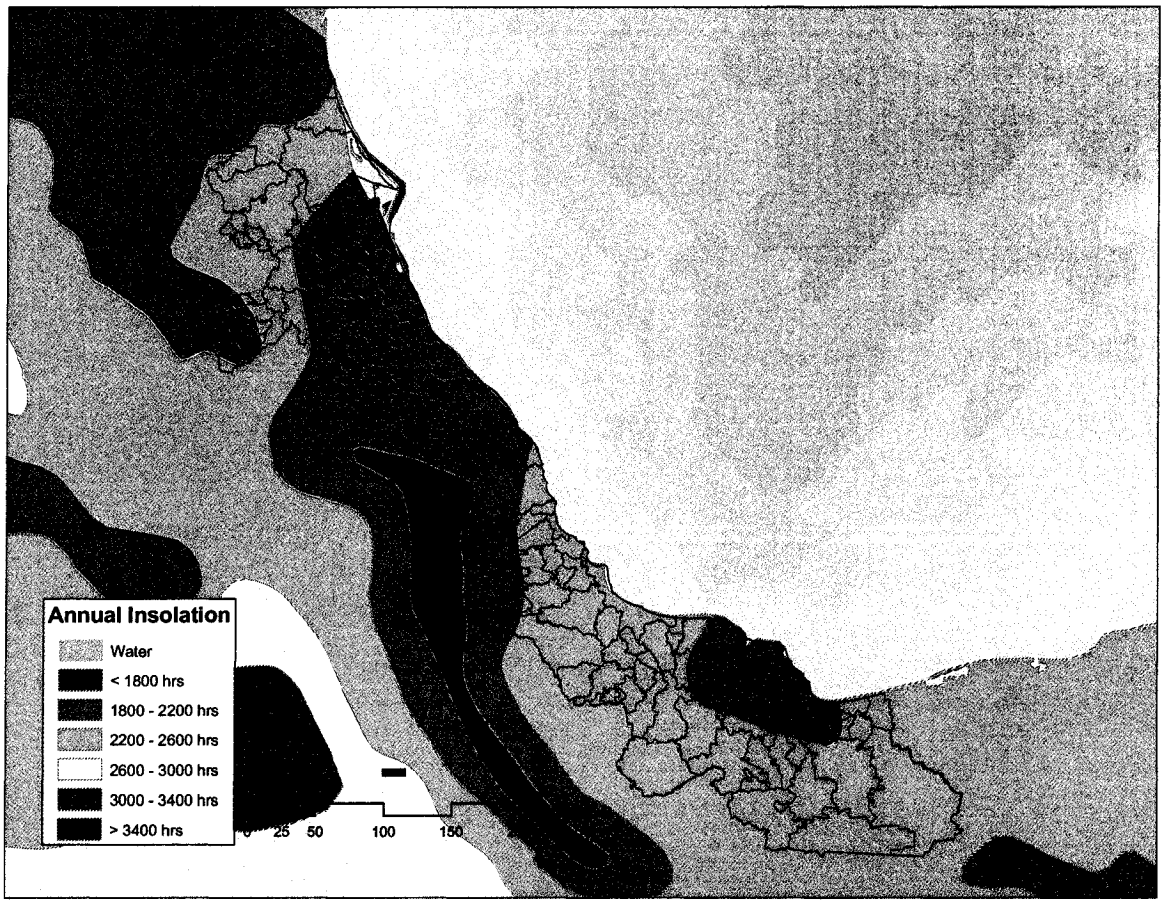


Figure 41. Map of annual solar radiation. Solar radiation is based on days with precipitation. Areas with small annual insolation have more days with precipitation. Annual radiation = Average (Maximum Radiation (Year*i*) – Minimun Radiation (Year*i*)) / 2.

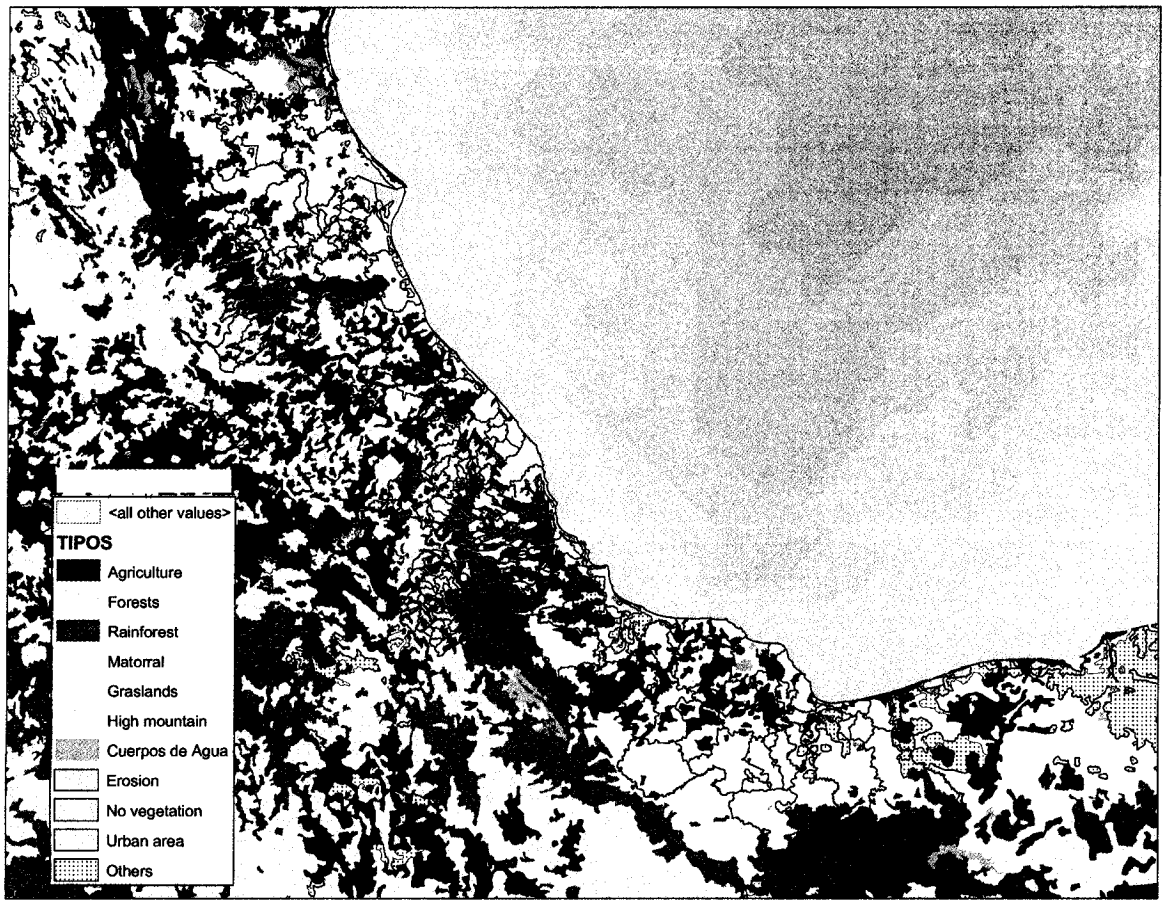


Figure 42. Land use coverage. Many vegetation types were group into larger classes for the purpose of display. Agriculture and induced grasslands clearly are the predominant vegetation types. Larger areas of rainforest can be found south of the Neovolcanic axis.

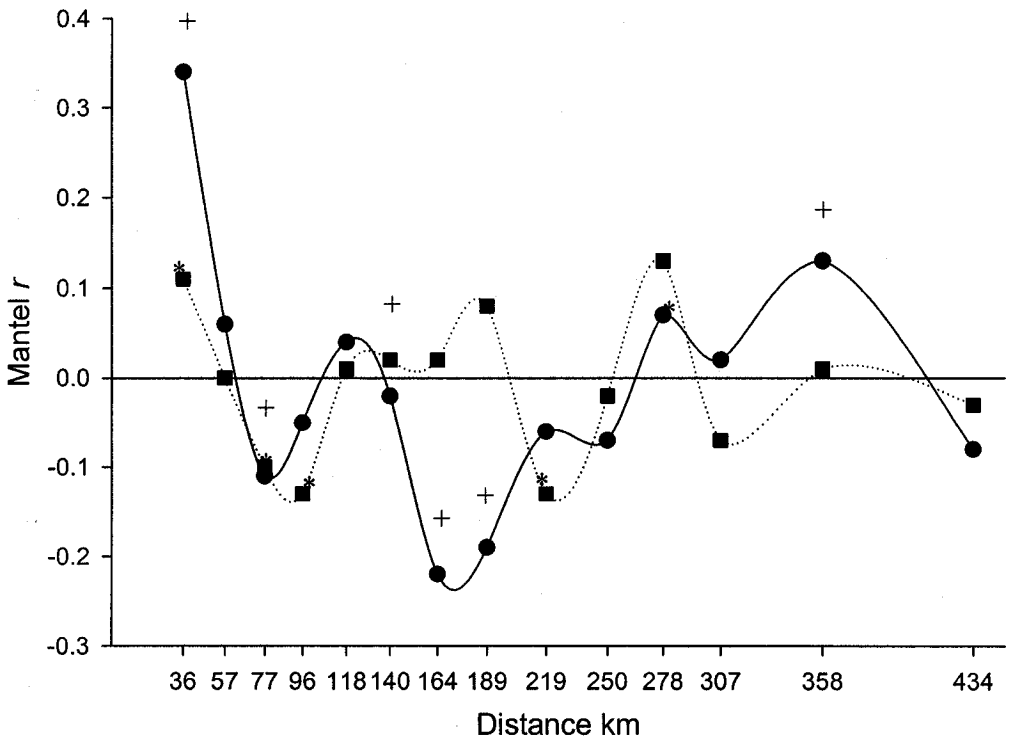


Figure 43. Mantel correlogram for dengue fever cases (circles) and dengue hemorrhagic (squares) cases in the state of Veracruz. Distances indicate the upper limit of the distance class. + denotes a DF significant distance class and * denotes a DHF significant distance class (Mantel permutation, $p < 0.05$)

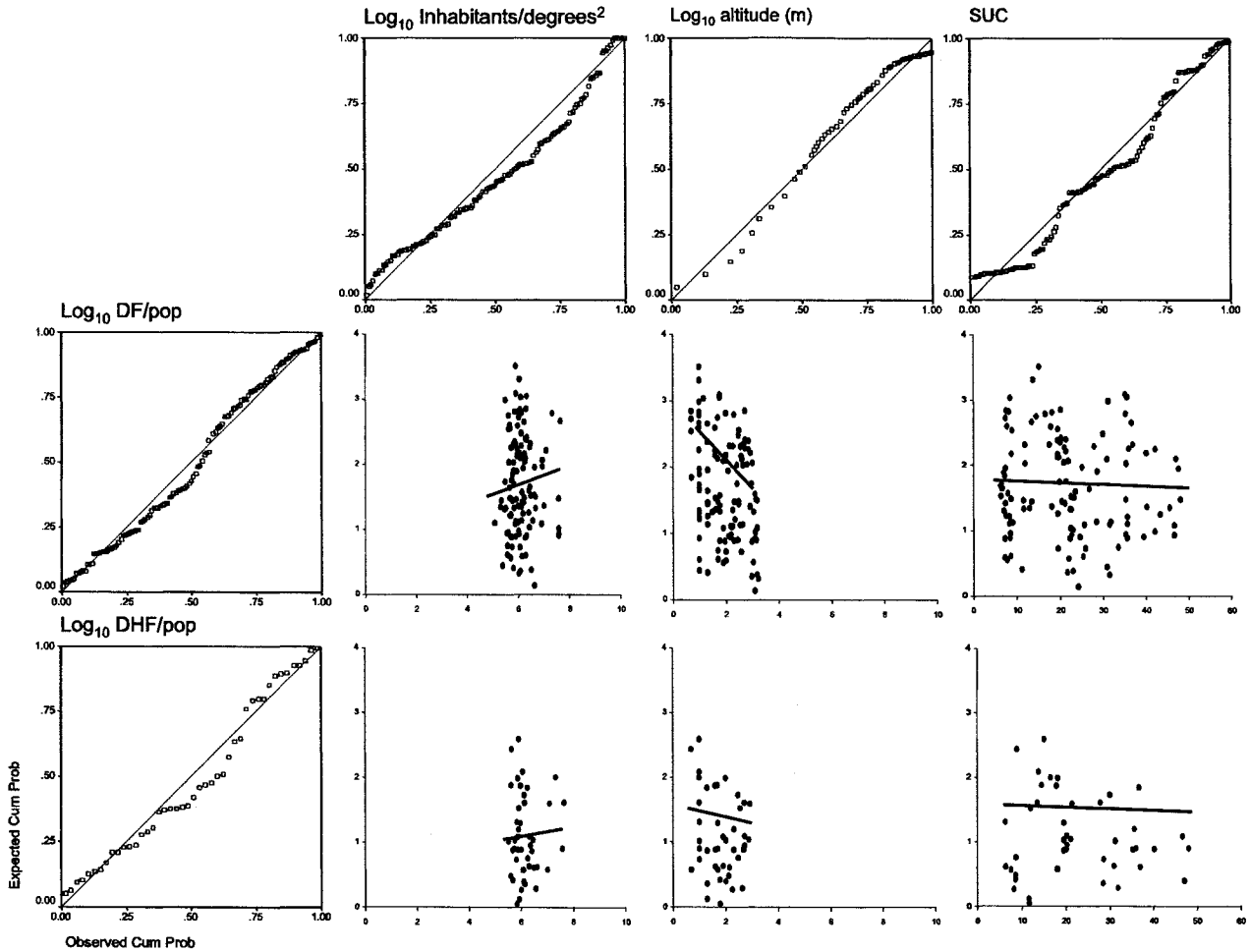


Figure 44. Linear regression of dengue cases (DF and DHF) vs. inhabitant density, altitude and mosquito susceptibility (interior plots, *x*-axis correspond to the top variables. *y*-axis to the left variables). Outer plots are P-P plots to assses normality (observed cumulative probability vs. expected cumulative probability).