Testimony for "Science of Zika: The DNA of an Epidemic"
House of Representatives
Committee on Science, Space, and Technology

Dr. Daniel Neafsey
Associate Director, Genomic Center for Infectious Disease
Broad Institute of MIT and Harvard
Cambridge, MA

SUMMARY
The Zika epidemic has caught the world off guard. We lack fundamental understanding of how the virus moves from person to person via mosquitoes, information that is crucial for an effective response to the epidemic. Though the virus was little studied prior to the recent epidemic that began in Brazil, the *Aedes* mosquito that spreads it is better known, because *Aedes* mosquitoes also spread the viruses that cause Yellow Fever, Chikungunya, and Dengue Fever. Despite the mosquito's importance, we lack foundational resources to pursue DNA-based studies of the biology and transmission of Zika. This resource is gap is critical. Infectious disease epidemiology has been transformed by DNA during the last 10 years into a rich digital information science, allowing biologists and public health agencies track the spread of outbreaks over time and space, and learn about what mosquito and human factors contribute to disease spread. We can now tackle emergent infectious diseases like Zika using efficient and innovative genetic tools, and the scientific community stands ready to develop and apply these tools to Zika to protect vulnerable populations within our borders and around the world.

RATIONALE FOR DNA-BASED EPIDEMIOLOGY
Insight into the biology of how *Aedes* mosquitoes are able to spread disease can be gained through sequencing and mapping the *Aedes* genome (the entirety of its DNA), by studying the DNA of mosquitoes collected across time and space during the outbreak, and by looking for DNA factors that impact interventions and Zika transmission. DNA sequencing will generate foundational resources that benefit a diverse array of scientific studies, leading to the improvement of existing disease interventions, the preservation of their efficacy, and the discovery of new interventions.

It is exceedingly important that we learn more about the nature of Zika transmission, and identify and invest in new approaches to block disease transmission. We do not yet have a drug to treat Zika patients, nor vaccines with which to protect people from infection. **Our primary disease intervention option at the present time is to stop *Aedes* mosquitoes from spreading the disease.** It is unlikely that we will ever eliminate *Aedes* mosquitoes within our own borders using conventional mosquito control measures, and much of the world lacks the socio-political infrastructure to even mount such an effort. Because the mosquitoes are here to stay, we must learn how to disrupt their biology or strategically suppress their populations in a way that impacts Zika transmission.

It is not a coincidence that Zika is spread by the same *Aedes* mosquitoes that transmit Yellow Fever, Chikungunya, and Dengue Fever, but a consequence of the remarkable predilection *Aedes* mosquitoes have for feeding on humans, and their capacity to utilize human-created environments for reproduction. Out of almost 4000 mosquito species on

this planet, very few have assembled the combination of molecular, behavioral, and ecological traits that make *Aedes* such an effective disease vector.

DNA sequencing of mosquito genomes has already given us a glimpse into some of the biological traits that favor human disease transmission by *Aedes*, *Anopheles,* and other mosquitoes[1–3]. We can now appreciate mosquito-born disease transmission as a product of complex biological interactions between a mosquito, the pathogen, and humans. Because Zika is a new and understudied disease, knowledge about the details of these interactions remains thin for this virus.

There are many ways to translate knowledge of mosquito/human/pathogen interactions into disease control measures. Examples include:

1) Insecticide resistance detection and surveillance
Insecticides can be very effective at controlling mosquito populations. The historic elimination of Yellow Fever, Malaria, and other mosquito-borne disease from the US can be attributed to mosquito control. Global Malaria mortality fell by 60% between 2000 and 2015, in large part due to the intensive use of insecticide-treated bednets and indoor insecticidal spraying[4].

The routine consequence of intensive insecticide campaigns is resistance. Scientists are concerned that further gains against Malaria will be limited by the high prevalence of insecticide resistance in many regions of Africa[5]. Resistance to many insecticides is already locally common in *Aedes* mosquitoes[6,7]. The genetic factors behind this resistance are not well understood. DNA-based studies comparing resistant mosquitoes versus susceptible mosquitoes can provide indicators of resistance to use in surveillance programs. Similar to strategies being used to limit antimicrobial resistance in bacterial pathogens, we can maintain insecticide efficacy for longer through rational, judicious use of insecticides, informed by DNA-based markers to track the origin and regional spread of resistance.

2) Mosquito population studies
Using DNA to understand which mosquito populations are connected vs. isolated can tell us much about the Zika epidemic. By comparing the DNA of mosquitoes sampled from different locations, we can address a wide range of fundamental questions. For example: Is the geographic spread of Zika due solely to human movement, or do mosquitoes also play a role? Which mosquito populations harbor high levels of insecticide resistance? Will resistance spread from one region to another? Our DNA-based understanding of *Aedes* mosquito populations is very limited compared to *Anopheles* Malaria mosquitoes, where such information has illuminated many aspects of disease transmission and spread. For example, this information has influenced our understanding of how mosquito populations rebound from control programs and restore themselves after dry seasons, and which mosquito populations are smaller (and therefore ripe targets for control)[8]. A recent small-scale DNA study has determined that *Aedes* mosquitoes have established a year-round population of mosquitoes in the Capital Hill neighborhood[9]. Where else are new *Aedes* mosquito populations establishing themselves, and what is their source?

3) Genetic modification of the mosquito
The prospect of using genetically modified mosquitoes to control disease spread is no longer science fiction. For example, Hadyn Parry's testimony will likely describe the

Oxitec method of inserting 'self-limiting' genes into mosquitoes that are released in large numbers, to achieve sharp reductions in local mosquito populations.

Recently, a powerful new DNA tool has arrived with the potential to translate biological insights into control measures very quickly. The CRISPR-*Cas9* DNA editing system gives us unprecedented power to insert, turn off, or modify genes in virtually any organism, including mosquitoes. When coupled with a 'gene drive', another DNA-based tool that can rapidly spread a CRISPR mutation through a mosquito population, we have the power to modify wild mosquito populations to restore sensitivity to insecticides, or to bite other animals instead of humans, or to kill pathogens they ingest during a blood meal and not transmit them to the next person they bite. CRISPR gene drives have been demonstrated to be viable in the laboratory with fruit flies[10] and *Anopheles* Malaria mosquitoes[11,12]. CRISPR gene drives are an example of the increasingly direct connection between DNA and vector-borne disease control.

IMPROVING THE *AEDES* MOSQUITO GENOME MAP

One example of a specific DNA-based resource that will assist the Zika response is an improved genome map for *Aedes aegypti*, containing virtually all of the DNA sequence in long pieces for all of the chromosomes. The current genome map for the *Aedes aegypti* mosquito was a significant advancement when it was first released in 2007[1], but it was far from perfect. DNA sequences are long strings of nucleotides, or 'letters': A, T, G, and C. At 1.5 billion nucleotides long, the *A. aegypti* genome is far larger than that of many other mosquitoes, and much of that extra length is composed of repetitive DNA sequences, (eg ATC-ATC-ATC-ATC . . .). Most modern DNA sequence is read in small pieces of 100 nucleotides in length. Assembling a non-repetitive genome map from short nucleotide sequences can be a computationally difficult task, akin to putting together a large jigsaw puzzle without a picture to look at as a guide. Assembling a large, *repetitive* genome from short sequencing reads is more like putting together a jigsaw puzzle with no picture, *where most of the pieces have the same color and similar shape.* As a result, the existing reference genome assembly for *A. aegypti* is in over 36 thousand pieces (despite having only 3 chromosomes), and is estimated to be missing as many as 20% of the genes. Furthermore, there is evidence that some of the assembled regions have been put together incorrectly.

While this genome map has led to some important breakthroughs in understanding the vectorial capacity of *A. aegypti*, we can do better. **DNA sequencing technology has improved in quality and price dramatically since 2007, making it more than 10,000 times cheaper to sequence a nucleotide of DNA now compared to then**. In recognition of this opportunity, Dr. Leslie Vosshall at Rockefeller University organized the *Aedes* Genome Working Group (AGWG) in January 2016, to coordinate efforts in the vector genomics research community aimed at improving the genome assembly map[13]. With no central funding, and organized via Twitter, this upstart group has made rapid progress exploring a wide range of approaches to employ new sequencing technology in pursuit of an improved *Aedes* reference genome. For a price tag of a few hundred thousand dollars, this group aims to produce a much improved version of the genome map released in 2007 at a cost of $18 million. Examples of the new technologies that will enable this feat include:

1) DNA sequencing machines that produce very long strings of nucleotides.
Pacific Biosciences, a member of the AGWG, manufactures sequencers that can read DNA in strings of more than 10,000 nucleotides, far longer than the 700 nucleotide

strings produced for the 2007 genome map. To return to the puzzle analogy, this is equivalent to increasing the size of the puzzle pieces and reducing their number.

2) Machines and techniques that tell us which pieces of sequenced DNA belong to the same territory in the map. Three biotechnology companies (10X Genomics, BioNano Genomics, and Dovetail Genomics) have been recruited to the AGWG because their technology can inform which DNA sequences are proximal to each other on chromosomes. Application of these tools will mean that genome map puzzle pieces are no longer randomly shuffled in one big pile, but organized into small piles of pieces that correspond to different regions of the puzzle.

3) More sophisticated software to assemble the map. Compared to 2007, there are now many more tools available to put together genome maps. The previously mentioned AGWG corporate partners have their own software to apply their respective data types to the task of genome assembly. This software is often optimized for the attributes of the human genome, however, and may be inexpert at combining different, complementary data types in order to produce the best possible map. Work at academic research institutes in the US to explore the optimal computational use of the new data is therefore a vital component of the AGWG, and is being carried out by investigators with key expertise in several locations, including Rockefeller University, Yale University, the National Human Genome Research Institutes, the University of California San Francisco, Virginia Tech, and the Broad Institute. To use the puzzle metaphor a final time, this advance is equivalent to producing a robot that can scan a field of pieces and quickly find the missing piece to fill a gap.

The AGWG is still in the process of evaluating and optimizing these technologies. The group aims to produce a new genome map for *Aedes aegypti* by late summer of 2016. Finding and labeling all of the genes in the new map will take additional effort, and relating the function of those genes to insecticide resistance and other mosquito traits influencing disease spread will take years of work from dedicated members of the vector research community. This technologically opportune project to produce an improved DNA map will not only be a foundational investment in the control of Zika, however, but also Yellow Fever, Dengue, Chikungunya, and future emerging viral diseases likely to hitch a ride between hosts via *Aedes* mosquitoes.

CONCLUSION

The work of the AGWG represents both a triumph and a tragedy. It is a triumph because of the open and collaborative spirit of the endeavor, and the rapidity with which the community has responded to address the Zika epidemic. It is a tragedy because it took a public health emergency like the Zika epidemic for the research community to remedy the poor quality of a foundational resource for understanding disease transmission. Though the AGWG was founded without a solid base of funding, the maturation of its efforts will require focused investment of resources to ensure the quality and integrity of the new genome map it hopes to deliver. Innovative and effective disease control efforts that make use of the improved *Aedes* genome map are an exciting prospect, but will also require committed investment. **The Zika epidemic can become a proving ground for the power of new DNA-based epidemiological and intervention tools. We have the opportunity to demonstrate to the world how new technologies will let us understand, anticipate, and control the spread of an epidemic, and we have an obligation to vulnerable populations to seize this opportunity.**

REFERENCES

1. Nene, V. *et al.* Genome sequence of Aedes aegypti, a major arbovirus vector. *Science* **316,** 1718–1723 (2007).
2. Neafsey, D. E. *et al.* Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. *Science* **347,** 1258522 (2015).
3. McBride, C. S. *et al.* Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **515,** 222–227 (2014).
4. WHO | World Malaria Report 2015. *WHO* Available at: http://www.who.int/malaria/publications/world-malaria-report-2015/report/en/. (Accessed: 20th May 2016)
5. Ranson, H. & Lissenden, N. Insecticide Resistance in African Anopheles Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain Malaria Control. *Trends Parasitol.* **32,** 187–196 (2016).
6. Deming, R. *et al.* Spatial variation of insecticide resistance in the dengue vector Aedes aegypti presents unique vector control challenges. *Parasit. Vectors* **9,** 67 (2016).
7. Dusfour, I. *et al.* Multiple insecticide resistance in Aedes aegypti (Diptera: Culicidae) populations compromises the effectiveness of dengue vector control in French Guiana. *Mem. Inst. Oswaldo Cruz* **106,** 346–352 (2011).
8. O'Loughlin, S. M. *et al.* Genomic signatures of population decline in the malaria mosquito Anopheles gambiae. *Malar. J.* **15,** 182 (2016).
9. Lima, A., Lovin, D. D., Hickner, P. V. & Severson, D. W. Evidence for an Overwintering Population of Aedes aegypti in Capitol Hill Neighborhood, Washington, DC. *Am. J. Trop. Med. Hyg.* **94,** 231–235 (2016).
10. Gantz, V. M. & Bier, E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* **348,** 442–444 (2015).
11. Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E6736–6743 (2015).
12. Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat. Biotechnol.* **34,** 78–83 (2016).
13. Harmon, A. Team of Rival Scientists Comes Together to Fight Zika. *The New York Times* (2016).