Species are the fundamental units of biological diversity; therefore, understanding

speciation patterns and mechanisms is paramount to evolutionary biology. Speciation generates all the biodiversity on the planet, but what keeps them from merging all back together? Defining what a species is and what allows species to persist is a controversial topic. Commonly, a species is split into two when some geographic barrier presents itself, such as a mountain range, a river formation, or even the building of a railway This new barrier causes reduction of interbreeding between the two populations and over time two distinct groups develop. If the two populations are unable to interbreed when later brought together, this is referred to as allopatric speciation. However, what occurs when species that are not bound by geographically isolating barriers and share overlapping ranges? The conundrum of what maintains a species once they form is a popular research topic. It is widely considered that linkage disequilibrium (LD) is what keeps species distinct types, however what causes the LD to accrue in the genome2 .Does creating LD in ways besides reproductive isolation facilitate the persistence of species? What are the driving factors of speciation and the maintenance of species? The current state of biology makes for an interesting time. There is a lot of data out in the world and questions that can be answered the original researchers never even fathomed would be asked. Here I provide an attempt to answer an age old question in the field, using publically available data and new methodologies.

Old world monkeys are the most widely spread non-human primate3. They share a last

common ancestor with humans 25 million years ago (MYA)3. Macaques have 19 recognized

species comprising four groups (Sylvanus , Silenus , sinica and fascicularis), making them unique

in speciation studies. They provide not only vast biomedical functionality, but through their diversity are ecologically significant as well3. Previous genetic work in my focal species Macaca mulatta (CR) and M.fascicularis (CE) shows asymmetrical hybridization from M.mulatta into M. fascicularis4, with an estimated divergence time of 1-2.5 MYA6,7. Both species are broadly distributed over southeast Asia in a parapatric geographic distribution with adjoining ranges in Indochina The observation of hybridization between these two species conflicts with the simple model of allopatric speciation8. In most plants and animals, hybrid sterility acts as a post-fertilization barrier to hybridization5. As such, this system is of great importance to study, and can potentially shed light on the mechanism behind this.

). SV’s, such as an insertion, deletion, inversion, or translocation, are

common in. genomes and like single nucleotide polymorphisms can act as genetic markers.

While serving as genetic markers, SV’s have have also been implicated as contributing factors

to speciation10. Insertions or deletions may occur near promotors/enhancers/repressors

facilitating speciation or a translocation may occur causing genes to shuffle to another

chromosome and be under a different set of genomic controlling factors. While all structural

variants will be analised, previous work in the literature has highlighted the importance of

inversions in speciation, as such, they will be the initial focus of the project.

Inversions have been shown to capture locally adapted alleles when two populations

are hybridizing. They prevent single crossovers in the F1 hybrid generation, and as such, areas

of hybrid sterility should show less introgression and admixture of the two genomes. It is

possible that M. mulatta and M. fascicularis might exist as two distinct gene pools because

inversions have been the driving force of LD between them, allowing recombination to

homogenize these two species outside inversions. There is complete recombination

suppression up to 2.5mb outside inversion breakpoints12. As such, regions of high divergence

should match with regions of suppressed recombination, hinting at inverted regions of the

genome. Recombination rate is a parameter that varies both across the genome and among

species. It determines the genomic scale over which patterns of differentiation and admixture

may vary12. Inversions may facilitate speciation by protecting locally adapted alleles inside the

inversions from gene flow with other populations, allowing further divergence between

species. Creating LD in ways besides reproductive isolations (inversions reducing

recombination) facilitates species persistence. However, inversions are not a complete block, so

selection/local adaptation must also be implicated. Several gene categories might be linked to

contributing to speciation; genes related to sexual phenotypic characteristics, immunological

differences (such as MHC or RH factors) and ecological gene families. Converting genomic

patterns into useful data about speciation will not only take computational knowledge, but also

key facets of biological insight; like the dynamic nature of gene flow, selection acting on hybrid

populations, and recombination rate variation. Genomic studies of primate models can shed

light on the complex admixture that could have led to the formation of the chimeric human

genome.