Characterization of Vomeronasal Type-1 Receptor (V1R) Repertoire Diversity in Mouse Lemurs (Genus *Microcebus*)

by

Amy Michelle Kam May Heilman

Department of Biology
Duke University

Date:_____

Approved:

_____
Anne D. Yoder, Supervisor

_____
John Willis

_____
Robert Brandon

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Department of
Biology in the Graduate School
of Duke University

2013

ABSTRACT

Characterization of Vomeronasal Type-1 Receptor (V1R) Repertoire Diversity in Mouse Lemurs (Genus *Microcebus*)

by

Amy Michelle Kam May Heilman

Department of Biology
Duke University

Date:_____

Approved:

_____
Anne D. Yoder, Supervisor

_____
John Willis

_____
Robert Brandon

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Biology in the Graduate School of Duke University

2013

Abstract

An organism's ability to detect and interpret chemical signals from the surrounding environment is vital to its survival. In mammals, the vomeronasal organ provides an integral part of this requirement as it allows for the detection of inter-individual chemical signals. Located within the vomeronasal epithelium is a group of highly specialized receptors called vomeronasal type-1 receptors (V1Rs). These receptors bind pheromones and kairomones, which are chemical cues from conspecifics and heterospecifics, respectively. Recently, V1Rs have been recognized as highly variable across mammalian taxa. One extraordinary example is the mouse lemur, in which ~83% of the estimated 259 genes are expected to be intact. This relatively young and cryptically diverse prosimian group has a controversial history of species delimitation and it is possible that patterns in the rapidly evolving and highly diverse V1R gene family may provide insight into species boundaries. This study generated broad and deep coverage of the two most diverse V1R subfamilies across multiple *Microcebus* species. By utilizing Pacific Bioscience's single molecule real time sequencing technology, this study produced sequences at a fraction of the time and cost of the traditional method of Sanger-sequenced clones. The results advance our understanding of the high diversity–at least 5.6% and 9.9% sequence divergence intra- and inter-individually for V1R*IX* and V1R*strep*, respectively–and rapid evolution of V1Rs in mouse lemurs. Comparing these results with draft genome mined sequences demonstrates the limited utility of low coverage genomes for identifying variation in this gene family. The patterns uncovered here provide no clear evidence for the role of these two characterized subfamilies in maintaining or generating species boundaries, but may

instead point to an alternative mechanism of species recognition that is more complex and intricate than previously envisioned.

Contents

List of Tables

List of Figures

Introduction

Recent research has identified a remarkably diverse gene family responsible for detecting and interpreting chemical signals: the vomeronasal type-1 receptor (V1R) gene family. V1R genes have been found in a wide range of vertebrates, from lampreys to humans,[1] and there is a striking amount of variation in number of V1R gene copies, gene subfamilies and ratio of intact genes to pseudogenes across vertebrates as a whole. In particular, the primate order demonstrates the stark contrasts in V1R repertoire diversity. For example, although humans have 116 V1R loci, only 3 are potentially functional whereas mouse lemurs are estimated to have 259 loci, 214 of which are expected to be functional.[2]

The unusually high number of potential V1R loci combined with the high proportion of intact to pseudogenized copies (~83% putatively functional), makes mouse lemurs an intriguing system in which to study the patterns of V1R evolution. The benefit, however, is not to V1R research alone; discovering and recognizing the true level of mouse lemur species diversity has been a contentious issue over the last twenty years.[3] Studying V1R diversity may provide insight into the evolution of the reproductive isolating mechanisms contributing to the establishment or reinforcement of mouse lemur divergence events. Thus, capturing and characterizing mouse lemur V1R diversity may help to define species boundaries within the genus *Microcebus*. Two main lines of evidence reinforce this hypothesis: identification of a diverse V1R repertoire in *M.*

---

[1] Grus & Zhang, 2009
[2] Young et al., 2010
[3] Tattersall, 2007

murinus[4] and demonstration of chemical signals affecting mouse lemur behavior.[5] More broadly, simultaneously examining evolution at both the molecular and species level could provide key insights into the interaction of selection between these two levels of organization.

In this thesis, I will present background information, study aims, methods, and major results. Chapter 1 summarizes what is presently known about mouse lemurs, both phylogenetically and behaviorally, as well as V1R diversity and function in mammalian taxa. In reviewing previous studies focused on V1R function as well as the importance of chemosensory communication in mouse lemurs, I intend to present compelling evidence as to why characterizing V1R gene diversity will serve to advance our current understanding of the evolutionary history of one of the most specious lineages within the suborder Lemuriformes. Chapter 2 outlines the robust methods and materials by which I aim to intensively examine the V1R repertoire diversity in mouse lemurs. Chapter 3 reports the major findings, which will provide preliminary evidence motivating future studies.

[4] Young et al., 2010
[5] Perret, 1995

1. Background

1.1 Chemoreception

Chemoreception is the oldest and most widespread organismal sensory system.[1] This ability is vital, as organisms need to detect and interpret chemical cues from the surrounding environment in order to modulate appropriate behavioral responses, such as locating food and mates or avoiding predators and other dangers.[2] These chemical signals are called odorants. Typically, odorants are volatile chemical compounds that are less than 300 Daltons.[3] They vary in polarity and size, both of which determine their volatility and solubility. These two properties determine the medium of transmission, air or liquid, and molecular lifespan.[4] The total number of odorants in existence is unknown, but every living organism, from single-celled bacteria to plants to animals, releases chemical odorants.[5]

Odorants are categorized by function. Two of the most well studied types responsible for inter-individual interactions are kairomones and pheromones. Kairomones are interspecific signals, which benefit the receiver but not the transmitter, for example prey chemical cues attracting a predator or predator cues alarming prey.[6] Pheromones are intraspecific signals, which convey information such as individual, social, reproductive, and sexual status, which may trigger behavioral responses such as attraction, avoidance, or aggression.[7] These behaviors directly produce differential

---

[1] Niimura, 2013
[2] Ache & Young, 2005
[3] Touhara & Vosshall, 2009
[4] Brennan & Zufall, 2006
[5] Touhara & Vosshall, 2009
[6] Chamero, Leinders-Zufall & Zufall, 2012
[7] Johnston & delBarco-Trillo, 2009

survival and reproduction, both of which are necessary conditions for evolution via natural selection.[8]

In terrestrial mammals, chemical signals are detected by two functionally overlapping yet anatomically divided sensory organs: the main olfactory epithelium of the main olfactory system and the vomeronasal organ of the accessory olfactory system.[9] In the main olfactory system, stimuli are carried by the nasal airstream to the main olfactory epithelium, which is located within the nasal cavity and covered in ciliated olfactory sensory neurons.[10] The olfactory sensory neurons express olfactory receptors, which are primarily generalists being broadly tuned to a wide range of odorant molecules.[11] Furthermore as the concentration of stimuli increases, more receptors are activated.[12]

In the accessory olfactory system, the sensory apparatus is the vomeronasal organ, which is located at the base of the nasal septum in the nasopalatine canal.[13] Normally isolated from the regular airstream, stimuli only gain access via vascular pumping, which occurs in response to novel stimuli detected by the main olfactory system.[14] Vomeronasal sensory neurons (VSNs) express vomeronasal receptors that are highly sensitive (detecting concentrations as low as 10-11 Molar) and uniquely tuned to specific pheromones and kairomones.[15] Two different types are expressed in two distinct anatomical compartments of the vomeronasal epithelium: vomeronasal type-1 receptors (V1Rs) located in the apical layer and vomeronasal type-2 receptors (V2Rs)

---

[8] Darwin, 1859
[9] Brennan & Zufall, 2006
[10] ibid.
[11] Wyatt, 2003
[12] ibid.
[13] ibid.
[14] Keverne, 1999
[15] Wyatt, 2003

located in the basal layer.[16] The terminal microvilli of the VSNs bind the odorant and each neuron expresses a single, or at most a few, of the vomeronasal receptor genes from one randomly chosen allele.[17] The axons of the VSNs transmit the signal to the accessory olfactory bulb which then sends the signal to a specific vomeronasal nucleus in the amygdala, and finally, to the hypothalamus.[18] This system uses a completely separate pathway from that of the main olfactory system. Most strikingly, it completely bypasses higher cognitive centers, and instead triggers innate and stereotyped behaviors controlled by the limbic system.[19]

Both vomeronasal receptor families, as well as olfactory receptors, belong to the superfamily of G-protein coupled receptor proteins.[20] These proteins span seven transmembrane domains and couple with and activate G-proteins co-localized in the vomeronasal epithelium (Figure 1). Beyond sharing the general characteristics of the GPCR superfamily, both vomeronasal receptor families show very little homology, thereby suggesting independent evolutionary trajectories and pressures.[21] V1R genes are single exon coding, have short amino-terminal domains and greatest sequence variation occurs in their transmembrane domains.[22] In contrast, V2R genes have a complex intron/exon organization with multiple splicing variants, as well as a particularly long and highly variable extracellular amino-terminus.[23] Given the similarity of the long extracellular amino tail to glutamate and calcium-sensing receptors, it is predicted that

---

[16] Touhara & Vosshall, 2009
[17] Dulac & Axel, 1995
[18] ibid.
[19] ibid.
[20] Brennan & Zufall, 2006
[21] ibid.
[22] Keverne, 1999
[23] Touhara & Vosshall, 2009

V2Rs are responsible for detecting non-volatile proteins and peptides like major urinary proteins (MUP), major histocompatibility complex proteins (MHC), and exocrine gland peptide (EGP).[24]

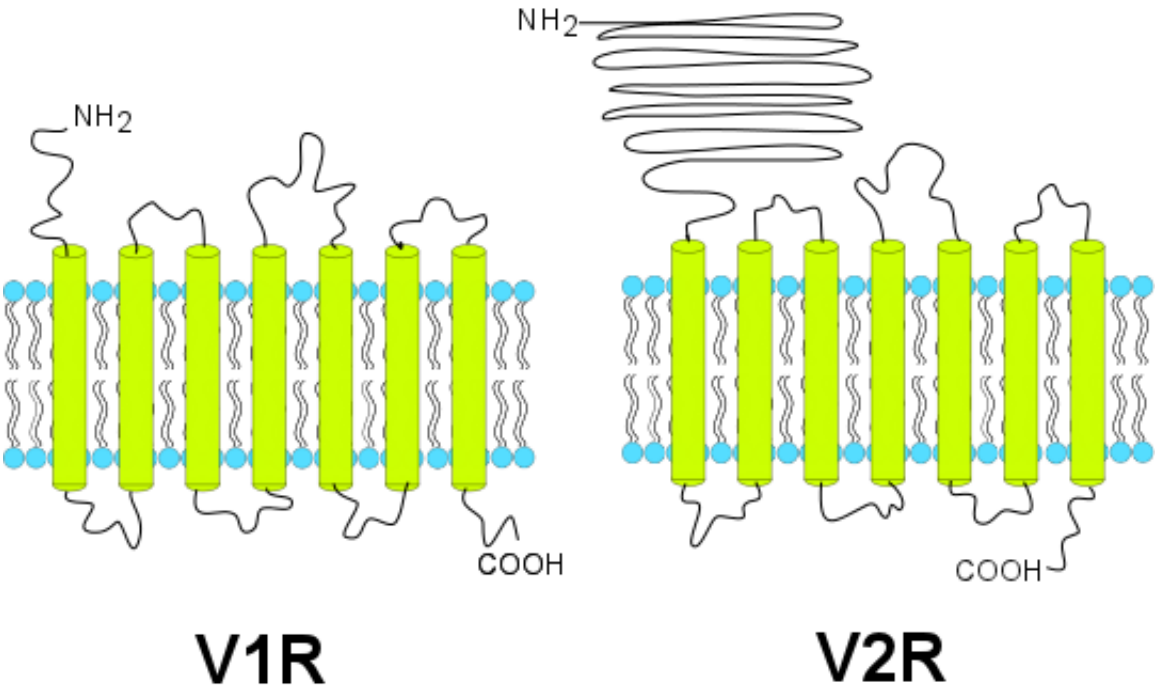

Figure 1: V1R structure and location in the cell membrane. Adapted from Mombaerts (2004).

These two gene families show remarkable diversity at all levels studied thus far (i.e. multiple subfamilies and ample inter- and intraspecific variation). Moreover, there is striking variation in the number of putatively functional, i.e. intact copies, versus the number of gene copies that have premature stop codons, non-sense mutations, frameshifts, and, in V2Rs, incomplete exon coding regions. The V1R genes are much easier to compare among different organisms given their short length (~900 bp) and their

[24] Rodriguez, 2008

single exon coding. As a result, more V1R sequence data has been generated from a variety of organisms, especially mammalian species.

1.2 Mammalian vomeronasal type-1 receptors

Though the anatomical components of the vomeronasal system are only present in tetrapods, V1R genes have been identified in teleosts, a cartilaginous fish, and a jawless vertebrate, therefore dating the origin of V1Rs to approximately 450MYA.[25] Previously, it was thought that the expansion of the V1R repertoire (as opposed to V2R) reflected the transition from land to water.[26] A recent study, however, demonstrated that terrestrial squamates have a much greater number of V2R genes than V1R genes.[27] This is in complete contrast to several terrestrial mammalian taxa possessing much greater V1R repertoire diversity than V2R diversity.[28] In fact, few V2Rs have been identified in mammals that are non-rodent.[29]

The most extensive work regarding characterization of V1R repertoires has been performed in the rodents *Mus musculus* and *Rattus norvegicus*. Over 300 V1R genes, which are distributed among 12 families, have been identified in the mouse genome (*M. musculus*).[30] Similarly, 219 V1R genes across 11 families have been found in the rat. Both rodents share 10 subfamilies, but there is one rat-specific lineage and two mouse-specific lineages.[31] In most shared subfamilies, genes form species-specific clades that

[25] Rodriguez, 2008
[26] Shi & Yang, 2007
[27] Bryzkczynska et al., 2013
[28] Young et al., 2010
[29] Grus et al., 2005
[30] Shi et al., 2005
[31] ibid.

may be the result of gene conversion or post-divergence duplications.[32] In fewer shared subfamilies, mouse and rat V1R genes are interspersed, suggesting that a relatively low number of gene duplications occurred prior to the mouse-rat split.[33]

Whether gene copies are interspersed between species or exclusively clustered within one species, V1R evolution is characterized by exceptionally fast gene-turnover, likely due to the gain and loss of individual genes.[34] Though approximately 86-94% of autosomal genes in rat have one-to-one orthologs in mouse, only 19% of rat V1R genes have one-to-one orthologs in mouse.[35] This has led to the hypothesis that V1R genes are responding to rapid changes in chemical signals such as pheromones and kairomones, which may be caused by neutral changes accumulating in chemical signals within isolated populations.[36] Moreover, the V1R adaptive response, at least across these two species, appears to be accomplished by tandem duplications, possibly mediated LINE repeats, which are thought to have been most active at the same time.[37]

The complete characterization of V1R genes in the mouse model has been complemented by molecular studies that demonstrate the behavioral and physiological function of V1Rs. Prior to the discovery of vomeronasal receptors, Clancy et al.[38] removed the VNO of virgin male mice, which disrupted sexual behaviors resulting in reduced frequency of matings. They also showed that lesions in the vomeronasal system severely reduced male-specific aggressive behaviors. After the discovery of

---

[32] Shi et al., 2005
[33] ibid.
[34] Grus & Zhang, 2004
[35] ibid.
[36] Shi et al., 2005
[37] Lane et al., 2002
[38] Clancy et al., 1984

vomeronasal receptors, Krieger et al.[39] differentiated the functional roles of V1R and V2R activation. Hypothesizing that structural differences between these two receptor gene families lead to differential binding affinities, they isolated various molecular components of male mouse urine. They discovered that only the G-protein Gαi (co-localized with V1Rs) was activated when exposed to lipophilic molecules, which were very likely volatile, and that only the G-protein Gαo (co-localized with V2Rs) was activated by α2u-globulin, a mouse urinary protein.[40] In order to test whether regulatory elements are responsible for discriminatory vomeronasal sensory neuron activation as opposed to diversity in the coding sequence, Stewart and Lane[41] interrogated the promoter regions for evidence as to how each sensory neuron is able to express only one allele of a single locus from the vast repertoire of V1Rs present in the mouse genome. They found that V1R subfamilies have broad and well-conserved promoter regions, which control transcription, and the mechanism of stochastic competition likely causes transcription factors to stably bind with only one gene at a time.[42] A more recent study by Isogai et al.[43] was able to elucidate the molecular mechanisms underlying the causal relationship between chemical signal exposure and stereotyped behavior. Their *in situ hybridization* experiments clearly showed the highly exclusive relationship between known stimulus and sensory neuron activation of specific V1R genes and V2R genes. By exposing a given mouse to bedding from ethologically relevant sources, they showed that V1Rs detect heterospecific information, such as those from congenerics

[39] Krieger et al., 1999
[40] ibid.
[41] Stewart & Lane, 2007
[42] ibid.
[43] Isogai et al., 2011

and predators, as well as the physiological status of a conspecific. [44] In other words, V1Rs can detect both pheromones and kairomones. These studies therefore provide irrefutable evidence for the highly specialized function of vomeronasal receptors, which had previously been based solely on correlations relating genotype to behavioral phenotype.

In addition to examining the physiological and organismal aspect of V1R, a number of researchers have explored the potential role of V1R genes in species divergence. Kurzweil et al.[45] investigated the hypothesis that V1Rs are related to species level divergence events by comparing two mouse lineages, *Mus spretus* and *Mus musculus*, which diverged only ~1MYA. They found that V1R orthologs were subject to diversifying selection and gene conversion. This result contrasts with the predominantly duplication-driven V1R repertoire differences between mouse and rat. [46] Karn et al.[47] considered the importance of V1Rs in assortative mating between two incipient species in secondary contact. They screened over 300 genes for evidence of high levels of divergence and found one candidate sub-species-specific V1R gene which appeared to segregate non-randomly with a proposed subspecies recognition pheromone, salivary androgen-binding protein (*abp*). Park et al.[48] examined a similar situation in a pair of closely related mouse species, *Mus musculus* and *Mus domesticus*, which form a narrow hybrid zone. They found evidence of purifying selection, though relatively weak. As a result, they proposed that neutral genetic drift creates new variants and purifying

[44] Isogai et al., 2011
[45] Kurzweil et al., 2009
[46] ibid.
[47] Karn et al., 2010
[48] Park et al., 2011

selection then reduces the number.[49] The studies presented here demonstrate the role of V1R diversity in speciation, but whether it establishes or reinforces divergence remains to be studied. Moreover, it appears that a number of diversifying mechanisms affect V1R diversity. It will therefore be important to explore a number of different organisms at different levels of phylogenetic relationships in order to find potential common V1R evolutionary patterns.

More recent research has questioned whether V1R repertoire diversity within and between mouse and rat is an anomaly or a widespread phenomenon. The broadest species comparison to date was conducted by Young et al.[50] who surveyed full genome assemblies, draft genome assemblies (2X and up), and trace archive data. Using robust methods, they were able to create confident estimates, though possibly underestimates, of V1R repertoires in species with only draft or Trace Archive data sets.

The result of their 37 mammalian species comparison revealed species-specific expansions across mammalian taxa, which resulted in "semi-private" V1R repertoires for the majority of mammals characterized.[51] Interestingly, approximately 80% of all the V1R genes found are most similar to V1R genes from the same genome rather than from any other included species genome or, in other words, paralogs are more similar than orthologs. Table 1 demonstrates the impressive range of V1R genes identified in a representative subset of mammalian species. The most prominent variations are seen within the primate order. The complete human genome contains 116 V1R-like copies, only 3% of which are potentially functional, while in complete contrast, the mouse lemur

[49] Park et al., 2011
[50] Young et al., 2010
[51] ibid.

(*Microcebus murinus*) is estimated to have at least 259 copies, 83% intact of which are expected to be intact. Other Old World and New World primates show similar levels of total V1R genes and putatively functional copies to human (Table 1). Moreover, the high number and widespread representation of V1R pseudogenes suggest that V1R loss may have begun in the common ancestor of Old and New World primates. In contrast, the tarsier and bushbaby have many V1R gene copies with a higher proportion intact, which is similar to, though not as high as, those belonging to the mouse lemur.

Table 1: Representative set of species V1R repertoire diversity from Young et al. (2010). Only mouse and human have complete genomes; other V1R numbers are adjusted based on the number of V1R genes mined and on the number of sequences available and coverage.

Species	Total V1Rs	Number of Subfamilies	Percent intact
Rattus norvegicus	219*	11	~49%
Mus musculus	392	12	~61%
Homo sapiens	116	?	~3%
Pan troglodytes	106*	?	~4%
Gorilla gorilla	115*	?	~3%
Microcebus murinus	259*	9	**~83%**
Otolemur garnetti	133*	?	~58%
Tarsius syrichta	266*	?	~15%

The high copy number and diversity of potential intact V1R genes in *Microcebus* has attracted recent attention. Hohenbrink et al.[52] detected positive selection as a diversifying driver in mouse lemur V1Rs across multiple levels of organization. Using the 107 gene copies identified in Young et al.,[53] they performed phylogenetic analyses to examine *M. murinus* V1R diversity. The phylogenetic reconstruction revealed nine monophyletic clusters whose membership varied from 4 to 29 gene copies. Additionally, they chose seven loci across the known V1R repertoire for targeted amplification and DNA sequencing in up to 10 mouse lemur species. Overall, they found strong evidence

[52] Hohenbrink et al., 2012
[53] Young et al., 2010

for positive selection across the *M. murinus* V1R repertoire and ongoing selection in loci amplified between species. There was substantial heterogeneity in the strength of selection for individual loci and clusters throughout the repertoire and between individual loci and clusters, however.[54] In regards to protein domains under selection, codons in the extracellular regions, specifically 2nd and 3rd were found to be under significant positive selection. They proposed that these domains are important for ligand binding. Moreover, Hohenbrink et al.[55] tentatively assigned functions from mouse V1R subfamilies to mouse lemur V1R clusters based on homology. Although this approach may be overly speculative, they hypothesized that certain clusters were associated with the detection of female, male, and heterospecific cues. From this study, it is clear that the evolution and divergence of mouse lemur V1Rs is quite complex as no single mechanism applies across species, individuals, gene subfamilies, or specific genes.

The results of the Hohenbrink et al.[56] study inspire intriguing questions. What degree of diversity is typical within and between multiple clusters? Are there differences in cluster composition within the genus *Microebus*? Are ecological and phylogenetic differences associated with V1R diversity and composition? A closer look at the phylogenetic relationships and social systems will provide clues as to the intricate and complex interaction between the molecular level (V1Rs) and species level (mouse lemurs).

---

[54] Hohenbrink et al., 2012
[55] ibid.
[56] ibid.

1.3 Mouse lemurs

Mouse lemurs (genus *Microcebus*) are a relatively young (~8.4 MYA)[57] and cryptically diverse group within the Family Cheirogaleidae.[58] These small-bodied and nocturnal prosimian species are nearly indistinguishable based on direct visual observation.[59] As such, only in the last decade and via the assessment of genetic divergence have we recognized their high level of diversity, increasing from only two species in 1994[60] to at least 19 currently.[61] However, despite the recognition of multiple species within this radiation, we know little about speciation processes within *Microcebus*. Mouse lemurs occur in every forest habitat found in Madagascar, including littoral, dry deciduous, spiny, transitional, gallery, and low montane rainforests.[62] Some of these species are wide-ranging, occurring across a number of ecological environments, while others are very restricted and prefer one forest type.[63] Moreover, a number of species are found in sympatry,[64] while others form hybrid zones in transitional forest.[65] It has also been found that putative biogeographic barriers separate closely related species.[66]

The present study focuses on a subset of the 19 species that are currently recognized: *Microcebus murinus*, *M. myoxinus*, *M. berthae,* and *M. rufus*. Though *M. myoxinus*, *M. berthae,* and *M. rufus* are always recovered as reciprocally monophyletic

---

[57] Yoder & Yang, 2004
[58] Radespiel et al., 2011
[59] Heckman et al., 2007
[60] Schmid & Kappeler, 1994
[61] Radespiel at al., 2011
[62] Radespiel et al., 2007
[63] Mittermeier et al., 2010
[64] Dammhahn & Kappeler, 2005
[65] Hapke et al., 2011
[66] Weisrock et al., 2010

in mitochondrial analyses,[67] they are never resolved as such with nuclear loci.[68] Instead, nuclear data indicate incomplete lineage sorting, thus making these species ideal candidates for species boundaries analysis.[69] Their species delimitation is strongly supported by their complete geographic separation as these species occupy distinct habitats separated by significant biogeographic barriers. *Microcebus murinus* inhabits the entire western coast forest and occurs sympatrically with both *M. myoxinus* north of the Tsiribihina River and *M. berthae* south of the river (Figure 2).[70] Alternatively, *M*. *rufus* occurs in the humid low montane rainforests of eastern Madagascar and is separated from the other three species by a formidable north-south mountain range, which accounts for the stark difference in rainfall characterizing the humid eastern and dry western halves of the island (Figure 2).[71] All phylogenetic studies to date, including those based on both mitochondrial [72] and nuclear [73] DNA sequence data, place *M. murinus* within a lineage sister to the remainder of the genus (Clade A; Figure 3). *Microcebus myoxinus*, *M. berthae,* and *M. rufus* form a monophyletic clade and are distantly related with respect to *M. murinus* (Clade B; Figure 3). The phenotypic variation across these species is minor. *Microcebus murinus*, commonly known as the grey mouse lemur, ranges between 58-67g in weight and has a brownish-greyish pelage with reddish tones.[74] At half the weight (~30g), *M. berthae*, or Madame Berthae's mouse lemur, is

[67] Heckman et al., 2007
[68] Weisrock et al., 2012
[69] Degnan & Rosenberg, 2009
[70] Mittermeier et al., 2010
[71] Weisrock et al., 2010
[72] ibid.
[73] Weisrock et al., 2012
[74] Rasoloarison, 2000

recognizable by its rufous coat.[75] The slightly larger (43-55g) pygmy mouse lemur (*M. myoxinus*) is also distinguished by its rufous-brown coat, however it has a distinct reddish-brown midline stripe.[76] Lastly, *M. rufus*, or the brown mouse lemur, maintains a midrange weight (40-48g) and mixes of reddish-brown on the head and limbs with a grayish-brown back and tail.[77]

[75] Dammhahn & Kappeler, 2008
[76] Mittermeier et al., 2010
[77] Atsalis, 2000

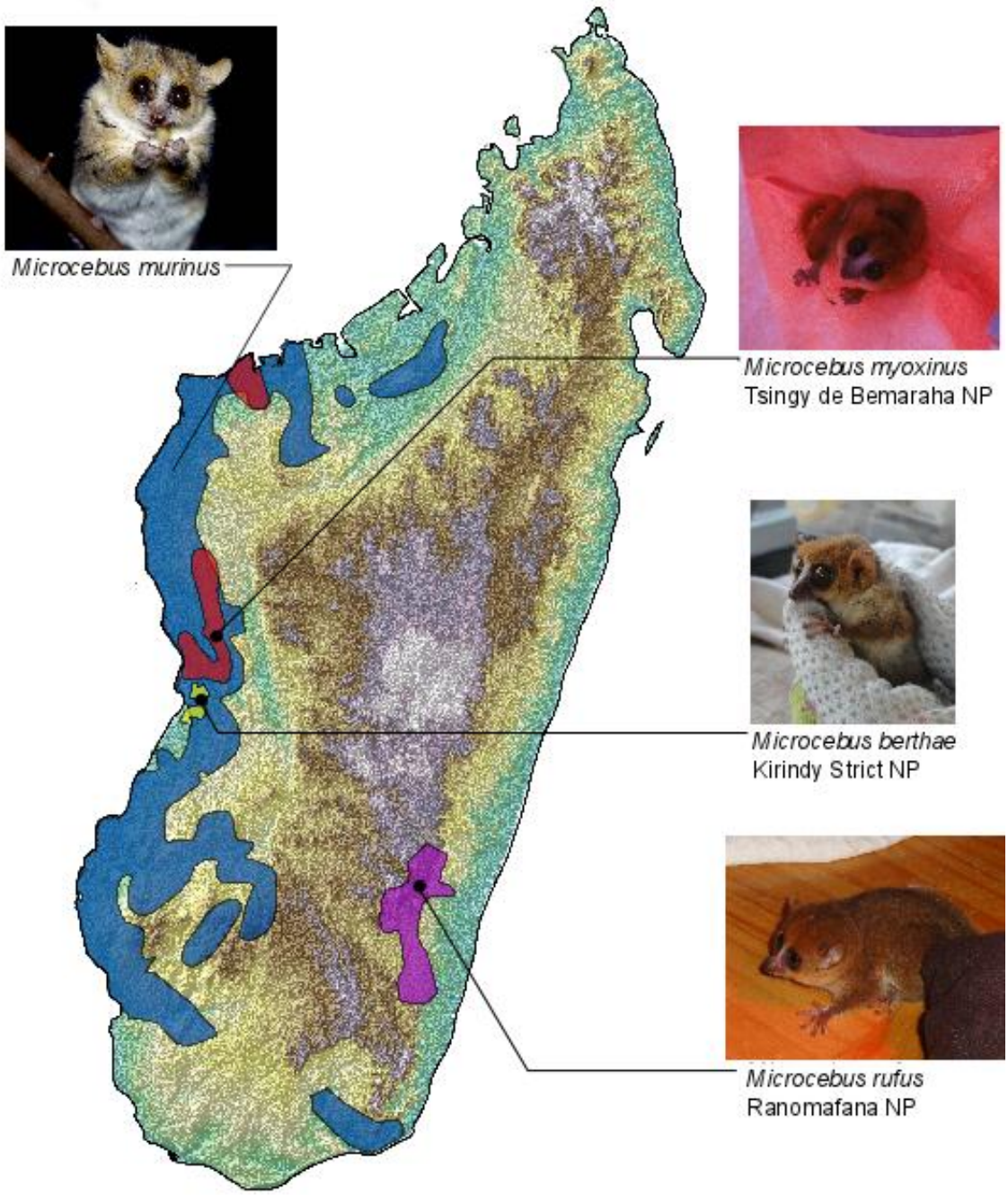

Figure 2: Physical map of Madagascar depicting IUCN ranges of each species in blue (*M. murinus*), red *(M. myoxinus*), yellow *(M. berthae*), and purple (*M. rufus*). Range for *M. rufus* was modified to reflect the range described in Mittermeier et al. (2010). Dots refer to the specific locations of sampled individuals for *M. myoxinus*, *M. berthae,* and *M. rufus*.

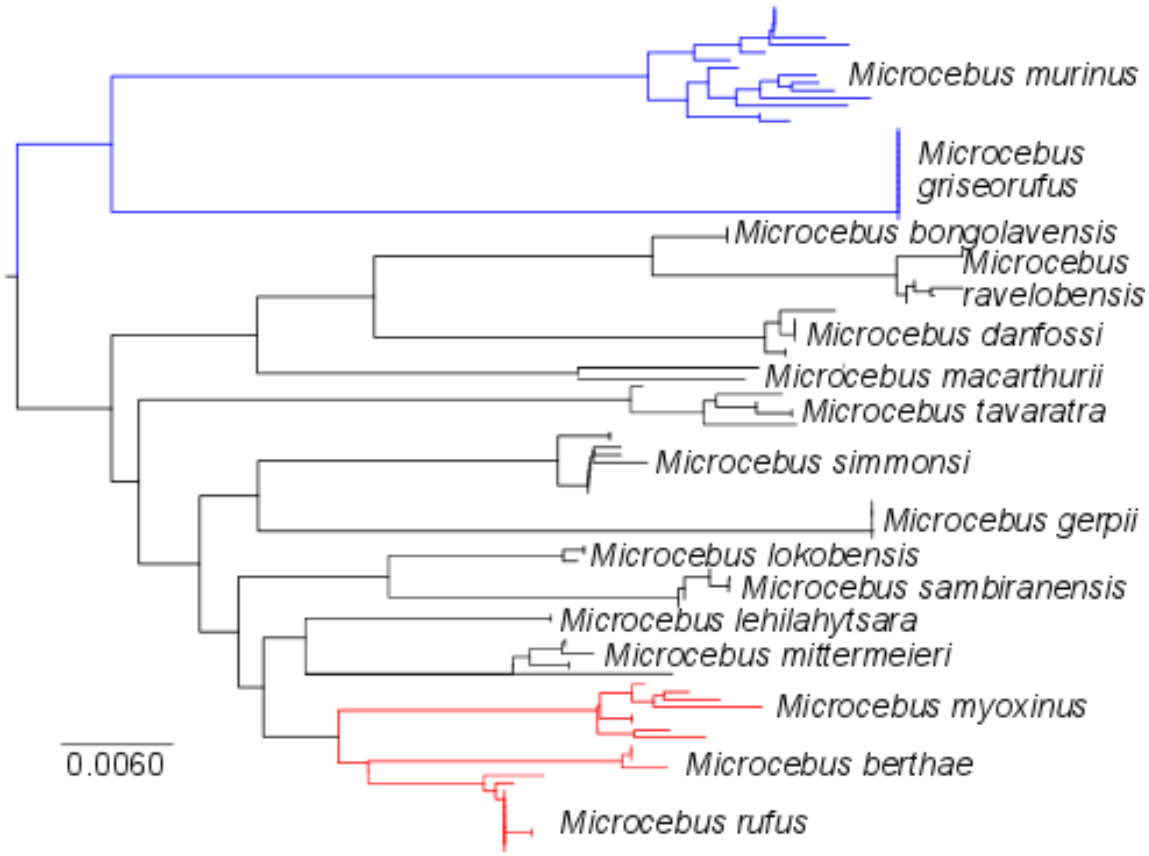

Figure 3: mtDNA gene tree based on COII sequence data gathered from GenBank. Blue branches identify species of "Clade A" while red branches identify species in "Clade B." Note, this is not meant to represent a species tree as the true phylogenetic ordering is still contentious. The important attribute to notice is the distant relationship between *M. murinus* and the monophyletic trio (*M. rufus, M. berthae,* and *M. myoxinus*).

Studies in the field and in the laboratory over the past 30 years have demonstrated the importance of chemosensory communication in mouse lemurs. As stated above, these species are nocturnal and, as such, chemical molecules likely provide more reliable signals than auditory cues or visual cues in their dark and dense

arboreal habitats.[78] Findings from both observational and experimental settings have demonstrated the importance and prevalence of chemosignalling in mouse lemurs.

In the wild, *Microcebus* individuals spend most of their nighttime activity foraging and traveling alone and, during the dry season (May-October), form sleeping groups with other conspecifics.[79] At the beginning of the wet season (November-April), which coincides with the start of the mating period, behavioral activity increases substantially.[80] There is also a shift in behavior most notable in males, as they establish home ranges to overlap with multiple females.[81] Mouse lemurs appear to have a promiscuous system whereby individuals attempt to overlap ranges with many potential mates.[82] In *M. berthae*, intense male-male competition has been observed as a result of this substantial overlap.[83] The timeframe for successful insemination is limited; the ovarian cycle in *M. murinus* is only 48-69 days with a fertile phase of only 2-4 hours.[84] Therefore, it is very important that females are able to advertise their reproductive status and location to males. Blanco & Meyer[85] analyzed fecal samples and noticed an increase in estradiol hormone leading up to estrus and a decline after. Perret[86] found a similar pattern in urinary estradiol. Additionally, *M. murinus* females are known to increase anogenital-marking behavior primarily on the day of estrus.[87] These olfactory cues may be

---

[78] Perret, 1995
[79] Glatston, 1983
[80] ibid.
[81] Perret, 1995
[82] Kappeler & Rasoloarison, 2003
[83] Dammhahn & Kappeler, 2005
[84] Bueschling et al., 1998
[85] Blanco & Meyer, 2009
[86] Perret, 2005
[87] Bueschiling et al., 1998

important in “delayed communication” as well as providing a trail for a male to find a receptive female.[88]

Captive studies have further demonstrated the importance of this critical period for reproduction as individuals must find and attract mates efficiently. A number of studies suggest that this necessity is mediated by chemical signaling, particularly by behaviors such as mouth-wiping, anogenital rubbing and urine-washing.[89] By manipulating photoperiod in the lab to match that of the wild to induce mating behaviors, Glatston[62] observed that dominant *M. murinus* males performed marking behaviors much more frequently than subordinate males, especially when in multi-male groups with females present. In females, he found that increased mouth-wiping and anogenital rubbing behaviors in females coincided with increased sexual receptivity.[90] Another study by Schilling and Perret[91] isolated urine from *M. murinus* males and determined that a chemical factor in the lipid portion of urine in dominant males depressed subordinate male testosterone levels. Interestingly, Perret[92] demonstrated two contrasting behavioral responses to chemical cues. First, she observed the intersexual stimulatory effects of male urine inducing synchrony of female estrous. Second, she identified the intrasexual inhibitory effects of female urine lengthening estrous and increasing proportion of male progeny produced. Two studies focused on the anatomical components of the vomeronasal system. The first study by Aujard[93] showed that the removal of the VNO from sexually experienced male *M. murinus* severely reduced the frequency of important

---

[88] Bueschiling et al., 1998
[89] Glatston, 1983
[90] ibid.
[91] Schilling & Perret, 1987
[92] Perret, 1995
[93] Aujard, 1997

precopulatory behaviors, such as anogenital investigation, mounting, and inter-male aggression. The second study by Araújo[94] found that *M. murinus* males with intact VNO were able to discriminate female chemical cues from control substances, males in which the VNO had been removed could discriminate using the main olfactory system but were less efficient, and bulbectomized males were unable to discern olfactory cues all together.

Another vital function of the chemosensory system is kairomone (or heterospecific cue) detection. Given their small body size, *Microcebus* species are susceptible to high predation pressure.[95] In a captive study, Sündermann[96] exposed individuals of predator-naïve *M. murinus* (i.e. those born and kept in captivity) to feces from endemic Malagasy predators, introduced predators, and endemic non-predators and found that individuals were able to discriminate between predator and non-predator. In a wild study, two sympatric species, *M. murinus* and *M. ravelobensis*, displayed anti-predator responses to feces from fossa but not barn owl, brown lemur or sifaka.[97] These results support an innate mechanism (i.e. genetic) for predator recognition in the olfactory system, which may be based on sulfurous metabolites produced during digestion of meat as opposed to experientially learned behaviors or recognition of specific predators.[98] Indeed, it would be evolutionarily advantageous to recognize chemicals associated with carnivores innately as the opportunities to learn and still survive are limited.

---

[94] Araújo, 2003
[95] Goodman, 2003
[96] Sünderman, 2008
[97] Kappel et al., 2011
[98] Sünderman, 2008

The demonstrated importance of chemoreception in altering intersexual, intrasexual, and heterospecific response behaviors in *Microcebus* species, coupled with the extraordinary size of the known and predicted intact *Microcebus* V1R repertoire, provides convincing evidence for the potentially central role of V1R gene diversity in survival, reproductive success, and speciation within *Microcebus*. The clear similarities and differences in the life histories of these four species in relation to one another fall within the range of what is generally understood for all characterized mouse lemurs. As such, this subset provides an appropriate example of what we might observe across the entire genus. Moreover, patterns that relate to ecological similarities and differences between given pairs or triplets of these species may be reflected in chemoreceptor genes. The specific reporting of chemosignalling and presence of chemical cues in bodily secretions provides strong evidence for the role of chemoreception mediating reproduction, which may have played a role in the divergence of allopatric species as well as the current role of assortative mating in sympatric species. Therefore it is possible that patterns in V1R repertoire diversity, especially given the rapid rate of evolution within this gene family, could provide an opportunity to detect potential species differences, which may have reproductive, behavioral, and ecological implications.

1.4 Objectives

Though an *in situ hybridization* experiment clearly demonstrating a one-to-one relationship between ligand, receptor, and neuron (like Isogai et al.[99]) would be ideal, these studies have yet to be designed due to the ethical and animal care concerns relating to terminal experimentation in mouse lemurs. Thus, the molecular evolutionary

[99] Isogai et al., 2011

approach taken here is fundamental to understanding the functional and evolutionary patterns of V1Rs in mouse lemurs. A study thoroughly interrogating the two most diverse subfamilies will help bring us closer to recognizing the entire repertoire. Moreover, a comparison between multiple mouse lemur species may reveal patterns of private, or species-specific, gene lineages, which would indicate the potential importance of V1Rs in maintaining species boundaries. Overall, a comprehensive sampling of diversity will provide foundational data upon which further studies can build and advance our appreciation of both V1R and mouse lemur evolution in concert.

The aim of this study is to advance our understanding of the complex and potentially informative patterns of V1R subfamily diversity in multiple species of mouse lemur. Broadly, I seek to extensively survey the two most diverse V1R subfamilies in order to quantify the level of variation of V1R genes in these subfamilies. By characterizing multiple individuals from multiple mouse lemur species, both closely and distantly related, I seek to compare patterns associated with different degrees of relatedness (within and between individuals, species, and clades). This investigation may reveal species-specific gene copies or gene lineages, which may in turn indicate V1R subfamily function delimiting species boundaries within *Microcebus*. In order to generate the data necessary for this study cheaply and efficiently, I employ the Pacific Biosciences SMRT cell CCS technology to broadly and deeply interrogate this multi-gene family. I validate my use of this method by directly comparing sequences captured with the SMRT cell CCS method to those data obtained by the tried-and-true Sanger-sequenced clone method and draft genome mined sequences.

2. Materials and methods

Sequencing and fully describing the variation of large, closely-related gene families is difficult because of the potential for extensive gene duplication and high levels of sequence identity between individual genes. This difficulty can be reduced with access to a high quality reference genome. Even so, the potential for copy number variation between conspecific individuals remains problematic. Knowing this, surveys of V1R gene variation may fail to confidently identify the exact number of V1R genes present within an individual.[1] Vector cloning, and subsequent Sanger sequencing, of PCR amplicons have proven useful for characterizing gene diversity of large and closely related gene families.[2] Unfortunately, this approach requires rigorous and time-consuming laboratory preparation and expense, making it cost and time prohibitive.

In order to scrutinize the diversity of the V1R repertoire and identify potential species-specific V1R genes in *Microcebus*, I implemented an efficient and cost-effective method for gene capture and sequencing: single molecule real-time (SMRT) circular consensus sequencing (CCS) using the Pacific Biosciences (PacBio) *RS* instrument. This technology is capable of sequencing thousands of target sequences that can vary in length from ~300 bp to ~6,000 bp, which can be simultaneously detected in a relatively short time span.[3] SMRT CCS technology provides a minimum of 3X coverage per 1000 bp per sequencing read, and is therefore ideally suited to rare variant detection.[4]

---

[1] Young et al., 2010
[2] Steidl et al., 2012
[3] Eid et al., 2009
[4] Travers et al., 2010

These features make SMRT CCS technology not only appropriate, but ideal for the present study because V1R genes are approximately 900 bp in length.[5] Moreover, recent upgrades to PacBio *RS* chemistry have increased sequencing accuracy, thus allowing for multiple amplicons to be sequenced in one run.[6] This advancement makes it possible to sequence the V1R repertoires of multiple individuals at once. In addition, known barcode sequences can be attached to the end of a target sequence to allow for downstream matching of a particular amplicon to a particular individual. Herein, V1R amplicons from five individuals, two *M. murinus*, one *M. myoxinus,* one *M. berthae,* and one *M. rufus*, were pooled. As part of my experimental design for data validation, individuals for which V1R sequence data have previously been generated were included: (1) a *M. murinus* individual wherein a V1R subfamily was targeted using traditional cloning and Sanger sequencing[7] and (2) the *M. murinus* used for the 2X draft genome (from which previous studies have mined V1R genes[8]). Including these latter samples will broaden existing data sets thereby furthering our understanding of V1R gene family repertoire diversity.

2.1 Data collection

Wild-caught individuals of *M. berthae* (10 individuals), *M. myoxinus* (10 individuals), and *M. rufus* (10 individuals) were sampled during an expedition to Madagascar in September 2011 (Appendix A). Sampling was conducted at the following localities: Tsingy de Bemaraha (*M*. *myoxinus*), Kirindy (*M. berthae*), and Ranomafana

[5] Dulac & Axel, 1995
[6] http://www.pacificbiosciences.com/, accessed 28 February 2013
[7] Chan et al., in prep
[8] Young et al., 2010

(*M. rufus*) (Figure 2). Hair samples were collected and preserved in 2mL tubes with RNAlater. When possible, microchips were used to identify and prevent resampling. All sampling methods were approved by the Duke University IACUC (registry number A042-11-02 and the CITES permit is No. 11US43685A/9). A liver sample from a captive-bred *M. murinus* individual (DLC 7013) was provided by the Duke Lemur Center. Dr. Jeff Rodgers (Baylor School of Medicine) provided whole genomic DNA of the *M. murinus* individual used for the 2X draft genome (herein referred to as Mmur2X).[9]

Genetic material was extracted from hair samples belonging to three wild-caught individuals, Mmyo020 (*M. myoxinus*), Mber040 (*M. berthae*), and Mruf060 (*M. rufus*), using a modified QIAamp DNA Mini Kit protocol (QIAGEN). The following modifications were made to the first two steps of the QIAamp protocol: ~100 hair strands, (see Appendix B for justification of number of strands), were placed in a 2mL microtube with 180µL ATL buffer and 20µL DTT (1M).[10] The recommended amount of Proteinase K was added to the mixture and samples were incubated overnight at 56C. The remaining steps followed the protocol. DNA was extracted from the DLC7013 (*M. murinus*) tissue sample using the standard protocol provided with the QIAgen DNeasy Tissue and Blood kit (QIAGEN).

This study focused on two of the most diverse gene subfamilies, which are known to have at least 29 (Cluster I Hohenbrink et al. 2012 or V1R*strep* sensu Chan et al. in prep) and 19 (Cluster IX Hohenbrink et al. 2012; herein referred to as V1R*IX*) copies each. V1R degenerate primers were designed by aligning sequences identified in

[9] Anne Yoder, pers. comm.
[10] Tara McDaneld, pers. comm.

Young et al.[11] and by grouping copies into the subfamilies identified by Hohenbrink et al.[12] Conserved regions coinciding with 1st and 7th transmembrane protein domains, as predicted by HMMTOP software,[13] were selected for primer design using the Primer3 plug-in within Geneious software (version 6.0.6; Biomatters Ltd.). Additionally, degenerate primers designed by Chan et al.,[14] mapping to the 2nd and 7th transmembranes, were utilized for comparison. Five forward barcodes designed by Pacific Biosciences were incorporated onto the 5' end of the forward primers; this allowed for multiplexing of 5 different individuals (Appendix C).

Amplicons were obtained using High Fidelity Taq DNA Polymerase (Sigma-Aldrich). PCR reactions were conducted in 50µL reactions with the following final concentrations: 1X High Fidelity buffer, 2mM MgCl2, 200µM each dNTP, 0.8µM primers, 0.625 unit Taq, and ~15ng DNA template. The following touchdown thermocycler profile was used for all sequences: initial denaturation 95C for 3 min followed by 15 cycles of 95C for 1 min, 60C (1C decrease per cycle) for 1 min, 72C for 1 min 30 sec, then another 20 cycles of 95C for 1 min, 45C for 1 min, 72C for 1 min 30 sec, and a final extension of 72C for 10 min.

The PCR products were then visualized on a 2% agarose gel with SYBR Green I Nucleic Acid Gel Stain (1 µL per 10mL of 1XTAE; Lonza). Bands of the expected size (~800bp) were extracted using UltraClean GelSpin DNA Extraction Kit (MoBio Laboratories). Two PCR amplifications were performed for each primer pair targeting specific subfamilies for each individual. The duplicate amplifications were then pooled

[11] Young et al., 2010
[12] Hohenbrink et al., 2010
[13] Tusnády & Simon, 2001
[14] Chan et al., in prep

together. The paired PCR reaction samples were then purified and concentrated using the QIAquick PCR Purification Kit (QIAGEN). The concentration of the resulting elution was quantified using a NanoDrop spectrophotometer (Thermo Scientific). A total of 16 different purified PCR products were multiplexed in equal nanogram proportions. This pooled sample was then submitted to Duke University's IGSP Genome Sequencing & Analysis Core Facility for AMPure bead cleanup, library preparation, and PacBio RS sequencing across 2 SMRT cells using 120 minute movie lengths.

2.2 Data analysis

CCS reads were first subjected to quality control using the Galaxy platform.[15] Reads between 700–900bp in length, which had a minimum PHRED 15 score for at least 80% of base pairs, were retained. These filtered sequences were then split by barcode using the FASTX toolkit barcode splitter (http://hannonlab.cshl.edu/fastx_toolkit/), creating individual data files corresponding to each *Microcebus* individual. Files were then imported into Geneious for alignment and manual trimming of the barcode. These trimmed sequences were then clustered based on 99% similarity[16] using the CD-HIT-EST webserver.[17] Only clusters with at least four individual reads (i.e. 4X coverage) were kept and sequences were retrieved and aligned with Mafft[18] using a custom python script designed by Jose Luis Villanueva-Caña (University of Barcelona).

---

[15] Goecks et al., 2010
[16] I decided to use 99% instead of 98% similarity threshold given that the majority of errors should have been eliminated by the highly stringent parameters in other filtering steps.
[17] Li & Godzik, 2006
[18] Katoh, 2005

To visualize patterns in *Microcebus* V1R repertoire diversity, consensus sequences were manually called for each of the 99% similar CCS read clusters in Geneious. Within the 99% similar clusters, all SNPs with three or more identical copies were differentiated. Given the depth and breadth of coverage achieved by this study, it is likely that the variation captured includes orthologs, paralogs, recent duplications, and heterozygous alleles, but without sequence from noncoding flanking regions to differentiate loci, it is difficult to distinguish between different gene types with high confidence. Therefore distinct sequences identified in this study are referred to as "gene copies." For comparison of repertoire diversity, consensus sequences for the V1R*IX* cluster were aligned by subfamily and a NJ tree was created in Geneious to visualize the relationships. For V1R*strep* repertoire diversity, the five individuals in this study were aligned with additional wild-caught *Microcebus* individuals (*M. griseorufus, M. simmonsi*, and *M. murinus*) from Chan et al.[19] Genetic distances were computed using the F84+Gamma substitution model with the alpha shape parameter fixed to 0.5 in the BASEML implementation of PAML.[20]

Gene copies generated for Mmur2X and DLC7013 in this study were aligned with 2X draft reference gene copies and Sanger sequenced clones, respectively. In doing so, I was able to provide a more inclusive depiction of total V1R subfamily repertoire diversity for two *Microcebus* individuals.

[19] Chan et al., in prep
[20] Yang, 2007

3. V1R subfamily repertoire diversity in mouse lemurs

3.1 Quality filtering

In the first step of quality filtering, ~86% of the CCS sequences were within our quality standards (Figure 4). After separating the reads by barcode, there were between 3,000-6,000 filtered reads per individual (including both subfamilies). Clustering at 99% similarity produced between 79-144 consensus sequences for each individual. To differentiate intact copies from pseudogenes, all gene copies were translated in Geneious and those without premature stop codons were considered intact. The results of these steps revealed between 25 and 40 intact V1R*IX* genes and between 42 and 77 intact V1R*strep* genes for each individual (Table 2).

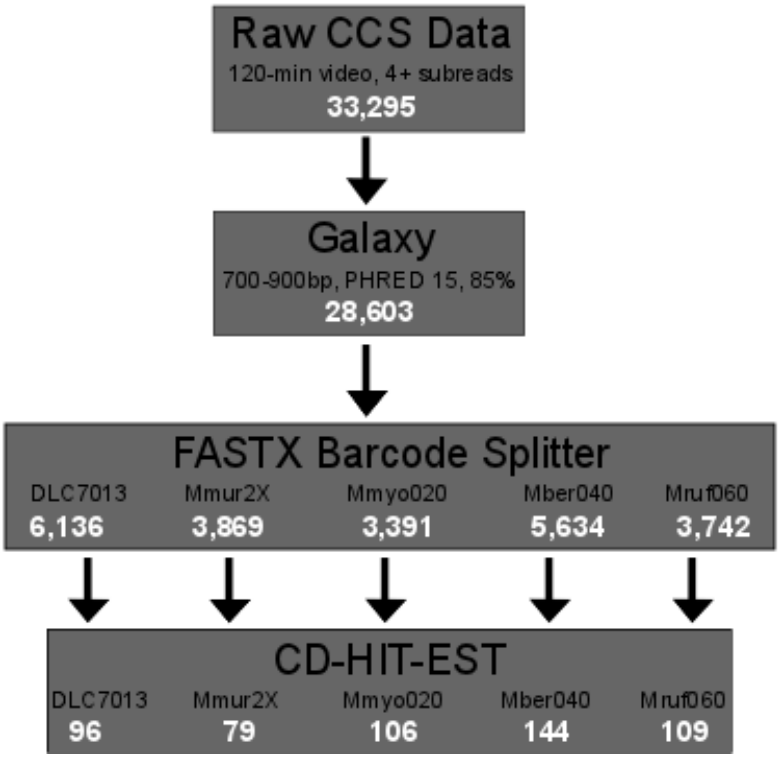

Figure 4: Flowchart of data filtering programs, filtering parameters and remaining reads.

Table 2: Gene copies per subfamily per individual. Numbers outside parentheses represent sequences with open reading frames whereas numbers inside the parentheses are pseudogenes

Species	Individual	V1R*IX*	V1R*strep*
Microcebus murinus	DLC7013	28(1)	57(10)
Microcebus murinus	Mmur2X	34(0)	42(3)
Microcebus myoxinus	Mmyo020	35(0)	66(5)
Microcebus berthae	Mber030	40(0)	77(27)
Microcebus rufus	Mruf060	25(2)	76(6)

*3.2 V1R*IX *repertoire diversity between four species*

All intact gene copies from the five different individuals were aligned. From the alignment, a neighbor-joining tree was constructed to visualize patterns (Figure 5; Appendix D includes bootstrap support). Both individuals of *M. murinus* are colored in blue while the three closely related species are colored in orange (*M. berthae*), red (*M. myoxinus*), and brown (*M. rufus*). Black internodes represent different species at connecting terminal nodes.

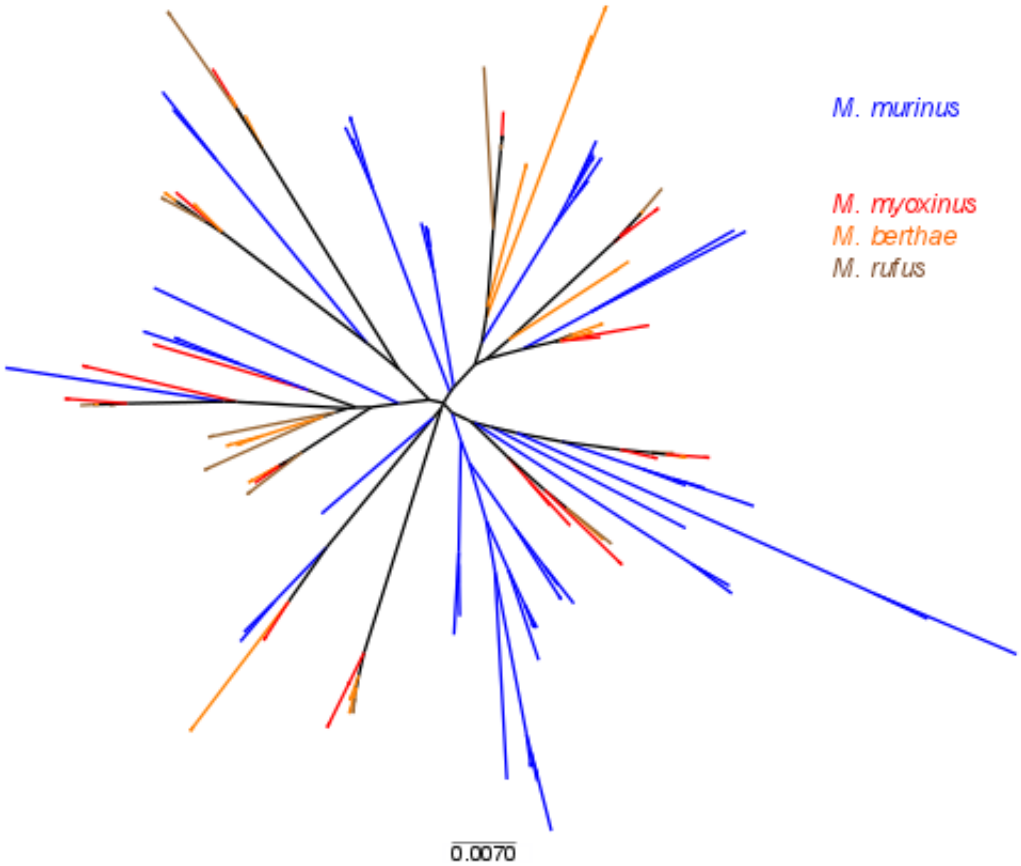

Figure 5: Unrooted phylogram for V1R*IX* for five individuals belonging to four *Microcebus* species.

I identified a total of 165 *V1RIX* gene copies across five individuals belonging to four species: *M. murinus* (29 and 34), *M. myoxinus* (35), *M. berthae* (40), and *M. rufus* (27). Approximately 98% of these gene copies were intact and, given the sequence similarity to known functional V1R copies in rodents, are likely functional.[1] The unrooted phylogram revealed patterns of intermingled diversity between gene copies belonging to different species (Figure 5). Of the four species, only *M. murinus* appeared to have a few private lineages, one of which appeared to include nine gene copies (lower middle blue lineage in Figure 5). Most gene copies belonging to *M. myoxinus*, *M. berthae*, and *M.*

[1] Young et al., 2010

rufus shared lineages with gene copies from at least one of the other two species. Some of these lineages were Clade B-specific with more basal *M. murinus* gene copies, while others were intermingled with *M. murinus* gene copies. These patterns suggest some bias toward clade-specific lineages. This provides minor evidence for a link between degree of phylogenetic relationship and V1R gene sequence similarity as more closely related species share more closely related gene copies. *Overall, the lack of species-specific lineages provides no evidence for the role of V1RIX genes in maintaining species boundaries.* Considered another way, this result may suggest an alternative mechanism by which V1Rs maintain premating barriers; instead of particular gene copies, it may be that suites of V1R gene copies are responsible for detecting conspecifics. Future resources such as the complete genome for *M. murinus* will help to clarify whether this combinatorial mechanism is responsible for maintaining species boundaries.

The genetic distance values were computed for V1R*IX* within individuals and within *M. murinus* (for which I have two samples) using gamma-corrected F84 model in PAML (Table 3). All intra-individual sequence divergence values ranged between 5.6% and 6.8%. The percent of sequence divergence between the two *M. murinus* individuals was 6.5%.

Table 3: Genetic distance values for V1R*IX* repertoire diversity intra-individually (*M. murinus*, *M. myoxinus, M. berthae,* and *M. rufus)* and inter-individually (*M. murinus).*

Species	ID	Number of Sequences	Nucleotide distance (F84+G)
1) *Microcebus murinus*		62	0.065
a) *Microcebus murinus1*	MmurDLC7013	28	0.068
b) *Microcebus murinus2*	Mmur2X	34	0.064
3) *Microcebus myoxinus*	Mmyo020	35	0.056
4) *Microcebus berthae*	Mber030	40	0.061
5) *Microcebus rufus*	Mruf060	25	0.061

The genetic distance calculations revealed common intra-individual sequence divergences (average of 6.2%) for all four species. By analyzing a data set combining two *M. murinus* individuals, I was able to compare intra- and inter-individual variation for this species (Table 3). The intra-individual genetic distances for both *M. murinus* individuals, DLC7013 and Mmur2X, were 6.8% and 6.4%, respectively, and the inter-individual genetic distance was 6.5% (Table 3). The similar levels of sequence divergence between different species, both intra- and inter-individually, are remarkable. Also interesting was the similarity of intra-individual sequence divergence between the wild caught individuals and both captive-bred individuals; the captive-bred individuals may share some ancestry given that they were previously members of the Duke Lemur Center, though decades apart. This result substantiated the value of including captive mouse lemurs in this and additional V1R *Microcebus* comparisons. The overall similarity and level of the intra-individual genetic distances for the other three *Microcebus* species

corresponded to previously reported variation in rhesus macaque MHC class I (~2-6% sequence divergence).[2]

*3.3 V1R*strep *repertoire diversity between six species*

Eight individuals from six species were included for the V1R*strep* diversity analyses. Chan et al.[3] characterized two species not included in this study (*M. griseorufus* and *M. simmonsi*) as well as an additional *M. murinus*, which was wild-caught. Phylogenetically, *M. griseorufus* is sister to *M. murinus* and both comprise Clade A, whereas the placement of *M. simmonsi* is basal to Clade B (Figure 3). Intact gene copies from these eight individuals were aligned and a neighbor joining tree revealed two distinct lineages within V1Rstrep (Figure 6). Due to the high similarity of the gene copies in this family, short branch lengths created polytomies and reduced bootstrap support (data not shown). Chan et al.[4] observed this general pattern of *V1Rstrep* diversity within a broader survey of lemurs, and were able to recover bootstrap support for the two distinct lineages by aligning multiple taxa across the cheirogaleid family (*Cheirogaleus major*, *C. minor*, *Allocebus trichotis*, and *Phaner pallescens)*. The colors used for species from this study are consistent with the V1R*IX* unrooted phylogram while the two new species are colored with cyan (*M. griseorufus*), and green (*M. simmonsi*).

[2] Daza-Vamenta et al., 2004
[3] Chan et al., in prep
[4] ibid.

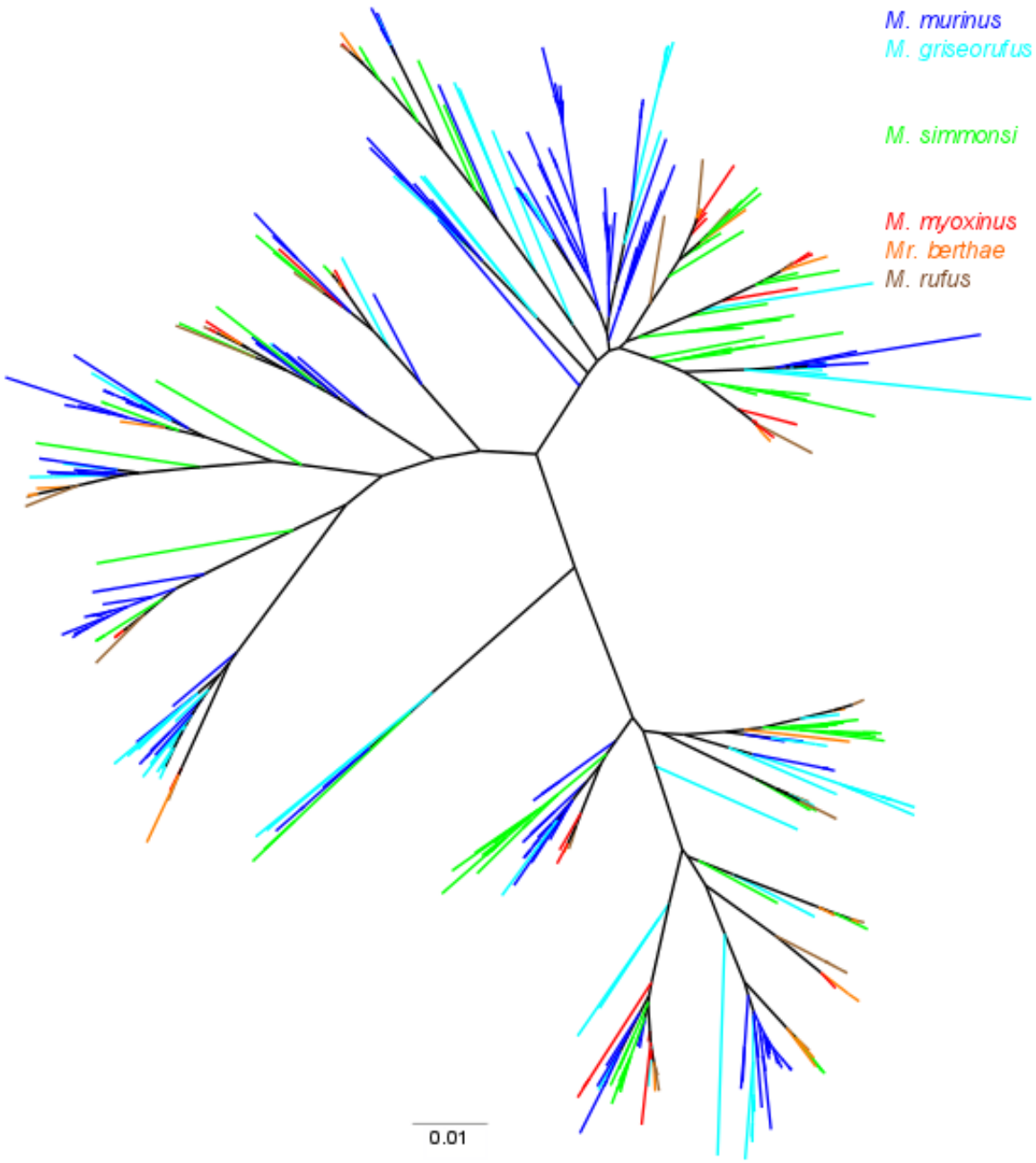

Figure 6: Unrooted phylogram for V1R*strep* for eight individuals belonging to six *Microcebus* species.

For V1R*strep*, a total of 369 gene copies, 86% of which were potentially functional, were identified across the four species: *M. murinus* (112), *M. myoxinus* (71),

M. berthae (104), and *M. rufus* (82). Similar to the V1R*IX* subfamily, many of the gene copies were interspersed among all six species included in the alignment and tree construction (Figure 6). There were a few Clade A-specific lineages, but no single species-specific lineages. Generally, gene copies from both members of Clade A appeared together more often than they appeared apart (separated by gene copies from outside the clade). For Clade B, there appeared to be only one clade-specific lineage (lower right-hand corner in Figure 6), which in fact contained all three species. The majority of Clade B clusters were intermingled with either or both Clade A gene copies and gene copies belonging to *M. simmonsi*. Though no species-specific lineages were detected for this subfamily, there was some clade membership bias; however, like V1R*IX*, there was little evidence for clear V1R*strep* divergence between Clade A and Clade B. Additionally, gene copies belonging to *M. simmonsi*, which phylogenetically lies between the two clades (Figure 2), emerged within lineages for both clades more or less equally.

Unlike the V1R*IX* subfamily, two distinct lineages emerged in the neighbor-joining reconstruction of V1R*strep* gene copies (upper pair of branches in contrast to lower set of branches in Figure 6). Chan et al.[5] also observed two distinct lineages within V1R*strep* for all lemurs included in their study, which they named V1R*strep*-α and V1R*strep*-β. Merging gene copies from this study with other cheirogaleid genera reinforced the findings of both Chan et al.[6] and this study (Appendix E). The presence of these two lineages suggests that there was an ancient duplication event prior to the divergence of cheirogaleids. Interestingly, these two lineages have not only survived

[5] Chan et al., in prep
[6] ibid.

through time but exhibit a high degree of diversification in each species. This rapid expansion of both α- and β-lineages in V1R*strep* as well as the current evidence suggesting that this is a strepsirrhine-specific subfamily[7] advocate the potential importance of this V1R subfamily in the evolutionary history of cheirogaleids.

Intra-individual genetic distances for all individuals and inter-individual genetic distance for the three individuals of *M. murinus* were computed using gamma corrected F84 model in PAML (Table 4). Percent sequence divergence ranged between 9.9% and 11.6% for all eight intra-individual calculations. The inter-individual genetic distance between the three *M. murinus* individuals was 11.1%.

Table 4: Genetic distance values for V1R*strep* repertoire diversity intra-individually (*M. murinus, M. myoxinus, M. berthae, M. rufus, M. griseorufus*, and *M. simmonsi*) and inter-individually (*M. murinus*). Last three individuals are from Chan et al. (in prep). Letters in parenthesis differentiate captive, (c), and wild-caught, (w).

Species	ID	Number of Sequences	Nucleotide distance (F84+G)
1) *Microcebus murinus*		174	0.111
a) *Microcebus murinus1*	MmurDLC7013 (c)	57	0.104
b) *Microcebus murinus2*	Mmur2X (c)	42	0.103
c) *Microcebus murinus3*	RMR49 (w)	75	0.106
3) *Microcebus myoxinus*	Mmyo020	66	0.099
4) *Microcebus berthae*	Mber030	77	0.108
5) *Microcebus rufus*	Mruf060	76	0.107
7) *Microcebus griseofurus*	RMR65	68	0.116
8) *Microcebus simmonsi*	BET5	83	0.105

Genetic distances for this subfamily revealed high levels (~10.6%) of intra-individual sequence divergence, which was similar among all eight individuals (Table 4).

[7] Hohenbrink et al., 2012

Genetic distances among individuals were nearly equal to those within, including *M. murinus* from which we have data from 3 individuals (one wild-caught). This within species comparison between wild and captive bred individuals provided further credence to the value of utilizing captive bred mouse lemurs in V1R studies. Part of this striking level of sequence divergence came from the large number of gene copies identified and the variation in copy number, an observation coinciding with previous studies on individual copy number variation in chemosensory genes generally.[8]

3.4 Combined methods provide conservative estimate of V1R diversity

Intact gene copies for DLC7013 using unbarcoded primers targeting V1R*strep* from both Sanger-sequencing clones from Chan et al.[9] and this study's PacBio CCS were aligned. A Jukes-Cantor neighbor-joining tree was constructed in Geneious to visualize overlap and differences. Approximately 50% of the sequences were recovered by both methods. An additional comparison was made by aligning these gene copies with PacBio CCS intact sequences with 2X coverage and up (Appendix F). Following standards from previously reported V1R genetic data for other taxa, this alignment was then clustered at 98% similarity, which are considered identical as minor differences may represent sequencing error, heterozygous gene copies and recent gene duplications.[10] This resulted in 69 V1R*strep* gene copies for one *Microcebus* individual (Figure 7).

[8] Nowaza & Nei, 2008
[9] Chan et al., in prep
[10] Rodriguez et al., 2002

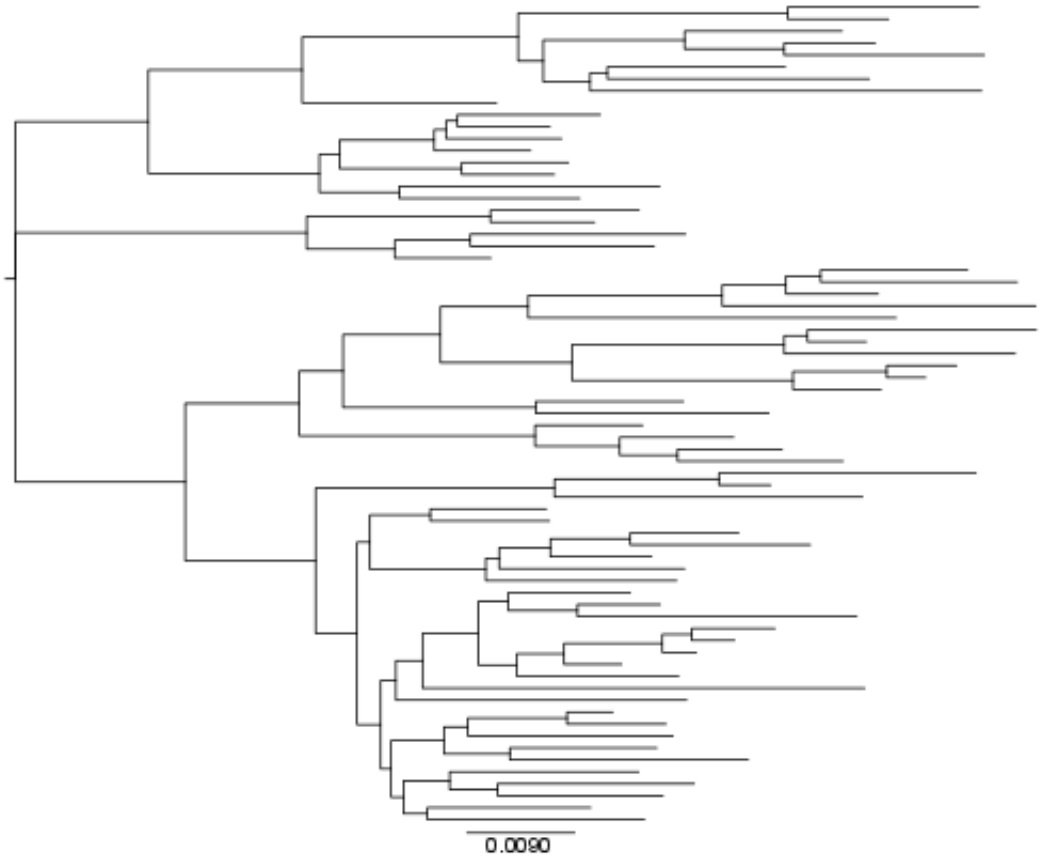

Figure 7: NJ tree of V1R*strep* gene copies clustered at 98% for one *M. murinus* individual (DLC7013) from sequences in this study and Chan et al (in prep).

Both PacBio CCS and cloning with Sanger sequencing produced high levels of V1R diversity. This study identified at least 57 highly confident distinct sequences (at least 4X coverage) and Chan et al.[11] identified at least 68 distinct clones. Based on these results, it appeared that both methods provided a limited perspective given that the coverage of the entire subfamily afforded by each was low despite the 75 clones sequenced per individual by Chan et al.[12] and ~2,000 CCS reads in this study. It is possible that some of the disagreement is due to error. Therefore, in order to estimate

[11] Chan et al., in prep
[12] ibid.

the most comprehensive number of distinct *Microcebus* V1R*strep* gene copies, we combined gene copies identified by each method and clustered those at 98% similarity (standard used in reporting V1R diversity literature).[13] The result was a conservative estimate of 69 distinct gene copies, which may not include recent duplications and heterozygous alleles. *This supports a minimum conservative estimate of approximately 69 distinct intact gene copies constituting the V1Rstrep subfamily in mouse lemurs, which is over twice the number of the previously recognized diversity in this subfamily.*

Intact gene copies from both subfamilies identified for Mmur2X were aligned to the 105 unique V1R sequences identified from the 2X draft genome by Young et al.[14] A neighbor-joining tree was constructed from this alignment to visualize overlap and dissimilarities (Figure 8).

[13] Rodriguez et al., 2002
[14] Young et al., 2010

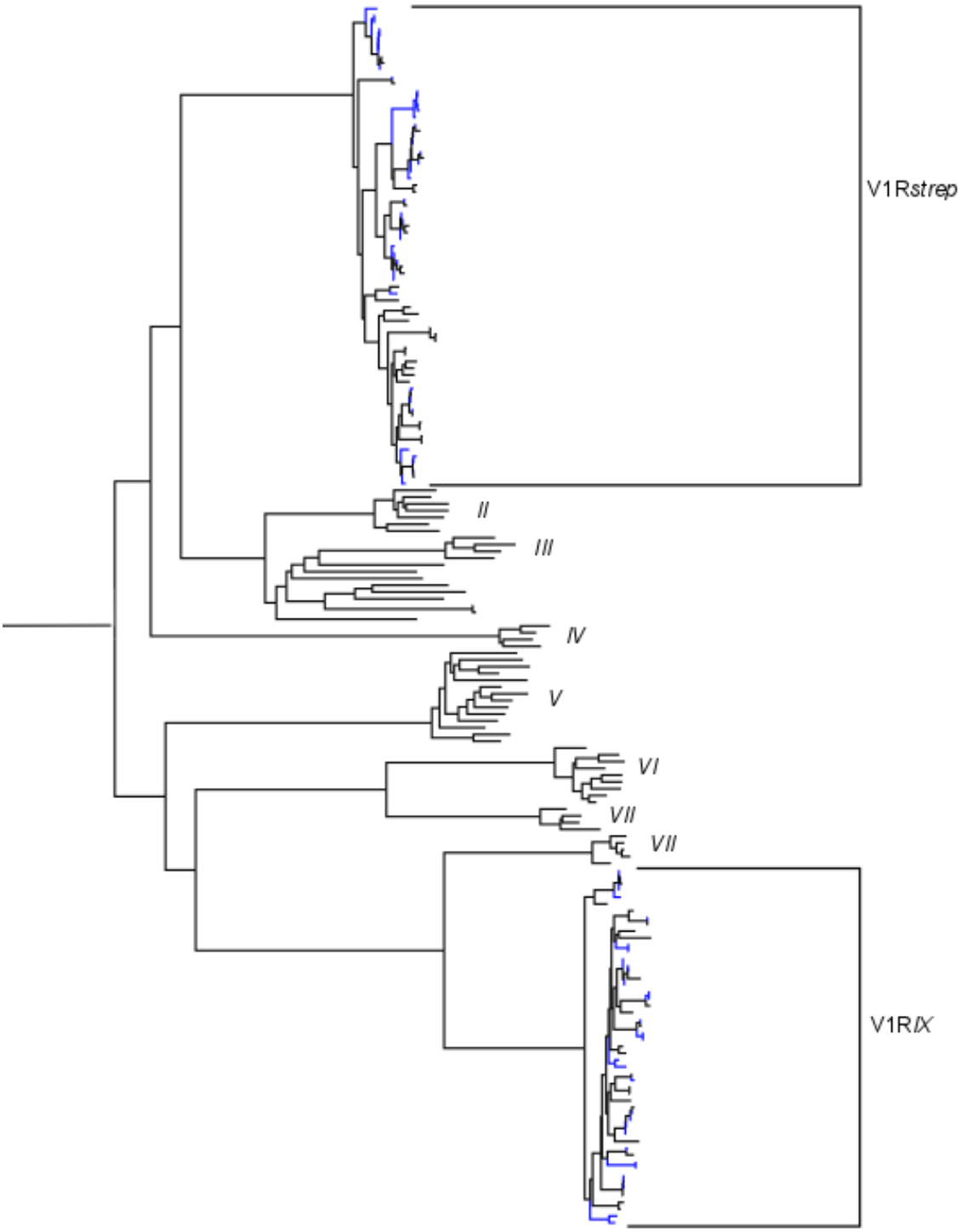

Figure 8: V1R gene family tree. PacBio sequences from Mmur2X (in blue) aligned with Young et al. (2010) identified V1R genes. Both targeted subfamilies are identified with brackets. The remaining subfamilies are labeled according to Hohenbrink et al. (2012).

It is likely that limitations of current sampling methods apply to the differences between gene copies identified for Mmur2X in this study and sequences mined from the draft genome. It should be noted that 7 V1R*IX* and 12 V1R*strep* gene copies were essentially identical (at least 99% similar) to the reference sequences, thereby providing proof-of-method (Figure 9). Clustering all identified gene copies for these two subfamilies for Mmur2X at 98% percent resulted in a conservative estimate of 22 V1R*IX* and 25 V1R*strep* individual gene copies. *The much lower estimate for V1Rstrep in this individual (as compared to the 69 V1Rstrep gene copies for DLC7013) indicates issues in using draft genome sequences to gain an accurate representation of diverse gene families. Both low coverage and lack of assembly prevent complete characterization or, in this case, provide a considerable underestimation.* Young et al.[15] acknowledged similar limitations to adequately characterizing V1R variation and stated that their analyses provide only "approximate surveys rather than accurate estimates of gene numbers." This underestimation is likely to be an issue for other taxa in which low coverage sequences were used to discover V1R diversity. As a result, V1R diversity may be even greater than we currently appreciate in a wide range of taxa.

[15] Young et al., 2010

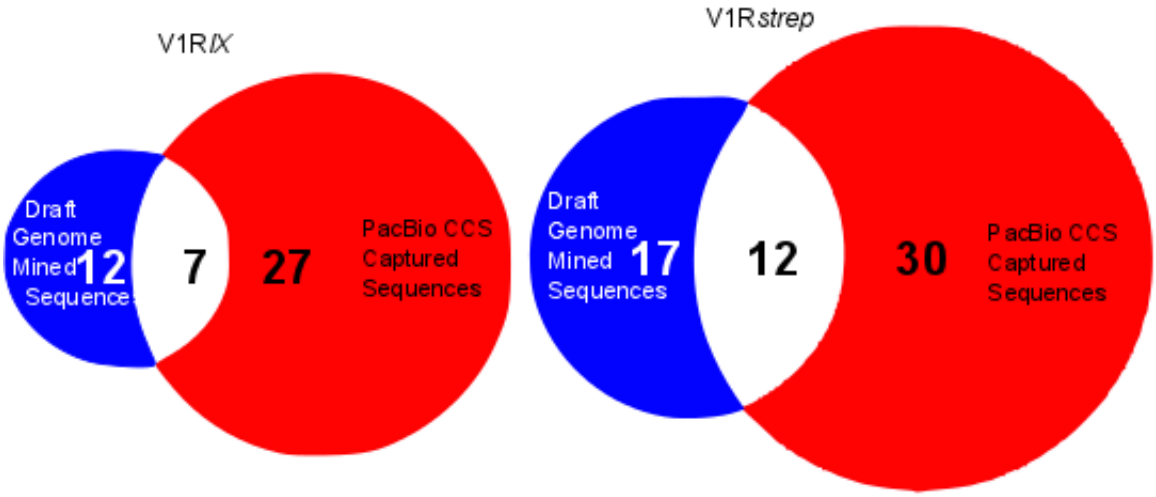

Figure 9: Venn diagram of the sequences identified from mining draft genome scaffolds and those by PacBio CCS. Numbers in the blue portion represent sequences only captured by mining draft whereas numbers in red portion were sequences captured only by PacBio CCS. Numbers in white region represent overlap in sequences captured.

3.5 Genetic diversity in Microcebus V1R subfamilies

The phylogenetic reconstructions and the high intra-individual sequence divergence demonstrated the overall extraordinary V1R diversity in *Microcebus* V1R repertoires. As mentioned above, this variation encompassed different types of variation (e.g. heterozygous, paralogous and orthologous genes). The depth and breadth of coverage achieved by SMRT CCS provided enough resolution to capture heterozygous loci. If every V1R gene copy identified is considered to come from a heterozygous site, the number of distinct loci for each subfamily is still high. The comprehensive and conservative estimate of 69 distinct gene copies for V1R*strep* suggests a lower limit of 35 separate loci (this estimate excludes possible recent duplications). Clusters of multiple gene copies from the same individual, most of which occur at the tips of the two trees, indicated paralogous alleles, which were generated post-divergence. The

interspersion of the gene copies from different species provides evidence for the persistence of multiple orthologous genes, which were a result of vertical descent from the common ancestor of four species characterized in this study and the two additional species from Chan et al.[16] Though there is strong evidence for the existence of heterozygous, paralogous and orthologous alleles contributing to the overall variation in *Microcebus* V1R subfamily diversity, it is not possible at this time to differentiate which gene copy is of which type.

Copy number variation (CNV) also contributed to the diversity detected between species for both subfamilies. The number of V1R copies in V1R*IX* ranged from 25 (Mruf060) to 40 (Mber030) whereas in V1R*strep* it ranged from 42 (Mmur2X) to 77 (Mber030) (Table 2). Young et al.[17] discovered high CNV in a broad evaluation between very distantly related mammalian taxa (order-level comparison). This study demonstrated that this phenomenon may also extend down to the species-level. Tentatively, this study revealed the potential for individual-level CNV within *M. murinus*. Individual CNV has been found in previous studies on OR genes in humans, in which >30% of the ~800 OR genes were polymorphic with respect to copy number.[18,19] Though more exhaustive work needs to be done to confirm individual CNV for V1R in mouse lemurs, this possibility also has support from the observed rapid gene birth and death in

[16] Chan et al., in prep
[17] Young et al., 2008
[18] Nozawa et al., 2007
[19] Young et al., 2008

of the V1R gene family[20] as well as from comparisons between the human genomes revealing 76 unique copy number polymorphisms between 20 individuals.[21]

A type of genetic diversity identified but not included in my analyses was the pseudogene. Far fewer pseudogenes than intact genes were identified in this study; pseudogenes were less than 2% in V1R*IX* and only ~14% in V1R*strep* of the total gene copies found (Table 2). This result tentatively supported the finding by Young et al.[22] suggesting that *Microcebus* is remarkable in its high number of intact genes to pseudogenes (predicted 19% pseudogenized). This contrasts with the at least 39% pseudogenized copies in other mammalian species with greater V1R diversity, including mouse (153 of 392), rabbit (162 of 283), and platypus (122 of 1405).[23] One caveat to this result may be that few pseudogenes were identified in this study for two primary reasons. First, gene copies identified as intact in this study may not truly be intact because the loss-of-function mutation occurred outside the target region. Second, the mutations occurring within the region of either or both the primer sequence would prevent primer annealing so that these pseudogenes would not be amplified. In the future, it will be useful to capture the entire exon so that it will be possible to capture all gene copies and differentiate between fully intact genes and pseudogenes. Even so, targeting ~800 of the ~900 total bp, this study provided a reasonable, preliminary assessment of the relative number of pseudogenes within these two families.

---

[20] Rodriguez, 2008
[21] Sebat et al., 2004
[22] Young et al., 2010
[23] ibid.

3.6 Utility of PacBio CCS

The PacBio RS system proved ideal for targeting this diverse gene family. No assembly was required as the long reads covered the entire ~800 bp amplicon. At least 33,000 sequencing cells (based on the outputted number of quality raw reads) across two SMRT cells in just two hours provided broad and deep coverage of multiple targeted sequences. Including barcoded primer amplification provided further efficiency as multiple individuals were characterized in one run. Though this technology has been criticized for its high error rate (~16%), CCS overcomes the high per-read error rate (<0.1%).[24] Recently, Pacific Biosciences improved their sequencing chemistry and this greatly increased sequencing accuracy. Errors associated with sequence repeats, multiple cytosines in particular, remained an issue, but artificial insertions and deletions were easily removed by manually calling consensus sequences.

For nearly 30 years, Sanger chain-terminated sequencing was the dominant sequencing technology.[25] This method, coupled with bacterial vector cloning, provided a breakthrough in which multiple alleles could be resolved.[26] By combining the low amplification error rate, innate proof-reading and repair mechanisms of the bacterial host with the low error rate of Sanger sequencing (~1 in 10,000), researchers were able to confidently assign true sequence reads. Unfortunately, this process is time-consuming and costs about $500 per Mb on average.[27] Alternatively, this PacBio CCS approach amounted to approximately $28 per Mb. The direct comparison between results from these two technologies demonstrated that multiple interrogation techniques are needed

---

[24] Glenn, 2011
[25] Schuster, 2008
[26] Kircher & Kelso, 2010
[27] ibid.

to better quantify the high levels of diversity within this gene family. Overall, the methods used in this study provide future studies targeting V1R and other diverse gene families with a cheaper and more efficient alternative to traditional cloning and Sanger sequencing.

3.7 Non-invasive samples for downstream next generation sequencing

Another interesting result of this study was the effective use of non-invasive samples for amplification of multiple closely related genes and for downstream next generation sequencing. Both locally acquired tissue- and field collected hair-extracted genetic material provided very similar numbers of filtered CCS reads, which remained after stringent filtering (Figure 4). This result demonstrated feasibility of extracting high quality DNA from hair, which provided enough DNA for reliable amplification of V1R gene copies. Filtering of those amplicons demonstrated that they were likely no different than those that were derived from tissue. Overall, this lends power to non-invasive techniques, which are vital to conservation studies on wild populations.

Conclusions

This study documents high levels of genetic diversity within two *Microcebus* V1R subfamilies. The results reported herein reinforce previous studies hypothesizing that V1R diversity in *Microcebus* is underestimated.[1] By targeting two subfamilies in four species, this study brings us closer to identifying the full diversity of V1R within *Microcebus*. This improved awareness of the extraordinarily diverse V1R repertoire in mouse lemurs may represent a broader scale underestimation across mammalian taxa. V1R research may have only touched the surface of the enormous V1R diversity across vertebrates (excluding those organisms with complete reference genomes). The results of this study emphasize the primary utility of using more limited reference sequences for preliminary estimations from which further investigation can and should be pursued. Future research on V1R diversity should intensify sampling strategies, even those that have been previously considered exhaustive. Researchers should not limit themselves to current approximations when designing their sampling strategies as they may miss a substantial portion of the total diversity.

Results also uncover a pattern of rapid evolution for V1R subfamilies, which strengthens the argument that these genes are important to *Microcebus* evolution. The identification of many more intact than non-functional copies provides preliminary evidence for a greater rate of gene birth than death, which thus far appears to be most dramatic in mouse lemurs in comparison with other characterized mammalian V1R repertoires. Previous studies have found evidence for different selective forces for V1Rs,

[1] Young et al., 2010

such as positive selection for both *Microcebus*[2] and *Mus*[3] and genomic drift with weak purifying selection in *Mus.*[4, 5] Further characterizations of the V1R diversity building upon this study will need to test these alternative hypotheses.

The method itself for capturing V1R diversity in this study also provides an important contribution to V1R research. The PacBio *RS* platform with CCS is able to target sequences with high efficiency and provide high accuracy at a fraction of the cost and time required for traditional cloning with Sanger sequencing. Future studies will benefit from the development of and proof-of-concept for this method.

Though there was no obvious evidence for the role of these two subfamilies contributing to species boundaries, the high diversity in these two subfamilies supports their selective advantage, perhaps indicating functional importance for detection of kairomones as opposed to pheromones. Regardless, the extreme diversity and considerable intra- and inter-individual genetic distances provide compelling evidence for their essential function in some capacity related to species survival and evolution. The clear-cut *in vivo* study conducted by Isogai et al.[6] was instrumental in demonstrating the highly specific relationship between ligand, receptor, and neuron. This information can be used to make tentative predictions based on sequence homology, but generally, the deep divergence between *Mus musculus* and other taxa also containing high V1R diversity prevents extension of functional conclusions in *Mus* to V1Rs in other organisms with any degree of certainty. Given the extraordinarily diverse V1R repertoire in mouse

---

[2] Hohenbrink et al., 2012
[3] Shi et al., 2005
[4] Park et al., 2011
[5] Kurzweil et al., 2009
[6] Isogai et al., 2011

lemurs, there is strong potential for this system to be quite astonishing in regards to V1R function.

One of the most remarkable observations in this study, which was not foreseen, was the substantial degree of polymorphism. The resolution of the PacBio CCS methodology as well as the combination of multiple "samplings" from different methods provided highly confident sequences from which I identified 29 unique V1R*IX* gene copies and 20 unique V1R*strep* gene copies among the five individuals characterized in this study (numbers resulting from clustering all gene copies within each subfamily at 98%). The high polymorphism and large genetic distances discovered here for V1R gene copies has been reported in few other gene families, one of which is the major histocompatibility complex (MHC), a gene family essential to immune defense.[7]

Overall, the advancement in our knowledge of accurate diversity in multiple mouse lemur provides compelling groundwork, both in material and methods, for further research exploring: (1) ecological pressures and evolutionary processes driving such extreme diversity; (2) diversity within and between additional individuals, populations, and species; and (3) experiments to discover the exact function of given subfamilies.

[7] Figueroa et al., 1988

Appendix A

Table 5: Collection information for wild-caught samples.

Sample Name	Species	Sample type	Weight (in grams)	Location Name	Latitude	Longitude	Collection Date
AMH001	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.13282	E044.81077	9/4/11
AMH002	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.13284	E044.81059	9/4/11
AMH003	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.13399	E044.81136	9/4/11
AMH004	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.13886	E044.81217	9/4/11
AMH005	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.13121	E044.81081	9/5/11
AMH006	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.13138	E044.81013	9/5/11
AMH007	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.13143	E044.81026	9/5/11
AMH008	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.13290	E044.81067	9/5/11
AMH009	Microcebus myoxinus	Hair	0.01	Bekopaka	S19.13263	E044.81098	9/5/11
AMH010	Microcebus myoxinus	Hair	0.01	Petit Tsingy	S19.14362	E044.79471	9/6/11
AMH021	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05552	E044.66446	9/14/11
AMH022	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05588	E044.66354	9/15/11
AMH023	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05539	E044.66402	9/15/11
AMH024	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05453	E044.66418	9/15/11
AMH025	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05532	E044.66384	9/15/11
AMH026	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05286	E044.66524	9/15/11
AMH027	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05396	E044.66426	9/15/11
AMH028	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05684	E044.66375	9/16/11

AMH029	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05229	E044.66488	9/16/11
AMH030	Microcebus berthae	Hair	0.01	Kirindy (N5)	S20.05588	E044.66354	9/16/11
AMH051	Microcebus rufus	Hair	0.01	Talatakely	S2115.492	E04725.284	9/21/11
AMH052	Microcebus rufus	Hair	0.01	Talatakely	S2115.766	E04725.300	9/21/11
AMH053	Microcebus rufus	Hair	0.01	Campsite	S2115.228	E04725.232	9/22/11
AMH054	Microcebus rufus	Hair	0.01	Campsite	S2115.232	E04725.252	9/22/11
AMH055	Microcebus rufus	Hair	0.01	Campsite	S2115.191	E04725.200	9/22/11
AMH056	Microcebus rufus	Hair	0.01	Talatakely	S2115.642	E04725.181	9/23/11
AMH057	Microcebus rufus	Hair	0.01	Campsite	S2115.232	E04725.252	9/24/11
AMH058	Microcebus rufus	Hair	0.01	Campsite	S2115.221	E04725.225	9/24/11
AMH059	Microcebus rufus	Hair	0.01	Campsite	S2115.171	E04725.146	9/24/11
AMH060	Microcebus rufus	Hair	0.01	Campsite	S2115.175	E04725.125	9/24/11

Appendix B

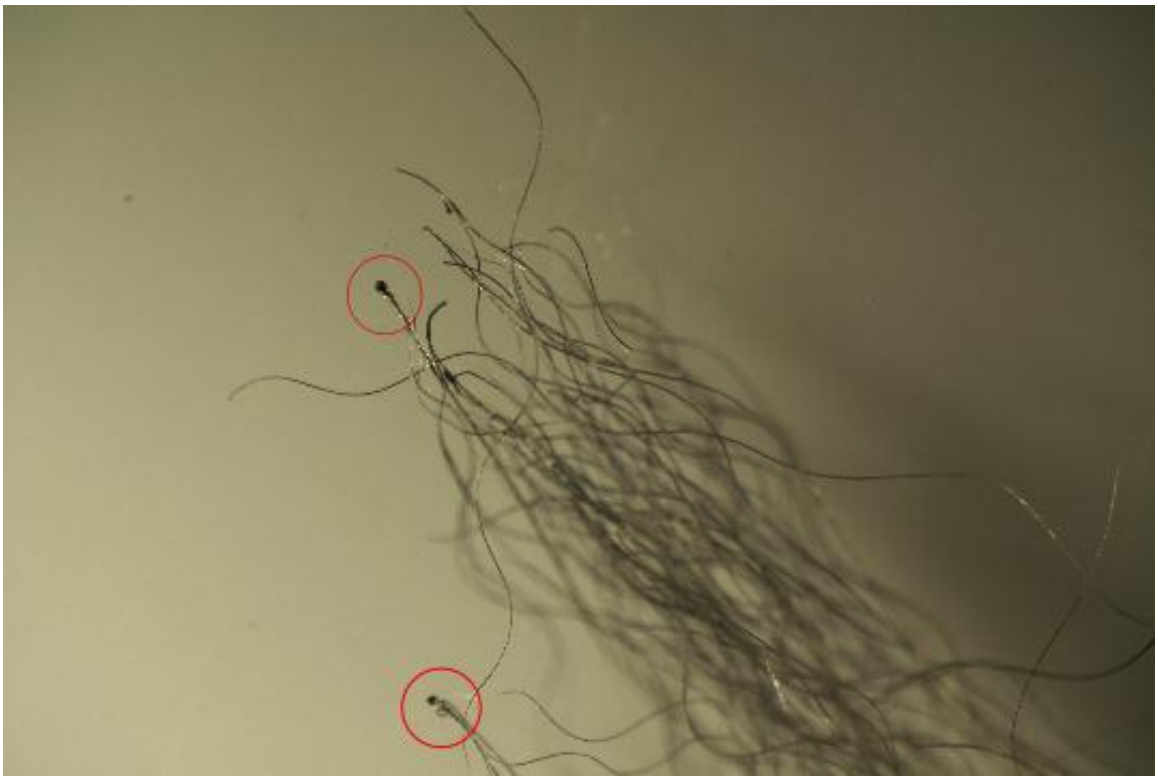

Figure 10: Photo from dissecting microscope of hair from Duke Lemur Center *Microcebus murinus* individual. Attached follicles are circled in red.

Appendix C

Table 6: Barcoded and unbarcoded primers. Primers beginning with "micMur…" were designed in this study. "V1RG" primers are from Chan et al. (in prep). Barcoded sequences are lower case letters and are from Pacific Biosciences website.

Primer Name	Primer Sequence
micMurV1R.I.RVR.PacBio	TGACAATGAACASAAAGG
micMurV1R.I.FWD.PacBio.F0	gcgctctgtgtgcagcTGGAATCYYTGGAAAT
micMurV1R.I.FWD.PacBio.F1	tcatgagtcgacactTGGAATCYYTGGAAAT
micMurV1R.I.FWD.PacBio.F2	tatctatcgtatacgcTGGAATCYYTGGAAAT
micMurV1R.I.FWD.PacBio.F3	atcacactgcatctgTGGAATCYYTGGAAAT
micMurV1R.I.FWD.PacBio.F5	tgtgagtcagtacgcgTGGAATCYYTGGAAAT
micMurV1R.IX.RVR.PacBio	ATTATCACHAAAGGACTYA
micMurV1R.IX.FWD.PacBio.F0	gcgctctgtgtgcagcCAYAGCCAAYAYCTT
micMurV1R.IX.FWD.PacBio.F1	tcatgagtcgacactCAYAGCCAAYAYCTT
micMurV1R.IX.FWD.PacBio.F2	tatctatcgtatacgCAYAGCCAAYAYCTT
micMurV1R.IX.FWD.PacBio.F3	atcacactgcatctgCAYAGCCAAYAYCTT
micMurV1R.IX.FWD.PacBio.F5	tgtgagtcagtacgcgCAYAGCCAAYAYCTT
V1RG1R	GACAATGAACACAAAGGGGCTGAA
V1RG1F	CTCAACCAGCTGGTCTTAGCYAAC
V1RG1F_F0	gcgctctgtgtgcagcCTCAACCAGCTGGTCTTAGCYAAC
V1RG1F_F1	tcatgagtcgacactaCTCAACCAGCTGGTCTTAGCYAAC
V1RG1F_F2	tatctatcgtatacgcCTCAACCAGCTGGTCTTAGCYAAC
V1RG1F_F3	atcacactgcatctgaCTCAACCAGCTGGTCTTAGCYAAC
V1RG1F_F4	acgtacgctcgtcataCTCAACCAGCTGGTCTTAGCYAAC

Appendix D

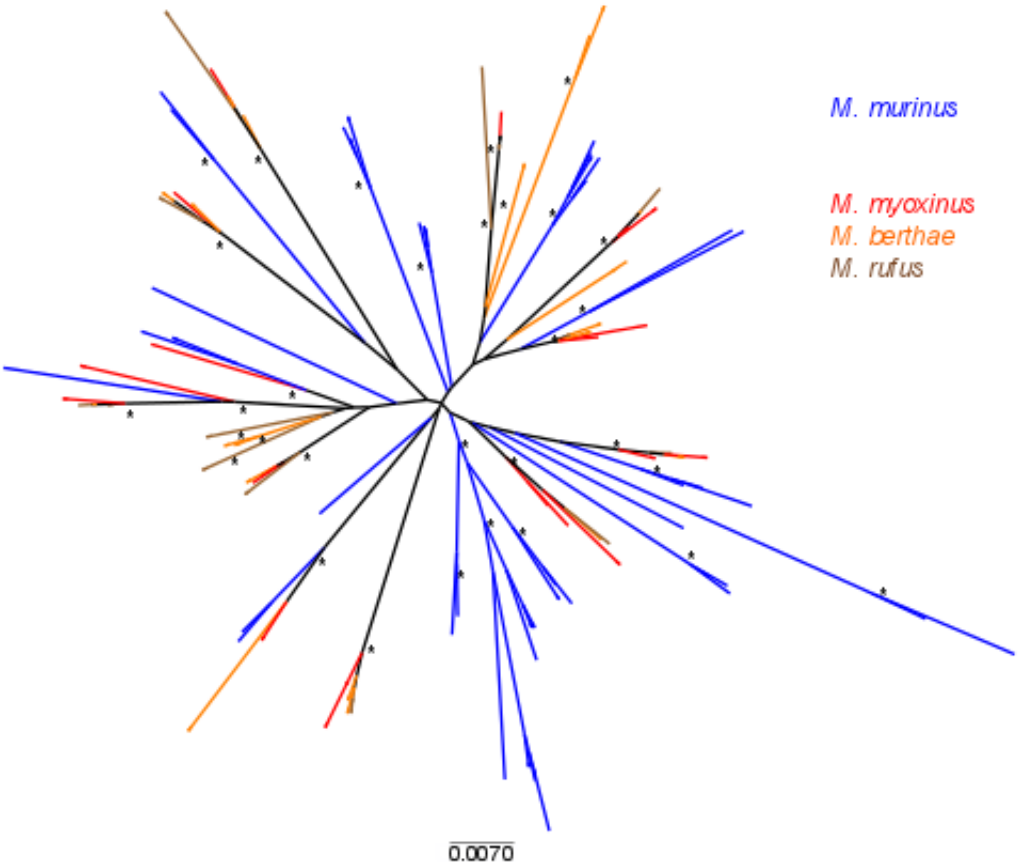

Figure 11: **Unrooted phylogram with bootstrap support for V1RIX.** Neighbor-joining (Jukes-Cantor distances), 1,000 bootstrap replications consensus tree. Nodes with bootstrap scores 50% and higher denoted by asterisk.

Appendix E

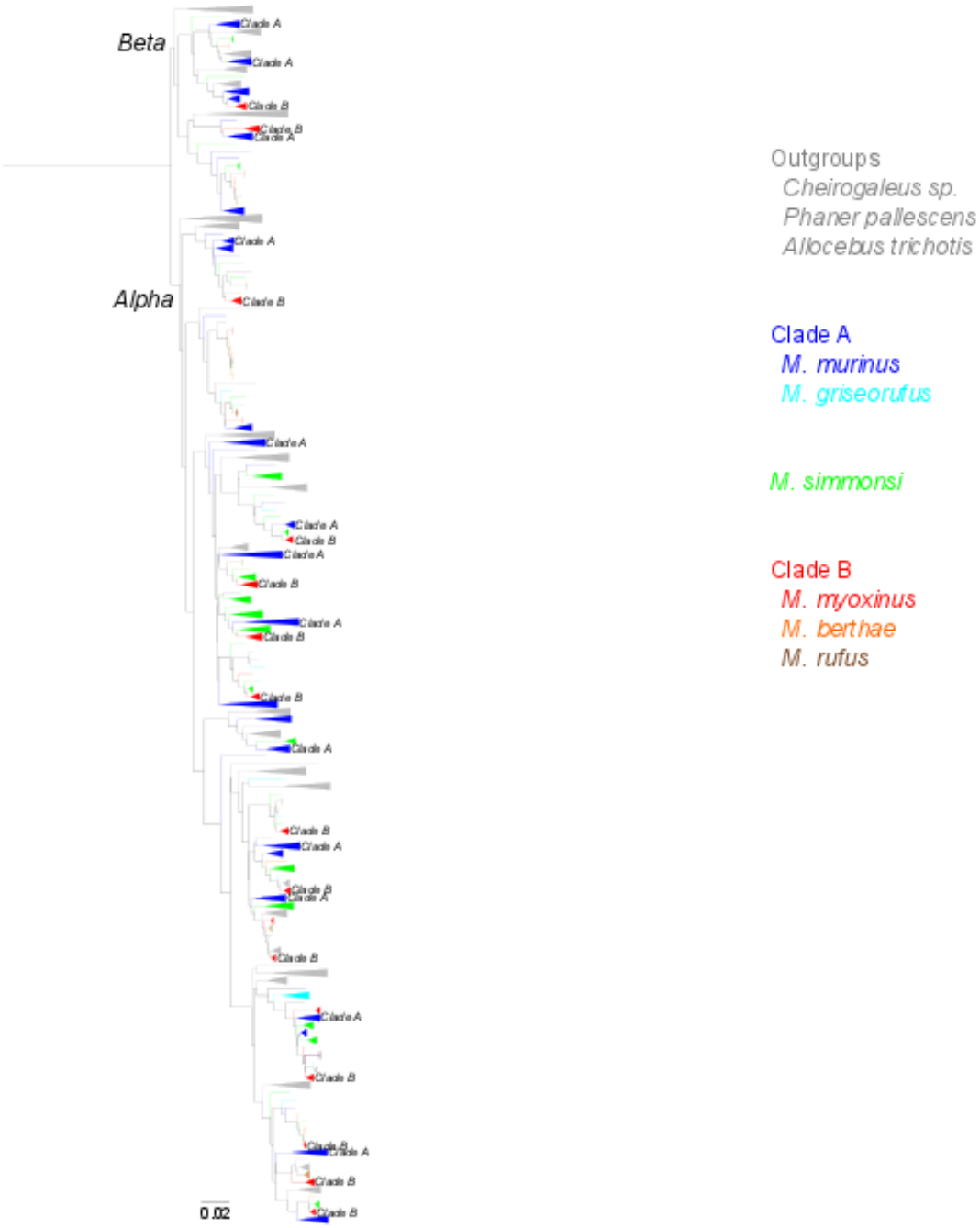

Figure 12: V1R*strep* consensus tree from 500 bootstrap replicates. Triangular termini are collapsed gene copies; collapsed gene copies from more than one species are labeled by clade. Basal nodes forming “Beta” and “Alpha” lineages are labeled**.**

Appendix F

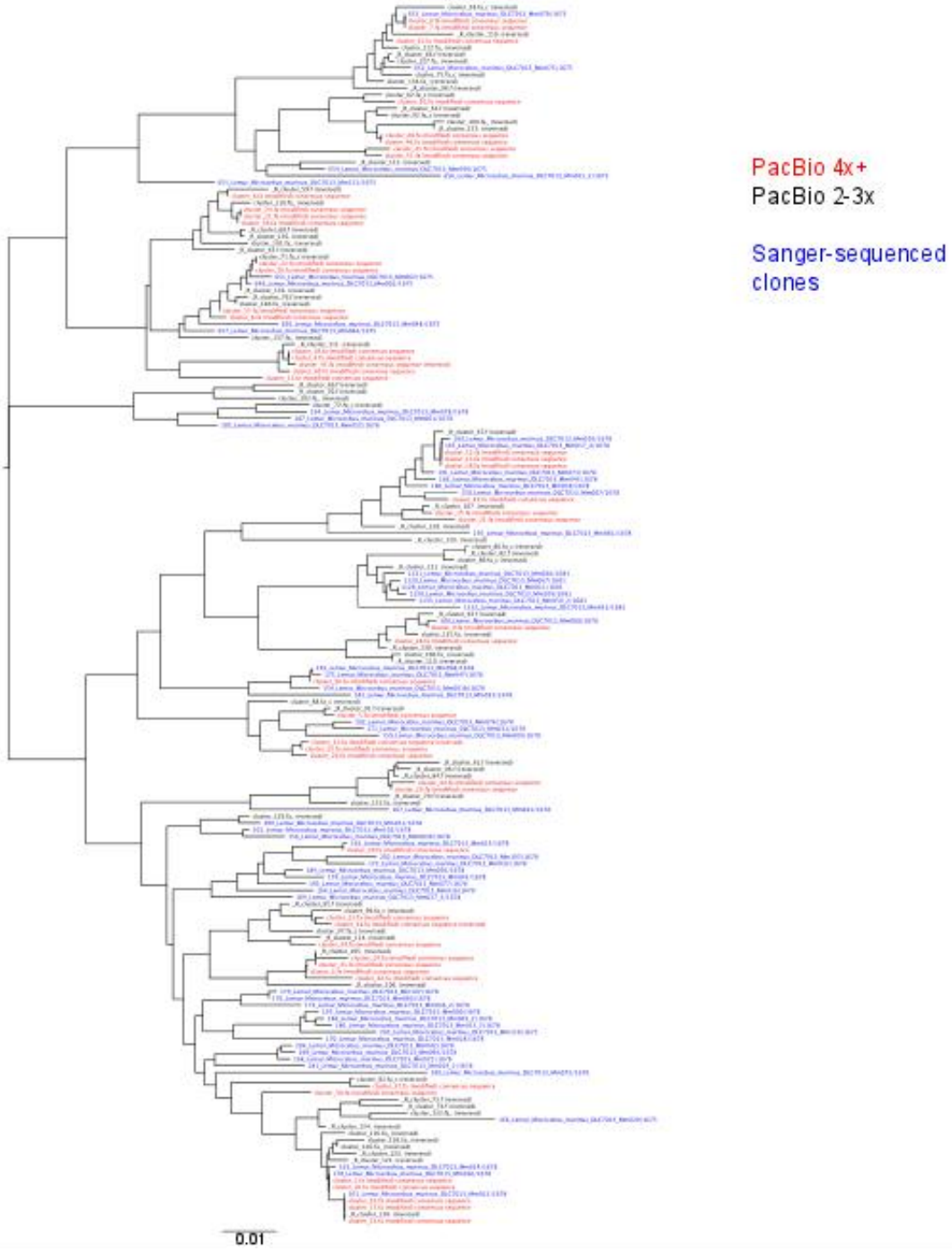

Figure 13: Neighbor-joining tree of gene copies from PacBio (4X+ in red; 2X-3X in black) and Sanger-sequenced clones (blue) for V1R*strep* in DLC7013.

References

Ache, BW and JM Young. “Olfaction: Diverse Species, Conserved Principles.” *Neuron* 48.3 (2005): 417–430.

Araújo, A. “Male Behavioral Response to the Urine Odor of Females in Lesser Mouse Lemur (*Microcebus murinus* Miller, 1777)(*Cheirogaleidae*, Primates).” *Revista Brasileira de Zoociências* 5.1 (2009): 55-77.

Atsalis, S. “Spatial Distribution and Population Composition of the Brown Mouse Lemur (*Microcebus rufus*) in Ranomafana National Park, Madagascar, and Its Implications for Social Organization..” *American Journal of Primatology* 51.1 (2000): 61–78.

Aujard, F. “Effect of Vomeronasal Organ Removal on Male Socio-Sexual Responses to Female in a Prosimian Primate (*Microcebus murinus*).” *Physiology & behavior* 62.5 (1997): 1003–1008.

Blanco, MB and JS Meyer. “Assessing Reproductive Profiles in Female Brown Mouse Lemurs (*Microcebus rufus*) From Ranomafana National Park, Southeast Madagascar, Using Fecal Hormone Analysis.” *American Journal of Primatology* 71.6 (2009): 439–446.

Brennan, PA and F Zufall. “Pheromonal Communication in Vertebrates.” *Nature* 444.7117 (2006): 308–315.

Brykczynska, U et al. “Contrasted Evolution of the Vomeronasal Receptor Repertoires in Mammals and Squamate Reptiles.” *Genome Biology and Evolution* (2013): 1-28.

Buesching, CD et al. “Multimodal Oestrus Advertisement in a Small Nocturnal Prosimian, Microcebus Murinus.” *Folia Primatologica* 69.1 (1998): 295–308.

Chamero, P, T Leinders-Zufall, and F Zufall. “From Genes to Social Communication: Molecular Sensing by the Vomeronasal Organ.” *Trends in Neurosciences* (2012): 1–10.

Chan, LM et al. "Ancient Gene Duplication and other Surprising Patterns of Allelic Diversity of V1R genes in Toothcombed Primates (suborder *Strepsirrhini*)." In prep.

Clancy, AN, A Coquelin, F Macrides, RA Gorski, and EP Nobel. Sexual behavior and aggression in male mice: involvement of the vomeronasal system. *Journal Neuroscience* 4, (1984): 2222-2229.

Dammhahn, M and PM Kappeler. “Social System of *Microcebus berthae*, the World’s Smallest Primate.” *International Journal of Primatology* 26.2 (2005): 407–435.

Dammhahn, M and PM Kappeler. “Small-Scale Coexistence of Two Mouse Lemur Species (*Microcebus berthae and M. murinus*) Within a Homogeneous Competitive Environment.” *Oecologia* 157.3 (2008): 473–483.

Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life*. London: John Murray, Albemarle Street, 1859.

Daza-Vamenta, Riza et al. “Genetic Divergence of the Rhesus Macaque Major Histocompatibility Complex.” *Genome Research* 14.8 (2004): 1501–1515.

Degnan, JH and NA Rosenberg. “Gene Tree Discordance, Phylogenetic Inference and the Multispecies Coalescent.” *Trends In Ecology & Evolution* 24.6 (2009): 332–340.

Dulac, C, and R Axel. “A Novel Family of Genes Encoding Putative Pheromone Receptors in Mammals.” *Cell* 83.2 (1995): 195–206.

Eid, J et al. “Real-Time DNA Sequencing From Single Polymerase Molecules.” *Science* 323.5910 (2009): 133–138.

Fastx toolkit. http://hannonlab.cshl.edu/fastx_toolkit

Figueroa F, Gunther E, Klein J. MHC polymorphism pre-dating speciation. *Nature* 335(1988):265– 67.

Geneious version 6.1 created by Biomatters. Available from http://www.geneious.com/

Glatston, A. "Olfactory Communication in the Lesser Mouse Lemur (*Microcebus murinus*)." *Perspectives in Primate Biology*. By P. K. Seth. New Delhi: Today & Tomorrow’s Printers, 1983. 63-73.

Glenn, TC. “Field Guide to Next-Generation DNA Sequencers.” *Molecular Ecology Resources* 11.5 (2011): 759–769.

Goecks, J, A Nekrutenko, J Taylor and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* 11.8 **(**2010):R86.

Goodman, SM. "Predation on Lemurs." *The Natural History of Madagascar*. Chicago: University of Chicago, 2003. 1221-228.

Grus, WE and J Zhang. “Origin of the Genetic Components of the Vomeronasal System in the Common Ancestor of All Extant Vertebrates.” *Molecular Biology and Evolution* 26.2 (2009): 407–419.

Grus, WE et al. “Dramatic Variation of the Vomeronasal Pheromone Receptor Gene Repertoire Among Five Orders of Placental and Marsupial Mammals.”

Proceedings Of The National Academy Of Sciences Of The United States Of America 102.16 (2005): 5767-5772.

Grus, WE and J Zhang. “Rapid Turnover and Species-Specificity of Vomeronasal Pheromone Receptor Genes in Mice and Rats.” *Gene* 340.2 (2004): 303–312.

Hapke, A et al. “Hybridization of Mouse Lemurs: Different Patterns Under Different Ecological Conditions.” *BMC Evolutionary Biology* 11.1 (2011): 297.

Heckman, KL et al. “Multiple Nuclear Loci Reveal Patterns of Incomplete Lineage Sorting and Complex Species History Within Western Mouse Lemurs (*Microcebus*).” *Molecular Phylogenetics and Evolution* 43.2 (2007): 353–367.

Hohenbrink, P, U Radespiel, and NI Mundy. “Pervasive and Ongoing Positive Selection in the Vomeronasal-1 Receptor (V1R) Repertoire of Mouse Lemurs.” *Molecular Biology and Evolution* 29.12 (2012): 3807–3816.

Isogai, Y et al. “Molecular Organization of Vomeronasal Chemoreception.” *Nature* 478.7368 (2011): 241–245.

Johnston, RE and J DelBarco-Trillo. "Communication by Chemical Signals: Behavior, Social Recognition, Hormones and the Role of the Vomeronasal and Olfactory Systems." *Hormones, Brain and Behavior*. Ed. D. W. Pfaff, A. P. Arnold, A. M. Etgen, and R. T. Rubin. By S. E. Fahrbach. 2nd ed. Vol. 1.: Elsevier, 2009. 395-441.

Kappel, P, S Hohenbrink, and U Radespiel. “Experimental Evidence for Olfactory Predator Recognition in Wild Mouse Lemurs.” *American Journal of Primatology* 73.9 (2011): 928–938.

Kappeler, PM and RM Rasoloarison. “Morphology, Behavior and Molecular Evolution of Giant Mouse Lemurs (*Mirza Spp*.) Gray, 1870, with Description of a New Species.” *Primate Report* (2005): 1–24.

Kappeler, PM and RM Rasoloarison. "*Microcebus*, Mouse Lemurs, Tsidy." *The Natural History of Madagascar*. By Steven M. Goodman. Chicago: University of Chicago, 2003. 1310-315.

Karn, RC, JM Young, and CM Laukaitis. “A Candidate Subspecies Discrimination System Involving a Vomeronasal Receptor Gene with Different Alleles Fixed in *M. m. domesticus* and *M. m. musculus*.” Ed. Sebastian D Fugmann. *PLoS ONE* 5.9 (2010): e12638.

Katoh, K. “MAFFT Version 5: Improvement in Accuracy of Multiple Sequence Alignment.” *Nucleic Acids Research* 33.2 (2005): 511–518.

Keverne, EB. “The Vomeronasal Organ.” *Science* 286.5440 (1999): 716–720.

Kircher, M and J Kelso. “High-Throughput DNA Sequencing - Concepts and Limitations.” *BioEssays* 32.6 (2010): 524–536.

Krieger, J et al. “Selective Activation of G Protein Subtypes in the Vomeronasal Organ Upon Stimulation with Urine-Derived Compounds.” *The Journal of Biological Chemistry* 274.8 (1999): 4655–4662.

Kurzweil, VC et al. “Dynamic Evolution of V1R Putative Pheromone Receptors Between *Mus musculus* and *Mus spretus*.” *BMC Genomics* 10.1 (2009): 74.

Lane, RP et al. “Sequence Analysis of Mouse Vomeronasal Receptor Gene Clusters Reveals Common Promoter Motifs and a History of Recent Expansion.” *Proceedings Of The National Academy Of Sciences Of The United States Of America* 99.1 (2002): 291–296.

Li W and A Godzik. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(2006): 1658-1659.

Mittermeier, RA et al. *Lemurs of Madagascar*. Arlington, VA: Conservation International, 2010.

Mombaerts, Peter. “Genes and Ligands for Odorant, Vomeronasal and Taste Receptors.” *Nature Reviews Neuroscience* 5.4 (2004): 263–278.

Niimura, Y. “Chapter 4: Evolution of Chemosensory Receptor Genes in Primates and Other Mammals.” In *Post-Genome Biology of Primates* (2012): 43-78.

Nozawa M and M Nei. 2008. Genomic drift and copy number variation of chemosensory receptor genes in humans and mice. *Cytogenetic Genome Research* 123: 263-269.

Nozawa, M, Y Kawahara and M Nei. “Genomic Drift and Copy Number Variation of Sensory Receptor Genes in Humans.” *Proceedings of the National Academy of Sciences* 104.51 (2007): 20421–20426.

Pacific Biosciences. http://www.pacificbiosciences.com/, Accessed 28 February 2013.

Park, SH et al. “The Microevolution of V1r Vomeronasal Receptor Genes in Mice.” *Genome Biology and Evolution* 3.0 (2011): 401–412.

Perret, M. “Relationship between urinary estrogen levels before conception and sex ratio at birth in a primate, the gray mouse lemur. ” *Human Reproduction* Vol.20, No.6 (2005): 1504–1510.

Perret, M. “Chemocommunication in the reproductive function of mouse lemurs.” Alterman, L., Gerald A. Doyle, and M. Kay. Izard. *Creatures of the Dark: The Nocturnal Prosimians*. New York: Plenum, (1995): 377-392.

Radespiel, U. “Ecological Diversity and Seasonal Adaptations of Mouse Lemurs (*Microcebus Spp*.).” *Lemurs* (2007): 211–234.

Radespiel, U et al. “First Indications of a Highland Specialist Among Mouse Lemurs (*Microcebus Spp*.) and Evidence for a New Mouse Lemur Species From Eastern Madagascar.” *Primates* 53.2 (2011): 157–170.

Rasoloarison, RM, SM Goodman, and JU Ganzhorn. “Taxonomic Revision of Mouse Lemurs (*Microcebus*) in the Western Portions of Madagascar.” *International Journal of Primatology* 21.6 (2000): 963–1019.

Rodriguez, I. “4.44 Genomics of Vomeronasal Recepts.” (2008): 815–819.

Rodriguez, I et al. “Multiple New and Isolated Families Within the Mouse Superfamily of V1r Vomeronasal Receptors.” *Nature Neuroscience* 5.2 (2002): 134–140.

Sebat, J. “Large-Scale Copy Number Polymorphism in the Human Genome.” *Science* 305.5683 (2004): 525–528.

Schilling, A and M Perret. “Chemical Signals and Reproductive Capacity in a Male Promisian Primate (*Microcebus murinus*).” *Chemical Senses* (1987): 143-158.

Schmid, J and PM Kappeler. “Sympatric Mouse Lemurs (*Microcebus* Spp.) in Western Madagascar.” *Folia Primatologica* 63.3 (1994): 162–170.

Schuster, SC. “Next-Generation Sequencing Transforms Today’s Biology.” *Nature* 5.1(2008):16-18.

Shi, P and J Zhang. “Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land.” *Genome Research* 17(2007):166-174.

Shi, P et al. “Adaptive Diversification of Vomeronasal Receptor 1 Genes in Rodents.” *Journal of Molecular Evolution* 60.5 (2005): 566–576.

Steidl, C et al. “MHC Class II Transactivator CIITA Is a Recurrent Gene Fusion Partner in Lymphoid Cancers.” *Nature* 471.7338 (2012): 377–381.

Stewart, R and R P Lane. “V1R Promoters Are Well Conserved and Exhibit Common Putative Regulatory Motifs.” *BMC Genomics* 8.1 (2007): 253.

Sündermann, D, M Scheumann and E Zimmermann. “Olfactory Predator Recognition in Predator-Naïve Gray Mouse Lemurs (*Microcebus murinus*).” *Journal of Comparative Psychology* 122.2 (2008): 146–155.

Tattersall, I. “Madagascar's Lemurs: Cryptic Diversity or Taxonomic Inflation?” *Evolutionary Anthropology: Issues, News, and Reviews* 16.1 (2007): 12–23.

Touhara, K and LB Vosshall. “Sensing Odorants and Pheromones with Chemosensory Receptors.” *Annual Review of Physiology* 71.1 (2009): 307–332.

Travers, KJ et al. “A Flexible and Efficient Template Format for Circular Consensus Sequencing and SNP Detection.” Nucleic Acids Research 38.15 (2010): e159–e159.Wyatt, Tristram D. *Pheromones and Animal Behaviour: Communication by Smell and Taste*. Cambridge, UK: Cambridge UP, 2004.

Tusnády, GE and I Simon. “The HMMTOP Transmembrane Topology Prediction Server.” *Bioinformatics* 17.9 (2001): 849–850.

Weisrock, DW et al. “Concatenation and Concordance in the Reconstruction of Mouse Lemur Phylogeny: an Empirical Demonstration of the Effect of Allele Sampling in Phylogenetics.” *Molecular Biology and Evolution* 29.6 (2012): 1615–1630.

Weisrock DW, Rasoloarison RM, Fiorentino I, Ralison JM, Goodman SM, et al. Delimiting Species without Nuclear Monophyly in Madagascar’s Mouse Lemurs. *PLoS ONE* 5.3(2010): e9883.

Wyatt, TD. *Pheromones and Animal Behaviour: Communication by Smell and Taste*. Cambridge, UK: Cambridge UP, 2004.

Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and *Evolution* 24(2007): 1586–1591.

Yoder, A and Z Yang. “Divergence Dates for Malagasy Lemurs Estimated From Multiple Gene Loci: Geological and Evolutionary Context.” *Molecular Ecology* (2004): 757-773.

Young, JM et al. “Extreme Variability Among Mammalian V1R Gene Families.” *Genome Research* 20.1 (2010): 10–18.

Young, JM et al. “Extensive Copy-Number Variation of the Human Olfactory Receptor Gene Family.” *The American Journal of Human Genetics* 83.2 (2008): 228–242.