View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by K-State Research Exchange

This is the author's final, peer-reviewed manuscript as accepted for publication. The publisher-formatted version may be available through the publisher's web site or your institution's library.

Genetic characterization and phylogenetic analysis of skunk-associated rabies viruses in North America with special emphasis on the central plains

Rolan Davis, Susan A. Nadin-Davis, Michael Moore and Cathleen Hanlon

How to cite this manuscript

If you make reference to this version of the manuscript, use the following information:

Davis, R., Nadin-Davis, S. A., Moore, M., & Hanlon, C. (2013). Genetic characterization and phylogenetic analysis of skunk-associated rabies viruses in North America with special emphasis on the central plains. Retrieved from http://krex.ksu.edu

Published Version Information

Citation: Davis, R., Nadin-Davis, S. A., Moore, M., & Hanlon, C. (2013). Genetic characterization and phylogenetic analysis of skunk-associated rabies viruses in North America with special emphasis on the central plains. Virus Research, 174(1-2), 27-36.

Copyright: © 2013 Elsevier B.V.

Digital Object Identifier (DOI): doi:10.1016/j.virusres.2013.02.008

Publisher's Link: http://www.sciencedirect.com/science/article/pii/S0168170213000580

This item was retrieved from the K-State Research Exchange (K-REx), the institutional repository of Kansas State University. K-REx is available at http://krex.ksu.edu

Genetic characterization and phylogenetic analysis of Skunk-Associated Rabies Viruses in North America with special emphasis on the Central Plains

Rolan Davis[1], Susan A. Nadin-Davis[2], Michael Moore[2] and Cathleen Hanlon[1]

Kansas State University Rabies Laboratory, 2005 Research Park Circle, Manhattan, KS 66502, USA[1]; Animal Health Microbiology Research, Ottawa Laboratory (Fallowfield), Canadian Food Inspection Agency, Ottawa, Ontario, K2J 4S1, Canada[2]

* communicating author R. Davis, Email rdavis@vet.k-state.edu

Key words: rabies virus, phylogeography, phylogeny, nucleoprotein, glycoprotein

15 **Abstract**

16 Across North America the skunk acts as a reservoir for several rabies virus variants. Some of
17 these variants are geographically restricted in range as is the case for the California skunk variant
18 and two distinct variants present in Mexico. In contrast the North Central and South Central
19 skunk rabies viruses are dispersed in overlapping ranges over large areas of the Midwestern
20 region of the United States with the former extending into southern parts of the Canadian
21 prairies. Despite this extensive range, there has been only very limited molecular characterization
22 of these two viral variants. This study has examined the genetic diversity of the rabies viruses
23 associated with North American skunks, with particular emphasis on the South Central skunk
24 variant which was found to comprise three distinct geographically restricted groups of viruses
25 that could in some cases be further sub-divided. The phylogenetic relationships of these groups
26 and sub-groups allowed us to infer the likely direction of spread of these variants in some
27 instances. Patterns of amino acid replacement of North American skunk-associated rabies viruses
28 for both the nucleoprotein and glycoprotein products are also examined. These patterns reflect
29 the virus phylogeny but no amino acid residues associated specifically with the skunk host were
30 identified.

31 **1. Introduction**

32 Rabies virus is the prototype species of the *Lyssavirus* genus (ICTV, 2011). Despite its
33 small 12 Kb single-stranded negative sense RNA genome and limited five gene coding capacity
34 (Wunner, 2007), infection with this virus results in rabies, a feared neurological disease that is
35 universally fatal except in extremely rare circumstances (Jackson, 2007). These viruses can
36 infect any mammalian species but they are maintained in association with particular reservoir
37 hosts (Hanlon et al., 2007). Although both human and animal disease can be prevented through
38 time-tested vaccination and post-exposure prophylactic regimens, these viruses continue to
39 present a risk of disease around the world today, even in many developed countries.

40 Prior to the advent of animal control measures coupled with effective vaccination
41 campaigns, the main public health threat in the United States was primarily from canine rabies
42 virus variants transmitted by domestic dogs. As canine rabies virus variants were driven to
43 extinction in the US, public awareness of the role of wildlife in rabies transmission increased
44 (Price et al., 1961) and, by 1960, diagnosed cases in wildlife grew to outnumber cases in
45 domestic animals(Eng et al., 1989). Skunks were the most commonly diagnosed species in the
46 United States during the years 1961 to 1989, but then were superseded in the 1990s by the
47 raccoon rabies epizootic in the mid-Atlantic and north-eastern states (Blanton et al., 2009).

48 In the United States, rabies cases among skunks were first reported in California as early
49 as 1826(Hovey, 1874). More widespread epizootics began to emerge in the late 1950s with four
50 distinct regions being affected by 1960 (Charlton et al., 1975). These regions included: 1) the
51 north-central states and southern regions of the Canadian-prairie Provinces, 2) Texas in the
52 south, 3) California in the west, and 4) north-eastern states neighboring the Canadian provinces

53 of Ontario and Quebec. Following the advent of monoclonal antibody panel typing (Smith et al.,
54 1986), the distinctive nature of the viruses responsible for these outbreaks was established and
55 later confirmed by molecular epidemiological studies (Nadin-Davis et al., 2002). The viruses
56 responsible for these outbreaks are currently referred to as the North Central skunk (NCSK), the
57 South Central Skunk (SCSK) and the California skunk (CASK) variants respectively while
58 skunks in Ontario and New York State were shown to harbor viruses similar to the arctic fox
59 type spread by red foxes in the same region, although it has been proposed that the skunk acts as
60 a secondary host for this variant (Nadin-Davis et al., 2006). These variants represent no less than
61 three of the seven major rabies virus lineages identified world-wide (Bourhy et al., 2008). NCSK
62 and CASK are branches of the cosmopolitan lineage thought to have been spread from Europe to
63 many parts of the world during the colonial period, the SCSK variant is a member of the
64 American Indigenous lineage found only on that continent, and the arctic fox variant is a member
65 of the Arctic/Arctic-like lineage which circulates in northern climes and across large parts of
66 Asia (Nadin-Davis et al., 2007, 2012; Kuzmin et al., 2008). In the US and Canada, the striped
67 skunk (*Mephitis mephitis*), which is the principal skunk species diagnosed with rabies, is
68 believed to be the maintenance host for these three skunk-associated rabies epizootics although
69 the disease has been documented in other species such as hog-nosed (*Conepatus leuconotus*) and
70 hooded skunks (*Mephitis macroura*) (Hass and Dragoo, 2006). In Mexico, additional viral
71 strains associated primarily with the spotted skunk (*Spilogale putorius)* have been identified
72 Aranda and Lopez-de Buen, 1999; Velasco-Villa et al., 2002; Nadin-Davis and Loza-Rubio,
73 2006).

74 Although the last reported human death due to skunk transmitted rabies occurred in the
75 United States in 1981 (Krebs et al., 2000), cases of infected domestic animals help to underscore

76 the public health importance of skunk transmitted rabies virus variants (Rupprecht et al., 1995).
77 Thus while skunks are the predominant reservoir species across the US Midwest, with 56% of
78 reported cases from these states occurring within skunks in 2008 (Blanton et al., 2009), cases
79 within domestic species, presumably caused by skunk transmitted rabies, accounted for another
80 14% of reported cases. Rabies cycles through the skunk population of the Great Plains with
81 regular peaks and troughs in the number of reported cases (Pool and Hacker, 1982; Oertli et al.,
82 2009). The factors that define these peaks are not well understood but it is speculated that rabies
83 cases may be directly tied to fluctuations in the skunk population density. Indeed cyclical
84 patterns of rabies cases have been observed in wildlife species elsewhere. In Ontario, fox rabies
85 incidence patterns of varying periodicity defined several discrete geographical units differing in
86 host species distribution and persistence (Tinline and MacInnes, 2004); it was speculated that
87 host meta-population structure plays a key role in disease persistence. In Europe, fox population
88 density, turnover and social interactions were identified as the most important ecological factors
89 influencing disease patterns (Steck and Wandeler, 1980). A detailed description of the molecular
90 epidemiology of the rabies viruses circulating in skunks across the Great Plains was expected to
91 reveal useful information about the spread of the disease in this population that is crucial to
92 planning and implementation of effective rabies control and prevention strategies directly in this
93 reservoir species. However, prior to the start of this study there was insufficient viral gene
94 sequence data available to support such analysis, as relatively few examples of genomic
95 sequences from skunk rabies viruses had appeared in the literature (Nadin-Davis et al., 1997;
96 Velasco-Villa et al., 2008) or were available in the public databases. To address this deficiency
97 and generate the first detailed analysis of the phylogeography of the SCSK rabies virus variant, a
98 substantial database of nucleotide sequences of skunk-associated rabies viruses from six states

99 (Arkansas, Kansas, Missouri, Nebraska, Oklahoma and South Dakota) in the US Midwest has
100 been compiled. Sequence information for the N (nucleoprotein) gene (all samples) and for the G
101 (glycoprotein) gene (most samples) was generated since the targeting of these two genes for
102 phylogenetic studies allows comparison with many other sequences in the public databases;
103 moreover variation at the G protein could be functionally significant with respect to host cell
104 binding, cell entry and pathogenesis (Wunner, 2007). This sequence information, when
105 combined with that from other skunk-associated rabies viruses, extends our knowledge of (1) the
106 level of genetic diversity, (2) phylogeny and (3) evolutionary processes operating on the proteins
107 of viruses that are associated with such a permissive, ubiquitous, and vagile host.

108

109

110 **2. Materials and Methods**

111 *2.1. Samples*

112 The rabies-positive samples examined in this study, which originated from several US
113 states, the prairie provinces of Canada and distinct regions of Mexico, were processed at two
114 different laboratories. The 78 samples examined by the Kansas State University (KSU) Rabies
115 Laboratory included 32 isolates from Kansas, 14 from Nebraska and a single sample each
116 originally from Colorado and Florida, all collected during early 2009. Additional US samples
117 were solicited from primary diagnostic facilities in the states of Arkansas (12), Missouri (four),
118 Oklahoma (six) and South Dakota (six), with an additional two samples originating from
119 Minnesota but provided by the facility in South Dakota. All 78 samples were received for
120 routine diagnostic testing rather than active surveillance investigations and their designations
121 were generated thus: first a two letter code to indicate state of origin followed by a two letter host
122 species code and a two digit number indicating year of submission and finally a four digit
123 submission number. Full details of these samples are provided in supplementary material (Table
124 S1). The ZIP code from where the submission originated was recorded and mapped.

125 For comparative analysis, an additional 22 rabies-positive samples were characterized at
126 the Ottawa laboratory (OLF); this included 14 CASK variant specimens from California and a
127 sample from a Mexican skunk (designated by V followed by a three digit number and the variant
128 type), and an additional seven samples from Western Canada, representing the northernmost
129 range of the NCSK variant, that are designated by L followed by a six digit number, that includes
130 the year of isolation (two digits) and a submission number (four digits), and the suffix WSK.
131 Details of these samples, together with all other isolates accessed through GenBank and used for
132 phylogenetic analyses, are listed in supplementary material (Table S2).

133 *2.2. RNA extraction, Reverse Transcription and PCR*

134 Rabies-positive samples were stored at -70°C until processed for RNA extraction.
135 Approximately 5-10 mg of brain material was added to 100µL of a lysis buffer (10mM TRIS
136 HCl, 150 mM NaCl, 1.5 mM MgCl2 and 0.65% NP40 substitute) to rupture the cells, 1mL of
137 TRIzol reagent (Invitrogen, Carlsbad, CA) was added and the sample refrigerated overnight (~18
138 hours). Lysates were then processed according to the instructions provided by the manufacturer
139 of the TRIzol reagent and each final dried RNA pellet was re-suspended in 100µl of RNase–free
140 water and stored at -70°C.

141 For those samples processed at the KSU rabies facility, amplification of the viral genome
142 at both the nucleoprotein (N) and the glycoprotein (G) genes was performed by generating two
143 overlapping amplicons for each gene using the collection of primers described in supplementary
144 material (Table S3). One of four oligonucleotide primers was annealed to the viral RNA target in
145 a reaction that contained 5 µl of the purified RNA and 1 µl of reverse transcription (RT) primer
146 (5 µM / 10 µM for degenerate primers). Mixtures were heated to 94°C for 90 seconds and then
147 cooled quickly on ice. Tubes received 14 µl of a RT reaction buffer mix containing 100 mM
148 Tris (pH 8.3), 10 mM MgCl, 0.5 mM each dNTP, 10 units of reverse transcriptase and 16 units
149 RNase inhibitor (both produced by Roche Diagnostics, Indianapolis, IN). Reactions were
150 incubated at 42°C for 90 minutes. Following RT, 80 µl of PCR pre-mix was added to the 20 µl
151 RT product. The pre-mix contained 69 µl of distilled water, 8 µl of Tris (1 M, pH 8.3), 0.5 µl
152 AmpliTaq DNA polymerase (Applied Biosystems, Branchburg, NJ), 1 µl of forward primer (20
153 µM / 40 µM for degenerate primers) and 1.5 µl of reverse primer (20 µM) / 40 µM for
154 degenerate primers. Thermal cycling was performed in a Thermo PxE 0.2 with an initial
155 denaturation at 94°C for 60 seconds, followed by 39 cycles of 94°C for 30 seconds, 42°C for 30

156 seconds, 72°C for 90 seconds followed by a final cycle ending with a seven-minute elongation
157 period at 72°C.
158 For those samples processed at the OLF rabies facility, RT-PCR of the rabies virus N
159 gene was performed essentially as described previously (Nadin-Davis, 1998) except that primer
160 RabN1 was replaced by primer RVfor2, (5'-gtACGCTTAACAACAARAYCARAGAA-3'
161 targeting bases 1-24 at the 3' end of the genomic RNA) for both the RT and PCR. RT-PCR of
162 the G gene was described previously (Nadin-Davis et al., 1997, 1999).
163 *2.3. PCR product purification and sequencing*
164 Each PCR was screened for successful amplification by analysis of an aliquot by standard
165 agarose gel electrophoresis. Rabies virus-specific amplicons were recovered from the remaining
166 PCR product using a Wizard PCR Prep DNA Purification System as recommended by the
167 manufacturer Promega (Madison, WI) either directly (OLF) or after size fractionation by
168 electrophoresis through 2% NuSieve low melting point agarose (KSU).
169 Sequencing of products generated at KSU was performed on Applied Biosystems 3730xl
170 or 3730 DNA analyzers at the University of Kentucky's Advanced Genetic Testing Center
171 (AGTC), Lexington, KY. Products prepared at OLF were sequenced in-house using a NEN
172 4200L automated sequencing system (LiCor Biosciences, Lincoln, NE) with IR-dye labeled
173 primers and a SequiTherm EXCEL™ II DNA sequencing kit (Epicentre Biotechnologies)
174 obtained from Interscience, ON.
175 *2.4. Sequence Analyses*
176 Sequences received from the AGTC facility were aligned and edited using Bioedit (Hall,
177 2011). Sequences generated at OLF were compiled and edited using Eseq version 3.0 (LiCor
178 Biosciences). Final sequence databases of complete coding regions for the N gene (1350 bases),

179 G gene (1572 bases) or the concatenated data from both genes (2922 bases) were aligned using
180 CLUSTALX version 1.8 (Thompson et al., 1997). *MEGA* version 4 (Tamura et al., 2007) was
181 used for generation of phylogenies using the neighbor joining (NJ) and maximum parsimony
182 (MP) methods, for computation of transition/transversion ratios and
183 synonymous/nonsynonymous nucleotide substitution rates (dN/dS) using the Kumar method and
184 for translation of nucleotide sequences to protein. Modeltest (Posada and Crandall, 1998) was
185 used to identify the General Time Reversible gamma model with invariant sites (GTR + G + I) as
186 the generally applicable nucleotide substitution model that best fit both the N and G gene
187 sequence databases and this model was used for generation of maximum likelihood (ML)
188 phylogenies by PhyML, version 3.0 (Guindon et al., 2010).
189
190 **3. Results**

191 *3.1. Phylogeography of skunk rabies viruses in the Midwestern US*

192 Initial efforts focused on full length N gene analysis, but since these data yielded
193 phylogenies that were only modestly supported, additional studies targeting the G gene were
194 undertaken. Analysis was completed on 59 unique N sequences, on 64 unique G sequences and
195 68 unique sequences when N & G were concatenated as previously described (Kuzmin et al.,
196 2010); this database included a newly generated sequence from an isolate of raccoon variant
197 (FLRC090148).

198 Results of a neighbor joining (NJ) analysis for these 68 concatenated N and G sequences
199 and an EBLV2 isolate included as an out-group to the rabies virus clade are shown in Figure 1.
200 Phylogenetic trees generated by analysis of each individual coding region all exhibited a similar

201 topology; however, several lineages were more highly supported when both genes were analyzed
202 in tandem rather than individually (data not shown). Samples represented both skunk transmitted
203 viral variants circulating in the central plains. All samples from South Dakota and Minnesota
204 carried the NCSK variant together with a single outlying sample isolated from a dog from
205 Arkansas. The remaining 11 samples from Arkansas along with all samples from Kansas,
206 Missouri, Nebraska and Oklahoma were members of the SCSK variant. Neighbor-joining
207 analysis identified considerable diversity within this group and delineated three strongly
208 supported clades (Figure 1). The most outlying clade (SCSK I), is comprised entirely of samples
209 from Oklahoma. Clade SCSK II is entirely populated by isolates from Arkansas and Missouri
210 with well supported sub-division into IIA and IIB types. A third clade (SCSK III), made up of
211 samples primarily distributed across Kansas and Nebraska, is further subdivided into two types
212 with some samples from Kansas (IIIA) clearly differentiated with strong support from the
213 remaining members of the clade (IIIB). As illustrated in Figure 2, each of these clades occupies a
214 distinct geographical range with no observed spatial overlap of these viral types.

215 *3.2. Phylogenetics of North American skunk rabies viruses*

216 To place these results into a broader context additional phylogenetic analysis was
217 undertaken on a broader dataset that included additional newly sequenced skunk rabies virus
218 samples from Canada and the state of California. To allow inclusion of the greatest diversity of
219 viral samples, including those from bats and other terrestrial sources available in GenBank (see
220 list in Table S2), only N gene sequences were employed but this yielded robust phylogenies due
221 to the degree of genetic variation across the dataset. Figure 3 illustrates the results of a
222 maximum likelihood (ML) analysis completed on 122 rabies virus N genes (1350 bases) using a
223 European bat lyssavirus type 2 (EBLV-2) as an outgroup; similar phylogenies were generated

224 using NJ and maximum parsimony (MP) methods (data not shown). As expected the ML tree
225 clearly divides all samples into two main lineages with strong support. The NCSK, CASK and
226 South Baja California (SBC) skunk variants all fall within a large group previously identified as
227 the cosmopolitan lineage (Nadin-Davis et al., 2002). The American indigenous lineage clearly
228 segregates into two groups of viruses associated with bats, including the Arizona skunk (AZ SK)
229 variant recently derived from a bat reservoir (Leslie et al., 2006; Kuzmin et al., 2012), and
230 terrestrial species respectively with sub-division of the latter into clades defining the raccoon,
231 central Mexican skunk and the US SCSK variants. Phylogenetic analyses using a smaller dataset
232 of G gene sequences of representative viruses generated similar trees (data not shown).

233 *3.3. Substitution patterns for N and G genes*

234 The nucleotide substitution patterns observed for both genes were examined using data
235 from representative skunk-associated viruses but excluding the Arizona skunk variant. Overall
236 transition/transversion ratios were high at 5.0403 (N gene) and 4.6102 (G gene). N gene dN/dS
237 ratios for each variant within the dataset ranged from 0.0357 (SCSK) to 0.0925 (SBC skunk)
238 with other variants yielding intermediate values (Table 1). It is unclear if the relatively high
239 value for the SBC skunk variant is significant or if it is a consequence of the limited numbers of
240 isolates of this type examined. Values for dN/dS ratios for the G gene tended to be 2-3 times
241 greater due to a higher level of non-synonymous substitutions at this less conserved locus. These
242 values overwhelmingly supported the operation of purifying selection on these genes ($p = 0.0000$
243 for all groups analysed using the Kumar method in MEGA) rather than neutral or positive
244 selective evolutionary forces.

245 *3.4. Coding differences within the N gene*

246 Conversion of the nucleotide sequence data generated from the various skunk variants in
247 this study to deduced N protein coding sequences enabled a detailed comparison of this viral
248 protein. For all skunk-associated viral variants, excluding the Arizona skunk variant, pairwise N
249 protein distance values ranged from 0 to 0.08, corresponding to amino acid differences ranging
250 between 0 and 34. An alignment of these nucleoprotein sequences identified many highly
251 conserved residues as well as some variable positions that appear to reflect the phylogenetic
252 relationships between these variants. Figure S1 (Supplemental data) shows such an alignment
253 using representatives of all the skunk variants together with sequences from a few other viruses
254 that circulate in other reservoir hosts as illustrated in Figure 3. The most notable coding
255 differences found within the skunk-associated rabies variants are identified in Figure S1 and
256 discussed further below.

257 *3.5. Coding differences within the G gene*

258 Analysis of the predicted glycoprotein for all skunk-associated viral types, excluding the
259 Arizona skunk variant, indicated pairwise distance values ranging from 0 to 0.17, corresponding
260 to amino acid differences of between 0 to 79. A glycoprotein sequence alignment of skunk-
261 associated rabies viruses together with representatives of other rabies virus variants that circulate
262 in the central US (Figure S2) identified the 19 amino acid N-terminal signal peptide and the
263 hydrophobic trans-membrane domain (amino acids 439-461) as areas of high variability as
264 previously documented (Badrane et al., 2001). Differences between members of the
265 cosmopolitan lineage and the SCSK variant viruses were especially pronounced across these
266 regions. However substitutions at particular positions were retained within some variants and
267 viral types; the more notable are illustrated in Figure S2 and discussed further below.

268 **4. Discussion**

269 Detailed molecular epidemiological studies of rabies viruses are increasingly providing
270 insights into the emergence, history and transmission dynamics of rabies enzootics and
271 epizootics (Holmes et al., 2002; Bourhy et al., 2008; Talbi et al., 2010). Since rabies has become
272 entrenched within multiple wildlife species within North America, an understanding of how the
273 virus was introduced and then spread within each host population may hold the key to control
274 and eventual elimination of this disease. Rabies viruses within skunks have been noted for many
275 decades and although phylogenetic studies have helped to trace the historical origins of the
276 variants associated with this host (Velaco-Villa et al., 2008), a detailed phylogeographic study of
277 the genomic diversity of skunk rabies variants within the central United States has not been
278 published previously. This study has explored the diversity of rabies viruses associated with
279 skunks across this region in which both NCSK and SCSK variants were found.

280 *4.1. SCSK rabies viruses*

281 Rabies case records suggest that the current SCSK variant emerged from a focus of skunk
282 rabies cases in Texas in the mid 1950s, followed by subsequent spread of this outbreak
283 throughout the southern Great-Plains region of the United States. While isolates of the SCSK
284 variant virus recovered from the states of Arkansas, Oklahoma, Missouri, Kansas and Nebraska
285 are not differentiated by classical antigenic methods, they do exhibit marked genetic diversity
286 that allows their sub-division into three major viral clades (Figure 1). Significant clustering of
287 isolates according to the state of submission was observed (Figure 2).

288 All viruses from across much of the states of Kansas and Nebraska collectively form the
289 largest monophyletic grouping designated as clade III. Moreover, there is a well defined division

290 between eight isolates labeled as IIIA which come primarily from eastern Kansas and the rest of
291 the samples (IIIB) of this clade. The identification of two viruses recovered from Colorado
292 (COSK090005) and Missouri (MOSK090041) which also group within clade IIIB suggest that
293 this type is responsible for the recent expansion of SCSK into these states, in particular in
294 Colorado where the epizootic has reached as far as the foothills of the Rocky Mountains. Isolate
295 OKBV090073 within this group is reported as occurring approximately 150 miles away from the
296 nearest isolate of the same type. The bovine host of this case had no history of travel so this viral
297 isolate likely represents a southern extension of type IIIB.

298 Clade II, comprised of samples from Arkansas and all but one of the samples from
299 Missouri, is also divided into two types (IIA and IIB) with strong support. The Arkansas viruses
300 of type IIB originate from the southern half of the state while the type IIA specimens originate
301 from the northern half of Arkansas and Missouri; the limited genetic variation of all samples
302 from Missouri, illustrated by their tight clustering within a branch of a much larger clade
303 dispersed across Arkansas, is consistent with spread of the disease northwards from neighboring
304 Arkansas. The distinct ranges of these two types are separated by the Arkansas River which may
305 serve as a barrier for transmission of viruses of this clade.

306 A small group of viruses, all originating exclusively from Oklahoma, comprised the
307 outlying clade I of the SCSK variant. This was the only variant recovered from this state with the
308 exception of specimen OKBV090073 that appears to represent an incursion of clade IIIB. Some
309 historical samples from Texas and New Mexico included in a broader phylogenetic analysis also
310 clustered as outliers of the SCSK variant, while additional samples from Texas segregated on
311 distinct branches within other parts of the SCSK clade (Figure 3). The oldest characterized virus

312 of the SCSK variant, from a skunk recovered in Texas in 1968, appears to have a common
313 ancestry with clade III.

314 While the geographically distinct ranges of SCSK types are a unique finding of the
315 present study this was unexpected since a recent study of striped skunks within the Midwestern
316 United States showed that gene flow is high between animals from the Dakotas through to
317 Oklahoma (Barton et al., 2010). Accordingly admixture of viral biotypes should be observed.
318 However, the molecular epidemiology described in the present study, albeit from a limited time
319 period, seems to show that such mixing of viral types is not the case. These findings may
320 indicate that the viral variants move across the landscape in wave-fronts with localized genetic
321 drift leading to emergence of sub-types in particular areas. Alternatively it might indicate
322 adaptation of various biotypes to distinct habitats across the region although to date no evidence
323 for such evolutionary factors exists. As this sample set has demonstrated only a snapshot of sub-
324 type distribution across the landscape further retrospective and prospective studies will explore
325 temporal changes to this pattern and help to better understand the contributing factors.

326 *4.2. NCSK rabies viruses*

327 The epizootic due to the NCSK variant was first recognized in the late 1940s; Missouri reported
328 28 cases of rabies within skunks in 1959, apparently due to spread of an epizootic front moving
329 south from the Dakotas (Parker, 1975). All South Dakota and Minnesota terrestrial isolates
330 examined in this study were exclusively of the NCSK variant. They clustered closely with the
331 Canadian samples recovered from the provinces of Saskatchewan and Manitoba that are located
332 directly north of the states of Montana, North Dakota and Minnesota. Indeed this sample set was
333 relatively homogeneous with no strongly supported phylogenetic structure, indicating that the

334 virus in this border region has probably experienced little impediment to its spread across the
335 landscape. Especially notable is the limited variation observed within the Canadian specimens
336 despite inclusion of samples recovered over a 13 year period. However, viruses from further
337 afield exhibited greater diversity. For example, sample ARDG090042, recovered from a dog in
338 north-eastern Arkansas in 2009, was an outlier to this group based on analysis of both N and G
339 gene sequences, as were samples from Kentucky and Wisconsin (Figure 3). Indeed previous
340 samples from this region of north-eastern Arkansas and south-eastern Missouri show a pocket of
341 NCSK circulating among skunks in the region (unpublished data). Current national data (Blanton
342 et al., 2010) show a southern extension of this variant within central Kentucky and Tennessee but
343 not as far west as Arkansas and Missouri. Regardless, the single Arkansas isolate evaluated here
344 is not from an area contiguous with the more northern regions affected by this viral variant and
345 its divergence is thus not unexpected. Analysis of more samples from states where this rabies
346 virus variant occurs is needed to complete our understanding of its range and genetic diversity.

347 *4.3. CASK rabies viruses*

348 Prior studies on rabies in California have used antigenic typing tools and genetic methods
349 based on PCR and restriction endonuclease analysis to explore the diversity of the virus
350 circulating in terrestrial species in the state (Crawford-Miksza et al., 1999). Subsequently
351 nucleotide sequence analysis has been undertaken on a limited number of isolates (Velasco-Villa
352 et al., 2008) but without the benefit of detailed spatial information on the source of those isolates
353 so as to allow correlation with the earlier studies. This study genetically characterized a small
354 set of viruses from terrestrial species from different regions of California to allow comparison
355 with other skunk-associated viruses and to explore their regional variation. Only the northern
356 half of the state was represented in this sample set (see Figure 4) since skunk rabies is rarely if

357 ever reported in the southern counties (Crawford-Miksza et al., 1999). All California viruses
358 formed a monophyletic clade (CASK) which can readily be sub-divided into three types that
359 exhibit geographical localization. Furthermore the identification of several amino acid coding
360 differences between these viruses support the conclusion that these three types represent some of
361 the discrete antigenic types proposed previously (Crawford-Miksza et al., 1999). CASK type a
362 comprises isolates from Mariposa county (San Joachim valley variant); type b consists of
363 specimens from the north-eastern region of the state including the Sonoma/North coast regions
364 and some inland areas (Trinity and Yolo counties); type c from Glenn, Sutter, Colusa and
365 Amador counties corresponds to the Sacramento Valley variant which was previously described
366 as being particularly distinctive with respect to its monoclonal antibody binding pattern
367 (Crawford-Miksza et al., 1999).

368 *4.4. Mexican skunk rabies viruses*

369 Included in our analysis are seven rabies isolates from Mexico that segregate into two
370 discrete clades representing variants localized to South Baja California (SBC skunk) and central
371 Mexico, also referred to elsewhere as the MEXSK-2 and MEXSK-1 variants respectively
372 (Velasco-Villa et al., 2008). These variants circulate predominantly in spotted skunks and
373 possibly also hog-nosed skunks. The SBC skunk variant is closely related to the CASK variant,
374 perhaps not surprising given the geographical proximity of the areas where they circulate, while
375 the central Mexican variant clusters as an outlying group to both the SCSK and raccoon strains.
376 The predominant role of spotted skunks in this enzootic may be significant. During the mid 19th
377 century when rabies transmitted by skunks, or *Rabies Mephitica* as it was designated (Hovey,
378 1874), was common in Kansas and Colorado, both the spotted skunk and the striped skunk were
379 responsible. While historically the spotted skunk apparently played a significant role in disease

380 transmission, today this species is nearly extirpated from most of its historical range in the
381 central United States and all of the recent US isolates detailed here are from striped skunks. We
382 speculate that the viruses circulating currently in central Mexican spotted skunks are remnants of
383 the virus that predominated in the Great Plains over 150 years prior and gave rise to the SCSK
384 variant. Moreover, the position of the raccoon variant within this cluster of skunk-associated
385 viruses is consistent with the hypothesis that the raccoon strain emerged after a host shift from a
386 skunk-associated virus rather than directly from a bat reservoir. Indeed all bat-associated viral
387 variants appear to group well outside of the cluster of viruses associated with terrestrial hosts. In
388 contrast, the recently emerged Arizona skunk variant, known to have arisen by host shift events
389 from a bat reservoir, clusters closely with the responsible big brown bat variant (Leslie et al.,
390 2006; Kuzmin et al., 2012).

391 *4.5. Viral evolutionary processes*

392 The database of N and G gene sequences generated in this study showed that the patterns
393 of nucleotide substitution exhibited by skunk-associated rabies viruses are similar to those
394 observed in prior studies on lyssavirus diversity in general (Bourhy et al., 2008; Delmas et al.,
395 2008). Changes are predominantly synonymous in nature and most nonsynonymous mutations
396 result in very conservative amino acid substitutions. However, this study did identify some
397 amino acid substitutions that are associated with particular viral variants or clades. Within the N
398 gene amino acid replacements of particular note are at the following positions: residues specific
399 to all members of the SCSK variant occur at positions 3 (Thr in place of Ala), 93 (Asp in place
400 of Gly), and 448 (Asn in place of Ser); replacement of Met by Leu at position 126 in SCSK and
401 the Central Mexican variants; differences specific to the SCSK III variant at residue 209 (Ala in
402 place of Thr), to the SCSK IIIB variant at amino acid 135 (Gln in place of Pro), and at position

403 182 (Ile in place of Val) for both SCSK variants II and III. Several amino acids were represented
404 at residue 40 with either Cys or Ser predominating in most variants while Gly was restricted to
405 SCSK IIa and III; distinct substitutions at residues 254 (Lys) and 428 (Gly) observed in the
406 SCSK IIb variant further reinforced the distinctive nature of this viral group. The SBC skunk
407 viruses exhibited distinctive amino acid residues at positions 11 (Tyr), 36 (Ser), 84 (Ile) and
408 407(Ala) while the Central Mexican viruses were distinctive at residues 99 (Gln), 181 (Val) and
409 388 (Asp). Interestingly, with just a few exceptions, several coding differences between the
410 CASK variants were observed. Specific residues were associated with variants "a" at positions
411 40 (Phe), 128 (Met) and 410 (Ile), with variants "b" at residues 9 (Arg except for V640) and 84
412 (Pro) and with variants "c" at positions 13 (His), 202 (Ser) and 255 (Asp) while a His residue
413 was located at position 369 for variants "b" and "c", a substitution also shared with samples
414 V212TXSK and 2311Mxzacsk01 (central Mexican skunk). As a group, the NCSK samples
415 exhibited much less variability within the nucleoprotein with the exception of the more outlying
416 samples such as ARDG090042, 421Kydg07 and 3789SK. These three samples individually and
417 collectively exhibited a number of amino acid replacements (e.g. at residues 126 and 433)
418 compared to other members of the NCSK clade. However many of these substitutions were also
419 observed in viruses of the other clades examined suggesting that these positions had sufficient
420 flexibility to allow genetic drift. The Arizona skunk variant retained the coding capability of the
421 original bat variant and there was no evidence of adaptation to a skunk variant within the N
422 protein sequence.

423 Within the G gene the following substitutions are of particular note: at amino acids 37-
424 39, a glycosylation site within antigenic site II was conserved in all members of the cosmopolitan
425 lineage but not in many SCSK isolates nor in any bat-associated variants; moreover a Ser-Thr

426 substitution at position 201 also within antigenic site II was found in all members of the SCSK
427 III clade; the core sequence of the conserved linear G5 epitope, "HDFR" at residues 261-264
428 (Cai et al., 2010), was substituted to HDLH in all SCSK variants and indeed in this sample set
429 only the first two amino acids of this epitope were conserved. In contrast the linear epitope
430 "WXXXDI" at residues 14-19 (Mansfield et al., 2004), represented here as WSPIDI, was highly
431 conserved together with its flanking sequences. While most skunk-associated viruses were
432 conserved across antigenic site III (residues 330-338), some variation occurred around this site,
433 most notably in bat-associated variants where residue 333 was either the more common Arg or in
434 some cases Lys; however, either residue at this site maintains viral virulence (Tuffereau et al.,
435 1989). The distinctive Arkansas sample (ARDG090042) had 10 unique substitutions in the
436 glycoprotein compared to other US members of the NCSK variant. Several other substitutions
437 observed only in particular viral variants are identified in Figure S2. As noted for the
438 nucleoprotein analysis, the viruses of the Arizona skunk variant exhibit glycoprotein sequences
439 highly characteristic of the bat viral variant from which they are derived and are quite distinctive
440 from the SCSK viruses described in this report.

441 These changes appear most likely to have arisen through chance mutation followed by
442 fixation in the absence of selective pressure. No amino acids specific to the skunk-associated
443 viruses and distinctive from the viruses of other hosts were identified in this study in either the N
444 or G proteins and it would thus appear that these products are unlikely to harbor residues that
445 confer host specificity, a finding consistent with that reported by others (Velasco-Villa et al.,
446 2008). Indeed, it has been proposed that high levels of synonymous substitutions may allow
447 rabies virus to become "pre-adapted to replicate in a wide range of species" (Holmes et al., 2002;
448 Gordon et al., 2004; Velasco-Villa et al., 2008). However, the variation in these gene and protein

449 sequences documented here may have practical utility for the development of virus sub-typing
450 methods employing either monoclonal antibodies or molecular-based methods directed to these
451 variable sites. Such tools would allow variant and sub-type tracking to facilitate further
452 epidemiologic analysis of the spread of these viruses across the landscape. Moreover, given that
453 the striped skunk appears to be relatively susceptible to being infected by rabies, as well as
454 permissive to maintaining the virus as a reservoir host, continued epidemiologic vigilance is
455 warranted to permit early detection of future host jumps of other rabies virus variants into this
456 species.

457 *4.6. Conclusions*

458 In summary, this study has shown that rabies virus variants associated with North American
459 skunk populations have emerged from two distinct ancestral sources and are subject to purifying
460 selection which significantly restricts genetic drift. There is no evidence that viruses associated
461 with this host bear specific and unique amino acid residues in either the N or G proteins but
462 genetic signatures of both the CASK and SCSK variants identify sub-types with discrete
463 geographical ranges. Such information provides a baseline for subsequent molecular
464 epidemiological studies exploring the direction and speed of spread of the virus across the
465 landscape, information that may facilitate any future control efforts against this disease.

466 **References**

467 Aranda, M. and Lopez-de Buen, L., 1999. Rabies in skunks from Mexico. J. Wildl. Dis. 35, 574-
468 577.

469 Badrane, H., Bahloul, C., Perrin, P., Tordo, N., 2001. Evidence of two lyssavirus phylogroups
470 with distinct pathogenicity and immunogenicity. J. Virol. 75, 3268-3276.

471 Barton, H.D., Gregory, A.J., Davis, R., Hanlon, C.A., Wisely, S.M., 2010. Contrasting landscape
472 epidemiology of two sympatric rabies virus strains. Mol. Ecol. 19, 2725-2738.

473 Blanton, J.D., Palmer, D., Rupprecht, C.E., 2010. Rabies surveillance in the United States during
474 2009. J. Am. Vet. Med. Assoc. 237, 646-657.

475 Blanton, J.D., Robertson, K., Palmer, D., Rupprecht, C.E., 2009. Rabies surveillance in the
476 United States during 2008. J.Am.Vet.Med.Assoc. 235, 676-689.

477 Bourhy, H., Reynes, J.-M., Dunham, E.J., Dacheux, L., Larrous, F., Huong, V.T.Q., Xu, G., Yan,
478 J., Miranda, M.E.G., Holmes, E.C., 2008. The origin and phylogeography of dog rabies
479 virus. J. Gen. Virol. 8, 2673-2681.

480 Cai, K., Feng, J.-N., Wang, Q., Li, T., Shi, J., Hou, X.-J., Gao, X., Liu, H., Tu, W., Xiao, L.,
481 Wang, H., 2010. Fine mapping and interaction analysis of a linear rabies virus neutralizing
482 epitope. Microbes Infect. 12, 948-955.

483 Charlton K.M., Webster, W.A., Casey, G.A., 1975. Skunk rabies. In: Baer, G.M. (Ed.), The
484 Natural History of Rabies, Vol. 2. Academic Press, New York, pp. 307-324.

485 Crawford-Miksza, L.K., Wadford, D.A., Schnurr, D.P., 1999. Molecular epidemiology of
486 enzootic rabies in California. J. Clin. Virol. 14, 207-219.

487 Delmas, O., Holmes, E.C., Talbi, C., Larrous, F., Dachent, L., Bouchier, C., Bourhy, H., 2008.
488 Genomic diversity and evolution of the lyssaviruses. PLoS One 3(4), e2057.

489 Eng, T.R., Hamaker, T.A., Dobbins, J.G., Tong, T.C., Bryson, J.H., Pinsky, P.F., 1989. Rabies
490 surveillance, United States, 1988. MMWR CDC Surveill Summ. 38, 1-21.

491 Gordon, E.R., Curns, A.T., Krebs, J.W., Rupprecht, C.E., Real, L.A., Childs, J.E., 2004.
492 Temporal dynamics of rabies in a wildlife host and the risk of cross-species transmission.
493 Epidemiol. Infect. 132, 515-524.

494 Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New
495 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the
496 Performance of PhyML 3.0. Systematic Biol. 59, 307-21.

497 Hall, T.A., 2011. BioEdit: a user-friendly biological sequence alignment editor and analysis
498 program for Windows 95/98/NT. Available at
499 http://www.mbio.ncsu.edu/BioEdit/page2.html. Accessed August 14, 2012.

500 Hanlon, C., Niezgoda, M., Rupprecht, C. 2007. Rabies in Terrestrial Animals, In: Jackson, A.C.,
501 Wunner, W. (Eds.), Rabies. 2nd ed. Academic Press, London, UK, pp. 201-258.

502 Hass, C.C., Dragoo, J.W., 2006. Rabies in hooded and striped skunks in Arizona. J. Wildl. Dis.
503 42, 825-829.

504 Holmes, E.C., Woelk, C.H., Kassis, R., Bourhy, H., 2002. Genetic constraints and the adaptive
505 evolution of rabies virus in nature. Virology 292, 247-257.

506 Hovey, H.C., 1874. Rabies Mephitica. Am. J. Sci. Arts 7, 478-483.

507 International Committee on Taxonomy of Viruses. 2011. ICTV Files and Discussions. ICTV
508 2011 Master Species List – Version 1, February 21, 2012. Available at:
509 http://talk.ictvonline.org/files/ictv_documents/m/msl/4090.aspx. Accessed March 7, 2012.

510 Jackson, A.C.**,** 2007. Pathogenesis. In: Jackson, A.C., Wunner, W. (Eds.), Rabies. Academic
511 Press, London, UK, p. 341-381.

512 Krebs, J.W., Smith, J.S., Rupprecht, C.E., Childs, J.E., 2000. Mammalian reservoirs and
513 epidemiology of rabies diagnosed in human beings in the United States, 1981-1998. Ann.
514 N. Y. Acad. Sci. 916, 345-353.

515 Kuzmin, I.V., Hughes, G.J., Botvinkin, A.D., Gribencha, S.G., Rupprecht, C.E., 2008. Arctic and
516 Arctic-like rabies viruses: distribution, phylogeny and evolutionary history. Epidemiol.
517 Infect. 136, 509–519.

518 Kuzmin, I.V., Mayer, A.E., Niezgoda, M., Markotter, W., Agwanda, B., Breiman, R.F.,
519 Rupprecht, C.E., 2010. Shimoni bat virus, a new representative of the Lyssavirus genus.
520 Virus Res. 149, 197-210.

521 Kuzmin, I.V., Shi, M., Orciari, L.A., Yager, P.A., Velasco-Villa, A., Kuzmina, N.A., Streicker,
522 D.G., Bergman, D.L., Rupprecht, C.E., 2012. Molecular inferences suggest multiple hosts
523 shifts of rabies viruses from bats to mesocarnivores in Arizona during 2001-2009. PLoS
524 Pathogens 8(6), e1002786.

525 Leslie, M.J., Messenger, S., Rohde, R.E., Smith, J., Cheshier, R., Hanlon, C., Rupprecht, C.E.,
526 2006. Bat-associated rabies virus in Skunks. Emerg. Infect. Dis. 12, 1274-1277.

527 Mansfield, K.L., Johnson, N., Fooks, A.R., 2004. Identification of a conserved linear epitope at
528 the N terminus of the rabies virus glycoprotein. J. Gen. Virol. 85, 3279-3283.

529 Nadin-Davis, S.A., 1998. Polymerase chain reaction protocols for rabies virus discrimination. J.
530 Virol. Methods 75, 1-8.

531 Nadin-Davis, S.A., Loza-Rubio, E., 2006. The molecular epidemiology of rabies associated with
532 chiropteran hosts in Mexico. Virus Res. 117, 215-226.

533 Nadin-Davis, S.A., Abdel-Malik, M., Armstrong, J., Wandeler, A.I., 2002. Lyssavirus P gene
534 characterisation provides insights into the phylogeny of the genus and identifies structural
535 similarities and diversity within the encoded phosphoprotein. Virology 298, 286-305.

536 Nadin-Davis, S.A., Huang, W., Wandeler, A.I., 1997. Polymorphism of rabies viruses within the
537 phosphoprotein and matrix protein genes. Arch.Virol. 142, 979-992.

538 Nadin-Davis, S.A., Sampath, M.I., Casey, G.A., Tinline, R.R., Wandeler, A.I., 1999.
539 Phylogeographic patterns exhibited by Ontario rabies virus variants. Epidemiol. Infect.
540 123, 325-336.

541 Nadin-Davis, S.A., Muldoon, F., Wandeler, A.I., 2006. Persistence of genetic variants of the
542 arctic fox strain of Rabies virus in southern Ontario. Can. J. Vet. Res. 70, 11-19.

543 Nadin-Davis, S.A., Turner, G., Paul, J.P.V., Madhusudana, S.N., Wandeler, A. I., 2007.
544 Emergence of Arctic-like rabies lineage in India. Emerg. Infect. Dis. 13, 111-116.

545 Nadin-Davis, S.A., Sheen, M., Wandeler, A.I., 2012. Recent emergence of the Arctic rabies virus
546 lineage. Virus Res. 163, 352–362.

547 Oertli, E.H., Wilson, P.J., Hunt, P.R., Sidwa, T.J., Rohde, R.E., 2009. Epidemiology of rabies in
548 skunks in Texas. J. Am. Vet. Med. Assoc. 234, 616-620.

549 Parker, R.L. 1975. Rabies in Skunks, In: Baer, G.M. (Ed.), The Natural History of Rabies, Vol.
550 2. Academic Press, New York, pp. 41-51.

551 Pool, G.E., Hacker, C.S., 1982. Geographic and seasonal distribution of rabies in skunks, foxes
552 and bats in Texas. J. Wildl. Dis. 18, 405-418.

553 Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution.
554 Bioinformatics 14, 817-818.

555 Price, E.R., Blenden, D.C., Logue, J.T., 1961. Rabies in Missouri 1950-1960. Mol. Med. 58,
556 460-466.

557 Rupprecht, C.E., Smith, J.S., Fekadu, M.,, Childs, J.E., 1995. The ascension of wildlife rabies: a
558 cause for public health concern or intervention? Emerg. Infect. Dis. 1, 107-114.

559 Smith, J.S., Reid-Sanden, F.L., Roumillat, L.F., Trimarchi, C., Clark, K., Baer, G.M., Winkler,
560 W.G., 1986. Demonstration of antigenic variation among rabies virus isolates by using
561 monoclonal antibodies to nucleocapsid proteins. J. Clin. Microbiol. 24, 573-580.

562 Steck, F., Wandeler, A., 1980. The epidemiology of fox rabies in Europe. Epidemiol. Rev. 2, 71-
563 96.

564 Talbi, C., Lemey, P., Suchard, M.A., Abdelatif, E., Elharrak, M., Jalal, N., Faouzi, A.,
565 Echevarría, J.E., Morón, S.V., Rambaut, A., Campiz, N., Tatem, A.J., Holmes, E.C.,
566 Bourhy, H., 2010. Phylodynamics and human-mediated dispersal of a zoonotic virus. PLoS
567 Path. 610: e1001166.

568 Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics
569 Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.

570 Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The
571 CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided
572 by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.

573 Tinline, R.R., MacInnes, C.D., 2004. Ecogeographic patterns of rabies in southern Ontario based
574 on time series analysis. J. Wildl. Dis. 40, 212-221.

575 Tuffereau, C., Leblois, H., Bénéjean, J., Coulon, P., Lafay, F., Flamand, A., 1989. Arginine or
576 lysine in position 333 of ERA and CVS glycoprotein is necessary for rabies virulence in
577 adult mice. Virology 172, 206-212.

578 Velasco-Villa, A., Gomez-Sierra, M., Hernandez-Rodriguez, G., Juarez-Islas, V., Melendez-
579 Felix, A., Vargas-Pino, F., Velazquez-Monroy, O., Flisser, A., 2002. Antigenic diversity
580 and distribution of rabies virus in Mexico. J. Clin. Microbiol. 40, 951-958.

581 Velasco-Villa, A., Reeder, S.A., Orciari, L.A., Yager, P.A., Franka, R., Blanton, J. D., Zuckero,
582 L., Hunt, P., Oertli, E.H., Robinson, L.E., Rupprecht, C.E., 2008. Enzootic rabies
583 elimination from dogs and reemergence in wild terrestrial carnivores, United States.
584 Emerg. Infect. Dis. 14, 1849-1854.

585 Wunner, W., 2007. Rabies Virus. In: Jackson, A.C., Wunner, W. (Eds.), Rabies. Academic Press,
586 London, UK, pp. 23-68.

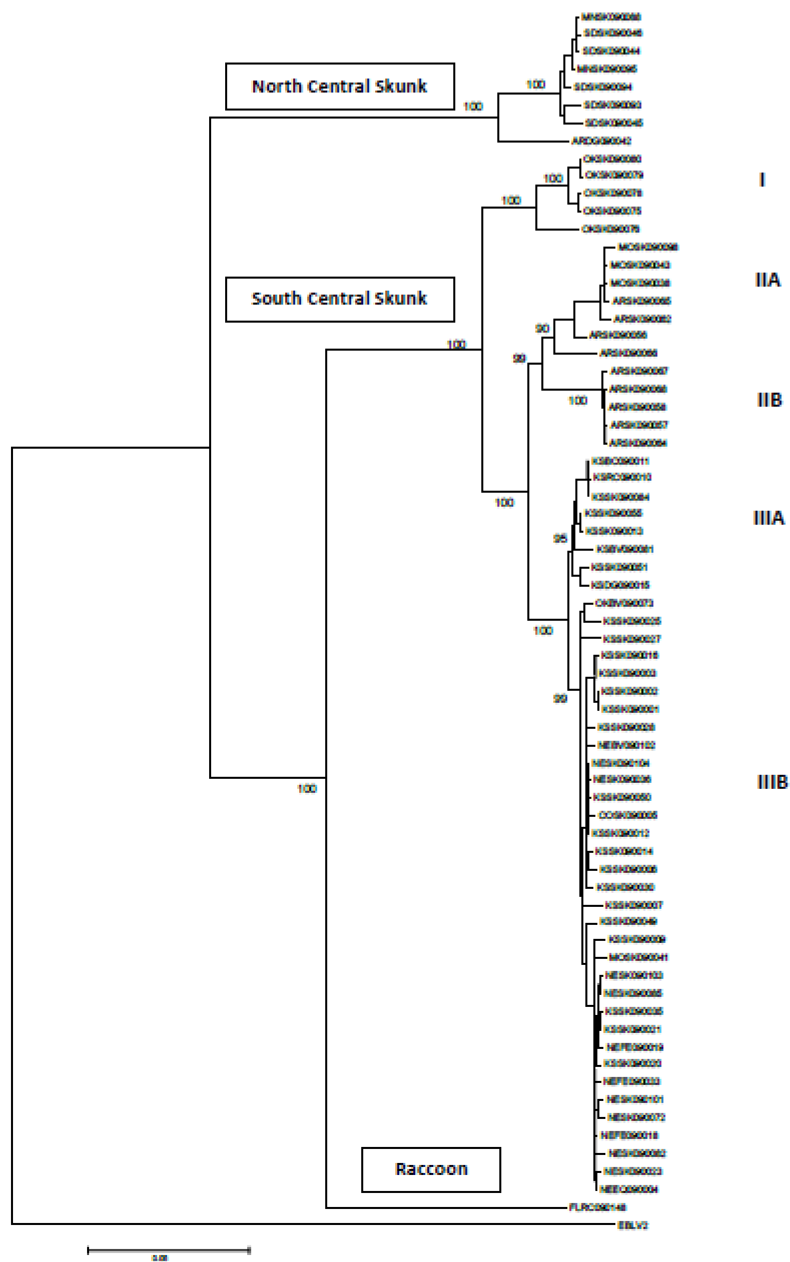
587

588

589

590 Figure legends

591 <u>Figure 1</u>

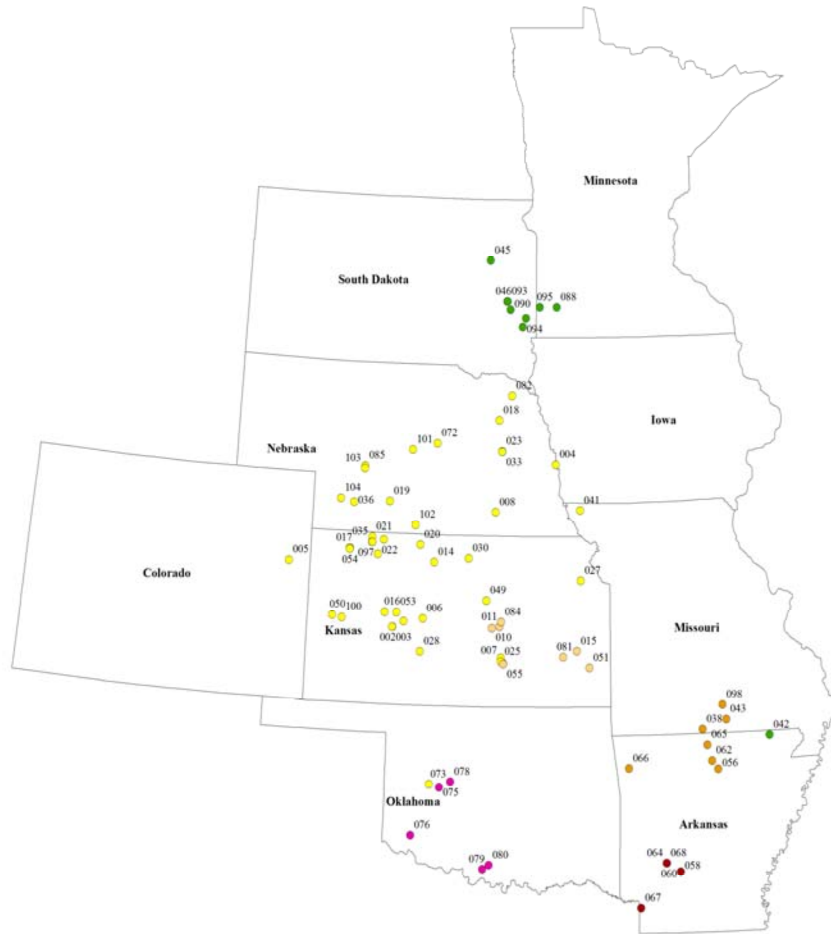


592

593 Figure 1. A phylogenetic tree of rabies virus sequences generated from 67 specimens infected
594 with skunk variants. A NJ analysis was performed on concatenated N and G gene sequences;
595 corresponding sequences for one raccoon virus variant and an EBLV2 isolate were included as
596 outgroups. Bootstrap values for major branch points are shown within the tree. The names of
597 the main variants are shown in boxes while the designations of the different clades and types of
598 the SCSK variant as described in the text are provided to the right of the tree. A distance scale is
599 shown at bottom left.

600 Figure 2

601



602

603 Figure2. Spatial distribution of all skunk variant isolates from the US Midwest examined in this study.
604 Using ZIP code information each sample was mapped with a pin color coded based on the clade or type
605 in which it clustered in the phylogenetic tree (Figure 1). Green pins represent the NCSK variant, while
606 the SCSK variant clades are represented by pink (SCSK I), orange (SCSK IIA), red (SCSK IIB), peach (SCSK
607 IIIA) and yellow (SCSK IIIB). Samples are designated with the final 3 digits as detailed in Table S1.
608

Figure 3

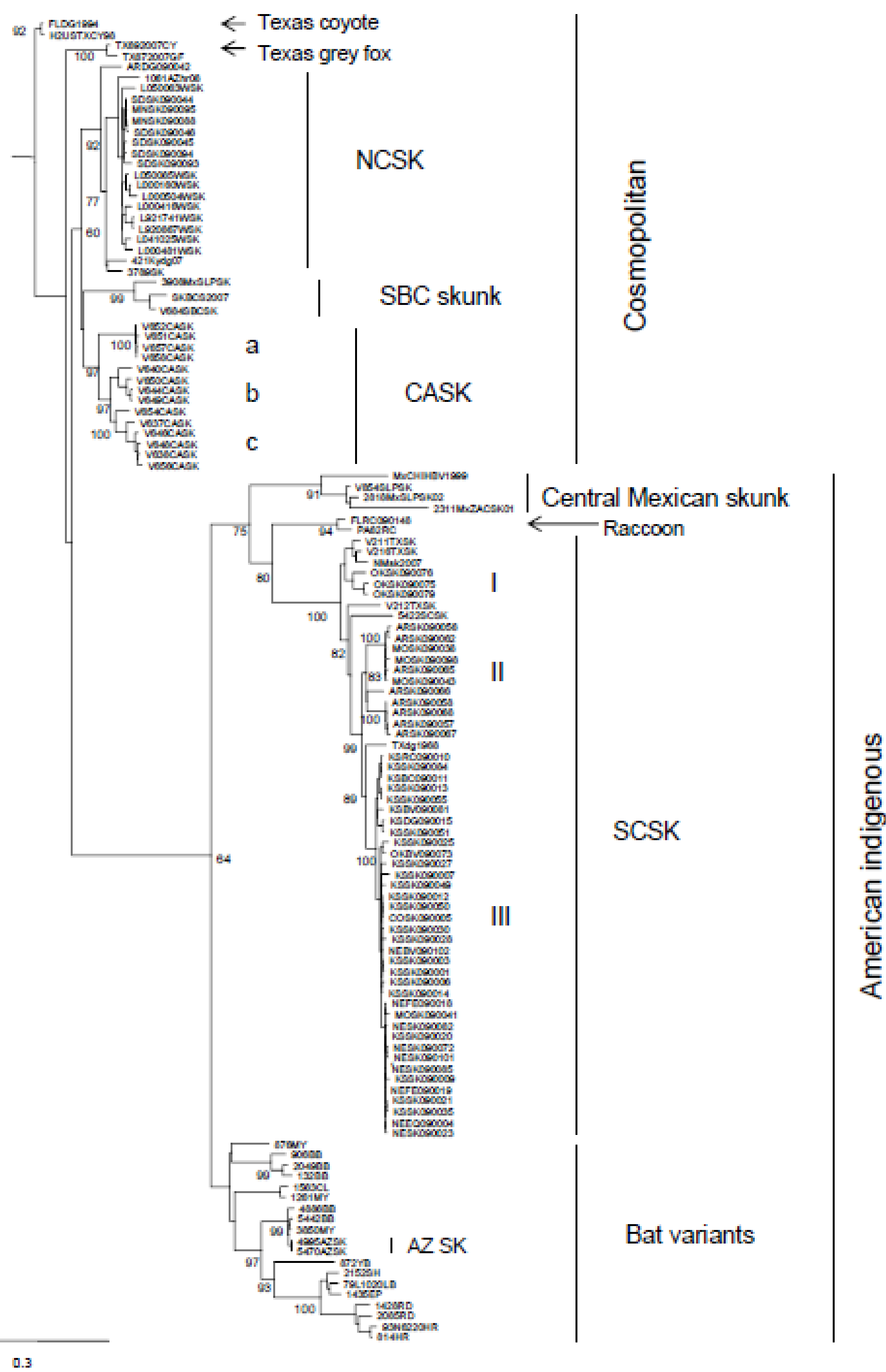
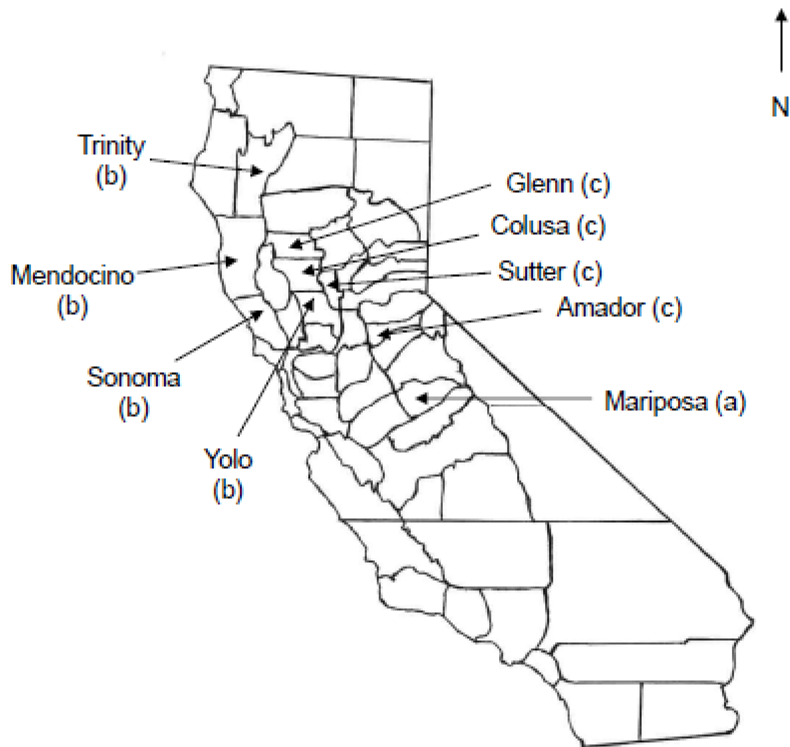


Figure 3. A maximum likelihood analysis of N gene sequences of North America skunk-associated rabies viruses and representative isolates from other sympatric viral variants. The phylogenetic tree is rooted to an EBLV2 outgroup (not shown). Bootstrap values for major branch points are shown in the tree. The lineage variant type and type designations are shown to the right of the tree. A distance scale is shown at bottom left.

616

Figure 4



617

618 Figure 4. Map of the state of California showing the counties from which CASK rabies virus
619 variant isolates were characterized in this study. The viral types (a, b and c) identified by
620 phylogenetic analysis are indicated after the county name.

621

622 <u>Table 1</u>

Table 1. Patterns of synonymous and nonsynonymous nucleotide differences between skunk-associated rabies viruses

	N gene				G gene			
	No. of samples in group	dS	dN	dN/dS	No. of samples in group	dS	dN	dN/dS
NCSK	20	0.0649	0.0041	0.0632	14	0.0645	0.0085	0.1318
CASK	14	0.1399	0.0064	0.0457	2	0.2740	0.0352	0.1285
SCSK	56	0.1037	0.0037	0.0357	58	0.0922	0.0108	0.1171
Central Mexico SK	4	0.2295	0.0142	0.0619	1	N/A	N/A	N/A
SBC SK	3	0.0951	0.0088	0.0925	1	N/A	N/A	N/A
All variants	97	0.5392	0.0144	0.0267	76	0.4529	0.0341	0.0753