1   Article (Discoveries)

2

3   Coordinated evolution between N2 neuraminidase and H1 and H3 hemagglutinin genes

4   increased influenza A virus genetic diversity in swine

5

6   Michael A. Zeller,[ab] Jennifer Chang,[c] Amy L. Vincent,[c] Phillip C. Gauger,[a] Tavis K. Anderson[c,#]

7

8   [a]Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary

9   Medicine, Iowa State University, Ames, Iowa, USA;

10   [b]Bioinformatics and Computational Biology Program, Iowa State University, Ames, Iowa, USA;

11   [c]Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames, Iowa,

12   USA.

13

14   [#]Address correspondence to: Tavis K. Anderson, e-mail: tavis.anderson@usda.gov

15

16

17 **ABSTRACT**

18 The neuraminidase (NA) and hemagglutinin (HA) of influenza A virus (IAV) are essential

19 surface glycoproteins. In this study, the evolution of subtype N2 NA paired with H1 and H3

20 subtype HA in swine was evaluated to understand if genetic diversity of HA and NA were

21 linked. Using time-scaled Bayesian phylodynamic analyses, the relationships of paired swine N2

22 with H1 or H3 from 2009 to 2018 were evaluated. These data demonstrated increased relative

23 genetic diversity within the major N2 clades circulating in swine (N2.1998 between 2014-2017

24 and N2.2002 between 2010-2016). Relative genetic diversity of NA-HA pairs (e.g., N2.1998B/

25 H1.Delta1B) were correlated, suggesting intergene epistasis. Preferential pairing was observed

26 among specific NA and HA genetic clades and this was associated with gene reassortment

27 between cocirculating influenza A strains. Using the phylogenetic topology of inferred N2 trees,

28 the expansion of genetic diversity in the NA gene was quantified and increases in diversity were

29 observed subsequent to NA-HA reassortment events. The rate of evolution among NA-N2 clades

30 and HA-H1 and HA-H3 clades were similar. The frequent regional movement of pigs and their

31 influenza viruses is a possible explanation driving this pattern of drift, reassortment, and rapid

32 evolution. Bayesian phylodynamic analyses demonstrated strong spatial patterns in N2 genetic

33 diversity, and that frequent interstate movement of N2 clades homogenized diversity. The

34 reassortment and evolution of NA and its influence on HA evolution may affect antigenic drift,

35 impacting vaccine control programs and animal health.

36 **Key-words:** epistasis, genomic epidemiology, influenza A virus, swine, vaccine

37   **INTRODUCTION**

38   Influenza A virus (IAV) is an important swine respiratory pathogen with high economic

39   consequences in commercial swine production systems, with a risk for zoonotic transmission

40   (Vincent, Ma, et al. 2008; Dykhuis-Haden, et al. 2012). Clinical signs of IAV in swine include

41   cough, fever, lethargy, and a reduction of appetite that may lead to weight loss. IAV induced

42   lung pathology often predisposes to secondary bacterial infections, resulting in further

43   production losses and increased risk for mortality (Gramer 2006; Wang, et al. 2013). Due to the

44   high morbidity associated with clinical disease, prevention and control of IAV is necessary to

45   minimize animal suffering, mitigate production loss, and protect public health.

46   Vaccines have historically been formulated based on the hemagglutinin (HA) protein,

47   which is the primary target of protective immune responses (Sandbulte, et al. 2015; Vincent, et

48   al. 2017). Vaccines that contain HA antigens that are well-matched to the genetic diversity

49   circulating IAV in swine have demonstrated efficacy, but this is greatly reduced when applied to

50   control genetically drifted IAV (Vincent, Lager, et al. 2008). Further, the H1 and H3 HA

51   subtypes of IAV that are detected in US swine contain at least 14 genetic clades with distinct

52   antigenic properties (Anderson, et al. 2013; Zeller, Anderson, et al. 2018). Consequently, a

53   potential strategy to increase the breadth of protection offered by IAV vaccines is the inclusion

54   of additional proteins. Research has suggested that the neuraminidase (NA) gene may also

55   contribute to vaccine efficacy through neuraminidase inhibiting antibodies (Monto and Kendal

56   1973; Sandbulte, et al. 2016).

57   N1 and N2 subtype NA genes circulate in the US, with the N2 subtype being detected in

58   approximately two-thirds of swine IAV reported in surveillance programs (Anderson, et al. 2013;

59   Zeller, Anderson, et al. 2018). Within these NA subtypes, there are two N1 genetic clades,

3

60   including the classical swine lineage (N1.Classical) which emerged coincident with the 1918

61   H1N1 introduction into swine (Koen 1919), and a Eurasian swine lineage N1 gene

62   (N1.Pandemic) that emerged in the US associated with the 2009 H1N1 human pandemic

63   (Dawood 2009; Smith, et al. 2009). Contemporary N2 genetic clades circulating in swine are

64   derived from human-to-swine transmission episodes: the first associated with human seasonal

65   IAV in the late 1990s (N2.1998), and the second in the early 2000s (N2.2002). The N2.1998

66   lineage was introduced into swine with the triple reassortant IAV that evolved into the

67   H3.Cluster IV HA (Zhou, et al. 1999). The N2.2002 lineage was introduced during a human to

68   swine spillover coincident with the H1.Delta clades (Vincent, et al. 2009). Although veterinary

69   diagnostic labs continue to occasionally detect sporadic human seasonal NA in swine (Zeller, Li,

70   et al. 2018), these have yet to become widely established.

71   The quantification of genetic and antigenic diversity described in IAV in swine has

72   largely focused on the HA gene. However, the NA and HA function in a coordinated effort to

73   replicate within and transmit between hosts. Functionally, a balance in NA and HA cellular

74   interactions are necessary for IAV infection to result in successful host-to-host transmission

75   (Mitnaul, et al. 2000; Das, et al. 2011). Links between these genes have been documented in

76   phylogenetic studies on human seasonal H3N2 which have suggested that mutation in one gene

77   affects mutation in the other (i.e., inter-gene epistasis) (Neverov, et al. 2014). In swine, a similar

78   type of evolutionary dynamic is plausible given the 14 different HA genetic clades and 4 NA

79   lineages with considerable genetic diversity within each, but the degree to which this affects

80   observed diversity is unknown.

81   The objective of this study was to assess the NA genetic diversity of IAV circulating in

82   US swine and evaluate how evolution and reassortment to pair with new HA subtypes or clades

4

83   may affect the genetic diversity and evolutionary trajectory of the N2 gene. Understanding how

84   HA pairing and reassortment impacts the evolutionary dynamics and extent of genetic diversity

85   of the NA in swine IAV is critical to understanding how the genomic context affects the

86   emergence and transmission of IAV in swine and will facilitate control efforts that include NA in

87   vaccine strain selection.

88   **RESULTS**

89   **Increased N2 relative genetic diversity**

90   Effective population size (EPS) was used to quantify changes in relative genetic diversity in

91   swine IAV N2 lineages and clades (Figure 1). The N2.1998 demonstrated an increase in diversity

92   between 2014 and 2017 whereas the N2.2002 lineage demonstrated an increase in diversity

93   between 2010 and 2016. Four statistically supported monophyletic clades were detected within

94   these broad swine N2 lineages, N2.1998 split into N2.1998A and N2.1998B circa 2008 (2006.46

95   – 2008.78 95% HPD), and N2.2002 split into N2.2002A and N2.2002B circa 2006 (2004.92 –

96   2007.14 95% HPD). The relative genetic diversity for each of these clades are presented in

97   Figure S1. The average within genetic clade distance was 3% and the between clade distance was

98   greater than 6% for each of the monophyletic clades in each lineage; but, there was clade specific

99   variation in diversity. The N2.1998A had a constant EPS, compared to the increasing EPS

100   observed in the N2.1998B clade. The N2.2002A demonstrated an initial increase that peaked in

101   mid-2012, before being gradually superseded by the N2.2002B by mid-2014 and peaking in

102   2015. These data suggest a general trend of increasing relative genetic diversity observed in both

103   N2 lineages. The N2.1998B clade appears to be the major driver of the increased EPS of the

104   N2.1998 lineage whereas the N2.2002A and N2.2002B demonstrated different temporal trends

105   that both contributed to the overall increase in EPS across the study period.

N2 and HA reassortment and non-random pairing

107 NA-HA pairing was inferred from the N2 trees annotated by HA clade. Distinct clustering of HA

108 clade groups was evident, along with shifts in HA clade annotation on the NA genes as a likely

109 result of reassortment. We defined establishment of the IAV HA and N2 clade pairing within the

110 swine population when they were observed with greater than 10 detections across multiple years.

111 Nine such reassortment events were found and annotated (Figure 2). Longer branch lengths were

112 observed after reassortment, suggesting an increase in N2 genetic diversity. A Chow test (Chow

113 1960) was used to statistically validate if there was a change in the rate of mutation of the N2

114 neuraminidase before and after HA and NA reassortment (Figure S2). Of the nine reassortment

115 events observed, three had significant support for differences in evolutionary rate. The

116 reassortment event labeled as "1" in Figure 2A, a transition in N2.1998A from H3.ClusterIVF to

117 H1.Delta1B ($p < 0.001$) (Figure S2A), showed the rate of mutation slowed following

118 reassortment. The reassortment event labeled as "7" in Figure 2B described an N2.2002B paired

119 with H3.ClusterIVB reassorting to Cluster IVA ($p < 0.001$) (Figure S2E): in this case, the rate of

120 mutation increased following reassortment. The reassortment event labeled "8" in Figure 2B was

121 an N2.2002B gene that changed pairing from H3.ClusterIVA to H1.Delta1A ($p < 0.001$) (Figure

122 S2F); for this event, the rate of mutation remained the same after reassortment, but there was a

123 increase in branch length to the oldest gene in the genetic clade of these pairings. The remaining

124 reassortment events with onward transmission in the US swine population shown in Figure 2 had

125 a general trend towards increased mutation rates following reassortment but this was not

126 statistically significant (Fig S2).

127   A tanglegram consisting of a subsampled pairing of N2 and HA sequences was

128 generated. Lines connecting N2 to their constituent HA within strain tended to be paired by

129 clade, suggesting these pairings were not random (Figure 3). In Figure 3, the connecting lines

130 indicate NA-HA pairings: the frequency of detection and the topology of these trees

131 demonstrated some pairing pattern shifts that subsequently became fixed in the population

132 (Figure 3). Prior to 2013, detections of H1.Delta1A were paired with N2.2002A. Following

133 reassortment in 2013, this HA gene was frequently paired with N2.2002B and became the

134 predominant H1.Delta1A pair by 2015 (Figure S3A). Prior to 2011 H3.ClusterIVA were paired

135 with N2.2002A, and after 2011 this HA was paired with N2.2002B and became the predominant

136 pairing by 2013 (Figure S3B). The H3.2010.1 was paired with the N2.2002A predominantly

137 from 2014-2017, after which N2.2002B became predominantly detected (Figure S3C).

138   The non-random pairing observed in the tanglegram was statistically significant (chi-

139 squared (p < 0.0001); Fisher's exact test (p < 0.0001)). The distribution of N2-HA pairings

140 (Figure S4A) were subjected to a post hoc analysis using the standardized residuals from the chi-

141 squared test (Figure S4B). This demonstrated disproportionate pairing of N2 clades with their

142 constituent HAs. N2.1998A were predominantly paired with H1.Delta1B and H3.ClusterIVF.

143 This pairing demonstrated a temporal trend, as most of the N2.1998A detections were associated

144 with the H1.Delta1B rather than H3.ClusterIVF following reassortment (Figure 2). N2.1998B

145 were almost exclusively paired with H1.Delta2. The N2.2002A were disproportionately paired

146 with H1.Delta1B or H3.ClusterIVE (Figure S4B): both clades were infrequently detected (Figure

147 S4A). The N2.2002B was disproportionately paired with H1.Delta1A, H1.Alpha, and the

148 H3.ClusterIVA-D (Figure S4B). The H1.Delta1A and H3.ClusterIVA HA clades were also

149 associated with N2.2002A indicating these HA genes permit multiple N2 clade pairings (Figure

150 S4B). The H3.2010.1 clade viruses were paired with either N2.2002A or N2.2002B in

151 approximately equal numbers, but data collected over the last 2 years demonstrated an increase

152   of N2.2002B pairing (Figure S3C). The non-random pairing observed between N2 and HA genes

153   through the chi-squared test indicated a potential correlation in relative genetic diversity of each

154   gene. The relative genetic diversity of the N2.1998B clade paired with the H1.Delta2 clade and

155   the N2.2002A paired with the H1.Delta1 clade were selected for analysis due to their near

156   exclusive pairing. Detection of the H1.Delta2 paired with the N2.1998B first occurred in 2011

157   and thereafter became the predominant N2-HA pairing of these clades. After 2012, the EPS of

158   the N2.1998B and the H1.Delta2 became superimposed both temporally and in magnitude,

159   suggesting that changes in the genetic diversity of these two genes are linked (Figure S5A).

160   Paired N2.2002A and H1.Delta1B demonstrated similar temporal aspects and magnitude of their

161   respective measures of relative genetic diversity (Figure S5B).

162   Similarity in the relative genetic diversity of N2-HA pairs indicated that the mutation rate

163   of N2 clades and paired HA clades are likely similar. Comparing the mean substitution rate per

164   codon of N2 clades and select HA clades demonstrated similar mutation rates, N2.1998A:

165   0.0044, N2.1998B: 0.0046, N2.2002A: 0.0041, N2.2002B: 0.0045, H1.Gamma: 0.0047,

166   H1.Delta2: 0.0049, H3.ClusterIVA: 0.0049, H3.2010.1: 0.0050 sites/mutation/year (Figure S6).

167   Collectively, the overall NA mutation rate was slightly lower than the overall HA mutation rate,

168   with a difference of 0.000475 sites/mutation/year.

169   **Diversifying selection in the N2**

170   A mixed effects model of evolution selection analysis demonstrated that multiple amino

171   acid sites were under positive selection for each N2 clade (Table 1). The N2.1998A had 4 amino

172   acid sites under positive selection in both the transmembrane and globular head regions.

173   Although the analysis indicated position 414 as under positive selection, glycine was present at

174   position 414 in all but three N2.1998.A genes and did not appear to be truly under positive

175 selection. The N2.1998B had one transmembrane, five stalk, and four globular head sites under

176 positive selection. There were two distinct monophyletic groups nested within the N2.1998B

177 clade with the sites under positive selection also split across each clade, suggesting these two

178 groups diverged prior to 2009 (the earliest data used in our analysis). The N2.2002A had one

179 stalk and three globular head sites under positive selection, but no transmembrane sites. The

180 N2.2002B had one transmembrane, one stalk, and seven globular head sites under positive

181 selection. These data supporting the presence of positive selection across each of the N2 NA

182 clades indicates that functional amino acid mutations are likely occurring in the gene as it is

183 evolving.

184 **Directionality of reassortment**

185   Reassortment resulting in HA and N2 gene pairings did not appear to be random (Figure

186 4). There was limited reassortment of the N2.1998A clade NA genes from H3.ClusterIVF to

187 H1.Delta1B viruses (Figure 4). The N2.1998B clade genes were almost exclusively paired with

188 H1.Delta2, with very few sporadic dead-end reassortment events detected. The few reassortment

189 events within the N2.1998B clade were given high and uniform posterior probabilities as well as

190 Bayes factors, but our maximum likelihood trees indicated that despite N2.1998B reassortment,

191 it was not consistently detection or sustained in swine. The N2.2002A genes demonstrated

192 reassortment from H1.Delta1A, H1.Delta1B, H3.ClusterIVA, and H3.2010.1 to multiple other

193 HA clades, indicating the N2.2002A genes may pair with multiple HA genes. The N2.2002B

194 demonstrated the greatest breadth of reassortment with the gene pairing with H1.Delta1A,

195 H1.Alpha, H3.ClusterIVA, H3.ClusterIVB, and H3.2010.1. These results indicate that some NA

196 genes are permissive of multiple HAs (i.e., approaching random), whereas other NA genes are

9

197 only detected with specific HA genes. This Bayesian approach is consistent with the clade

198 pairing patterns observed in the tanglegram and chi-squared analysis (Figure 3 and Figure S5).

199 **Spatial dissemination of N2 genetic clades**

200   The spatial detection of N2 clades from 2009-2018 indicated differences in geographic

201 distribution (chi-squared test p < 0.0001: Figure 5A) that reflect general patterns of swine

202 agriculture in the US. Viruses containing N2.1998A had the fewest detections and had limited

203 geographic distribution in the Midwest, primarily in Iowa or adjacent states. The N2.1998B was

204 detected most frequently in the Southeast (North Carolina), with a few detections in the Midwest

205 including Indiana and Iowa. The N2.2002A was detected most frequently in the Midwest and

206 was rarely detected in other regions of the US. The N2.2002B was detected in the Midwest as

207 well as the Southeast (North Carolina). On a state basis, differences in detection of viruses

208 containing specific NA clades was observed, for example viruses detected in Iowa contained NA

209 clades in descending order of frequency: N2.2002B, N2.2002A, then N2.1998B.

210   Interstate N2 gene movement, inferred by Bayesian phylodynamic techniques (Figure

211 5B), demonstrated that even infrequently detected N2 had geographic movement despite the low

212 overall distribution presented in Figure 5A. IAV containing N2.1998A moved between

213 Minnesota, Iowa, and Nebraska. IAV with N2.1998B moved from North Carolina into the

214 Midwest as well as from Iowa to Minnesota and Missouri. Despite movement to the Midwest,

215 the N2.1998B was not frequently detected in that region. IAV with N2.2002A were moved

216 (likely transported with live swine) from Iowa and Indiana to multiple locations throughout the

217 Midwest, including Nebraska and Kansas. IAV with N2.2002B moved from North Carolina to

218 Indiana and Pennsylvania, with Indiana, Iowa, and Illinois also acting as sources of N2 gene

219 distribution. Nebraska appeared as a possible source for N2.2002B to move into Kansas. Despite

10

220   the inference of transmission of almost all N2 genetic clades to all geographic regions across the

221   US, the detection of N2 clades within surveillance data suggested that many of these clades do

222   not persist.

223   **DISCUSSION**

224   Multiple cocirculating clades of swine IAV NA were reported, with a significant amount

225   of genetic diversity that challenges control efforts (Anderson, et al. 2013; Zeller, Anderson, et al.

226   2018). Our data demonstrated that diversity in the NA gene has increased dramatically over the

227   past 10 years. Previously described N2.1998 and N2.2002 lineages genetically diverged circa

228   2008 and 2006 respectively, warranting the naming of four statistically supported monophyletic

229   clades (N2.1998A, N2.1998B, N2.2002A, N2.2002B). Our data suggested that this observed

230   diversity has occurred due to the concurrent forces of drift, reassortment that resulted in novel

231   NA-HA pairings, and interstate movement of IAV. These dynamics played a critical role in

232   establishing predominant NA-HA pairings. The interstate movement of IAV with pigs may

233   perpetuate the cycle of drift, reassortment and diversification through the regular introduction of

234   novel NA-HA pairings to different regions; and this creates significant challenges to the

235   formulation of well-matched vaccines and other control efforts.

236   Viruses containing N2.1998 and N2.2002 lineage genes had multiple reassortment

237   events, leading to the emergence of genetically distinct N2 clades. Our data demonstrated that

238   N2 genes may persist at low detection levels, exchange HA genes via reassortment, and then

239   through diversification and spatial dissemination the "minor" clades may rapidly become

240   predominant. This is best exemplified by the frequency of detection of the N2.1998B and

241   H1.Delta2 genes. The H1.Delta2 HA gene was originally paired with an N2.2002 when first

242   detected in the US swine population, and this combination occurred at relatively low levels prior

243 to 2011. In late 2011, a reassortment event that resulted in the pairing of the N2.1998B with the

244 H1.Delta2 HA occurred, this coincided with correlated increases in the relative genetic diversity

245 of both the H1.Delta2 and N2.1998B (Figure S4A), and subsequent substantive increases in the

246 number of detections of the N2.1998B gene, with this combination representing 16% of H1

247 detections in 2017 (Zeller, Anderson, et al. 2018). Similarly, the N2.1998A clade persisted

248 paired with H1.Delta1B subsequent to reassortment with H3.ClusterIVF, whereas H3.ClusterIVF

249 IAV became infrequently detected. It is plausible to suggest that reassortment generating new

250 NA-HA pairings could result in diversification of both HA and NA genes, adaptation, and

251 potential wide dissemination. Our results suggested that NA-HA pairs and their evolution in

252 swine are coordinated and likely play a role in alternating predominance of specific N2 clades

253 and their paired HA clades.

254   An increase in detection frequency of specific NA-HA pairs following reassortment may

255 be related to a favorable balance between the NA and HA protein activity (Hensley, et al. 2009),

256 a better match between the surface genes and internal gene segments for replication (Gao, et al.

257 2017), or novel antigenic properties that evade population immunity (Abente, et al. 2016). The

258 tanglegram in this study indicated that reassortment was followed by genetic divergence in

259 connected NA and HA genes, evidenced in longer branch lengths and complete replacement of

260 one NA-HA pairing with another over time. Newly paired HA and NA genes may require

261 adaptation, and an additional collateral consequence of reassortment (antigenic shift) may be an

262 increase in antigenic drift in both of the surface proteins. These data highlight the necessity to

263 monitor genetic changes in both NA and HA, and patterns of NA and HA pairings through IAV

264 surveillance in swine: these data will facilitate vaccine antigen selection and provide warning of

265 potential rapid diversification that may result in a less effective vaccine formulation.

266   Our data revealed non-random pairing between NA and HA clades. We detected

267   reassortment events that were a single (dead-end) detections or followed by multiple detections.

268   The increased frequency of detection of novel NA-HA pairs following reassortment was

269   indicative of sustained transmission of that NA-HA pairing within the swine population. Some

270   reassortment events with subsequent transmission were also characterized by subsequent

271   increased genetic diversity of both the N2 and HA in a coordinated manner based on the

272   correlation in increase and decrease in relative genetic diversity across an 8 year period. A

273   possible explanation for the coordinated changes in diversity genes is that the activity of these

274   two proteins requires a balance for an IAV infection that results in transmission (Mitnaul, et al.

275   2000; Das, et al. 2011). The coordinated increase in diversity may arise due to selection pressure

276   to optimize the HA and NA balance, evidence for this is a lag before the increase in detection

277   numbers following reassortment, i.e., a novel NA and HA requires a period of adaptation

278   (detected as an increase in relative genetic diversity) prior to increasing in detection frequency

279   (Figure S3).

280   The genetic evolutionary rates of the N2 NA and the HA genes in swine were shown to

281   have similar rates. In a limited number of studies, the mutation rate for the HA has been

282   consistently similar to the NA across studies when measured empirically (Saitou and Nei 1986;

283   Air, et al. 1990) and computationally (Khandaker, et al. 2013; Westgeest, et al. 2014). However,

284   the general paradigm in IAV biology is a focus on the evolution of the HA and the role of

285   immune-driven selection in its dynamics. Our data fit the growing body of literature that suggest

286   that the NA evolves in a similar manner to the HA, and that immune driven selection may lead to

287   functional amino acid changes (Correia, et al. 2018). We detected 26 amino acid sites that

288   demonstrated a signal of diversifying selection, this included sites that were pervasively

13

289 observed across different clades, some of which may have functional consequences. However,

290 relative to data on selection analyses in HA genes, the number of sites in the NA undergoing

291 diversifying selection is more limited, possibly as a compensatory mechanism for genetic

292 changes in the HA.

293   The four monophyletic N2 clades circulating in U.S. swine had different geographic

294 patterns and frequencies of reassortment. The N2 clades that were paired with more HA clades

295 were the N2.2002A and N2.2002B clades, and these both had a higher relative genetic diversity

296 when compared to the N2.1998A and B clades, as well as being frequently detected over the

297 study period. This observation may reflect a "sampling effect" where a larger pool of genetic

298 diversity in the N2 increases the likelihood of a reassortment event with an HA that may

299 demonstrate onward transmission. Additionally, our data suggested that that genetic diversity

300 increased following reassortment events; again, following reassortment, diversity may increase

301 as a consequence of the genes being under coordinated evolutionary pressure to replicate within

302 and transmit between hosts.

303   Detections of N2 clades were observed to have regional patterns. The N2.1998A was

304 limited to 3 states in the Midwest, while the majority of N2.1998B detections were in the

305 Southeast (North Carolina). The N2.2002A was detected across the Midwest, the N2.2002B was

306 detected in the Midwest and Southeast. These patterns were derived from over eight years' worth

307 of data and support general swine agricultural practices where the Midwest appears to be a

308 central point for pig shipment. Prior research suggested that the transmission of IAV in swine

309 was observed in directed patterns in the US, namely from the Southwest and Southeast to the

310 Midwest (Nelson, et al. 2011), and our data on the movement of the N2 genes is in general

311 agreement with these prior findings. The implication of this is that directed movement of viruses

14

312   could influence the opportunity of reassortment and subsequent patterns of diversity: areas that

313   largely receive pigs and their viruses may provide more opportunities for reassortment.

314   Consequently, a location such as Iowa would be expected to have the highest IAV genetic

315   diversity as well as the most reassortment events. However, there are exceptions, some N2 clades

316   remained geographically restricted despite interstate movement. The N2.1998B was introduced

317   to the Midwest multiple times and to multiple states, but was not successfully established,

318   indicating that factors (e.g., genetic compatibility or host-related) have prevented this clade from

319   expanding across the US.

320   **Conclusion**

321   This study demonstrates that changes within the NA and HA of the influenza virus in

322   swine in the US were not independent of each other. While this study focused on the N2 due to

323   its frequency of detection and apparent reassortment and genetic diversity, understanding the

324   evolution of N1 and N2 subtypes will benefit vaccine development. If the HA vaccine antigen

325   does not match currently circulating strains, a vaccine that contains a matched NA antigen may

326   still reduce clinical signs and shedding of the virus. Further, monitoring the NA may serve as an

327   early indicator of phenotypic change based on detected reassortment events. IAV reassortment

328   may increase genetic and antigenic drift (Vijaykrishna, et al. 2011), and the current movement of

329   swine between geographic locations in the US may increase these events. Future surveillance

330   efforts with increased monitoring and surveillance of the genetic diversity of NA will be additive

331   to the HA in control efforts to protect swine from influenza and reduce the chance for zoonotic

332   transmission.

333   **MATERIALS AND METHODS**

334   **Data acquisition and phylogenetic clade assignment**

335   Approximately 4000 paired nucleotide N2 NA and HA sequences detected in swine with

336   accompanying metadata were downloaded from the Influenza Research Database (IRD) on

337   October 18th, 2018 (Squires, et al. 2012; Yun Zhang, et al. 2017). Data was limited to IAV swine

338   sequences from the US from August 9th, 2009 to September 20th, 2018. All NA sequences less

339   than 1200 nt were removed (representing less that 85% of the coding region). Sequences

340   associated with agricultural fair events and pigs were removed as they do not represent IAV from

341   the general commercial swine population (i.e., sequences annotated with 'OSU') (Bowman, et al.

342   2017).

343   Phylogenetic clade classifications for H1 viruses were designated using the Swine H1

344   Clade Classification Tool provided on the IRD (Anderson, et al. 2016). H3 and NA clades were

345   determined through maximum-likelihood phylogenetic analysis using reference sequences

346   (Chang, et al. 2019). Reference nucleotide sequences and downloaded IAV from swine were

347   aligned using MAFFT v7.27 (Katoh and Standley 2013). Maximum likelihood trees were

348   inferred using FastTree2 v2.1.9 (Price, et al. 2010) for each gene alignment using a general time

349   reversible model of nucleotide substitution with each IAV gene classified to clade based upon

350   the nearest neighbor in the reference gene dataset. Within and between clade distances were

351   calculated using MEGA X v10.1 (Kumar, et al. 2018).

352   **Quantifying relative genetic diversity in swine IAV N2 NA genes**

353   Estimates of relative diversity for the N2.2002 and N2.1998 genetic lineages were

354   determined through time-scaled Bayesian phylodynamic analyses and the inference of effective

355   population size. Based upon computational limitations, we randomly sampled 620 genes from

16

356   the N2.2002 lineage (from n=3650 total genes). These genes were aligned using MAFFT v7.27

357   (Katoh and Standley 2013) and a maximum-likelihood tree was inferred using FastTree2 v2.1.9

358   (Price, et al. 2010). Root-to-tip divergence was analyzed using TempEST v1.5.1 (Rambaut, et al.

359   2016) and genes with incongruent divergence and sampling date were removed (data available

360   via GitHub). The resulting final dataset consisted of 600 non-identical N2.2002 genes. All of the

361   available N2.1998 genes (n=596 taxa) were aligned, those that weren't at least 50% of the gene

362   were removed, and a maximum likelihood tree was inferred with root-to-tip divergence analyzed

363   in TempEST v1.5.1 with two taxa removed as outliers, resulting in 583 taxa for analysis.

364   These datasets were analyzed using Bayesian phylogenetic methods in BEAST v1.8.4

365   (Drummond, et al. 2012) with the BEAGLE library v3.1.2 (Ayres, et al. 2011), implementing a

366   generalized time reversible (GTR) nucleotide substitution model (Tavaré 1986) with gamma-

367   distributed site heterogeneity (Yang 1994). We employed an uncorrelated relaxed clock with

368   lognormal distribution (Drummond, et al. 2006), and a GMRF Bayesian skyride with time

369   aware-smoothing as the coalescent model (Minin, et al. 2008). The MCMC chain length was set

370   to 100 million iterations with sampling every 10,000 iterations. Results were analyzed using the

371   GRMF skyride reconstruction in Tracer v1.6 (Rambaut and Drummond 2007).

372   **Measuring concurrent changes in the diversity of conserved NA and HA gene pairings**

373   The frequency of particular NA and HA pairings was measured by splitting the data into

374   the N2.2002 and N2.1998 monophyletic clades. NA-HA pairings that did not occur more than 10

375   times were removed from analysis, and the significance of gene pairings was measured with a

376   post-hoc analysis following a Pearson's chi square test for independence using the standard

377   residuals in R v3.3.3 (R Core Team 2015). A confirmatory Fisher's exact test was performed due

378   to sparsity of some NA-HA pairings

379   Of these documented NA and HA pairings, N2.1998B was almost exclusively paired with

380   H1.Delta2 HA genes (449 of 462 genes). Consequently, we analyzed the evolutionary dynamics

381   of 462 taxa of N2.1998B clade NA genes and 522 taxa of H1.Delta2 HA genes separately. These

382   genes were aligned, poor quality data were removed, and root-to-tip divergence was analyzed

383   with divergent sequences removed. This process resulted in 458 N2.1998B genes and 517

384   H1.Delta2 genes (data available on GitHub). These data were analyzed using BEAST v1.8.4

385   implementing a GTR substitution model with gamma-distributed rate variation, an uncorrelated

386   relaxed clock with lognormal distribution, and the GMRF Bayesian skyride coalescent. The

387   MCMC chain was run for 100 million iterations with sampling every 10,000 iterations. The

388   effective population size (EPS) for both clades N2.1998B and H1.Delta2 was plotted to

389   determine correlated changes in relative genetic diversity. A similar analysis with the same

390   analytical settings was conducted on the N2.2002A genes that were paired with H1.Delta1B HA

391   (following screening, n=842 N2.2002A and H1.Delta1B taxa for analysis).

392   **Estimating the evolutionary rate of swine IAV N2 NA and HA genes**

393   To estimate the evolutionary rate of N2 NA, we created alignments of each of the major

394   of genetic clades: N2.1998A, N2.1998B, N2.2002A, and N2.2002B (n = 115, 458, 262, 310,

395   respectively). To estimate the evolutionary rate of H1 and H3 HA, we generated alignments of

396   H1.Gamma, H1.Delta2, H3.2010.1, and H3.ClusterIVA (n=300, 300, 299, 298). Using BEAST 2

397   v2.5.1 (Bouckaert, et al. 2014), we implemented an unlinked codon substitution model split into

398   independent partitions (1 + 2 + 3) and unlinked clock models for each codon position. The

399   analysis used a general time reversible (GTR) nucleotide substitution model with a gamma site

400   heterogeneity, an uncorrelated relaxed clock with lognormal distribution, and an exponential

401   population growth coalescent tree prior. The MCMC chain length was set to 100 million

18

402   iterations with sampling every 10,000 iterations. Mean rate of substitution was calculated by

403   averaging the values for the three codon position's mean rate of substitution in terms of

404   sites/mutation/year with distribution for each and visualized using ggplot2 (Wickham 2016).

405   **Detecting diversification following reassortment of N2 NA in swine IAV**

406   Reassortment in N2 NA gene was identified using topological incongruity in gene trees

407   (Boni, et al. 2010). Maximum likelihood trees were inferred for N2.1998 and N2.2002 using

408   RAxML v8.2.11 (Stamatakis 2014) with a GTR model with a gamma-distributed rate variation

409   and statistical support values determined using rapid bootstrapping with the autoMRE criterion.

410   Trees were visualized in FigTree v1.4.3 (Rambaut 2012), and we defined sustained transmission

411   of a reassorted pairing as one where a monophyletic clade of NA genes acquired a novel HA

412   genetic clade, and the NA-HA combination was maintained within the monophyletic clade for

413   more than 10 subsequent branches. To facilitate visualization, a randomly sampled set of 300 HA

414   and NA genes were sorted and transformed into proportional branch length cladograms and built

415   into a tanglegram in R v3.3.3 (R Core Team 2015) using the dendextend v1.9.0 package (Galili

416   2015). Branches on these constituent maximum-likelihood trees were colored by phylogenetic

417   clades with lines connecting trees colored based on HA clade to emphasize reassortment events.

418   Branch lengths of the NA tree were compared to assess whether reassortment affected

419   diversity. To achieve this, we identified and extracted monophyletic clades that contained a

420   reassortment event. Extracted clades were realigned using MAFFT v7.27 (Katoh and Standley

421   2013), and a maximum likelihood trees were inferred using FastTree2 v2.1.9 (Price, et al. 2010).

422   We were able to generate 7 phylogenies that demonstrated an NA gene acquiring a novel HA

423   gene that then persisted in the US swine population. The 7 trees were loaded into R v3.3.3, and

424   the patristic distance from each NA gene to the oldest leaf node was determined using APE v5.2

425   (Paradis, et al. 2004). A linear model of $y = \beta0 + \beta1x$ was fitted to these data, where y was

426   patristic distance, $\beta1$ was date of collection of the NA gene, and $\beta0$ represents an intercept. A

427   linear model was fit for patristic distance for NA genes that had the prior paired HA clade, and

428   for the NA genes that had acquired a novel HA gene. The fitted models for each of reassortment

429   event (i.e., 14 models, 2 for each of the 7 reassortment events) were compared using the Chow

430   Test from the gap package v1.2.2 in R (Zhao 2007). To determine whether using one or two

431   linear models was appropriate, we measured whether there was significant reduction in the

432   residuals at the 95% confidence level.

433   Within the NA genes, we used MEME (Murrell, et al. 2012), part of the HyPhy package

434   v2.3.14, to test for instances of episodic and pervasive selection by using a mixed-effects model

435   of evolution. Results from MEME were confirmed by conducting concurrent analyses with by

436   FEL and SLAC (Kosakovsky Pond and Frost 2005), and reviewed via HyPhy vision v2.4.3.

437   **Inferring the directionality of N2 and HA genes following reassortment in swine IAV**

438   To quantitatively characterize the direction of the exchange of NA and HA genes, we

439   adapted a phylogeographic approach, using NA genetic clade information as our "location state."

440   The same data subset and setup was used as in our analyses on quantifying changes in relative

441   genetic diversity. Each NA gene had an HA genetic clade designated as a trait, and transitions

442   from one category to another (e.g., from N2.1998B/H1.Gamma to N2.1998B/H3 2010.1) were

443   inferred along the internal branches representing the evolutionary history of the virus. These

444   transitions are termed Markov jumps and state change counts for the HA clade trait were

445   reconstructed in BEAST v1.8.4. The sites for the trait were calculated using an asymmetric

446   substitution model, a Bayesian stochastic search variable selection (BSSVS) (Edwards, et al.),

447   and an uncorrelated relaxed clock model with lognormal distribution. Analysis output was

20

448   assessed in Tracer v1.6.0 to confirm convergence and SPREAD3 v0.9.6 (Bielejec, et al. 2016)

449   was used to compute the Bayes factors with 10% burn-in. Bayes factors were plotted using the

450   inferno color scheme from the R viridis package v0.5.1 using the ggplot2 v3.1.0 and cowplot

451   v0.9.4.

452   **Assessing the spatial distribution of N2 NA genes in swine IAV**

453   Swine IAV N2 data were reduced to subsets that included only sequences that included a

454   US state. The number of detections of each N2 genetic clade was rendered on a US map using

455   the R maps module v3.3.0, colored to emphasize differences using the viridis color scheme and

456   rendered by ggplot2 v3.1.0. Statistical association between N2 genetic clade and state of

457   collection was assessed with a Pearson's chi square test for independence, and then a post hoc

458   analysis of chi square test using the standard residuals in R v3.3.3 (R Core Team 2015). A

459   Fisher's exact test was performed to confirm the results of the initial chi square test. To quantify

460   the movement of the NA genes across the US, we used a phylogeographic approach with US

461   state as our "location state." Each NA gene had an US state designated as a trait, and transitions

462   from one category to another (e.g., from North Carolina to Iowa) were inferred along the internal

463   branches representing the evolutionary history of the virus. These number of state change counts

464   for the US state trait were reconstructed in BEAST v1.8.4 following the same methods described

465   above.

21

482   **REFERENCES**

483   Abente EJ, Santos J, Lewis NS, Gauger PC, Stratton J, Skepner E, Anderson TK, Rajao DS,

484   Perez DR, Vincent AL. 2016. The molecular determinants of antibody recognition and

485   antigenic drift in the H3 hemagglutinin of swine influenza A virus. Journal of virology

486   90:8266-8280.

487   Air GM, Gibbs AJ, Laver WG, Webster RG. 1990. Evolutionary changes in influenza B are not

488   primarily governed by antibody selection. Proceedings of the National Academy of

489   Sciences 87:3884-3888.

490   Anderson TK, Macken CA, Lewis NS, Scheuermann RH, Van Reeth K, Brown IH, Swenson SL,

491   Simon G, Saito T, Berhane Y. 2016. A Phylogeny-Based Global Nomenclature System

492   and Automated Annotation Tool for H1 Hemagglutinin Genes from Swine Influenza A

493   Viruses. mSphere 1:e00275-00216.

494   Anderson TK, Nelson MI, Kitikoon P, Swenson SL, Korslund JA, Vincent AL. 2013. Population

495   dynamics of cocirculating swine influenza A viruses in the United States from 2009 to

496   2012. Influenza Other Respir Viruses 7 Suppl 4:42-51.

497   Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, Huelsenbeck JP, Ronquist F,

498   Swofford DL, Cummings MP. 2011. BEAGLE: an application programming interface

499   and high-performance computing library for statistical phylogenetics. Systematic biology

500   61:170-173.

501   Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. 2016. SpreaD3:

502   interactive visualization of spatiotemporal history and trait evolutionary processes.

503   Molecular biology and evolution 33:2167-2169.

504   Boni MF, de Jong MD, van Doorn HR, Holmes EC. 2010. Guidelines for identifying

505   homologous recombination events in influenza A virus. PloS one 5.

506   Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A,

507   Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis.

508   PLoS computational biology 10:e1003537.

509   Bowman AS, Walia RR, Nolting JM, Vincent AL, Killian ML, Zentkovich MM, Lorbach JN,

510   Lauterbach SE, Anderson TK, Davis TC. 2017. Influenza A (H3N2) virus in swine at

511   agricultural fairs and transmission to humans, Michigan and Ohio, USA, 2016. Emerging

512   infectious diseases 23:1551.

513   Chang J, Anderson TK, Zeller MA, Gauger PC, Vincent AL. 2019. octoFLU: Automated

514   Classification for the Evolutionary Origin of Influenza A Virus Gene Sequences Detected

515   in US Swine. Microbiology resource announcements 8:e00673-00619.

516   Chow GC. 1960. Tests of equality between sets of coefficients in two linear regressions.

517   Econometrica: Journal of the Econometric Society:591-605.

518   Correia V, Abecasis AB, Rebelo-de-Andrade H. 2018. Molecular footprints of selective pressure

519   in the neuraminidase gene of currently circulating human influenza subtypes and

520   lineages. Virology 522:122-130.

521   Das SR, Hensley SE, David A, Schmidt L, Gibbs JS, Puigbò P, Ince WL, Bennink JR, Yewdell

522   JW. 2011. Fitness costs limit influenza A virus hemagglutinin glycosylation as an

523   immune evasion strategy. Proceedings of the National Academy of Sciences 108:E1417-

524   E1422.

525   Dawood FS. 2009. Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans.

526   The New England Journal of Medicine 360:2605-2615.

527   Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with

528   confidence. PLoS biology 4:e88.

529   Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti

530   and the BEAST 1.7. Molecular biology and evolution 29:1969-1973.

531   Dykhuis-Haden C, Painter T, Fangman T, Holtkamp D. 2012. Assessing production parameters

532   and economic impact of swine influenza, PRRS and *Mycoplasma hyopneumoniae* on

533   finishing pigs in a large production system. American Association of Swine

534   Veterinarians; Denver, Colorado. p. 75-76.

535   Edwards CJ, Suchard MA, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC,

536   Coxon P, Monaghan N. Supplemental Information Ancient Hybridization and an Irish

537   Origin for the Modern Polar Bear Matriline.

538   Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of

539   hierarchical clustering. Bioinformatics 31:3718-3720.

540   Gao S, Anderson TK, Walia RR, Dorman KS, Janas-Martindale A, Vincent AL. 2017. The

541   genomic evolution of H1 influenza A viruses from swine detected in the United States

542   between 2009 and 2016. Journal of General Virology 98:2001-2010.

543   Gramer M. 2006. Swine influenza virus: the only constant is change.

544   Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K,

545   Raman R, Sasisekharan R, Bennink JR. 2009. Hemagglutinin receptor binding avidity

546   drives influenza A virus antigenic drift. Science 326:734-736.

547   Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:

548   improvements in performance and usability. Molecular biology and evolution 30:772-

549   780.

550   Khandaker I, Suzuki A, Kamigaki T, Tohma K, Odagiri T, Okada T, Ohno A, Otani K,

551   Sawayama R, Kawamura K. 2013. Molecular evolution of the hemagglutinin and

552   neuraminidase genes of pandemic (H1N1) 2009 influenza viruses in Sendai, Japan,

553   during 2009–2011. Virus Genes 47:456-466.

554   Koen J. 1919. A practical method for field diagnosis of swine diseases. Am J Vet Med 14:468-

555   470.

556   Kosakovsky Pond SL, Frost SD. 2005. Not so different after all: a comparison of methods for

557   detecting amino acid sites under selection. Molecular biology and evolution 22:1208-

558   1222.

559   Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary

560   genetics analysis across computing platforms. Molecular biology and evolution 35:1547-

561   1549.

562   Minin VN, Bloomquist EW, Suchard MA. 2008. Smooth skyride through a rough skyline:

563   Bayesian coalescent-based inference of population dynamics. Molecular biology and

564   evolution 25:1459-1471.

565   Mitnaul LJ, Matrosovich MN, Castrucci MR, Tuzikov AB, Bovin NV, Kobasa D, Kawaoka Y.

566   2000. Balanced hemagglutinin and neuraminidase activities are critical for efficient

567   replication of influenza A virus. Journal of virology 74:6015-6020.

568   Monto AS, Kendal AP. 1973. Effect of neuraminidase antibody on Hong Kong influenza. The

569   Lancet 301:623-625.

570   Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012.

571   Detecting individual sites subject to episodic diversifying selection. PLoS Genet

572   8:e1002764.

573   Nelson MI, Lemey P, Tan Y, Vincent A, Lam TT-Y, Detmer S, Viboud C, Suchard MA,

574   Rambaut A, Holmes EC. 2011. Spatial dynamics of human-origin H1 influenza A virus

575   in North American swine. PLoS pathogens 7:e1002077.

576   Neverov AD, Lezhnina KV, Kondrashov AS, Bazykin GAJPg. 2014. Intrasubtype reassortments

577   cause adaptive amino acid replacements in H3N2 influenza genes. 10:e1004037.

578   Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R

579   language. Bioinformatics 20:289-290.

580   Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for

581   large alignments. PLoS One 5:e9490.

582   R Core Team. 2015. R: A language and environment for statistical computing.

583   Rambaut A. 2012. FigTree v1. 4. Molecular evolution, phylogenetics and epidemiology.

584   Edinburgh, UK: University of Edinburgh, Institute of Evolutionary Biology.

585   Rambaut A, Drummond AJ. 2007. Tracer v1. 6 http://beast. bio. ed. ac. uk. In: Tracer.

586   Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of

587   heterochronous sequences using TempEst (formerly Path-O-Gen). Virus evolution

588   2:vew007.

589   Saitou N, Nei M. 1986. Polymorphism and evolution of influenza A virus genes. Molecular

590   biology and evolution 3:57-74.

591   Sandbulte MR, Gauger PC, Kitikoon P, Chen H, Perez DR, Roth JA, Vincent AL. 2016.

592   Neuraminidase inhibiting antibody responses in pigs differ between influenza A virus N2

593   lineages and by vaccine type. Vaccine 34:3773-3779.

594   Sandbulte MR, Spickler AR, Zaabel PK, Roth JA. 2015. Optimal Use of Vaccines for Control of

595   Influenza A Virus in Swine. Vaccines (Basel) 3:22-73.

596   Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL,

597   Raghwani J, Bhatt S, et al. (1709 co-authors). 2009. Origins and evolutionary genomics

598   of the 2009 swine-origin H1N1 influenza A epidemic. Nature 459:1122-1125.

599   Squires RB, Noronha J, Hunt V, Garcia-Sastre A, Macken C, Baumgarth N, Suarez D, Pickett

600   BE, Zhang Y, Larsen CN, et al. 2012. Influenza research database: an integrated

601   bioinformatics resource for influenza research and surveillance. Influenza Other Respir

602   Viruses 6:404-416.

603   Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of

604   large phylogenies. Bioinformatics 30:1312-1313.

605   Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences.

606   Lectures on mathematics in the life sciences 17:57-86.

607   Vijaykrishna D, Smith GJ, Pybus OG, Zhu H, Bhatt S, Poon LL, Riley S, Bahl J, Ma SK,

608   Cheung CL. 2011. Long-term evolution and transmission dynamics of swine influenza A

609   virus. Nature 473:519.

610   Vincent AL, Lager KM, Janke BH, Gramer MR, Richt JA. 2008. Failure of protection and

611   enhanced pneumonia with a US H1N2 swine influenza virus in pigs vaccinated with an

612   inactivated classical swine H1N1 vaccine. Veterinary microbiology 126:310-323.

613   Vincent AL, Ma W, Lager KM, Gramer MR, Richt JA, Janke BH. 2009. Characterization of a

614   newly emerged genetic cluster of H1N1 and H1N2 swine influenza virus in the United

615   States. Virus Genes 39:176-185.

616   Vincent AL, Ma W, Lager KM, Janke BH, Richt JA. 2008. Swine influenza viruses: a North

617   American perspective. Advances in virus research 72:127-154.

618   Vincent AL, Perez DR, Rajao D, Anderson TK, Abente EJ, Walia RR, Lewis NS. 2017.

619   Influenza A virus vaccines for swine. Vet Microbiol 206:35-44.

620   Wang Y, Gagnon CA, Savard C, Music N, Srednik M, Segura M, Lachance C, Bellehumeur C,

621   Gottschalk M. 2013. Capsular sialic acid of Streptococcus suis serotype 2 binds to swine

622   influenza virus and enhances bacterial interactions with virus-infected tracheal epithelial

623   cells. Infection and immunity 81:4498-4508.

624   Westgeest KB, Russell CA, Lin X, Spronken MI, Bestebroer TM, Bahl J, van Beek R, Skepner

625   E, Halpin RA, de Jong JC. 2014. Genomewide analysis of reassortment and evolution of

626   human influenza A (H3N2) viruses circulating between 1968 and 2011. Journal of

627   virology 88:2844-2857.

628   Wickham H. 2016. ggplot2: elegant graphics for data analysis: Springer.

629   Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable

630   rates over sites: approximate methods. Journal of Molecular evolution 39:306-314.

631   Yun Zhang, Brian D. Aevermann, Tavis K. Anderson, David F. Burke, Gwenaelle Dauphin,

632   Zhiping Gu, Sherry He, Sanjeev Kumar, Christopher N. Larsen, Alexandra J. Lee, et al.

633   2017. Influenza Research Database: An integrated bioinformatics resource for influenza

634   virus research. Nucleic Acids Res 45:D466-D474.

635   Zeller MA, Anderson TK, Walia RW, Vincent AL, Gauger PC. 2018. ISU FLUture: a veterinary

636   diagnostic laboratory web-based platform to monitor the temporal genetic patterns of

637   Influenza A virus in swine. BMC Bioinformatics 19:397.

638   Zeller MA, Li G, Harmon KM, Zhang J, Vincent AL, Anderson TK, Gauger PC. 2018. Complete

639   Genome Sequences of Two Novel Human-Like H3N2 Influenza A Viruses,

640   A/swine/Oklahoma/65980/2017 (H3N2) and A/Swine/Oklahoma/65260/2017 (H3N2),

641   Detected in Swine in the United States. Microbiol Resour Announc 7:e01203-01218.

642   Zhao JH. 2007. A Genetic Analysis Package with R. Journal of Statistical Software 23:1-18.

643   Zhou NN, Senne DA, Landgraf JS, Swenson SL, Erickson G, Rossow K, Liu L, Yoon K-j,

644   Krauss S, Webster RG. 1999. Genetic reassortment of avian, swine, and human influenza

645   A viruses in American pigs. Journal of virology 73:8851-8856.

646

647

648   **FIGURE LEGENDS**

649   **Figure 1.** Relative genetic diversity, measured as effective population size (EPS), of the

650   predominant N2 neuraminidase clades in influenza A viruses detected from 2009-2018 in swine

651   in the United States (N2.1998 in blue and N2.2002 in red). Median EPS denoted by solid colored

652   lines with the 95% HPD shaded in the same color. Relative genetic diversity increased linearly in

653   the N2-1998 between 2014 and 2017, and between 2010 and 2016 in the N2.2002.

654   **Figure 2.** Maximum likelihood trees of the (A) N2.1998 and (B) N2.2002 lineage neuraminidase

655   (NA) genes with N2 sequences from August 2009 to July 2018. Phylogenetic trees show

656   sustained (> 10 detections) paired hemagglutinin (HA) clade transitions through colored changes,

657   being either black to red or red to blue. NA genetic clades are annotated on the right of each

658   panel. A numbered arrow indicates a reassortment event where an N2 gene was paired with a

659   new HA genetic clade with subsequent sustained detection in the swine population. The 10

660   numbered arrows represent HA gene exchanges as follows: (1) H3 C-IVF to H1 Delta1B; (2) H1

661   Delta1B to H1 Delta1A; (3) H1 Delta1B to H3 C-IVA; (4) H1 Delta1B to H3 2010.1; (5) H3 C-

662   IVB to H1 Alpha; (6) H1 Alpha to H3 2010.1; (7) H3 C-IVB to H3 C-IVA; (8) H3 C-IVA to H1

663   Delta1A; and (9) H1 Delta1A to H3 2010.1. Each sustained reassortment event demonstrated

664   increased genetic divergence, visualized as longer branch lengths in the phylogeny. The scale bar

665   represents approximately 10 nucleotide changes in the neuraminidase gene.

666   **Figure 3.** Tanglegram of the hemagglutinin (HA) and paired N2 neuraminidase (NA) genes in

667   US swine influenza A virus. HA and N2 NA lineages are indicated by color and are presented on

668   defining branches of the phylogeny. Connecting lines in the tanglegram are colored by the HA

669   clade and demonstrate a link between HA and N2 present in the same virus. Maximum

670   likelihood trees are presented as cladograms with branch lengths scaled proportionally to the

30

671 changes per site. Multiple, repetitive connecting lines between specific HA and N2 clades was

672 suggestive of non-random pairing. Reassortment is indicated when connecting lines move across

673 multiple clades. For example, the H3.Cluster IVA and H1.Delta2 demonstrated connecting lines

674 with multiple N2 clades, suggesting reassortment events occurred that subsequently became

675 fixed in the swine population. Genetic divergence of HA was observed during these reassortment

676 events based on branch length.

677 **Figure 4.** Bayes factors (Ln BF) representing the confidence of transition events of the

678 hemagglutinin (HA) gene by clade for neuraminidase (NA) gene by N2 clades from 2009 to

679 2018. Higher Ln Bayes factors represent in warmer colors are higher evidence for reassortment;

680 uncolored squares present an absence of the transition between genes. The N2.1998A clade

681 demonstrated the least frequent reassortment of the N2 clades, but there was a transition from

682 H3.ClusterIVF to H1.Delta1B. The small numbers of detected reassortment events for N2.1998B

683 had strong support due to near exclusive pairing between N2.1998B and H1.Delta2. Both

684 N2.2002A and N2.2002B NA genes demonstrated evidence for multiple reassortment events.

685 **Figure 5.** (A) Geographic representation of each N2 clade from 2009-2018. N2 clade was

686 dependent on location (chi squared: $p < 0.001$). N2.1998A was observed with few detections in

687 limited locations in the Midwest. N2.1998B is detected primarily in North Carolina, with few

688 detections present in the Midwest including Indiana and Iowa. N2.2002A was detected in the

689 Midwest but was rarely detected in North Carolina. N2.2002B was detected in the Midwest as

690 well as North Carolina. (B) Bayes factors for inferred state-to-state movement of N2 clades

691 demonstrating the frequency with which particular genetic clades are moving from particular

692 states on the y-axis to other states on the x-axis.

31

693   SUPPLEMENTARY FIGURE LEGENDS

694   **Figure S1**. Effective population size representing the relative genetic diversity of the N2 clades

695   detected in North American swine from 2009-2018, N2.1998A, N2.1998B, N2.2002A, and

696   N2.2002B. Temporal dominance in genetic diversity of NA N2 was apparent after separating the

697   clades. The N2.2002A genetic diversity peaked in mid-2012, N2.2002B peaked in mid-2015, and

698   N2.1998B peaked at approximately mid-2016. N2.2002A and N2.2002.B maintained moderate

699   genetic diversity at the end of the study period.

700   **Figure S2.** Regression models fitted to patristic distance from the oldest gene on a phylogenetic

701   tree for neuraminidase (NA) from a donor hemagglutinin in red (HA) to a recipient HA in blue.

702   The slope of the regression line is representative of the mutation rate of the NA paired with

703   either the donor or recipient HA. (A) IAV containing N2.1998A paired with H3.ClusterIVF

704   donating NA genes to IAV with H1.Delta1B (Figure 2, reassortment event 1), (B) N2.2002A

705   paired with H1.Delta1B donating NA genes to H3.2010.1 (Figure 2, reassortment event 4), (C)

706   N2.2002A paired with H3.ClusterIVB donating NA genes to IAV with H1.Alpha (Figure 2,

707   reassortment event 5), (D) N2.2002A paired with H1.Alpha donating NA genes to IAV with

708   H3.2010.1 (Figure 2, reassortment event 6), (E) N2.2002B paired with H3.ClusterIVB donating

709   NA genes to IAV with H3.ClusterIVA (Figure 2, reassortment event 7), (F) N2.2002B paired

710   with H3.ClusterIVA donating NA genes to IAV with H1.Delta1A (Figure 2, reassortment event

711   8), (G) N2.2002B paired with H3.ClusterIVA donating NA genes to IAV with H1.2010.2 (Figure

712   2, reassortment event 9). The Chow test indicated a significant break, indicating that two linear

713   models was more valid than one for subsets A, E, and F. Linear breaks may indicate a temporal

714   shift in mutation rate.

715   **Figure S3.** Neuraminidase clades over time paired with (A) H1.Delta1a, (B) H3.ClusterIVA, (C)

716   H3.2010.1 in IAV detected from 2009-2018. Graphs depict the number of detections of each HA

717   and NA pairing. A change in predominant HA and NA pairings occurred over time.

718   **Figure S4.** (A) The observed number of HA and N2 neuraminidase clade pairings represented in

719   the dataset. (B) The standardized residuals of the HA and N2 neuraminidase clade pairings based

720   on the chi squared test for independence ($p < 0.001$) performed separately on the N2.1998 and

721   N2.2002 clades. The post hoc analysis demonstrated individual contribution to the p-value of

722   each clade pairing. Red cells represent HA-N2 pairing observed more than expected and blue

723   cells represent pairings observed less than expected if pairings were random. The results of this

724   analysis support that HA pairing with N2 was not random.

725   **Figure S5.** (A) Relative diversity of the N2.1998B clade and the H1.Delta2 clade over time from

726   2009-2018. Relative diversity of H1.Delta2 and N2.1998B was not correlated until after 2012,

727   prior to which the H1.Delta2 was primarily paired with N2.2002 . (B) Relative diversity of the

728   N2.2002A clade and the H1.Delta1B clade that were paired from 2009-2018. Median EPS were

729   denoted by lines with the 95% higher posterior density shaded in the same color. Temporally

730   matched changes at similar magnitudes suggested a correlation between the diversity of the

731   shown NA-HA.

732   **Figure S6.** The mean nucleotide substitution rate of H1, H3, and N2 clades. Substitution rates

733   were calculated using BEAST2. Each codon position represented a different log normal relaxed

734   clock model, and a GTR+ Γ substitution model with four categories. The three clock rates for

735   each codon position were averaged to determine the mean nucleotide substitution rate across the

736   entire gene. N2.2002A had the lowest mean substitution rate at 0.0041, while H3.2010.1 had the

737 highest substitution rate at 0.0050. The mean HA substitution rate was 0.0005

738 nucleotides/site/year faster than the NA mean substitution rate.

739
740
741
742

Table 1. Amino acid sites detected as undergoing positive selection in the neuraminidase N2 gene of swine influenza A viruses. Instances of episodic and pervasive selection were determined using a mixed-effects model of evolution (MEME) part of the HyPhy package. Sites under positive selection were also assessed with the FEL and SLAC methods in HyPhy. Sites common between MEME and FEL were annotated with '*' and those overlapping between MEM and SLAC were annotated with '+.'

Group	Site	Structural Domain	alpha	beta+	p+	p value	LRT	# branches	Amino Acid Polymorphisms
N2.1998A (n = 115)	23	transmembrane	1.589	167.178	0.018	0.0479	4.54	1	I 91.3%, L 7.0%, F 0.9%, Y 0.9%
	147*	globular head	0	11.394	0.309	0.0547	4.28	4	D 93.0%, G 4.3%, N 2.6%
	414	globular head	2.024	20.691	0.113	0.0712	3.77	3	G 97.4%, S 1.7%, L 0.9%
	434	globular head	0	51.322	0.075	0.0424	4.77	3	T 84.3%, A 14.8%, N 0.9%
N2.1998B (n = 453)	19*	transmembrane	0	3.368	0.354	0.0435	4.72	7	A 76.4%, T 21.2%, G 1.1%, V 1.1%, S 2.2%
	40*	stalk	0	1.312	1	0.0799	3.55	7	C 76.8%, Y 22.1%, S 0.7%, F 0.2%, H 0.2%
	66	stalk	0	248.872	0.005	0.002	10.76	2	V 94.5%, M 5.3%, F 0.2%
	77*	stalk	0	1.415	1	0.0797	3.56	9	K 77.9%, I 18.1%, E 0.7%, M 2.6%, N 0.2%, T 0.2%, V 0.2%
	82*	stalk	0	1.509	0.646	0.0657	3.93	6	A 96.7%, S 1.8%, V 1.1%, T 0.4%
	86*	stalk	0	1.225	1	0.087	3.39	8	N 81.2%, S 17.7%, T 0.7%, D 0.2%, R 0.2%
	149	globular head	0	81.658	0.009	0.0077	8.11	2	F 78.6%, V 21.2%, C 0.2%
N2.2002A (n = 256)	269*+	globular head	0	2.943	1	0.0121	9.81	16	I 67.2%, L 20.6%, V 9.8%, M 1.3%, S 0.7%, R 0.4%
	313*+	globular head	0	1.873	0.997	0.0144	7.21	8	V 85.0%, I 6.2%, A 3.1%, T 2.2%, M 1.8%, D 1.3%, G 0.4%
	468*	globular head	0	0.879	1	0.0894	6.88	5	P 97.6%, L 0.7%, S 1.8%
	50*+	stalk	0	1.229	1	0.0267	3.34	6	A 73.8%, V 12.9%, T 11.3%, S 2.0%
	126*	globular head	0	13.975	0.105	0.0282	5.57	7	P 95.7%, L 2.3%, S 0.8%, H 0.8%, R 0.4%
N2.2002B (n = 310)	127*+	globular head	0	1.262	1	0.0295	5.47	10	D 87.1%, N 7.8%, E 2.0%, G 2.0%, V 1.2%
	312*+	globular head	0.479	6.151	0.325	0.0798	3.56	7	I 87.1%, V 9.0%, T 2.0%, A 2.0%
	19	transmembrane	0.583	324.387	0.006	0.0052	8.88	2	T 97.4%, S 1.3%, L 0.6%, A 0.3%, I 0.3%
	41	stalk	0.946	41.509	0.023	0.0827	3.49	1	D 57.7%, N 41.6%, G 0.3%, S 0.3%
	199*+	globular head	0.518	4.964	1	0.0017	11.05	0	K 64.5%, E 20.3%, N 5.5%, G 5.5%, R 3.9%, Q 0.3%
	331	globular head	0.656	27.71	1	0.0689	3.84	1	S 89.7%, R 8.7%, K 0.6%, N 0.6%, I 0.3%
	332	globular head	0.944	213.11	0.006	0.0148	6.82	2	F 82.3%, L 17.1%, S 0.6%
	344*+	globular head	0	1.094	1	0.0467	4.59	5	E 84.8%, K 14.5%, R 0.6%
	399	globular head	0.827	13.423	0.151	0.0627	4.02	13	G 42.6%, D 26.8%, E 24.5%, S 5.8%, K 0.3%

743

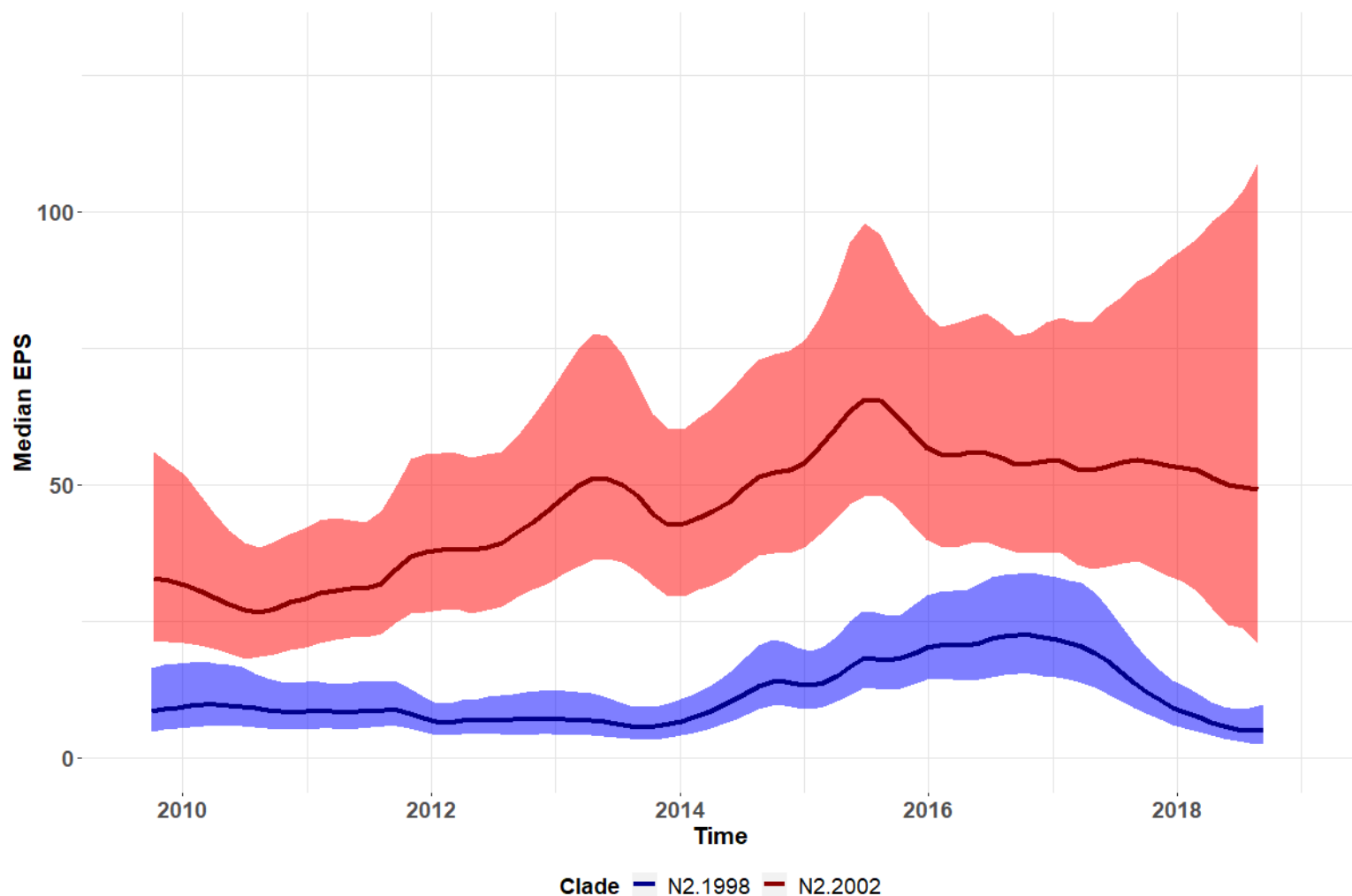| 402* | globular head | 0 | 0.877 | 1 | 0.0484 | 4.52 | 7 | N 95.8%, D 3.2%, H 0.3%, K 0.3%, S 0.3% |
| 412* | globular head | 0 | 1.629 | 0.599 | 0.0514 | 4.4 | 7 | V 71.9%, I 28.1% |

Figure 1. Relative genetic diversity, measured as effective population size (EPS), of the predominant N2 neuraminidase clades in influenza A viruses detected from 2009-2018 in swine in the United States (N2.1998 in blue and N2.2002 in red). Median EPS denoted by solid colored lines with the 95% HPD shaded in the same color. Relative genetic diversity increased linearly in the N2-1998 between 2014 and 2017, and between 2010 and 2016 in the N2.2002.
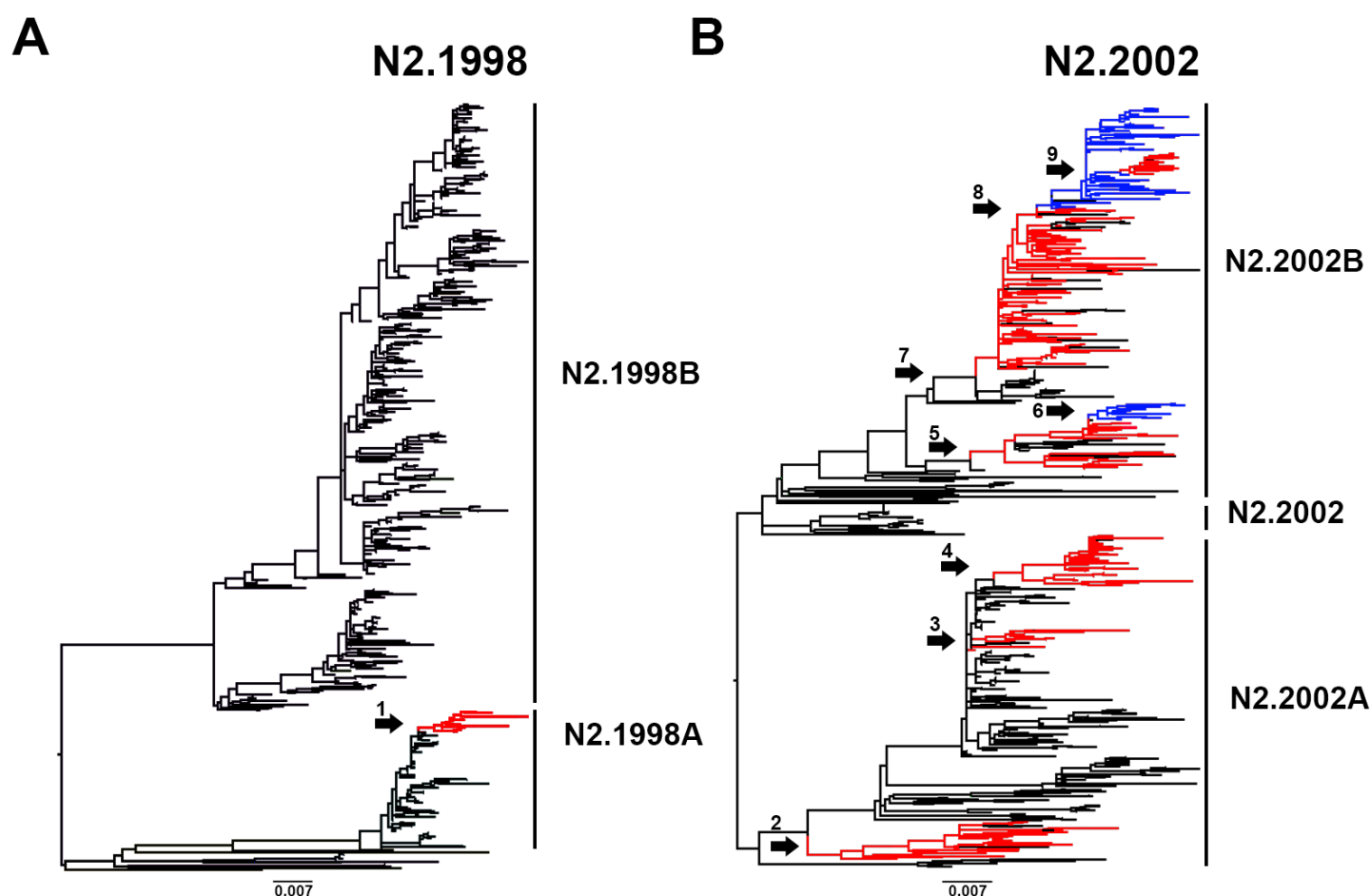
Figure 2. Maximum likelihood trees of the (A) N2.1998 and (B) N2.2002 lineage neuraminidase (NA) genes with N2 sequences from August 2009 to July 2018. Phylogenetic trees show sustained (> 10 detections) paired hemagglutinin (HA) clade transitions through colored changes, being either black to red or red to blue. NA genetic clades are annotated on the right of each panel. A numbered arrow indicates a reassortment event where an N2 gene was paired with a new HA genetic clade with subsequent sustained detection in the swine population. The 10 numbered arrows represent HA gene exchanges as follows: (1) H3 C-IVF to H1 Delta1B; (2) H1 Delta1B to H1 Delta1A; (3) H1 Delta1B to H3 C-IVA; (4) H1 Delta1B to H3 2010.1; (5) H3 C-IVB to H1 Alpha; (6) H1 Alpha to H3 2010.1; (7) H3 C-IVB to H3 C-IVA; (8) H3 C-IVA to H1 Delta1A; and (9) H1 Delta1A to H3 2010.1. Each sustained reassortment event demonstrated increased genetic divergence, visualized as longer branch lengths in the phylogeny. The scale bar represents approximately 10 nucleotide changes in the neuraminidase gene.
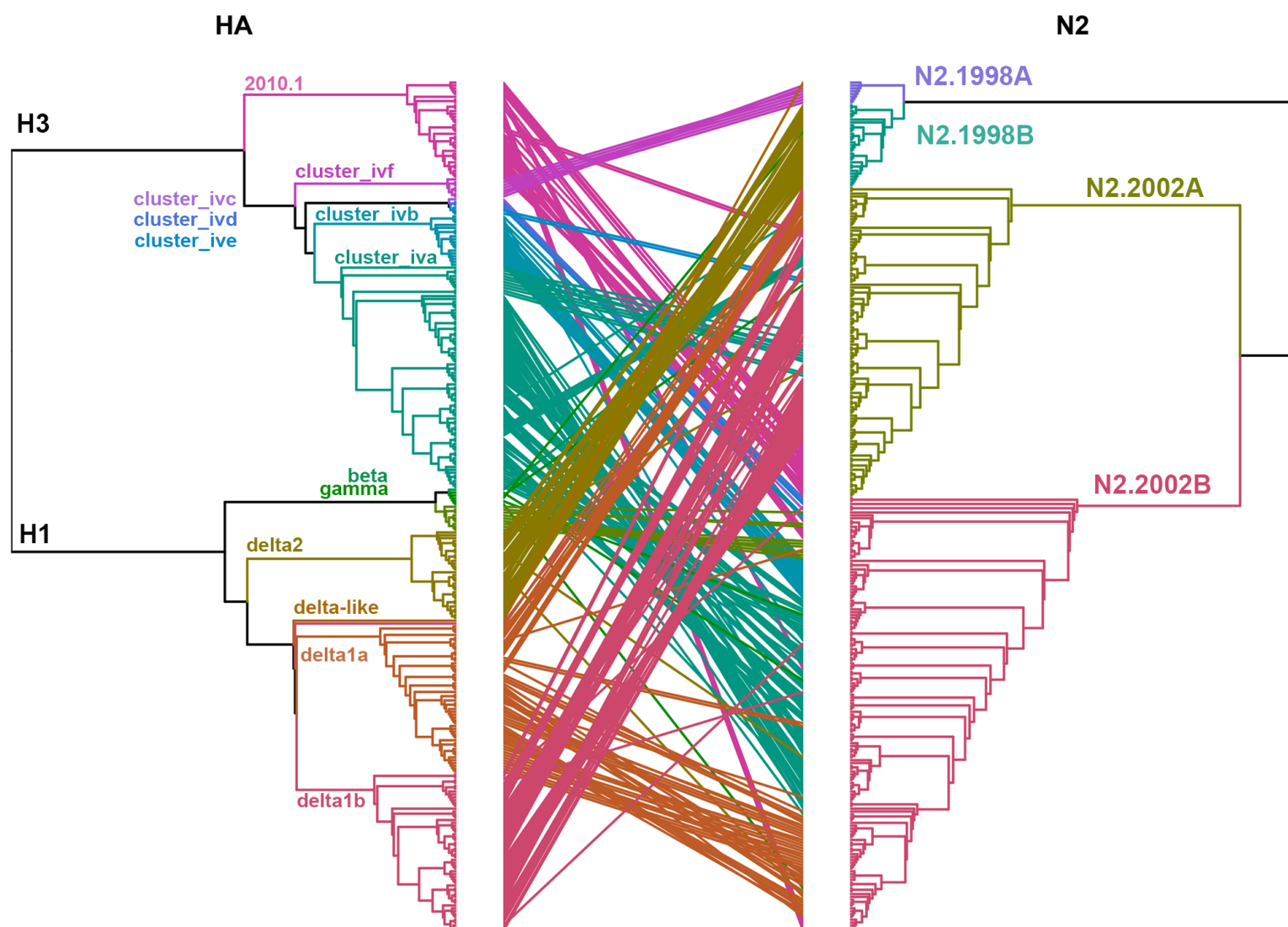
Figure 3. Tanglegram of the hemagglutinin (HA) and paired N2 neuraminidase (NA) genes in US swine influenza A virus. HA and N2 NA lineages are indicated by color and are presented on defining branches of the phylogeny. Connecting lines in the tanglegram are colored by the HA clade and demonstrate a link between HA and N2 present in the same virus. Maximum likelihood trees are presented as cladograms with branch lengths scaled proportionally to the changes per site. Multiple, repetitive connecting lines between specific HA and N2 clades was suggestive of non-random pairing. Reassortment is indicated when connecting lines move across multiple clades. For example, the H3.Cluster IVA and H1.Delta2 demonstrated connecting lines with multiple N2 clades, suggesting reassortment events occurred that subsequently became fixed in the swine population. Genetic divergence of HA was observed during these reassortment events based on branch length.

Figure 4. Bayes factors (Ln BF) representing the confidence of transition events of the hemagglutinin (HA) gene by clade for neuraminidase (NA) gene by N2 clades from 2009 to 2018. Higher Ln Bayes factors represent in warmer colors are higher evidence for reassortment; uncolored squares present an absence of the transition between genes. The N2.1998A clade demonstrated the least frequent reassortment of the N2 clades, but there was a transition from H3.ClusterIVF to H1.Delta1B. The small numbers of detected reassortment events for N2.1998B had strong support due to near exclusive pairing between N2.1998B and H1.Delta2. Both N2.2002A and N2.2002B NA genes demonstrated evidence for multiple reassortment events.
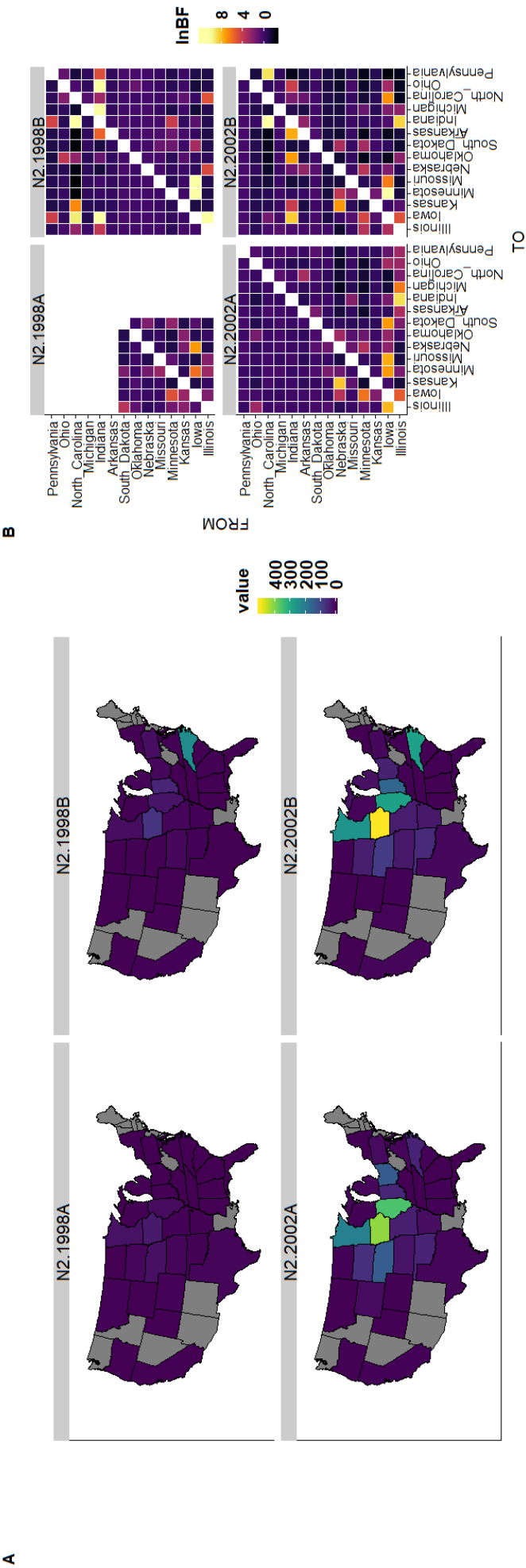
Figure 5. (A) Geographic representation of each N2 clade from 2009-2018. N2 clade was dependent on location (chi squared: p < 0.001). N2.1998A was observed with few detections in limited locations in the Midwest. N2.1998B is detected primarily in North Carolina, with few detections present in the Midwest including Indiana and Iowa. N2.2002A was detected in the Midwest but was rarely detected in North Carolina. N2.2002B was detected in the Midwest as well as North Carolina. (B) Bayes factors for inferred state-to-state movement of N2 clades demonstrating the frequency with which particular genetic clades are moving from particular states on the y-axis to other states on the x-axis.