

# Symptom evolution following the emergence of maize streak virus

Adérito L. Monjane<sup>1,2,\*</sup>, Simon Dellicour<sup>3,4,\*†</sup>, Penelope Hartnady<sup>5</sup>, Kehinde A. Oyeniran<sup>5</sup>, Betty E. Owor<sup>6</sup>, Marion Bezeidenhout<sup>7</sup>, Daphne Linderme<sup>7</sup>, Rizwan A. Syed<sup>7</sup>, Lara Donaldson<sup>7</sup>, Shane Murray<sup>7</sup>, Edward P. Rybicki<sup>7</sup>, Anders Kvarnheden<sup>2</sup>, Elhman Yazdkhasti<sup>2</sup>, Pierre Lefeuvre<sup>8</sup>, Rémy Froissart<sup>9</sup>, Philippe Roumagnac<sup>10,11</sup>, Dionne N. Shepherd<sup>7,12</sup>, Gordon W. Harkins<sup>13</sup>, Marc A. Suchard<sup>14</sup>, Philippe Lemey<sup>3</sup>, Arvind Varsani<sup>15,16</sup>, Darren P. Martin<sup>5,†</sup>

<sup>1</sup> Fish Health Research Group, Norwegian Veterinary Institute, Norway

<sup>2</sup> Department of Plant Biology, Swedish University of Agricultural Sciences, Sweden

<sup>3</sup> Department of Microbiology, Immunology and Transplantation, Rega Institute, Laboratory for Clinical and Epidemiological Virology, KU Leuven - University of Leuven, Leuven, Belgium

<sup>4</sup> Spatial Epidemiology Laboratory (SpELL), Université Libre de Bruxelles, Brussels, Belgium

<sup>5</sup> Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Observatory, 7925, Cape Town, South Africa

<sup>6</sup> Department of Agricultural Production, School of Agricultural Sciences, Makerere University, Kampala, Uganda

<sup>7</sup> Molecular and Cell Biology Department, University of Cape Town, Rondebosch, 7700, Cape Town, South Africa

<sup>8</sup> CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>9</sup> University of Montpellier, Centre National de la Recherche Scientifique (CNRS), Institut de recherche pour le développement (IRD), UMR 5290, Maladie Infectieuses & Vecteurs : Écologie, Génétique Évolution & Contrôle" (MIVEGEC), Montpellier, France

<sup>10</sup> CIRAD, BGPI, Montpellier, France

<sup>11</sup> BGPI, INRA, CIRAD, SupAgro, Univ Montpellier, Montpellier, France

<sup>12</sup> Research Office, University of Cape Town, Rondebosch, 7700, Cape Town, South Africa

<sup>13</sup> South African MRC Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Bellville 7535, South Africa

<sup>14</sup> Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

<sup>15</sup> The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, USA

<sup>16</sup> Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Rondebosch, 7700, Cape Town, South Africa

\* These authors contributed equally to this work

† Corresponding authors: [simon.dellicour@ulb.ac.be](mailto:simon.dellicour@ulb.ac.be), [darrenpatrickmartin@gmail.com](mailto:darrenpatrickmartin@gmail.com)

**Keywords:** virulence, pathogenicity, evolutionary trade-off, chlorosis

40 **Abstract**

41 For pathogens infecting single host species evolutionary trade-offs have previously been  
42 demonstrated between pathogen-induced mortality rates and transmission rates. It remains unclear,  
43 however, how such trade-offs impact sub-lethal pathogen-inflicted damage, and whether these  
44 trade-offs even occur in broad host-range pathogens. Here, we examine changes over the past 110  
45 years in symptoms induced in maize by the broad host-range pathogen, maize streak virus (MSV).  
46 Specifically, we use the quantified symptom intensities of cloned MSV isolates in differentially  
47 resistant maize genotypes to phylogenetically infer ancestral symptom intensities and check for  
48 phylogenetic signal associated with these symptom intensities. We show that whereas symptoms  
49 reflecting harm to the host have remained constant or decreased, there has been an increase in how  
50 extensively MSV colonizes the cells upon which transmission vectors feed. This demonstrates an  
51 evolutionary trade-off between amounts of pathogen-inflicted harm and how effectively viruses  
52 position themselves within plants to enable onward transmission.

53

54 **Impact statement**

55 Phylogenetic analyses demonstrate an evolutionary trade-off between the amount of harm inflicted  
56 by a broad host-range virus and how effectively the virus positions itself within plants to enable  
57 onward transmission.

58

## 59 Introduction

60 Maize was first introduced to Africa from South America in the early 1500s, and within 300 years  
61 became one of the most important food crops on the continent (McCann, 2001). The rapid  
62 intensification of maize cultivation in the 1800s coincided with the emergence of maize streak  
63 disease in southern Africa (Fuller, 1901; Harkins et al., 2009); a disease that has subsequently spread  
64 to all maize growing parts of Africa and today persists as one of the most serious threats to the food  
65 security of sub-Saharan Africans (Bosque-Pérez, 2000; Martin and Shepherd, 2009). The causal  
66 agent, a maize-adapted strain of maize streak virus (MSV; genus *Mastrevirus*, family *Geminiviridae*),  
67 is apparently a recombinant of two wild-grass adapted MSV strains: MSV-B and MSV-F/G (Varsani et  
68 al., 2008). Today the geographical range of this “MSV-A” strain (Krabberger et al., 2017) coincides  
69 with the combined ranges of its various insect transmission vector species, all of which are  
70 leafhoppers in the genus *Cicadulina*, and includes most of Africa and various Atlantic and Indian  
71 Ocean islands (Bosque-Pérez, 2000). Crucially, the host-range of MSV-A is also broadly coincident  
72 with that of its vector species, and includes more than 100 different grass species (Damsteegt, 1983;  
73 Krabberger et al., 2017).

74

75 The rise of MSV-A is a typical viral emergence story, and as such, efforts to reconstruct what  
76 happened after MSV-A first began infecting maize could potentially yield generalisable insights into  
77 how pathogens evolve following successful host-range changes. It would be particularly informative  
78 to determine how the symptoms that MSV-A induces in maize have evolved since the virus first  
79 emerged in the mid to late 1800s. Although it is expected that pathogens like MSV-A will evolve over  
80 time to maximise their probability of transmission, there is a spectrum of possible ways in which this  
81 transmission optimisation imperative might impact the evolution of pathogen-mediated molecular  
82 processes that underlie infection symptoms (Bull and Luring, 2014; Cressler et al., 2016;  
83 Doumayrou et al., 2013; Escriu et al., 2003; Mauck et al., 2010; Read, 1994). Various theories,  
84 focusing primarily on trade-offs between host mortality rates, infection durations and pathogen  
85 transmission rates, predict that under natural selection, transmission probabilities and hence the

86 basic reproductive number of a pathogen should be optimised at intermediate rates of pathogen-  
87 induced host mortality (Alizon et al., 2009; Anderson and May, 1982).

88

89 Because of their general disregard for sub-lethal amounts of host harm, it is unclear how relevant  
90 such theories might be for explaining the evolution of MSV-induced disease symptoms. MSV and  
91 many other plant viruses only rarely kill their hosts (i.e. pathogen attributed mortality rates are  
92 generally close to zero) and infections generally persist for the remainder of a host's life. Also,  
93 selection pressures acting on crop-infecting viruses such as MSV will not necessarily yield the same  
94 outcomes as for pathogens that infect humans, other animals or uncultivated plants because crop  
95 plants frequently have either very short natural lifespans or are destructively harvested long before  
96 their natural lifespans expire. Regardless of how these factors might influence the evolution of  
97 symptoms over time, it is expected that infection symptoms that are associated with decreased  
98 transmission probabilities - perhaps high host mortality rates or reduced host reproduction rates -  
99 should evolve to either remain mild, or to become less severe. Conversely, symptoms that are  
100 associated with increased transmission probabilities - such as increased lesion sizes or lesion colours  
101 that are more attractive to insect vector species - should evolve to become more extreme.  
102 Therefore, all that would be required to detect relationships between transmission probabilities and  
103 the intensity of virus-induced disease symptoms would be to first identify quantifiable symptom  
104 types that might impact transmission, and then to determine whether these vary in some concerted  
105 way during the adaptation of a virus such as MSV to a new host species such as maize.

106

107 The most obvious symptom of a MSV-A infection that is likely to impact transmission is the  
108 occurrence of distinct yellow streaks centred along the primary, secondary and tertiary veins of  
109 maize leaves. These streaks arise as a consequence of impaired chloroplast formation within MSV-  
110 infected photosynthesising cells that surround veins (Engelbrecht, 1982; Pinner et al., 1993). The  
111 proportion of a leaf's surface that is covered in streaks reflects, at least in part, the degree to which  
112 viruses originating in the vasculature manage to infect surrounding photosynthesising tissues during  
113 the early developmental stages of the leaf's formation (Lucy, 1996). The intensity of chlorosis within



114 streaks varies both between MSV variants infecting the same host, and between different host  
115 genotypes that are infected by the same MSV variant (Martin et al., 1999; Pinner et al., 1993;  
116 Schnippenkoetter et al., 2001). Depending on the number of intact chloroplasts within MSV-infected  
117 photosynthesising cells, streaks can range in colour from pale green through yellow to white  
118 (Engelbrecht, 1982; Pinner et al., 1993). Therefore, whereas the spatial arrangements and  
119 dimensions of chlorotic lesions on MSV-infected maize leaves directly indicate the locations and  
120 numbers of leaf cells that are infected by MSV (Lucy, 1996), the colour of chlorotic streaks indicates  
121 the degree to which photosynthesis is impaired within infected cells (Pinner et al., 1993).

122

123 Another easily quantifiable symptom of MSV infections that may impact the probability of virus  
124 transmission is leaf stunting. While leaf stunting is often attributed to reduced carbohydrate  
125 production due to virus-mediated impairment of photosynthesis (Shepherd et al., 2010), it is also  
126 probably a consequence of MSV directly interfering with the cell cycle progression control  
127 mechanisms of infected cells (Lucy, 1996; McGivern et al., 2005; Shepherd et al., 2005). MSV-  
128 induced cell cycle abnormalities could also account for other less commonly described bi-lateral leaf  
129 asymmetry symptoms. These symptoms can include off-centre leaf midribs, bending or twisting of  
130 leaves where one half of a leaf is more stunted than the other, and curling or corrugation of leaves  
131 caused by varying degrees of veinal swelling along the leaf-length.

132

133 Whereas measurements of leaf stunting or leaf deformation might be used as indicators of MSV-  
134 mediated harm to host plants, measurements of chlorotic areas - the proportion of leaf surface  
135 areas that are covered by streaks - might be used as indicators of how effectively MSV colonizes host  
136 cells. Since MSV can only be acquired by its insect transmission vectors when they feed directly on  
137 chlorotic lesions (Peterschmitt et al., 1992), and no report exists of any successful MSV transmissions  
138 from MSV-infected plants that do not display streak symptoms, chlorotic leaf areas are likely  
139 correlated to some degree with transmission probabilities. Further, measurements of streak colours  
140 could be indicators of both harm to the host, and the transmission fitness of MSV. Specifically,  
141 whereas the intensity of chlorosis within streaks reflects the extent of chloroplast destruction within

142 infected photosynthesising cells, it could also be associated with the probability of onward MSV  
143 transmission. This is because the sap-feeding insects that transmit viruses such as MSV tend to  
144 prefer pale-yellow over green when selecting feeding sites on leaves (Fereres and Moreno, 2009;  
145 Hodge and Powell, 2008; Isaacs et al., 1999).

146

147 Irrespective of the precise underlying genetic or physiological causes of MSV-induced disease  
148 symptoms, any associations that exist between the fitness of MSV and the intensity of the infection  
149 symptoms that it causes, should be detectable by quantitative analyses of different symptom types  
150 that are produced in maize by an assortment of MSV isolates sampled over the ~110 years following  
151 the emergence of the virus as a maize pathogen.

152

153 Here we undertake such an analysis. Since there are no available MSV samples from before 1979, we  
154 use the quantified chlorotic areas, intensities of chlorosis, leaf deformation and leaf stunting  
155 symptoms caused by 59 diverse field isolates of MSV-A sampled between 1979 and 2007 to  
156 phylogenetically infer the evolution of symptoms produced by MSV-A in maize between ~1900 and  
157 2007. Finally, to verify the accuracy of these symptom inferences we infer the genome sequences of  
158 seven ancestral MSV-A variants, synthesise infectious clones of these *in vitro* and directly compare  
159 actual and inferred symptom intensities. What we find is that patterns of MSV-A symptom evolution  
160 over the last century are broadly consistent with there being a trade-off between transmission  
161 probabilities and the amount of harm MSV inflicts on infected maize.

162

## 163 **Results and Discussion**

164 *Increased within host virus spread is not necessarily associated with harm to the host*

165 To determine the degrees to which symptoms of MSV-A infections have evolved since the time of  
166 the most recent common ancestor (MRCA) of all known MSV-A isolates, we quantified four infection  
167 symptom types (chlorotic areas, intensities of chlorosis, leaf deformation and leaf stunting) for 59  
168 MSV-A isolates and one MSV-B isolate in three differentially MSV-resistant maize genotypes. These  
169 were Golden Bantam (Sakata, South Africa) which is very susceptible to infection by MSV, displays

170 severe MSV infection symptoms and will hereafter be referred to as the “S” (for susceptible) maize  
171 genotype; STAR174 (Starke Ayres, South Africa) which has a moderate degree of resistance to MSV,  
172 displays moderate MSV infection symptoms and will hereafter be referred to as the “M” (for  
173 moderate) maize genotype, and PAN77 (Pannar Seed Company, South Africa) which has a high  
174 degree of resistance to MSV, displays mild MSV infection symptoms, and will hereafter be referred  
175 to as the “R” (for resistant) maize genotype.

176

177 We first attempted to determine whether evidence existed of correlations between the measured  
178 symptom types (Figure 1). It was apparent that in the S maize, degrees of leaf stunting caused by  
179 MSV were positively correlated with chlorotic areas and leaf deformation. In the R maize there was  
180 also a strong positive correlation between leaf stunting and chlorotic areas, but there was a negative  
181 correlation between leaf stunting and the intensity of chlorosis. Therefore, for both the S and R  
182 maize the higher the proportion of photosynthesising leaf cells that are infected by MSV, the greater  
183 the amount of leaf stunting that occurs. Chlorotic areas were negatively correlated with the intensity  
184 of streak chlorosis in the M and R maize (Figure 1). This indicates that as greater numbers of  
185 photosynthesising cells in the leaves of the more MSV resistant maize genotypes become infected,  
186 the amount of chloroplast destruction within the infected cells tends to decrease.

187

188 Collectively, these results indicate that symptoms which are indicative of harm to the host - such as  
189 the intensity of chlorosis, leaf deformation and leaf stunting - are not, as one might expect, always  
190 positively correlated with chlorotic areas; a symptom type indicative of how successfully viruses  
191 have colonized photosynthesising cell populations within maize leaves (Lucy, 1996). It is apparent,  
192 therefore, that harm to the host could, to some degree at least, be mitigated without strongly  
193 impacting the capacity of MSV to infect and replicate within host cells. Put simply, selection acting  
194 on particular symptom types will not necessarily impact how the other symptom types evolve.

195

196 *MSV symptom measurements contain detectable phylogenetic signals*

197 All of the symptom measurements for the 59 phenotyped MSV-A isolates contained detectable  
198 phylogenetic signals, in that sequences that were associated with similar symptom intensities  
199 tended to cluster together to some degree within the MSV-A phylogenetic tree (Figures 2 and 3, see  
200 also Figure 2-figure supplements 1 and 2). Specifically, Pagel's  $\lambda$  values - where 0 = no correlation  
201 and 1 = perfect correlation between symptom measurements and phylogenetic placement - for  
202 every combination of symptom and host type ranged between 0.251 for leaf deformation in the M  
203 maize genotype, and 0.706 for intensity of chlorosis in the M maize genotype, with the 95%  
204 credibility intervals of these values excluding zero for every symptom x host comparison (Figures 2  
205 and 3).

206

207 This association between the symptoms induced by MSV isolates and their phylogenetic placement  
208 indicated that, just as it is possible to infer ancestral genome sequences based on the phylogenetic  
209 relationships of an observed sample of genome sequences, it should be possible to meaningfully  
210 infer the symptom intensities that were induced by ancestral MSV variants based on the observed  
211 symptom intensities of the 59 phenotyped MSV-A isolates.

212

### 213 *Increase of MSV-induced chlorotic leaf areas over the past ~110 years of MSV evolution*

214 We used the continuous diffusion model implemented in BEAST 1.10 (Suchard et al., 2018) to infer  
215 the intensities of infection symptoms produced by ancestral MSV-A sequences since the emergence  
216 of MSV-A as a serious maize pathogen in the late 1800s. To identify general trends in symptom  
217 intensity changes across all of the sampled MSV-A lineages, we mapped symptom intensities to the  
218 time-scaled MSV phylogeny (i.e. the maximum clade consensus tree; Figures 2 and 3) and then  
219 examined the averaged symptom intensities associated with phylogenetic tree branches through  
220 time (Figure 4). This revealed that since 1979, the time of the earliest sampled MSV-A isolates  
221 investigated here, MSV-A lineages have been progressively inducing, on average: (1) higher  
222 percentage chlorotic areas in all tested hosts, but most notably in the S and M maize genotypes; (2)  
223 lower intensities of streak chlorosis in the M and R maize genotypes; (3) more leaf deformation in

224 the S maize genotype; and (4) slightly lower degrees of leaf stunting in the S and M maize genotypes  
225 (Figure 4).

226

227 Of all of the analysed symptom types, only average chlorotic areas on all three maize genotypes  
228 (Figure 4A), and leaf stunting in the R genotype (Figure 4D), were predicted to have progressively  
229 increased between 1900 and 1979. It is important to stress, however, that the credibility intervals of  
230 these average symptom intensities are particularly broad for the “pre-sampling” time-period  
231 between 1900 and 1979. This is because: (1) the uncertainty associated with predicted ancestral  
232 virus symptom intensity estimates is expected to increase as we extrapolate backwards in time  
233 before the earliest sampling dates; and (2), at any given time-point, different lineages were not all  
234 predicted to induce similar symptom intensities such that the averaged symptom intensities across  
235 these lineages are associated with an additional among-lineage degree of variance.

236

237 When considering the predicted intensity of the various symptom types along specific virus lineages,  
238 rather than averaging these across the entire sample of viruses, it is apparent that changes in  
239 symptom intensity were not predicted to have all occurred concomitantly across all MSV-A lineages.  
240 For example, when the predicted chlorotic area and leaf deformation estimates are overlaid on the  
241 MSV-A phylogeny (Figure 2A and 3A, respectively), it is apparent that multiple independent large  
242 increases in leaf deformation and decreases in chlorotic area are predicted to have occurred across  
243 various terminal branches of the phylogeny (Figures 2 and 3). The sampled viruses at the tips of  
244 these branches tend to display large increases in leaf deformation or decreases in induced chlorotic  
245 areas relative to closely related viruses, frequently in two or three of the maize genotypes – and are  
246 dispersed throughout the phylogeny (i.e. they are not all close relatives to each other). The fact that  
247 there is no clustering within the phylogeny of viruses that display either large decreases in chlorotic  
248 area or large increases in leaf deformation symptoms compared to their nearest relatives, suggests  
249 that both increased leaf deformation and decreased chlorotic area phenotypes are maladaptive.

250

251 Conversely, the phylogenetic clustering of viruses that all share similar increases or decreases in the  
252 intensity of particular symptoms relative to the other sampled viruses would be consistent with  
253 ancestral sequences having undergone adaptive changes in the intensity of the symptoms that they  
254 induce. There are multiple phylogenetic clusters containing viruses displaying decreased intensity of  
255 chlorosis, and/or increased chlorotic area phenotypes relative to those predicted for ancestral  
256 viruses (Figures 2 and 3): this is an indicator that both decreased intensities of chlorosis, and  
257 increased chlorotic areas might be adaptive. It is noteworthy, however, that some of these clusters  
258 of viruses that induce low intensities of chlorosis and/or high chlorotic areas, are not consistent  
259 between the three different maize hosts. This again suggests that such adaptations might, to some  
260 degree at least, be host-genotype specific.

261

262 Although there was an overall positive correlation between the chlorotic area and leaf stunting  
263 symptoms of the 59 phenotyped MSV isolates (especially in the S and R maize genotypes; Figure 1),  
264 there were no clear instances where both chlorotic areas and leaf stunting were predicted to have  
265 concomitantly increased in extent across individual internal branches of the MSV phylogeny. There  
266 were, however, a number of instances across terminal branches of the phylogeny where predicted  
267 decreases in chlorotic area were associated with concomitant predicted decreases in leaf stunting.  
268 The viruses at the tips of these branches induce lower chlorotic areas and/or less leaf stunting than  
269 their nearest relatives in two or three of the tested maize genotypes (Figure 3B).

270

#### 271 *Potential impact of isolate sampling*

272 It should be stressed here that each of the 59 phenotyped MSV-A isolates was a single clone that  
273 was sampled at random from a genetically heterogeneous population of viruses infecting an  
274 individual maize plant. It is well known that up to 10% of MSV genomes that are cloned from an  
275 infected plant will induce significantly reduced chlorotic areas and leaf stunting relative to other  
276 clones from the same infection that might better represent the dominant virus population within  
277 that plant (Boulton et al., 1991a, 1991b). It is plausible, therefore, that the overall correlation  
278 between chlorotic areas and leaf stunting observed for the 59 phenotyped virus clones (Figure 1) is

279 not solely a consequence of adaptive increases in induced chlorotic areas being causally linked with  
280 increased leaf stunting. Rather, the detected positive correlation between chlorotic areas and leaf  
281 stunting is probably attributable, at least in part, to the decreased chlorotic areas and leaf stunting  
282 symptoms produced by the approximately 10% (i.e.  $\sim 6/59$ ) of examined MSV-A isolates which are  
283 expected to be slightly defective. Even if only affecting approximately 10% of the isolates, we  
284 acknowledge that this artefactual correlation could also, to some extent, impact the continuous  
285 character mapping of symptom intensities (Figures 2 and 3) and inferred changes in overall symptom  
286 intensities (Figure 4). However, because these 10% of isolates are a priori randomly distributed in  
287 the tree, their inclusion is mainly expected to add some random noise in the comparative analysis.  
288 Therefore, we do not expect a systematic bias in generating the trends we estimate and compare.

289

#### 290 *Observed ancestral virus symptoms confirm symptom inferences*

291 We selected seven nodes within the MSV-A tree and inferred the genome sequences of the ancestral  
292 viruses that are represented by these nodes (Figures 2 and 3; nodes labelled "A0" to "A6"). We then  
293 chemically synthesised these genome sequences, and directly measured the symptom phenotypes  
294 that they induced in the S, M and R maize genotypes (see the Material and Methods for further  
295 details).

296

297 We compared the observed symptoms produced by the synthesised ancestral viruses against those  
298 inferred by the Bayesian continuous trait mapping approach for these same viruses. This revealed  
299 that 16/21 of the chlorotic area, 14/21 of the intensity of chlorosis, 11/21 of the leaf deformation,  
300 and 16/21 of the leaf stunting measurements of the synthesised ancestral viruses were within the  
301 95% confidence intervals of those inferred based on the phenotypes of contemporary MSV  
302 sequences (Figure 5). However, when comparing measured and inferred symptom intensities within  
303 each genotype separately, we did not find any significant correlation (determination coefficients of  
304 linear regressions all associated with a p-value  $> 0.05$ ). This is unsurprising given (i) only seven  
305 viruses were examined in this way and (ii) the high variance associated with both the inferred and  
306 observed estimates of symptom intensities. Observed symptoms deviated most notably from

307 inferred symptoms for the R maize genotype, with 12/27 of the discordant symptom inferences  
308 being obtained for this host. In particular, 6/7 of the leaf deformation inferences for the R maize  
309 genotype were outside of the observed 95% confidence intervals of the measured estimates (Figure  
310 5C).

311

312 Overall, the best correlation between the observed and inferred symptom measurements were  
313 obtained for the leaf stunting (Pearson  $R^2 = 0.74$ ; slope = 0.87) and chlorotic area (Pearson  $R^2 = 0.58$ ;  
314 slope = 0.84) measurements, indicating good agreement between the inferred and observed  
315 measurements of these symptom types in the three different maize genotypes. The agreement  
316 between the inferred and observed leaf deformation and intensity of chlorosis measurements was  
317 less convincing (Figure 5B and Figure 5C). The greater degree of correlation between the observed  
318 and inferred chlorotic area and leaf stunting measurements is probably a consequence of these  
319 measurements containing more phylogenetic information than the others. For example, whereas  
320 the leaf stunting and chlorotic area measurements yielded the highest average Pagel's  $\lambda$  values, the  
321 leaf deformation measurements yielded both the lowest  $\lambda$  values and the lowest degree of  
322 correlation between observed and inferred symptom measurements (Supplementary File 1).

323

324 It should be stressed that disagreement between the observed and inferred symptom intensities for  
325 the synthesised ancestral viruses does not imply that the inferred estimates are incorrect. Ancestral  
326 sequences cannot usually be inferred with perfect accuracy: see Supplementary File 2 for a list of  
327 genome sites in the inferred ancestral genomes that could not be resolved with a high degree of  
328 confidence. It is therefore very unlikely than any of the inferred ancestral genomes that we have  
329 tested actually ever existed. The synthesised viruses are, at best, close approximations of actual  
330 ancestral viruses.

331

332 Nevertheless, the relative chlorotic areas and intensities of chlorosis induced by these synthetic  
333 approximations of actual ancestral viruses are broadly consistent with the computational inference  
334 that MSV-A is evolving towards the induction of increased chlorotic areas and decreased intensities



335 of chlorosis (Figure 2-figure supplement 3). For example, whereas the “oldest” of the synthesised  
336 ancestral viruses, A0 (which is estimated to have existed in ~1904), induced either the lowest or  
337 second lowest chlorotic areas of all the synthesised ancestral viruses on all of the maize genotypes  
338 tested (Figure 2-figure supplement 3A), it also induced the highest intensities of chlorosis in all of  
339 these genotypes (Figure 2-figure supplement 3B).

340

341 The degrees of leaf deformation and leaf stunting induced by the synthesised ancestral viruses on  
342 the three maize genotypes also indicated possible trends in the evolution of these symptom types  
343 that were not revealed by the Bayesian continuous trait mapping approach. Specifically, whereas  
344 successively more recent MSV variants have tended to induce higher degrees of leaf deformation in  
345 the S and M maize genotypes (Figure 2-figure supplement 3C), these variants have tended to induce  
346 lower degrees of leaf stunting in the S maize genotype and higher degrees of leaf stunting in the M  
347 genotype (Figure 2-figure supplement 3D).

348

349 *MSV-A is apparently evolving towards a maximization of host colonization with a minimization of*  
350 *host harm*

351 If the maize genotypes analysed here are in any way representative of those cultivated throughout  
352 Africa since 1900, these results suggest that MSV-A has over the past century been evolving towards  
353 the induction in most maize genotypes of increased chlorotic leaf areas and decreased intensities of  
354 chlorosis. Although in some maize genotypes this evolution has possibly been associated with  
355 increased degrees of leaf deformation and/or stunting, in others this does not seem to have been be  
356 the case.

357

358 Since MSV can only be acquired by its insect transmission vectors when they feed directly on  
359 chlorotic lesions (Peterschmitt et al., 1992), it is very likely that, by increasing the numbers of  
360 photosynthesizing maize leaf cells that it infects (i.e. by increasing induced chlorotic areas), MSV-A  
361 has increased the probability of its acquisition and onward transmission by insects.

362

363 The reason that decreased intensities of chlorosis might have been favoured by selection is not as  
364 obvious. Although it is plausible that selection has favoured an intensity of chlorosis that is  
365 maximally attractive to transmission vectors, it is similarly plausible that selection has also/instead  
366 favoured less chloroplast destruction so as to foster sufficient carbohydrate synthesis to prevent  
367 excessive host stunting. In this regard, further insect behavioural studies will be required to  
368 determine the relationships between the intensity of chlorosis, host stunting and transmission rates.

369

370 In any event, the changes in symptom intensities over time that we have both inferred and observed  
371 for MSV-A in maize are consistent with the hypothesis that the maximization of transmission  
372 probabilities which characterizes the adaptation of pathogens to their hosts (Alizon et al., 2009;  
373 Anderson and May, 1982), involves optimizing the balance between the amount of harm inflicted on  
374 the host and the extent to which pathogens colonize the host cell populations from which they will  
375 be transmitted to new hosts. Since MSV-A first arose in the mid- to late 1800s, it appears that the  
376 virus has achieved this balance by increasing the proportions of photosynthesising leaf cells that it  
377 infects, while concomitantly reducing the amount of chloroplast destruction that occurs within these  
378 cells.

379

380 The trade-off we identify is of significant importance because it implies that even when pathogens  
381 have host ranges that include over 100 species, within the context of individual host species, they  
382 may still conform to evolutionary imperatives that are similar to those of narrow host-range  
383 pathogens. Possibly more significant, however, is that a trade-off between sub-lethal symptom  
384 severities and transmission rates in MSV alludes to a more general form of the popular transmission-  
385 virulence trade-off hypothesis (Alizon et al., 2009; Anderson and May, 1982) which has, until now,  
386 focused almost exclusively on the relationship between pathogen-induced host mortality and  
387 pathogen transmission rates.

388

## 389 **Materials and Methods**

### 390 *Virus symptom quantification*

391 A set of 59 MSV field isolates (Supplementary File 3), representing the breadth of known diversity  
392 within the MSV-A strain, were used for symptom evaluation (Figure 2-figure supplement 4). These  
393 isolates were chosen from a collection of 704 fully sequenced MSV-A cloned genomes that were  
394 available in the University of Cape Town's MSV genome bank in December 2010. Their selection was  
395 based on sampling dates to ensure the broadest possible range of sampling times - from 1979  
396 through to 2007.

397 Specifically, isolates were selected as follows: (i) we made and tested agroinfectious constructs for  
398 every virus that was sampled before 1990 and for which we had a molecular clone sampled directly  
399 from a source plant, (ii) we obtained and tested all agroinfectious clones that had previously been  
400 constructed in the 1990s and 1980s, and (iii) for viruses sampled between 2000 and 2007, we chose  
401 isolates that were representative of the major phylogenetic clades of viruses in eastern, western and  
402 southern Africa (Figure 2-figure supplement 4). Infectious clones of eight of the 59 selected isolates  
403 have been previously described (Boulton et al., 1989; Martin et al., 2001; Peterschmitt et al., 1996).  
404 Infectious clones of the remaining 51 isolates were constructed for this work as described by (Martin  
405 et al., 2001). Briefly, this involved taking linearised MSV genomes cloned in an *E. coli* plasmid vector,  
406 producing head-to-tail tandem copies of these genomes within their respective vectors using a series  
407 of partial restriction enzyme digests, transferring these genome "dimers" into an *E. coli* / *Rhizobium*  
408 *radiobacter* (updated species name for *Agrobacterium tumefaciens*) binary plasmid vector (pBI121)  
409 and then transferring these recombinant plasmids, referred to as infectious constructs, into *R.*  
410 *radiobacter*.

411  
412 Symptoms produced by each of the 59 MSV-A isolates and one MSV-B isolate (AF239960; MSV-B  
413 [ZA\_VW\_1989]) - a closely-related grass-adapted MSV strain (Willment et al., 2002) - were examined  
414 in three different maize genotypes that are known to vary in their degree of MSV resistance: (i)  
415 Golden Bantam that is very susceptible to infection by MSV ("S" genotype), (ii) STAR174 that has a  
416 moderate degree of resistance to MSV ("M" genotype), and (iii) PAN77 that has a high degree of  
417 resistance to MSV ("R" genotype) maize genotype. Whereas the R genotype is a white maize hybrid  
418 containing the MSV-1 resistance gene that has been deployed throughout sub-Saharan Africa within

419 improved “MSV-resistant” maize varieties since the 1980s (Bosque-Perez, 2000), the S and M maize  
420 genotypes were chosen as representatives of less MSV resistant maize varieties. The degrees of MSV  
421 resistance displayed by these three maize genotypes span those observed for commercial cultivars  
422 that have been widely used throughout Africa since a least the 1980s (Martin et al., 1999). Crucially,  
423 despite the widespread availability since the 1980s of MSV resistant maize genotypes, most of the  
424 maize genotypes that are currently in use throughout Africa have no, or only low levels of, MSV  
425 resistance (Bosque-Pérez, 2000; Martin and Shepherd, 2009).

426

427 For every virus isolate and maize genotype combination, between 36 and 72 three-day old maize  
428 seedlings were inoculated in three to five independent experiments using *R. radiobacter* mediated  
429 delivery of infectious constructs as described by (Martin et al., 2001). In each experiment between  
430 12 and 16 plants were inoculated with only *R. radiobacter* growth medium to serve both as leaf size  
431 standards relative to which leaf stunting in infected plants would be calculated, and as uninfected  
432 cross-contamination controls. Inoculated plants were each grown for 35 days with 16 hours of light  
433 per day in approximately 150 grams of soil at between 20 and 25 degrees centigrade in a biosafety  
434 level 1 plant growth room. Leaves four, five and six from each maize seedling were harvested at 21,  
435 28 and 35 days post-inoculation, respectively. The second quarter from the base of each leaf was  
436 used for automated symptom quantification by image analysis. Chlorotic leaf area measurements of  
437 this leaf segment using image analysis have previously been shown to correlate with a five-point  
438 scale widely used by breeders for visually rating the severity of MSV symptoms when assessing the  
439 MSV resistance of maize genotypes in the field (Bosque-Pérez, 2000; Martin et al., 1999).

440

441 Leaf segments were scanned at 300 dpi on a HP Scanjet 5590P flatbed scanner (Hewlett Packard,  
442 USA), and Windows 32 bit bitmap images were analysed with the custom-written image analysis  
443 program, DIA (Martin, 2019). All leaf images were captured against a blue background with a circular  
444 size standard (a South African 50c coin). For each analysed image, the program DIA (Martin, 2019)  
445 initially identified and counted the number of pixels in the size standard and then identified the leaf

446 segments as the remaining “non-blue” objects in the image that were greater than 500 pixels in size.

447 For each identified leaf segment DIA quantified four different symptom phenotypes:

448 (1) The area of the leaf segment relative to the size standard. This was achieved by dividing the  
449 number of pixels within the leaf segment image by the number of pixels within the size  
450 standard image. Given that the coin has a known area it is then straightforward to adjust the  
451 unit of leaf area measurement to  $\text{mm}^2$ . For each agroinoculation experiment in a particular  
452 maize genotype, the leaf segment areas thus determined were each individually divided by  
453 the average leaf segment areas determined for all of the uninfected control plants of the  
454 same genotype from that experiment. For each leaf segment this number was then  
455 subtracted from one to yield the final leaf stunting score (such that lower values of this score  
456 would mean less stunting and higher values would mean more stunting).

457 (2) The degree of leaf deformation. This was achieved by fitting the smallest possible rectangle  
458 around the leaf segment and then determining the proportion of pixels within this rectangle  
459 that fell outside the leaf segment.

460 (3) The proportion of the leaf covered by chlorotic lesions. This was achieved by firstly  
461 identifying the main pixel colour categories in each leaf image and then determining which  
462 of these pixel colour categories represent chlorotic lesions. The main pixel colour categories  
463 were determined by randomly sampling 1000 pixels from the image of the leaf segment,  
464 calculating the Euclidean distance between the red, green and blue colour scores of every  
465 pair of pixels, constructing a UPGMA dendrogram from these scores and, from this  
466 dendrogram, identifying the ten main pixel colour categories. Each pixel in the leaf image  
467 was then assigned to one of the ten categories based on the Euclidean distance between the  
468 pixel red, green and blue colour scores and the mean of the red, green and blue colour  
469 scores of the ten pixel categories. Once all pixels were assigned to categories the average  
470 red, green and blue colour scores of each of the ten categories were determined. To  
471 determine which of the pixel colour categories represented chlorotic lesions, the ten pixel  
472 colour categories were partitioned into the two sets that maximized the difference between  
473 the average red+blue colour scores of the sets. The set of pixels with the highest average

474 red+blue score was taken as the “chlorotic lesion set” and the set of pixels with the lower  
475 score as the “non-chlorotic leaf area set”. The proportion of pixels assigned to the chlorotic  
476 lesion set divided by the total number of pixels in the leaf segment image was the  
477 proportion of the leaf area falling within chlorotic lesions.

478 (4) The average intensity of chlorosis. This was achieved by taking the average of the red, green  
479 and blue colour scores of every pixel in the chlorotic lesion set and dividing by 255 (such that  
480 pure white = 1 and pure black = 0).

481

482 Outlier measures were identified using Tukey’s method (Tukey, 1977): symptoms measures below  
483 the 1.5 interquartile range were discarded. Averages of chlorotic area, intensity of chlorosis, leaf  
484 deformation and leaf stunting measurements across leaves 4, 5 and 6 for each virus in each host  
485 were later used to infer the magnitudes of these symptom types in ancestral MSV-A variants.

486

#### 487 *Inference of ancestral MSV sequences*

488 An alignment containing the 59 MSV-A isolates for which symptoms were quantified (hereafter  
489 referred to as “phenotyped” isolates), together with an additional 630 MSV-A sequences and 182  
490 sequences belonging to MSV strains B through K (all of which can be obtained from GenBank), was  
491 constructed using MUSCLE with default settings (Edgar, 2004) and edited by eye using the suite of  
492 editing tools available in IMPALE (Khoosal and Martin, 2015). This 871 MSV full genome sequences  
493 alignment was screened for recombination using RDP4.46 (Martin et al., 2015) in a two-stage  
494 process. In an initial screen, default settings were used and all evidence of recombination within  
495 MSV strains other than MSV-A was removed from the alignment by identifying recombinationally-  
496 derived sequence fragments and removing these by replacing the associated nucleotide characters  
497 in the alignment file with the standard “gap” character, “-”. In the second screen, other than the use  
498 of no multiple testing p-value correction and a p-value cut-off of  $10^{-6}$ , default RDP4.46 settings were  
499 again used. This Bonferroni uncorrected p-value cut-off is still reasonably conservative due to the  
500 fact that during the ancestral sequence reconstructions we were specifically focused only on the

501 small subset of all detectable recombination events that impacted the inference of the seven  
502 ancestral MSV-A sequences that would later be chemically synthesized.

503

504 The ancestral sequence inference was carried out using RDP4.46 (Martin et al., 2015) and MrBayes  
505 3.2 (Ronquist et al., 2012) with recombination being accounted for in two distinct ways:

506 (1) Whenever the most recent common ancestor (MRCA) of a particular group of MSV-A  
507 sequences was to be determined, and every member of that group was inferred to possess  
508 evidence of one or more ancestral recombination events (i.e. the MRCA likely also contained  
509 evidence of that recombination event), the ancestral sequences were inferred with  
510 partitions at the identified recombination breakpoint sites, across which tree topologies  
511 were free to vary.

512 (2) For all other recombination events (i.e. those with no evidence of being present within the  
513 MRCA sequence being inferred), every nucleotide within all recombinationally-derived  
514 genome fragments were replaced in the sequences being analysed using “-” symbols (in this  
515 case denoting missing data).

516

517 The inference of MRCA sequences was carried out on subsets of between 252 and 346 of the original  
518 871 sequences, as the full 871 sequence alignment could not be processed in a reasonable amount  
519 of time (Supplementary File 4; alignments in Supplementary Data 1-3). These subsets were selected  
520 based on both the phylogenetic relatedness of sequences to the MRCA being inferred, and the  
521 degrees of relatedness of sequences to others in the analysed dataset. Specifically, sequences in  
522 MSV-A clades that branched within three nodes to the node of the MRCA being inferred were  
523 iteratively removed following two simple rules: (1) if two sequences differed by five or fewer sites  
524 and neither was one of the 59 phenotyped sequences, then one was randomly discarded; and (2), if  
525 a non-phenotyped sequence differed at five or fewer sites to a phenotyped sequence, then the non-  
526 phenotyped sequence was discarded. For those MSV-A sequences falling within clades that  
527 branched more than three nodes from the MRCA sequence being inferred, this same iterative  
528 sequence removal approach was used except that the rules were applied to sequences differing at

529 ten or fewer sites, instead of five or fewer sites. For the MSV sequences belonging to strains other  
530 than MSV-A, sequences were iteratively removed until there were no pairs of non-MSV-A sequences  
531 that differed from one another at 20 or fewer sites.

532

533 The MRCA sequences that were selected for inference were (i) the MRCA of all the currently  
534 sampled MSV-A sequences (designated A0), (ii) the MRCA of all the MSV-A sequences sampled from  
535 mainland Africa (designated A1), (iii) the MRCA of all currently sampled members of the MSV-A<sub>1</sub>  
536 subtype which is currently the most widely distributed MSV-A lineage found in Africa and which is  
537 currently the most prevalent subtype found in all parts of Africa other than southern Africa  
538 (designated A2), (iv) the MRCA of all currently sampled members of the MSV-A<sub>3</sub> subtype which is the  
539 second most prevalent MSV-A subtype found in East Africa to date (designated A5), (v) the MRCA of  
540 all members of the MSV-A<sub>4</sub> subtype which is the most prevalent MSV-A subtype found in southern  
541 Africa (designated A4), (vi) the MRCA of all currently sampled members of the second most  
542 prevalent MSV-A<sub>1</sub> sub-lineage in Africa and the most prevalent MSV-A<sub>1</sub> sub-lineage found on  
543 Madagascar (designated A5), and (vii) the MRCA of all currently sampled members of the most  
544 prevalent MSV-A<sub>1</sub> sublineage in Africa (designated A6).

545

546 Each MRCA sequence was inferred with MrBayes3.2 using either two or three independent runs,  
547 with sampling carried out every 500 generations and all but the last 1000 samples discarded as burn-  
548 in. Convergence was monitored using standard deviations of split frequencies, ranging between  
549 0.064 and 0.007 after between 12.2 million and 44.6 million generations. Each MRCA sequence was  
550 determined based on the average posterior probabilities of the nucleotide states at each genome  
551 site that were yielded by the two/three independent runs (Supplementary File 4). At nucleotide sites  
552 where no single nucleotide state had an associated posterior probability > 0.8 in any of the  
553 replicated runs (such as occurred at sites likely to be gap characters), the MRCA sequence inferred  
554 by maximum parsimony using PHYLIP (Felsenstein, 1993) from a maximum likelihood tree  
555 constructed using RAxML8 (Stamatakis, 2014) was used either as a tie-breaker, where it was likely  
556 that a site should have an associated nucleotide, or to infer that a site should have no associated



557 nucleotide, reflecting what should be a gap character in the MRCA sequence within the context of  
558 the analysed alignment. Ancestral genome sequences (Supplementary Data 4) were synthesized at  
559 Epoch Life Science (USA). Infectious clones of these reconstructed genomes were produced as in  
560 (Martin et al., 2001), and were used to infect the S, M and R maize as described above.

561

#### 562 *Phylogenetic inference of past virus infection symptoms*

563 Prior to fitting a dated-tip molecular clock model in a Bayesian inference approach, we evaluated  
564 the temporal signal using regressions of root-to-tip genetic distances against sequence sampling  
565 times. The analyses were based on maximum likelihood trees inferred with the program FastTree 2  
566 (Price et al., 2010) and the determination coefficients ( $R^2$ ) of the linear regression were estimated  
567 with the program TempEst (Rambaut et al., 2016). The p-values were calculated using the approach  
568 of (Murray et al., 2016) and based on 1,000 random permutations of the sequence sampling dates  
569 (Navascués et al., 2010). This root-to-tip regression analysis confirmed the presence of a significant  
570 temporal signal ( $R^2 = 0.196$ , p-value < 0.001).

571

572 Phylogenetic inference was performed with BEAST 1.10 (Suchard et al., 2018) using a combination of  
573 flexible models: a relaxed (uncorrelated lognormal) molecular clock with dated tips (Rambaut, 2000)  
574 and, for the continuous diffusion analysis (i.e. continuous character mapping) of symptom  
575 measurements, a relaxed random walk diffusion model with an underlying lognormal distribution to  
576 represent among-branch heterogeneity (Lemey et al., 2010). In both cases, posterior estimates for  
577 the standard deviations of the lognormal distributions, from which the branch-specific evolutionary  
578 rates and the rate scalars for the branch-specific precision matrices are drawn, indicated a significant  
579 deviation from a strict molecular clock and a homogeneous Brownian diffusion model. We further  
580 opted for a simple constant size coalescent model as tree prior to avoid mixing and convergence  
581 problems associated with high-dimensional model parameterisations. We analysed an alignment of  
582 60 sequences from which recombinant regions were removed. The alignment – containing the 59  
583 phenotyped MSV-A isolates and a single phenotyped MSV-B isolate used as an outgroup (alignment  
584 in Supplementary Data 4) - was constructed as above using MUSCLE and edited using IMPALE.

585 During phylogenetic inference, clades descending from inferred ancestral sequences were  
586 constrained in order to ensure that the node corresponding to the seven reconstructed ancestral  
587 sequences existed within the final trees. We ran an MCMC of 500 million generations, sampling  
588 every 0.1 million generations and removing the first 2% of samples as burn-in. A maximum clade  
589 credibility tree (MCC tree) was then inferred from the posterior tree distribution using  
590 TreeAnnotator 1.10 from the BEAST package (Suchard et al., 2018). For the comparison,  
591 phylogenetic inference of past virus infection symptoms was performed with and without the  
592 outgroup (MSV-B) sequence and associated symptom measures. Including or excluding this outgroup  
593 in the analysis had no noticeable effect on the ancestral reconstruction of symptom intensities  
594 (results not shown).

595

596 Finally, we displayed the results of the continuous diffusion analyses using two different approaches:  
597 (1) by colouring MCC trees according to inferred symptom values, and (2) by generating figures  
598 reporting averaged inferred symptom values associated with phylogenetic branches through time.  
599 Note that while this later approach allows estimating general evolutionary trends in symptom  
600 intensities through time, it is informed by the number of branches occurring at each time slice.  
601 Therefore, when going back in time, the number of branches available to estimate these global  
602 trends decreases, and they are increasingly distant for the tip measurements informing these  
603 estimates, which reduces the power to infer any deviation from apparent symptom stability. To  
604 avoid a strong contribution to inferred symptom intensities by the MSV-B outgroup isolate, these  
605 summaries were produced for the phylogenetic analysis performed without including the MSV-B  
606 outgroup.

607

608 We investigated the phylogenetic signal associated with symptom measurements by estimating  
609 Pagel's  $\lambda$  values (Pagel, 1999) for every combination of symptom and host type. Pagel's  $\lambda$  is a  
610 commonly used measure of phylogenetic signal in continuous traits such as the symptom  
611 measurements examined here. Whereas a Pagel's  $\lambda$  value of 0 reflects independence across trait  
612 observations and therefore an absence of phylogenetic signal, a value of 1 suggests that traits arise

613 according to a classic Brownian diffusion model in a manner that is perfectly consistent with the  
614 phylogenetic relationships of the organisms from which the trait measurements were taken.  
615

616 **Data availability**

617 All data and R code used for analyses in this study are available on the following public repository:

618 [https://github.com/sdellicour/msv\\_symptom\\_evolution](https://github.com/sdellicour/msv_symptom_evolution)

619

620 **Acknowledgements**

621 ALM was supported by a Swedish Institute scholarship (00448/2014) to conduct this research at The Swedish  
622 University of Agricultural Sciences (Uppsala), and is currently supported by the Research Council of Norway  
623 (267978/E40). SD is supported by the Fonds National de la Recherche Scientifique (FNRS, Belgium) and was  
624 previously funded by the Fonds Wetenschappelijk Onderzoek (FWO, Belgium). KAO is supported by the South  
625 African National Research Foundation (NRF) and The World Academy of Science (TWAS). PL1 is supported by  
626 the European Union: European Regional Development Fund (ERDF), by the Conseil Régional de La Réunion and  
627 by the Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD).  
628 GWH is supported by a South African National Research Foundation Grant number TTK1207122745. PR was  
629 supported by EU grant FP7-PEOPLE-2013-IOF (N° PIOF-GA-2013-622571). PL2 is supported by the European  
630 Research Council under the European Union's Horizon 2020 research and innovation programme (grant  
631 agreement no. 725422-ReservoirDOCS), and the Research Foundation Flanders (Fonds voor Wetenschappelijk  
632 Onderzoek Vlaanderen, G066215N, G0D5117N and G0B9317N).

633

634 **References**

- 635 Alizon S, Hurford A, Mideo N, Van Baalen M. 2009. Virulence evolution and the trade-off hypothesis: history, current state  
636 of affairs and the future. *J Evol Biol* **22**:245–259. doi:10.1111/j.1420-9101.2008.01658.x
- 637 Anderson RM, May RM. 1982. Coevolution of hosts and parasites. *Parasitology* **85**:411–426.  
638 doi:10.1017/s0031182000055360
- 639 Bosque-Pérez NA. 2000. Eight decades of maize streak virus research. *Virus Res* **71**:107–121. doi:10.1016/s0168-  
640 1702(00)00192-1
- 641 Boulton MI, Buchholz WG, Marks MS, Markham PG, Davies JW. 1989. Specificity of *Agrobacterium*-mediated delivery of  
642 maize streak virus DNA to members of the Gramineae. *Plant Mol Biol* **12**:31–40. doi:10.1007/BF00017445
- 643 Boulton MI, King DI, Donson J, Davies JW. 1991a. Point substitution in a promoter-like region and the V1 gene affect the  
644 host range and symptoms of maize streak virus. *Virology* **183**:114–121. doi:10.1016/0042-6822(91)90124-t
- 645 Boulton MI, King DI, Markham PG, Pinner MS, Davies JW. 1991b. Host range and symptoms are determined by specific  
646 domains of the maize streak virus genome. *Virology* **181**:312–318. doi:10.1016/0042-6822(91)90497-y
- 647 Bull JJ, Luring AS. 2014. Theory and empiricism in virulence evolution. *PLoS Pathog* **10**:e1004387.  
648 doi:10.1371/journal.ppat.1004387
- 649 Cressler CE, McLEOD DV, Rozins C, VAN DEN Hoogen J, Day T. 2016. The adaptive evolution of virulence: a review of  
650 theoretical predictions and empirical tests. *Parasitology* **143**:915–930. doi:10.1017/S003118201500092X
- 651 Damsteegt V. 1983. Maize streak virus: I. Host range and vulnerability of maize germ plasm. *Plant Disease* **67**:734–736.
- 652 Doumayrou J, Avellan A, Froissart R, Michalakos Y. 2013. An experimental test of the transmission-virulence trade-off  
653 hypothesis in a plant virus. *Evolution* **67**:477–486. doi:10.1111/j.1558-5646.2012.01780.x
- 654 Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf*  
655 **5**:113. doi:10.1186/1471-2105-5-113
- 656 Engelbrecht A. 1982. Chloroplast development in streak infected *Zea mays*. *S Afr J Bot* **1**:80–80.
- 657 Escriu F, Fraile A, García-Arenal F. 2003. The evolution of virulence in a plant virus. *Evolution* **57**:755–765.
- 658 Felsenstein J. 1993. PHYLIP (phylogeny inference package), version 3.5. URL:  
659 <http://evolution.genetics.washington.edu/phylip.html>.
- 660 Fereres A, Moreno A. 2009. Behavioural aspects influencing plant virus transmission by homopteran insects. *Virus Res*  
661 **141**:158–168. doi:10.1016/j.virusres.2008.10.020
- 662 Fuller C. 1901. Mealic variegation. *First Report of the Government Entomologist Natal 1899-1900* 17–19.
- 663 Harkins GW, Martin DP, Duffy S, Monjane AL, Shepherd DN, Windram OP, Owor BE, Donaldson L, van Antwerpen T, Sayed  
664 RA, Flett B, Ramusi M, Rybicki EP, Peterschmitt M, Varsani A. 2009. Dating the origins of the maize-adapted strain  
665 of maize streak virus, MSV-A. *J Gen Virol* **90**:3066–3074. doi:10.1099/vir.0.015537-0
- 666 Hodge S, Powell G. 2008. Do plant viruses facilitate their aphid vectors by inducing symptoms that alter behavior and  
667 performance? *Environ Entomol* **37**:1573–1581. doi:10.1603/0046-225x-37.6.1573
- 668 Isaacs R, Willis MA, Byrne DN. 1999. Modulation of whitefly take-off and flight orientation by wind speed and visual cues.  
669 *Physiol Entomol* **24**:311–318. doi:10.1046/j.1365-3032.1999.00144.x
- 670 Khoosal A, Martin DP. 2015. IMPALE: Improved Alignment Editor. URL: <https://github.com/super-cache-money/IMPALE>.
- 671 Kraberger S, Saumtally S, Pande D, Khoodoo MHR, Dhayan S, Dookun-Saumtally A, Shepherd DN, Hartnady P, Atkinson R,  
672 Lakay FM, Hanson B, Redhi D, Monjane AL, Windram OP, Walters M, Oluwafemi S, Michel-Lett J, Lefevre P,  
673 Martin DP, Varsani A. 2017. Molecular diversity, geographic distribution and host range of monocot-infecting  
674 mastreviruses in Africa and surrounding islands. *Virus Res* **238**:171–178. doi:10.1016/j.virusres.2017.07.001
- 675 Lemey P, Rambaut A, Welch JJ, Suchard MA. 2010. Phylogeography takes a relaxed random walk in continuous space and  
676 time. *Mol Biol Evol* **27**:1877–1885.
- 677 Lucy AP. 1996. Tissue specificity of *Zea mays* infection by maize streak virus. *Mol Plant Microbe Interact* **9**:22–31.
- 678 Martin DP. 2019. DIA (program). URL: <http://web.cbio.uct.ac.za/~darren/DIAWindowsSetup.exe>.

679 Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in  
680 virus genomes. *Virus Evol* **1**:vev003. doi:10.1093/ve/vev003

681 Martin DP, Shepherd DN. 2009. The epidemiology, economic impact and control of maize streak disease. *Food Sec* **1**:305–  
682 315. doi:10.1007/s12571-009-0023-1

683 Martin DP, Willment JA, Billharz R, Velders R, Odhiambo B, Njuguna J, James D, Rybicki EP. 2001. Sequence diversity and  
684 virulence in *Zea mays* of Maize streak virus isolates. *Virology* **288**:247–255. doi:10.1006/viro.2001.1075

685 Martin DP, Willment JA, Rybicki EP. 1999. Evaluation of Maize streak virus pathogenicity in differentially resistant *Zea mays*  
686 genotypes. *Phytopathology* **89**:695–700. doi:10.1094/PHYTO.1999.89.8.695

687 Mauck KE, De Moraes CM, Mescher MC. 2010. Deceptive chemical signals induced by a plant virus attract insect vectors to  
688 inferior hosts. *Proc Natl Acad Sci USA* **107**:3600–3605. doi:10.1073/pnas.0907191107

689 McCann J. 2001. Maize and Grace: History, Corn, and Africa’s New Landscapes, 1500-1999. *Comp Stud Soc Hist* **43**:246–272.

690 McGivern DR, Findlay KC, Montague NP, Boulton MI. 2005. An intact RBR-binding motif is not required for infectivity of  
691 Maize streak virus in cereals, but is required for invasion of mesophyll cells. *J Gen Virol* **86**:797–801.  
692 doi:10.1099/vir.0.80689-0

693 Murray GGR, Wang F, Harrison EM, Paterson GK, Mather AE, Harris SR, Holmes MA, Rambaut A, Welch JJ. 2016. The effect  
694 of genetic structure on molecular dating and tests for temporal signal. *Methods Ecol Evol* **7**:80–89.  
695 doi:10.1111/2041-210x.12466

696 Navascués M, Depaulis F, Emerson BC. 2010. Combining contemporary and ancient DNA in population genetic and  
697 phylogeographical studies. *Mol Ecol Resour* **10**:760–772. doi:10.1111/j.1755-0998.2010.02895.x

698 Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**:877–884. doi:10.1038/44766

699 Peterschmitt M, Granier M, Frutos R, Reynaud B. 1996. Infectivity and complete nucleotide sequence of the genome of a  
700 genetically distinct strain of maize streak virus from Reunion Island. *Arch Virol* **141**:1637–1650.  
701 doi:10.1007/bf01718288

702 Peterschmitt M, Quiot JB, Reynaud B, Baudin P. 1992. Detection of maize streak virus antigens over time in different parts  
703 of maize plants of a sensitive and a so-called tolerant cultivar by ELISA. *Ann Appl Biol* **121**:641–653.  
704 doi:10.1111/j.1744-7348.1992.tb03473.x

705 Pinner MS, Medina V, Plaskitt KA, Markham PG. 1993. Viral inclusions in monocotyledons infected by maize streak and  
706 related geminiviruses. *Plant Pathology* **42**:75–87. doi:10.1111/j.1365-3059.1993.tb01472.x

707 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One*  
708 **5**:e9490. doi:10.1371/journal.pone.0009490

709 Rambaut A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into  
710 maximum likelihood phylogenies. *Bioinformatics* **16**:395–399. doi:10.1093/bioinformatics/16.4.395

711 Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using  
712 TempEst (formerly Path-O-Gen). *Virus Evol* **2**. doi:10.1093/ve/vev007

713 Read AF. 1994. The evolution of virulence. *Trends Microbiol* **2**:73–76.

714 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012.  
715 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*  
716 **61**:539–542. doi:10.1093/sysbio/sys029

717 Schnippenkoetter WH, Martin DP, Hughes FL, Fyvie M, Willment JA, James D, von Wechmar MB, Rybicki EP. 2001. The  
718 relative infectivities and genomic characterisation of three distinct mastreviruses from South Africa. *Arch Virol*  
719 **146**:1075–1088. doi:10.1007/s007050170107

720 Shepherd DN, Martin DP, McGivern DR, Boulton MI, Thomson JA, Rybicki EP. 2005. A three-nucleotide mutation altering  
721 the Maize streak virus Rep pRBR-interaction motif reduces symptom severity in maize and partially reverts at  
722 high frequency without restoring pRBR-Rep binding. *J Gen Virol* **86**:803–813. doi:10.1099/vir.0.80694-0

723 Shepherd DN, Martin DP, Van Der Walt E, Dent K, Varsani A, Rybicki EP. 2010. Maize streak virus: An old and complex  
724 "emerging" pathogen. *Mol Plant Pathol* **11**:1–12. doi:10.1111/j.1364-3703.2009.00568.x  
725 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies.  
726 *Bioinformatics* **30**:1312–1313. doi:10.1093/bioinformatics/btu033  
727 Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data  
728 integration using BEAST 1.10. *Virus Evol* **4**:vey016. doi:10.1093/ve/vey016  
729 Tukey JW. 1977. Exploratory data analysis. Reading, Mass. : Addison-Wesley Pub. Co.  
730 Willment JA, Martin DP, Van der Walt E, Rybicki EP. 2002. Biological and genomic sequence characterization of maize  
731 streak virus isolates from wheat. *Phytopathology* **92**:81–86. doi:10.1094/PHYTO.2002.92.1.81  
732  
733

734 **Figure Legends**

735 **Figure 1. Correlations between MSV-A symptom measurements.** For each specific maize genotype and pair of  
736 traits, correlation estimates accounting for the phylogenetic relatedness of viral isolates are indicated by the  
737 colour gradient. A correlation estimate was considered significant (indicated by an asterisk) if the 95% HPD  
738 interval of the posterior distribution of the estimate excluded zero.

739

740 **Figure 2. Continuous character mapping of symptom intensities, chlorotic area (A) and intensity of chlorosis**  
741 **(B), onto the MSV-A maximum clade credibility tree.** Symptom intensities of ancestral viruses were  
742 phylogenetically inferred based on those observed in sensitive (S), moderately resistant (M) and resistant (R)  
743 maize genotypes infected by 59 MSV-A isolates and one MSV-B isolate (the outgroup), sampled between 1979  
744 and 2010. The intensities of symptoms are represented by coloured branches on a grey-scale. “A0-A6” indicate  
745 the ancestral nodes of the constrained clades for which MRCA sequences were inferred, synthesised and used  
746 to verify symptom intensity estimates. For each maximum clade credibility tree, we also report Pagel’s  $\lambda$  values  
747 (Pagel 1999), which indicate the degree of phylogenetic signal associated with the various symptom  
748 measurements. A version of the tree with taxon labels is given in Figure 2-figure supplement 1.

749

750 **Figure 3. Continuous character mapping of symptom intensities, leaf deformation (A) and leaf stunting (B),**  
751 **onto the MSV-A maximum clade credibility tree.** Symptom intensities of ancestral viruses were  
752 phylogenetically inferred based on those observed in sensitive (S), moderately resistant (M) and resistant (R)  
753 maize genotypes infected by 59 MSV-A isolates and one MSV-B isolate (the outgroup), sampled between 1979  
754 and 2010. The intensities of symptoms are represented by coloured branches on a grey-scale. “A0-A6” indicate  
755 the ancestral nodes of the constrained clades for which MRCA sequences were inferred, synthesised and used  
756 to verify symptom intensity estimates. For each maximum clade credibility tree, we also report Pagel’s  $\lambda$  values  
757 (Pagel 1999), which indicate the degree of phylogenetic signal associated with the various symptom  
758 measurements.

759

760 **Figure 4. Inferred changes in MSV-A symptom intensities since 1900.** The solid lines indicate the median  
761 values across the tree through time, and the shaded polygons represent the 80% highest posterior density  
762 (HPD) credible intervals: purple for the sensitive maize genotype (S), orange for the moderately resistant maize  
763 genotype (M) and green for the resistant maize genotype (R). Symptom intensities were here obtained from  
764 the phylogenetic analysis performed without including the MSV-B outgroup.

765

766 **Figure 5. Regression analysis of the inferred and observed symptom intensities of synthesised ancestral**  
767 **MSV-A variants.** Dots are coloured according to the maize genotype in which symptom intensities were  
768 inferred: purple for the sensitive maize genotype (S), orange for the moderately resistant maize genotype (M)  
769 and green for the resistant maize genotype (R). Vertical error bars reflect 95% credibility intervals of symptom  
770 intensity inferences for ancestral MSV-A variants, horizontal error bars represent 95% confidence intervals of  
771 the mean of observed symptom intensity measurements for synthesised versions of the ancestral MSV  
772 variants. The grey line has a slope = 1.0 and a y-intercept = 0: whereas points centred on this line would  
773 represent perfectly accurate inferences (i.e. instances where inferred symptom intensities = measured



774 symptom intensities), those with associated error bars that intersect this line have credibility/confidence  
775 intervals that include coordinates where the inferred and observed symptom intensities are equal. Fitted  
776 regression lines are given in orange with slope parameters, Y-intercept parameters and Pearson's  $R^2$  values  
777 given on the bottom right of each graph.  
778

779 **Figure Supplement Legends**

780 **Figure 2-figure supplement 1. Maximum clade credibility tree (from Figure 2) with taxon labels and posterior**  
781 **probabilities displayed at internal nodes.**

782

783 **Figure 2-figure supplement 2. Maximum clade credibility tree (from Figure 2) with taxon labels and 95%**  
784 **highest probability density (HPD) intervals reflecting the uncertainty of inferred node ages (green bars).**

785

786 **Figure 2-figure supplement 3. Changes in MSV-A symptom intensities over time across viral lineages**  
787 **descending from the most recent common ancestor of all known MSV-A isolates.** Included here are mean  
788 symptom measurements for individual synthesised approximations of ancestral MSV-A isolates (all of which  
789 existed prior to 1975) and means of symptom measurements of different viruses descending from these  
790 ancestral sequences (all of which were sampled between 1979 and 2007). Symptom measurements were  
791 made in sensitive (S; in purple), moderately resistant (M; in orange) and resistant (R; in green) maize  
792 genotypes. Lines link ancestral sequences (on the left) with direct descendants (on the right). Error bars on  
793 each point falling before 1980 are the 95% confidence intervals of the mean of repeated symptom  
794 measurements made for an individual ancestral virus in a single maize genotype. Error bars on each point  
795 falling after 1980 are the 95% confidence intervals of the mean of averaged symptom measurements in a  
796 single maize genotype made for multiple sampled viruses that all descended from a single ancestral virus.

797

798 **Figure 2-figure supplement 4. Maximum likelihood phylogenetic tree indicating the evolutionary**  
799 **relationships between the 59 MSV-A isolates selected in this study and all other available MSV-A isolates.**  
800 Selected MSV-A isolates are highlighted with green tip circles and the MSV-B outgroup is represented with a  
801 red tip circle.

802

803

804 **Supplementary Files**

805 **Supplementary File 1.** Table: Regression analysis of mean inferred vs observed symptom trait values.

806

807 **Supplementary File 2.** Table: Sites within ancestral sequences that could not be determined with a high  
808 degree of confidence (sites with a posterior probability of <0.8 in all attempted ancestral sequence inferences).

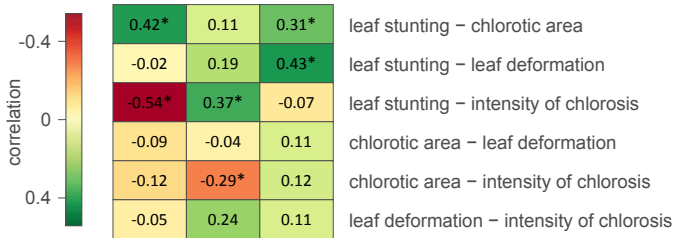
809

810 **Supplementary File 3.** Table: MSV isolates for which symptoms were quantified.

811

812 **Supplementary File 4.** Table: Dataset properties, numbers of generations used and standard deviations (SD) of  
813 split frequencies achieved during Bayesian inference of ancestral sequences.

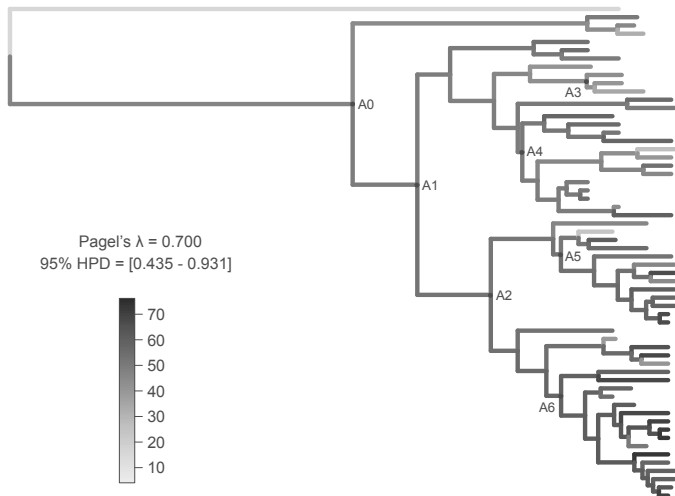
## Maize genotypes



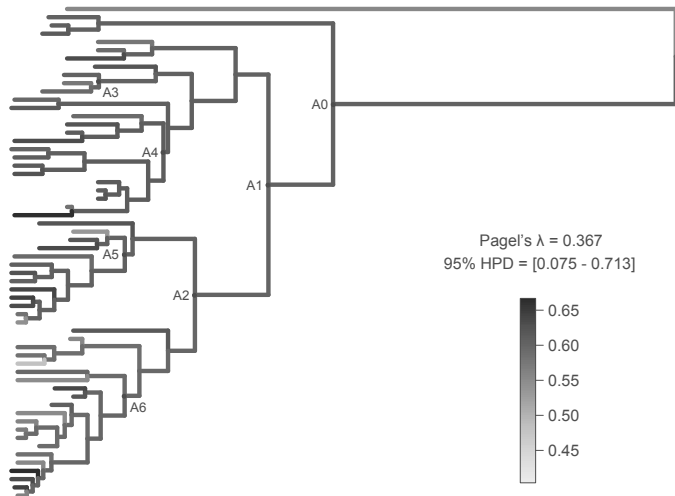
**A**

Chlorotic area

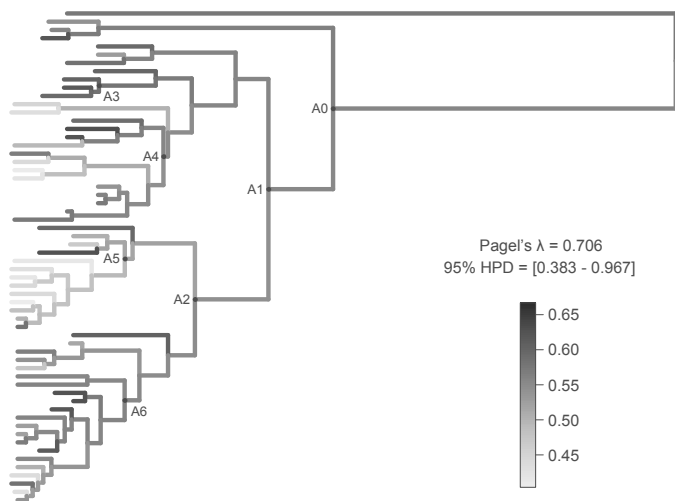
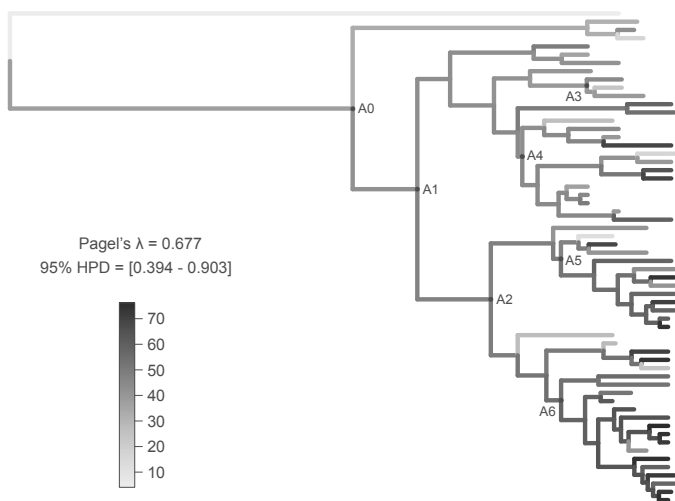
S maize genotype



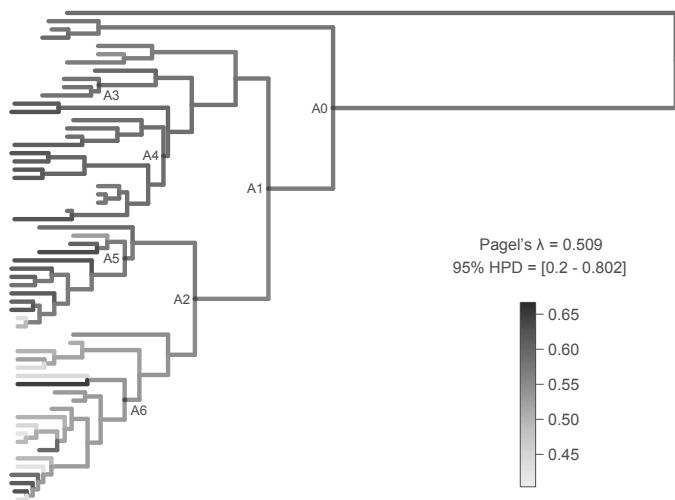
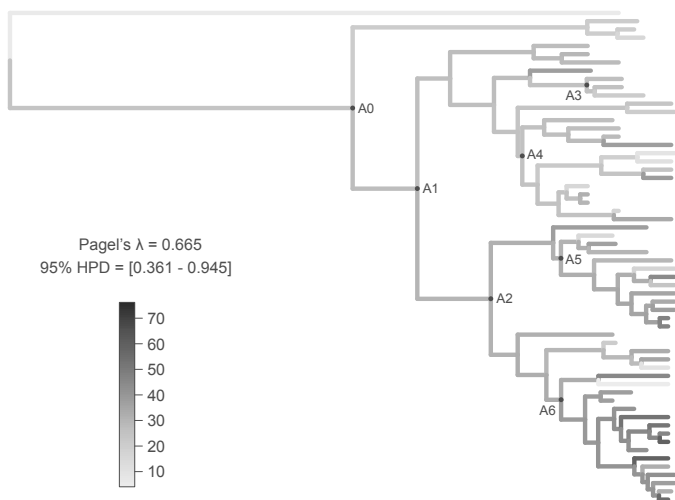
Intensity of chlorosis

**B**

M maize genotype



R maize genotype

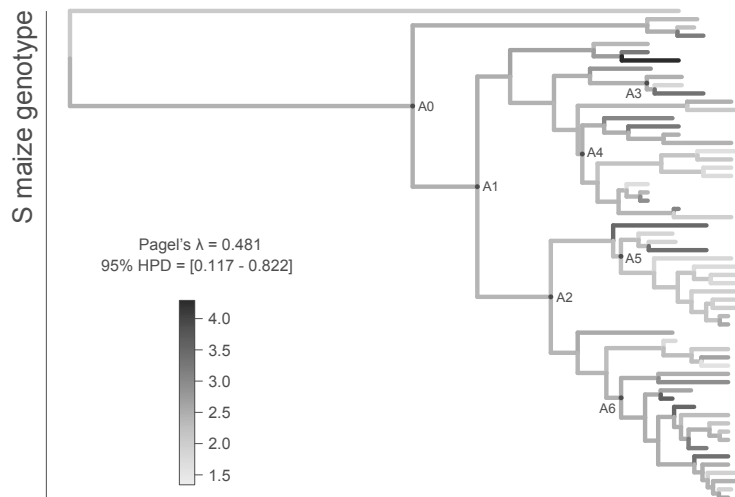


1800 1850 1900 1950 2000

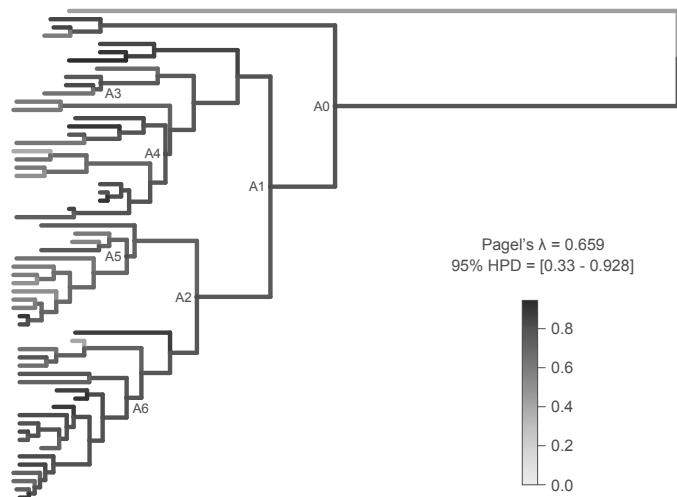
2000 1950 1900 1850 1800

**A**

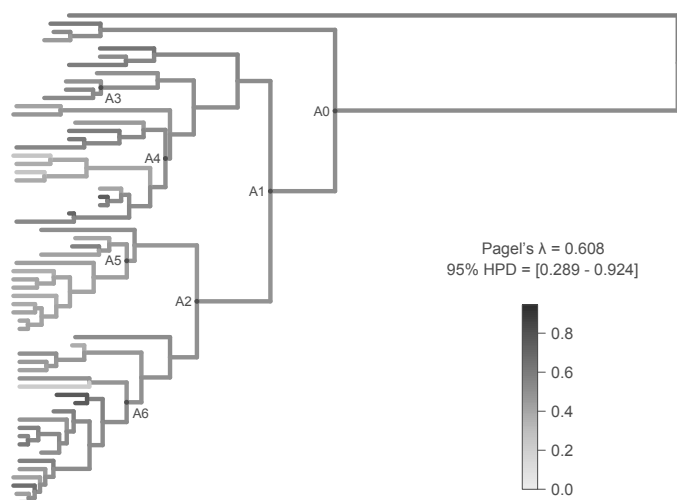
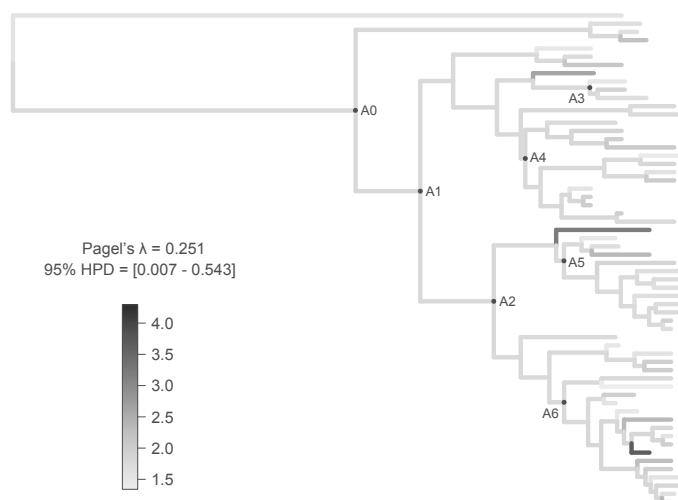
Leaf deformation



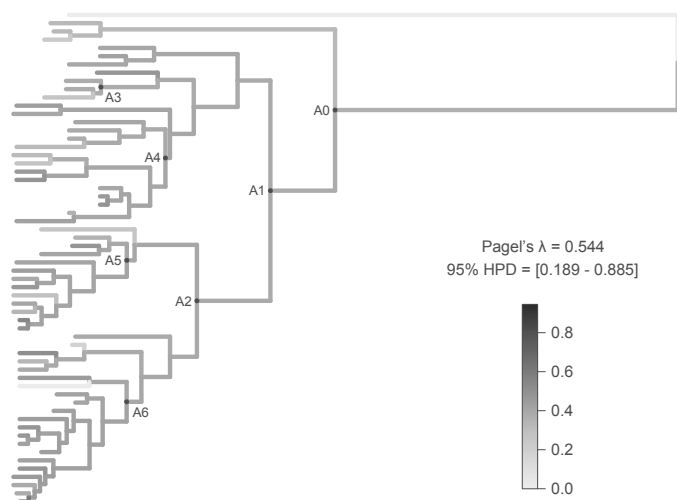
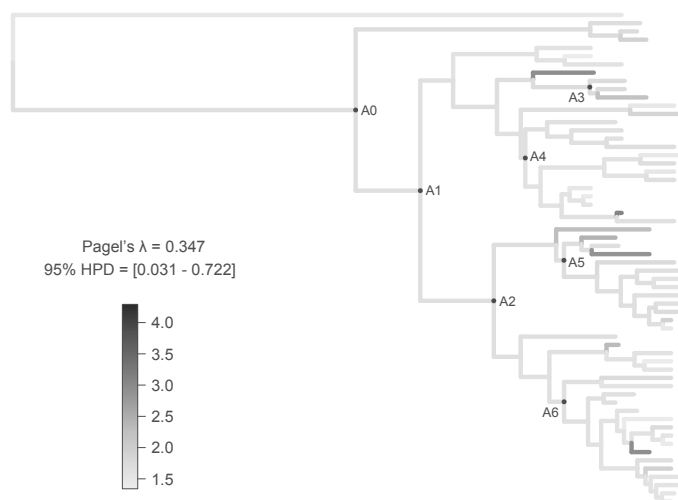
Leaf stunting

**B**

M maize genotype



R maize genotype



1800 1850 1900 1950 2000

2000 1950 1900 1850 1800

