1
2
3
4
5
6

7 **Characterization by Next Generation Sequencing Reveals the Molecular**

8 **Mechanisms Driving the Faster Evolutionary rate of *Cassava brown streak virus***

9 **Compared with *Ugandan cassava brown streak virus***

10

11 Titus Alicai[1], Joseph Ndunguru[2], Peter Sseruwagi[2], Fred Tairo[2], Geoffrey Okao-Okuja[1], Resty

12 Nanvubya[1], Lilliane Kiiza[1], Laura Kubatko[3], Monica A. Kehoe[4] and Laura M. Boykin[5,*]

13 [1]National Crops Resources Research Institute, P.O. Box 7084, Kampala, Uganda
14
15 [2]Mikocheni Agricultural Research Institute, Coca cola Road, Box 6226, Dar es Salaam, Tanzania
16
17 [3]The Ohio State University, 154W 12th Avenue, Columbus, Ohio 43210, USA

18 [4]Crop Protection Branch, Department of Agriculture and Food, Western Australia, Bentley Delivery Centre, Perth,
19 6983, Western Australia, Australia.

20 [5] The University of Western Australia, ARC Centre of Excellence in Plant Energy Biology and School of Chemistry
21 and Biochemistry, Crawley, Perth 6009, Western Australia, Australia.
22
23 * Corresponding author: laura.boykin@uwa.edu.au
24
25

26 **Keywords:** cassava, Uganda, *Manihot esculenta,* smallholder farmer, *Cassava brown streak*

27 *virus*, species tree estimation, SVD Quartets, nonsynonomous mutations

28
29
30
31

32   **Abstract**

33   Cassava is a major staple food for 800 million people. Cassava brown streak disease (CBSD), is

34   caused by *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV)

35   is suppressing cassava yields in East Africa at an alarming rate. Previous studies have

36   documented CBSV is more devastating than UCBSV. This is because CBSV is harder to breed

37   resistance for, causes more infections and yield losses in cassava, and its species delimitation is

38   more challenging. We set out to characterize the CBSV and UCBSV whole genomes from the 26

39   previously published genomes and three new from Uganda, using NGS data with the goal of

40   uncovering genetic patterns that explain the observed biological differences. In this paper, we

41   report phylogenetic relationships, rates of synonymous and non-synonymous substitutions, and

42   whole genome-based evolutionary rates for CBSV and UCBSV. Using the whole genome

43   sequences we produced the first coalescent based species tree estimation for CBSV and UCBSV

44   which supports previously published studies pointing to multiple species of both CBSV and

45   UCBSV. This new species framework led to the finding that CBSV has a faster rate of evolution

46   when compared with UCBSV. The genes responsible for CBSV's rapid rate of evolution are NIa,

47   6K2, NIb and P1. Furthermore, we have discovered that for CBSV, rates of nonsynonomous

48   substitutions are more predominant than synonymous substitution and occur across the entire

49   genome. All comparative analyses between CBSV and UCBSV presented suggests CBSV is

50   outsmarting the cassava immune system, thus is more devastating and harder to control.

51

52

53

54

55

1

Introduction

Cassava (*Manihot esculenta* Crantz) is a major staple food crop for 800 million people in over 100 tropical and sub-tropical countries [1]. In sub-Saharan Africa, it is the main source of dietary calories for approximately 300 million people [2]. The tuberous storage roots of cassava are rich in carbohydrates and can be cooked or processed for human food, animal feeds and a wide range of industrial products. The crop is relatively drought tolerant and can yield well even in less fertile soils, hence, its importance to poor families farming marginal lands [3]. Cultivation of cassava is most adversely affected by two viral diseases; cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) [4], which together were reported to cause production losses of more than US$1 billion every year [5] in Africa.

Serious yield losses due to CMD were first observed on mainland East Africa in the 1920s [6]. Recorded epidemics of CMD later occurred in the 1930s, 1940s and from 1990s to date [7,8]. By contrast, for about 70 years since it was first described [9], CBSD was confined to low altitudes (below 1000 meters above sea level) along coastal eastern Africa in Kenya, Tanzania and Mozambique. However, in the early 2000s, outbreaks of CBSD were reported over 1000 km inland at mid-altitude locations (above 1000m) in multiple countries all around Lake Victoria in Uganda [10], western Kenya [11] and northern Tanzania [4]. Where it is already established in eastern Africa, the current CBSD epidemic prevails as the main cause of losses in cassava production. Over the last 10 years, the CBSD epidemic has expanded to other countries in East and Central Africa such as Rwanda, Burundi, Congo, DR Congo and South Sudan [12-14]. This has significantly increased the risk to countries in central and west Africa which are among the world's leading cassava producers, and where CBSD does not occur.

80   CBSD is caused by *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus*

81   (UCBSV). Both viruses are (+) ssRNA viruses in the genus *Ipomovirus* and family *Potyviridae* [15-18],

82   and are often together referred to as cassava brown streak viruses (CBSVs). The CBSVs have

83   genomic organization of 10 segments, total size of approximately 8.9 to 10.8 kb, and coding for

84   a polypeptide with about 2,900 amino acid residues [15,17,18]. The complete genome of a CBSD

85   causal virus was first sequenced in 2009 [18], and to date there are only 26 publicly available [19].

86   Currently there are two species recognized by the ICTV, but Ndunguru et al. [19] have suggested

87   further speciation in the UCBSV clade. Both viruses are transmitted in a semi-persistent manner

88   by the whitefly *Bemisia tabaci* [20] and mechanically [21]. Symptoms of CBSD on cassava vary with

89   cultivar, virus or plant age, but typically include leaf veinal choloris, brown stem lesions, as well

90   as constrictions, fissures and necrosis of the tuberous storage roots [22,23].

91

92   Although CBSD has become established in eastern Africa, there is limited knowledge on the

93   diversity of causal viruses, their distribution and evolutionary potential. Therefore, it is

94   necessary to obtain several full genome sequences of CBSD viral isolates, better understand the

95   causal viruses and design long term control approaches for the disease.

96

97   In contrast to the growing knowledge on the causal agents of CBSD, host-pathogen interactions

98   are less clear. As such, little is known about specific responses of different cassava varieties to

99   prevailing species or strains of CBSD viral pathogens. Development and dissemination of CBSD-

100   tolerant varieties has been the main means adopted for CBSD control in eastern Africa. With

101   significant efforts geared at breeding for CBSD-resistant varieties, it is of great interest to know

102   if such resistance protects cassava against one or both CBSVs. Such resistance may be

103   expressed as several related features including restricted infection, systemic spread or recovery

104 of infected plants from disease and the possibility that stem cuttings taken from these may give

105 rise to progeny that are virus-free (reversion). Recent studies have shown CBSV to be the more

106 aggressive virus, infecting both tolerant and susceptible cultivars as single or mixed infections

107 with UCBSV [15,24,25]. In contrast, tolerant varieties were infected with only CBSV, but free of

108 UCBSV, suggesting their resistance to the latter. Compared with UCBSV, CBSV isolates have

109 been reported to be more detectable, having higher infection rates by graft inoculation and

110 inducing more severe symptoms [26]. It has also been shown that plants of CBSD tolerant or

111 resistant cultivars graft-inoculated with UCBSV developed milder symptoms and a significantly

112 higher proportion of the progenies were virus-free (reverted) compared to those infected with

113 CBSV [27]. To date, the underlying reasons for this more aggressive nature of CBSV compared

114 with UCBSV are not known.

115

116 In this study, CBSV and UCBSV molecular diversity was investigated by using next generation

117 sequencing to understand new complete genomes of three isolates from Uganda. The

118 sequences obtained were analyzed to determine species composition, CBSV and UCBSV

119 evolutionary rates, role of such changes in virus-host interactions, resulting into cassava cultivar

120 susceptibility or resistance. We set out to answer the following questions:

121   1) How do the three new complete genomes from Uganda compare to those already

122   published [19]?

123   2) Are CBSV and UCBSV distinct species and is there further speciation?

124   3) Why is CBSV more aggressive and harder to breed resistance for than UCBSV?

125
126 **Results**

127 **CBSD Field Symptoms Associated with CBSV and UCBSV Isolates**

4

128   Categorisation of CBSD foliar symptom distribution on symptomatic plants assessed revealed

129   that the most frequently encountered type was LL - symptoms only on lower leaves (68.4%),

130   followed by SW - systemic and on the whole plant (26.3%), and SL – systemic but localized

131   (5.3%) (table 1). Based on CBSVs detected and CBSD leaf symptom severity scores for 57

132   sampled plants, whereas the majority of plants infected by UCBSV alone as determined by RT-

133   PCR had mild chlorosis (severity score 2), CBSV infections (single or mixture with UCBSV) tended

134   to have moderate to severe symptoms (scores 3-4) in same proportion to those exhibiting score

135   2 (fig. 1, table 1). Regarding the three isolates used here for whole genome sequencing, U8

136   (UCBSV) was from a plant with CBSD score 3 and LL symptom type. Both CBSV isolates (U1 and

137   U4) were from plants with severity scores 2 and 3, symptom types LL and SL, respectively.

138

139   **Next Generation Sequencing**

140   The three samples from Uganda produced raw reads ranging from 21,844,716 to 23,648,990.

141   After trimming for quality using CLCGW, these numbers were reduced to 21,582,374 to

142   23,373,606 (table 2). Following *de novo* assembly of the trimmed reads using CLCGW, the

143   numbers of contigs produced were 621-1,008. The contigs of interest from *de novo* assembly

144   were of lengths 2,214 to 8,954nt, with average coverage 24 to 366. After mapping to a

145   reference genome in Geneious, the lengths of the consensus sequences were 8,893 to 9,563

146   with average coverages of 25 to 393. The final sequences consisted of a consensus between

147   the *de novo* and the mapped consensus with lengths of 8,700 to 8,748.

148

149   **Genomic Variability and Positive Selection**

150   The CBSV genomes included in this study were more variable when compared with those of

151   UCBSV (supplementary figs. S1 and S2). Characterizing amino acid usage at each position in the

5

152   whole genome revealed that CBSV genomes have non-synonymous substitutions present across

153   their entire genome (fig. 2), and predominating when compared to synonymous substitutions.

154   In contrast, UCBSV had near equal non-synonymous and synonymous substitutions across the

155   entire genome. Genes in the UCBSV genomes with non-synonymous substitutions at a higher

156   frequency were; P1, NIb and HAM1 (fig. 2).

157

158   CBSV had 68 positively selected sites and 66 negatively selected sites, UCBSV had zero

159   positively selected sites (codons) and 558 negatively selected sites (table 3). Analyzed together

160   there are 3 positively selected sites and 1383 negatively selected sites. The coat protein (CP) of

161   CBSV had the highest number of positively selected sites (16) while 6K2 had zero.

162

163   **Rates of Evolution**

164   CBSV and UCBSV have different rates of evolution (table 4). We tested two hypothesis using

165   CODEML. The null hypothesis tested was CBSV and UCBSV have equal rates of evolution while

166   the null hypothesis was that CBSV and UCBSV have different rates of evolution (two omegas;

167   model = 2). The Likelihood Ratio Test was used to test for significance if the difference in

168   likelihood was greater than 3.84 (based on the Chi-squared distribution and one degree of

169   freedom) we rejected the null hypothesis that the rates between CBSV and UCBSV are equal.

170

171   CBSV whole genome sequences showed it is evolving 5 times faster than UCBSV. The genes

172   contributing to this accelerated rate of evolution for CBSV are NIa (D=29.95), followed by 6K2

173   (D=6.74), NIb (D=5.18) and P1 (4.61) (table 4 and fig. 4). The transition/transversion ratios were

174   also estimated using CODEML and show the 6K1 (19.6) and CP (13.2) genes have the highest

175   estimates while the remaining 8 genes ranged from 5.05 – 9.93.

6

176

Species Tree Estimation - SVDQ

178   The species phylogeny (fig. 3) shows strong support for a split into two primary viral clades, one

179   consisting of CBSV (fig. 3 clades A and B) and the other consisting of UCBSV (fig. 3 clades E-G),

180   with 100% bootstrap support separating the two clades. Figure 3 shows clades labeled A–G

181   which correspond to; 1) labels A-F from Ndunguru et al [19], and 2) a new clade G defined in this

182   study. Within the CBSV clade, there are several additional clades with 100% bootstrap support,

183   including the two new CBSV whole genomes from Uganda (U1 and U4). These are the first CBSV

184   whole genomes sequences from Uganda. The other CBSV grouping with 100% bootstrap

185   support labeled B in Figure 3 contains 4 Tanzania samples KoR6, Tan 79, Tan 19 1 and Nal 07. In

186   the UCBSV clade there are 6 nodes supported with a 100% bootstrap, including the new UCBSV

187   whole genome added from this study (U8) which is sister to Kab 07 from Uganda. In addition,

188   the CBSV clade had all samples from a given country grouping together while the UCBSV clade

189   had monophyletic clades from different countries (the multi-colored lines in fig. 3).

190

Comparison of Gene Trees to Species Tree

192   Clades A and B, which partition the CBSV isolates into two groups, are consistently present with

193   high support in all genes except HAM1 and CP (table 5). Clades D and G, which each consist of a

194   pair of UCBSV isolates, have high support across all genes, while clades C and E have relatively

195   high support across a majority of genes. Clade F is strongly supported by the CI gene, which is

196   relatively long, but is not found in the phylogenetic tree estimated for any of the other genes.

197

198   The whole genome concatenated analysis using MrBayes shows strong support (posterior

199   probability 1.0) for all clades (table 5). However, this analysis does not take into account the

7

200   possibility of variation in the evolutionary processes across the individual genes. The SVDQ

201   analysis, on the other hand, uses a coalescent-based method to estimate the overall species

202   tree, and properly accounts for variation in the evolutionary history for each gene. In viewing

203   the bootstrap support values for each of the clades from the SVDQ analysis, we see that the

204   level of support for each clade across the genome is more accurately represented by the

205   corresponding bootstrap proportion. For example, clade F, which was found only in the

206   phylogeny of the CI gene, shows a bootstrap proportion of 0.44 for the SVDQ analysis (as

207   compared to 1.0 for the MrBayes concatenated analysis) (table 5). Similarly, the SVDQ analysis

208   gives a bootstrap proportion of 0.87 for clade E, which showed posterior probabilities below 0.8

209   for 3 of the 10 genes, as compared to a posterior probability of 1.0 for the concatenated

210   analysis with MrBayes. All other clades are supported with bootstrap values of 1.0, consistent

211   with the MrBayes analysis.

212

213   **Sliding Window SVD Score**

214   The SVD Score Sliding Window analysis (fig. 4) shows several interesting patterns. First, note

215   that the gene boundaries track well with shifts in the magnitude of the SVD Score, indicating

216   that individual genes are subject to specific evolutionary processes that vary from gene to gene.

217   In particular, several genes show strong support for the primary CBSV/UCBSV split, as indicated

218   by their low scores, while other genes show variation from this basic process, as indicated by

219   increases in the scores. In addition, fig. 4 shows the test statistic associated with the hypothesis

220   test of a shift in the rate of evolution between the two groups, with `*' indicating that the rate

221   difference between the two groups is statistically significant. It is readily apparent from the

222   graph that genes that show strong support (low SVD Score) for the primary CBSV/UCBSV split

223   also show strong evidence for statistically significant differences in evolutionary rate. These

8

224   results support the overall hypothesis that certain genes in CBSV have accelerated rates of

225   evolution that contribute to the increased aggressiveness of the virus.

226
227

228

229   **Discussion**

230   In this study we analyzed the molecular mechanisms underlying the field and laboratory

231   observations that CBSV more readily infects cassava plants and tends to display severe

232   symptoms when compared with those infected with UCBSV. Our analyses included

233   characterizing three new complete CBSV (2) and UCBSV (1) genomes, which were combined

234   with the 26 previously published. Our major findings show further speciation of CBSV and

235   UCBSV, a larger genetic landscape for CBSV, including many nonsynonomous sites, and that

236   CBSV has a faster rate of evolution compared with UCBSV (table 4 and fig. 4).

237

238   **Genes with Accelerated Rates of Evolution in CBSV**

239   We have identified P1, 6K2, NIb and NIa as the genes with accelerated rates of evolution in

240   CBSV. The function of P1 is as an RNA silencing suppressor (RSS), and there is also the

241   suggestion that it may be involved in virion binding to the whitefly stylet via a "bridge"

242   formation by a virus-encoded P1 protein for both CBSV and UCBSV. 6K2 is associated with

243   cellular membrane and is responsible for systemic infection and viral long distance movement

244   [28]. The NIb encodes for a nuclear inclusion polymerase and the NIA for a nuclear inclusion

245   protease [18,29].

246

247   In Potyviruses generally, when NIa and VPg are associated together they are located in the

248   cytoplasm and nucleus of infected cells. When 6K2-VPg-NIa forms a larger product, the VPg

249   plays a role in viral RNA replication [30]. Even though VPg is not one of the genes with a higher

250   evolution rate, both 6K2 and NIa are a part of the complex which affects replication, and this

251   may go some way to explaining their apparent accelerated evolution rate. Is it possible that the

252 accelerated rates of evolution for genes involved in replication could even be a response to the

253 relatively recent interaction of the viruses and cassava? These viruses are not present in South

254 America where cassava originates so the viruses must be native to Africa. It would appear that

255 the adaptation is still occurring and the cassava immune system does not know how to fight

256 these infections yet. Cassava was introduced to East Africa in the 18th century through oceanic

257 movement. The first reports of brown streak disease in Tanzania occurred 1936 [9,13]. There has

258 been little opportunity for the co-evolution of the viruses and the host, therefore a natural

259 resistance would be hard prospect. This raises the possibility of the original host of these

260 viruses, a non-cassava host which may be harboring these viruses or the most recent common

261 ancestor of these viruses. This in turn leads us to wonder just how old these viruses and their

262 ancestors are, and the best way to answer that is to sequence more virus genomes from both

263 cassava and non-cassava hosts wherever they are found.

264

265 **How Can CBSV Still Function with Such a Large Genetic Landscape?**

266 CBSV and UCBSV have different evolutionary patterns as observed by characterizing the whole

267 genome sequences of CBSV and UCBSV separately. CBSV is genetically more diverse when

268 compared with UCBSV, as evident by the greater amino acid usage (supplementary fig. 1), the

269 faster rates of evolution across the entire genome (table 4), and greater number of

270 nonsynonymous sites across the entire genome (Figure 2). How can CBSV still function with

271 such a large genetic landscape? RNA viruses walk a very fine line of having the genetic arsenal

272 to overcome the host immune system and diverging to a point that key functions of genes are

273 lost [31]. Recent studies [32,33] have shown that viruses with a large genetic landscape adapt to host

274 changes much quicker and can overcome the host immune system faster. Viruses that occupy a

275 large portion of the possible sequence space might be less fit but they outcompete the fitter

276   strain when the host immune system shifts and hence these viruses have been described as

277   adapted to "survival of the flattest" [34,35]. This means that a virus that covers the most sequence

278   space will be able to adapt to host immune system faster than those with smaller spaces.

279   Viruses that are adapted in this category ("survival of the flattest") are going to be harder to

280   breed resistance for because the virus has a larger ability to adapt to changes. It is clear that in

281   our case, CBSV is the virus that has a larger sequence space (Supplemental Figure 1) when

282   compared to that of UCBSV, which is clearly smaller (Supplemental Figure 2). CBSV is one of the

283   RNA viruses that can be described as adapted to "survival of the flattest", while UCBSV is not.

284   Therefore, CBSV is more devastating because it has a larger genetic arsenal which it uses

285   overcome the changes breeders are introducing into cassava.

286

287   Not only are the CBSV genomes more genetically diverse, but are also characterized by a large

288   number of nonsynoymous changes in the genome (Figure 2). An excess of nonsynonymous over

289   synonymous substitutions at individual amino acid sites is signifies that positive selection has

290   affected the evolution of a protein between the extant sequences under study and their most

291   recent common ancestor [36]. Positive selection is the process by which new advantageous

292   genetic variants sweep a population and is the mechanism Darwin described to drive evolution.

293   This is further evidence that CBSV has a greater capacity to evade the cassava immune system

294   as compared with UCBSV. CBSV had 66 sites under positive selection (Table 4) while UCBSV had

295   none. The CBSV sites under positive selection are found not only in the regions that have gained

296   the most attention, CP and HAM1-like [13], but are also found in all other genes except 6K2. This

297   is further support for CBSV's ability to outsmart the cassava immune system. Every gene in the

298   CBSV genome (except 6K2) has sites under positive selection indicating effective RNA silencing

299   of the virus will need to encompass many loci.

12

300

301   Using computational methods combined with field observations we have concluded that CBSV

302   is more devastating than UCBSV. This assertion is also supported by two recent biological

303   studies. The first was a test of reversion in three different cassava varieties (Albert, Kaleso and

304   Kiroba) infected with CBSV and UCBSV. Reversion is a type of resistance mechanism where by

305   virus-infected plants will naturally recover from infection over time, and their progeny from

306   stem cuttings are virus-free. A reversion event infers the host immune system was able to clear

307   or restrict the virus from systemic movement. It was shown that UCBSV infected cassava had a

308   higher rate of reversion when compared to plants infected with CBSV [27] indicating the plants

309   infected with UCBSV recovered more often than those infected with CBSV. This is another line

310   of evidence that CBSV is more devastating and the cassava immune systems of the three

311   varieties tested are struggling to resist the virus.

312

313   The second study that supports the hypothesis that CBSV is more aggressive than UCBSV

314   analyzed virus-derived small RNAs within three cassava varieties (NASE 3, TME204 and 60444).

315   Plants infected with viruses are known to trigger RNAi antiviral defense that can be measured

316   by quantifying the abundance of 21-24 nucleotide (nt) segments produced by the dicer enzyme

317   [37]. Cassava varieties were infected with either CBSV or UCBSV, NGS was used to detect virus-

318   derived small RNAs [24], and the 21-24 nt dicer fragments were mapped to either CBSV or UCBSV

319   depending on which virus was used to infect the plant. The results showed that CBSV infection

320   triggered a stronger immune response as measured by greater abundance of virus derived

321   small RNA fragments across the entire CBSV genome compared with UCBSV. In addition, across

322   all three genotypes they observed that cassava grafted with CBSV-infected buds showed more

323   severe symptoms compared to UCBSV-infected plants [24]. This is further evidence that CBSV is a

13

324   more aggressive virus and breeding for resistance to CBSV and UCBSV will require different

325   experimental approaches.

326

327   **Implications of the Species Tree for CBSV and UCBSV**

328   We have produced the first species tree estimation of the CBSD causal virus species using whole

329   genome sequences and the coalescent based SVD Quartets species tree estimation algorithm.

330   Differences in the evolutionary history of the two viruses are seen in the branching patterns in

331   Figure 3. CBSV has diverged into two main clades A and B, while UCBSV has several well

332   supported clades but the backbone is still unresolved, indicating more sampling is needed to

333   fully understand the diversity and evolutionary history of UCBSV. The species tree (Figure 3) is

334   similar to the concatenated whole gene tree reported in Ndunguru et al. [19], except addition of

335   the clade labeled "G", and lack of support for clades E and F in the UCBSV species. It is well

336   documented that concatenating genes without using the coalescent based models can produce

337   misleading results [38,39]. In our case, only CI supports clade F, and it is also the longest gene

338   (1,883 bp), therefore swamps signals of other genes. The whole genome concatenation

339   recovers clade F with a posterior probability of 1.00 (Table 4). With regards to clade E, the

340   SVDQ tree was more reflective of the individual gene tree signal by producing a bootstrap value

341   of 0.87 versus 1.00 for the whole genome concatenated tree (Table 4). These results suggest

342   that the topology in the UCBSV species will change as more samples are added.

343

344   Our integrative approach of species tree estimation coupled with analyzing rates of evolution

345   has lead to a new framework for CBSV and UCBSV, which includes analyzing and treating these

346   two groups of viruses as separate species. Multiple putative species of both CBSV and UCBV

347   have been identified which means cassava needs to be resistant to the virus species that are

14

348   prevalent if farmers' fields. We argue that this genomic diversity and faster rate is what is

349   causing the breeders to struggle with breeding resistant varieties and also why the diagnostic

350   primers are not working consistently. CBSV also has more positively selected sites than UCBSV.

351   It was first thought that CBSD was restricted to the coastal areas and below 1000 m [23] but as

352   more genetic data is gathered CBSV and UCBSV are found at all elevations in many ecozones

353   throughout East Africa [4,10,13,15,19,40]. We are still in the discovery phase with CBSV and UCBSV

354   species as there are only 29 (now with the three new included here) whole genome sequences

355   and other new species of both viruses are likely to be discovered. As we move forward it is

356   important to include all known samples and use appropriate species tree estimation methods

357   such as SVDQ.

358

359   Finally, the traditional gene regions (CP and HAM1-like) used to delimit species and are the

360   targets for diagnostic primers do not recover the species tree (Table 4). We recommend

361   designing new diagnostic regions for other genes that recover the species tree and also do not

362   have an accelerated rate of molecular evolution (Figure 4), such as CI or P3 for species level

363   diagnoses. It is possible that the spread of CBSV and UCBSV could have been exacerbated

364   through dissemination of infected cuttings, as virus indexing with primers targeting CP may

365   have misleadingly returned negative results.

366

367   **Implications of the Results for Cassava Breeding**

368   During the last three decades worldwide, agricultural production has been compromised by a

369   series of epidemics caused by new variants of classic viruses that show new pathogenic and

370   epidemiological properties. An important determinant of the fitness of a virus in a given host is

371   its ability to overcome the defenses of the host. Overcoming plant resistance by changes in the

15

372   pathogenicity of viral populations represents a specific and important case of emergence, with

373   tremendous economic consequences since it jeopardizes the success and durability of

374   resistance factors in crops as an anti-viral control strategy. In this study, we found CBSV to be

375   more variable, have more positively selected sites and evolving five times faster than UCBSV.

376   These findings have huge implications for cassava improvement efforts in Africa where CBSV is

377   widely present. Field and laboratory results have proven CBSV to be more virulent and more

378   devastating than UCBSV. Knowledge of specific virus species an improved cassava variety is

379   resistant to will determine where to screen, multiply and deploy such varieties. Cassava

380   breeders have to take into consideration the evolutionary and biological differences between

381   CBSV and UCBSV in the breeding programs. For example, cassava breeders can breed varieties

382   that are resistant to CBSV that can be strategically deployed in areas where CBSV is more

383   prevalent, and similarly for UCBSV. Furthermore, it becomes more appropriate to always screen

384   cassava materials against CBSV as a minimum, even if UCBSV is the more prevalent virus. Such

385   strategy will in effect ensure durable resistance as opposed to the indiscriminate screening and

386   distribution of the improved CBSD resistant cassava varieties, without knowledge of the virus

387   species in the area.

388
389   **Methods**

390   **Field Plant Sample Collection**

391   Farmers' fields in Uganda with cassava plants 3-6 months old were surveyed for CBSD in 20

392   districts. In each field, cassava plants were visually assessed to confirm typical CBSD symptoms

393   on leaves and stems. CBSD leaf symptom severity was scored on a 1-5 scale [41,42]; 1 = no visible

394   symptoms, 2 = mild vein yellowing or chlorotic blotches on some leaves, 3 =

395   pronounced/extensive vein yellowing or chlorotic blotches on leaves, but no lesions or streaks

16

396   on stems, 4 = pronounced/extensive vein yellowing or chlorotic blotches on leaves and mild

397   lesions or streaks on stems, 5 = pronounced/extensive vein yellowing or chlorotic blotches on

398   leaves and severe lesions or streaks on stems, defoliation and dieback. CBSD symptoms were

399   also categorized based on distribution of leaf chlorosis and stem lesions on the plant; systemic

400   and on the whole plant (SW), systemic on leaf or stem parts but localized (SL), only on lower

401   leaves (LL). On selected symptomatic plants, portions of the third fully expanded leaf on a shoot

402   were picked as samples, air-dried by pressing between sheets of newsprint and stored pending

403   RNA extraction.

404

405   **RNA Extraction**

406   About 0.25 g cassava leaf samples were frozen in liquid nitrogen, then ground using a mortar

407   and pestle. 2 ml CTAB lysis buffer (2% CTAB; 100 mM Tris–HCl, pH 8.0; 20 mM EDTA; 1.4 M 134

408   NaCl; 1% sodium sulphite; 2% PVP) was added and samples homogenized. The 1 ml of the

409   homogenate was incubated at $65^{°}C$ for 15 min, an equal volume of chloroform: isoamyl alcohol

410   (24:1) was added, and the sample was centrifuged for 10 min at approximately 14,500rpm.

411   800μl of the aqueous layer was transferred to a new tube with an equal volume of 4 M LiCl and

412   incubated at $-20^{°}C$ for 2 hrs. The samples were centrifuged for 25 min at 14,500 rpm and the

413   supernatant was poured off. The pelleted RNA was re-suspended in 200 μl TE buffer containing

414   1% SDS, 100 μl of 5M NaCl. 300 μl of ice-cold isopropanol were added and incubated at $-20^{o}C$

415   for 30 min. The sample was centrifuged at 13,000 rpm for 10 min and the aqueous layer was

416   decanted and RNA pellets washed in 500 μl of 70% ethanol by centrifuging at 13,000 rpm for 5

417   min. The ethanol was decanted off and RNA pellet dried to remove residual ethanol. The RNA

418   was re-suspended in 50 μl nuclease-free water and stored at −80 °C prior to testing.

419

CBSV and UCBSV Detection by RT-PCR

All samples were tested for presence of CBSV and UCBSV by a two-step RT-PCR assay [43]. The PCR mixture consisted of 16.0 µl nuclease free water, 2.5 µl PCR buffer, 2.5 µl MgCl$_2$ (2.5 mM), 0.5µl dNTPs (10 mM), 1.0 µl of each primer (10mM) [forward CBSDDF2 5'-GCTMGAAATGCYGGRTAYACAA-3' and reverse CBSDDR 5'-GGATATGGAGAAAGRKCTCC-3'], 0.5 µl Taq DNA polymerase and 1.0 µl of cDNA. The PCR thermo profile consisted of: 94°C for 2 min followed by 35cycles of 94°C (30 s), 51°C (30 s) and 72°C (30 s) for denaturation, annealing and extension, respectively. PCR products were analysed by electrophoresis in a x1 TAE buffer on a 1.2% agarose gel, stained with ethidium bromide, visualized under UV light and photographed using a digital camera.

Sample Selection for Sequencing

From the data obtained in the diagnostic tests, samples for sequencing were selected to represent different geographical regions, symptom types and severities. Three samples that tested positive for either CBSV (2) or UCBSV (1) were selected for this study. The two samples for which presence of CBSV was confirmed (U1 and U4) had been collected from different farmer's fields in Mukono district, central Uganda. The sample with UCBSV (U8) selected for further analysis originated was collected from a field in Mayuge district, eastern Uganda.

Generation of the Transcriptomes

The three samples were transported to the laboratory and extracted as detailed above. Total RNA was blotted on to FTA cards and later extracted using methods previously described [44]. Total RNA from each sample was sent to the Australian Genome Research Facility (AGRF) for

18

443   library preparation and barcoding before 100 bp paired-end sequencing on an Illumina

444   HiSeq2000.

445

446   **De novo Sequence Assembly and Mapping**

447   For each sample, reads were first trimmed using CLC Genomics Workbench 6.5 (CLCGW) with

448   the quality scores limit set to 0.01, maximum number of ambiguities to two and removing any

449   reads with <30 nucleotides (nt). Contigs were assembled using the *de novo* assembly function

450   of CLCGW with automatic word size, automatic bubble size, minimum contig length 500,

451   mismatch cost two, insertion cost three, deletion cost three, length fraction 0.5 and similarity

452   fraction 0.9. Contigs were sorted by length and the longest subjected to a BLAST search (blastn

453   and blastx) [45]. In addition, reads were also imported into Geneious 6.1.6 [46] and provided with

454   reference sequences obtained from Genbank (KR108828 for CBSV and KR108836 for UCBSV).

455   Mapping was performed with minimum overlap 10%, minimum overlap identity 80%, allow

456   gaps 10% and fine tuning set to iterate up to 10 times. A consensus between the contig of

457   interest from CLCGW and the consensus from mapping in Geneious was created in Geneious by

458   alignment with MAFFT [47]. Open reading frames (ORFs) were predicted and annotations made

459   using Geneious. Finalized sequences were designated as "complete" based on comparison with

460   the reference sequences used in the mapping process, and "coding complete" if some of the 5'

461   or 3' UTR was missing but the coding region was intact [48,49], and entered into the European

462   Nucleotide Archive (WEBIN ID number Hx2000053576).

463

464   **Genome Alignment and Annotation**

465   Twenty-six whole genomes (12 CBSV and 14 UCBSV) were downloaded from GenBank and

466   imported into Geneious [46], and the MAFFT plugin [47]was used to align them with the 3 new

19

467 whole genome sequences obtained in this study. Nucleotide alignments were translated into

468 protein using the translate align option in Geneious and then visually inspected for quality.

469 Annotations were transferred to the 3 new genomes from the 26 previously published genomes

470 using the live annotation option in Geneious.

471

472 **Characterizing the Genetic Diversity in CBSV and UCBSV Genomes**

473 CBSV and UCBSV are distinct species (Figure 2) therefore the genomes were treated separately

474 in the analyses in characterizing the genomes. Characterizing the genetic diversity of CBSV and

475 UCBSV was done using the Synonymous Non-synonymous Analysis Program (SNAP v2.1.1)

476 implemented in the Los Alamos National Laboratory HIV-sequence database

477 (http://www.hiv.lanl.gov) [50]. SNAP calculates synonymous and non-synonymous substitution

478 rates based on a set of codon-aligned nucleotide sequences. This program is based on the

479 simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide

480 substitutions of [51], and incorporating a statistic developed for computing variances and

481 covariances of dS's and dN's [52]. An application of the SNAP package in HIV-1 research has also

482 been developed [53].

483

484 **Estimating Rates of Evolution**

485 To further characterize the CBSV and UCBSV genomes, we estimated the rates of molecular

486 evolution using CODEML implemented in PAML (Phylogenetic Analysis by Maximum Likelihood)

487 [54]. PAML is a package of programs for analysis of DNA or protein sequences by using maximum

488 likelihood methods in a phylogenetic framework. The null hypothesis tested was CBSV and

489 UCBSV have equal rates of evolution (one omega; model = 0) while the alternative hypothesis

490 was that CBSV and UCBSV have different rates of evolution (two omegas; model = 2). The

491   Likelihood Ratio Test was used to test for significance if the difference in likelihood was greater

492   than 3.84 (based on the Chi-squared distribution and one degree of freedom) we then rejected

493   the null hypothesis that the rates between CBSV and UCBSV are equal. Initial analyses were

494   carried out for the entire genome and showed CBSV has a higher rate of evolution (Table 4;

495   Figure 4). To identify which gene or genes were contributing to the faster rate of evolution we

496   analyzed the individual genes separately testing the hypotheses and parameters utilized for the

497   complete genome.

498

499   **Testing for Positive Selection**

500   Sites under positive selection were identified using SLAC [55] implemented on the

501   http://www.datamonkey.org web server [56]. The settings used to run SLAC were as follows: the

502   best fitting model (GTR) was specified global dN/dS value was estimated and the significance

503   level was set to 0.01.

504

505   **Gene Tree Estimation**

506   Individual gene trees were estimated using MrBayes 3.2.1 [57] run in parallel on Magnus (Pawsey

507   Supercomputing Centre, Perth, Western Australia) utilizing the BEAGLE library [58]. MrBayes 3.2.1

508   was run utilizing 4 chains for 30 million generations and trees were sampled every 1000

509   generations. All runs reached a plateau in likelihood score, which was indicated by the

510   standard deviation of split frequencies (0.0015), and the potential scale reduction factor (PSRF)

511   was close to one, indicating the MCMC chains converged.

512

513   **Species Tree Estimation**

514   The SVDQ method [59] implemented in PAUP* [60] was used to analyze the whole-genome data.

515   This method allows analysis of multi-locus data in a coalescent framework that allows for

516   variation in the phylogenetic histories of individual genes. The method was run with all possible

517   quartets (23,751) sampled in each of 100 bootstrap replicates, and the consensus across all

518   bootstrap replicates was used as the estimate of the species tree.   Bootstrap support values for

519   each node were used to quantify uncertainty in the species tree estimate. The entire analysis

520   took approximately 2.5 minutes on a MacBook Pro running OSX 10.11.2 with a 2.2 GHz Intel

521   Core i7 processor.

522

523   **Comparison of Gene Trees to Species Tree**

524   We compared the single-gene phylogenies constructed using MrBayes with the overall species

525   tree phylogeny estimated using SVDQ and the concatenated phylogeny estimated by MrBayes.

526   For each tree, we evaluated presence or absence of the clades identified by Ndunguru et al.[19]

527   labeled A-F in Figure 3. We identified an additional clade (clade G, Figure 3) that we noticed to

528   be consistently present across genes and methods. For each of these clades present in a

529   particular tree, we recorded the posterior probability (for trees constructed by MrBayes) or the

530   bootstrap proportion (for the tree estimated by SVDQ) in Table 5.

531

532   **Sliding Window SVD Score**

533   The SVD Score [61] was used to quantify support for two viral clades for portions of the genome in

534   a sliding window analysis. Briefly, the SVD Score measures the extent to which the data support

535   a phylogenetic "split" – a division of the taxa into two groups with specified group membership.

536   Low values of the SVD Score indicate strong support for the split of interest, while larger values

537   indicate *either* a lack of support for the split or a shift in the underlying evolutionary process

22

538   (see Allman et al. (2016) for details and examples). We computed the SVD Score with the split

539   defined by CBSV vs. UCBSV across the genome in windows of 500 bp, sliding in increments of

540   100 bp, and plotted the resulting SVD Scores across the genome, with boundaries between

541   genes marked with vertical lines. The computations took less than one minute on a MacBook

542   Pro running OSX 10.11.2 with a 2.2 GHz Intel Core i7 processor.

543

544

545 **Acknowledgments**

551

552 **Conflict of Interest**

553 The authors declare that they have no conflicts of interest with the contents of this article.

554

555 **References**
556

557 1   Thresh, J. M. Control of tropical plant virus diseases. *Adv Virus Res* **67**, 245-295,
558   doi:10.1016/S0065-3527(06)67007-3 (2006).

559 2   Nweke, F. I. A Cash Crop in Africa. COSCA Working Paper No. 14. Collaborative Study of
560   Cassava in Africa, International Institute of Tropical Agriculture, Ibadan, Nigeria.
561   (1996).

562 3   Robertson, A. I. & Ruhode, T. The potential advantages of cassava over hybrid maize as
563   a food security crop and a cash crop in the Southern Africa semi-arid zone. *African*
564   *Crop Science Crop Science Conference Proceedings* **5**, 539-542 (2001).

565 4   Legg, J. P. *et al.* Comparing the regional epidemiology of the cassava mosaic and
566   cassava brown streak virus pandemics in Africa. *Virus Res* **159**, 161-170,
567   doi:10.1016/j.virusres.2011.04.018 (2011).

568 5   Legg, J. P., Owor, B., Sseruwagi, P. & Ndunguru, J. Cassava mosaic virus disease in East
569   and Central Africa: epidemiology and management of a regional pandemic. *Adv Virus*
570   *Res* **67**, 355-418, doi:10.1016/S0065-3527(06)67010-3 (2006).

571   6   Hall, F. W. Annual Report of the Department of Agriculture, Uganda. *Government*

572   *Printer, Entebbe*, 35 (1928).

573   7   Jameson, J. D. Cassava mosaic disease in Uganda. *East African Agricultural and Forestry*

574   *Journal* **29**, 208-213 (1964).

575   8   Otim-Nape, G. W. *et al.* The current pandemic of cassava mosaic virus disease in East

576   Africa and its control. *NARO/NRI/DFID Publication. Chatham, UK. 100pp* (2000).

577   9   Storey, H. H. Virus diseases of East African plants: VI, A progress report of studies of

578   the diseases of cassava. *East African Agricultural Journal* **2**, 34-39 (1936).

579   10   Alicai, T. *et al.* Re-emergence of cassava brown streak disease in Uganda. *Plant Disease*

580   **91**, 24-29 (2007).

581   11   Ntawuruhunga, P. & Legg, J. P. New Spread of Cassava Brown Streak Virus Disease and

582   Its Implications for The Movement of Cassava Germplasm in The East and Central

583   African Region. International Institute of Tropical Agriculture-Uganda & Eastern Africa

584   Root Crops Research Network Report. (2007).

585   12   Bigirimana, S., Barumbanze, P., Ndayihanzamaso, P., Shirima, R. & Legg, J. P. First

586   report of cassava brown streak disease and associated Ugandan cassava brown streak

587   virus in Burundi. *New Disease Reports* **24**, 26 (2011).

588   13   Mbanzibwa, D. R. *et al.* Evolution of cassava brown streak disease-associated viruses. *J*

589   *Gen Virol* **92**, 974-987, doi:10.1099/vir.0.026922-0 (2011).

590   14   Mulimbi, W. *et al.* First report of Ugandan cassava brown streak virus on cassava in

591   Democratic Republic of Congo. *New Disease Reports* **26** (2012).

592   15   Winter, S. *et al.* Analysis of cassava brown streak viruses reveals the presence of

593   distinct virus species causing cassava brown streak disease in East Africa. *J Gen Virol*

594   **91**, 1365-1372, doi:10.1099/vir.0.014688-0 (2010).

595   16   Monger, W. A., Seal, S., Isaac, A. M. & Foster, G. D. Molecular characterization of

596   cassava brown streak virus coat protein. *Plant Pathol* **50**, 527-534 (2001).

597   17   Monger, W. A. *et al.* The complete genome sequence of the Tanzanian strain of
598   Cassava brown streak virus and comparison with the Ugandan strain sequence. *Arch*
599   *Virol* **155**, 429-433, doi:10.1007/s00705-009-0581-8 (2010).

600   18   Mbanzibwa, D. R. *et al.* Genetically distinct strains of Cassava brown streak virus in the
601   Lake Victoria basin and the Indian Ocean coastal area of East Africa. *Arch Virol* **154**,
602   353-359, doi:10.1007/s00705-008-0301-9 (2009).

603   19   Ndunguru, J. *et al.* Analyses of Twelve New Whole Genome Sequences of Cassava
604   Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa:
605   Diversity, Supercomputing and Evidence for Further Speciation. *PLoS One* **10**,
606   e0139321, doi:10.1371/journal.pone.0139321 (2015).

607   20   Maruthi, M. N. *et al.* Transmission of Cassava brown streak virus by Bemisia tabaci
608   (Gennadius). *J Phytopathol* **153**, 307-312 (2005).

609   21   Lister, R. M. Mechanical transmission of cassava brown streak virus. *Nature* **183**, 1588-
610   1589 (1959).

611   22   Jennings, D. L. Observations on virus diseases of cassava in resistant and susceptible
612   varieties. II. Brown streak disease. *Empire Journal of Experimental Agriculture* **28**, 261-
613   269 (1960).

614   23   Nichols, R. F. W. The brown streak disease of cassava: Distribution, climatic effects and
615   diagnostic symptoms. *East African Agricultural Journal* **15**, 154-160 (1950).

616   24   Ogwok, E., Ilyas, M., Alicai, T., Rey, M. E. & Taylor, N. J. Comparative analysis of virus-
617   derived small RNAs within cassava (Manihot esculenta Crantz) infected with cassava
618   brown streak viruses. *Virus Res* **215**, 1-11, doi:10.1016/j.virusres.2016.01.015 (2016).

619   25   Patil, B. L. *et al.* RNAi-mediated resistance to diverse isolates belonging to two virus
620   species involved in Cassava brown streak disease. *Mol Plant Pathol* **12**, 31-41,
621   doi:10.1111/j.1364-3703.2010.00650.x (2011).

622 26   Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J. & Maruthi, M. N. The
623   symptom and genetic diversity of cassava brown streak viruses infecting cassava in
624   East Africa. *Adv Virol* **2012**, 795697, doi:10.1155/2012/795697 (2012).

625 27   Mohammad, I. U., Ghosh, S. & Maruthi, M. N. Host and virus effects on reversion in
626   cassava affected by cassava brown streak disease. *Plant Pathology* **65**, 593-600 (2016).

627 28   Jiang, J., Patarroyo, C., Garcia Cabanillas, D., Zheng, H. & Laliberte, J. F. The Vesicle-
628   Forming 6K2 Protein of Turnip Mosaic Virus Interacts with the COPII Coatomer Sec24a
629   for Viral Systemic Infection. *J Virol* **89**, 6695-6710, doi:10.1128/JVI.00503-15 (2015).

630 29   Dombrovsky, A., Reingold, V. & Antignus, Y. Ipomovirus--an atypical genus in the family
631   Potyviridae transmitted by whiteflies. *Pest Manag Sci* **70**, 1553-1567,
632   doi:10.1002/ps.3735 (2014).

633 30   Revers, F. & Garcia, J. A. Molecular biology of potyviruses. *Adv Virus Res* **92**, 101-199,
634   doi:10.1016/bs.aivir.2014.11.006 (2015).

635 31   Sanjuan, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-
636   nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A* **101**, 8396-8401,
637   doi:10.1073/pnas.0400146101 (2004).

638 32   Lauring, A. S., Frydman, J. & Andino, R. The role of mutational robustness in RNA virus
639   evolution. *Nat Rev Microbiol* **11**, 327-336, doi:10.1038/nrmicro3003 (2013).

640 33   Acevedo, A., Brodsky, L. & Andino, R. Mutational and fitness landscapes of an RNA
641   virus revealed through population sequencing. *Nature* **505**, 686-690,
642   doi:10.1038/nature12861 (2014).

643 34   Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. Evolution of digital
644   organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331-333,
645   doi:10.1038/35085569 (2001).

646 35   Sanjuan, R., Cuevas, J. M., Furio, V., Holmes, E. C. & Moya, A. Selection for robustness
647   in mutagenized RNA viruses. *PLoS Genet* **3**, e93, doi:10.1371/journal.pgen.0030093
648   (2007).

649   36   Massingham, T. & Goldman, N. Detecting amino acid sites under positive selection and
650   purifying selection. *Genetics* **169**, 1753-1762, doi:10.1534/genetics.104.032144 (2005).

651   37   Blevins, T. *et al.* Four plant Dicers mediate viral small RNA biogenesis and DNA virus
652   induced silencing. *Nucleic Acids Res* **34**, 6233-6246, doi:10.1093/nar/gkl886 (2006).

653   38   Kubatko, L. S. & Degnan, J. H. Inconsistency of phylogenetic estimates from
654   concatenated data under coalescence. *Syst Biol* **56**, 17-24 (2007).

655   39   Degnan, J. H. & Salter, L. A. Gene tree distributions under the coalescent process.
656   *Evolution* **59**, 24-37 (2005).

657   40   Monger, W. A. *et al.* The complete genome sequence of the Tanzanian strain of
658   Cassava brown streak virus and comparison with the Ugandan strain sequence. *Arch*
659   *Virol* **155**, 429-433 (2010).

660   41   Hahn, S. K., Isoba, J. C. G. & Ikotun, T. Resistance breeding in root and tuber crops at
661   the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria. *Crop*
662   *Protection* **8**, 147-168 (1989).

663   42   Hillocks, R. J., Raya, M. D. & Thresh, J. M. The association between root necrosis and
664   above ground symptoms of brown streak virus infection of cassava in southern
665   Tanzania. *International Journal of Pest Management* **42**, 285-289 (1996).

666   43   Mbanzibwa, D. R. *et al.* Simultaneous virus-specific detection of the two cassava brown
667   streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed
668   infections, and infections in Manihot glaziovii. *J Virol Methods* **171**, 394-400,
669   doi:10.1016/j.jviromet.2010.09.024 (2011).

670   44   Ndunguru, J., Legg, J. P., Aveling, T. A., Thompson, G. & Fauquet, C. M. Molecular
671   biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses
672   in Africa and evidence for East Africa being a center of diversity of cassava
673   geminiviruses. *Virol J* **2**, 21, doi:10.1186/1743-422X-2-21 (2005).

674   45   Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
675   search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

676   46   Geneious v5.1. Available from http://www.geneious.com/ (2010).

677   47   Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple

678   sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066

679   (2002).

680   48   Kehoe, M. A., Coutts, B. A., Buirchell, B. J. & Jones, R. A. Split personality of a Potyvirus:

681   to   specialize   or   not   to   specialize?   *PLoS   One*   **9**,   e105770,

682   doi:10.1371/journal.pone.0105770 (2014).

683   49   Kehoe, M. A., Coutts, B. A., Buirchell, B. J. & Jones, R. A. Plant virology and next

684   generation   sequencing: experiences   with   a   Potyvirus.   *PLoS   One*   **9**,   e104580,

685   doi:10.1371/journal.pone.0104580 (2014).

686   50   Korber, B. in *Computational Analysis of HIV Molecular Sequences*   (eds A.G. Rodrigo &

687   G. H. Learn) Ch. 4, 55-72 (Kluwer Academic Publishers, 2000).

688   51   Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and

689   nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418-426 (1986).

690   52   Ota, T. & Nei, M. Variance and covariances of the numbers of synonymous and

691   nonsynonymous substitutions per site. *Mol Biol Evol* **11**, 613-619 (1994).

692   53   Ganeshan, S., Dickover, R. E., Korber, B. T., Bryson, Y. J. & Wolinsky, S. M. Human

693   immunodeficiency virus type 1 genetic evolution in children with different rates of

694   development of disease. *J Virol* **71**, 663-677 (1997).

695   54   Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood.

696   *Comput Appl Biosci* **13**, 555-556 (1997).

697   55   Kosakovsky Pond, S. L. & Frost, S. D. Not so different after all: a comparison of

698   methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208-1222,

699   doi:10.1093/molbev/msi105 (2005).

700 56 Delport, W., Poon, A. F., Frost, S. D. & Kosakovsky Pond, S. L. Datamonkey 2010: a suite
701   of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455-2457,
702   doi:10.1093/bioinformatics/btq429 (2010).

703 57 Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model
704   Choice Across a Large Model Space. *Systematic Biology* **61**, 539-542,
705   doi:10.1093/sysbio/sys029 (2012).

706 58 Ayres, D. L. *et al.* BEAGLE: an application programming interface and high-performance
707   computing library for statistical phylogenetics. *Systematic Biology* **61**, 170-173,
708   doi:10.1093/sysbio/syr100 (2012).

709 59 Chifman, J. & Kubatko, L. Quartet inference from SNP data under the coalescent
710   model. *Bioinformatics* **30**, 3317-3324, doi:10.1093/bioinformatics/btu530 (2014).

711 60 Phylogenetic Analysis using Parsimony (* and other methods) v. (open source version
712   4.0a147 downloaded from http://people.sc.fsu.edu/~dswofford/paup_test/ on March 25,
713   2016). (Sinauer Associates, Sunderland, MA 2002).

714 61 Allman, E. S., Kubatko, L., Pearl, D. K. & Rhodes, J. A. Singular Value Decomposition of
715   Site Pattern Frequency Matrices as a Tool to Quantify Phylogenetic Signal in Genome-
716   Scale Data. **in press** (2016).

717
718

719   Table 1: CBSD leaf symptom severities and types on plants infected by *Cassava brown streak*
720   *virus* and *Ugandan cassava brown streak virus*
721

Virus species	Number of plants with CBSD foliar symptom severity score[1]				Number of plants with CBSD foliar symptom type[2]		
	2	3	4	5	SW	LL	SL
CBSV	5	7	1	0	3	8	2
UCBSV	24	5	0	0	7	21	1
CBSV + UCBSV	9	6	0	0	5	10	0
Total	38	18	1	0	15	39	3
Percentage	66.7	31.6	1.7	0.0	26.3	68.4	5.3

722
723

724   [1]Foliar CBSD symptom severity score based on 1-5 scale; 1 = no visible symptoms, 2 = mild vein
725   yellowing or chlorotic blotches on some leaves, 3 = pronounced/extensive vein yellowing
726   or chlorotic blotches on leaves, but no lesions or streaks on stems, 4 = pronounced/extensive
727   vein yellowing or chlorotic blotches on leaves and mild lesions or streaks on stems, 5 =
728   pronounced/extensive vein yellowing or chlorotic blotches on leaves and severe lesions or
729   streaks on stems, defoliation and dieback.
730   [2]Types of foliar CBSD symptoms based on distribution of leaf chlorosis and stem lesions on the
731   plant; systemic and on the whole plant (SW), systemic on leaf or stem parts but localized (SL),
732   only on lower leaves (LL).
733

31

4 Table 2. Next generation sequencing data for samples from cassava brown streak disease symptomatic plants collected in Uganda

5

Sample ID	Accession number	Virus	No. of reads obtained	No. of reads after trimming	Number of contigs produced (CLC)	Contig length (CLCGW, nt)	Average coverage (CLCGW)	Number of reads mapped to contig of interest	Ref seq. used for mapping	Length of consensus sequence from mapping (Geneious)	No. reads mapped to ref. sequence	Average coverage (Geneious)	Final sequence length (Coding region only)
U1		*CBSV*	23,335,344	23,053,082	726	3919, 2214	31, 24	1264, 549	KR108828	8,893	2,233	25	8,748
U4		*CBSV*	21,844,716	21,582,374	621	8,949	255	23,658	KR108828	8,949	22,987	256	8,748
U8		*UCBSV*	23,648,990	23,373,606	1,008	8,954	366	33,778	KR108836	9,563	178,117	393	8,700

6

7

738 Table 3. *Cassava brown streak virus* (CBSV) amino acid (AA) sites under positive selection
739 (analyses method: SLAC Hy-Phy). There were no sites under positive selection for
740 *Ugandan cassava brown streak virus* (UCBSV).
741
742

Gene	CBSV AA site under positive selection
P1	44, 46, 50, 174, 224, 230, 250, 283, 288, 349, 358
P3	415, 455, 467, 472, 499, 525, 618
6K1	658, 678
CI	735, 761, 820, 827, 848, 852, 894, 935, 1218
VPg	1465
NIa	1620, 1645, 1704, 1754, 1785
Nib	1879, 1880, 1890, 1907, 1929, 2109, 2145, 2156, 2161, 2285
HAM1	2320, 2345, 2404, 2432, 2453, 2475, 2519
CP	2550, 2555, 2588, 2611, 2631, 2635, 2640, 2659, 2728, 2745, 2783, 2818, 2843, 2860, 2877, 2884

743
744
745
746
747
748
749
750
751
752
753
754
755
756
757

758 Table 4. Rates of evolution tested using CODEML implemented in PAML. H_O was CBSV and
759 UCBSV have equal rates of evolution (one omega; model = 0), while H_1 was that CBSV
760 and UCBSV have different rates of evolution (two omegas; model = 2).
761

Gene	Assumptions	K (ts/tv rate ratio)	w (omega Dn/Ds) 0	w (omega Dn/Ds)1	Likelihood Ratio Test (if greater than 3.84 reject H_0) H_0=equal rates
WGS	UCBSV and CBSV equal rates	5.90944	0.06358		
	UCBSV and CBSV different rates	5.9598	0.05518	**0.07622**	26.29*
P1	UCBSV and CBSV equal rates	5.07336	0.10394		
	UCBSV and CBSV different rates	5.05456	0.09047	**0.12203**	4.61*
P3	UCBSV and CBSV equal rates	5.17197	0.08635		
	UCBSV and CBSV different rates	5.20559	0.0764	**0.10198**	2.22
6K1	UCBSV and CBSV equal rates	19.54143	0.00969		
	UCBSV and CBSV different rates	19.69676	**0.01527**	0.00316	3.01
CI	UCBSV and CBSV equal rates	9.9388	0.01722		
	UCBSV and CBSV different rates	8.06276	0.0155	**0.01977**	0.73
6K2	UCBSV and CBSV equal rates	8.40649	0.04684		
	UCBSV and CBSV different rates	8.82738	0.02354	**0.11057**	6.74*
VPg	UCBSV and CBSV equal rates	5.852	0.05759		
	UCBSV and CBSV different rates	5.87009	0.054	**0.06323**	0.29
NIa	UCBSV and CBSV equal rates	8.01105	0.02932		
	UCBSV and CBSV different rates	8.68283	0.01408	**0.06719**	29.95*
NIb	UCBSV and CBSV equal rates	6.07872	0.05329		
	UCBSV and CBSV different rates	6.14047	0.0452	**0.06508**	5.18*
HAM1	UCBSV and CBSV equal rates	7.25177	0.16144		
	UCBSV and CBSV different rates	7.2343	**0.17929**	0.14007	2.23
CP	UCBSV and CBSV equal rates	13.09297	0.06075		
	UCBSV and CBSV different rates	13.26752	0.05606	**0.07155**	1.29
			faster rate UCBSV	**faster rate CBSV**	* Rates are different

34

762

763   Table 5. Support for Clades A – G (Figure 3) in individual gene trees and whole genome analyses. Table entries represent
764   posterior probabilities from analysis with MrBayes, except values reported for SVDQ, which are bootstrap
765   proportions. Support values below 95% are indicated in bold, and '--' indicates that the clade was not present.

766

767

Genomic region	Clade A	Clade B	Clade C	Clade D	Clade E	Clade F	Clade G
P1	0.9998	0.9998	0.9698	0.9972	0.9954	--	0.9998
P3	0.9998	0.9998	0.9857	0.9998	0.9998	--	0.9998
6K1	**0.9466**	0.9985	0.9887	0.9995	--	--	0.9842
CI	0.9998	0.9996	0.9976	0.9998	0.9998	0.9999	0.9999
6K2	0.9765	0.9998	**0.7152**	**0.9170**	**0.6646**	--	0.9875
Vpg	0.9998	0.9999	--	0.9918	**0.6455**	--	0.9999
NIa	0.9996	0.9999	0.9998	0.9941	**0.7631**	--	**0.8934**
NIb	0.9999	0.9998	0.9998	0.9999	0.9998	--	0.9998
HAM1-like	--	--	--	0.9989	0.9983	--	0.9855
CP	--	--	--	0.9959	0.9950	--	0.9989
Whole genome	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SVDQ	1.00	1.00	1.00	1.00	**0.87**	**0.44**	1.00

768

769

770 Figure 1: Cassava brown streak disease symptoms on leaves and stems of sampled plants; **(a)**
771 Chlorosis along secondary and tertiary leaf veins of CBSV-infected plant of cultivar TME 204
772 (severity score 3), **(b)** Cultivar TME 14 plant with dual CBSV+UCBSV infection showing chlorosis
773 on secondary or tertiary veins, reverse chlorosis (general chlorosis and green area along veins)
774 (severity score 3), **(c)** UCBSV-infected plant of cultivar TME 204 exhibiting chlorosis on
775 secondary veins, reverse chlorosis, chlorotic spots and mild stem lesions (severity score 3), **(d)**
776 Very severely diseased plant (severity score 5) of cultivar TME 14 infected with both CBSV and
777 UCBSV, and having chlorosis on leaves, severe stem lesions/brown streaks, defoliation, stem
778 dieback.

779

780

781   Figure 2. Genetic diversity of CBSV and UCBSV using the <u>S</u>ynonymous <u>N</u>on-synonymous <u>A</u>nalysis <u>P</u>rogram (SNAP v2.1.1) implemented in the
782   Los Alamos National Laboratory HIV-sequence database (http://www.hiv.lanl.gov) [50]. UCBSV is on the top panel, CBSV at the
783   bottom. The 10 gene segments are labeled from P1-CP.
784

785

786   Figure 3. Species tree generated from SVD Quartets using the whole genome sequences. Colors
787   at the tips are based on country of origin. Branches with mixed colors indicate a clade
788   that contains samples with mixed country of origin. For example, the ancestral branch
789   of UCBSV TZ Tan 23 KR108839 and UCBSV UG MI B3 FJ039520 is colored red and
790   orange to indicate a clade with sampled with mixed country of origin.

791

Figure 4. Computed SVD Score with the split defined by CBSV vs. UCBSV across the genome in windows of 500 bp, sliding in increments of 100 bp, and resulting SVD Scores plotted across the genome. Boundaries between genes are marked with vertical lines to further characterize the CBSV and UCBSV genomes. Rates of molecular evolution were estimated using CODEML implemented in PAML (Phylogenetic Analysis by Maximum Likelihood) [54]. The results are shown for each gene and D represents the difference in likelihoods from the null hypothesis (CBSV and UCBSV have equal rates) and the alternative hypothesis (CBSV and UCBSV have different rates).

Supplemental Figure 1. CBSV Amino Acid variability obtained using datamonkey.org. Once files are uploaded to the site, the images below are obtained from the "Information from upload" tab and the pdf is downloaded.

Supplemental Figure 1. CBSV Amino Acid variability (positions 5–320), showing per-position residue counts.

```
        325       330       335       340       345       350       355       360
E₉ H₁₃D₉ M₉ F₁₀D₉ F₉ V₉ D₉ D₉ M₉ C₉ V₉ V₉ Q₉ G₉ K₁₀N₉ K₉ I₇ S₉ G₉ K₉ I₇ E₉ N₉ A₉ L₉ V₉ S₁₀K₉ S₉ W₉ D₉ E₉ L₉ Q₉ Q₆ I₈ E₉
F₄ M₁ F₅ I₇ D₄ G₄ M₄ C₄ V₄ V₄ Q₄ G₄ K₄ N₄ K₄ S₃ S₃ G₄ R₄ G₁ E₄ N₄ A₄ L₄ V₄ L₃ K₄ S₅ S₂ E₃ D₃ L₄ S₃ E₄ I₄ E₄ L₄ Y₄ S₄ F₄
D₁       V₂   V₁ N₁ D₁ M₁ C₁ V₁ V₁ Q₁ G₁ E₁ E₁ N₁ I₁ S₁ V₁ K₁ I₁ E₁ N₁ A₁ P₂ V₁   K₁ D₁ W₁ D₁ G₁ L₁ Q₁ Q₁ I₁ H₁ M₁ Y₁
     N₁                                 N₁                                       N₁   E₁  E₁                   E₁
                                                                                D₁
        365       370       375       380       385       390       395       400
M₉ Y₉ S₁₃F₉ D₉ L₉ S₉ W₉ A₉ R₁₀S₉ R₉ D₈ D₉ F₁₃F₉ K₈ Y₉ F₉ D₉ E₉ D₉ V₉ G₉ S₉ L₉ I₃ S₉ T₉ C₉ C₉ T₉ P₉ S₈ T₉ L₁₁W₉ L₉ Y₉ A₉
D₄ L₄ D₁ W₄ A₄ R₄ T₄ R₅ D₄ N₄ F₄ F₃ R₄ H₄ K₁ D₄ D₄ D₅ V₄ G₄ K₄ L₄ I₆ S₄ T₄ C₄ C₄ T₄ P₄ S₄ T₄ L₄ W₄ L₅ Y₄ A₄ K₄ K₄ A₄ S₄
S₁ F₁   L₁ S₁ W₁ A₁    S₁   D₁ D₁ F₁ F₁   Y₁ F₁   E₁ N₁ V₁ S₁ S₁ L₁ I₁ S₁ T₁ C₁ C₁ T₁ P₁ S₁ T₁ T₁ W₁   Y₁ A₁ K₁ K₁
                Y₁ N₁
        405       410       415       420       425       430       435       440
K₉ K₉ A₉ S₉ F₉ Y₁₀K₉ Y₉ V₉ D₉ H₈ M₉ I₇ L₉ K₉ G₉ S₉ P₉ L₁₃V₁₃D₉ I₇ L₉ V₉ K₉ M₉ E₉ Y₉ V₉ G₉ K₈ H₉ L₉ E₉ M₉ F₉ N₉ S₉ V₉ E₉
F₄ Y₄ K₄ Y₅ V₄ D₄ H₄ M₄ M₄ L₄ K₄ G₄ S₄ P₄ L₄ V₄ D₄ M₄ D₁ I₇ K₄ M₄ E₄ Y₄ V₄ G₄ K₄ H₆ L₄ E₄ M₄ F₄ N₄ S₄ V₄ E₄ D₄ V₄ C₄ L₄
A₁ S₁ F₁   K₁   V₁ D₁ H₁ M₁ I₁ L₁ K₁ G₁ S₁ P₁ L₁ V₁   L₁ V₁ K₁ M₁ E₁ Y₁ V₁ G₁ K₁ H₁ L₂ E₁ M₁ F₁ N₁ S₁ V₁ E₁ D₁ V₁
        445       450       455       460       465       470       475       480
D₉ V₉ C₉ L₉ E₉ Y₉ A₉ H₉ F₉ M₉ Q₄D₉ L₉ M₉ Q₅ D₁₃H₈ V₉ T₈ D₉ Q₈ S₉ N₈ E₉ D₈ I₇ L₉ R₉ V₉ K₉ N₉ L₉ I₇ R₉ S₁₃Y₇ F₉ D₉ S₁₃V₉
E₄ Y₄ A₄ H₄ F₄ M₄ Q₄ D₄ M₂ I₇ L₁ E₄ Q₅ V₃ N₄ V₁ Q₅ S₄ D₄ E₄ N₄ M₄ M₄ R₄ V₄ K₄ N₄ L₄ I₄ R₄ S₄ Y₄ F₄ D₄ F₁ V₄ I₄ E₄ I₁ N₄
C₁ L₁ E₁ Y₁ A₁ H₁ F₁ M₁ V₂ D₁   M₁    I₁ C₁    T₁ D₁ Q₁ S₁ H₁ E₁ D₁ I₁ L₁ R₁ V₁ K₁ N₁ L₁ I₁ R₁ S₁ Y₁   D₁ S₁ V₁   E₁
              Q₁         D₁       N₁   S₁   T₁   N₁                               M₁
        485       490       495       500       505       510       515       520
I₇ E₉ A₉ N₈ K₉ Y₉ E₉ L₉ I₇ D₁₂R₉ I₇ I₇ D₈ K₉ K₉ T₈ Q₁₃L₉ E₉ A₉ Q₉ E₉ I₇ I₇ S₉ R₉ E₉ L₉ I₇ K₉ H₃ Q₉ Y₉ G₉ A₉ L₉ F₉ S₉ L₁₃
K₄ Y₄ E₄ L₄ I₄ D₄ S₄ I₄ R₁ I₄ K₄ K₄ S₄ E₅ L₄ E₄ A₄ E₁ E₄ V₄ I₄ S₄ R₄ E₄ L₄ I₄ K₃ H₄ Q₄ Y₄ G₄ D₄ L₄ F₄ S₄ L₄ R₄ E₄ R₅ E₁
A₁ N₁ K₁ S₁ E₁ L₁ I₁ D₁   E₁ I₁ E₁ K₁ K₁ T₁ Q₁ L₁   A₁ Q₁ E₁ I₁ I₁ S₁ R₁ E₁ L₁ I₁ K₁ H₁ Q₁ Y₁ G₁ A₁ L₁ F₁ S₁ L₁
         Y₁                        S₁                     S₁
        525       530       535       540       545       550       555       560
R₁₀E₉ R₉ I₇ S₉ I₇ N₉ I₇ G₉ F₉ D₉ S₁₄A₉ S₁₀L₉ S₁₁S₉ L₉ W₉ K₉ E₁₀S₉ E₉ S₈ L₉ S₁₃R₉ S₁₀D₉ S₁₃S₉ F₉ F₉ S₁₃C₉ L₉ M₉ S₉ S₉ P₉
L₄ I₄ N₄   G₄ F₄ D₄ S₄ A₄ K₄ L₄   S₄ L₄ W₄ E₁ E₄ S₄ E₅ S₅ V₄ K₄ Q₄ K₆ S₃ K₁ S₄ I₂ I₄ I₇ C₄ L₄ I₄ L₁ K₄ P₄ G₄ M₄ E₄ M₅
   M₁ S₁   N₁ V₁ G₁ F₁ D₁ S₁ A₁   L₁   S₁ L₁ W₁ K₁       M₁ L₁   G₁   D₁ V₂ S₁   F₁ S₁ C₁   M₁ S₁ S₁ P₁ G₁
                        Q₁                        R₁
        565       570       575       580       585       590       595       600
G₉ M₁₀E₉ M₉ C₉ A₉ N₈ W₉ I₇ L₈ S₈ V₉ C₉ S₁₄S₉ S₁₀Y₉ S₉ N₉ T₉ L₉ Y₉ L₉ V₉ D₉ K₈ G₉ V₉ V₈ F₉ F₉ W₉ S₁₃S₉ M₅ V₉ H₉ I₇ F₉ Q₉
L₃ A₄ N₄ W₄ I₄ S₄ E₄ V₃ F₄ K₄ N₄ C₄ Y₄   N₄ T₄ M₄ Y₄ L₅ V₄ D₄ R₄ S₄ M₄ M₄ F₄ I₄ W₄ N₄ R₄ M₄ V₄ V₁ I₄ F₄ N₄ S₄ C₄ V₄ Y₄
F₁   C₁ A₁ N₁ W₁ T₁ A₁ T₁ S₁ K₁ S₁ S₁   Y₁   N₁ I₁   Y₁ M₁ V₁ D₁ K₁ G₁ V₁ V₁ F₁ F₁ W₁ S₁ S₁   V₁ V₄ I₁ F₁ Q₁ T₁ C₁
E₁           S₁ L₁ S₁ V₁ C₁             Q₁                             S₁       I₁             H₁
        605       610       615       620       625       630       635       640
T₉ C₉ V₉ Y₁₀G₉ Y₉ W₉ N₉ L₉ W₉ F₉ S₉ K₉ A₉ V₁₀C₉ V₉ L₉ M₉ A₉ F₉ V₉ I₇ M₉ G₉ F₉ S₉ G₉ K₉ A₅ M₉ D₉ Y₉ L₉ K₁₃K₉ L₉ I₇ K₁₃N₉
G₄ Y₅ W₄ N₄ L₄ W₄ F₄ R₄ K₄ A₄ V₃ C₄ V₅ L₄ M₄ V₄ F₄ V₃ V₄ T₄ G₄ F₄ S₄ S₄ K₃ I₇ I₇ E₃ F₄ L₄ K₃ K₄ L₄ I₁ L₁ N₄ E₄ S₄ E₁ Q₄
V₁   G₁   W₁ N₁ L₁ W₁ F₁ S₁ M₁ A₁   C₁   L₁ M₁ A₁ F₁ V₁ I₇ M₁ G₁ F₁ S₂ G₁ K₁ V₁ M₁ V₄ Y₁ L₁ K₁ M₁   I₁ K₁ N₁   S₁
                        K₁         M₁                       D₁ D₁ S₁   K₁       M₁
        645       650       655       660       665       670       675       680
E₉ S₉ K₁₀Q₉ A₉ I₇ Q₉ Y₉ E₉ E₁₃G₉ L₉ V₁₀E₉ V₉ Q₉ G₉ S₁₃S₈ E₉ E₉ S₉ F₉ V₉ L₉ R₉ W₉ C₉ A₁₀F₁₄A₉ T₉ L₁₀F₉ L₉ S₉ F₉ I₇ N₉ Y₉
A₄ M₄ Q₃ F₄ E₄ E₄ G₄ L₄ V₃ L₁ V₅ Q₄ G₄ S₄ K₄ E₄ E₄ E₁ F₄ V₄ L₄ R₄ W₄ C₄ A₄ F₄ A₅ T₄ L₄ T₁ L₅ S₄ F₄ M₃ N₄ Y₄ D₄ W₄ A₄ V₄
S₁ Q₁ A₁ I₁ Q₁ Y₁ E₁ E₁ G₁       E₁   Q₁ G₁ S₁ K₁   K₁ S₁ F₁ V₁ L₁ R₁ W₁ C₁   F₁       F₁   I₁ F₁ I₁ N₁ Y₁ D₁ W₁
                  I₁                         E₁                                          S₁
        685       690       695       700       705       710       715       720
D₉ W₉ A₉ V₉ G₉ S₁₀V₉ S₉ A₉ I₇ G₉ K₁₀M₉ K₉ T₉ V₆ F₉ G₉ A₉ L₉ G₉ P₉ D₉ F₉ I₇ E₉ S₁₀Q₉ S₁₂G₉ D₁₃D₁₀N₉ D₉ E₉ L₉ S₆ F₉ T₉ T₉
G₄ S₄ V₄ S₅ A₄ I₇ G₄ K₄ M₄ K₅ T₄ M₄ F₄ G₄ A₄ L₄ G₄ P₄ D₄ F₄ I₇ E₄ S₄ Q₅ S₃ G₄ D₄ E₂ D₁ D₅ N₁ L₄ K₄ F₄ T₄ T₄ F₄ E₄ V₄ E₅
A₁ V₁ G₁   V₁   A₁ I₇ G₁   M₁   T₁ I₇ F₁ I₇ A₁ L₁ G₁ P₁ D₁ F₁ I₇ E₁ N₂ Q₁   D₂ N₁       E₁ L₁ S₁ F₁ K₃ T₁ F₁
                     G₁                                                    G₁                   T₁
```

41

```
        725      730      735      740      745      750      755      760
```

```
        765      770      775      780      785      790      795      800
```

```
        805      810      815      820      825      830      835      840
```

```
        845      850      855      860      865      870      875      880
```

```
        885      890      895      900      905      910      915      920
```

```
        925      930      935      940      945      950      955      960
```

```
        965      970      975      980      985      990      995     1000
```

```
       1005     1010     1015     1020     1025     1030     1035     1040
```

```
       1045     1050     1055     1060     1065     1070     1075     1080
```

```
       1085     1090     1095     1100     1105     1110     1115     1120
```

```
       1125     1130     1135     1140     1145     1150     1155     1160
```

42

```
        1165      1170      1175      1180      1185      1190      1195      1200
S₈ S₉ Y₉ L₉ A₉ T₉ N₉ M₉ S₉ E₉ W₉ S₉ R₈ V₉ C₉ E₉ Y₉ I₇ K₉ I₇ D₉ E₉ S₉ S₉ R₉ H₉ V₉ Q₉ E₉ V₉ K₈ I₇ P₉ W₉ Y₉ C₇ S₉ D₁₃M₉ S₉
A₄ T₄ N₄ M₄ S₄ E₄ W₄ S₄ R₄ V₄ C₄ E₄ Y₄ I₄ K₄ I₇ D₃ M₅ S₄ S₄ S₅ H₄ V₃ Q₄ E₄ V₄ K₄ M₄ P₄ W₄ Y₄ C₄ S₄ D₄ M₄ S₄ D₄ S₁ F₄ I₄
Y₁ L₁ A₁ T₁ N₁ M₁ S₁ E₁ W₁ S₁ R₁ V₁ C₁ E₁ Y₁ T₂ K₁ D₃ D₁ E₁    S₁ M₁ H₁ V₁ Q₁ E₁ V₁ K₁ I₇ P₁ W₁ Y₁ C₁ S₁ D₁ M₁    D₁ D₁
N₁                                    E₁ E₁    V₁       R₁                            S₁
        1205      1210      1215      1220      1225      1230      1235      1240
D₉ D₉ F₉ I₇ V₉ K₉ L₉ A₉ E₉ C₉ V₉ K₉ A₉ A₉ K₁₀P₉ K₉ L₈ S₁₂S₉ G₉ Y₉ K₉ V₉ D₉ N₉ V₉ D₉ F₉ H₁₁T₉ V₉ A₉ H₉ K₉ M₉ S₁₃V₉ G₉ E₉
V₄ K₄ L₄ A₄ E₄ C₄ V₄ K₄ A₄ S₄ K₄ P₄ K₅ L₄ S₄ S₄ G₄ Y₄ G₁ V₄ D₄ N₄ V₄ D₄ F₄ H₄ T₄ V₄ A₄ V₁ K₄ M₄ S₄ V₄ G₄ E₄ G₁ N₄ I₄ D₄
F₁ I₇ V₁ K₁ L₁ A₁ E₁ C₁ V₁ K₁ A₁ A₁    P₁    L₁ S₁ F₁ K₁ Y₁ K₁ V₁ D₁ N₁ V₁ D₁ F₁ H₁ T₁    A₁ H₁ K₁ M₁ S₁ V₁    E₁ S₁ N₁
                                       S₁
        1245      1250      1255      1260      1265      1270      1275      1280
S₉ N₉ I₇ D₉ E₉ S₉ S₉ A₉ L₉ V₉ A₉ T₉ I₇ L₉ D₉ E₉ V₉ K₉ Q₉ W₉ S₉ D₉ G₉ M₉ T₉ Y₉ H₉ S₉ S₉ T₉ P₉ S₉ N₉ K₉ S₉ L₉ M₉ S₉ L₉ M₉
E₄ S₄ S₄ S₅ L₄ V₄ A₄ T₄ I₄ L₄ D₄ E₄ V₄ K₄ Q₄ W₄ S₄ D₄ G₄ M₄ T₄ Y₄ H₄ S₄ S₄ T₄ P₄ R₄ N₄ K₃ S₄ L₄ M₄ S₄ L₄ M₄ V₄ G₄ W₄ I₄
I₇ D₁ E₁    S₁ A₁ L₁ V₁ A₁ T₁ I₇ L₁ D₁ E₁ V₁ K₁ Q₁ W₁ S₁ D₁ G₁ M₁ T₁ Y₁ H₁ S₁ S₁ T₁ P₁ S₂ N₁ K₁ S₁ L₁ M₁ S₁ L₁ M₁ V₁ G₁
        1285      1290      1295      1300      1305      1310      1315      1320
V₉ G₉ W₉ I₇ P₉ R₉ K₉ A₉ E₉ S₉ T₉ K₉ A₉ M₉ L₉ D₉ E₉ R₉ V₉ Q₉ S₉ L₁₄E₉ L₁₀L₁₃L₉ N₉ Q₉ L₉ N₉ G₉ V₉ K₉ G₉ V₉ D₉ D₉ Y₉ E₉ S₉
P₄ R₄ K₄ A₄ E₄ K₄ T₄ K₄ A₄ I₄ L₄ D₄ E₄ R₄ V₄ Q₄ R₄ L₄ E₄ L₅ L₄    N₄ Q₄ N₁ N₄ G₄ V₄ S₄ G₄ I₇ D₄ D₄ Y₄ E₄ S₄ L₄ V₄ S₄ F₄
W₁ I₇ P₁ R₁ K₁ A₁ E₁ S₁ T₁ K₁ A₁ M₁ L₁ D₁ E₁ R₁ V₁ Q₁ S₁    E₁    L₁    Q₁ L₁ N₁ G₁ V₁ K₁ G₁ V₁ D₁ D₁ Y₁ E₁ S₁ L₁ V₁
        1325      1330      1335      1340      1345      1350      1355      1360
L₉ V₉ S₉ F₉ F₉ S₉ E₉ N₉ P₉ H₉ S₉ A₉ E₉ Y₉ L₉ E₉ A₉ Q₉ C₉ A₉ S₉ D₉ Y₉ I₇ E₉ E₉ K₉ V₉ M₉ N₁₃V₉ K₉ K₁₃N₉ Y₉ D₉ K₉ P₉ I₁₀I₇
F₄ S₄ E₄ N₄ P₄ H₄ S₄ A₄ E₄ Y₄ L₄ E₄ A₄ Q₄ C₄ A₄ S₄ D₄ Y₄ I₇ E₄ E₄ K₄ V₄ M₄ N₂ V₄ K₄ R₃ K₁ Y₄ D₄ Y₃ P₄ M₄ I₇ I₇ G₄ L₄ V₄
S₁ F₁ F₁ S₁ E₁ N₁ P₁ H₁ S₁ A₁ E₁ Y₁ L₁ E₁ A₁ Q₁ C₁ A₁ S₁ D₁ Y₁ I₇ E₁ E₁ K₁ S₂ M₁ N₁ S₁    K₁ N₁    D₁ K₁ P₁ V₁ I₇    G₁
                                                   V₁       V₁                                        S₁
        1365      1370      1375      1380      1385      1390      1395      1400
I₇ G₉ L₉ V₁₃G₉ L₉ A₆ V₉ A₅ T₁₀G₉ T₉ F₉ A₉ Y₁₀W₉ Y₉ M₆ S₉ S₁₃N₉ A₉ A₉ S₉ E₉ V₉ V₈ E₈ S₉ Q₉ A₉ K₁₀H₉ K₉ Y₈ N₉ R₉ D₉ K₉ R₉
G₄ L₄ A₄ L₁ A₅ T₄ G₄ T₅ F₄ A₄ Y₄ W₄ Y₅ M₃ R₄ S₄ E₄ A₄ A₄ A₁ E₄ V₄ V₄ E₄ K₄ Q₄ A₄ K₄ H₄ K₅ Y₄ N₄ R₄ D₄ K₄ S₄ T₄ G₄ S₄ L₄
L₁ V₁ G₁    V₁ S₃    T₄    F₁ A₁    L₁    M₁ S₁ L₃ N₁    A₁ S₁ E₁ V₁ V₁ E₁ I₇ G₁ A₁    H₁    Y₁ N₁ F₁ D₁ K₁ R₁ T₁ G₁
             T₁    G₁          W₁          S₁                S₁ Q₁                R₁
        1405      1410      1415      1420      1425      1430      1435      1440
T₉ G₉ S₉ L₉ M₉ Y₉ D₁₀L₉ D₉ D₁₀Q₉ G₆ T₉ V₉ E₉ T₉ F₉ G₉ V₉ E₉ Y₉ S₉ D₉ A₉ V₉ M₁₂T₉ G₉ K₉ M₉ S₉ K₉ A₉ Q₉ K₉ E₁₀R₉ E₉ S₉ R₉
M₄ F₃ D₄ Y₅ S₃ D₄ Q₄ D₅ T₄ V₄ E₄ T₄ F₄ G₄ V₄ E₄ Y₄ S₄ D₄ A₄ V₄ I₇ T₄ G₄ K₄ G₁ S₄ K₄ A₄ Q₄ K₄ E₄ R₄ E₅ S₄ R₄ K₄ K₄ G₄ W₄
S₁ L₁ M₁    D₁ L₁       Q₁    T₁ S₃ E₁ T₁ F₁ G₁ V₁ E₁ Y₁ S₁ D₁ A₁ V₁ M₁ T₁ I₇ K₁ M₁ S₁ K₁ A₁ Q₁ K₁    R₁    S₁ R₁ K₁ K₁
   Y₁       N₁          V₁
        1445      1450      1455      1460      1465      1470      1475      1480
K₁₃K₉ G₁₃W₉ K₉ V₉ G₉ K₉ V₉ N₉ R₉ P₉ M₉ R₉ V₉ F₉ R₉ Q₉ L₉ Y₉ G₉ V₉ N₉ P₉ L₉ E₉ F₉ D₉ E₉ V₁₃V₉ M₉ R₉ V₉ G₉ E₉ L₉ T₉ S₁₀E₉
G₁ I₇ K₁ K₄ V₄ N₄ R₄ P₄ M₄ R₄ V₄ F₄ H₄ Q₄ L₄ Y₄ G₄ V₄ N₄ P₄ L₄ E₄ F₄ D₄ E₄ V₄ V₃ M₄ R₄ M₁ G₄ D₄ W₄ A₄ T₄ D₄ P₄ W₄ T₃ A₄
   W₁    V₁ G₁ K₁ V₁ D₁ R₁ P₁ M₁ R₁ V₁ F₁ R₁ Q₁ L₁ Y₁ G₁ V₁ N₁ P₁ L₁ E₁ F₁ D₁ I₇ V₁ V₁    R₁ V₁ G₁ E₁ L₁ T₁ S₁ E₁ P₁ W₁
                                                                    E₁
        1485      1490      1495      1500      1505      1510      1515      1520
P₉ W₉ S₉ A₉ K₉ D₁₄V₉ D₁₀I₇ D₉ N₉ M₉ L₉ V₇ E₉ L₉ D₁₀D₉ D₉ F₇ H₉ I₇ L₉ S₈ G₉ D₉ S₉ M₉ L₉ G₉ K₉ K₉ V₉ E₉ L₉ A₉ F₇ T₉ K₉ D₁₂
K₄ D₃ V₄ N₄ V₄    G₄ M₄ L₄ I₇ E₄ L₃ D₄ D₃ D₄ Y₄ H₄ I₇ L₃ K₄ D₄ D₄ R₄ M₄ L₃ G₃ K₉ K₁ V₄ E₄ L₃ A₄ F₄ T₄ K₃ S₁ G₄ S₄ G₂ S₁
S₁ A₁ K₁ D₁       I₇    N₁ M₁ L₁ V₁ E₁ I₇    D₁    Y₁ H₁ Y₂ L₁ S₁ G₁ K₁ F₁ M₁ L₁ G₁ K₁ K₁ V₁ E₁ L₁ A₁ F₁ T₁ K₁ D₁ S₂ N₁
                        L₁             I₇          D₁ S₁    E₁                   N₁
                                                           S₁
        1525      1530      1535      1540      1545      1550      1555      1560
G₉ S₉ N₉ E₉ E₉ T₉ V₉ V₉ K₉ L₉ T₉ P₉ H₉ R₉ S₁₃K₉ M₁₃A₉ S₉ S₁ M₉ S₁₀L₉ N₅ P₁ M₉ G₉ F₉ P₉ E₁₀E₉ E₉ G₉ R₉ W₉ S₁₂Q₉ T₉ G₉ S₉
E₄ T₄ V₄ V₃ K₂ L₄ T₃ P₄ H₄ R₄ S₄ S₃ M₄ A₄ M₁ S₄ S₁ S₅ L₄    P₄ M₄ G₃ F₄ G₁ E₄ E₄ E₅ G₄ R₄ W₄ S₄ Q₄ T₄ G₄ N₃ P₄ V₄ V₄ Q₃
S₁ E₁ E₁ T₁ Q₁ V₁ K₁ L₁ T₁ P₁ H₁ R₁ S₁ K₁    A₁    M₁    L₁    P₁ S₄    F₁ P₁    E₁    G₁ R₁ W₁ S₁ Q₁ T₁ G₁ S₁ P₁ R₁
      V₁                                                 M₁                                                V₁
```

```
        1565      1570      1575      1580      1585      1590      1595      1600
P9 V6 V9 Q9 K7 S8 S8 E8 E9 N7 Q9 S9 V12 E9 K13 Q5 V9 A9 K9 P9 E9 T9 T9 N9 P9 Y9 E9 H5 V10 L9 V9 R9 L9 G9 S9 A9 H5 L9 G9 T8
K4 S3 F2 E4 D3 G3 Q3 T4 V4 E4 M3 Q4 I1 A4 V1 P4 E4 T2 T4 N4 P4 Y4 E4 H4 V4 L4 V5 R4 L4 G4 S4 A4 H4 L4 G4 T4 S4 V4 L4 N4
V1 M2 K1 S1 S2 K1 E1 K1 Q1 S2 T1 E1 K1 Q1    A1 K1 V2 E1 T1 T1 N5 P1 Y1 E1 H1    L1    R1 L1 G1 S1 A1 H1 L1 G1 T1 S1 S1
   K1 T1    E2 D1 H1 N1    D1 V1              P1                                                                        V1
   L1 S1       E1 L1
   Q1
        1605      1610      1615      1620      1625      1630      1635      1640
S9 V9 L9 N9 C5 F9 F9 H5 G9 S9 K5 C5 V9 M9 P9 Y9 H5 L5 A9 E9 N10 G9 D9 A9 N9 E9 S13 L9 V9 M9 S9 T9 T9 R9 G9 Q9 F10 D9 F9 G9
C4 F4 F4 H4 G4 S4 K4 C4 V4 I4 P4 Y4 H4 L4 A4 E4 N4 G4 D4 L4 D3 K2 S3 L4 I1 I4 V1 T4 T4 R4 G4 Q4 F4 D4 F5 G4 P4 L4 K4 N4
L1 N1 C1 F1 F1 H1 G1 S1 K1 C1 V1 M1 P1 Y1 H1 L1 A1 E1 N1 G1 S1 E2 N1 E1 S1 L1    M1 S1 T1 T1 R1 G1 Q1    D1    G1 P1 L1
                                        A1 P1    V1
        1645      1650      1655      1660      1665      1670      1675      1680
P9 L9 K9 N9 I1 K10 C9 K9 K10 M9 K9 D10 Y9 D9 M9 T9 V9 C5 P10 L9 P9 R9 D9 V9 Q9 P9 F9 R9 A9 K9 I5 V9 F9 R9 E9 P9 K9 L9 G9 E9
I4 K4 C4 K5 K4 I4 K5 D4 Y4 D5 I5 T4 V3 C4 P4 L4 P5 S4 D4 V4 Q4 P4 F4 S4 A4 K4 M4 V4 F3 S4 E4 P4 K4 L4 G4 E4 E4 V4 V4 V5
K1 N1 I1    C1    M1       Y1    I1 T1 V1 C1    L1    R1 D1 V1 Q1 P1 F1 R1 A1 K1 I1 V1 F1 R1 E1 P1 K1 L1 G1 E1 E1
                 M1                                                            V1
        1685      1690      1695      1700      1705      1710      1715      1720
E9 V10 V10 V7 V9 Y9 F9 A9 S12 I1 D6 G9 S7 I1 V8 M9 K12 V7 S4 E9 K9 S13 N9 T9 Y9 Q9 A9 G9 G9 S9 F9 T12 H9 L9 W9 T9 Y9 Q9 H7 D9
V5 C2 F4 A4 S4 I1 D4 G4 R1 G1 V4 M4 K6 V3 S4 E4 S2 S4 N4 T4 Y4 T1 A4 G4 G4 L3 F4 T4 H4 L4 W4 A1 Y4 Q4 Y3 D4 G4 N4 P4 G4
   F1    I1 F1 A1 S1 I1 N1    N3 I1 V1 M1 I1 A1    A2 K1 S1 N1    Y1 Q1 A1 G1 G1 S1 F1 K1 H1 L1 W1 T1 H2 Q1 H1 D1 F2 N1
   Y1    Y1                   S1    A1 K1    E1                S1                                                    G1
        1725      1730      1735      1740      1745      1750      1755      1760
G9 N9 P9 G13 D9 C9 G9 G9 P9 I1 V9 A9 T9 S9 D9 M9 K9 V9 V9 G13 F9 H5 S13 G9 V9 V9 S9 N9 G10 A9 G9 E9 K9 L9 R9 A9 V13 Y9 T9 P9
D4 C4 G4 C1 P4 I4 V4 A4 T4 S4 D4 L4 K4 V4 V3 G4 F4 H4 S4 H1 V4 V3 V1 N4 G3 A3 G5 E4 K4 L4 R4 A4 V4 Y4 T4 P4 T1 N4 Q4 E4
P1 G1 D1    G1 G1 P1 I1 V1 A1 T1 S1 D1 M1 I1 V1 V1 G1 F1    S1 G1    V1 S2 E1    A1    E1 K1 L1 R1 A1 V1 Y1    P1 V1 N1
                                        K1          A1          N1
        1765      1770      1775      1780      1785      1790      1795      1800
V9 N8 H9 E9 L13 I9 A7 S10 L9 S9 T9 E9 I1 Q9 M9 T13 D9 F9 W9 T9 F9 N9 P9 D8 L9 V9 E9 W9 N9 S9 V9 A9 R9 V9 S9 T9 F5 F9 P9 M10
L4 I1 S4 S4 V1 S5 S4 E4 V4 Q4 M4 N2 D4 F4 W4 F1 F4 N4 P4 D2 L4 V4 E4 W4 N4 S4 V4 A4 R4 V4 S4 T4 F4 F4 P4 M4 Y4 K4 A3 I1
H1 E1 L1 I1    V2    T1 D1 V1 T1 M1 T1 D1    W1 T1 F1 K1 P1 D1 L1 N1 E1 W1 N1 S1 V1 A1 R1 V1 S1 T1 F1 F1 S4 M1 T1 K1
   K1       L1          Q1                   N1       V1                                                 P1    S1
                       D1                   A1
        1805      1810      1815      1820      1825      1830      1835      1840
S9 K9 A9 M9 N9 A8 I5 T9 V10 Q9 V9 N8 D9 G9 E9 E9 S7 I5 D9 G9 N9 L9 M9 M9 V9 G9 S8 V13 N10 S9 N13 V9 Y9 H15 N9 H5 V9 M9 K15 G9
N4 T4 M4 T4 V4 Q4 V5 D3 D4 G4 E4 E4 M4 I4 D4 G4 N4 L4 M4 M4 V4 G4 S4 V4 N4 S4 T5 S1 Y4 H4 Y1 H5 V4 M4 K4 G4 K5 S4 E3 S5
A1 M1 N1 A1 I1 S2    Q1    N1 D1 G1 E1 E1 L1 I1 L2 G1 N1 L1 M1 M1 V1 G1 S1 V1 N1    V1    N1    V1 M1    G1
                 N1       D1          D1
        1845      1850      1855      1860      1865      1870      1875      1880
K9 S10 E8 S9 F9 V9 R8 Y9 C5 E9 Q9 F9 P9 K9 S9 A9 F9 T9 P9 E9 L9 S9 D8 Q9 Y9 L13 P9 S9 I5 L9 S9 K9 P9 A9 F9 S9 K15 G9 L9 L9
F4 V4 R4 Y4 C4 E3 Q5 F4 P4 H5 S4 A4 F4 V4 P4 E4 L4 S4 D2 K3 Y4 L4 P4 S4 I5 S1 S4 K4 P4 A4 F4 R4 K4 G4 L4 L4 L1 Y4 N4 E4
E1    F1 V1 R1 Y1 C1 E1 Q1 F1 P1 K1 S1 A1 F1 M1 P1 E1 N5 S3 D1 Q2 G1 L1 P1    I1 L1 S1 K1 P1 A1 F1 S1 K1 G1    L1 K1 Y1
   Q1    N1             N1                   L1          Y1 V1
        1885      1890      1895      1900      1905      1910      1915      1920
K9 Y9 N9 E9 P9 V10 R9 V9 G9 L7 V9 N9 F9 E9 C9 L5 V9 K5 A9 Y5 L9 V9 V9 E13 K9 K9 L9 E9 E9 L9 G8 F9 M9 G9 N9 S9 G9 P13 Q5 W9
P4 V4 K4 V5 G4 F4 V4 N4 Y4 E4 C4 L4 I4 K2 A4 Y4 L4 V4 I4 E4 K3 K4 L4 K1 E3 L4 E4 F4 R2 G4 N4 S4 G4 P4 E4 W4 D4 W1 M4 E4
N1 E1 P1    R1    G1 L1 V1 S2 F1 E1 C1 S2 V1 K1 M5 S4 L5 V1 S1 E1 K1    D1 E1 G1 L1 H1 F1 E1 G1 N1 S1 G1 P1 Q1    D1 P1
                  N1          L1       A1 Y1       V1          L1          G1    V1
                                                                          C1
```

1925	1930	1935	1940	1945	1950	1955	1960

1965	1970	1975	1980	1985	1990	1995	2000

2005	2010	2015	2020	2025	2030	2035	2040

2045	2050	2055	2060	2065	2070	2075	2080

2085	2090	2095	2100	2105	2110	2115	2120

2125	2130	2135	2140	2145	2150	2155	2160

2165	2170	2175	2180	2185	2190	2195	2200

2205	2210	2215	2220	2225	2230	2235	2240

2245	2250	2255	2260	2265	2270	2275	2280

```
        2285        2290        2295        2300        2305        2310        2315        2320
```

```
        2325        2330        2335        2340        2345        2350        2355        2360
```

```
        2365        2370        2375        2380        2385        2390        2395        2400
```

```
        2405        2410        2415        2420        2425        2430        2435        2440
```

```
        2445        2450        2455        2460        2465        2470        2475        2480
```

```
        2485        2490        2495        2500        2505        2510        2515        2520
```

```
        2525        2530        2535        2540        2545        2550        2555        2560
```

```
        2565        2570        2575        2580        2585        2590        2595        2600
```

```
        2605        2610        2615        2620        2625        2630        2635        2640
```

Supplemental Figure 2. UCBSV AA variability determined using datamonkey.org.

525 530 535 540 545 550 555 560

565 570 575 580 585 590 595 600

605 610 615 620 625 630 635 640

645 650 655 660 665 670 675 680

685 690 695 700 705 710 715 720

725 730 735 740 745 750 755 760

765 770 775 780 785 790 795 800

805 810 815 820 825 830 835 840

845 850 855 860 865 870 875 880

885 890 895 900 905 910 915 920

925 930 935 940 945 950 955 960

965 970 975 980 985 990 995 1000

1005 1010 1015 1020 1025 1030 1035 1040

1045 1050 1055 1060 1065 1070 1075 1080

1085 1090 1095 1100 1105 1110 1115 1120

48

```
        1125      1130      1135      1140      1145      1150      1155      1160
E14A14N14Y14M14F14T14A13H14L14V14D14K14Q14G14F14M14P14S14P14V14F14E14L14M14K14N7L14L14L14H14T14D14A14V14G14V13S14S14T14
          T1                                                      S7                                    I?

        1165      1170      1175      1180      1185      1190      1195      1200
Y14L14A14T14N14M14S14G14W14S13S14L14K14E14Y14I?S12I?D14D14S10S14R14H14V14Q14E14V13Q14I?P14W14Y14C?S14D14M14S14D14D14
              K1              V1K2          N4                  I1

        1205      1210      1215      1220      1225      1230      1235      1240
F14I?V12K14L14A14E14C14V14K14A14A14K14P14K14S14Q4C?G14Y14K14V14D14N14V14D14F12H14T14V14A14H14K14M14S?V13G14E14S14N14
     I?                                                              L2                        M1

        1245      1250      1255      1260      1265      1270      1275      1280
M14D14E14S14S13A14L14V14A13T14I?L14D14E14V14K14Q?W14S14D14G14I?T14Y14H14S14S14T14P14S14N14K14S14L?M14S14L?M14V14G14
          R1            S1

        1285      1290      1295      1300      1305      1310      1315      1320
W14I?P14S14K14A14E14K14T14K14E14I?L14D14N14R12I?Q4R14L?E14L14L14L14N4Q14L14N14G14V14S14G14I?D14D14Y?E14S14L14V14
                                S2V2                                              V1

        1325      1330      1335      1340      1345      1350      1355      1360
R14F14F14S14E14N14P14H14S14A14E14Y14L?E14S14Q4C?A14S?D14Y14I?E14E14K14V14M14S12V14K14S14N12Y?D14K14S12L13I?L?G14
                                              N2          S2        P2M1M1

        1365      1370      1375      1380      1385      1390      1395      1400
M13V14G14L14A14V14A14T14G14T14F14A14Y14W14Y14M14S14S14S?A14A14V13E14L14V14E14K14Q4A14K14H?K14Y14N14S14D14K14R12T14G14
V1                                        D1        M1                                        S2

        1405      1410      1415      1420      1425      1430      1435      1440
S14L14M12F14D12M14D14D14Q13E14T14Y14E14N12F14G14P10E14Y14T14D14D14V14I?S14A13K14M14T13K14A14Q4K14E14S14D12S14K13K14K14
     L2   E2        S1        S2        A2              M3   G1        A1                    E2   S1
                                        S2

        1445      1450      1455      1460      1465      1470      1475      1480
G14W14K14A14G14K14M10N14R14P14M14S14V13F14H?Q14L14Y14G14V14N14P14L?E14F14D14E14V14V12M14S12V14G14K14L?E14T?E14P14W14
          V4            I1                              I?R2          A1

        1485      1490      1495      1500      1505      1510      1515      1520
D14V14K14E14L14N14V13D13A12M14M12I?E14L?D14D14D14Y14H14I?L14R14D14D14R12M14F13G14K14K12V14S14L?A14F14K13K14E14G14A12
          I1G1G2   L2              S2   Y1        T2              S3        S2

        1525      1530      1535      1540      1545      1550      1555      1560
D10E14E14T14I?V14N13L14T14P14H?R14S14K14M14A8 S14S14M14S14L14A14P14M14G14F14P14E14E14E14G14S14W14S14Q4T?G14A14P14L10
H2              S1              T6                                                                        V3
G1                                                                                                       M1
N1

        1565      1570      1575      1580      1585      1590      1595      1600
M11K10K14I?K8 K10E14D12E14V12E14V14Q?V14A14K14P?E14P12T14N14P14Y14D14D14H14I?L14V14S14L?G14S14A14H?L?G14T14R14V14L14
V2 S3   T1E4N2   E2  I?              L1          V2              R1
T1 Q1   S2T2                        S1

        1605      1610      1615      1620      1625      1630      1635      1640
N14C?F14F14H?G14S12K14C?V13I?P14Y14H14L?A14E14K14G14D14S14E14E14S12L14V10I?A14T14T14S?G14Q?F14D14F14G14P?M13K13
            A1       I1                              P2   I4                                      L1E1
            T1

        1645      1650      1655      1660      1665      1670      1675      1680
N14I?K14C14S12K14I?T14D14Y14D14I?T14I?C?P13L14P14N14D14V14Q14P14F14R14S13K14I?V14M14R14E14P14K14L14G14E14E14V14V14
          K2   V7         V3         S1                        A1
                          M1

        1685      1690      1695      1700      1705      1710      1715      1720
I?V14C12F14T12S14I?N14G14K14I?V14M14K14V13S14D14K14S14T12T14Y14P14A14G14G14Q4F14A14H14L?W14A14Y14K11Y13D14G14Q4P14
     Y2   A2              A1       A2                                              N3

        1725      1730      1735      1740      1745      1750      1755      1760
G14D14C?G14G14P14I?V14A14T14V14D14Q4K14V14V14G14F14H?S14G14V14I?S14N14S14S14E11E14K14L?R14A14V14Y14T14P?V14N4Q4
                                                          G3
```

```
            1765       1770       1775       1780       1785       1790       1795       1800
E₁₄L₁₄L₁₂N₁₄C₁₂I₁₀S₁₄G₁₂D₁₄I₁₀Q₁₄M₁₄T₁₄D₁₄F₁₄W₁₄T₁₄F₁₄N₁₄P₁₀D₁₄L₁₄V₁₄E₁₄W₁₄N₁₄S₁₄V₁₄A₁₄S₁₄V₁₄S₁₄T₁₄F₁₄F₁₄P₁M₁₄T₁₄K₁₄A₁₄
      M₂   Y₂ V₄    S₂

            1805       1810       1815       1820       1825       1830       1835       1840
I₁₀N₁₄T₁₄I₁₂T₁₄V₁₄Q₁₄A₁₄N₁₁E₁₄G₁₄E₁₄E₁₄L₁₃I₁₀D₁₄G₁₄N₁₄L₁₄M₁₄I₁₀V₁₄G₁₄Y₁₄V₁₄N₁₄R₁₄E₁₄V₁₄Y₁₄H₁₀N₁₄H₁₀V₁₃I₁₀K₁₄G₁₄K₁₂S₁₄E₁₄
      M₂         D₃         S₁              M₂                              I₁         S₂

            1845       1850       1855       1860       1865       1870       1875       1880
S₁₄F₁₄M₁₀S₁₃Y₁₄C₁₄E₁₄Q₁₄F₁₄P₁₄N₁₂C₁₄A₁₄F₁₄T₁₄K₁₄E₁₂L₁₄R₁₂D₁₄Q₁₃Y₁₄L₁₄P₁₄S₁₄I₁₀L₁₄S₁₄K₁₄P₁₄A₁₄F₁₄S₁₄K₁₄G₁₄L₁₄L₁₄K₁₄Y₁₄N₁₄
      T₄ K₁              S₂              D₂   S₂  K₁              V₃

            1885       1890       1895       1900       1905       1910       1915       1920
E₁₄P₁₄V₁₄R₁₃V₁₄G₁₄S₁₀V₁₁N₁₀F₁₄P₉C₁₄L₁₄M₁₂R₁₄A₁₄Y₁₄L₁₄K₁₂V₁₄E₁₄E₈M₁₄F₁₄E₁₃N₆L₁₀G₁₄F₁₄L₁₂K₇E₁₄A₁₄G₁₄P₁₄Q₁₄W₁₄D₁₄P₁₄I₁₀
         Q₁       L₄M₂D₄  S₅     I₂         N₂      G₄      K₁D₄      R₂E₅
                  L₁                               T₂      T₂        G₂
                                                           A₂

            1925       1930       1935       1940       1945       1950       1955       1960
E₁₄I₁₀L₁₄E₉D₁₄L₁₄N₁₄K₁₄K₁₄A₁₄A₁₄M₁₄G₁₄A₁₄L₁₄Y₁₄Q₅G₁₄K₁₄K₁₄Q₁₀D₁₄W₁₄L₁₃K₁₄S₁₄M₁₂E₁₄P₁₄A₁₃D₁₄F₁₄I₁₀T₁₀A₁₄V₁₄R₁₄E₁₄S₁₀F₁₄
      D₅                                              F₁      I₂  V₁      M₄

            1965       1970       1975       1980       1985       1990       1995       2000
K₁₄H₁₄L₁₄A₁₄G₁₄G₁₄D₁₄V₁₄G₁₄I₁₀W₁₄S₁₀G₁₄S₁₄L₁₄K₁₄A₁₄E₁₄L₁₄S₁₄P₁₂V₁₂E₁₀K₁₄V₁₄L₁₄E₁₄Q₁₄K₁₄T₁₄S₁₂V₁₄F₁₄T₁₆G₁₄A₁₄P₁₄I₁₀D₁₄L₁₄
                                            T₁ I₁ K₄                    R₂
                                            S₁

            2005       2010       2015       2020       2025       2030       2035       2040
L₁₄L₁₄G₁₄G₁₄K₁₄I₁₂L₁₄V₁₄D₁₄N₁₄F₁₄N₁₄H₁₀F₁₄F₁₄Y₁₄F₁₄N₁₄H₁₀L₁₄K₁₄G₁₄P₁₄W₁₄T₁₄V₁₄G₁₄I₁₀N₁₄K₁₄F₁₄N₁₄K₁₀G₁₄W₁₄D₁₄S₁₄L₁₄A₁₄S₁₂
          M₂                                                                             K₂

            2045       2050       2055       2060       2065       2070       2075       2080
Y₁₃F₁₄N₁₂H₁₆S₁₂W₁₄N₁₄F₁₄M₁₃D₁₄C₁₀D₁₄G₁₄S₁₄S₁₄F₁₄D₁₄T₁₄S₁₄L₁₄A₁₄P₁₄I₁₀L₁₄F₁₄Q₁₄L₁₄V₁₄C₁₄H₁₀M₁₄R₁₄E₁₄K₁₄F₁₆G₁₄N₄F₁₄D₁₄D₁₄
N₁   D₂   N₂        T₁

            2085       2090       2095       2100       2105       2110       2115       2120
I₁₀E₁₄S₁₂A₁₂A₁₄L₁₄R₁₄N₁₄L₁₄Y₁₄T₁₄Q₁M₁₄V₁₄Y₁₄T₁₄P₁₄I₁₀L₁₄T₁₄I₁₀D₁₄G₁₄Y₁₄I₁₀T₈K₁₄K₁₄H₁₄R₁₄G₁₄N₁₄N₁₄S₁₄G₁₄Q₁₄P₁₄S₁₄T₁₄V₁₄
V₁   K₂T₂                                         A₄
                                                 M₁
                                                 V₁

            2125       2130       2135       2140       2145       2150       2155       2160
V₁₄D₁₄N₁₄T₁₄I₁₀M₁₁L₁₄M₁₄I₁₀V₁₄V₁₄E₁₄Y₁₄C₁₀K₁₄A₁₄V₁₄M₁₄E₁₄S₁₂E₁₄G₁₄S₁₃T₄M₁₄Q₁₄F₁₄K₁₄Y₁₄M₁₄C₁₀N₁₄G₁₄D₁₄D₁₄L₁₄I₁₀L₁₀N₁₄V₁₄
          M₂ I₂                         T₂      K₁V₄                          M₂
                                               I₂
                                               E₂
                                               A₂

            2165       2170       2175       2180       2185       2190       2195       2200
P₁₄D₁₂D₁₄E₁₄V₁₄S₁₂M₁₄I₁₀Q₁₄S₁₄S₁₄F₁₄S₁₄E₁₁L₁₄F₁₄S₁₀E₁₄C₁₀G₁₄L₁₄D₁₄Y₁₄N₁₃F₁₄D₁₄D₁₄V₁₄H₁₄K₁₄S₁₄M₁₄E₁₄T₁₄I₁₀E₁₄Y₁₀M₁₄S₁₄H₁₄
      N₂         N₂   V₁         D₂                        T₁
                                N₁

            2205       2210       2215       2220       2225       2230       2235       2240
S₁₄F₁₄M₁₄L₁₃K₁₄D₁₄G₁₀V₁₂Y₁₄I₁₀P₁₄K₁₄L₁₄K₁₄K₁₂E₁₄S₁₄I₁₀V₁₄A₁₄I₁₀L₁₃E₁₄W₁₄E₁₄S₁₄G₁₄D₁₄E₁₄I₁₀M₁₄R₁₂T₁₄R₁₂S₁₄A₁₄L₁₀N₁₄A₁₄A₁₄
      M₁     D₄L₁              S₂   V₁                        M₄  S₂  S₂
             M₁                     M₁

            2245       2250       2255       2260       2265       2270       2275       2280
Y₁₄I₁₀E₁₄S₁₄Y₁₄G₁₄Y₁₄E₈D₁₄L₁₃M₁₄V₁₁E₁₄I₁₀E₁₄R₁₄Y₁₄A₁₄V₁₄F₁₄W₁₄A₁₄T₈E₁₃K₁₄G₁₄C₁₄E₁₄Y₁₄P₁₄L₁₄L₁₀D₁₄S₁₄K₁₄R₁₄V₁₄E₁₄G₁₄L₁₄
               D₆  P₁   I₁                  A₆G₁
                       T₁
                       M₁
```

```
            2285        2290        2295        2300        2305        2310        2315        2320
Y14 K10 D14 D14 H11 T14 D14 I13 N16 E14 E14 W14 L14 M6 G14 I13 L15 P14 P14 S14 F14 E14 H14 C14 Y14 V14 D14 T10 Q14 T15 K12 D14 L12 S13 G8 S9 E12 K10 P12 E14
    S4      Y2                          V3              M4    I1 E2      V2 K1 E4 G2 D2 Q2 L2
            N1                          I2                                  S2 E2    E2
                                        S2                                  K1
                                        T1

            2325        2330        2335        2340        2345        2350        2355        2360
L13 S13 I3 E12 S13 H1 D13 G11 V6 P5 Q1 M13 Q2 M14 K14 F14 P14 V14 T14 F14 V14 T14 G14 N14 L13 G14 K14 L14 A14 E14 V14 K13 S14 I3 L14 G14 I3 A13 N11 D14
F1 G1 V3 G2 N1 C2 N1 S3 T6 S5 R3 L1                      F2              S1              S1 S3
    T3       Y2          A1 L2
    P2       M1 T2
    A1

            2365        2370        2375        2380        2385        2390        2395        2400
V14 M9 A14 K8 N14 I13 D14 L14 P14 E14 V14 Q14 G14 T14 P14 D11 E14 I13 V13 S13 K14 K14 A14 Q14 L13 A14 V14 K14 M14 T14 N14 S12 P14 V14 L14 V14 E14 D14 T13 C14
    V5       S3                          E2    V1 A1 K2      F1          G2                      A1
             R1                          V1          M1

            2405        2410        2415        2420        2425        2430        2435        2440
L14 C14 F14 N14 A14 F14 N14 G14 L13 P13 G14 P14 Y14 I13 K14 W14 F14 L14 K14 K14 E14 L14 G14 L14 E8 G14 V14 V14 K12 M14 L14 S13 A14 F14 G8 D14 K14 S14 A14 Y14 A14
                                        M1                              D6          S2          E4
                                                                                                S2

            2445        2450        2455        2460        2465        2470        2475        2480
L14 C14 T14 F14 A14 Y14 V13 H14 N14 E14 L7 S14 D13 P14 I13 V14 F14 K14 G14 G14 V14 V14 N13 G14 E14 I13 V14 P14 P14 R14 G14 N12 N14 G14 F14 G14 W14 D14 P14 M9 F14
             A1      S1 S7      E1 V3          K1                  K2                                  I5

            2485        2490        2495        2500        2505        2510        2515        2520
K14 P14 D14 G9 C1 G7 C14 T14 F14 A14 E14 M14 P9 S14 S11 I13 K14 N14 D13 F14 S14 H14 S14 S14 S14 A14 L14 E14 K14 V14 V14 K14 L14 F10 L14 D14 N14 L14 V10 V14 K12
    E5       S6          S5    N2 M3      E1              K1                  S3 Y4              M4 S2
    E1                         G1 N1

            2525        2530        2535        2540        2545        2550        2555        2560
Q14 E14 E12 K12 K14 A13 K7 V13 A14 L14 T14 I13 D12 V14 Q3 A13 L14 N12 Q14 E14 E14 M9 E14 A12 E14 I13 T12 A14 L14 K14 K14 L13 W14 K9 D14 N14 G14 P14 T14 S13
    K2 Q1 S3 T1 S6 M1          E2          V1    D2          V4    E1          A1              M1 S5                  G1
    N1          G1                                          I1    T1          S1

            2565        2570        2575        2580        2585        2590        2595        2600
T14 R14 S14 P13 F14 E14 A14 S13 S14 L14 S14 A14 P12 Q2 V14 E14 R14 V14 N14 E14 L14 L14 Q14 K12 L14 K14 D13 E14 G14 L14 Q14 T14 K14 K14 S14 P14 C14 G14 E14 P14
    S2              K1              S2                                              Q2          N1

            2605        2610        2615        2620        2625        2630        2635        2640
D14 D14 G14 E14 V13 V13 D13 D14 D14 S14 D12 D13 G12 N14 N14 Q14 S14 S13 G12 K14 E14 V14 V8 D12 E14 S11 Q14 N14 N10 Q12 Q3 P12 S14 K14 P14 K14 F13 K14 I13 S14
    M1 M1 V1          N2 N1 D1 D3 K1          F1 E2    K2    I1 H1    N2    D1 D4 K2 H1 L1 K2 S1 S2 S1 L1       V2
             E1          D1                            N1    G1                      S1          L1
                                                                                                Q1

            2645        2650        2655        2660        2665        2670        2675        2680
G14 D14 G13 S14 A10 I3 S12 R14 D14 D14 I13 D14 K13 I13 P14 T14 N13 A14 L14 E14 I13 K14 K14 T14 F14 K14 P14 P14 K10 V14 S14 Q14 S14 A14 Y14 I13 W14 I13 P14 R14
    E1 N3 T4 V5 K2              S1          S1                              S4

            2685        2690        2695        2700        2705        2710        2715        2720
S14 Q14 S14 D14 N14 L14 T14 P14 D14 V14 I13 Q14 N13 F14 L13 A14 Y14 M10 P14 P14 S14 H13 A14 M14 D14 N14 Q14 L14 A14 S14 G14 I13 E14 V14 E14 N14 W14 A14 I13 E14
                        M1    S1              V4          Y1                              V1                      V2
                                                                                                                M1

            2725        2730        2735        2740        2745        2750        2755        2760
V14 A14 K14 A14 Y14 G14 V14 N14 M14 Q14 E14 F14 Y14 R13 T14 V10 L14 P14 A14 W14 M10 V14 N14 C14 M14 V14 N14 G14 T14 S14 D14 E14 S14 K14 N14 E14 K14 S14 W14 R14
                                        H1          I4          I1

            2765        2770        2775        2780        2785        2790        2795        2800
A14 V14 E14 L14 N14 S14 Q14 G14 E14 D14 V14 D14 D14 F14 E14 Y14 P14 M14 E14 P14 M14 Y14 K13 F14 A14 L14 P14 T14 M14 S14 K14 V11 M14 S14 N14 F14 S14 S14 Q14 A14
                                                                        S1                          I3
```