Characterization by Next Generation Sequencing Reveals the Molecular Mechanisms Driving the Faster Evolutionary rate of *Cassava brown streak virus* Compared with *Ugandan cassava brown streak virus*

Titus Alicai[1], Joseph Ndunguru[2], Peter Sseruwagi[2], Fred Tairo[2], Geoffrey Okao-Okuja[1], Resty Nanvubya[1], Lilliane Kiiza[1], Laura Kubatko[3], Monica A. Kehoe[4] and Laura M. Boykin[5,*]

[1]National Crops Resources Research Institute, P.O. Box 7084, Kampala, Uganda

[2]Mikocheni Agricultural Research Institute, Coca cola Road, Box 6226, Dar es Salaam, Tanzania

[3]The Ohio State University, 154W 12th Avenue, Columbus, Ohio 43210, USA

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

[5] The University of Western Australia, ARC Centre of Excellence in Plant Energy Biology and School of Chemistry and Biochemistry, Crawley, Perth 6009, Western Australia, Australia.

* Corresponding author: laura.boykin@uwa.edu.au

Keywords: cassava, Uganda, *Manihot esculenta,* smallholder farmer, *Cassava brown streak virus*, species tree estimation, SVD Quartets, nonsynonomous mutations

Abstract

Cassava is a major staple food for 800 million people. Cassava brown streak disease (CBSD), is caused by *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) is suppressing cassava yields in East Africa at an alarming rate. Previous studies have documented CBSV is more devastating than UCBSV. This is because CBSV is harder to breed resistance for, causes more infections and yield losses in cassava, and its species delimitation is more challenging. We set out to characterize the CBSV and UCBSV whole genomes from the 26 previously published genomes and three new from Uganda, using NGS data with the goal of uncovering genetic patterns that explain the observed biological differences. In this paper, we report phylogenetic relationships, rates of synonymous and non-synonymous substitutions, and whole genome-based evolutionary rates for CBSV and UCBSV. Using the whole genome sequences we produced the first coalescent based species tree estimation for CBSV and UCBSV which supports previously published studies pointing to multiple species of both CBSV and UCBSV. This new species framework led to the finding that CBSV has a faster rate of evolution when compared with UCBSV. The genes responsible for CBSV's rapid rate of evolution are NIa, 6K2, NIb and P1. Furthermore, we have discovered that for CBSV, rates of nonsynonomous substitutions are more predominant than synonymous substitution and occur across the entire genome. All comparative analyses between CBSV and UCBSV presented suggests CBSV is outsmarting the cassava immune system, thus is more devastating and harder to control.

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**

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

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 Nla (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.

176

177 **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

191 **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

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

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.

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

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

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

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

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°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

420 **CBSV and UCBSV Detection by RT-PCR**

421 All samples were tested for presence of CBSV and UCBSV by a two-step RT-PCR assay [43]. The
422 PCR mixture consisted of 16.0 µl nuclease free water, 2.5 µl PCR buffer, 2.5 µl $MgCl_2$ (2.5 mM),
423 0.5µl dNTPs (10 mM), 1.0 µl of each primer (10mM) [forward CBSDDF2 5'-
424 GCTMGAAATGCYGGRTAYACAA-3' and reverse CBSDDR 5'-GGATATGGAGAAAGRKCTCC-3'], 0.5
425 µl Taq DNA polymerase and 1.0 µl of cDNA. The PCR thermo profile consisted of: 94°C for 2 min
426 followed by 35cycles of 94°C (30 s), 51°C (30 s) and 72°C (30 s) for denaturation, annealing and
427 extension, respectively. PCR products were analysed by electrophoresis in a x1 TAE buffer on a
428 1.2% agarose gel, stained with ethidium bromide, visualized under UV light and photographed
429 using a digital camera.

430

431 **Sample Selection for Sequencing**

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

438

439 **Generation of the Transcriptomes**

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

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

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

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

Acknowledgments

This work was supported by the Bill and Melinda Gates Foundation Grant no. 51466 "Regional Cassava Virus Diseases Diagnostic Project" awarded to Mikocheni Agricultural Research Institute, Tanzania, and sub-grant to National Agricultural Research Organisation (Uganda). Computational resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia supported this work.

Conflict of Interest

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

References

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

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

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

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

5 Legg, J. P., Owor, B., Sseruwagi, P. & Ndunguru, J. Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Adv Virus 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

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

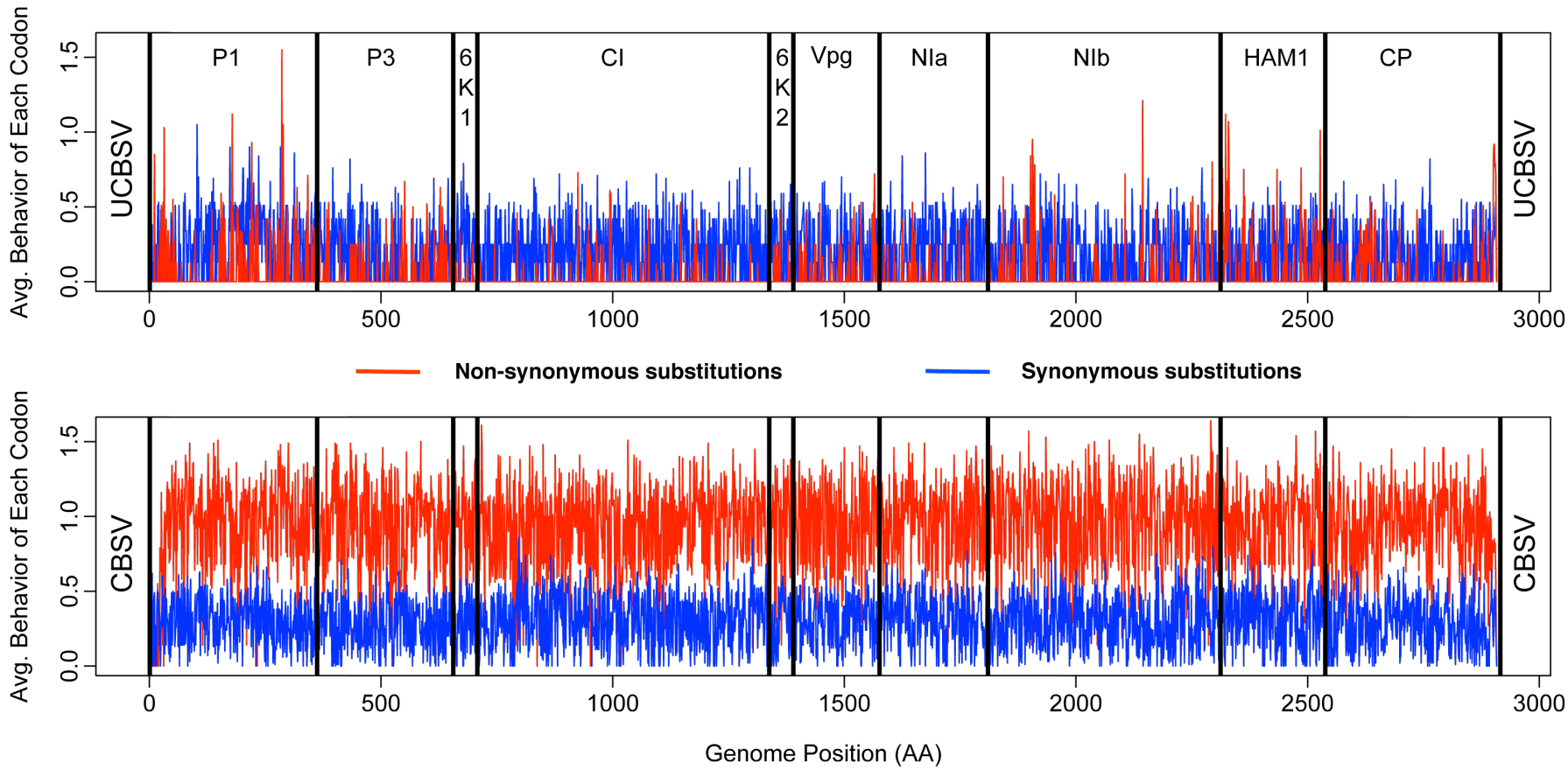
762
763 Table 5. Support for Clades A – G (Figure 3) in individual gene trees and whole genome analyses. Table entries represent
764 posterior probabilities from analysis with MrBayes, except values reported for SVDQ, which are bootstrap
765 proportions. Support values below 95% are indicated in bold, 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

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

781 Figure 2. Genetic diversity of CBSV and UCBSV using the Synonymous Non-synonymous Analysis Program (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

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

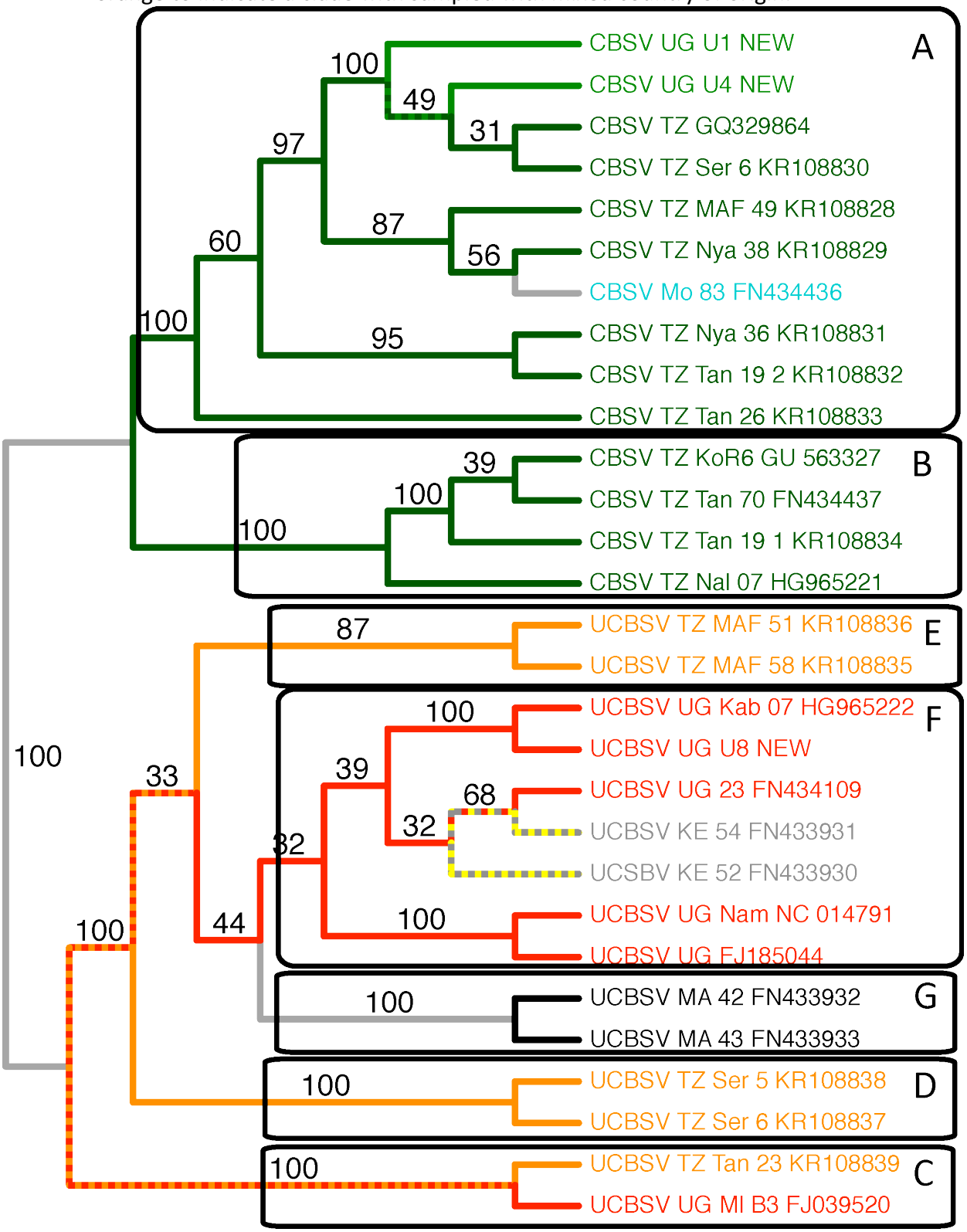
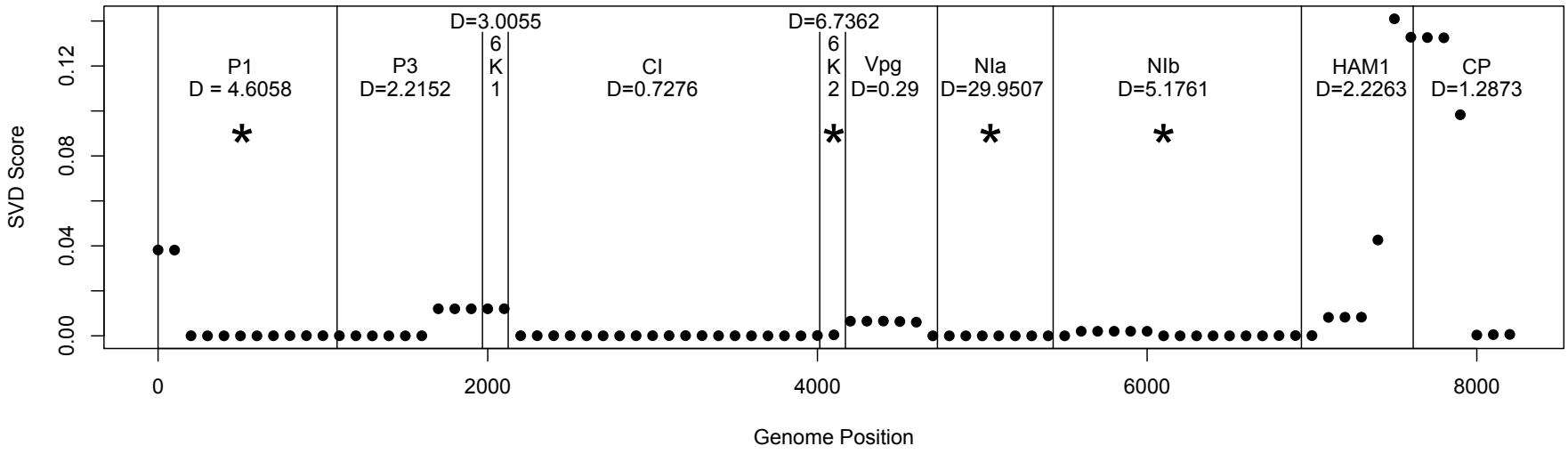
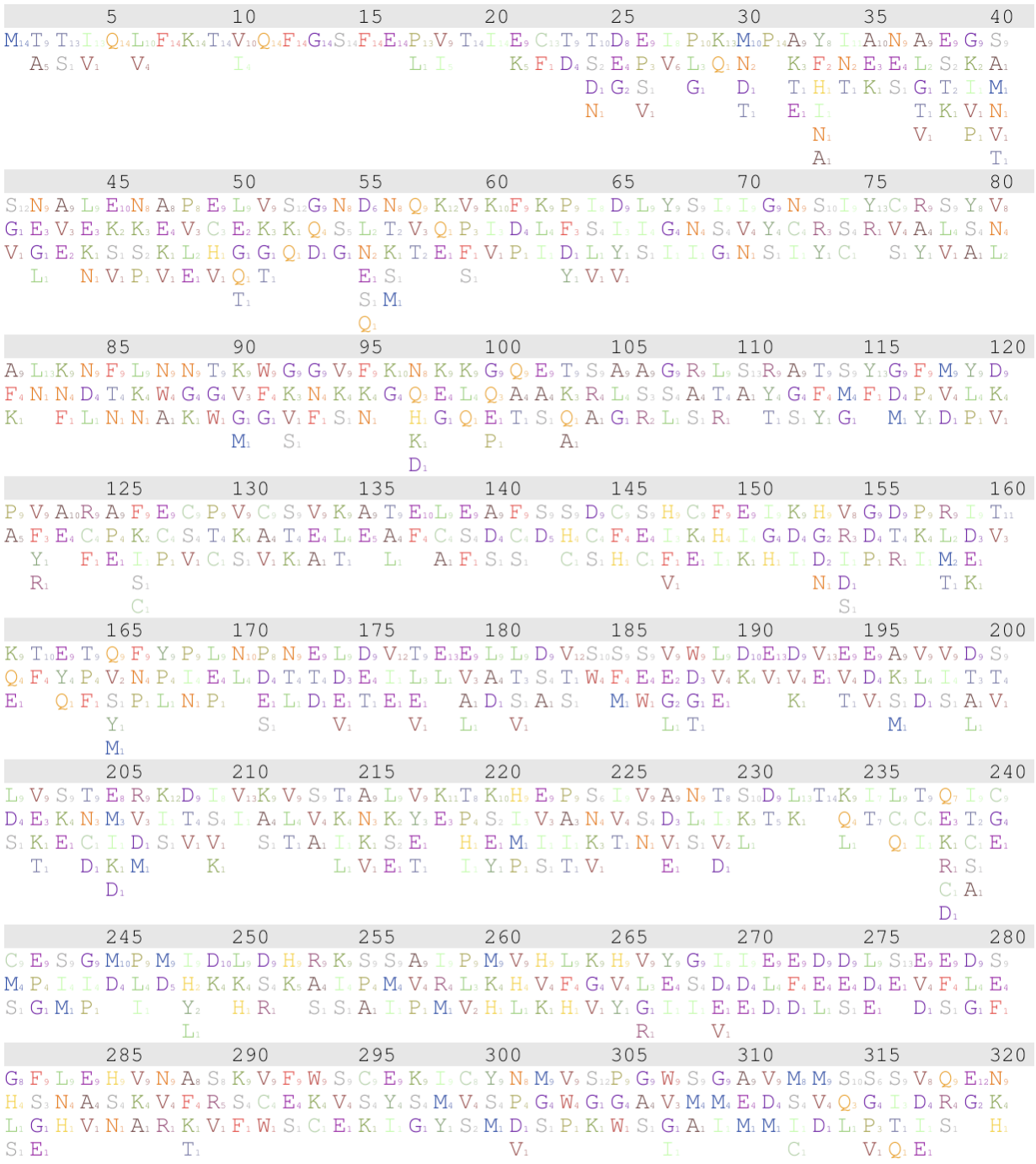
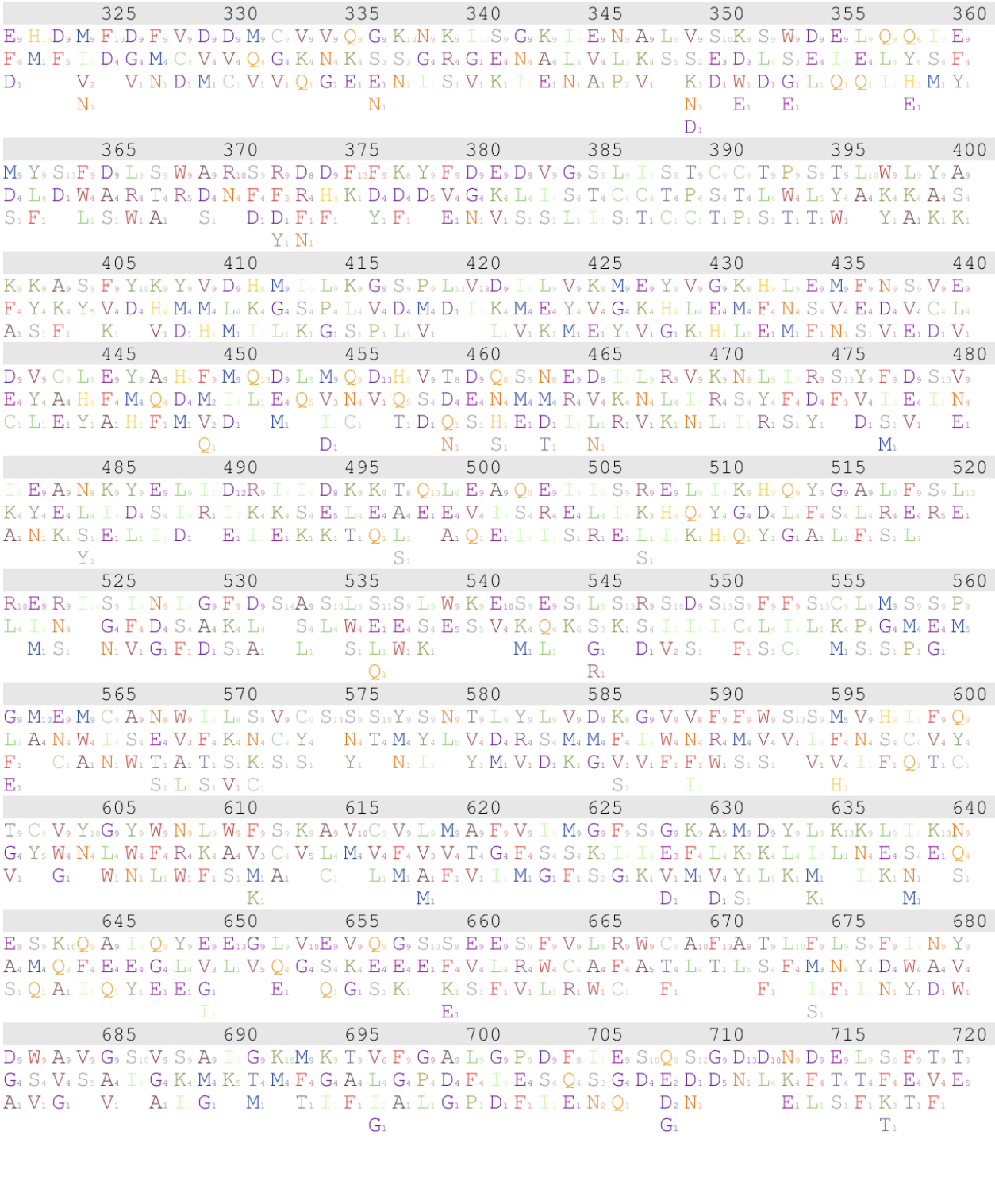


Figure 4. Computed SVD Score with the split defined by CBSV vs. UCBSV across the genome in windows of 500 bp, sliding in increments of 100 bp, and resulting SVD Scores plotted across the genome. Boundaries between genes are marked with vertical lines to further characterize the CBSV and UCBSV genomes. 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).

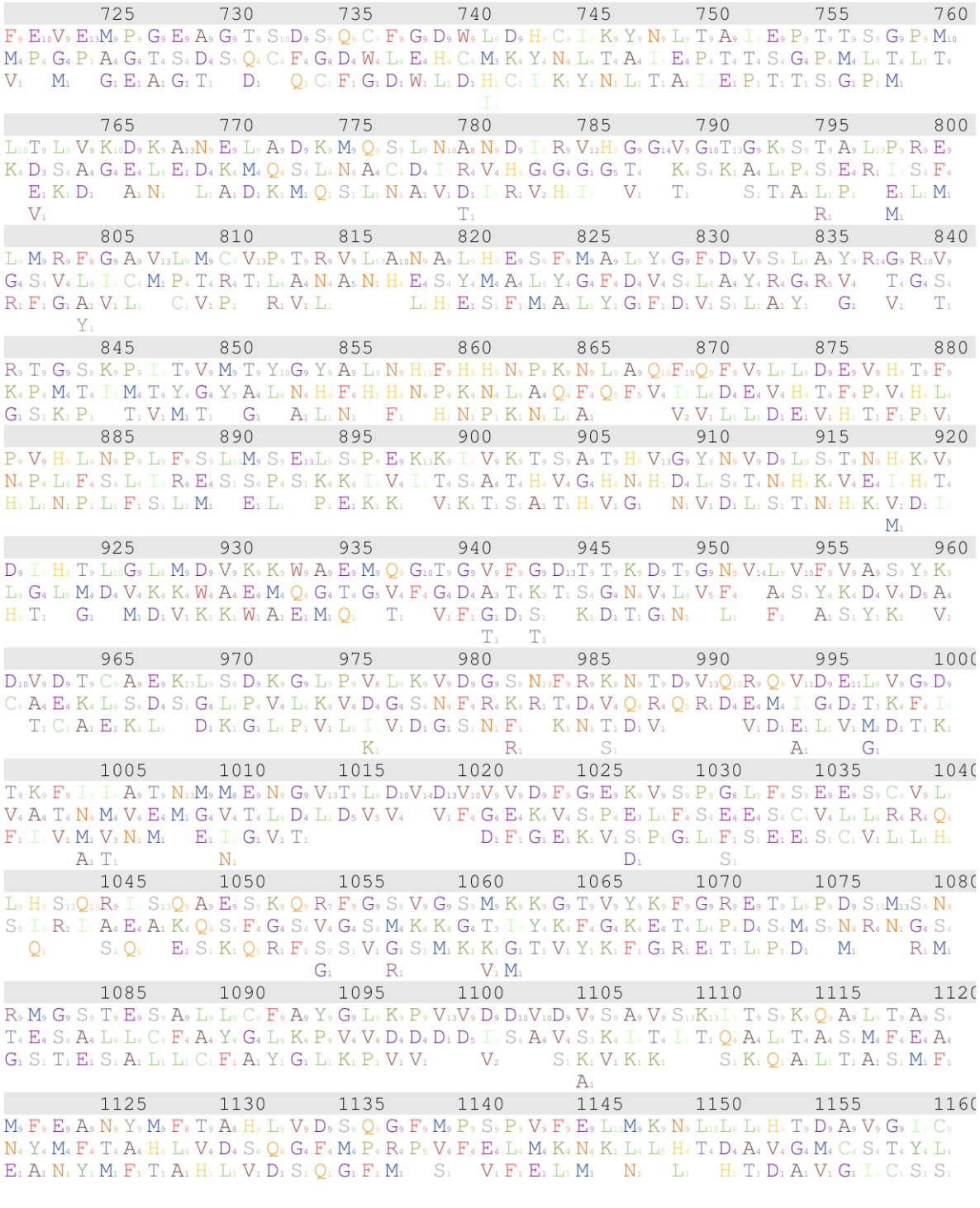


Supplemental Figure 1. CBSV Amino Acid variability obtained using datamonkey.org. Once files are uploaded to the site, the images below are obtained from the “Information from upload” tab and the pdf is downloaded.

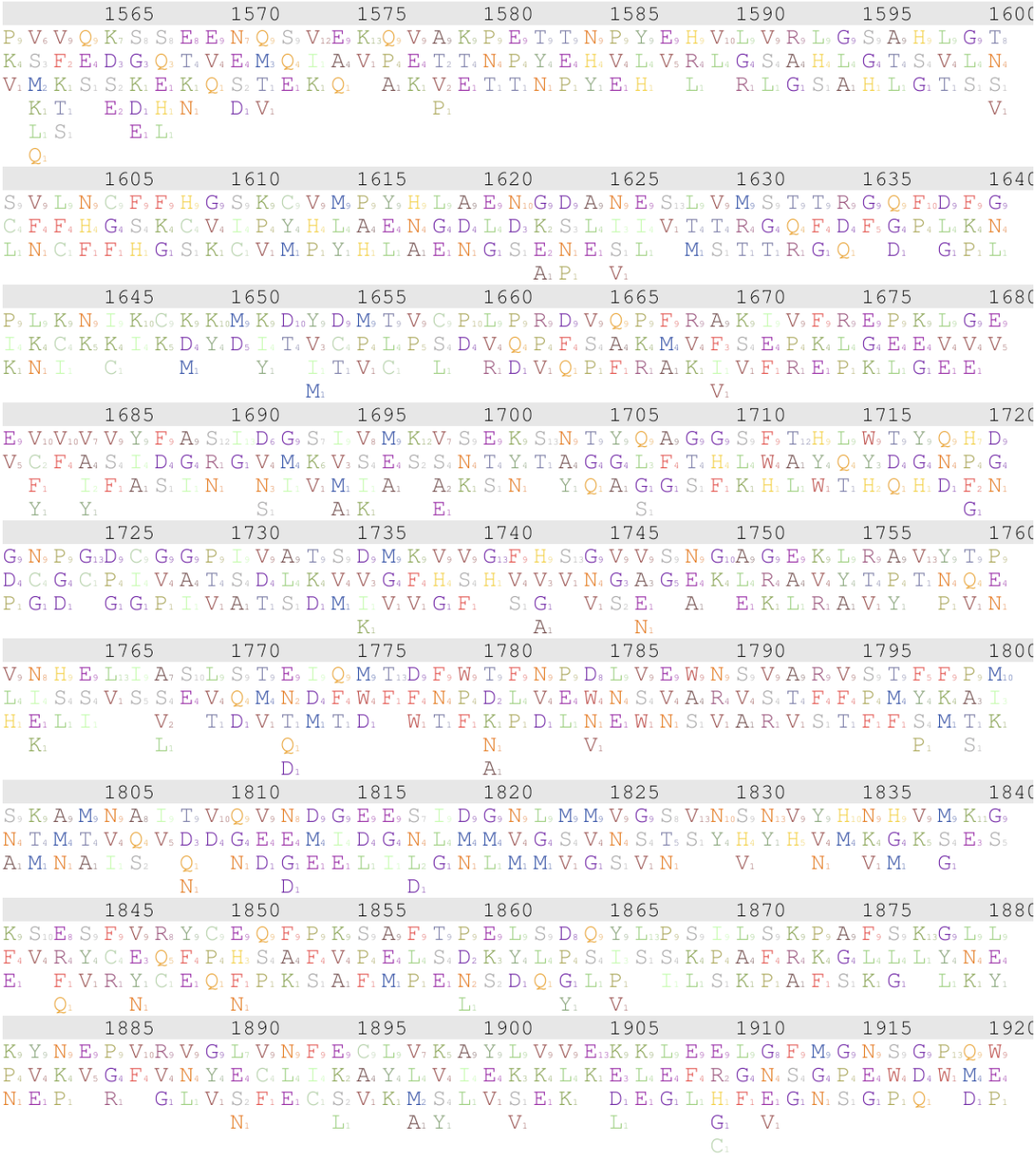




```
                325         330         335         340         345         350         355         360
E9 H13 D9 M9 F10 D9 F9 V9 D9 D9 M9 C9 V9 V9 Q9 G9 K10 N9 K9 I12 S9 G9 K9 I9 E9 N9 A9 L9 V9 S10 K9 S9 W9 D9 E9 L9 Q9 Q6 I9 E9
F4 M1 F5 I2 D4 G4 M4 C4 V4 V4 Q4 G4 K4 N4 K4 S3 S3 G4 R4 G1 E4 N4 A4 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 H3 M1 Y1
         N1                                  N1                                      N1    E1    E1            E1
                                                                                     D1
                365         370         375         380         385         390         395         400
M9 Y9 S13 F9 D9 L9 S9 W9 A9 R10 S9 R9 D8 D9 F13 F9 K8 Y9 F9 D9 E9 D9 V9 G9 S9 L9 I9 S9 T9 C9 C9 T9 P9 S8 T9 L10 W9 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 I4 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 Y10 K9 Y9 V9 D9 H8 M9 I9 L9 K9 G9 S9 P9 L13 V13 D9 I9 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 I1 K4 M4 E4 Y4 V4 G4 K4 H4 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 Q13 D9 L9 M9 Q9 D13 H8 V9 T8 D9 Q8 S9 N8 E9 D8 I9 L9 R9 V9 K9 N9 L9 I9 R9 S13 Y9 F9 D9 S13 V9
E4 Y4 A4 H4 F4 M4 Q4 D4 M2 I4 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 I3 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
I9 E9 A9 N8 K9 Y9 E9 L9 I13 D12 R9 I9 I9 D8 K9 K9 T8 Q13 L9 E9 A9 Q9 E9 I9 I9 S9 R9 E9 L9 I9 K9 H9 Q9 Y9 G9 A9 L9 F9 S9 L13
K4 Y4 E4 L4 I4 D4 S4 I4 R1 I1 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
R10 E9 R9 I14 S9 I9 N9 I9 G9 F9 D9 S14 A9 S10 L9 S11 S9 L9 W9 K9 E10 S9 E9 S8 L9 S13 R9 S10 D9 S13 S9 F9 F9 S13 C9 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 I1 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 M10 E9 M9 C9 A9 N8 W9 I8 L8 S8 V9 C9 S14 S9 S10 Y9 S9 N9 T9 L9 Y9 L9 V9 D9 K8 G9 V9 V8 F9 F9 W9 S13 S9 M5 V9 H9 I9 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                                          S1       I1                   H1
                605         610         615         620         625         630         635         640
T9 C9 V9 Y10 G9 Y9 W9 N9 L9 W9 F9 S9 K9 A9 V10 C9 V9 L9 M9 A9 F9 V9 I9 M9 G9 F9 S9 G9 K9 A5 M9 D9 Y9 L9 K13 K9 L9 I8 K13 N9
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 I4 I4 E3 F4 L4 K3 K4 L4 I3 L1 N4 E4 S4 E1 Q4
V1    G1    W1 N1 L1 W1 F1 S1 M1 A1    C1    L1 M1 A1 F1 V1 I1 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 K10 Q9 A9 I9 Q9 Y9 E9 E13 G9 L9 V10 E9 V9 Q9 G9 S13 S8 E9 E9 S9 F9 V9 L9 R9 W9 C9 A10 F13 A9 T9 L10 F9 L9 S9 F9 I9 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 S10 V9 S9 A9 I9 G9 K10 M9 K9 T9 V6 F9 G9 A9 L9 G9 P9 D9 F9 I9 E9 S10 Q9 S12 G9 D13 D10 N9 D9 E9 L9 S6 F9 T9 T9
G4 S4 V4 S5 A4 I4 G4 K4 M4 K5 T4 M4 F4 G4 A4 L4 G4 P4 D4 F4 I4 E4 S4 Q4 S3 G4 D4 E2 D1 D5 N1 L4 K4 F4 T4 T4 F4 E4 V4 E5
A1 V1 G1    V1    A1 I1 G1    M1    T1 I1 F1 I2 A1 L1 G1 P1 D1 F1 I1 E1 N2 Q1    D2 N1          E1 L1 S1 F1 K3 T1 F1
                                        G1                                  G1                                T1
```



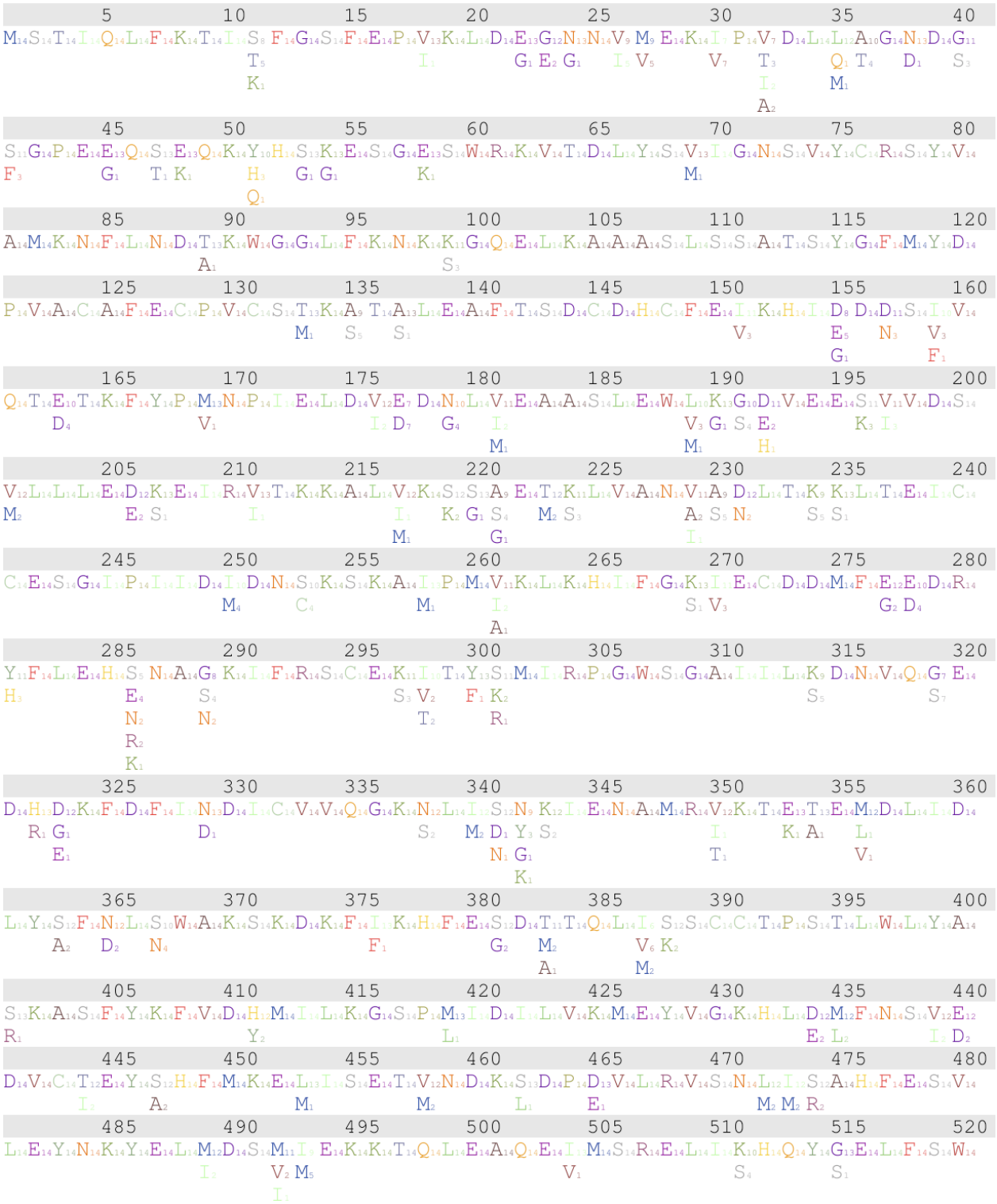
```
         1165      1170      1175      1180      1185      1190      1195      1200
S8 S9 Y9 L9 A9 T9 N9 M9 S9 E9 W9 S9 R8 V9 C9 E9 Y9 I5 K9 I8 D9 E9 S9 S9 R9 H9 V9 Q9 E9 V9 K8 I9 P9 W9 Y9 C9 S9 D13M9 S9
A4 T4 N4 M4 S4 E4 W4 S4 R4 V4 C4 E4 Y4 I4 K4 I3 D3 M5 S4 S4 S5 H4 V3 Q4 E4 V4 K4 M4 P4 W4 Y4 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 I1 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 I9 V9 K9 L9 A9 E9 C9 V9 K9 A9 A9 K10P9 K9 L8 S12S9 G9 Y9 K9 V9 D9 N9 V9 D9 F9 H13T9 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 I1 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 I9 D9 E9 S9 S9 A9 L9 V9 A9 T9 I9 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
I1 D1 E1    S1 A1 L1 V1 A1 T1 I1 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 I9 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 I4 D4 D4 Y4 E4 S4 L4 V4 S4 F4
W1 I1 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 I9 E9 E9 K9 V9 M9 N13V9 K9 K13N9 Y9 D9 K9 P8 I10I9
F4 S4 E4 N4 P4 H4 S4 A4 E4 Y4 L4 E4 A4 Q4 C4 A4 S4 D4 Y4 I4 E4 E4 K4 V4 M4 N2 V4 K4 R3 K1 Y4 D4 Y1 P4 M4 I4 I4 G4 L4 V4
S1 F1 F1 S1 E1 N1 P1 H1 S1 A1 E1 Y1 L1 E1 A1 Q1 C1 A1 S1 D1 Y1 I1 E1 E1 K1 S2 M1 N1 S1    K1 N1    D1 K1 P1 V1 I1    G1
                                                         V1       V1                                  S1
         1365      1370      1375      1380      1385      1390      1395      1400
I9 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 I1 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 I4 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 I1 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 I4 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 I1 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 D10I9 D9 N9 M9 L9 V7 E9 L9 D10D9 D9 F7 H9 I9 L9 S8 G9 D9 S8 M9 L9 G9 K9 K9 V9 E9 L9 A9 F9 T9 K9 D12
K4 D3 V4 N4 V4    G4 M4 L4 I3 E4 L3 D4 D3 D4 Y4 H4 I3 L3 K4 D4 D4 R4 M4 L3 G3 K5 K1 V4 E4 L3 A4 F4 T4 K3 S3 G4 S4 G2 S1
S1 A1 K1 D1       I1    N1 M1 L1 V1 E1 I2    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                I1          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 S13M9 S10L9 N5 P13M9 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 N1 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
```



```
         1925      1930      1935      1940      1945      1950      1955      1960
D9 P9 M9 E9 I9 L9 S5 D9 L9 N9 K9 K9 A13A9 M9 G9 A9 L9 Y9 Q13G9 S9 K9 Q9 D9 W9 L9 K9 S9 M9 T9 P9 E9 E9 F9 I9 M5 S13V9 R9
I4 L4 S3 D4 L4 N4 K4 K4 A4 A4 M4 G4 M1 L4 Y4 Q4 G4 S4 K4 S1 D4 W4 L4 K4 S4 M4 S5 P4 E4 D4 F4 V3 V4 S3 V5 R4 V5 R1 F4 K4
M1 E1 I1 L1 N1 D1 N3 N1 K1 K1 A1 A1    G1 A1 L1 Y1 Q1 G1    K1 Q1 D1 W1 L1 K1    M1 T1 P1 E1 I1 F1 I1    S1 E4    E1 S1
      N1        D1                                                                                E1    A1
         1965      1970      1975      1980      1985      1990      1995      2000
E9 S9 F9 K9 H9 L9 A9 G13G9 D9 V9 G13I9 W9 S10G9 S9 L13K9 A9 E9 L9 R9 P9 V9 E13K9 V9 S9 E9 M9 K9 T13R9 V9 F9 T9 G9 A9 P9
H4 L4 A4 G4 G4 D4 V3 D1 I4 W4 S4 W1 S5 L4 K4 A4 E4 A1 S4 P3 C2 E4 K4 V4 K5 V1 M4 K4 T4 S4 V4 F4 V1 G4 A4 P4 V4 D4 V5 L4
F1 K1 H1 L1 A1 G1 G1    V1 G1 I1       G1    L1 K1    E1 L1 R2 P1 V1 E1       S1 D1 M1 K1 T1 R1    F1 T1 G1 A1 P1    D1
                  I1                                   T1 Y1
         2005      2010      2015      2020      2025      2030      2035      2040
V10D9 V9 L9 L9 G9 G9 K9 V10L9 V9 D9 N10F13N9 H8 H9 F9 Y9 E9 N9 H9 L9 K9 G9 P9 W9 T9 V9 G9 M9 N9 K9 F9 N10S10G9 W9 D9 S9
L3 G4 G4 K4 I4 L4 V5 D4 N4 F4 N5 H4 H3 H1 Y4 E4 N4 H4 L4 K4 G4 P4 W4 T4 V4 G4 I4 N4 K4 F4 N4 R4 G4 W4 D3 K3 L4 A4 S3 Y4
   L1 L1 G1 G1 K1    L1    D1    F1 Y1    H1 F1 Y1 E1 N1 H1 L1 K1 G1 P1 W1 T1 V1 G1 M1 N1 K1 F1 N1 S1 G1 W1 D1 S1 E1 A1
                                                                                                                  L1
         2045      2050      2055      2060      2065      2070      2075      2080
L9 A9 S9 Y9 F9 N13H9 G9 W9 N9 F9 I9 D10C9 D9 G9 S9 R9 F9 D9 T9 S9 L13A9 P9 M9 L9 F9 Q9 L9 I7 C9 H9 M9 S10E9 S9 F9 G9 E8
F4 D2 H4 S4 W4 D1 F4 I3 D4 C4 D5 G4 S4 R4 F4 D4 T4 S4 L4 A4 P4 M4 P1 F4 Q4 L4 V4 C4 H4 M3 S4 E4 S5 F4 G4 G4 F4 D4 A4 M4
S1 N2 F1 N1 H1    W1 N1 F1 I1    C1    G1 S1 R1 F1 D1 T1 S1 L1 A1    M1 L1 F1 Q1 L1 I1 C1 V2 M1    E1    F1 G1 E1 F1 G1
   Y1             M1                                                 V1 H1                                          D1
         2085      2090      2095      2100      2105      2110      2115      2120
F9 D9 N5 I9 E9 T9 I6 A9 L8 S9 N9 L9 Y9 T9 Q9 I9 V9 Y9 T13P9 I9 L9 T9 I12D9 G9 Y9 M9 T8 K9 K9 H9 S13G13N9 N9 S13G9 Q9 P9
E4 T4 S5 A4 L4 S4 N4 L4 Y4 T4 Q4 I4 V4 Y4 T4 P4 I4 L4 I1 I4 D4 G4 Y4 G1 T3 K4 K4 H4 S4 G4 N4 N4 N1 N1 Q4 P4 Q2 T4 V4 V4
N1 I1 A3 V1 I1 A1 V2 S1 M1 L1 Y1 T1 Q1 I1 V1 Y1 T1 P1    L1 T1 I1 D1    F1 M1 T1 K1 K1 H1 S1 G1       S1 G1    P1 S1 T1
      E1 T1       A1 M1                                                F1        A1
                  L1 N1
         2125      2130      2135      2140      2145      2150      2155      2160
S9 T9 V9 V9 D9 N9 T9 M9 I9 L9 M9 I9 V9 V9 E13Y9 C9 S9 E9 V9 L9 S6 Q9 E9 G9 V8 S9 M9 K10Y10K7 Y9 M9 C9 N9 G9 D9 D9 L10I9
D4 N4 T4 M4 I4 L4 M4 I4 V4 V4 E4 Y4 C4 S4 C1 V4 M4 A4 H4 E4 G4 I1 E4 M4 K3 Y3 K5 Y5 M4 C4 N4 G4 D4 D4 L4 I4 L8 N4 A3 P4
V1 V1 D1 N1 T1 M1 I1 L1 M1 I1 V1 V1 E1 Y1    S1 E1 V1 L1 S1 Q1 L3 G1 V1 S2 F1          S2 C1 N1 G1 D1 D1    I1 V1 N1
                                            E1          I1          M1
                                            V1          M1
         2165      2170      2175      2180      2185      2190      2195      2200
L9 N9 A9 P9 D9 D9 E9 I10P9 I9 I4 Q9 S10R9 F13K9 D7 L9 F9 A9 E9 C9 G9 L9 N9 Y9 D10F9 D9 D9 V9 H9 K9 S9 I9 E9 T9 I9 E9 Y9
D4 E4 E4 I3 P4 M4 I4 Q4 S4 R4 F4 K4 N3 L4 D1 A4 E4 C4 G4 L4 N4 Y4 D4 F4 D5 D4 V4 H4 K4 E2 M4 D3 T4 M4 E4 Y4 M4 S4 H4 S5
A1 P1 D1 V1 E1 I1 P1    I1 Q1 S1 R1 F1 K1    L1 S2 A1 E1 C1 G1 L1 N1 Y1    F1    D1 V1 H1 K1 S2 I1 E1 T1 M1 E1 Y1 M1
         D1                                  F1                                K1
                                                                               S1
         2205      2210      2215      2220      2225      2230      2235      2240
M9 S10H9 S12F9 A9 Q9 R9 D8 G9 F9 F7 I9 P9 K10L9 K9 K9 E8 S9 I4 M6 A9 I7 L9 E10W9 E13R7 G9 D9 E9 V8 M9 S10T9 R7 S9 A13L9
F4 A4 R3 K1 D4 G4 F4 F4 I4 P4 K4 L4 K5 K4 E4 S4 M4 V4 A4 M3 L4 E4 W4 E5 S4 G4 D4 G1 I2 M4 S4 T4 S5 S4 A4 L4 S2 A4 N1 Y4
H1    F1 A1 Q1 R1 D1 G1 G1 Y1 I1 Y2    L1    K1 E1 S1 I1 I1 M2 V4 L1 M2 W1    S1    V2 E1 V1 M1 I1 T1    S1 N3 L1    A1
      Q1                F1       P1                D1 V1 A1                    S2                                A1
                                                     V1                       D1
         2245      2250      2255      2260      2265      2270      2275      2280
N9 A9 A9 Y9 I9 E9 S9 F9 G9 Y9 D9 D9 L9 M9 M8 E10I9 E9 R7 F7 A13V9 F9 W9 A9 S9 E11K9 G9 C9 E9 Y9 P9 L9 L8 D9 S9 K9 R9 V9
I9 E4 S4 F4 G4 Y4 D4 D4 L4 M4 T3 E4 I4 E5 S4 F2 A4 V4 F4 W4 F1 S4 E4 K3 G4 C4 G2 Y4 P4 L4 L4 D4 S4 K4 R4 V4 E4 G4 L4 Y4
A1 Y1 I1 E1 S1 F1 G1 Y1 D1 D1 A1 M1 T1    I1 Y2 S1 Y1 S2 Y2    W1 A1 S2 E1 K1 S1 C1 E1 Y1 P1 L1 L1 D1 M1 K1 R1 V1 E1 G1
                     L1          V1          A1 V1                                                   S1
```

```
         2285      2290      2295      2300      2305      2310      2315      2320
E9 G9 L9 Y11L9 D13E9 F6 T9 E11I9 N9 E9 E9 W9 L13S9 G9 I9 L9 P9 P9 S9 F9 E9 D7 C9 Y9 I9 D9 L9 Q9 V9 V9 D9 S12P9 Q9 S7 L7
L5 D4 E4 H2 T4 Y1 I4 N4 E4 D2 W4 L4 S4 G4 I4 G1 P4 P4 S4 F4 E4 N4 C4 Y4 I4 N6 L4 Q4 V4 V3 D4 S3 P4 Q3 S4 N1 S3 V3 A3 K4
   Y1 F1 D1 E1    T1 Y2 I4 N1 V2 E1 W1 L1 S1    I1 L1 P1 P1 S1 F1 E1 N1 C1 Y1 I1 D1 L1 M1 V1 K1 D1 S1 P1 Q3 K1 S1 T2 S2
                   D1       E1                                   M1          Q1    V1    R1       N3 M1 L1 N1       V1
                   V1
         2325      2330      2335      2340      2345      2350      2355      2360
N9 V8 A7 K8 S9 E9 E7 E9 V9 T9 S9 K9 F7 R9 M8 G9 I10E9 A9 P9 M7 T9 F9 V9 T9 G9 N9 A9 Q8 K10L9 K9 E9 V9 K9 Q9 I9 F9 G9 P9
S4 E4 E4 E5 V4 T4 S4 K3 I4 R4 M3 G3 I1 E3 A4 P4 A1 T4 F4 V4 T4 G4 N4 A4 Q4 K4 L4 K5 E4 V4 K4 Q4 I2 F4 G4 P4 T4 I4 P4 I5
A1 K1 V2 S1 E1 G1 K2 T1 S1 K1 I1 E1 V1 G1 I1 E1    P1 I1 T1 I2 V1 T1 G1 N1 A1 Q1    H1    E1 V1 M2 Q1 I1 F1 G1 P1 T1
   A1 S1       V1 S1       V1 R1 M1 A1 T1              F1                                L1          K1
         2365      2370      2375      2380      2385      2390      2395      2400
T9 I10P9 I9 V10S9 R8 K9 V5 D9 L9 P10E9 S7 Q9 G9 T9 I7 E9 E13I9 I9 K10E9 K9 A11R9 V9 A9 A9 E8 L7 V13G9 G9 P9 V10L9 V9 E9
V3 S4 R3 K4 I3 D4 L4 P4 E4 P5 Q4 G4 T4 V3 E4 E3 I4 V6 K4 I1 K5 A3 R4 V4 A4 S2 E4 L3 V3 G4 G4 P4 G1 L4 V5 E4 D4 T4 S4 L4
I1    Q1 S1 R1 K1 V1 D1 I3    E1    Q1 P2 T1 D1 E1 E1 I1       E1    A1 R1 V1 A1 F1 I1 L1 K1 I2    P1    L1    E1 D1 T1
P1    V1          Q1    F1          G1 A1 D1 A1                V1                A1 E1 V1 G1
                        L1          A1
         2405      2410      2415      2420      2425      2430      2435      2440
D9 T9 S9 L9 C9 F9 D9 A9 L9 N9 G9 L9 P10G9 P9 Y8 I8 K9 W9 F9 L7 E9 G10I4 G9 L9 E9 G9 L8 Y9 K9 L9 V9 E9 P9 Y9 Q9 N9 S7 M9
C4 F4 D4 A4 L4 N4 G4 L4 P4 G4 P5 Y4 I4 K4 W4 F4 M4 E4 G4 I4 G5 L4 E4 G4 L4 Y4 K4 L4 V4 E4 P4 Y4 Q4 N4 K4 M4 A4 S4 A5 L4
S1 L1 C1 F1 D1 A1 L1 N1 G1 L1    G1    Y1 I1 K1 W1 F1 L1 E1 M2 I1    L1 E1 G1 L1 Y1 K1 L1 V1 E1 P1 Y1 Q1 N1 K1 M1 K2 S1
                                                   V1
         2445      2450      2455      2460      2465      2470      2475      2480
A10S9 A9 L9 C9 V13F10A9 F9 V9 N9 K9 V9 G9 D9 D9 P9 M8 I9 F9 K9 G13V9 L9 S8 G9 E9 I7 V9 M9 P9 R9 G9 P6 N8 S8 F9 G9 W9 D9
C4 V4 F4 A4 F5 A1 N4 K4 V4 G4 D4 D4 P4 M4 I4 F4 K4 N4 V4 K4 G4 V4 L4 S4 L1 E4 I4 V4 M4 P4 R4 G4 P1 R1 K2 S3 F4 G4 S4 W4 D4 P4 I4 F4 Q4
   L1 C1 V1       V1 N1 K1 V1 G1 D1 D1 P1 M1 I1 G1 V1       S1 G1 K3 I1 V1 M3       G1 L1 P1       S3 G1 W1 D1 P1 I1
                                                 E1                   S1 N1                F1 L1
                                                                      N1
         2485      2490      2495      2500      2505      2510      2515      2520
P12I9 F9 Q9 P9 L9 N8 W9 K7 S8 T9 F9 A12E13M9 M9 T7 E9 E9 K9 N9 M8 M9 S9 H9 R10F9 R9 A9 L10S9 L9 V9 S9 D12F9 L9 K9 S9 S11
F1 L4 N4 W3 S3 S3 D4 F4 A4 E4 M4 M4 M1 M1 E4 K4 N4 M4 M4 S4 H4 R4 F4 R5 A4 L4 S4 L5 V4 S4 D4 F4 L4 K4 L1 S5 S4 Y4 F4 H1
S1 Q1 P1 L1 G1 K1 T4 S1 S2 K3 A1 E1 T1    A1 E1 A2 K1 N1 M1 M1 T1 H1    F1    A1    S1    V1 S1 D1 F1 S1    N1 S1 D2 T1
      M1 N1 W1 K1       T1 F1                E1                S1                                                    Y1
         2525      2530      2535      2540      2545      2550      2555      2560
S9 Y9 F10S9 F9 A9 K8 G8 L10D10R9 D11I8 F9 I9 D9 V9 Q9 A9 I9 D9 K9 D9 E14I9 E10A9 E9 M9 T7 K8 L9 K10E9 L9 W9 S7 S7 N9 K9
F5 A4 K4 G4 V3 D4 R4 D5 V2 F4 I4 F1 V4 Q4 A4 I2 D4 K4 D3 E4 I4 E5 A4    M4 T4 K4 L4 K5 E4 L4 W4 S3 N4 N4 K4 P4 T4 K2 T5
   S1    A1 K1 G1 E1    I1    F1 N1 L1 D1 V1 V2 A1 I1 G2 K1          I1 A1    M1 T1    A2 E1 E1 M1 W1 S1 N1 K2 N2 K1 S2
         L1       V1    R1                Q1                                             L1 Q1                N1       P1
         2565      2570      2575      2580      2585      2590      2595      2600
P9 T10K8 T9 R9 S13P9 F9 E9 S10S9 S9 L9 R8 A13P9 Q9 V9 A9 R9 V9 N9 E9 L13L9 K9 Q9 L9 K9 D9 A9 G9 M8 Q9 T9 S10K9 S9 P13C9
R4 S4 P4 F3 E4 F1 S4 S5 L4 R4 A4 P4 Q4 V4 Q1 R4 V4 N4 E4 L4 L4 N1 E1 K1 K4 D4 A4 G4 M2 Q4 T4 S4 K4 S5 P4 C4 G4 E4 G1 D4
S1    S1 S1 P1    E1    K1    L1 R1 A1 P1    V1 A1 R1 V1 N1 E1 Q2 Q4 L1 Q1 L1 K1 D1 I1 G1 M1 Q1    K1    P1 C1    E1
      R1                                                   L1                    V1
                                                                         A1
         2605      2610      2615      2620      2625      2630      2635      2640
G9 E8 P9 D9 E10G9 E9 V9 A8 S13P9 E9 S8 S9 E10D8 E9 E9 Q9 R6 K9 D7 K10G9 K9 A8 P7 M5 E13P7 P8 A8 E8 R8 S7 Q9 S9 E8 K8 S9
E4 G4 E5 V4 A3 S4 P4 E4 S4 E1 E4 D4 E5 E4 Q4 Q3 T2 D3 K4 G4 T5 A3 P4 M4 E4 P2 S2 V4 P1 K1 S3 Q4 K1 S2 E3 K3 S4 V2 G3 E3 G2
P1 D1    G1 V1 V1 A1 S1 P1    S1 S1    D1    G1 Q1 N1 T1 N1 M3 G1    A1 P1 S2 E1 T3    H1 E1 K1 T1 K1 F1 E1 K1 S2 M1 F1
                                           M1 T1 P1       L1 N1    P1 Q1 G1    M1    A1 D1
                                              A1          Q1 Q1       N1          T1
                                                          T1
                                                          S1
```

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



525 530 535 540 545 550 555 560
R14E14R14L11C11L12K12L14G14M14G14S14S14N14L14Y14T14Y14W14V12E14R14E14E14K10Q12S13E14K13S14S10A11I14S14C12I14I14S14K14P14
M3 Y3 M2 S2 E2 S2 H2 L1 S1 N4 V3 F2 Q2

565 570 575 580 585 590 595 600
G14M14E14M14L14I10N12W14V11S14E14V13C14K13S13K14Y14H14S14L14V14N14C14V14D14S13G14I12L13F14L14W14S14S14M11V13H14L14A12S10
M4 S2 A2 A1 S1 N1 N1 V2 M1 V2 I1 M2 K4 I1

605 610 615 620 625 630 635 640
V12S14V13Y14G14Y14W14N14L14W14F14S14Q14A14M11C14V14L14F14I12F14L14V12S14N14F14S14S7K14M12V14S11Y13L14K14K12L14I14V13S11
A2 I1 V2 L1 V1 I2 G5 C2 L2 N2 F3 G1 T2 I1 N3

645 650 655 660 665 670 675 680
E14K12K14L14A14M12K14N14E14E12G14F14V13E14V14Q14G14R14K14E14E14S13F14V14L14K14W14C14A14A11A14T14L14F14L14S14F14L14N14Y14
S2 I2 A2 I1 P1 V3

685 690 695 700 705 710 715 720
D14W14A14V14G14C14V13S14A14I14G14K14M14K14T14M12F14S12A14L14G14P14D14F14M12E14K10Q14D14G14D14D14D14L14K13F14T14T14F14E14
M1 I2 G2 I2 S4 S1

725 730 735 740 745 750 755 760
V14E14I14P14G14D14S14S14S14S14S14A14Q14T14F14G14E14W14I12E14H14C14I14K14F13N14L14V14S14I14E14P14T14T14S14G14P14M14L14T14
M2 Y1

765 770 775 780 785 790 795 800
L14E14R14G14K14A14N14E14L14A14D14Q14M12N14C14L14N14A14T14D14M13R14V14H14G14G14V14G14T14G14K14S14T14A11L14P14Y14E14L14M14
I2 I1 S2 T1

805 810 815 820 825 830 835 840
S14Y14G14A14V14L14V13C14V14P14T14S14V14L14A14N14A14L14H14E14S14F14M14S14L14F14G14F14D14V14S14L14A14Y14R12G13R14V14S14T14
M1 S2 S1

845 850 855 860 865 870 875 880
G14S14K14P14I14T14I14M14T14Y14G14Y14A14L14N14H14F14H14H14N14P14S10N14L14A14Q14F14Q14F14M11M12M14D14E14V14H14T14F14P14V14
K4 V2 L2 I1

885 890 895 900 905 910 915 920
H14L14N14P14L14F14S14L14L14R14E14L14S14P14D14K14K14M12I14K14T14S14A14T14H14V14G14H14N13V14D14L14S14T14N14H14K14V14D14M12
I2 S1 I2

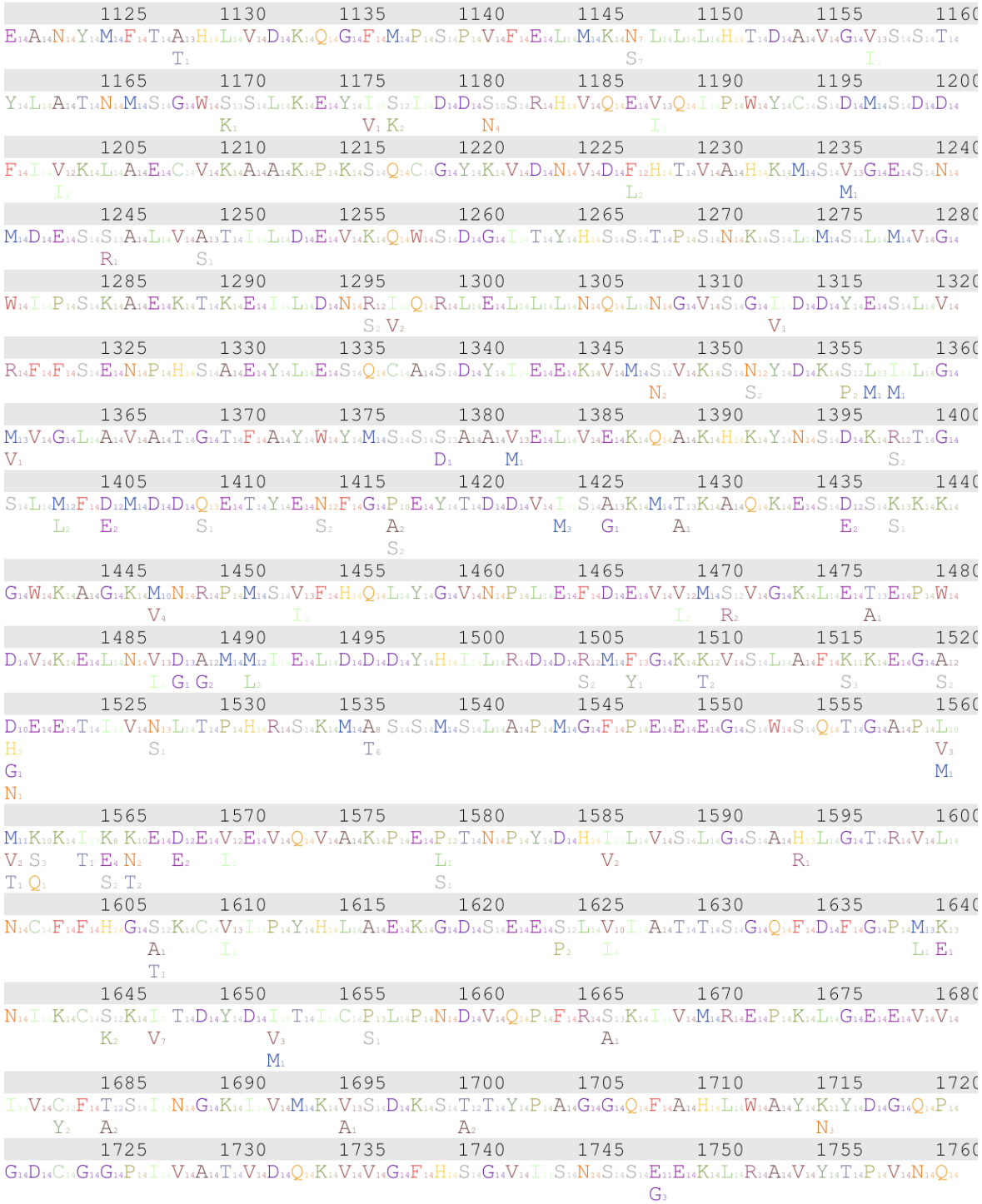
925 930 935 940 945 950 955 960
H14T14L14E14M9M14D14V14K14K14W14A14E14L14Q14G14T14S14V14F14G14D14V14T14K12E14P12G14N14V13L14V13F14V14A14S14Y14S14D13V14
V3 S2 H1 I1 I1 E1
L2 S1

965 970 975 980 985 990 995 1000
D14V14C14A13E12K14L14K14D14K14G14F14P14V14I12K14V14D14G14S12N14F14S14K14N14T14E14V14Q14K14M14V14D14G8M13Q10G14E14V14K14
S1 D2 M2 R2 E4 L1 R2 A2 K1 M1

1005 1010 1015 1020 1025 1030 1035 1040
F14I12V14A14T14N14I14I12E14N14G14V14T14L14D14V14D14V14V14V13D14F14G14E14R14M14S14P14N14L14C14S14E14D13S14C14I14L14M14Q14
M2 I1 E1

1045 1050 1055 1060 1065 1070 1075 1080
S14Q14S14I14S14Q14A14E14S14K14Q14S14F14G14S14V14G14S14M14K14S13G14S14V14Y14K14F14G14S14E12T14L14P14D14S14M14S14S9S14V14
K1 D2 N5

1085 1090 1095 1100 1105 1110 1115 1120
G14S14T14E14S14A14L14I14C14F14A14Y14G14L14K14P14V14V14D14D14V14D14M14G14S14V14S14S13V14T14Q14S14Q14A14L14T14A14S14M14F14
G1



```
         1765      1770      1775      1780      1785      1790      1795      1800
E14L14L12N14C12I10S14G12D14I14Q14M14T14D14F14W14T14F14N14P14D14L14V14E14W14N14S14V14A14S14V14S14T14F14F14P14M14T14K14A14
     M2   Y2 V4   S2
         1805      1810      1815      1820      1825      1830      1835      1840
I14N14T14I12T14V14Q14A14N11E14G14E14E14L13I14D14G14N14L14M14I12V14G14Y14V14N14R14E14V14Y14H14N14H14V13I14K14G14K12S14E14
       M2         D3         S1         M2                                I1         S2
         1845      1850      1855      1860      1865      1870      1875      1880
S14F14M10S13Y14C14E14Q14F14P14N12C14A14F14T14K14E12L14R12D14Q13Y14L14P14S14I11L14S14K14P14A14F14S14K14G14L14L14K14Y14N14
     T4 K1              S2         D2   S2   K1         V3
         1885      1890      1895      1900      1905      1910      1915      1920
E14P14V14R13V14G14S10V11N10F14P9 C14L14M12R14A14Y14L14K12V14E14E8 M14F14E13N6 L14G14F14L12K7 E14A14G14P14Q14W14D14P14I14
       Q1      L4 M2 D4   S5      I2         N2      G4      K1 D4       R2 E5
                L1                                 T2         T2          G2
                                                              A2
         1925      1930      1935      1940      1945      1950      1955      1960
E14I14L14E9 D14L14N14K14K14A14A14M14G14A14L14Y14Q14G14K14K14Q14D14W14L13K14S14M12E14P14A13D14F14I14T10A14V14R14E14S14F14
       D5                                                F1      I2      V1         M4
         1965      1970      1975      1980      1985      1990      1995      2000
K14H14L14A14G14G14D14V14G14I14W14S14G14S14L14K14A14E14L14S14P12V12E10K14V14L14E14Q14K14T14S12V14F14T14G14A14P14I14D14L14
                                                T1 I2 K4                R2
                                                S1
         2005      2010      2015      2020      2025      2030      2035      2040
L14L14G14G14K14I12L14V14D14N14F14N14H14F14F14Y14F14N14H14L14K14G14P14W14T14V14G14I14N14K14F14N14K14G14W14D14S14L14A14S12
            M2                                                                                                          K2
         2045      2050      2055      2060      2065      2070      2075      2080
Y13F14N12H14S12W14N14F14M13D14C14D14G14S14S14F14D14T14S14L14A14P14I14L14F14Q14L14V14C14H14M14R14E14K14F14G14N14F14D14D14
N1   D2   N2         T1
         2085      2090      2095      2100      2105      2110      2115      2120
I13E14S12A12A14L14R14N14L14Y14T14Q14M14V14Y14T14P14I14L14T14I14D14G14Y14I14T8 K14K14H14R14G14N14N14S14G14Q14P14S14T14V14
V1   K2 T2                                                A4
                                                          M1
                                                          V1
         2125      2130      2135      2140      2145      2150      2155      2160
V14D14N14T14I11M11L14M14I14V14V14E14Y14C14K14A14V14M14E14S12E14G14S13T4 M14Q14F14K14Y14M14C14N14G14D14D14L14I12L14N14V14
          M2 I3                                T2      K1 V4                                    M2
                                                          I2
                                                          E2
                                                          A2
         2165      2170      2175      2180      2185      2190      2195      2200
P14D12D14E14V14S12M14I13Q14S14S14F14S14E11L14F14S14E14C14G14L14D14Y14N13F14D14D14V14H14K14S14M14E14T14I14E14Y14M14S14H14
  N2         N2   V1         D2                   T1
                            N1
         2205      2210      2215      2220      2225      2230      2235      2240
S14F14M14L13K14D14G10V12Y14I14P14K14L14K14K12E14S14I12V14A14I14L14E14W14E14S14G14D14E14I14M14R12T14R12S14A14L14N14A14A14
       M1      D4 L1         S2      V1                                M4   S2   S2
                M1                  M1
         2245      2250      2255      2260      2265      2270      2275      2280
Y14I14E14S14Y14G14Y14E8 D14L13M14V11E14I14E14R14Y14A14V14F14W14A14T8 E13K14G14C14E14Y14P14L14L14D14S14K14R14V14E14G14L14
                D6   P1   I1                                A6 G1
                         T1
                         M1
```

```
          2285      2290      2295      2300      2305      2310      2315      2320
Y14K10D14D14H11T14D14I14N14E14E14W14L14M6 G14I14L14P14P14S14F14E14H14C14Y14V14D14T10Q14T13K12D14L12S13G8 S9 E12K10P12E14
  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
L13S13I5 E12S13H10D13G11V6 P5 Q11M13Q14M14K14F14P14V14T14F14V14T14G14N14L13G14K14L14A14E14V14K13S14I14L14G14I14A13N11D14
F1 G1 V3 G2 N1 C2 N1 S3 T6 S5 R3 L1                                    F2                S1                A13N11D14
      T3      Y2       A1 L2                                                                S1 S3
      P2               M1 T2
      A1
          2365      2370      2375      2380      2385      2390      2395      2400
V14M9 A14K8 N14I14D14L14P14E14V14Q14G14T14P14D11E14I13V13S11K14K14A14Q14L13A14V14K14M14T14N14S12P14V14L14V14E14D14T13C14
  V5    S5                                E2   V1 A1 K2                F1                G2                          A1
        R1                                V1        M1
          2405      2410      2415      2420      2425      2430      2435      2440
L14C14F14N14A14F14N14G14L14P14G14P14Y14I13K14W14F14L14K14E14L14G14L14E8 G14V14V14K12M14L14S14A14F14G8 D14K14S14A14Y14A14
                                        M1                        D6        S2                        E4
                                                                                                    S2
          2445      2450      2455      2460      2465      2470      2475      2480
L14C14T14F14A14Y14V13H14N13E14L7 S14D13P14I13V14F14K14G14V14V14N13G14E14I13V14P14P14R14G14N12N14G14F14G14W14D14P14M9 F14
              A1   S1   S7   E1   V3                K1                          K2                          I5
          2485      2490      2495      2500      2505      2510      2515      2520
K14P14D14G9 C14G7 C14T14F14A14E14M14P9 S14S11I10K14N14D13F14S14H14S14S14S13A14L14E14K14V14K14L11F10L14D14N14L14V10V14K12
        E5   S6                     S5   N2 M3       E1                K1                S3 Y4                M4   S2
             E1                          G1 N1
          2525      2530      2535      2540      2545      2550      2555      2560
Q14E14E12K12K11A13K7 V13A14L14T14I14D12V14Q14A13L14N12Q14E14E14M9 E14A12E14I14T12A14L14K14K14L13W14K9 D14N14G14P14T14S13
      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
T14R14S14P12F14E14A14S13S14L14S14A14P12Q14V14E14R14V14N14E14L14L14Q14K12L14K14D13E14G14L14Q14T14K14K14S14P14C14G14E14P14
        S2          K1                S2                        Q2       N1
          2605      2610      2615      2620      2625      2630      2635      2640
D14D14G14E14V13V13D13D14D14S14D12D13G12N11N12Q14S14S13G12K14E12V12V8 D12E14S14Q14N13N10Q13Q13P12S12K13P10K13F13K14I11S14
          M1 M1 V1        N2 N1 D1 D3 K1       F1 E2    K2       I4 H1       N2       D1 D4 K2 H1 L1 K2 S1 S2 S1 L1    V2
                            E1    D1                                N1       G1                S1          L1          L1
                                                                                                Q1
          2645      2650      2655      2660      2665      2670      2675      2680
G14D14G13S12A10I9 S12R14D14D14I13D14K13I13P14T14N13A14L14E14I13K14K14T14F14K14P14P14K10V14S14Q14S14A14Y14I13W14I14P14R14
      E1 N3 T4 V5 K2                  S1          S1                                S4
          2685      2690      2695      2700      2705      2710      2715      2720
S14Q14S14D14N14L14T14P14D14V14I13Q14N13F14L13A14Y14M10P14P14S14H13A14M14D14N14Q14L14A14S14G14I13E14V14E14N14W14A14I11E14
                            M1   S1          V4        Y1                                V1                      V2
                                                                                                              M1
          2725      2730      2735      2740      2745      2750      2755      2760
V14A14K14A14Y14G14V14N14M14Q14E14F14Y14R13T14V10L14P14A14W14M10V14N14C14M14V14N14G14T14S14D14E14S14K14N14E14K14S14W14R14
                                        H1   I4          I4
          2765      2770      2775      2780      2785      2790      2795      2800
A14V14E14L14N14S14Q14G14E14D14V14D14D14F14E14Y14P14M14E14P14M14Y14K13F14A14L14P14T14M14S14K14V11M14S14N14F14S14S14Q14A14
                                        S1                                I1
```