Full title:

PHYLODYNAMIC ANALYSIS of PORCINE CIRCOVIRUS TYPE 2 REVEALS GLOBAL WAVES of EMERGING GENOTYPES and the CIRCULATION of RECOMBINANT FORMS.

Giovanni Franzo[#1], Marti Cortey[#2], Joaquim Segalés[2], Joseph Hughes[3] and Michele Drigo[1].

[1]Department of Animal Medicine, Production and Health (MAPS), University of Padua, Viale dell'Università 16, 35020 Legnaro (PD), Italy; [2]Centre de Recerca en Sanitat Animal (CReSA), UAB-IRTA, Barcelona, Spain; [3] MRC-University of Glasgow Centre for Virus Research, Glasgow, Scotland, United Kingdom.

[#]corresponding authors: giovanni.franzo@unipd.it; marti.cortey@irta.cat

Abstract

Since the first description of *Porcine circovirus type 2* (PCV2), four genotypes (PCV2a, PCV2b, PCV2c and PCV2d) have been recognized and three of them have been shown to exhibit worldwide distribution. Here, the population dynamics of PCV2 has been reconstructed over time and the factors that have shaped its evolution determined. The results obtained confirm that PCV2 originated approximately at the beginning of the 20th century. The most recent common ancestor of genotypes PCV2a, PCV2b, PCV2c and PCV2d circulated in the 1950's, 1980's, 1960's and 1950's, respectively, and the population sizes of the individual genotypes remained low until the mid 90's, coinciding with the identification of PCV2 as a major pathogen of the pig industry.

The population dynamics of PCV2 have been characterized by the appearance of periodic waves of distinct genotypes that, after an initial rise, spread following major swine commercial routes and were then superseded by subsequent emerging genotypes. Various recombinant forms displayed comparable population dynamics and spreading routes to those of major genotypes, suggesting that recombinant strains are able to compete with parental ones. The capsid gene is subjected to immune selection and evasion of the host immune response seems to be a major force for the emergence and spread of new genotypes. In contrast, the evolution of other genes appears to be constrained by the particular genomic organization of PCV2. In summary, obtained results suggest that changes in farming strategies, international trade, host population immunity, recombination and the constraints imposed by genome organization have all played a major role in the evolutionary dynamics of PCV2.

Keywords

PCV2, phylodynamics, evolution, population dynamics, phylogeography, recombination.

Highlights

Different PCV2 genotypes circulated at low frequency until the 90's.

Farming strategy changes and international trade play important roles in PCV2 spread

Three major genotypes have successively emerged during the last 20 years.

Several recombinant clusters are distributed worldwide nowadays.

Recombination and immune derived selective pressure are important drivers of PCV2 evolution.

1. Introduction

Porcine Circovirus type 2 has emerged as one of the most devastating viral diseases of pigs, causing severe economic losses due to clinical and subclinical syndromes and costs associated to infection control (Alarcon, et al, 2013). PCV2 is a small non-enveloped circular single-stranded DNA virus (genome size about 1.7kb) that belongs to the family *Circoviridae*, genus *Circovirus.* This genus encompasses a group of other animal viruses such as *goose circovirus, canary circovirus, psittacine beak and feather disease virus, chicken anemia virus, and pigeon or columbid circovirus* . Although eleven PCV2-specific RNA transcripts have been predicted , only four of them (*ORF1, ORF2, ORF3* and *ORF4*) have been shown to be translated into proteins with relevant biological roles (Lv, et al, 2014). *ORF1* encodes –through alternative splicing– two proteins (i.e. *Rep and Rep'*), which are required for genome replication. *ORF2* encodes the *Cap* protein, the only component of the viral

capsid and the major target of the host immune response (Khayat, et al, 2011; Nawagitgul, et al, 2000). *ORF3* and *ORF4* are located in the same region as *ORF1* but in the antisense strand, and encode for two non-structural proteins involved in the regulation of viral replication and apoptosis (Lv, et al, 2014; Gao, et al, 2014; Karuppannan and Kwang, 2011; Mankertz, et al, 2004). Two intergenic regions (IR) oriented in opposite directions are located between *ORF1* and *ORF2*: the shorter is located between the 3' end of *ORF1* and *ORF2*, while the longer is located between their 5' ends and contains the origin of genome replication (Lv, et al, 2014).

PCV2 has been traditionally divided into four major genotypes: PCV2a, PCV2b, PCV2c and PCV2d. All but PCV2c have been reported worldwide (Xiao, et al, 2015a; Franzo, et al, 2015c), albeit with different prevalences. It is believed to be transmitted mainly through the oronasal route, even though any secretions and excretions can be involved in the viral transmission (Grau-Roma, et al, 2011). Although PCV2 has been reported to infect some wild population (i.e. wild boars, feral pigs and Peccaries) (Franzo, et al, 2015a), it affects mainly domestic pigs and has been a major challenge for the swine industry in the last 20 years. PCV2 causes a group of diseases collectively named porcine circovirus associated diseases (PCVD). One of them, PCV2-systemic disease (PCV2-SD) showed epidemic proportions in Europe and South East Asia by the late 1990's and in North-America by 2004-05 (Segales, et al, 2013). The increase in the frequency of outbreaks substantially mirrored an increase in PCV2b prevalence, which replaced PCV2a as the most widespread genotype (Dupont, et al, 2008; Carman, et al, 2008; Cheung, et al, 2007; Cortey, et al, 2011; Grau-Roma, et al, 2008; Wang, et al, 2009). More recently, the PCV2d genotype was identified in China and its wide circulation in that country is currently recognized (Ge, et al, 2012). Nevertheless, other studies have confirmed its presence, before its first recognition as an independent genotype, in Switzerland (1998) and its circulation in several countries has been reported (Xiao, et al, 2015a; Franzo, et al, 2015c). Some reports suggest that PCV2d may evade vaccine-induced immunity (Seo, et al, 2014; Xiao, et al, 2012), which had been traditionally

77 considered fully protective against previously circulating genotypes (Beach and Meng, 2012; Chae,
78 2012; Fort, et al, 2009). However, a recent report indicated that PCV2a based vaccines are also able
79 to protect against PCV2d infection (Opriessnig, et al, 2014a).

80 In the past years, other studies have provided remarkable knowledges of PCV2 epidemiology and
81 evolution, investigating its origin, dynamics and routes of spread (Xiao, et al, 2015b; Firth, et al,
82 2009; Vidigal, et al, 2012). Nevertheless, a comprehensive study on the epidemiology and
83 evolutionary dynamics of all known PCV2 genotypes is still lacking. Here, we aimed to elucidate
84 the molecular epidemiology of PCV2 at the global scale and determine the factors that have shaped
85 its evolution at the genomic level.

86 2. Material and methods

87 *2.1.1 Dataset*

88 A total of 925 PCV2 complete genome sequences with known collection dates and country of origin
89 were downloaded from GenBank (accessed 06/10/2014) and aligned using the MAFFT method
90 (Katoh and Standley, 2013). All poorly aligned sequences and those displaying degenerate
91 nucleotides or indels which caused reading frame alterations, suggesting sequencing errors, were
92 removed from the dataset.

93 *2.1.2 Recombination analysis*

94 The whole dataset was tested for recombination using two programs based on different approaches:
95 RDP4 (Martin, et al, 2010) and GARD (Kosakovsky Pond, et al, 2006). To decrease the
96 computational burden, the dataset was reduced using CD-HIT (Li and Godzik, 2006) to cluster
97 together sequences with a nucleotide identity threshold of 99% and a single sequence representative

of each cluster was selected. GARD analysis was performed using the program implemented in datamonkey (http://www.datamonkey.org/) and, to cope with the circular structure of the PCV2 genome, the complete genome alignment was concatenated twice. The RDP4 settings for each method were adjusted to account for the dataset features according to the RDP manual recommendations. In particular RDP, GENECONV, Chimaera and 3Seq were used in a primary scan while the full set of available methods was used for the analysis refinement. Only recombination events detected by more than 2 methods with a significance value lower than 10^{-5} (p-value $<10^{-5}$) and Bonferroni correction were accepted. The non-recombinant sequences as well as those sharing recombination events were split into separate datasets and expanded to their original size.

2.1.3 Genotyping and database preparation

The non-recombinant sequences were classified into genotypes PCV2a, PCV2b, PCV2c or PCV2d according to Franzo et al. 2015 (Franzo, et al, 2015b).

The most appropriate nucleotide substitution model was selected according to the results of the Akaike information criterion (AIC) score calculated using JModel Test 2.1.2 (Darriba, et al, 2012). A phylogenetic tree was reconstructed using the Maximum likelihood (ML) approach implemented in PhyML (Guindon, et al, 2010). The best tree search method included the combination of two branch swapping algorithms: nearest neighbor interchange (NNI) and subtree pruning and regrafting (SPR). The robustness of the monophyly of the taxa subsets was estimated with the fast non-parametric version of the aLRT (Shimodaira–Hasegawa [SH]-aLRT), developed and implemented in PhyML 3.0 (Anisimova, et al, 2011). On the basis of the recombination and phylogenetic analyses, sequences were divided into independent datasets, corresponding to different genotypes and CRFs (i.e. those including more than 30 sequences collected in two or more

121 countries). Every dataset was further divided in three regions, namely *ORF1*, *ORF2* and intergenic
122 region (obtained merging together the major and the minor intergenic regions) and a new alignment
123 was generated on each dataset. The coding regions were aligned at the amino acid level and then the
124 nucleotide sequences were back-translated using the MAFFT algorithm implemented in
125 TranslatorX (Abascal, et al, 2010).

126 *2.2 BEAST analysis*

127 The time to most recent common ancestor (tMRCA), substitution rates and population dynamics
128 were jointly estimated using a Bayesian serial coalescent approach implemented in BEAST 1.8.1
129 (Drummond and Rambaut, 2007). For this purpose, datasets for each genotype and each CRF were
130 analyzed independently. Each region (i.e. *ORF1*, *ORF2* and intergenic regions) were allowed a
131 different substitution model and molecular clock. Substitution and clock models were respectively
132 selected according to the results of the Bayesian Information Criterion (BIC) score calculated using
133 JModel Test 2.1.2 and to the Bayes Factor (BF) value, calculated through estimation of marginal
134 likelihood of the different models using the path sampling (PS) and stepping stones (SS) methods
135 (Baele, et al, 2012). To reconstruct population dynamics over time, the non-parametric skyline
136 model was selected (Drummond, et al, 2005). The timing of viral dispersal patterns among
137 countries were estimated with the same program using the discrete state phylogeographic approach
138 described by Lemey et al. (Lemey, et al, 2009a). This model allows character mapping in natural
139 time scale, under a molecular clock assumption and accounting for population size changes while
140 integrating them over phylogenetic uncertainty. Additionally, the implementation of the Bayesian
141 stochastic search variable selection (BSSVS) allowed a BF test that identified the most
142 parsimonious description of the spreading process. For this purpose, to avoid over-parametrization
143 of the model and achieve as much information as possible two models were used. The first one,
144 based on the complete datasets implemented a symmetric substitution model with BSSVS

considering each country as a discrete state. For the second analysis, countries were grouped in macro-areas (i.e. Asia_other than China, China, Europe, Oceania, North_America, South_Africa and South_America) and the total number of sequences was reduced by randomly selecting strains from each region to obtain a more balanced (comparable number of sequences) genotype and CRF datasets in terms of geographic origin. The best substitution model (i.e. symmetric vs asymmetric, with BSSVS) was selected considering the BF value, calculated through estimation of marginal likelihood of the different models using the path sampling (PS) and stepping stones (SS) methods (Baele, et al, 2012). Migration rates among countries were considered non-zero (i.e. well supported) when the BF, calculated using SPREADv1.0.6 was greater than 10. At least two independent runs of 200 million generations were performed sampling trees and parameters every 20 thousand generations. Tree, log and location rate files, after removal of the burn-in, were merged using log-combiner (BEAST package). The run results were accepted only if Estimate Sample Size exceeded 200 and mixing and convergence among different runs, evaluated by visually inspecting the trace plots using Tracer 1.5, were adequate. Parameter estimation was summarized in terms of mean and 95% Highest Posterior Density (95HPD). A similar approach was performed on all non-recombinant PCV2 sequences (i.e. PCV2a, PCV2b, PCV2c and PCV2d) to estimate the time to MRCA, substitution rates and population dynamics of PCV2 as a whole.

2.3 Selective pressures

The selective pressure on the viral proteins was estimated using different methods based on the ratio between non-synonymous and synonymous substitution rates (dN/dS). A dN/dS higher, equal or lower than 1 suggests diversifying, neutral and purifying selection, respectively.

Pervasive diversifying/purifying selection was estimated using SLAC, FEL and FUBAR method while episodic diversifying selection was evaluated using MEME (Kosakovsky Pond and Frost,

2005; Murrell, et al, 2013; Murrell, et al, 2012). The significance value was set to $p<0.05$ for the FEL and MEME methods and to $p<0.1$ for SLAC, which claims to be more conservative (Lemey, et al, 2009b). The results of FUBAR, based on a Bayesian approach, were accepted when the posterior probability was greater than 0.9. Biochemical properties driving substitutions at a given site were investigated using PRIME (http://www.datamonkey.org/help/citations.php), assuming the Conant-Stadler amino-acid property-specific distance measures. The Significance level was set to p-value < 0.05.

The action of selective pressures was compared among different genes using the *dNdSDistributionComparison.bf* implemented in HyPhy (Pond et al., 2005). Differences in the site-by-site selection patterns among different genotypes were investigated for each gene using the batch files *CompareSelectivePressure.bf* implemented in the same program. Ancestral state reconstruction of per site amino acid sequence was performed using the maximum likelihood approach of the *ape* package implemented in R (Paradis, et al, 2004) (see Supplementary data 3 in (Franzo, et al, 2016)).

3. Results

3.1 Datasets

A total of 898 sequences, sampled from 32 countries between 1980 and 2014, were included in the final dataset (see Supplementary data 1 in (Franzo, et al, 2016)). A total of 304 out of 898 (33%) sequences were identified as recombinants. Local recombination hot spots, defined using a permutation approach, were identified with a 99% confidence threshold in the intergenic regions while only a non-significant increase in the number of recombination sites were found in the middle of *ORF1* and *ORF2* genes (Fig. 1a).

These results were confirmed by GARD, which identified as recombination breakpoint positions 429 (*ORF1*), 995, 1190 (within the shorter intergenic region) and 1728 (within the longer intergenic region).

Recombinant sequences were classified in 15 different recombination events; 13 of those included more than one sequence and 9 were present in more than one country (Table 1). Sequences sharing recombination breakpoints are subsequently referred to as circulating recombinant forms (CRFs).

The non-recombinant sequences were classified into genotypes PCV2a, PCV2b and PCV2d, comprising 63, 310 and 217 sequences, respectively. PCV2a comprised sequences collected between 1993 and 2013 from 12 countries; the PCV2b dataset included strains collected between 2000 and 2013 from 25 countries; finally, PCV2d was detected between 1999 and 2014 in 13 countries. Due to the limited number of sequences (n = 4) and to the marginal epidemiological role of PCV2c, it was excluded from the genotype specific analyses. Three CRFs were also identified including 64, 141 and 45 sequences. CRF01 (recombination event 1) comprised sequences collected between 2000 and 2014 from 14 countries; CRF02 (recombination event 2) was detected in 7 countries between 2004 and 2014, while CRF03 (recombination event 9) was sampled between 2003 and 2013 in 13 countries. Sequence alignments are provided in Supplementary data 2 of (Franzo, et al, 2016)

3.2 Population dynamics and phylogeography

The time to most recent common ancestor (tMRCA), substitution rate, population dynamics and phylogeography were estimated for PCV2a, b and d and for the three CRFs. PCV2a and PCV2d tMRCAs were predicted towards the end of the 1950s, while PCV2b tMRCA was reported in the middle of the 70s (Table 2).

Comparable results were obtained when the tMRCA of all genotypes was estimated jointly in the same analysis, except for PCV2a which had a tMRCA predicted to be during the first half of the

214 past century (i.e. 1927; 95% Highest Posterior Density (HPD): 1879-1963). On the other hand, the
215 estimates of PCV2b (1988; 95%HPD: 1968-1997) and PCV2d (1957; 95%HPD: 1909-1984)
216 overlapped the estimates obtained with the single genotype datasets (Fig. 2 and Supplementary
217 figure 1). The origin of PCV2c was estimated during the 60s (i.e. 1964; 95%HPD: 1933-1977). All
218 regions of the genome showed a similar high substitution rate for each genotype (range of the mean
219 values: $3.05x10^{-4}$ - $1.59x10^{-3}$ substitutions per site per year subs/site/year) and the recombinant
220 forms (range of the mean values: $3.22x10^{-4}$ - $1.73x10^{-3}$ subs/site/year)(Table 2).
221 The analysis of the population dynamics revealed a pattern shared among the different genotypes.
222 In particular all of them were characterized by an initial growth followed by stabilization (logistic-
223 like pattern) and decrease (Fig. 3). Remarkably, recombinant clusters displayed similar trends,
224 characterized by a more or less rapid increase followed by stabilization (CRF01 and CRF03) or
225 decline (CRF02) (Fig. 3).
226 Despite the tMRCAs of PCV2 being predicted several decades before the appearance of the PCV2-
227 SD, relative genetic diversity remained low until the middle of the 1990s when PCV2a started to
228 rise (Fig. 3). At the end of that decade and in particular since the beginning of the new millennium,
229 a sharp rise in PCV2b was observed, which coincided with a decline of PCV2a over the same time
230 period. A comparable phenomenon was observed from about 2006, when an increase in PCV2d
231 genetic diversity - and possibly CRF02 and CRF03 as well - mirrored a PCV2b decline (Fig. 3).
232 The results obtained with the reduced dataset were comparable (overlapping 95%HPD) in terms of
233 tMRCA, evolutionary rates and population dynamics (data not shown).
234 Phylogeographic analysis revealed the worldwide distribution of all genotypes and CRFs, as well as
235 the presence of several well supported migration routes among countries (Fig. 4 and Supplementary
236 figure 2). Due to the limited number of countries included in the database of PCV2a based on
237 complete genomes, a broader database of 83 PCV2a *ORF2* sequences was assembled
238 (Supplementary data 1 of Franzo, et al, 2016). This dataset allowed the estimation of a tMRCA,

substitution rates and population dynamics fully comparable with those obtained with the complete genome dataset. Overall, 19, 20 and 8 rates yielded a Bayesian factor (BF) >10 for PCV2a, PCV2b and PCV2d, respectively. Even if different countries were sometimes involved, a common pattern was observed for all genotypes, which was characterized by dense interconnections both within and between continents. Several European and American countries seemed to be involved in independent intercontinental transmissions. In contrast, the intra- and intercontinental relations of Asia were, with few exceptions, mediated through China. A similar network, with less dense interconnection, was observed considering recombinant forms (Fig. 4 and Supplementary figure 2). When spreading patterns were analyzed by macro-areas, an asymmetric migration model was a better fit to the data only in the case of PCV2a (Supplementary Table 1). Transmission patterns estimated using these datasets revealed the same main trends previously described. However, due to the aggregation of several countries in macro-areas more contacts were inferred between Asia other than China and other regions (Supplementary figure 3 and 4).

3.3 Selective pressures

Analysis of the selective pressures revealed differences among genes at both the global and site by site levels. The *ORF3* gene displayed the highest dN/dS ratio (mean = 1.23, 95% CI = 1.08-1.39), followed by *ORF4* gene (mean = 0.60, 95% CI = 0.49-0.72), *Rep'* gene (mean = 0.28, 95% CI = 0.25-0.32), *Rep* gene (mean = 0.24, 95% CI = 0.21-0.26) and *Cap* gene (0.23, 95% CI = 0.21-0.25).

Within the *Cap* gene, 18 codons were reported under positive selection at the significance threshold used in this study. All but 3 of them were identified by MEME, while only 5 by more than one method (Table 3).

All the identified sites are exposed on the external surface of the capsid (Supplementary figure 6).

Within the *Rep* gene, seven codon positions were under diversifying selection, 4 located in the

region overlapping the *ORF3*-4 gene and 1 in the region overlapping the Rep'. The *ORF1* gene was mainly under purifying selection with the exception of the region overlapping *ORF3, ORF4* and the C-terminal (overlapping with the Rep' region) (Fig. 1 and Supplementary figure 6). The *Rep'* gene, as expected, displayed the same pattern as *Rep* in the N-terminal region, while in the C-terminal was under neutral selection (mean dN/dS = 0.77319, 95% CI = 0.587572-0.994536). The *ORF3* revealed the strongest positive selection pattern (11 significant sites out of 104), and *ORF4* presented 3 positively selected sites (Table 3). The comparison of selective forces acting on different genes (*Rep, Rep', ORF2, ORF3* and *ORF4*) demonstrated that models allowing independent parameters for different datasets (i.e. genes) were significantly better (p-value<0.05) than those constraining identical selective forces, proportion of selected sites and selective regimes (i.e. both proportion and selective strength). The *Rep* protein was analyzed using a similar approach, dividing it into different parts corresponding to: 1) the N-terminal part (without overlapping with other genes), 2) the region overlapping with the *ORF3* gene, 3) the region overlapping with the *ORF4* gene, and 4) the C-terminal (overlapping with the C terminal of the *Rep'*). A significant difference was found in the distribution of sites under selection between the non-overlapping part of Rep and the overlapping part with Rep' (p-value<0.001). The N-terminal part of the *Rep'* overlapping with *Rep* but in a different reading frame (i.e. Rep' is coded +2) is extremely conserved as demonstrated by nucleotide sequence entropy analysis (Supplementary figure 7). No differences were identified at a significance level lower than p-value<0.05 between the N-terminal part of the *Rep* protein and the region overlapping with *ORF3* as well as among the regions of the *Rep* protein overlapping with *ORF3, ORF4* and *Rep'*.

When divergence in the site-by-site selection patterns was evaluated, only *Cap* showed a large number of sites with differences among the genotypes. While most of the codons displayed different dN/dS only between one pair of genotypes, codons 30, 32, 79, 169, 191, 195 and 200 were under dissimilar selective pressure for different genotype combinations (Supplementary figure 8).

The ancestral state reconstruction revealed a remarkable variability of PCV2a *Cap* over time with major clades displaying different amino-acid (aa) profiles (Supplementary figure 9). On the contrary PCV2b appeared to be more homogenous and the aa variability affected mainly the terminal tips (data not shown). PCV2d revealed a particular scenario with the vast majority of the amino acid changes separating two major groups of strains, estimated to have occurred between the beginning and the middle of the 90s (Supplementary figure 10). Some of these amino acid mutations affected epitopic regions such as positions 59, 63 and 206.

4. Discussion

Currently a comprehensive understanding of PCV2 molecular epidemiology and evolution is limited due to: 1) the lack of data about its history as it has only been identified recently and it likely has an ancient origin (Firth *et al* ., 2009 and present study) and 2) the limited focus of many studies to date, both epidemiological and experimental. The present extensive study uses a phylodynamics approach to reconstruct the history, population dynamics and evolutionary forces acting on PCV2, which have and will continue to shape its epidemiology.

The origin of PCV2 was estimated between the middle of the 19th and the first half of the 20th century (mean=1883). Nevertheless, relative genetic diversity (i.e. Ne· t) remained low (Fig. 3) until the middle of the 1990s when a first rise was estimated, and coincided with the first reports of PCV2-SD outbreaks (Harding, et al, 1998; Harding, 1996). Despite passing undetected, PCV2 differentiated into several genotypes during the 20th century. The first genotype to diversify was PCV2a, whose tMRCA was estimated at the beginning of the 20th century, followed by PCV2d (~ 1950), PCV2c (~ 1960) and finally by PCV2b (~1980). Some epidemiological evidences support the results of our estimation: PCV2a was first identified in 1962 (Jacobsen, et al, 2009), within the

95%HPD of the tMRCA estimated in the present study. Similarly, PVC2b was already present in Switzerland in 1979 (Wiederkehr, et al, 2009), in agreement with the Bayesian estimation and supporting a slightly older origin compared to a previous estimation of 1989 (1980-1995) (Firth, et al, 2009). The most unexpected result is related to the early origin of PCV2d in the 50s, which is supported by the independent genotype analysis as well as the complete PCV2 analysis; however, it is in strong contrast to a recent study of PCV2d (i.e. beginning of 90s) (Xiao, et al, 2015b). The higher numbers of strains used in the present study (resulting in higher genetic variability) could explain the difference in the results.

The population dynamics reconstruction shows an initial rise of PCV2a diversity during the middle of the 90s that is followed by a rise in PCV2b, which rapidly supersedes the former about 10 years later (Fig. 3). This well-known phenomenon, described also by classical epidemiological studies (Dupont, et al, 2008; Carman, et al, 2008; Cheung, et al, 2007; Cortey, et al, 2011; Grau-Roma, et al, 2008; Wang, et al, 2009) and known as "genotype shift", had a remarkable impact on the swine industry. The rise of PCV2a and PCV2b broadly corresponds to the first descriptions of PCV2-SD around the world and to a dramatic change in the frequency and severity of clinical outbreaks, respectively (Cheung, et al, 2007; Carman, et al, 2008; Gagnon, et al, 2007; Dupont, et al, 2008). The results of the present study confirm what has recently been suggested by Xiao et al., (2015), namely the current occurrence of a second major "genotype shift", characterized by the emergence and spread of PCV2d, which appears to be replacing PCV2b (Fig 3).

A number of factors could explain why multiple genotypes were able to co-circulate undetected for decades and suddenly rise in succession as major threats for the swine industry. Firstly, genotypes that have not previously circulated in a swine population may replicate better than the most prevalent genotype to which the swine population is highly protected (Opriessnig, et al, 2013) resulting in the rise and fall of different genotypes. Additionally, the role of vaccines cannot be

ignored as the currently available commercial vaccines based on PCV2a may induce a reduced cross-protection against PCV2d (Seo, et al, 2014; Xiao, et al, 2012), potentially favoring the rise of the latter. However, the cross-protection efficacy of the current vaccines is a contentious issue (Beach and Meng, 2012; Chae, 2012). PCV2 was found to display an extremely high substitution rate, within the order of magnitude of 10^{-3}-10^{-4} substitutions/site/year, comparable to those of RNA viruses (Duffy et al, 2008). Such remarkable substitution rate can represent a substrate for selective pressures to act. Accordingly to expectations several sites in the capsid protein were demonstrated to be under diversifying selection. Since the vast majority of sites were detected by MEME and not by the other methods suggests that the action of the selective pressures was non-pervasive over time but acted throughout selective burst. Interestingly, almost all (17 out of 18) of the positively selective sites were located within recognized epitopic regions (Trible, et al, 2012; Lekcharoensuk, et al, 2004; Mahe, et al, 2000; Ge, 2013). Concordantly, many of the positions in which radical changes in chemical-physical properties were favored over conservative ones were predicted within epitopes (Supplementary table 2 and Supplementary figure 6). Finally, the Cap gene was the one with the highest number of codons located in epitope regions, characterized by different selective pressures acting on the different genotypes (Supplementary figure 8) which could be attributed to escape from the immune response during the prolonged circulation in a highly immune population. Accordingly, PCV2d most recently raised clade differs from the ancient one in some amino-acids located in epitopic regions (Supplementary figure 10) .

Overall, these results support the pivotal role of the immune derived selective pressure (natural infection and vaccine induced responses) on the capsid and consequently the role of its variability and progressive evolution in affecting PCV2 strains and genotypes fitness and spread.

Population dynamic patterns comparable to that of major genotypes were observed for three CRFs, which may have originated as early as the 1970s and are distributed worldwide, supporting the idea

358 of a rapid viral spread in a partially naïve population followed by increased population immunity
359 that counteracts further expansion. Interestingly, the vast majority of recombination breakpoints
360 were located within the intergenic regions (Fig. 1) and, to a lesser extent, within *ORF1* where they
361 are less likely to disrupt protein folding and/or affect the fitness of the virus. Globally, these
362 findings suggest that, at least in some circumstances, recombination can lead to the emergence of
363 strains characterized by high fitness, which are able to compete and potentially displace parental
364 strains.

365 The second major factor that has been claimed to favor the PCV2 rise is the development of modern
366 farming. Changes in the swine farming practice has led to huge and highly interconnected animal
367 populations (Segales, et al, 2013) enabling the spread of genotypes and CRFs worldwide that may
368 have been previously geographically isolated. In the present study, the network of spreading pattern
369 among countries was estimated using a Bayesian approach, which allowed reconstructing migration
370 rates and test for their significance accounting for the phylogenetic uncertainty. China seems to
371 have played a major role in the ancient spread of all major PCV2 genotypes (Supplementary figure
372 1). Although this hypothesis has already been proposed for PCV2d (Xiao, et al, 2015b) and it
373 cannot be excluded also for the other genotypes, it seems unlikely because China is mainly a swine
374 importer (Supplementary figure 5 and Vidigal *et al* ., 2012) and livestock and meat products
375 importation from China is forbidden in many countries (Wang, 2006). When macro-areas and a
376 more balanced dataset were considered a more variable scenario was discovered, with PCV2a,
377 PCV2d and CRF03 most likely originating in Europe, PCV2b and CR02 in China and CRF01 in
378 North_America. This is particularly interesting because the first PCV2d sequence was actually
379 sampled in Europe (Wiederkehr et al., 2009), supporting obtained results. However, even with this
380 approach and,with the only exception of PCV2a and CRF02, the root location posterior probability
381 was low (~0.5) (Supplementary figure 4) revealing an uncertainty that is not surprising considering
382 the large time frame between the estimated tMRCA and the oldest available sequences. Long

383 branches and the lack of historical data make it difficult to infer the spatial history of older viral
384 lineages with confidence. Additionally, the long branch length is likely to conceal additional spatial
385 movements between multiple locations during this time-frame (Nelson, et al, 2015).

386 Accounting for these factors, migration patterns shared by the different genotypes and CRFs have
387 been evaluated in terms of "contact" among countries, avoiding over-interpretation of their timing
388 and directionality. Altogether, three major "mixing" nuclei of local transmission (i.e. North
389 America, Europe and Asia) were identified (Fig. 4 and Supplementary figure 2). The strong linkage
390 between North American countries have already been involved in the spread of other pathogens like
391 PRRSV and swine influenza A viruses (swIAV) (Nelson, et al, 2015; Shi, et al, 2013). Similarly,
392 the European pig population has been considered as a meta-population with regards to swIAV and
393 several studies have reported the relevant pathogen spread among European countries (Franzo, et al,
394 2015c; Nguyen, et al, 2014; Artois, et al, 2001). Unlike in Europe, where several countries
395 participated in PCV2 dissemination, China seems to be the major PCV2 source for other Asian
396 countries (Fig. 4). Interestingly, such scenario differs from that described for swIAV, which seems
397 to be characterized by a negligible role of China in viral spread in Asia (Nelson, et al, 2015). It is
398 possible to hypothesize that the different biology of the two viruses and in particular the long
399 persistence of PCV2 in asymptomatic animals influenced the different patterns of spread. Different
400 intercontinental "contacts", which mirror the main commercial routes of live pigs (Supplementary
401 figure 5), have been estimated to link the European, North American and Asian countries (Fig. 4).
402 While North America and Europe participated respectively in the introduction of PCV2a and
403 PCV2b to Oceania, Canada and the USA were confirmed to be respectively the primary source of
404 PCV2a and b in Central America (Cuba) and to South Africa (Pérez, et al, 2011; Vidigal, et al,
405 2012) (Fig. 4 and Supplementary figure 2).

These results emphasize the role of live swine trade in the spreading of PCV2 strains around the world and the current limitations of biosecurity measures. Considering the subtle and often asymptomatic/subclinical nature of PCV2 infection (Segalés, 2012) and the presence of several direct and indirect transmission routes (Rose, et al, 2012), a focused effort should be devoted to the implementation of rigorous control measures that must be based on a widespread laboratory diagnostics.

Finally, the results of the present study demonstrate that PCV2 evolution, besides environmental factors, is determined by its own biology and genome organization. Interestingly, the *ORF3* and *ORF4* genes were under strong diversifying selection even in regions that have not been reported to be the target of the immune response. These results were particularly surprising, since these genes are involved in host-virus interaction (Lv, et al, 2014). This pattern can probably be attributed to the presence of different *ORFs* coded in differ reading frames (i.e. reading frame 1,-1 and -3 for *ORF1*, *ORF3* and *ORF4*, respectively). This causes substitutions that are synonymous in *Rep* to be mainly non-synonymous in the other *ORFs*. Since Rep is fundamental for virus replication while *ORF3* and *ORF4* are not (Gao, et al, 2014; Karuppannan and Kwang, 2011), it is highly probable that only synonymous mutations in Rep (mainly non-synonymous on *ORF3* and *ORF4*) are tolerated while synonymous mutations in *ORF3* and *ORF4* (mainly non-synonymous in *ORF1*) are highly deleterious for viral fitness and are purged by natural selection. Further evidences of this hypothesis come from the extremely conserved overlapping region of Rep and Rep', both proteins being fundamental for viral replication, it is expected that synonymous mutations are not tolerated either in Rep or in Rep'.

5. Conclusions

In summary, this study strengthens the idea that the complex PCV2 biology and epidemiology cannot be attributed to one or a few factors. A combination of (i) the PCV2 genomic constraints due to gene organization and virulence, (ii) the efficacy of population host immune responses, (iii) the vaccination pressure, (iv) the competition among different genotypes and (v) the availability of favorable environmental conditions (e.g. huge animal populations, international trades, poor control measures, etc.) (Drew, 2011) all play a role in determining the PCV2 epidemiology and evolution.

6. Funding information

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

7. Acknowledgments

The authors wish to especially acknowledge Dr Pablo R. Murcia for the precious help during the study and the draft of the manuscript. M Cortey is granted by a JIN fellowship from the Spanish Ministry of Economy and Competitivity.

8. References

Abascal, F., Zardoya, R., and Telford, M.J., 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38, W7-W13.

Alarcon, P., Rushton, J., and Wieland, B., 2013. Cost of post-weaning multi-systemic wasting syndrome and porcine circovirus type-2 subclinical infection in England - an economic disease model. Prev. Vet. Med. 110, 88-102.

448 Anisimova, M., Gil, M., Dufayard, J., Dessimoz, C., and Gascuel, O., 2011. Survey of branch support
449 methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes.
450 Syst. Biol. 60, 685-699.

451 Artois, M., Delahay, R., Guberti, V., and Cheeseman, C., 2001. Control of infectious diseases of wildlife in
452 Europe. The Veterinary Journal 162, 141-152.

453 Baele, G.F., Lemey, P.F., Bedford, T.F., Rambaut A FAU - Suchard, Marc,A., FAU, S.M., and Alekseyenko,
454 A.V., 2012. Improving the accuracy of demographic and molecular clock model comparison while
455 accommodating phylogenetic uncertainty. Molecular biology and evolution 29(9):2157-2167

456 Beach, N.M. and Meng, X.-., 2012. Efficacy and future prospects of commercially available and
457 experimental vaccines against porcine circovirus type 2 (PCV2). Virus Res. 164, 33-42.

458 Carman, S., Cai, H.Y., DeLay, J., Youssef, S.A., McEwen, B.J., Gagnon, C.A., Tremblay, D., Hazlett, M.,
459 Lusis, P., Fairles, J., et al, 2008. The emergence of a new strain of porcine circovirus-2 in Ontario and
460 Quebec swine and its association with severe porcine circovirus associated disease 2004-2006. Can. J. Vet.
461 Res. 72, 259-268.

462 Chae, C., 2012. Commercial porcine circovirus type 2 vaccines: Efficacy and clinical application. Veterinary
463 Journal 194, 151-157.

464 Cheung, A.K., Lager, K.M., Kohutyuk, O.I., Vincent, A.L., Henry, S.C., Baker, R.B., Rowland, R.R., and
465 Dunham, A.G., 2007. Detection of two porcine circovirus type 2 genotypic groups in United States swine
466 herds. Arch. Virol. 152, 1035-1044.

467 Cortey, M., Pileri, E., Sibila, M., Pujols, J., Balasch, M., Plana, J., and Segalés, J., 2011. Genotypic shift of
468 porcine circovirus type 2 from PCV-2a to PCV-2b in Spain from 1985 to 2008. Veterinary Journal 187, 363-
469 368.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D., 2012. JModelTest 2: More models, new heuristics and parallel computing. Nature Methods 9, 772.

Drew, T.W., 2011. The emergence and evolution of swine viral diseases: to what extent have husbandry systems and global trade contributed to their distribution and diversity? Rev. Sci. Tech. 30, 95-106.

Drummond, A.J. and Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Rambaut, A., Shapiro, B., and Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185-1192.

Duffy, S., Shackelton, L.A., and Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nature Reviews Genetics 9, 267-276.

Dupont, K., Nielsen, E.O., Baekbo, P., and Larsen, L.E., 2008. Genomic analysis of PCV2 isolates from Danish archives and a current PMWS case-control study supports a shift in genotypes with time. Vet. Microbiol. 128, 56-64.

Firth, C., Charleston, M.A., Duffy, S., Shapiro, B., and Holmes, E.C., 2009. Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. J. Virol. 83, 12813-12821.

Fort, M., Sibila, M., Perez-Martin, E., Nofrarias, M., Mateu, E., and Segalés, J., 2009. One dose of a porcine circovirus 2 (PCV2) sub-unit vaccine administered to 3-week-old conventional piglets elicits cell-mediated immunity and significantly reduces PCV2 viremia in an experimental model. Vaccine 27, 4031-4037.

Franzo, G., Cortey, M., Segalés, J., Hughes, J., & Drigo, M. (2016). Phylodynamic analysis of porcine circovirus type 2: Methodological approach and datasets. *Data in Brief*, *8*, 549-552.

Franzo, G., Cortey, M., de Castro, Alessandra Marnie Martins Gomes, Piovezan, U., Szabo, M.P.J., Drigo, M., Segalés, J., and Richtzenhain, L.J., 2015a. Genetic characterisation of porcine circovirus type 2 (PCV2)

492 strains from feral pigs in the Brazilian Pantanal: an opportunity to reconstruct the history of PCV2 evolution.
493 Vet. Microbiol 178, 158-162.

494 Franzo, G., Cortey, M., Olvera, A., Novosel, D., De Castro, A.M., Biagini, P., Segalés, J., and Drigo, M.,
495 2015b. Revisiting the taxonomical classification of Porcine Circovirus type 2 (PCV2): still a real challenge.
496 Virology journal 12, 131.

497 Franzo, G., Tucciarone, C.M., Dotto, G., Gigli, A., Ceglie, L., and Drigo, M., 2015c. International trades,
498 local spread and viral evolution: The case of porcine circovirus type 2 (PCV2) strains heterogeneity in Italy.
499 Infection, Genetics and Evolution 32, 409-415.

500 Gagnon, C.A., Tremblay, D., Tijssen, P., Venne, M.H., Houde, A., and Elahi, S.M., 2007. The emergence of
501 porcine circovirus 2b genotype (PCV-2b) in swine in Canada. Can. Vet. J. 48, 811-819.

502 Gao, Z., Dong, Q., Jiang, Y., Opriessnig, T., Wang, J., Quan, Y., and Yang, Z., 2014. ORF4-protein deficient
503 PCV2 mutants enhance virus-induced apoptosis and show differential expression of mRNAs in vitro. Virus
504 Res. 183, 56-62.

505 Ge, M., 2013. Epitope screening of the PCV2 Cap protein by use of a random peptide-displayed library and
506 polyclonal antibody. Virus Res. 177, 103-107.

507 Ge, X., Wang, F., Guo, X., and Yang, H., 2012. Porcine circovirus type 2 and its associated diseases in
508 China. Virus Res. 164, 100-106.

509 Grau-Roma, L., Crisci, E., Sibila, M., López-Soria, S., Nofrarias, M., Cortey, M., Fraile, L., Olvera, A., and
510 Segalés, J., 2008. A proposal on porcine circovirus type 2 (PCV2) genotype definition and their relation with
511 postweaning multisystemic wasting syndrome (PMWS) occurrence. Vet. Microbiol. 128, 23-35.

512 Grau-Roma, L., Fraile, L., and Segales, J., 2011. Recent advances in the epidemiology, diagnosis and control
513 of diseases caused by porcine circovirus type 2. Vet. J. 187, 23-32.

514 Guindon, S., Dufayard, J.-., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O., 2010. New algorithms
515 and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst.
516 Biol. 59, 307-321.

517 Harding, J., 1996. Postweaning multisystemic wasting syndrome: Preliminary epidemiology and clinical
518 findings. Proc Western Can Assoc Swine Pract . Saskatoon, Saskatchewan.October 1996:21.

519 Harding, J.C., Clark, E.G., Strokappe, J.H., Willson, P.I., and Ellis, J.A., 1998. Postweaning multisystemic
520 wasting syndrome: epidemiology and clinical presentation. Swine health and production 6, 249-256.

521 He Jialing, J., Cao, J., Zhou, N., Jin, J., Wu, J., and Zhou, J., 2013. Identification and functional analysis of
522 the novel ORF4 protein encoded by porcine circovirus type 2. J. Virol. 87, 1420-1429.

523 Jacobsen, B., Krueger, L., Seeliger, F., Bruegmann, M., Segalés, J., and Baumgaertner, W., 2009.
524 Retrospective study on the occurrence of porcine circovirus 2 infection and associated entities in Northern
525 Germany. Vet. Microbiol. 138, 27-33.

526 Karuppannan, A.K. and Kwang, J., 2011. ORF3 of porcine circovirus 2 enhances the in vitro and in vivo
527 spread of the of the virus. Virology 410, 248-256.

528 Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
529 improvements in performance and usability. Mol. Biol. Evol. 30, 772-780.

530 Khayat, R., Brunn, N., Speir, J.A., Hardham, J.M., Ankenbauer, R.G., Schneemann, A., and Johnson, J.E.,
531 2011. The 2.3-angstrom structure of porcine circovirus 2. J. Virol. 85, 7856-7862.

532 Kosakovsky Pond, S.L. and Frost, S.D., 2005. Not so different after all: a comparison of methods for
533 detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208-1222.

534 Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D., 2006. GARD: a genetic
535 algorithm for recombination detection. Bioinformatics 22, 3096-3098.

536 Lefeuvre, P., Lett, J.M., Varsani, A., and Martin, D.P., 2009. Widely conserved recombination patterns
537 among single-stranded DNA viruses. J. Virol. 83, 2697-2707.

538 Lekcharoensuk, P., Morozov, I., Paul, P.S., Thangthumniyom, N., Wajjawalku, W., and Meng, X.J., 2004.
539 Epitope mapping of the major capsid protein of type 2 porcine circovirus (PCV2) by using chimeric PCV1
540 and PCV2. J. Virol. 78, 8135-8145.

541 Lemey, P., Rambaut, A., Drummond, A.J., and Suchard, M.A., 2009a. Bayesian phylogeography finds its
542 roots. PLoS Comput. Biol. 5, e1000520.

543 Lemey, P., Salemi, M., and Vandamme, A. (Eds.), 2009b. *The Phylogenetic Handbook: A Practical*
544 *Approach to Phylogenetic Analysis and Hypothesis Testing.* CAMBRIDGE UNIVERSITY PRESS.

545 Li, W. and Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or
546 nucleotide sequences. Bioinformatics 22, 1658-1659.

547 Lv, Q.Z., Guo, K.K., and Zhang, Y.M., 2014. Current understanding of genomic DNA of porcine circovirus
548 type 2. Virus Genes 49, 1-10.

549 Mahe, D., Blanchard, P., Truong, C., Arnauld, C., Le Cann, P., Cariolet, R., Madec, F., Albina, E., and
550 Jestin, A., 2000. Differential recognition of ORF2 protein from type 1 and type 2 porcine circoviruses and
551 identification of immunorelevant epitopes. J. Gen. Virol. 81, 1815-1824.

552 Mankertz, A., Caliskan, R., Hattermann, K., Hillenbrand, B., Kurzendoerfer, P., Mueller, B., Schmitt, C.,
553 Steinfeldt, T., and Finsterbusch, T., 2004. Molecular biology of Porcine circovirus: analyses of gene
554 expression and viral replication. Vet. Microbiol. 98, 81-88.

555 Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefeuvre, P., 2010. RDP3: a flexible and
556 fast computer program for analyzing recombination. Bioinformatics 26, 2462-2463.

557 Monjane, A.L., Pande, D., Lakay, F., Shepherd, D.N., van der Walt, E., Lefeuvre, P., Lett, J.M., Varsani, A.,
558 Rybicki, E.P., and Martin, D.P., 2012. Adaptive evolution by recombination is not associated with increased
559 mutation rates in Maize streak virus. BMC Evol. Biol. 12, 252.

560 Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., and Scheffler, K.,
561 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol. Biol. Evol. 30,
562 1196-1205.

563 Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S.L., 2012.
564 Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8, e1002764.

565 Nawagitgul, P., Morozov, I., Bolin, S.R., Harms, P.A., Sorden, S.D., and Paul, P.S., 2000. Open reading
566 frame 2 of porcine circovirus type 2 encodes a major capsid protein. J. Gen. Virol. 81, 2281-2287.

567 Nelson, M.I., Viboud, C., Vincent, A.L., Culhane, M.R., Detmer, S.E., Wentworth, D.E., Rambaut, A.,
568 Suchard, M.A., Holmes, E.C., and Lemey, P., 2015. Global migration of influenza A viruses in swine.
569 Nature communications 6, 6696.

570 Nguyen, V., Kim, H., Moon, H., Park, S., Chung, H., Choi, M., and Park, B., 2014. A Bayesian
571 phylogeographical analysis of type 1 porcine reproductive and respiratory syndrome virus (PRRSV).
572 Transboundary and emerging diseases 61, 537-545.

573 Opriessnig, T., Gerber, P.F., Xiao, C., Mogler, M., and Halbur, P.G., 2014a. A commercial vaccine based on
574 PCV2a and an experimental vaccine based on a variant mPCV2b are both effective in protecting pigs against
575 challenge with a 2013 US variant mPCV2b strain. Vaccine 32, 230-237.

576 Opriessnig, T., O'Neill, K., Gerber, P.F., de Castro, A.M., Gimenéz-Lirola, L.G., Beach, N.M., Zhou, L.,
577 Meng, X., Wang, C., and Halbur, P.G., 2013. A PCV2 vaccine based on genotype 2b is more effective than a
578 2a-based vaccine to protect against PCV2b or combined PCV2a/2b viremia in pigs with concurrent PCV2,
579 PRRSV and PPV infection. Vaccine 31, 487-494.

580 Paradis, E., Claude, J., and Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R
581 language. Bioinformatics 20, 289-290.

582 Pérez, L.J., de Arce, H.D., Cortey, M., Domínguez, P., Percedo, M.I., Perera, C.L., Tarradas, J., Frías, M.T.,
583 Segalés, J., and Ganges, L., 2011. Phylogenetic networks to study the origin and evolution of porcine
584 circovirus type 2 (PCV2) in Cuba. Vet. Microbiol. 151, 245-254.

585 Pond, S.L.K., Frost S.D, Muse S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21,
586 676-679.

587 Rose, N., Opriessnig, T., Grasland, B., and Jestin, A., 2012. Epidemiology and transmission of porcine
588 circovirus type 2 (PCV2). Virus Res. 164, 78-89.

589 Segalés, J., 2012. Porcine circovirus type 2 (PCV2) infections: Clinical signs, pathology and laboratory
590 diagnosis. Virus Res. 164, 10-19.

591 Segales, J., Kekarainen, T., and Cortey, M., 2013. The natural history of porcine circovirus type 2: from an
592 inoffensive virus to a devastating swine disease? Vet. Microbiol. 165, 13-20.

593 Seo, H.W., Park, C., Kang, I., Choi, K., Jeong, J., Park, S., and Chae, C., 2014. Genetic and antigenic
594 characterization of a newly emerging porcine circovirus type 2b mutant first isolated in cases of vaccine
595 failure in Korea. Arch. Virol. 159, 3107-3111.

596 Shi, M., Lemey, P., Brar, M.S., Suchard, M.A., Murtaugh, M.P., Carman, S., D'Allaire, S., Delisle, B.,
597 Lambert, M., and Gagnon, C.A., 2013. The spread of type 2 porcine reproductive and respiratory syndrome
598 virus (PRRSV) in North America: a phylogeographic approach. Virology 447, 146-154.

599 Stevenson, L.S., Gilpin, D.F., Douglas, A., McNeilly, F., McNair, I., Adair, B.M., and Allan, G.M., 2007. T
600 lymphocyte epitope mapping of porcine circovirus type 2. Viral Immunol. 20, 389-398.

601 Timmusk, S., Wallgren, P., Brunborg, I.M., Wikström, F.H., Allan, G., Meehan, B., McMenamy, M.,
602 McNeilly, F., Fuxler, L., and Belák, K., 2008. Phylogenetic analysis of porcine circovirus type 2 (PCV2)
603 pre-and post-epizootic postweaning multisystemic wasting syndrome (PMWS). Virus Genes 36, 509-520.

604 Trible, B.R., Suddith, A.W., Kerrigan, M.A., Cino-Ozuna, A., Hesse, R.A., and Rowland, R.R.R., 2012.
605 Recognition of the different structural forms of the capsid protein determines the outcome following
606 infection with porcine circovirus type 2. J. Virol. 86, 13508-13514.

607 Vidigal, P.M., Mafra, C.L., Silva, F.M., Fietto, J.L., Silva Junior, A., and Almeida, M.R., 2012. Tripping
608 over emerging pathogens around the world: a phylogeographical approach for determining the epidemiology
609 of Porcine circovirus-2 (PCV-2), considering global trading. Virus Res. 163, 320-327.

610 Wang, F., Guo, X., Ge, X., Wang, Z., Chen, Y., Cha, Z., and Yang, H., 2009. Genetic variation analysis of
611 Chinese strains of porcine circovirus type 2. Virus Res. 145, 151-156.

612 Wang, R., 2006. China: pork powerhouse of the world. Adv.Pork Prod 17, 33-46.

613 Wiederkehr, D.D., Sydler, T., Buergi, E., Haessig, M., Zimmermann, D., Pospischil, A., Brugnera, E., and
614 Sidler, X., 2009. A new emerging genotype subgroup within PCV-2b dominates the PMWS epizooty in
615 Switzerland. Vet. Microbiol. 136, 27-35.

616 Xiao, C.T., Halbur, P.G., and Opriessnig, T., 2015a. Global molecular genetic analysis of porcine circovirus
617 type 2 (PCV2) sequences confirms the presence of four main PCV2 genotypes and reveals a rapid increase of
618 PCV2d. J. Gen. Virol. 96, 1830-1841

619 Xiao, C.T., Halbur, P.G., and Opriessnig, T., 2012. Complete genome sequence of a novel porcine circovirus
620 type 2b variant present in cases of vaccine failures in the United States. J. Virol. 86, 12469-12.

621 **Table 1. Count of strains belonging to each recombination event classified according to the**
622 **country of origin.**

Country	Recombination event															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Australia	1															1
Austria	4															4
Belgium									1							1
Brazil	1	1							1				2			6
Canada	9						1							1		11
China	13	102	2	1		2	1		7	2		1			1	132
Croatia							2		1		1				1	5
Denmark								1								4
France							1		3							4
Germany							2	1								3
Greece				1												1
Hungary	3						3									6
India							1									1
Italy	2	32		2		2	1		3			3				45
Japan	3															3
Korea	1															1
Lithuania		1							1							2
Malaysia									12							12
Netherlands									1							1
Portugal		1					5		3							9
Slovakia							1									1
South_Africa	1															1
South_Korea	8	1							10							19
Spain	3															3
Switzerland				1					1							3

Taiwan	1				4				1							6
USA	14			1				1								16
Vietnam		3														3
Total	64	141	2	6	4	4	19	3	45	2	1	4	2	1	2	304

Table 2. The time to Most Recent Common Ancestor (tMRCA) and substitution rates are reported for the major PCV2 genotypes and CRFs. All parameters have been estimated independently for each dataset.

Genotype	tMRCA(95%HPD)	Gene	Substitution per site per year (95%HPD)
PCV2a	1964 (1948-1974)	*ORF1*	5.45×10^{-4} (3.69×10^{-4}-7.21×10^{-4})
		ORF2	1.59×10^{-3} (9.8×10^{-4}-2.31×10^{-3})
		Intergenic	3.31×10^{-4} (3.18×10^{-4}-5.16×10^{-4})
PCV2b	1973 (1952-1996)	*ORF1*	3.12×10^{-4}(1.23×10^{-4}-4.87×10^{-4})
		ORF2	3.88×10^{-4}(1.6×10^{-4}-5.95×10^{-4})
		Intergenic	3.05×10^{-4}(1.18×10^{-4}-4.97×10^{-4})
PCV2d	1958 (1935-1979)	*ORF1*	4.89×10^{-4} (3.41×10^{-4}-6.53×10^{-4})
		ORF2	6.22×10^{-4} (4.47×10^{-4}-7.94×10^{-4})
		Intergenic	3.3×10^{-4} (2.26×10^{-4}-4.48×10^{-4})
CRF1	1973 (1957-1988)	*ORF1*	5.48×10^{-4} (3.05×10^{-4}-8.1×10^{-4})
		ORF2	1.07×10^{-3} (6.35×10^{-4}-1.58×10^{-3})
		Intergenic	4.42×10^{-4} (1.94×10^{-4}-6.94×10^{-4})
CRF2	1987 (1974-1987)	*ORF1*	5.83×10^{-4} (3.46×10^{-4}-8.61×10^{-4})
		ORF2	6.02×10^{-4} (3.61×10^{-4}-8.58×10^{-4})

		Intergenic	$3.22x10^{-4}$ ($1.65x10^{-4}$-$4.96x10^{-4}$)
		ORF1	$1.05x10^{-3}$ ($2.73x10^{-4}$-$1.89x10^{-3}$)
CRF3	1991 (1978-2001)	*ORF2*	$1.73x10^{-3}$ ($5.22x10^{-3}$-$10x10^{-3}$)
		Intergenic	$6.27x10^{-4}$ ($1.62x10^{-4}$-$1.15x10^{-3}$)

628

629 **Table 3. Codon positions detected under diversifying selection.**

Protein	Codon	SLAC dN-dS	SLAC p-value[a]	FEL dN-dS	FEL p-value[a]	MEME ω+	MEME p-value[a]	FUBAR dN-dS	FUBAR Post. Pr.[a]	N.Methods
Rep	77[b]	0.981	0.193	1.616	0.304	4.018	0.276	0.668	**0.921**	1
	91[b]	0.823	0.320	1.624	0.310	3.668	0.248	0.656	**0.943**	1
	146	0.623	0.345	1.354	0.255	77.615	**0.001**	0.482	0.900	1
	147	-0.415	0.888	-0.271	0.705	>100	**0.023**	-0.068	0.335	1
	150	0.219	0.480	0.050	0.964	62.402	**0.001**	0.052	0.431	1
	159	1.167	**0.097**	2.339	**0.031**	>100	**0.003**	0.710	**0.980**	4
	304	0.070	0.667	0.083	0.920	>100	**0.011**	-0.060	0.337	1
Rep'	77[b]	1.153	0.096	2.737	0.030	9.353	**0.034**	0.648	0.993	4
	138	0.419	0.409	0.834	0.123	>100	**0.009**	0.112	0.815	1
	163	0.877	0.170	1.803	0.080	4.707	0.090	0.574	0.981	1
	166	0.029	0.707	-0.182	0.748	>100	**0.016**	-0.057	0.322	1
	177	0.168	0.999	0.286	0.955	>100	**0.008**	-0.696	0.289	1
Cap	8	1.484	0.195	0.638	0.068	>100	0.066	0.181	**0.901**	1
	13[b]	1.150	0.270	0.464	0.114	>100	**0.034**	0.130	0.830	1
	30[b]	1.821	**0.028**	0.581	0.072	42.553	**0.048**	0.193	**0.927**	3
	59[b]	0.295	0.497	-0.138	0.801	>100	**0.000**	0.077	0.167	1
	63[b]	1.717	0.139	0.511	0.266	3.180	0.260	0.528	**0.921**	1
	68[b]	-1.225	0.911	-0.129	0.747	>100	**0.000**	-0.028	0.458	1
	72[b]	0.817	0.683	0.322	0.208	>100	**0.045**	0.070	0.704	1
	80[b]	1.160	**0.059**	0.494	**0.013**	>100	**0.013**	0.190	**0.948**	4
	88[b]	0.552	0.326	0.295	0.239	>100	**0.000**	0.053	0.646	1
	114[b]	-0.083	0.741	-0.111	0.544	>100	**0.006**	-0.110	0.121	1
	121[b]	1.015	0.149	0.366	**0.041**	>100	**0.025**	0.139	0.883	2
	130[b]	1.230	**0.090**	0.519	**0.017**	>100	**0.023**	0.193	**0.943**	4
	134[b]	-0.083	0.667	0.092	0.840	>100	**0.000**	0.033	0.579	1
	151[b]	1.325	**0.075**	0.463	0.128	4.281	0.119	0.147	0.876	1
	169[b]	1.644	0.138	0.558	0.339	71.949	**0.000**	0.476	0.788	1
	228[b]	0.382	0.652	0.126	0.409	>100	**0.004**	-0.060	0.327	1

	230[b]	0.082	0.626	0.069	0.748	>100	**0.023**	-0.014	0.423	1
	232[b]	2.325	**0.067**	0.873	**0.033**	>100	**0.021**	0.308	**0.961**	4
	19	2.076	**0.091**	2.991	**0.050**	>100	**0.047**	0.491	**0.956**	4
	29	2.436	**0.059**	3.429	**0.035**	>100	**0.027**	0.583	**0.973**	4
	41[b]	5.699	**0.004**	8.211	**0.003**	>100	**0.000**	1.668	**0.999**	4
	47[b]	1.338	0.141	2.530	**0.038**	>100	0.061	0.324	**0.909**	2
	60	1.749	0.132	2.350	0.053	>100	0.079	0.332	**0.922**	1
ORF3	62	2.202	0.124	3.533	0.068	>100	**0.041**	0.626	**0.962**	2
	63	1.748	0.132	3.211	0.055	>100	0.062	0.485	**0.941**	1
	65	0.906	0.539	1.031	0.259	>100	**0.022**	0.085	0.745	1
	78	2.457	**0.059**	4.335	**0.021**	>100	**0.024**	0.773	**0.981**	4
	87	2.295	0.197	3.255	0.106	>100	**0.044**	0.542	**0.949**	2
	103	7.112	**0.031**	9.576	0.074	5.533	**0.005**	3.103	**0.979**	3
	14	1.711	0.384	2.059	0.251	>100	**0.033**	0.345	**0.910**	2
ORF4	23	2.210	0.280	2.078	0.232	>100	**0.020**	0.404	**0.935**	2
	46	0.231	0.700	-1.622	0.575	>100	**0.043**	-0.061	0.571	1

Comparison between rates of non-synonymous and synonymous substitutions is expressed as dN–dS or ω (i.e. dN/dS). The statistical significance, highlighted in boldface, of positive over neutral selection is expressed in term of p-value (SLAC, FEL and MEME methods) or posterior probability (FUBAR method).

[a] Significance threshold was set to p-value<0,1,0,05,0,05 for SLAC, FEL and MEME methods and to Post. Pr.>0,9 for FUMAR method.

[b] codon position under positive selective pressure located in epitopic regions according to literature (Trible, et al, 2012; Lekcharoensuk, et al, 2004; Mahe, et al, 2000; Ge, 2013; Stevenson, et al, 2007;He Jialing, et al, 2013)

Figure captions

Fig 1. PCV2 genome organization, recombination breakpoints and selective pressures. a) Breakpoint Distribution Plot. All detectable breakpoint positions are indicated by small vertical

lines at the top of the graphs. A 300-nt window was moved along the alignments 1 nt at a time, and the count of breakpoints detected within the window region was plotted (solid lines). The white and dark gray areas indicate local 99% and 95% confidence thresholds respectively and they take into account local regional differences in sequence diversity that influence the ability of different recombination detection methods to identify recombination breakpoints. Regions where the solid line is above and below the 99% confidence threshold indicate recombination hot and cold spots, respectively. *ORF1* and *ORF2* have been highlighted in red and green.

b) Positive and negative dN/dS, colored respectively in green and red, have been reported for each ORF. To emphasize the interaction among *ORFs* encoded in the same region (represented in outer and inner circles according to their positive or negative sense), their dN/dS have been superimposed.

Fig 2. Maximum clade credibility phylogenetic tree. Maximum clade credibility phylogenetic tree (obtained through BEAST analysis) based on non-recombinant strains. Nodes displaying a posterior probability higher than 0.9 and between 0.9 and 0.7 are represented as black and grey circles, respectively.

Fig 3. PCV2 population dynamics plot. Relative genetic diversity (Ne·t) of different genotypes and CRFs over time.

Fig 4. PCV2 spreading and distribution over time.

a) Network displaying the spreading of PCV2 among different geographical regions. Edges connecting different countries within the same macro-areas have been represented with different curvature. Similarly, linkage within the same macro area is reported as separate loops. Edges are color coded according to genotype or CRF while the edge width is proportional to the Bayesian Factor (i.e. the statistical support) calculated for that migration route. A more detailed description of the countries involved is reported in Supplementary figure 2. b) Network displaying the pig trade among different macro-areas. Edges width is proportional to the pig trade (Log10 of head number).

A more detailed description of the international trades is reported in Supplementary figure 5. c) Estimated geographical distribution of different genotypes and CRF over time.

Supporting information captions

Supplementary figure 1. Bayesian maximum clade credibility phylogenetic trees of PCV2a, PCV2b and PCV2d. Tips are colored by the country of sampling, while the internal branches are colored based on the reconstructed ancestral state (country). Numbers at the nodes indicate the estimated mean age. Branch width is proportional to location posterior probability.

Supplementary figure 2. Maps of PCV2 genotype and CRF distribution and spreading routes.

Supplementary figure 3. Networks displaying the spreading of PCV2 genotypes and CRF among different geographical regions. Edges connecting different macro-areas are proportional to the Bayesian Factor (i.e. the statistical support) calculated for that migration route.

Supplementary figure 4. Bayesian maximum clade credibility phylogenetic trees of PCV2 PCV2a (Supplementary figure 4a), PCV2b (Supplementary figure 4b) and PCV2d (Supplementary figure 4c), CRF01 (Supplementary figur4d), CRF02 (Supplementary figure 4e) and CRF03 (Supplementary figure 4f). Tips are colored by the macro-area of sampling, while the internal branches are colored based on the reconstructed ancestral state (macro-area). Numbers at the nodes indicate the root location posterior probability.

Supplementary figure 5. Heat map of pairwise mean trade value of live swine between countries(1996-2014).

Supplementary figure 6. Graphical representation of dN-dS for different ORF.

Supplementary figure 7. Shannon entropy values have been calculated for each codon position in the *ORF1* gene. The region encoding for the *ORF3*, *ORF4* and Rep' are reported.

Supplementary figure 8. Codons reported to be under different selective forces between different genotypes combinations. Color coded bars associated to each codon position represent the presence of a statistically significant difference (p-value<0,05) in selective pressure acting on different genotypes pairs.

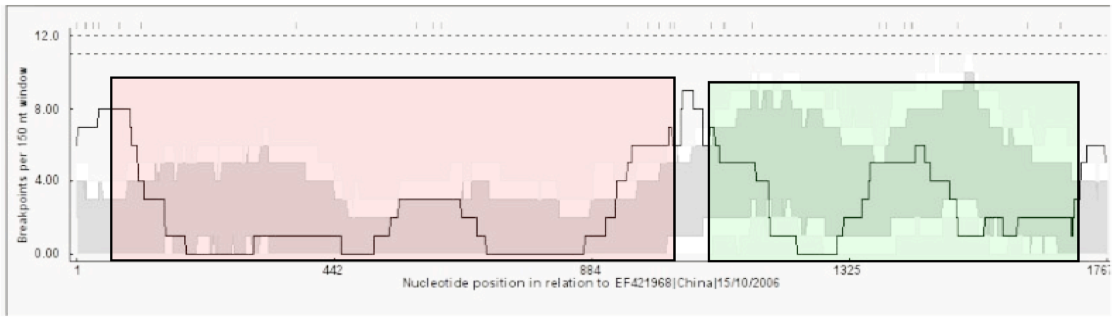
Supplementary figure 9. PCV2a maximum likelihood based reconstruction of ancestral states for Cap (*ORF2*) amino acids.

Supplementary figure 10. PCV2d maximum likelihood based reconstruction of ancestral states for Cap (*ORF2*) amino acids.

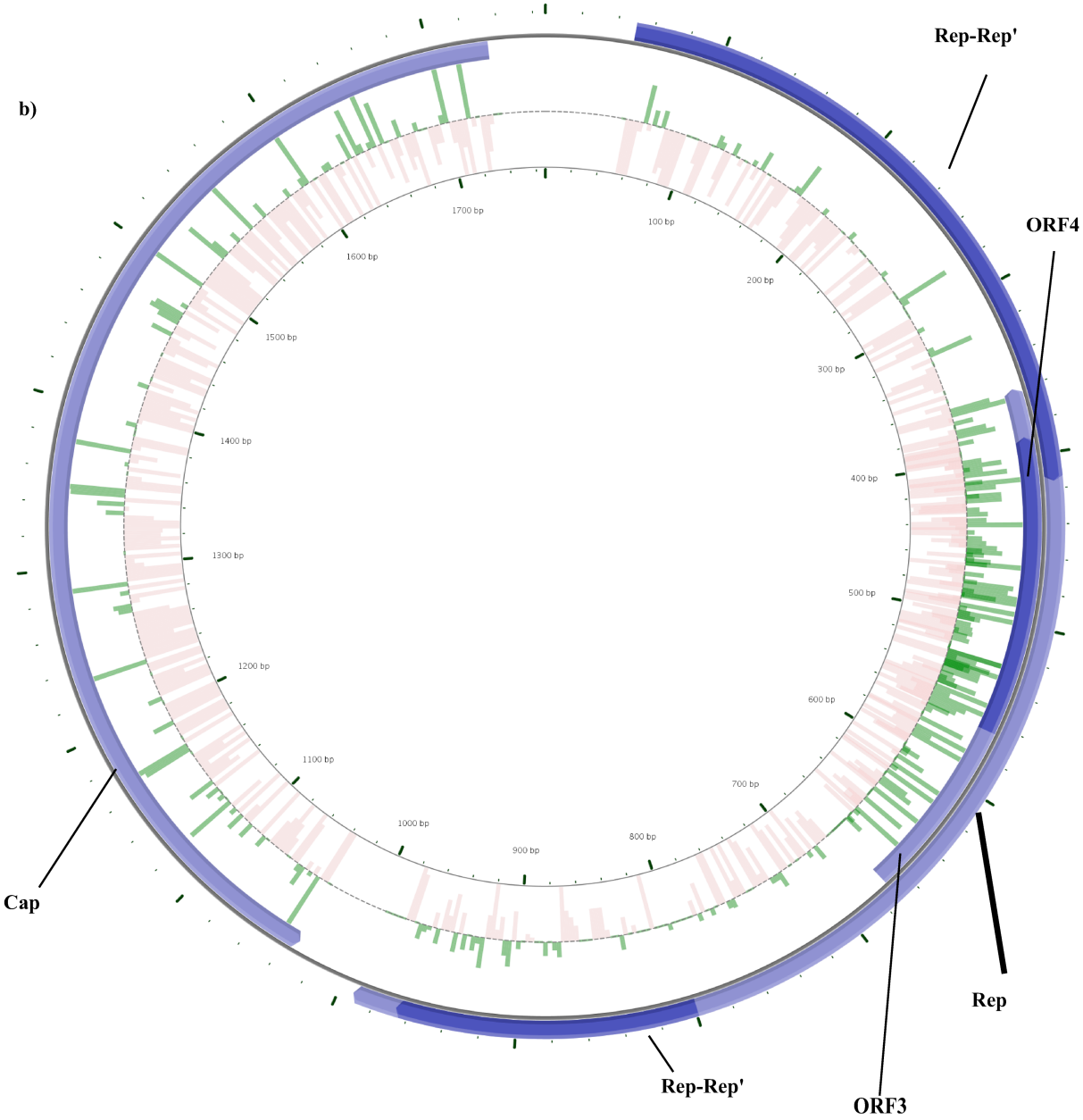
Supplementary Table 1. Logarithm of Marginal likelihood estimation (logMLE) calculated using path sampling (PS) and stepping-stone (SS) for symmetric and asymmetric spreading models.

Supplementary Table 2. Summary of PRIME analysis.

a)



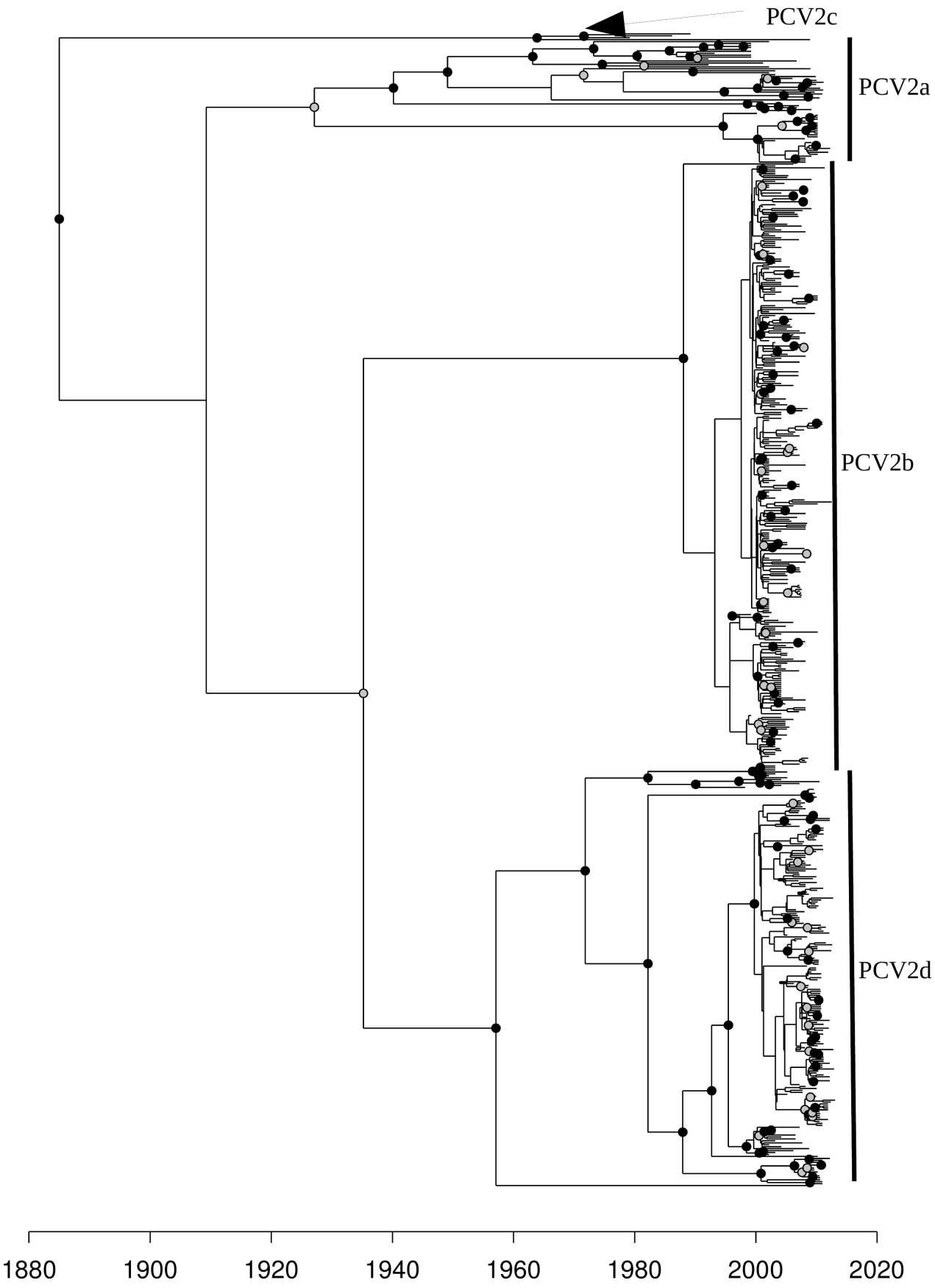
b)

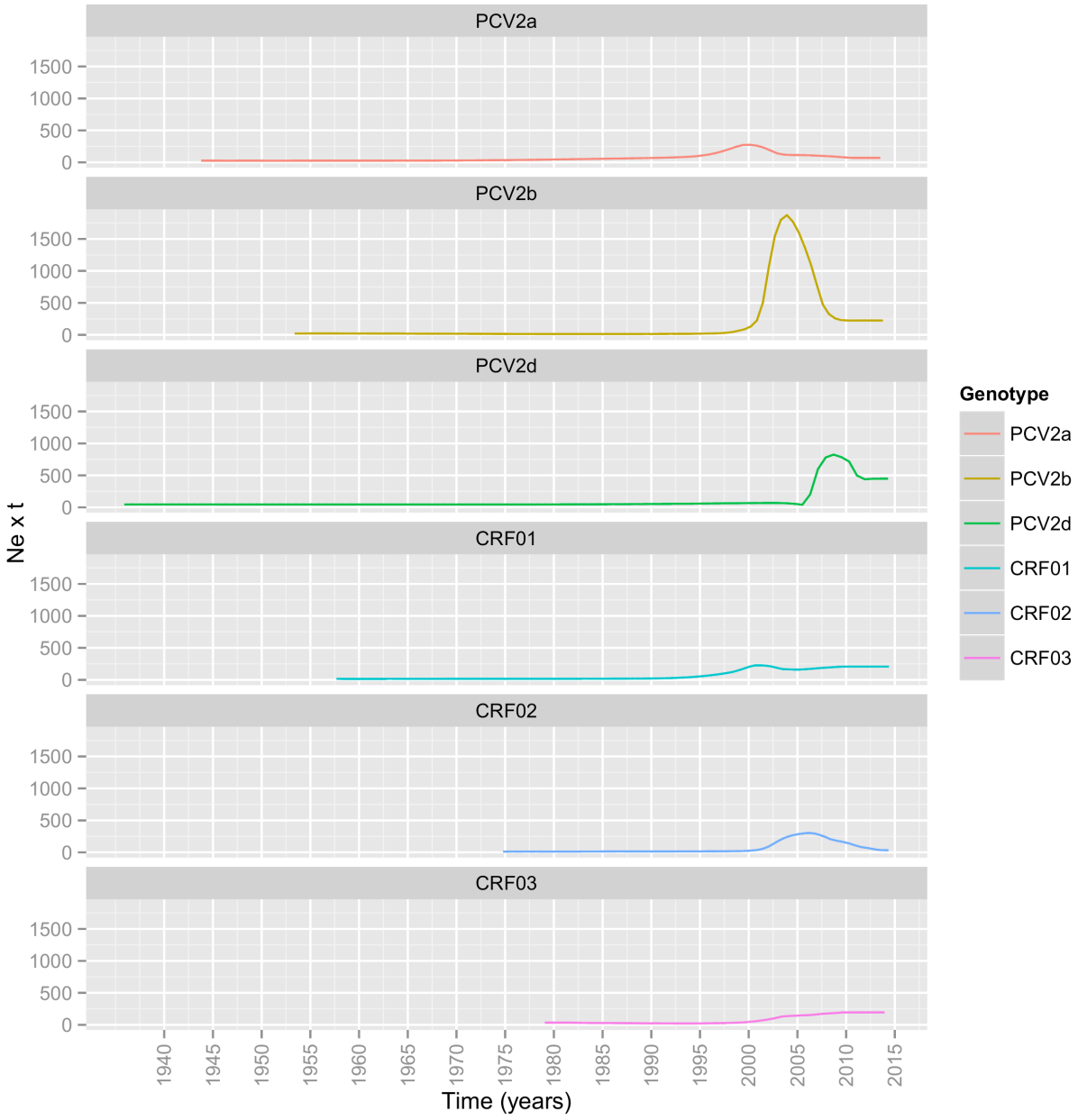


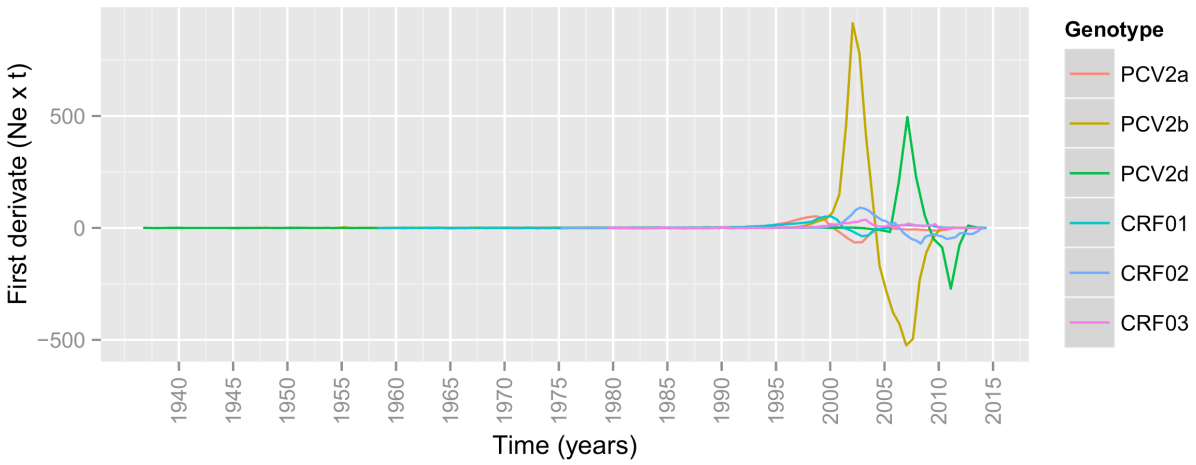
704

705

706


PCV2c
PCV2a
PCV2b
PCV2d
1880
1900
1920
1940
1960
1980
2000
2020


PCV2a
PCV2b
PCV2d
CRF01
CRF02
CRF03
1500
1000
500
0
Ne x t
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
Time (years)
Genotype
PCV2a
PCV2b
PCV2d
CRF01
CRF02
CRF03


First derivate (Ne x t)
500
0
−500
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
Time (years)
Genotype
PCV2a
PCV2b
PCV2d
CRF01
CRF02
CRF03

708

Supporting information captions

Supplementary figure 1. Bayesian maximum clade credibility phylogenetic trees of PCV2a, PCV2b and PCV2d. Tips are colored by the country of sampling, while the internal branches are colored based on the reconstructed ancestral state (country). Numbers at the nodes indicate the estimated mean age. Branch width is proportional to location posterior probability.

Supplementary figure 2. Maps of PCV2 genotype and CRF distribution and spreading routes.

Supplementary figure 3. Networks displaying the spreading of PCV2 genotypes and CRF among different geographical regions. Edges connecting different macro-areas are proportional to the Bayesian Factor (i.e. the statistical support) calculated for that migration route.

Supplementary figure 4. Bayesian maximum clade credibility phylogenetic trees of PCV2 PCV2a (Supplementary figure 4a), PCV2b (Supplementary figure 4b) and PCV2d (Supplementary figure 4c), CRF01 (Supplementary figur4d), C RF02 (Supplementary figure 4e) and CRF03 (Supplementary figure 4f). Tips are colored by the macro-area of sampling, while the internal branches are colored based on the reconstructed ancestral state (macro-area). Numbers at the nodes indicate the root location posterior probability.

Supplementary figure 5. Heat map of pairwise mean trade value of live swine between countries(1996-2014).

Supplementary figure 6. Graphical representation of dN-dS for different ORF.

Supplementary figure 7. Shannon entropy values have been calculated for each codon position in the *ORF1* gene. The region encoding for the *ORF3*, *ORF4* and Rep' are reported.

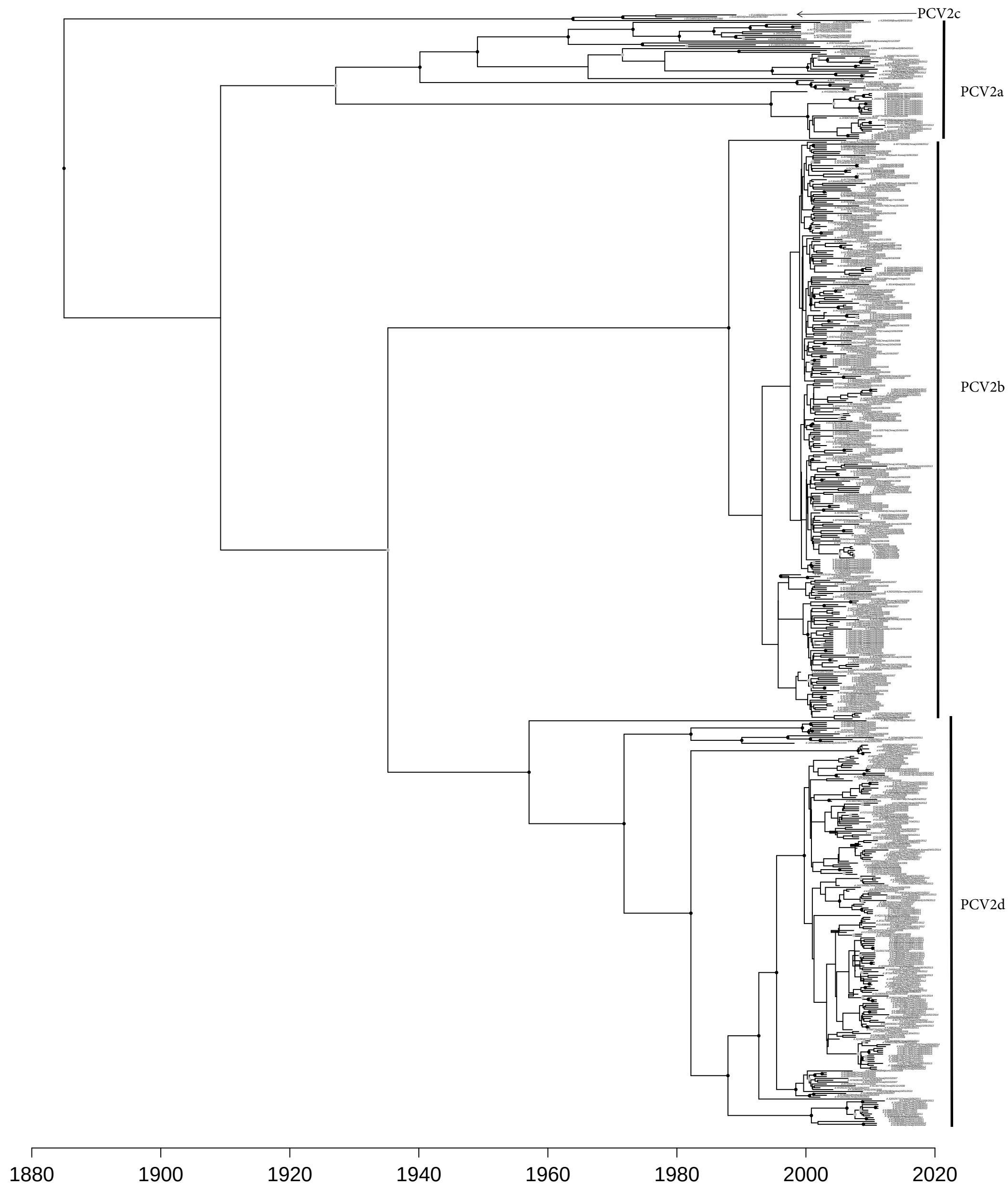
Supplementary figure 8. Codons reported to be under different selective forces between different genotypes combinations. Color coded bars associated to each codon position represent the presence of a statistically significant difference (p-value<0,05) in selective pressure acting on different genotypes pairs.

Supplementary figure 9. PCV2a maximum likelihood based reconstruction of ancestral states for Cap (*ORF2*) amino acids.

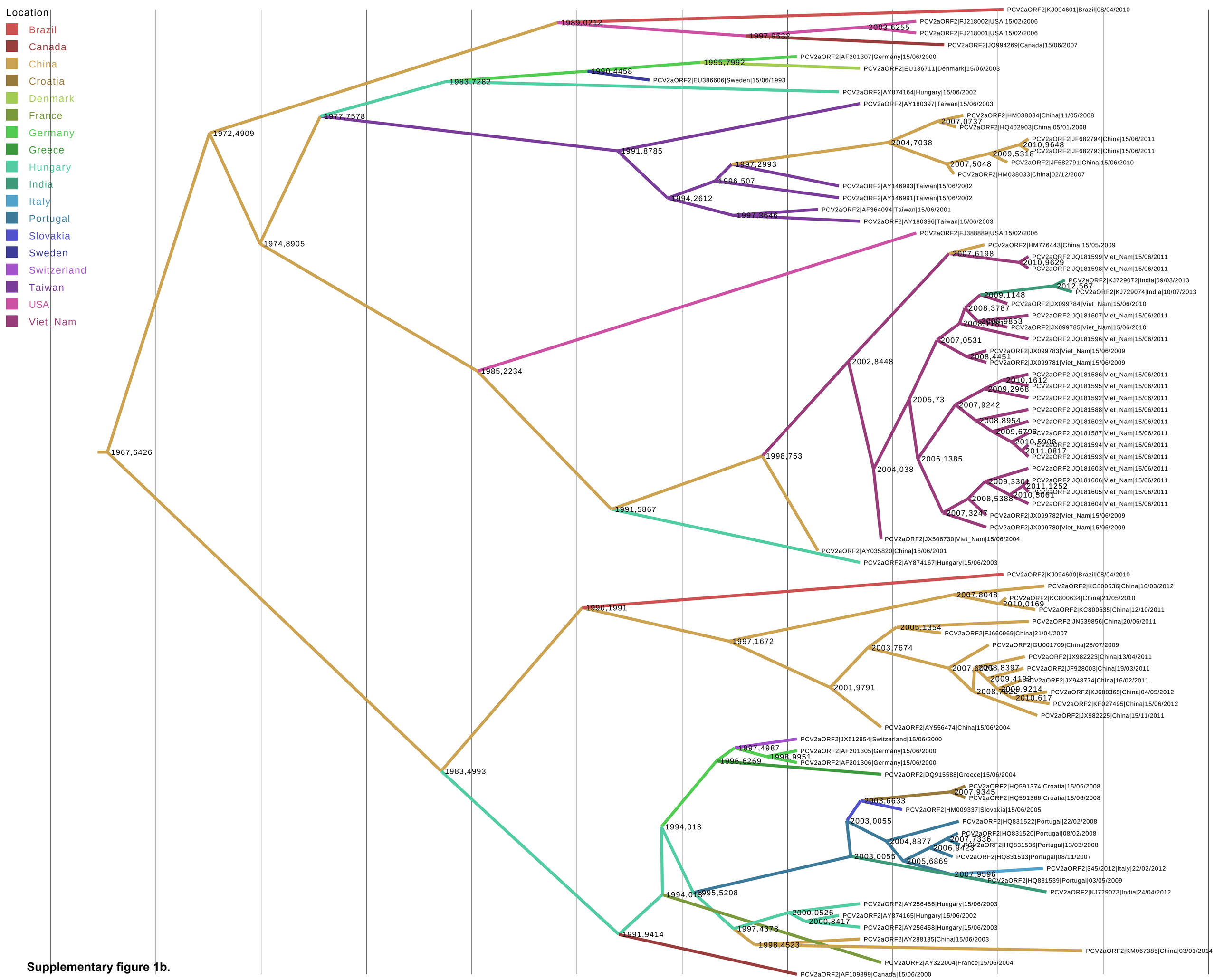
Supplementary figure 10. PCV2d maximum likelihood based reconstruction of amino acid ancestral state.

Supplementary Table 1. Logarithm of Marginal likelihood estimation (logMLE) calculated using path sampling (PS) and stepping-stone (SS) for symmetric and asymmetric spreading models.

Supplementary Table 2. Summary of PRIME analysis.

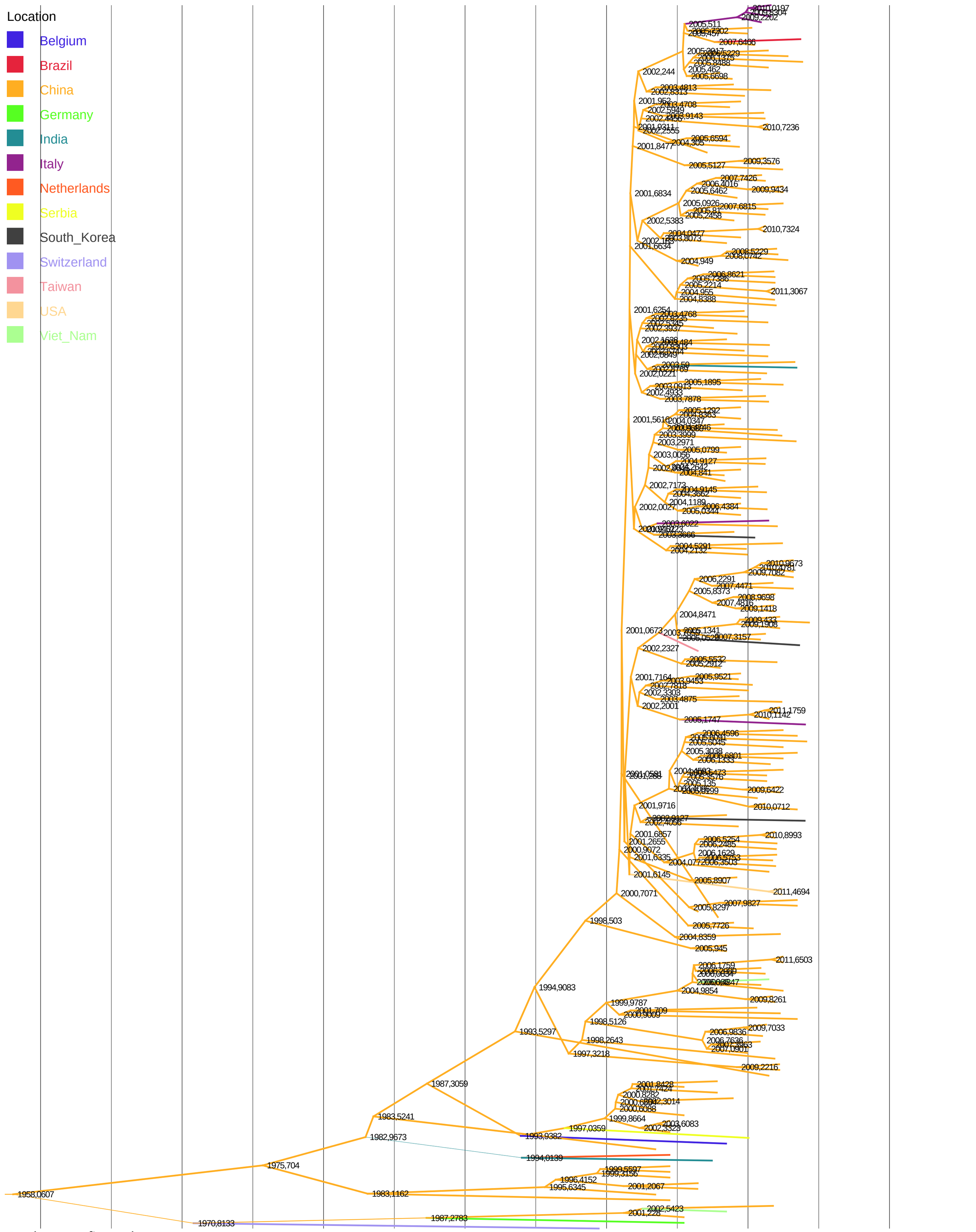


Supplementary figure 1a. Bayesian maximum clade credibility phylogenetic tree obtained based on non recombinant strains (BEAST software). Nodes displaying a posterior probability higher than 0,9 and comprised between 0,9 and 0,7 are represented as black and grey circles,respectively.It is possible to zoom in to appreciate further details and tip labels

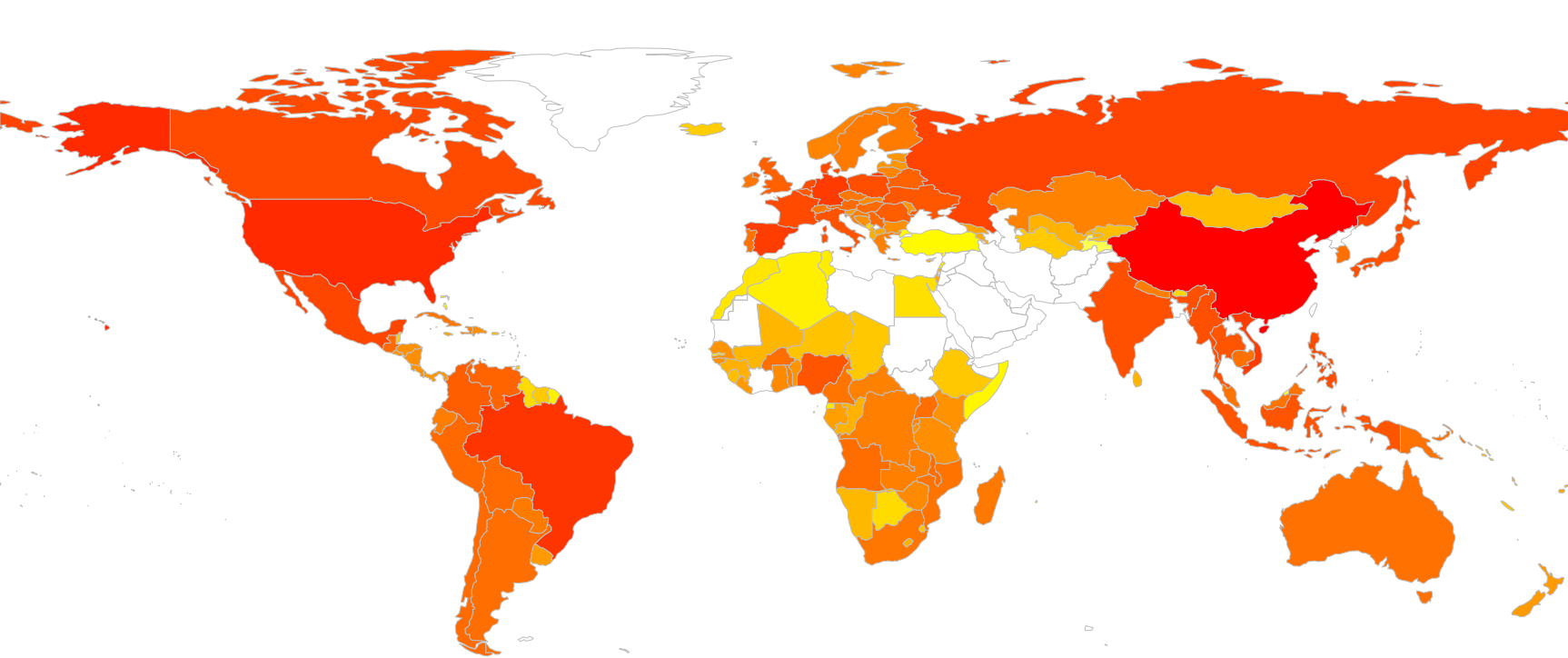


Supplementary figure 1b.

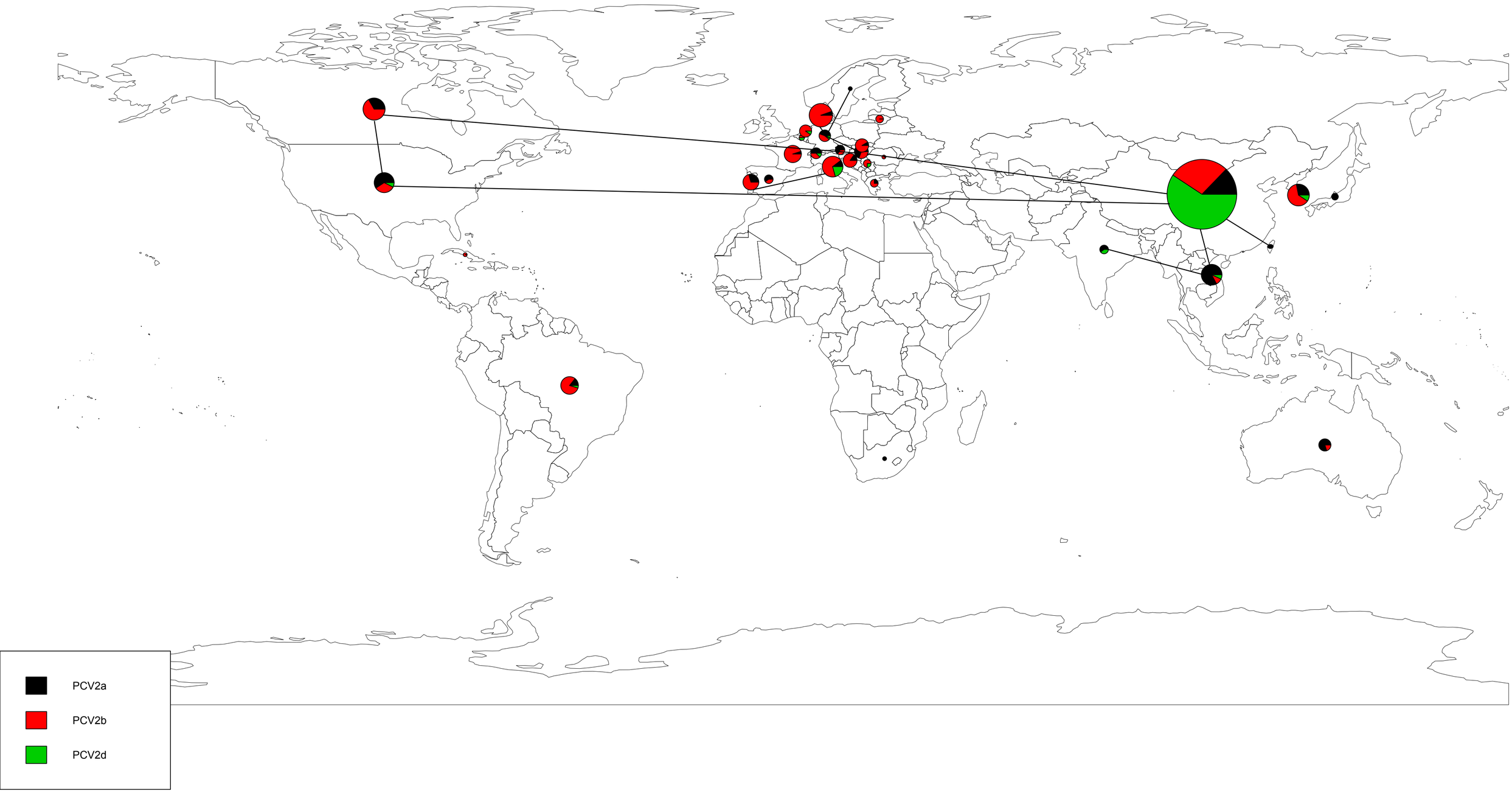

Location
Australia
Austria
Belgium
Brazil
Canada
China
Croatia
Czech
Denmark
France
Germany
Greece
Hungary
Italy
Lithuania
Netherlands
Portugal
Romania
Serbia
Slovakia
South_Korea
Spain
Switzerland
USA
Viet_Nam
Supplementary figure 1c.
Supplementary figure 1c.
1980
1990
2000
2010
2020



Supplementary figure 1d.

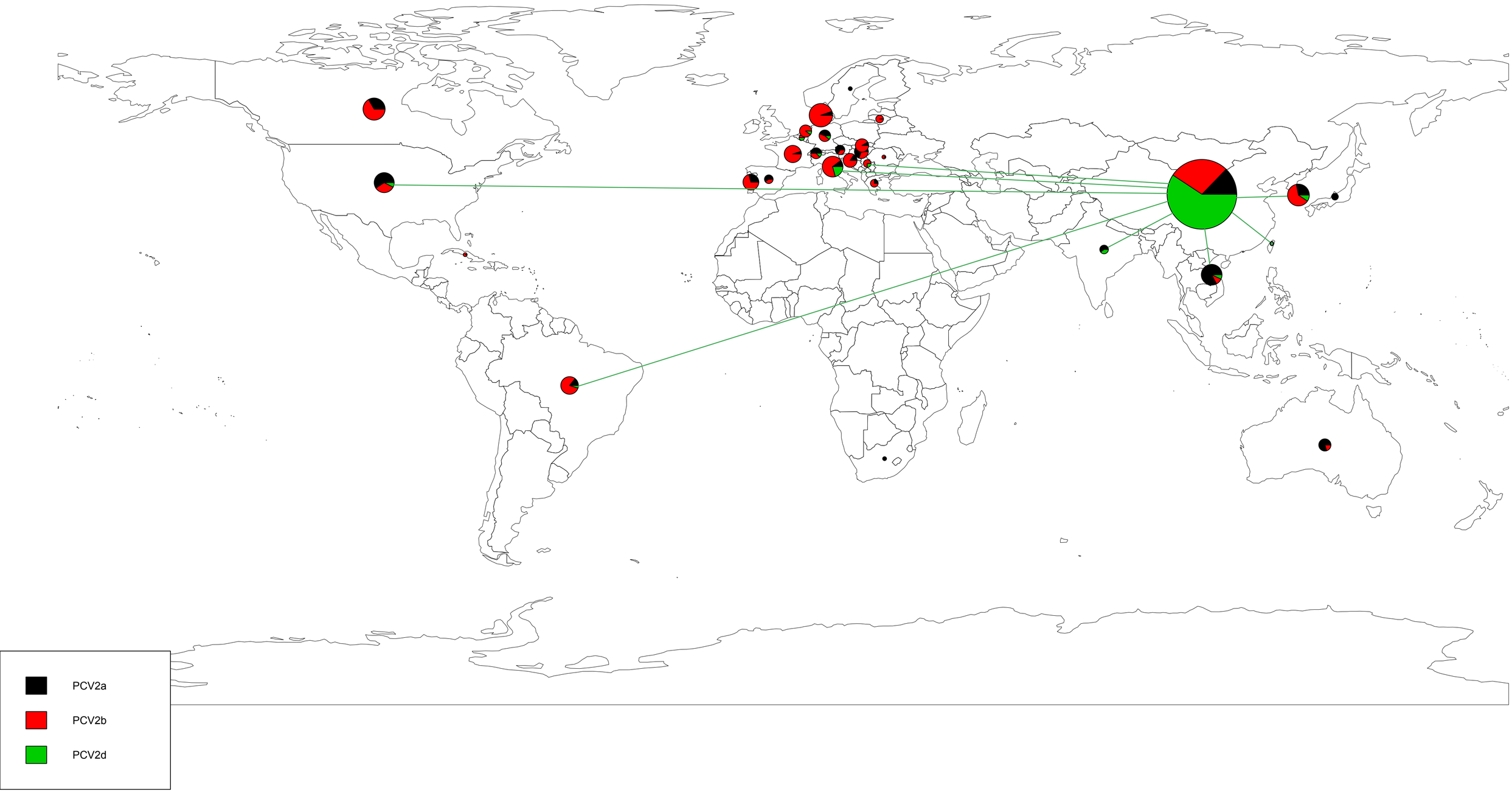


Supplementary figure 2a. Number of live pigs raised in differnt countries (average live pigs 1990-201 3) . Surce http: //faostat. fao. org/

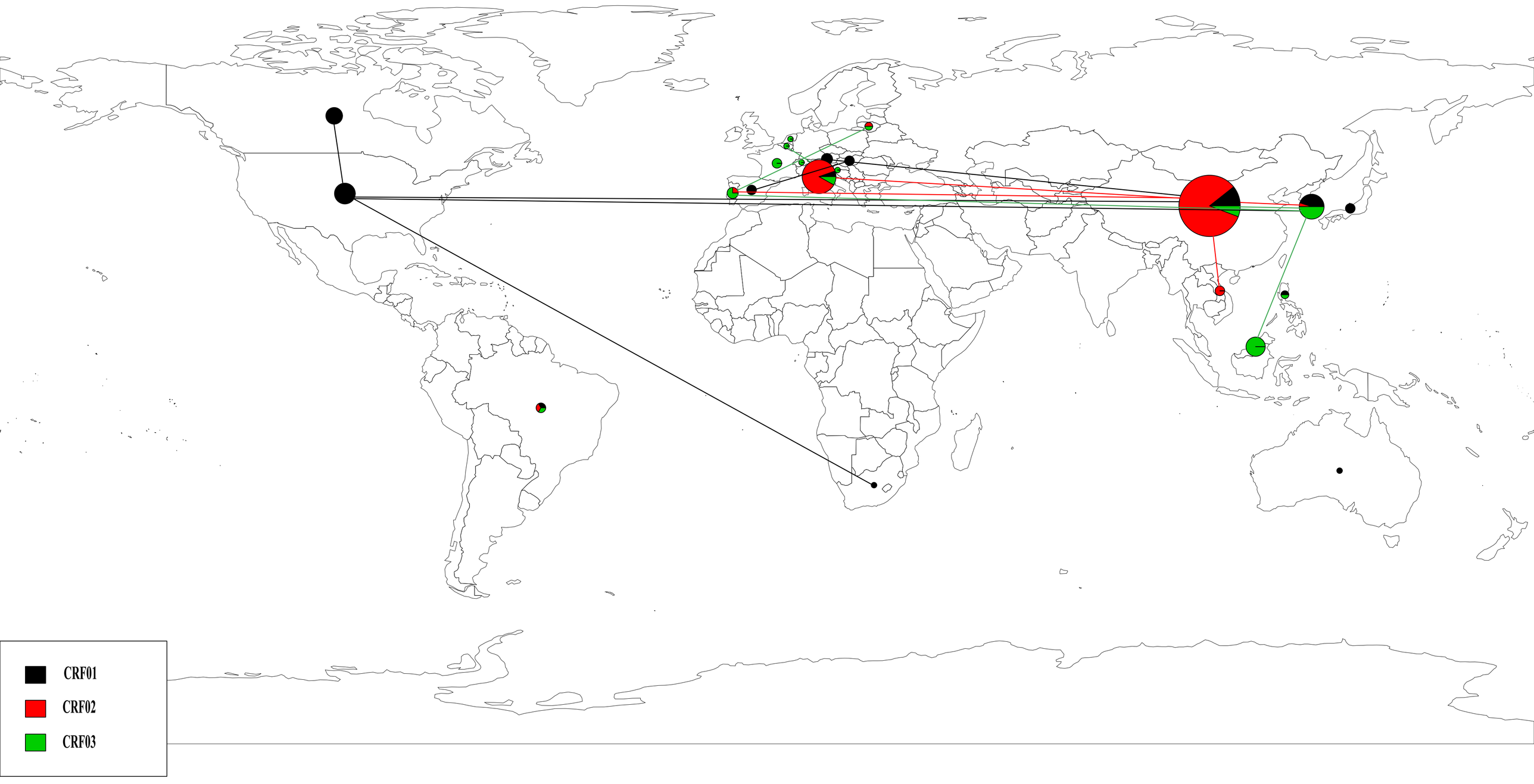


Supplementary figure 2b. Distribution of PCV2 genotyipes in different countries represented as pie charts. The radious of the circmferences is proportional to the number of sequences . Genotypes are representes as color-coded slices whose area is proportional to the genotype' s frequency. Well supported migration patterns between countries (i. e. BF>10) of PCV2a are represented as black lines.

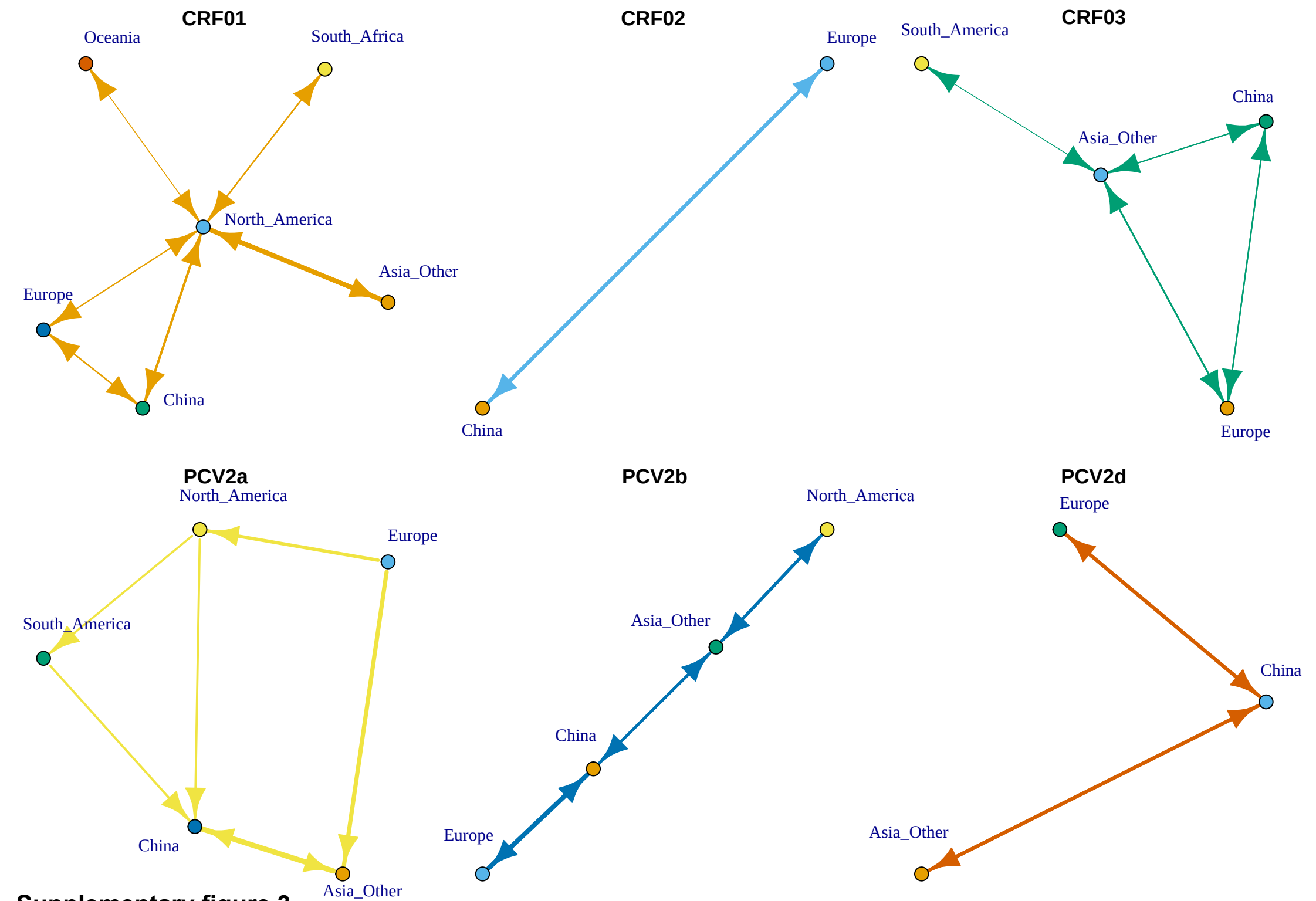
Supplementary figure 2c. Distribution of PCV2 genotyipes in different countries represented as pie charts. The radious of the circmferences is proportional to the number of sequences . Genotypes are representes as color-coded slices whose area is proportional to the genotype' s frequency. Well supported migration patterns between countries (i. e. BF>10) of PCV2b are represented as red lines.



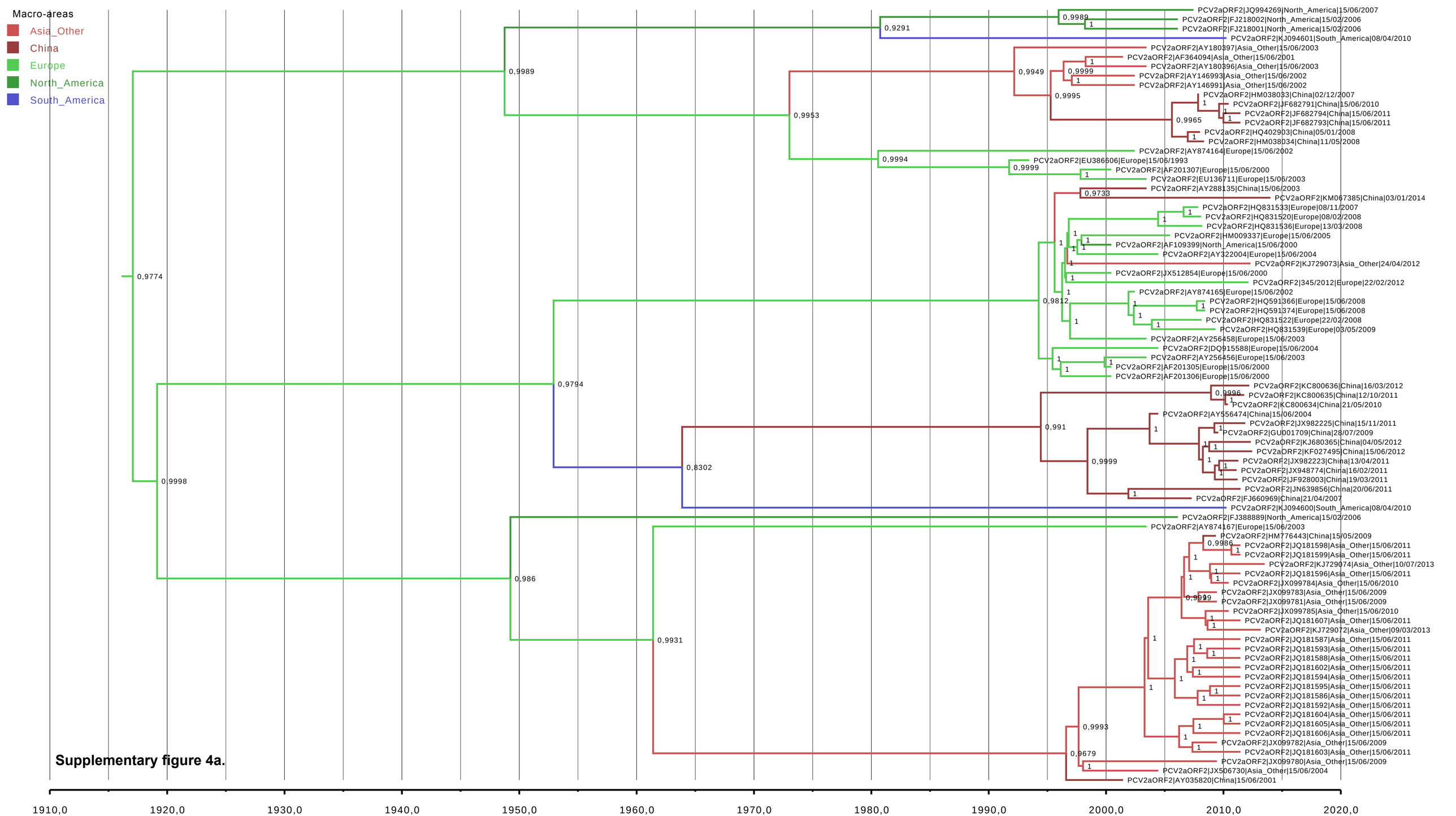
Supplementary figure 2d. Distribution of PCV2 genotyipes in different countries represented as pie charts. The radious of the circmferences is proportional to the number of sequences . Genotypes are representes as color-coded slices whose area is proportional to the genotype' s frequency. Well supported migration patterns between countries (i. e. BF>10) of PCV2d are represented as green lines.

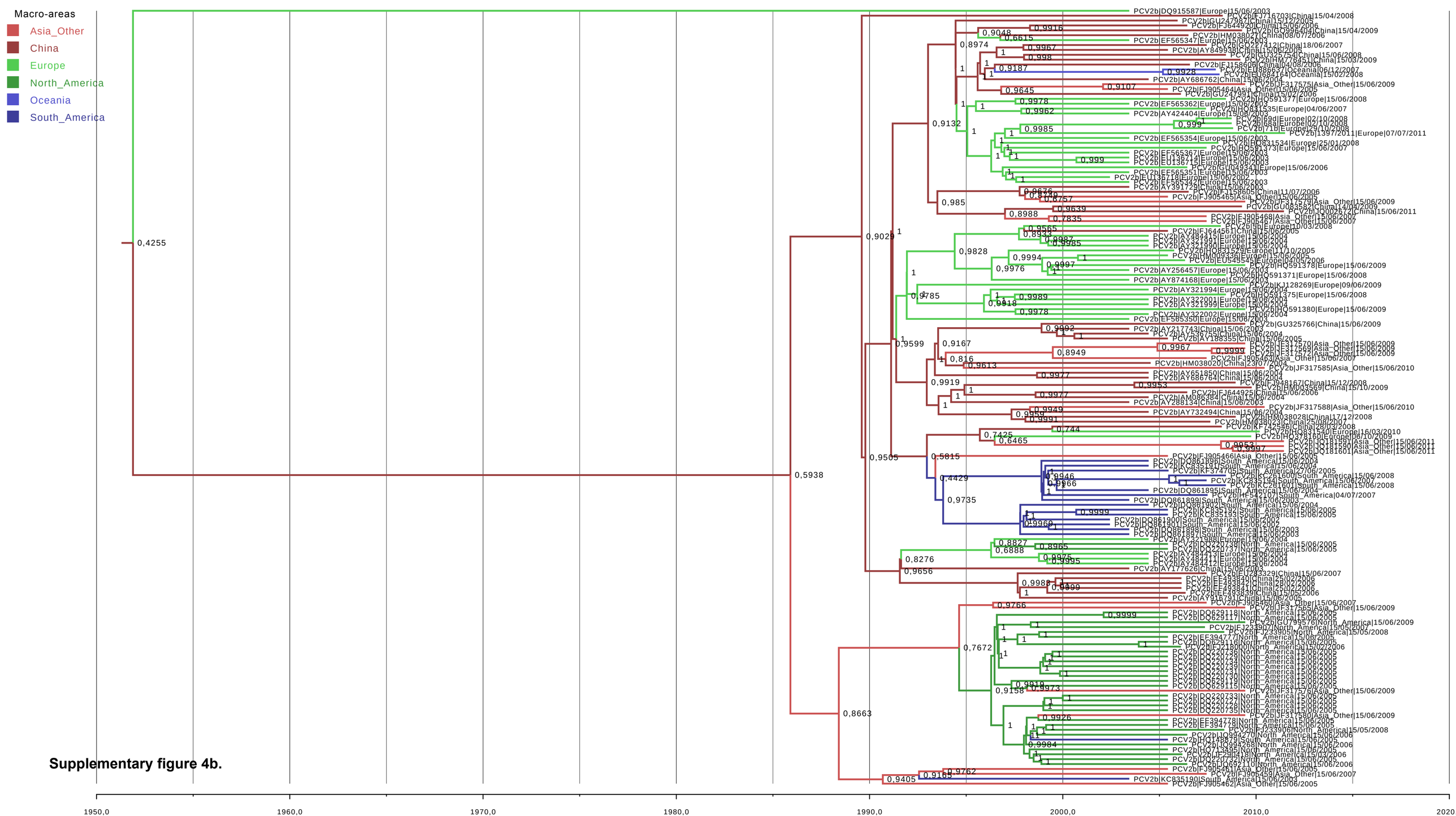


Supplementary figure 2e. Distribution of maj or recombinant clusters in different countries represented as pie charts. The radious of the circmferences is proportional to the number of sequences . Recombination events are representes as color-coded slices whose area is proportional to the recombinant frequency. Well supported migration patterns between countries (i. e. BF>10) of Event 1 , Event 2 and Event 9 are represented as black, red and green lines, respectively.

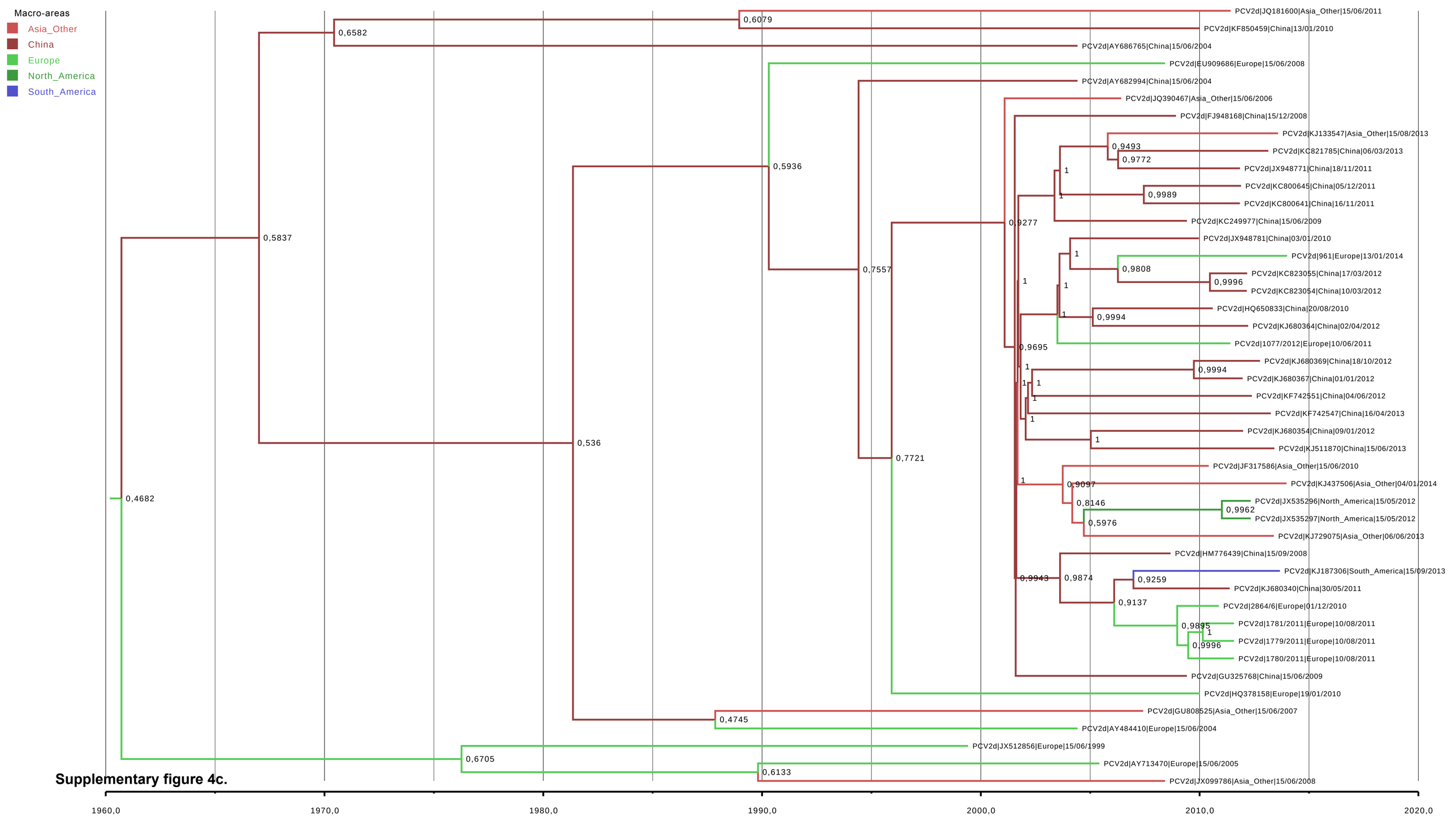


Supplementary figure 3.

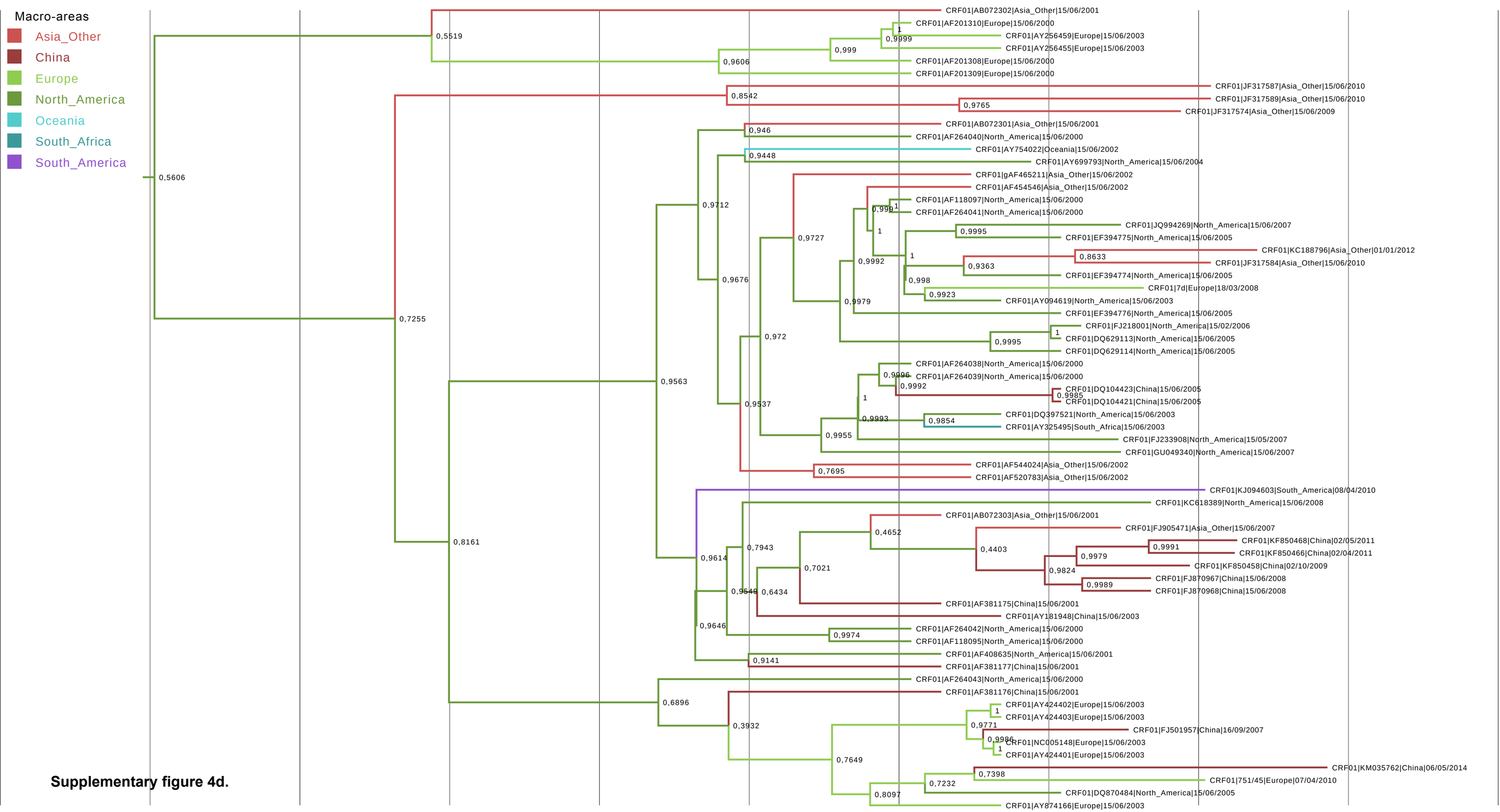

Macro-areas
Asia_Other
China
Europe
North_America
South_America
PCV2aORF2|JQ994269|North_America|15/06/2007
PCV2aORF2|FJ218002|North_America|15/02/2006
PCV2aORF2|FJ218001|North_America|15/02/2006
PCV2aORF2|KJ094601|South_America|08/04/2010
PCV2aORF2|AY180397|Asia_Other|15/06/2003
PCV2aORF2|AF364094|Asia_Other|15/06/2001
PCV2aORF2|AY180396|Asia_Other|15/06/2003
PCV2aORF2|AY146993|Asia_Other|15/06/2002
PCV2aORF2|AY146991|Asia_Other|15/06/2002
PCV2aORF2|HM038033|China|02/12/2007
PCV2aORF2|JF682791|China|15/06/2010
PCV2aORF2|JF682794|China|15/06/2011
PCV2aORF2|JF682793|China|15/06/2011
PCV2aORF2|HQ402903|China|05/01/2008
PCV2aORF2|HM038034|China|11/05/2008
PCV2aORF2|AY874164|Europe|15/06/2002
PCV2aORF2|EU386506|Europe|15/06/1993
PCV2aORF2|AF201307|Europe|15/06/2000
PCV2aORF2|EU136711|Europe|15/06/2003
PCV2aORF2|AY288135|China|15/06/2003
PCV2aORF2|KM067385|China|03/01/2014
PCV2aORF2|HQ831533|Europe|08/11/2007
PCV2aORF2|HQ831520|Europe|08/02/2008
PCV2aORF2|HQ831536|Europe|13/03/2008
PCV2aORF2|HM009337|Europe|15/06/2005
PCV2aORF2|AF109399|North_America|15/06/2000
PCV2aORF2|AY322004|Europe|15/06/2004
PCV2aORF2|KJ729073|Asia_Other|24/04/2012
PCV2aORF2|JX512854|Europe|15/06/2000
PCV2aORF2|345/2012|Europe|22/02/2012
PCV2aORF2|AY874165|Europe|15/06/2002
PCV2aORF2|HQ591366|Europe|15/06/2008
PCV2aORF2|HQ591374|Europe|15/06/2008
PCV2aORF2|HQ831522|Europe|22/02/2008
PCV2aORF2|HQ831539|Europe|03/05/2009
PCV2aORF2|AY256458|Europe|15/06/2003
PCV2aORF2|DQ915588|Europe|15/06/2004
PCV2aORF2|AY256456|Europe|15/06/2003
PCV2aORF2|AF201305|Europe|15/06/2000
PCV2aORF2|AF201306|Europe|15/06/2000
PCV2aORF2|KC800636|China|16/03/2012
PCV2aORF2|KC800635|China|12/10/2011
PCV2aORF2|KC800634|China|21/05/2010
PCV2aORF2|AY556474|China|15/06/2004
PCV2aORF2|JX982225|China|15/11/2011
PCV2aORF2|GU001709|China|28/07/2009
PCV2aORF2|KJ680365|China|04/05/2012
PCV2aORF2|KF027495|China|15/06/2012
PCV2aORF2|JX982223|China|13/04/2011
PCV2aORF2|JX948774|China|16/02/2011
PCV2aORF2|JF928003|China|19/03/2011
PCV2aORF2|JN639856|China|20/06/2011
PCV2aORF2|FJ660969|China|21/04/2007
PCV2aORF2|KJ094600|South_America|08/04/2010
PCV2aORF2|FJ388889|North_America|15/02/2006
PCV2aORF2|AY874167|Europe|15/06/2003
PCV2aORF2|HM776443|China|15/05/2009
PCV2aORF2|JQ181598|Asia_Other|15/06/2011
PCV2aORF2|JQ181599|Asia_Other|15/06/2011
PCV2aORF2|KJ729074|Asia_Other|10/07/2013
PCV2aORF2|JQ181596|Asia_Other|15/06/2011
PCV2aORF2|JX099784|Asia_Other|15/06/2010
PCV2aORF2|JX099783|Asia_Other|15/06/2009
PCV2aORF2|JX099781|Asia_Other|15/06/2009
PCV2aORF2|JX099785|Asia_Other|15/06/2010
PCV2aORF2|JQ181607|Asia_Other|15/06/2011
PCV2aORF2|KJ729072|Asia_Other|09/03/2013
PCV2aORF2|JQ181587|Asia_Other|15/06/2011
PCV2aORF2|JQ181593|Asia_Other|15/06/2011
PCV2aORF2|JQ181588|Asia_Other|15/06/2011
PCV2aORF2|JQ181602|Asia_Other|15/06/2011
PCV2aORF2|JQ181594|Asia_Other|15/06/2011
PCV2aORF2|JQ181595|Asia_Other|15/06/2011
PCV2aORF2|JQ181586|Asia_Other|15/06/2011
PCV2aORF2|JQ181592|Asia_Other|15/06/2011
PCV2aORF2|JQ181604|Asia_Other|15/06/2011
PCV2aORF2|JQ181605|Asia_Other|15/06/2011
PCV2aORF2|JQ181606|Asia_Other|15/06/2011
PCV2aORF2|JX099782|Asia_Other|15/06/2009
PCV2aORF2|JQ181603|Asia_Other|15/06/2011
PCV2aORF2|JX099780|Asia_Other|15/06/2009
PCV2aORF2|JX506730|Asia_Other|15/06/2004
PCV2aORF2|AY035820|China|15/06/2001
0,9989
0,9291
0,9989
0,9949
0,9999
0,9995
0,9965
0,9953
0,9994
0,9999
0,9733
0,9812
0,9774
0,9794
0,8302
0,9998
0,9996
0,991
0,9999
0,986
0,9931
0,9986
0,9999
0,9993
0,9679
1910,0
1920,0
1930,0
1940,0
1950,0
1960,0
1970,0
1980,0
1990,0
2000,0
2010,0
2020,0
Supplementary figure 4a.



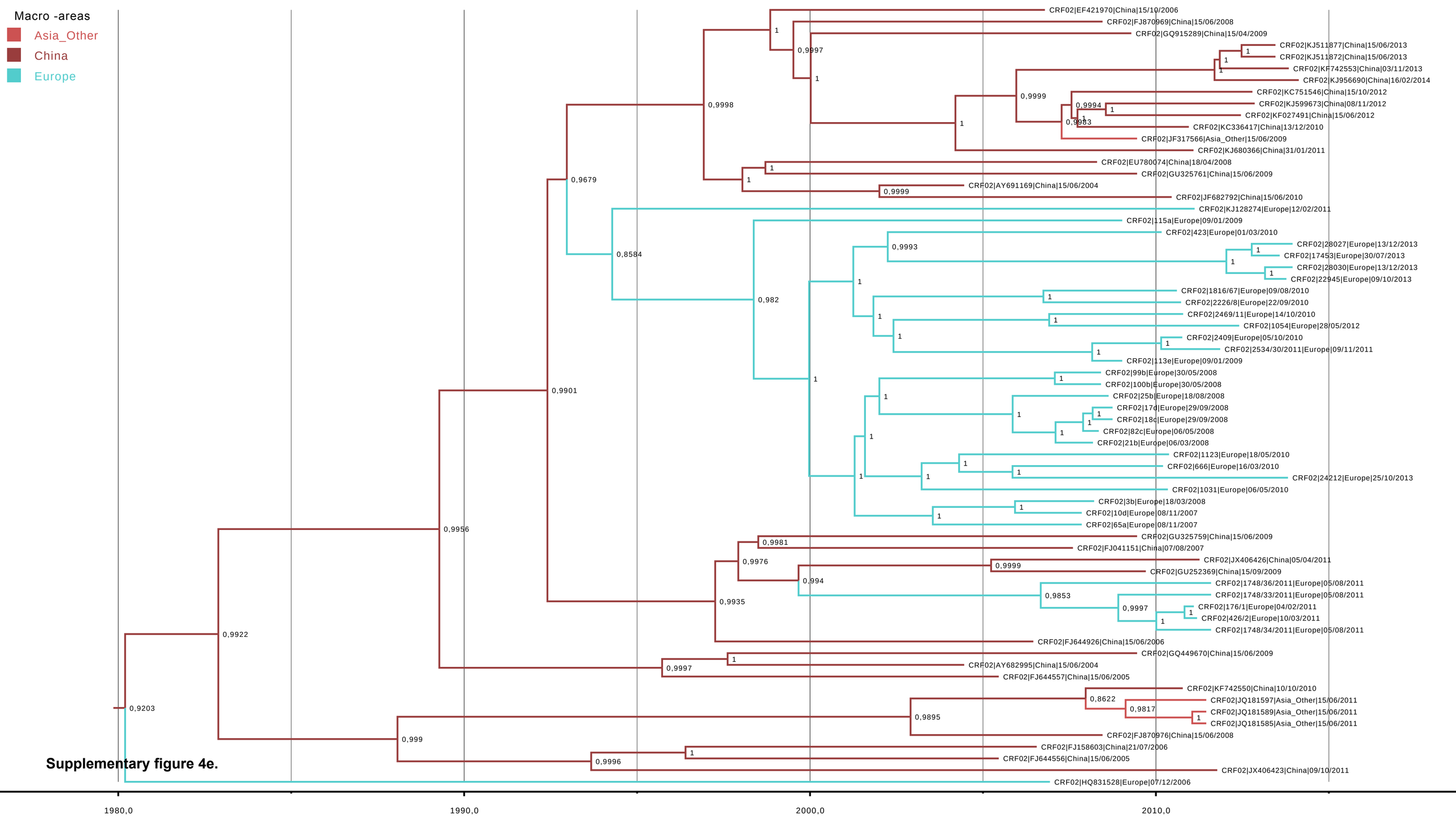
Supplementary figure 4b.



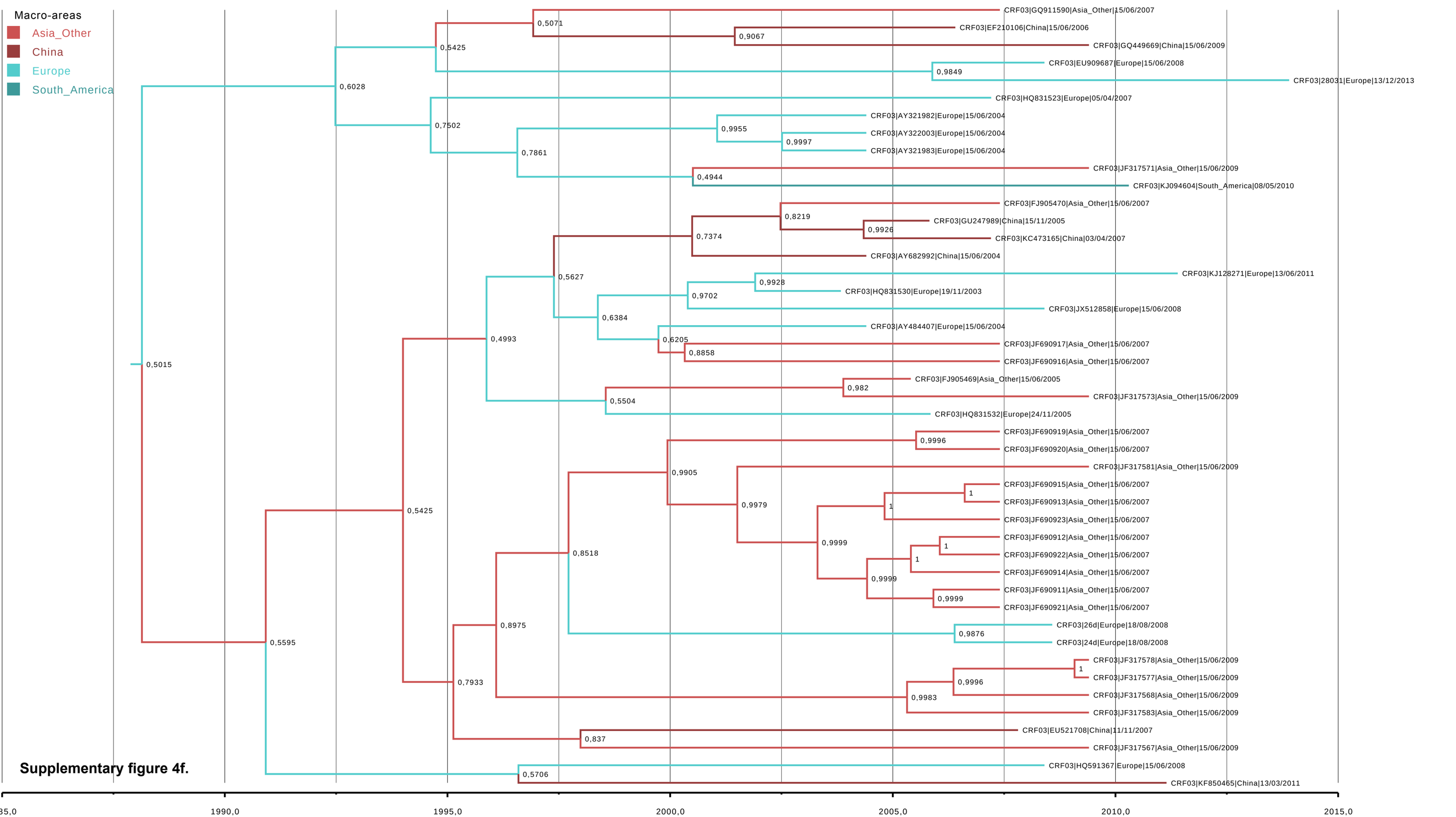
Supplementary figure 4c.



Supplementary figure 4d.

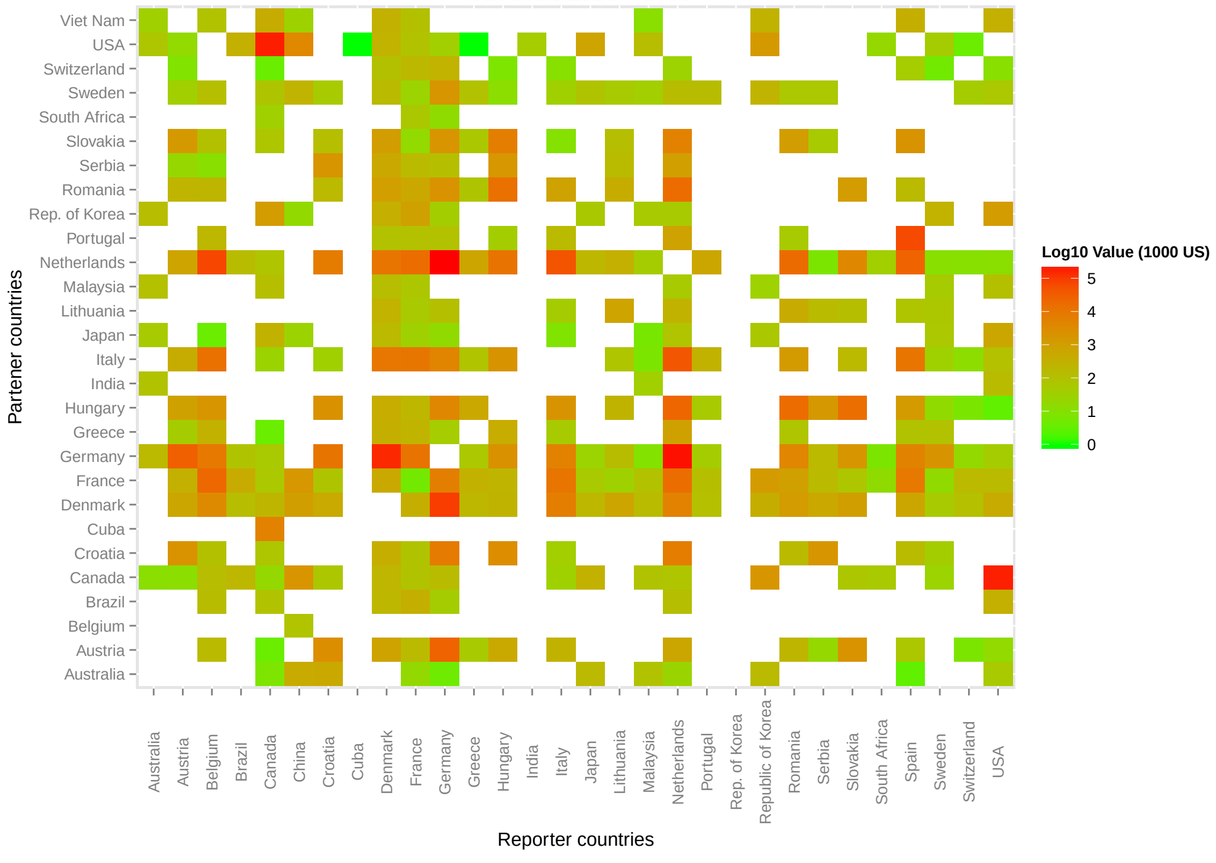


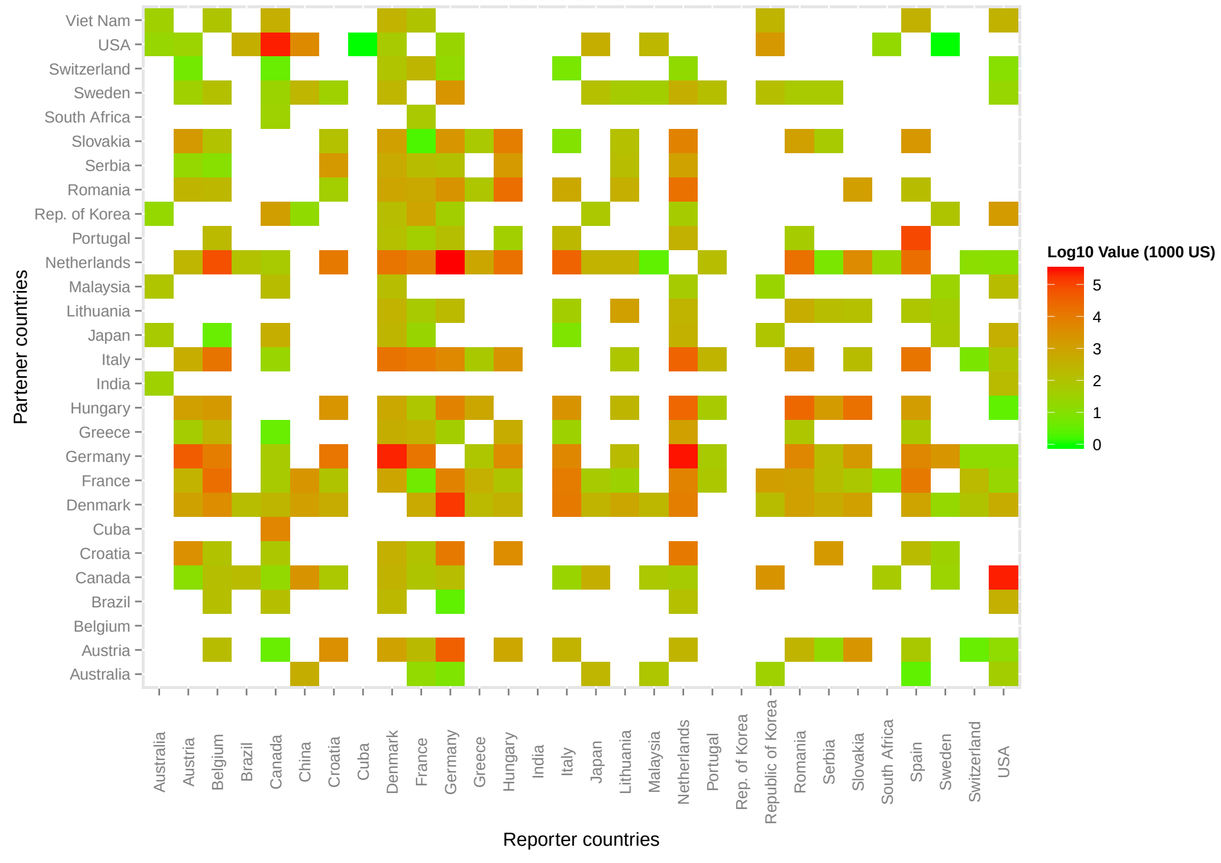
Supplementary figure 4e.

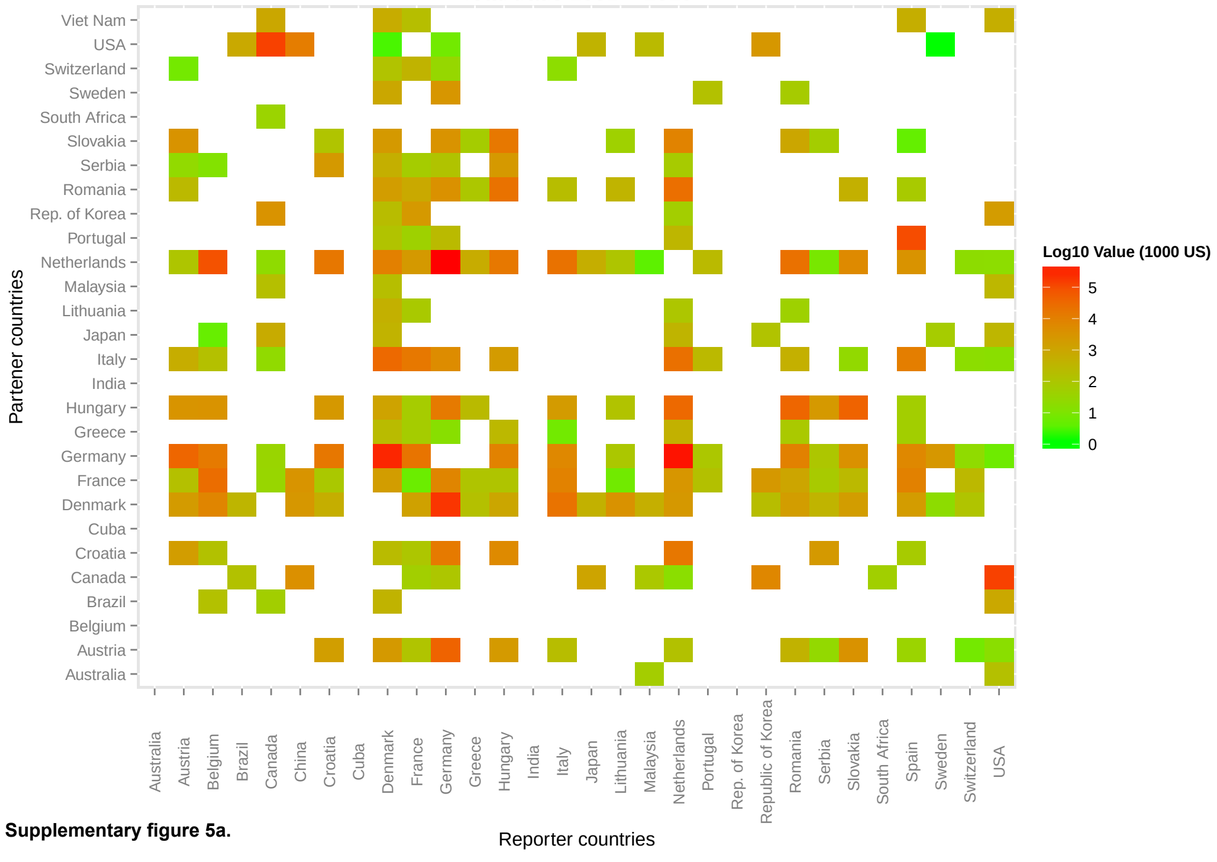


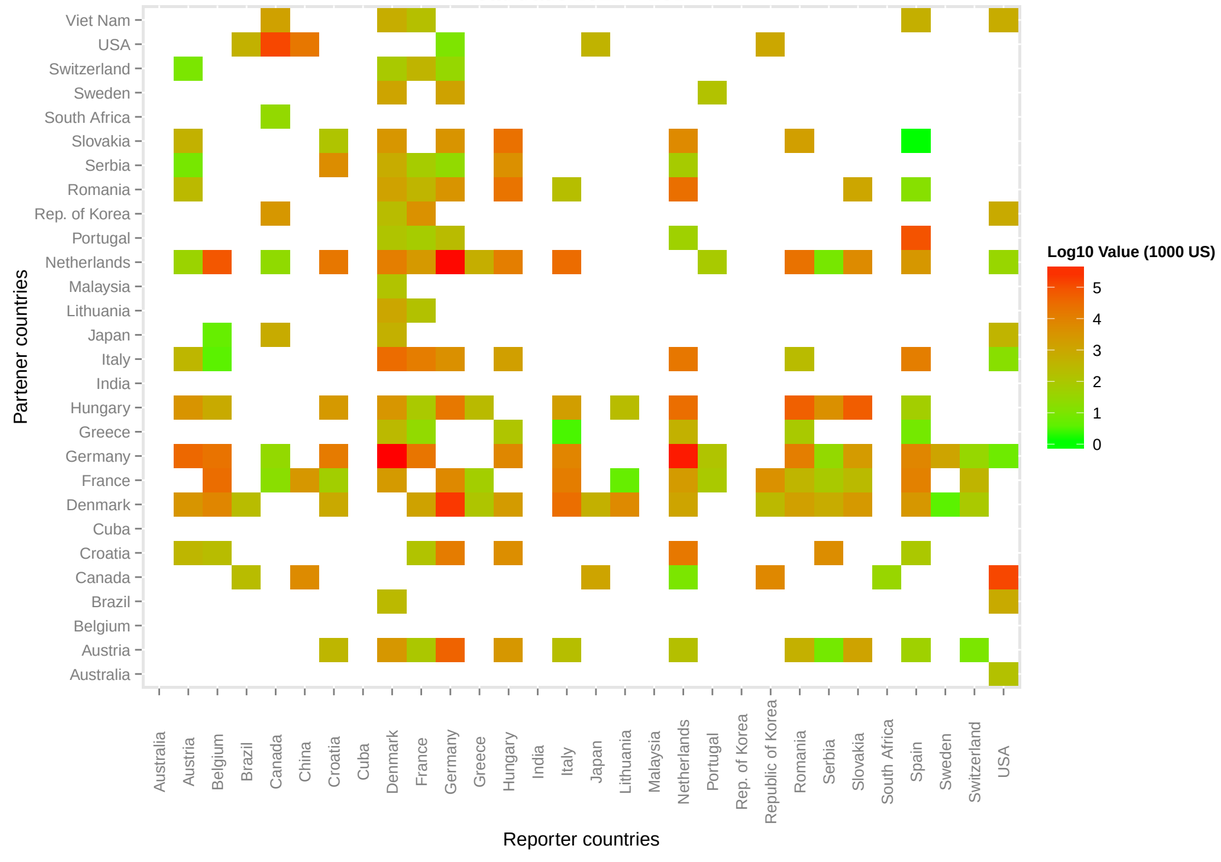
Supplementary figure 4f.

Supplementary figure 5. Heatmap pairwise mean trade value of live swine between countries (1996-2014) (http://comtrade.un.org/) calculated for different decades and type of trade flow: total trade (Supplementary figure 5a), total export (Supplementary figure 5b) and total import (Supplementary figure 5c).

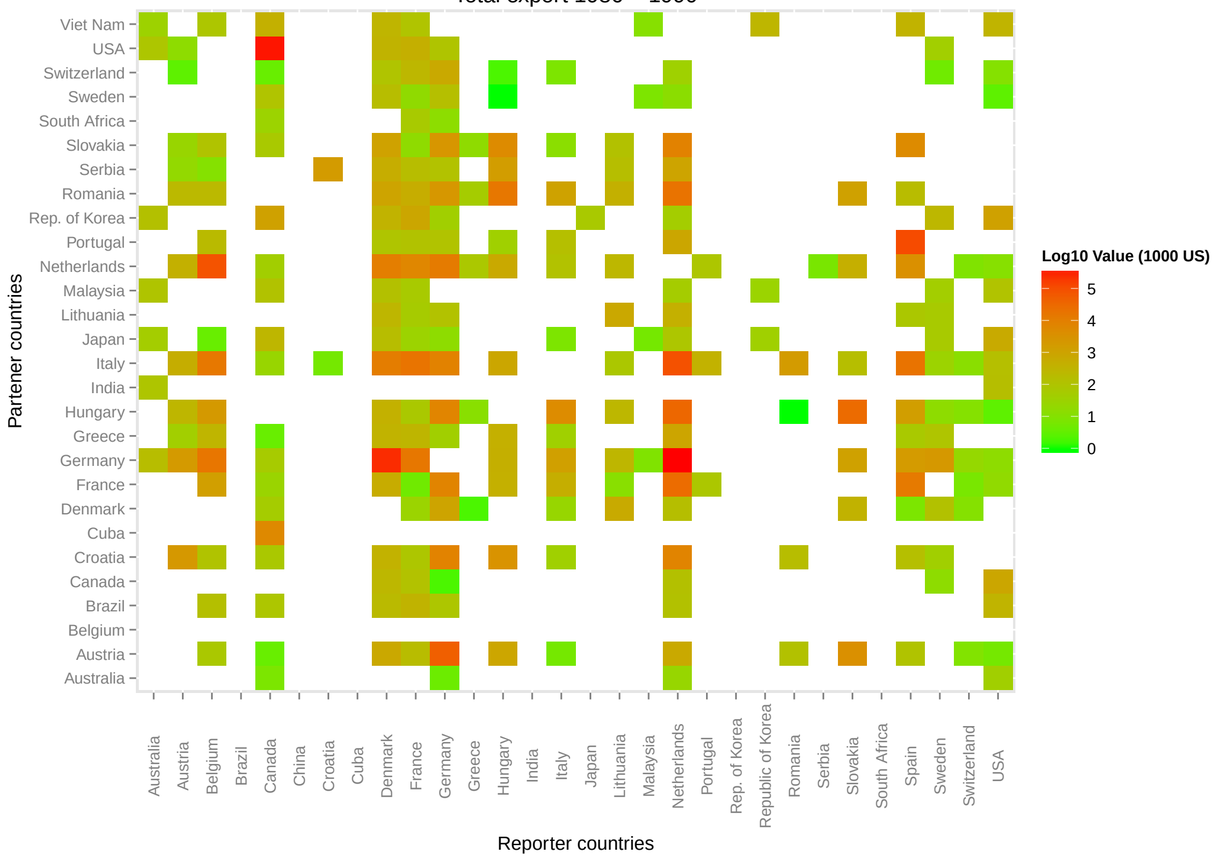


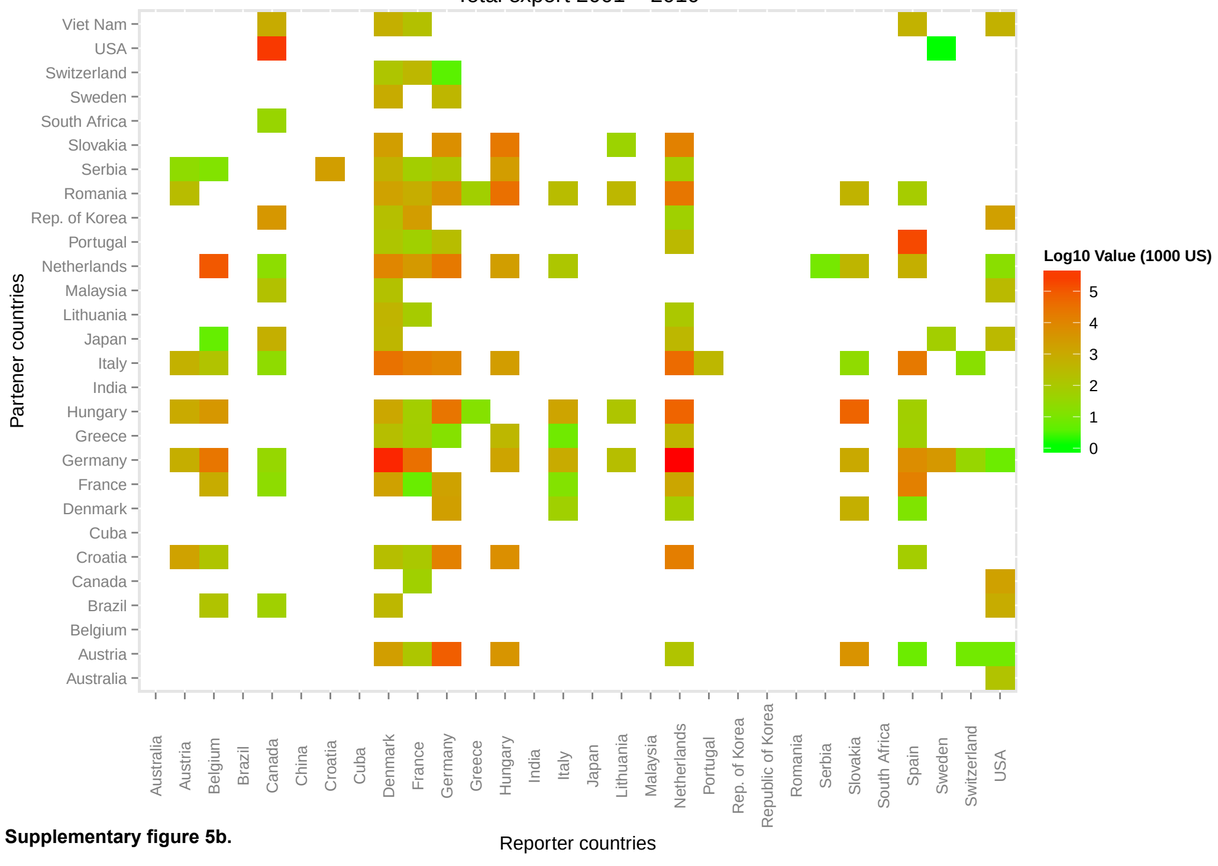




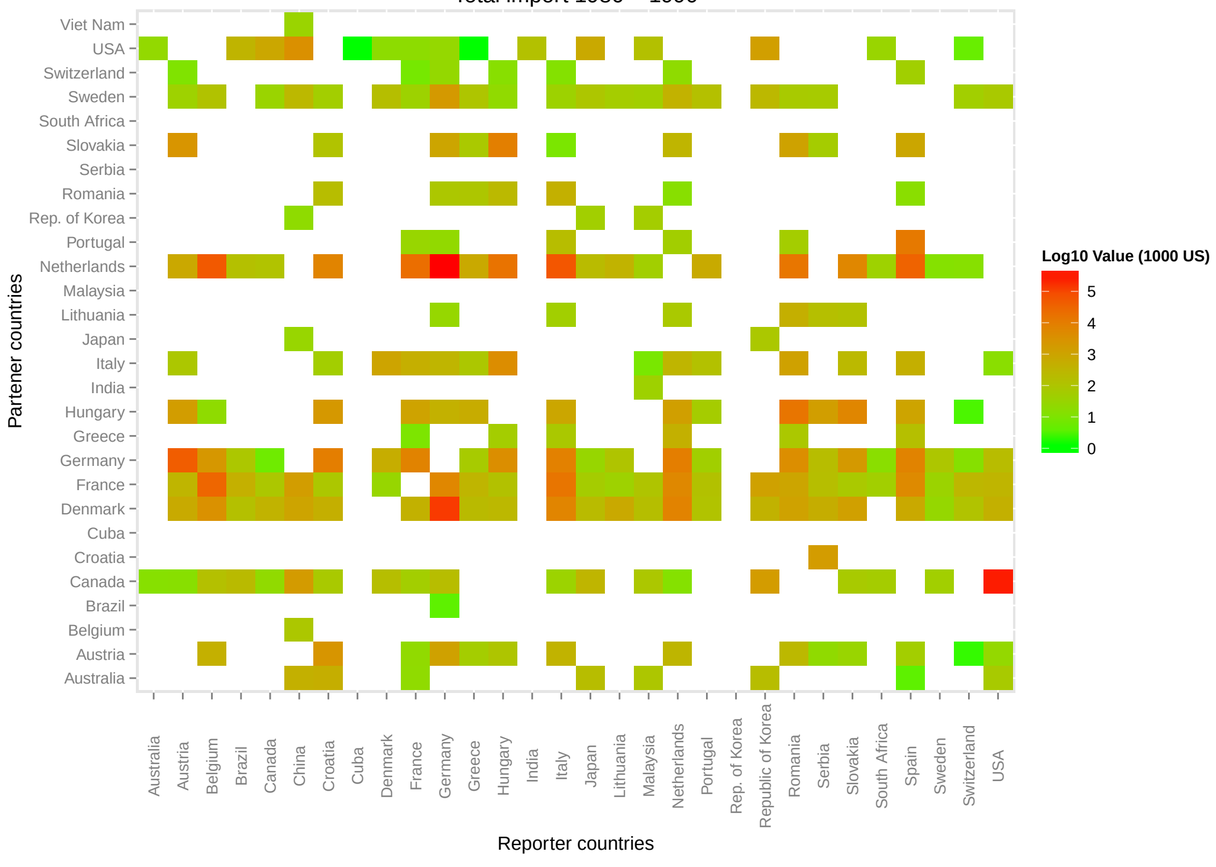


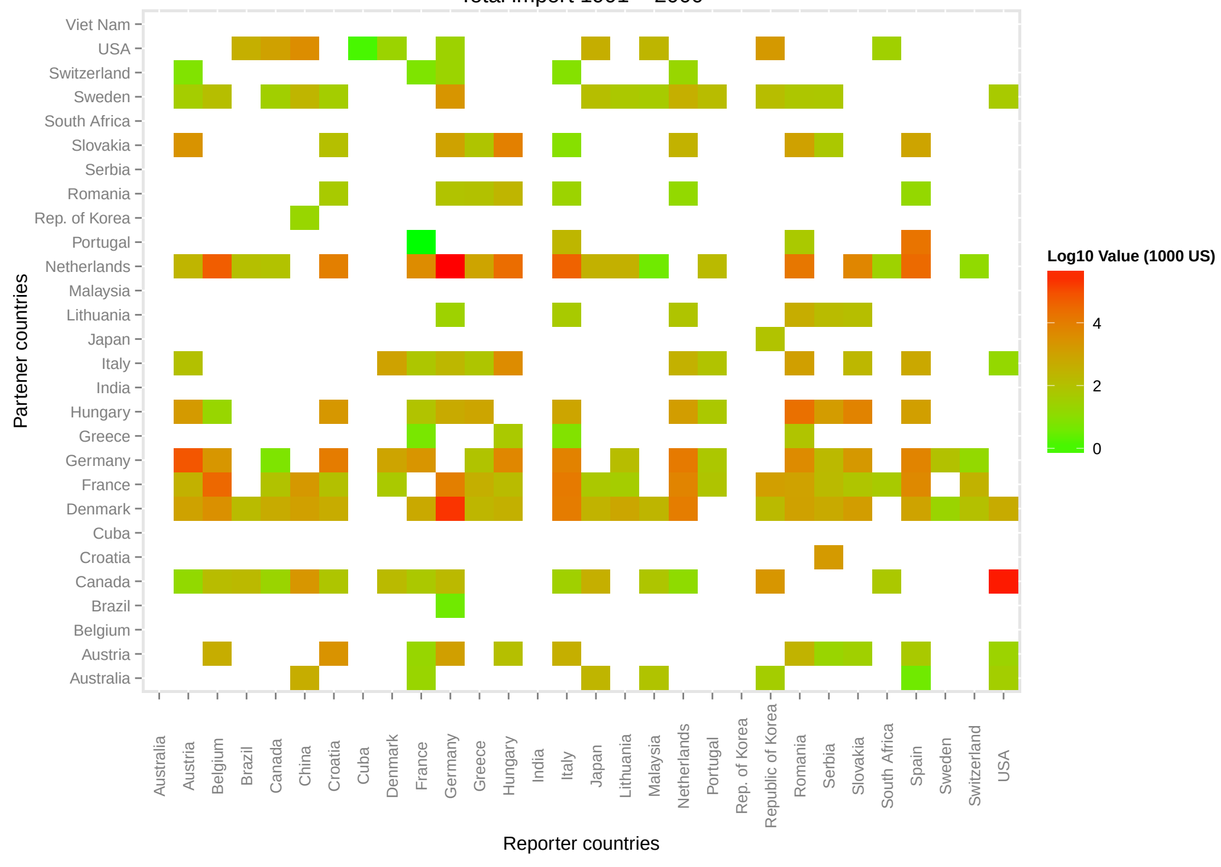
Supplementary figure 5a.

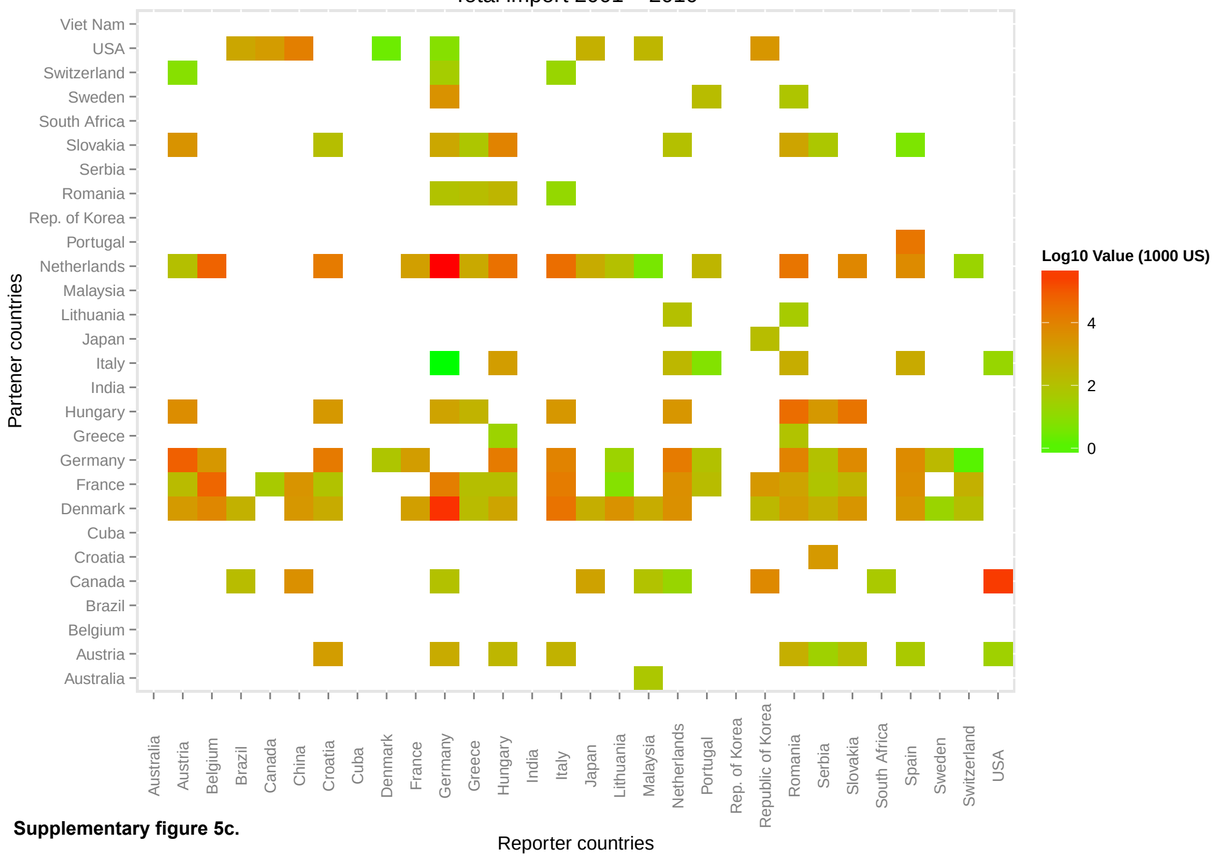


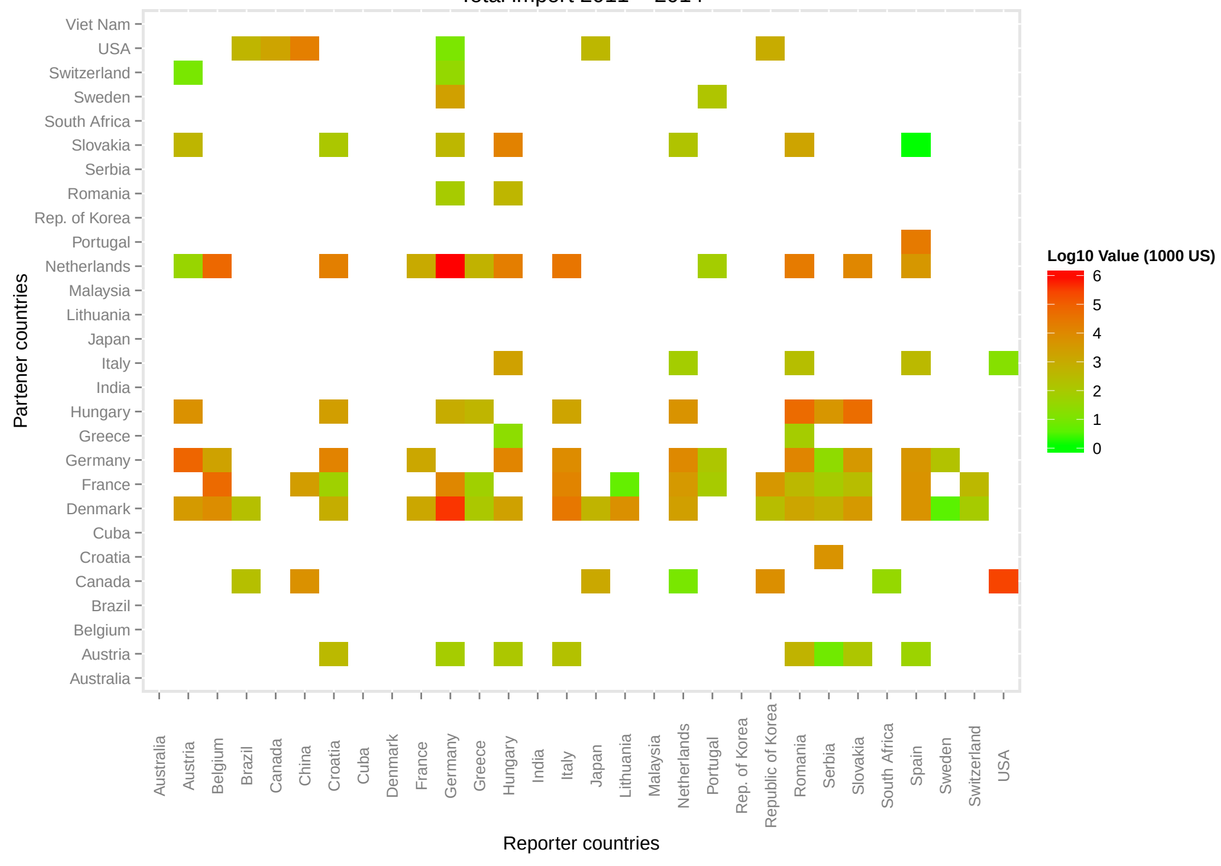


Supplementary figure 5b.

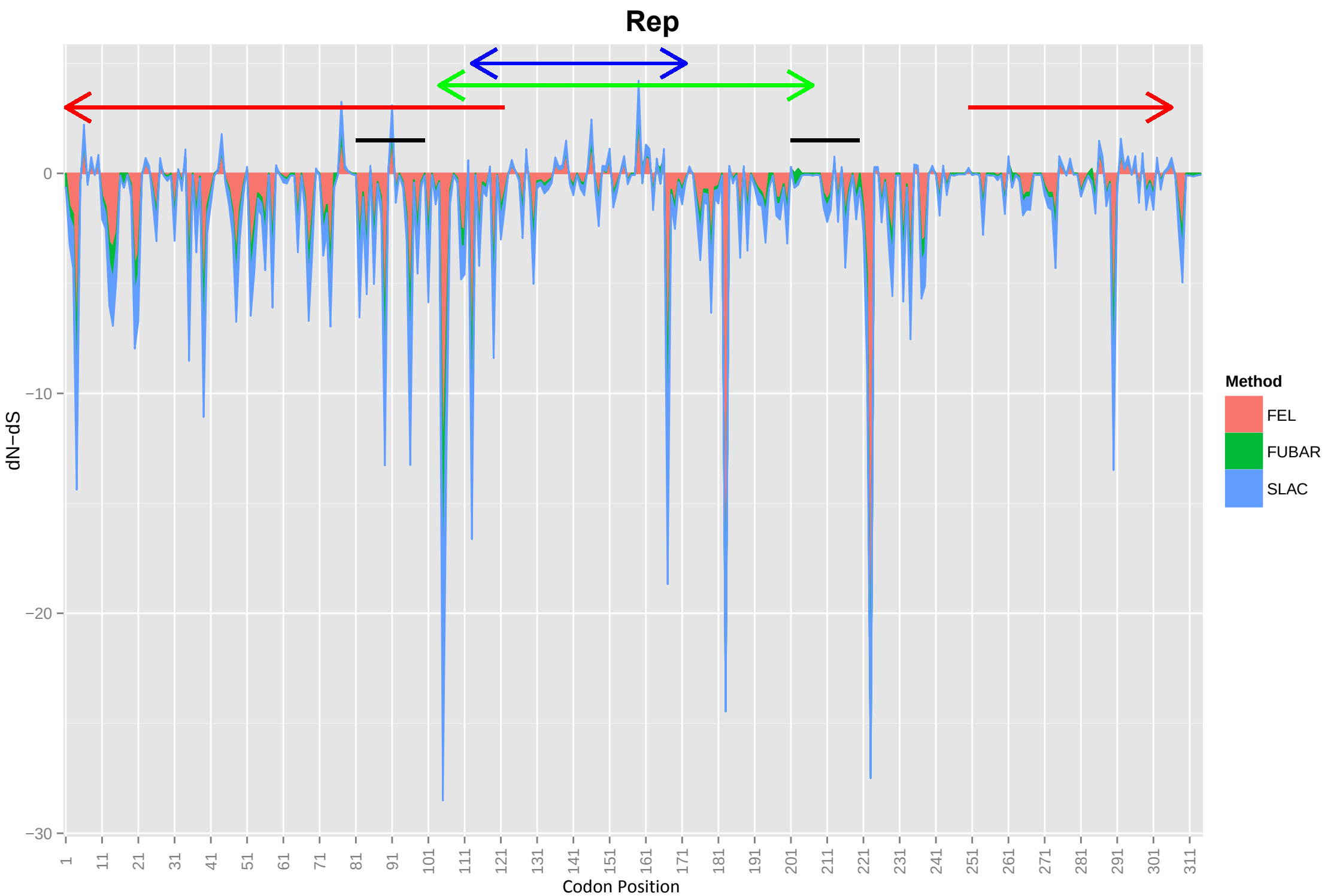




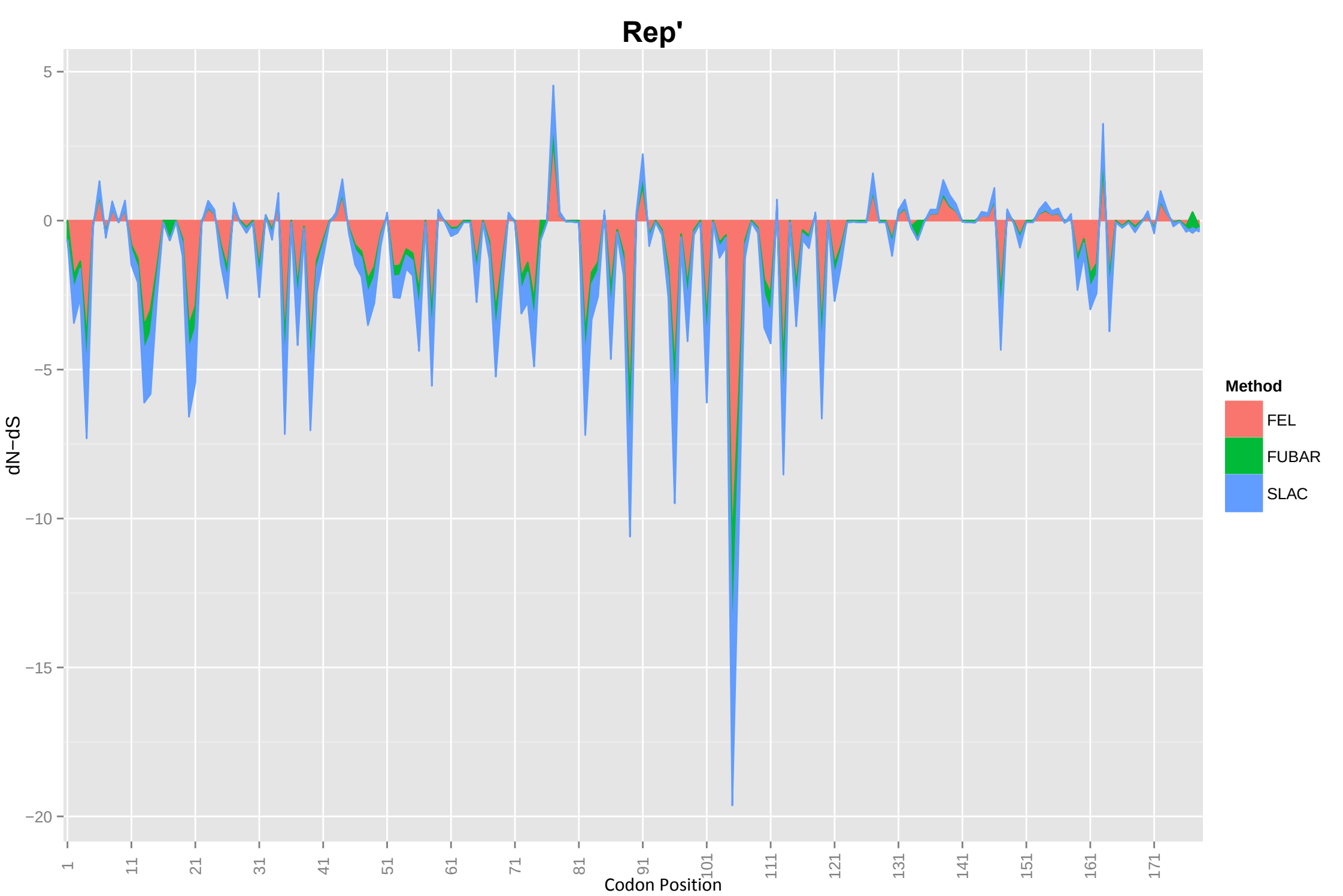




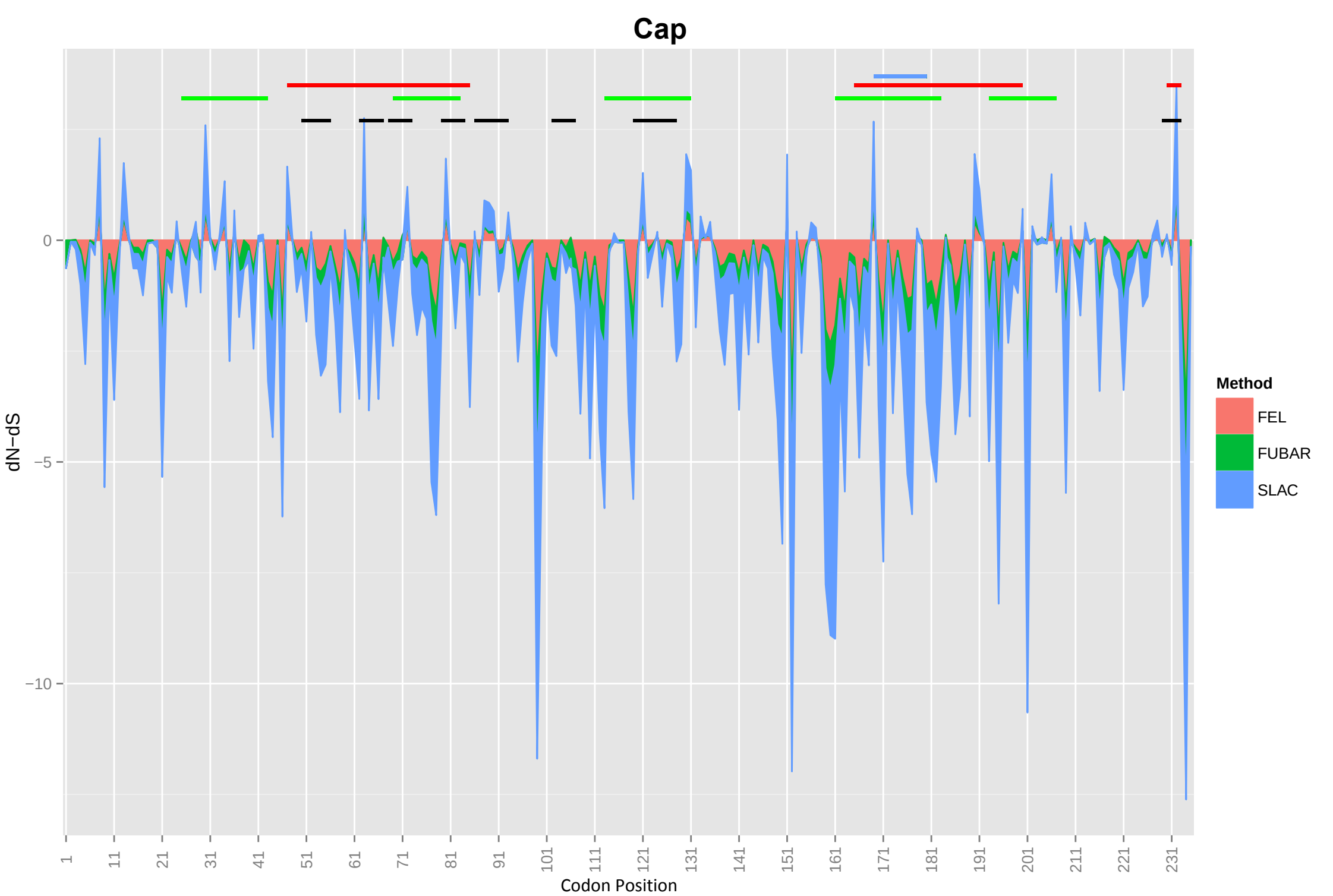
Supplementary figure 5c.



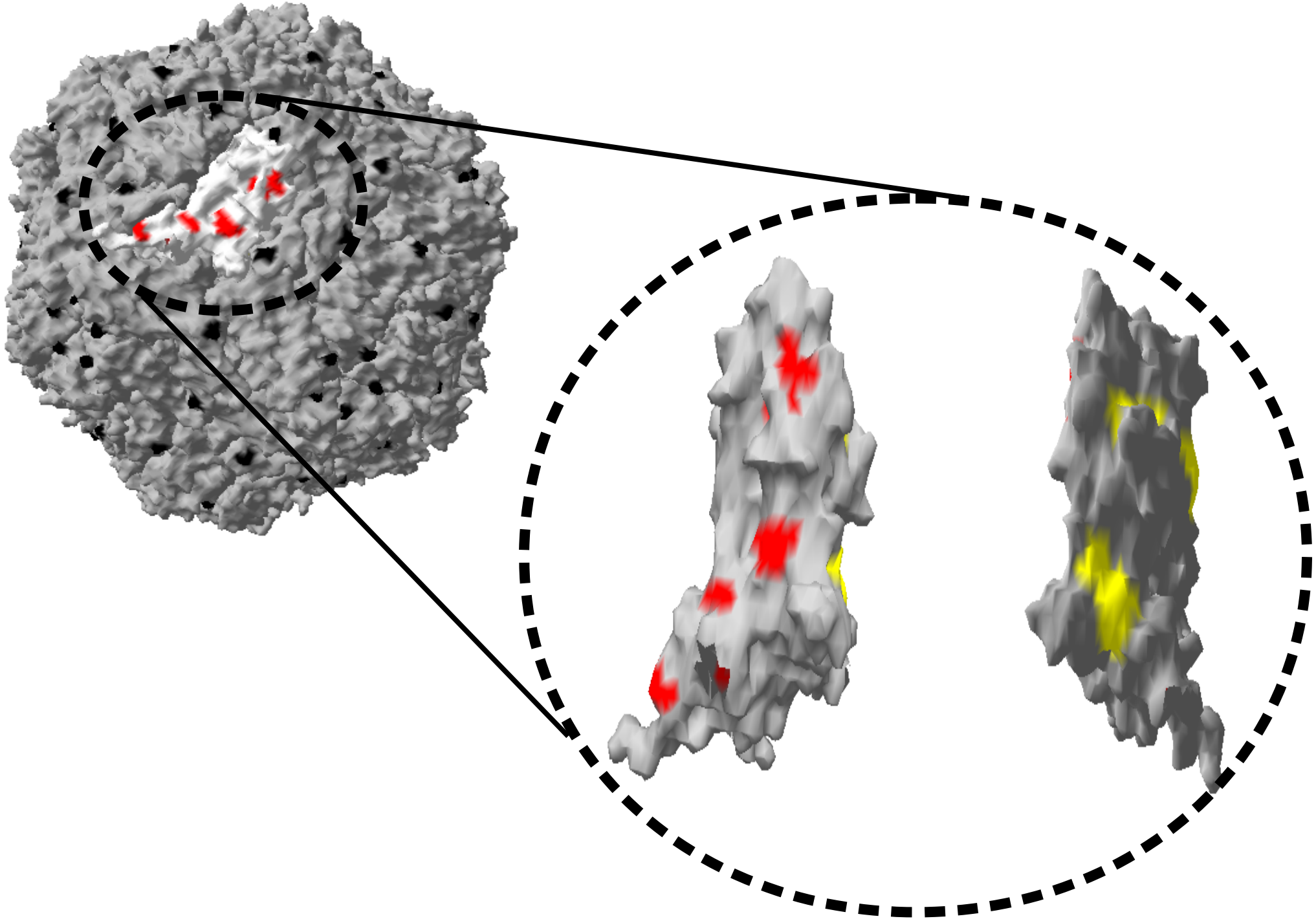
Supplementary figure 6a. Graphical display of the selective pressures (expressed as dN-dS) acting on each site of the Rep protein. Other overlapping ORFs, i.e. Rep',ORF3 andORF4, are represented as red, green and blue arrows, respectively. Epitopes identified within Rep protein are represented as black lines (Stevenson et al.,2007).



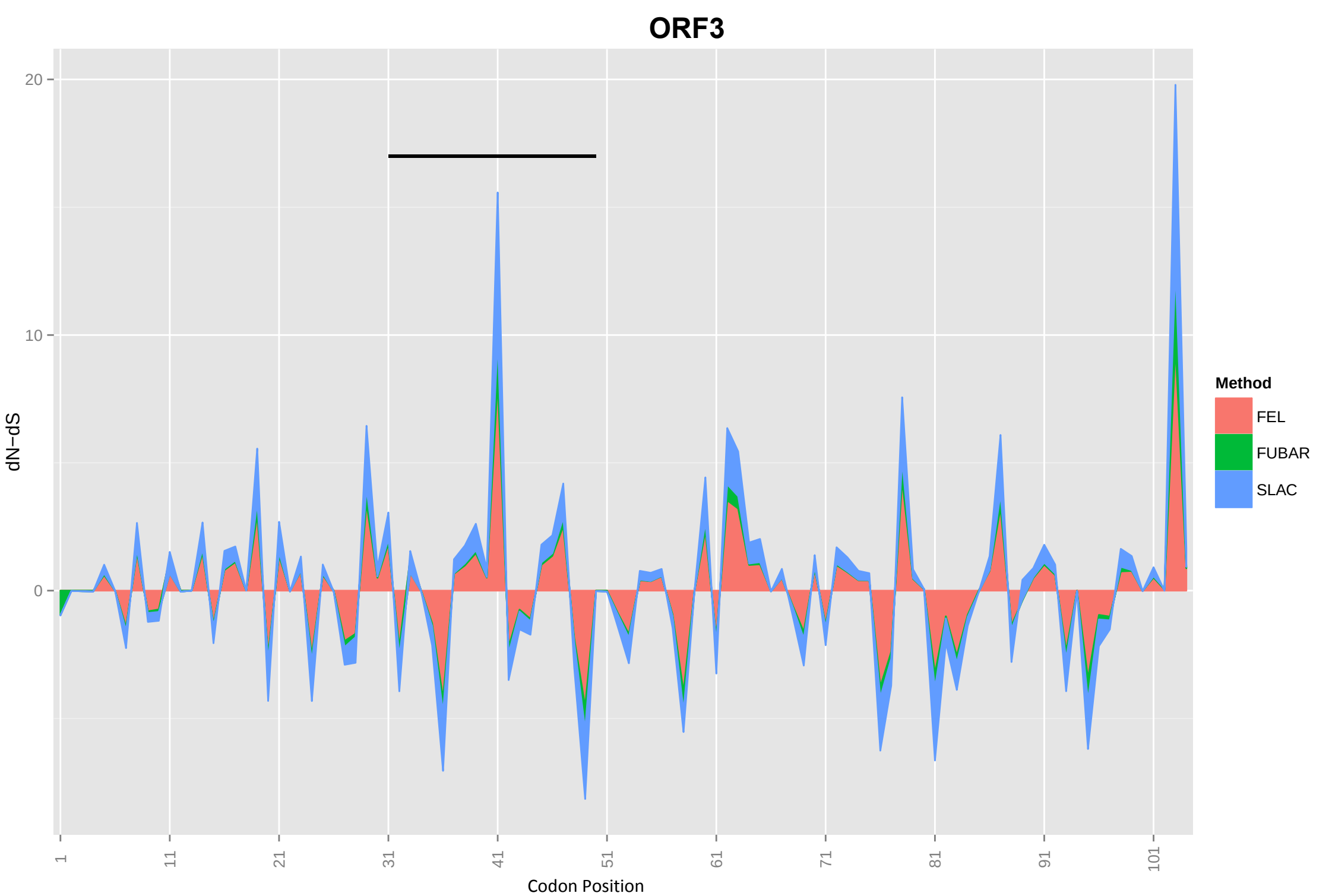
Supplementary figure 6b. Graphical display of the selective pressures (expressed as dN-dS) acting on each site of the Rep' protein.



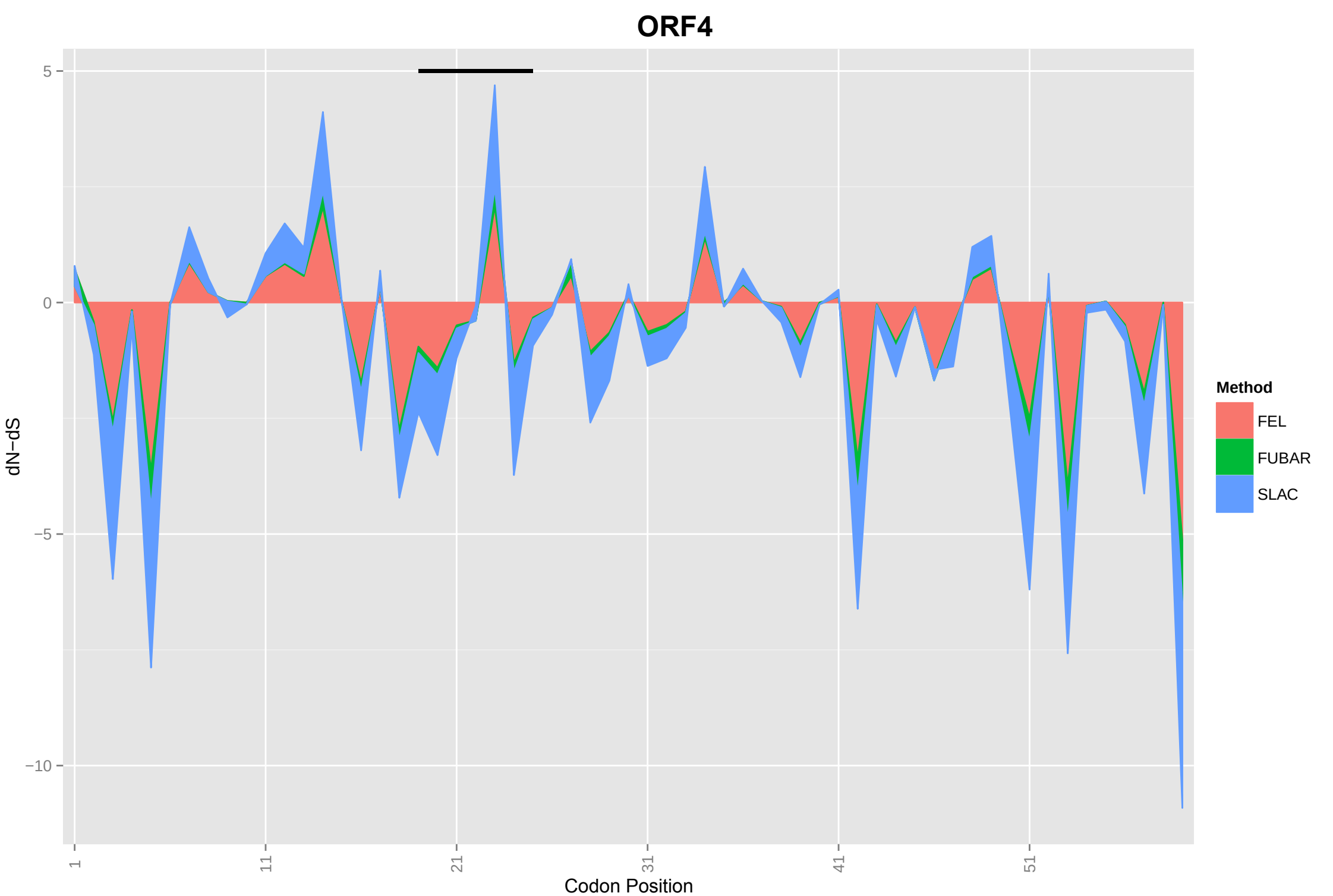
Supplementary figure 6c. Graphical display of the selective pressures (expressed as dN-dS) acting on each site of the Cap protein. Epitopes identified within Cap protein by Trible et al.,(2012), Lekcharoensuk et al.,(2004), Mahé et al.,(2000) and Meng Ge et al.,(2013) are represented as blue, red, green and black lines, respectively.



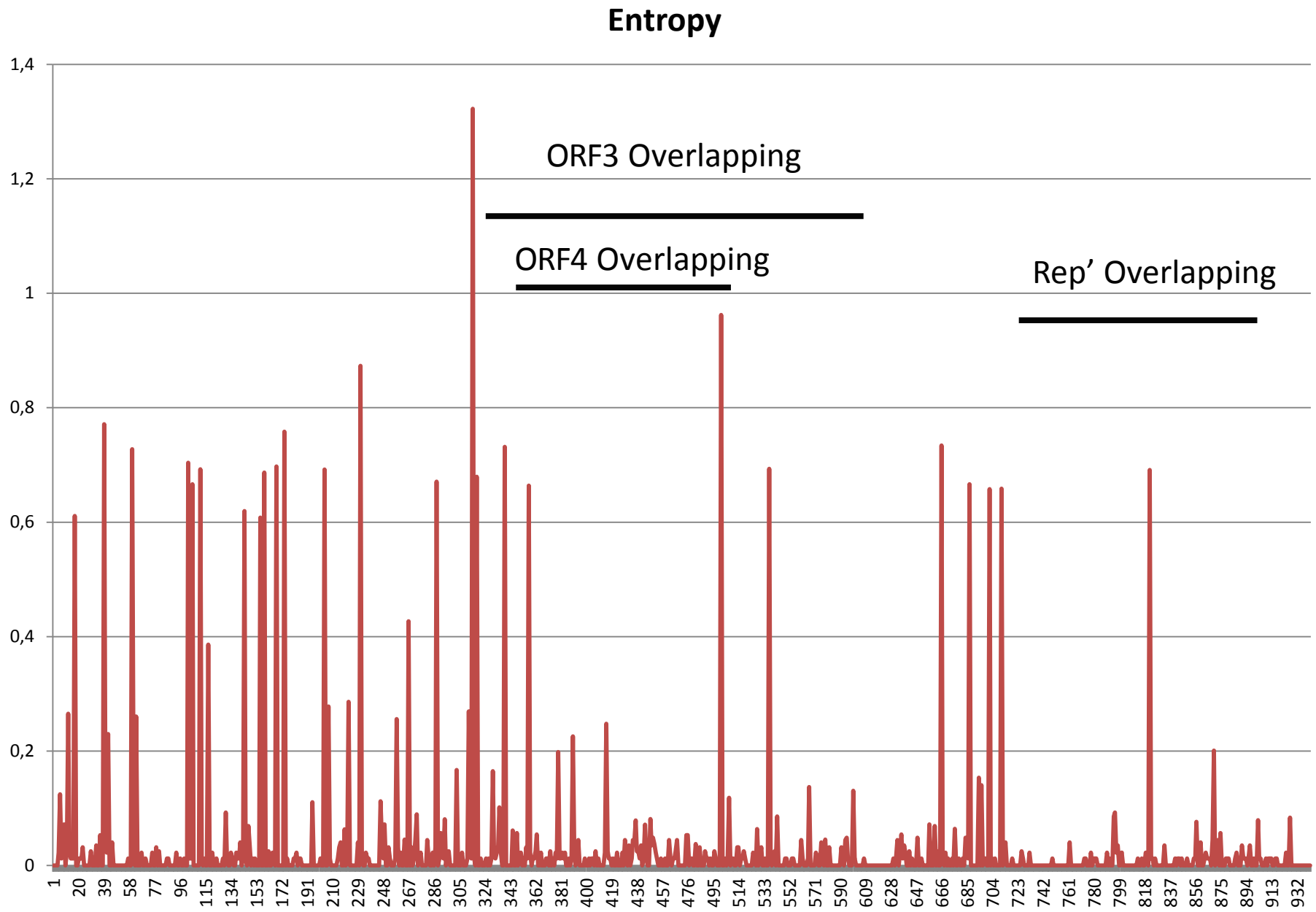
Supplementary figure 6d. The PCV2 capsid is represented highlighting in red the amino acids under positive selective pressure exposed on the virus's surface. In the bottom-left figure a single capsomer is magnified highlighting in red and yellow the sites under positive selection exposed respectively to the external and internal surface of the capsid.



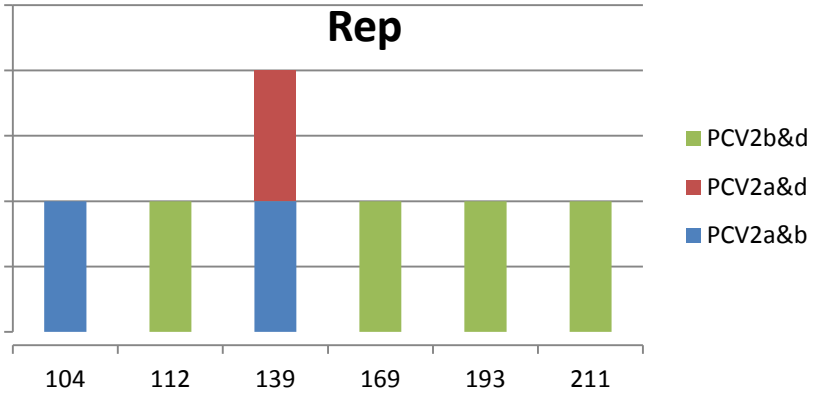
Supplementary figure 6e. Graphical display of the selective pressures (expressed as dN-dS) acting on each site of the ORF3 protein. Epitopes identified within ORF3 protein are represented as black lines (Stevenson et al.,2007).

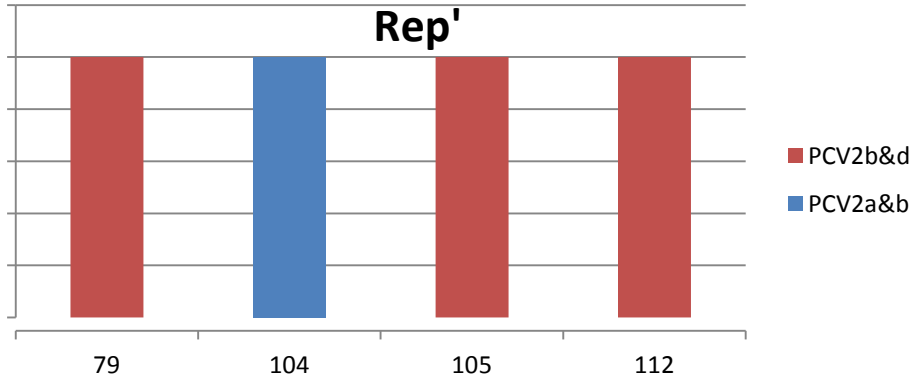


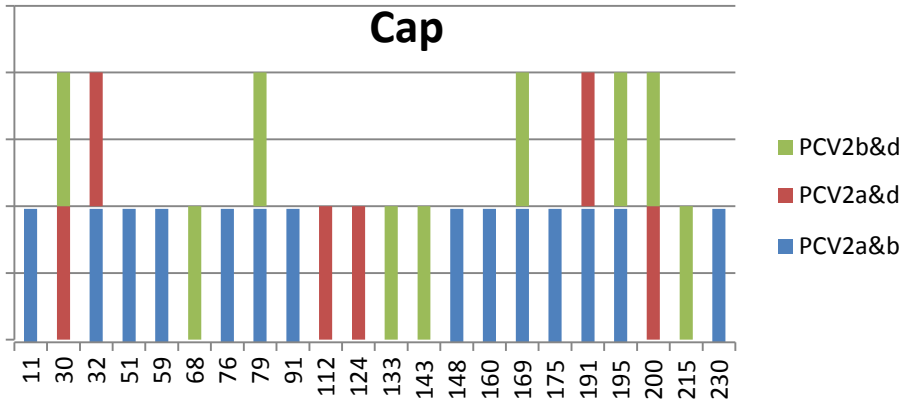
Supplementary figure 6f. Graphical display of the selective pressures (expressed as dN-dS) acting on each site of the ORF4 protein. Epitopes identified within ORF4 protein are represented as black lines (Jialing He et al.,2013).

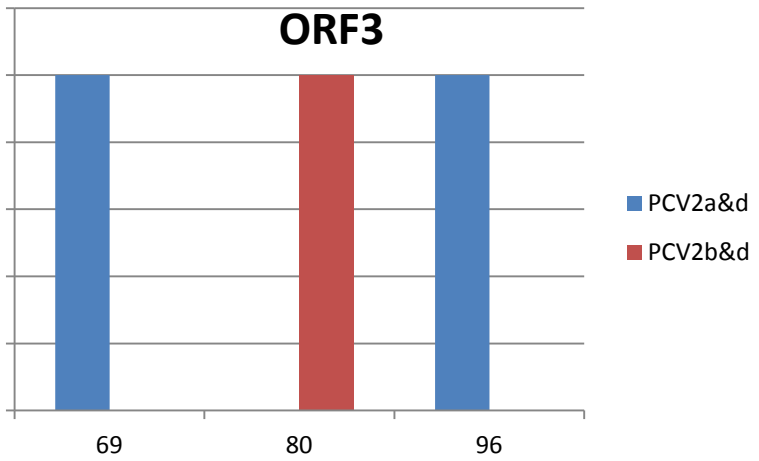


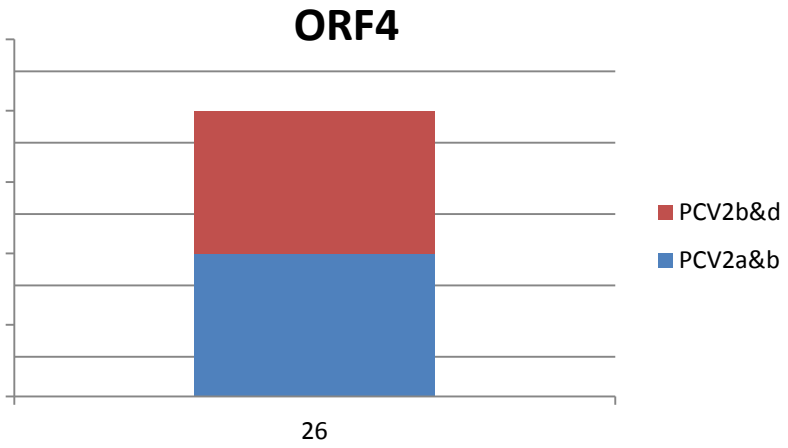
Supplementary figure 7. Shannon entropy values have been calculated for each codon position in the ORF1 gene. The region encoding for the ORF3, ORF4 and Rep' are reported.





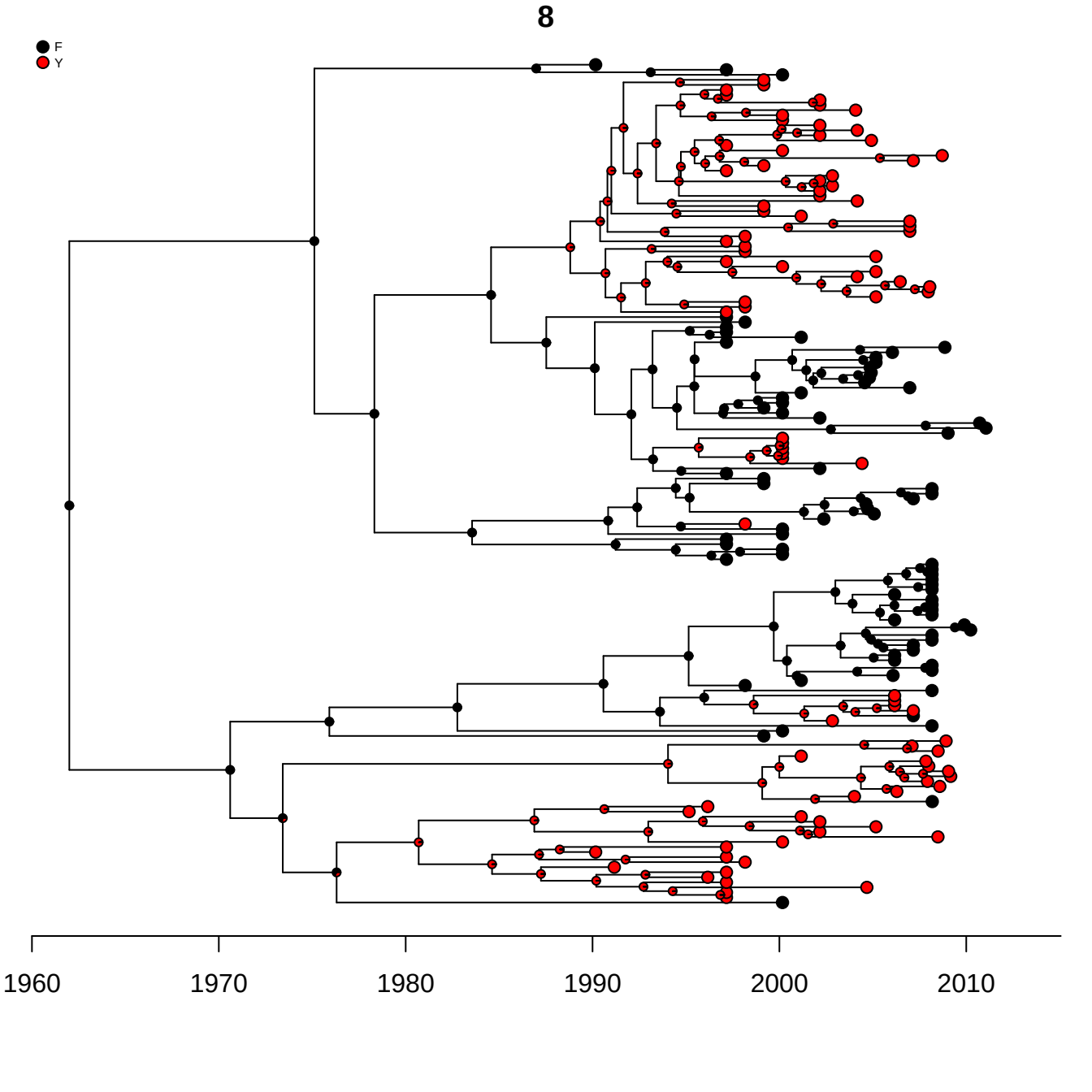


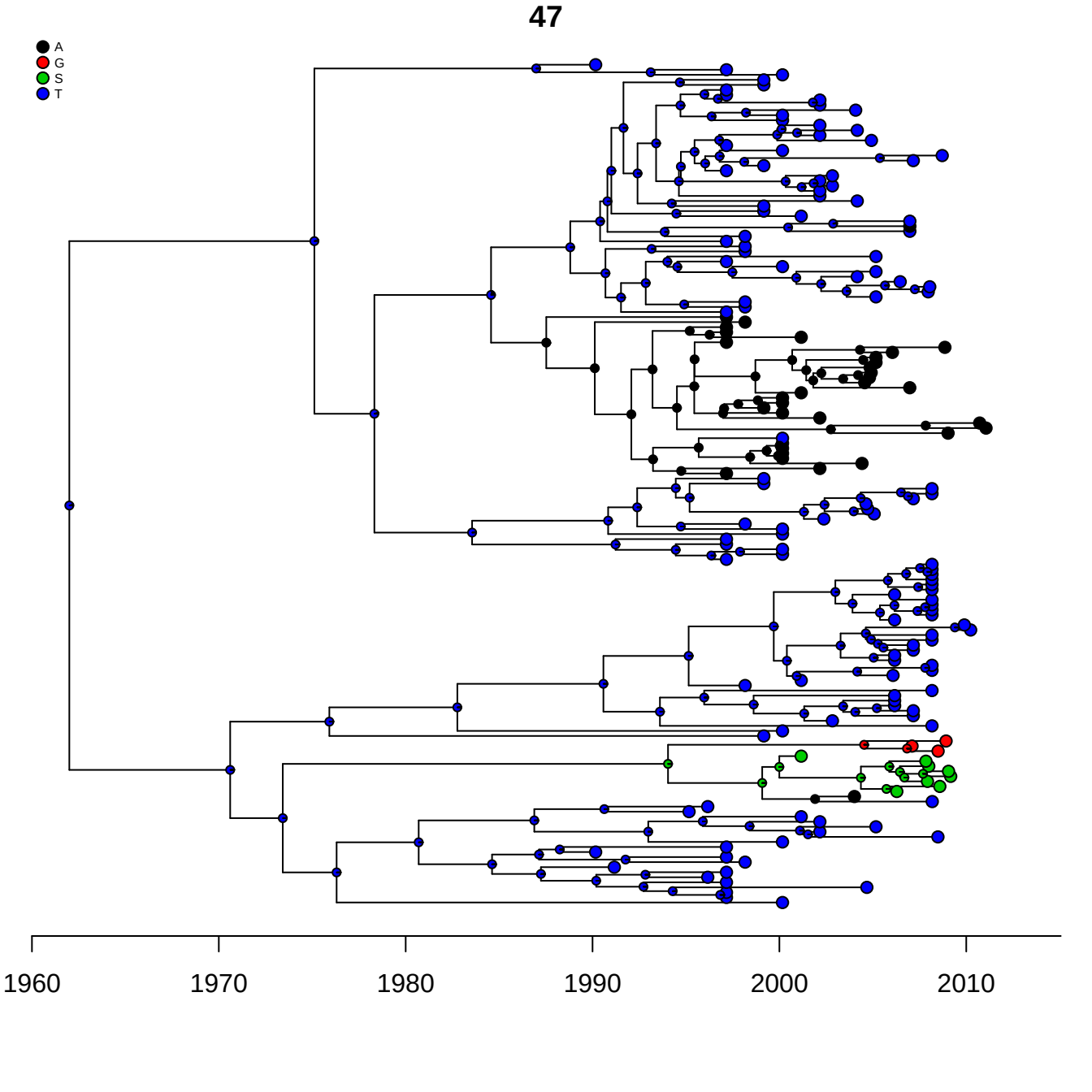


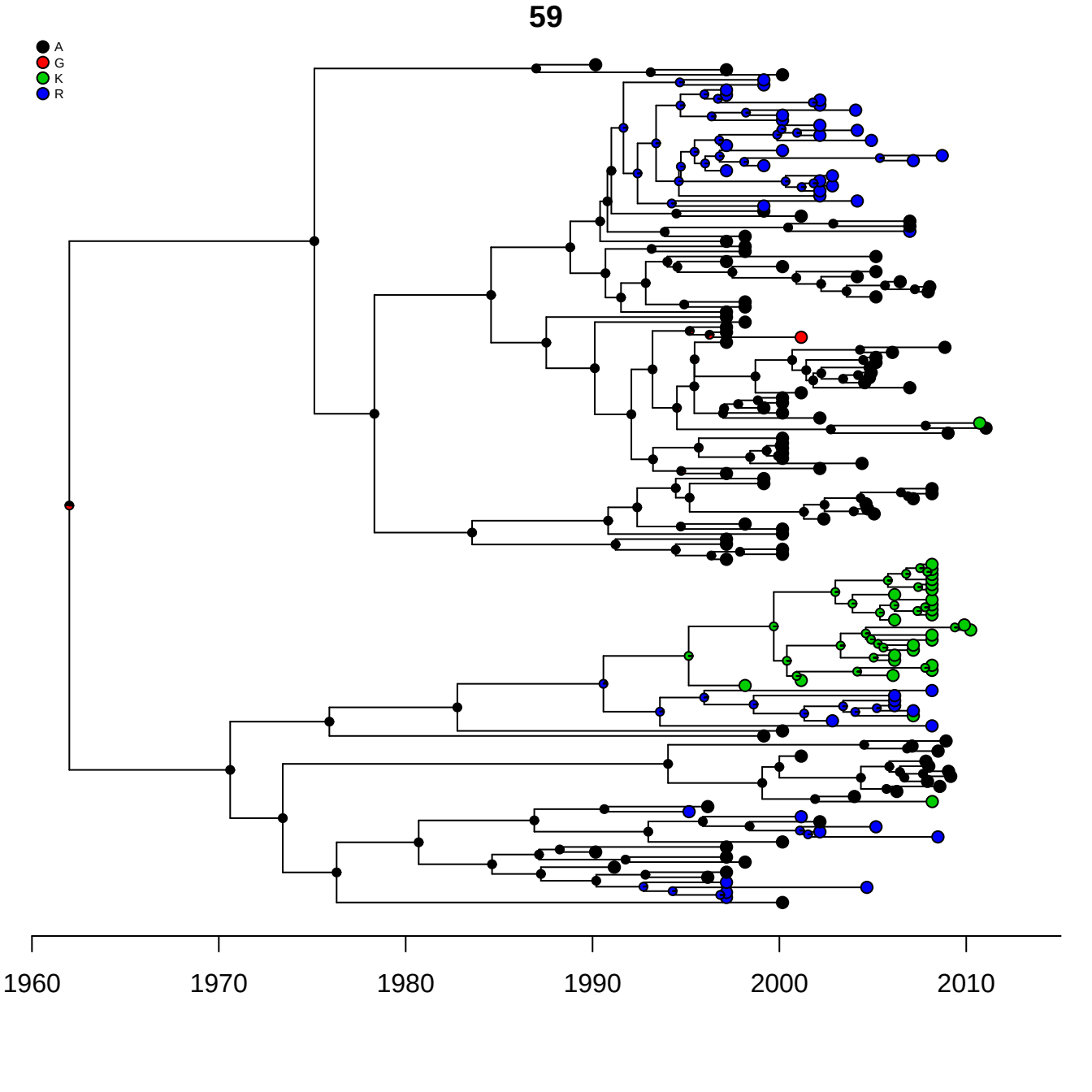


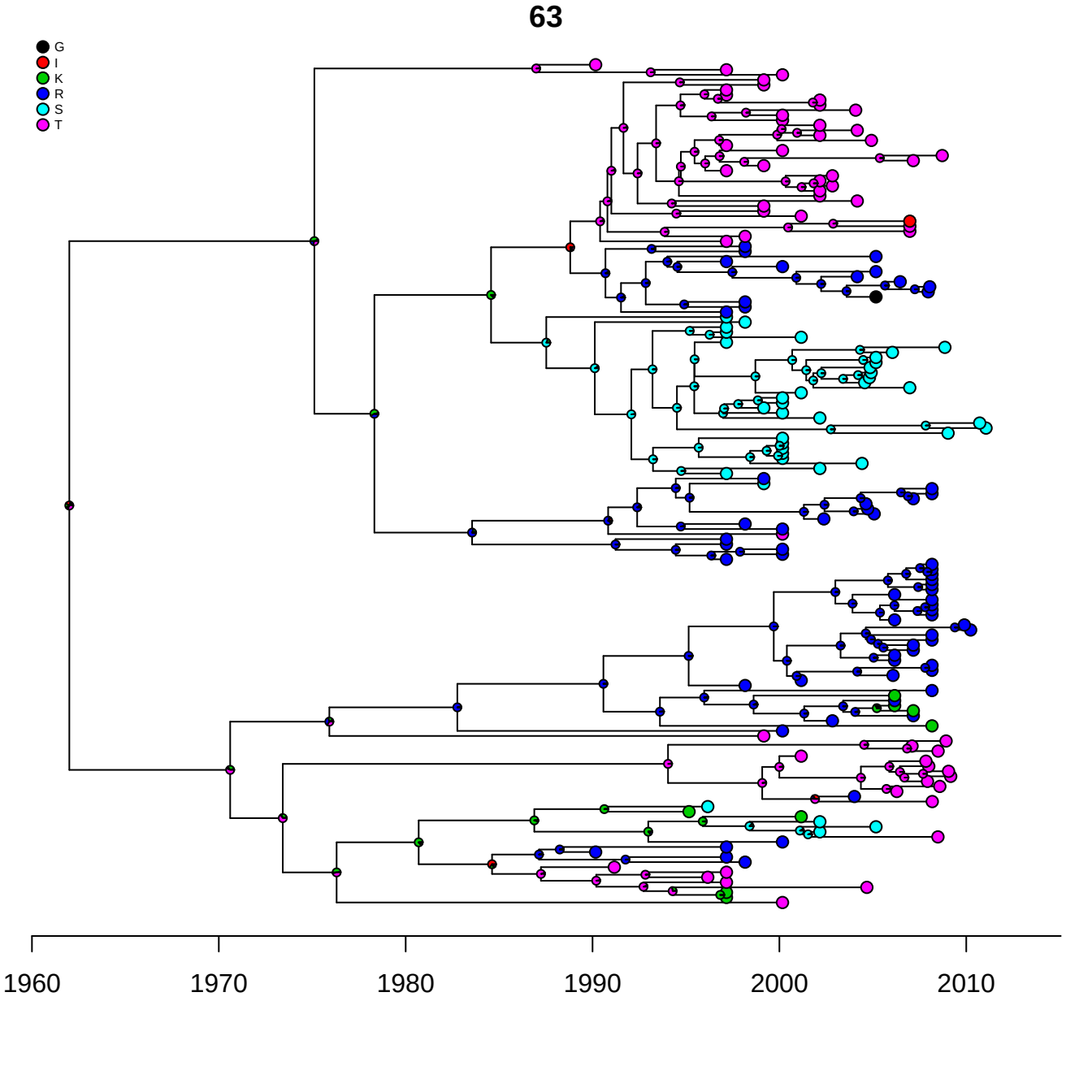
Supplementary figure8. Codons reported to be under different selective forces between different genotypes combinations. Color coded bars associated to each codon position represent the presence of a statistically significant difference (p-value<0,05) in selective pressure acting on different genotypes pairs.

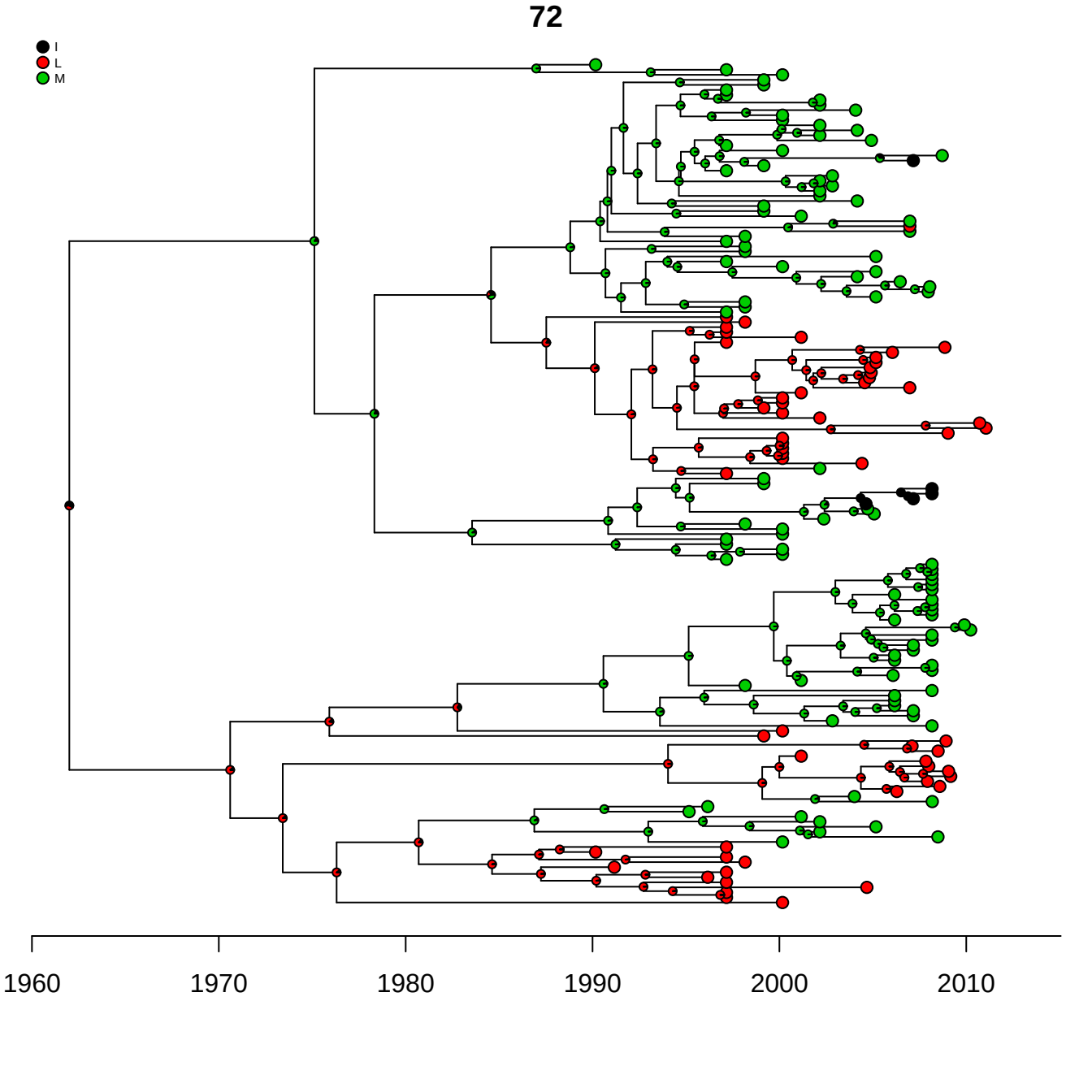
Supplementary figure 9. PCV2a maximum likelihood based reconstruction of amino acid ancestral states. Only variable positions are reported.

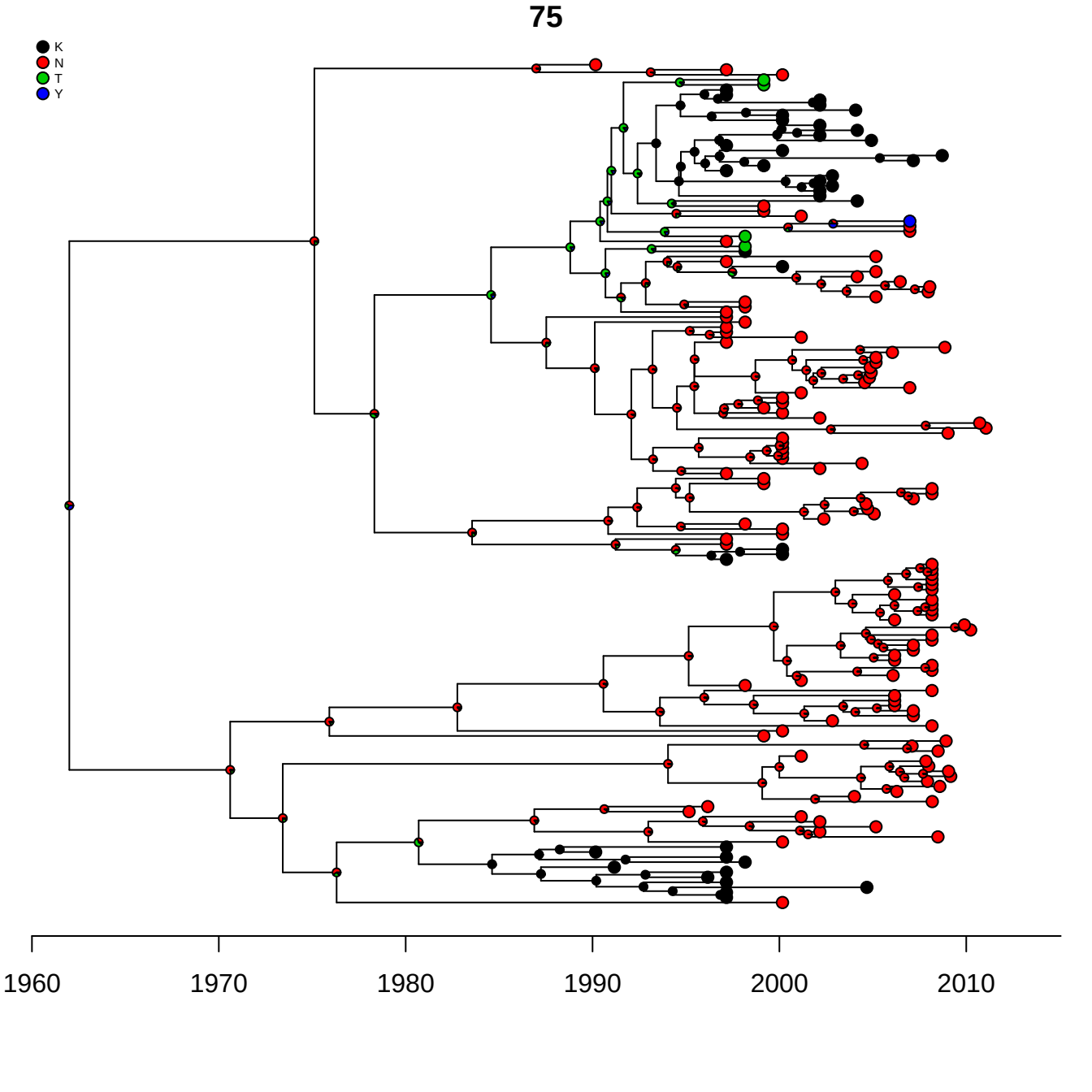

8
F
Y
1960
1970
1980
1990
2000
2010

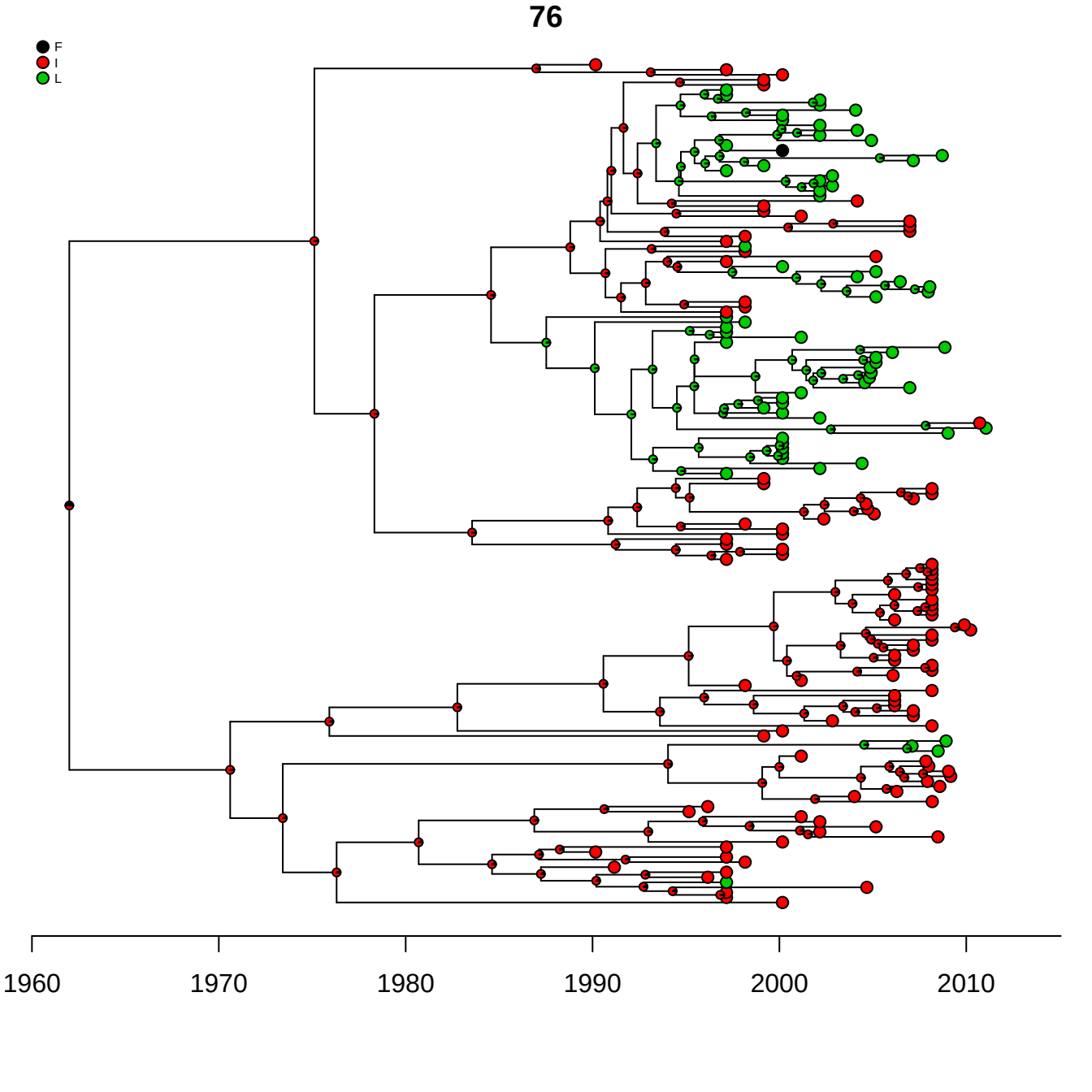

47
A
G
S
T
1960
1970
1980
1990
2000
2010

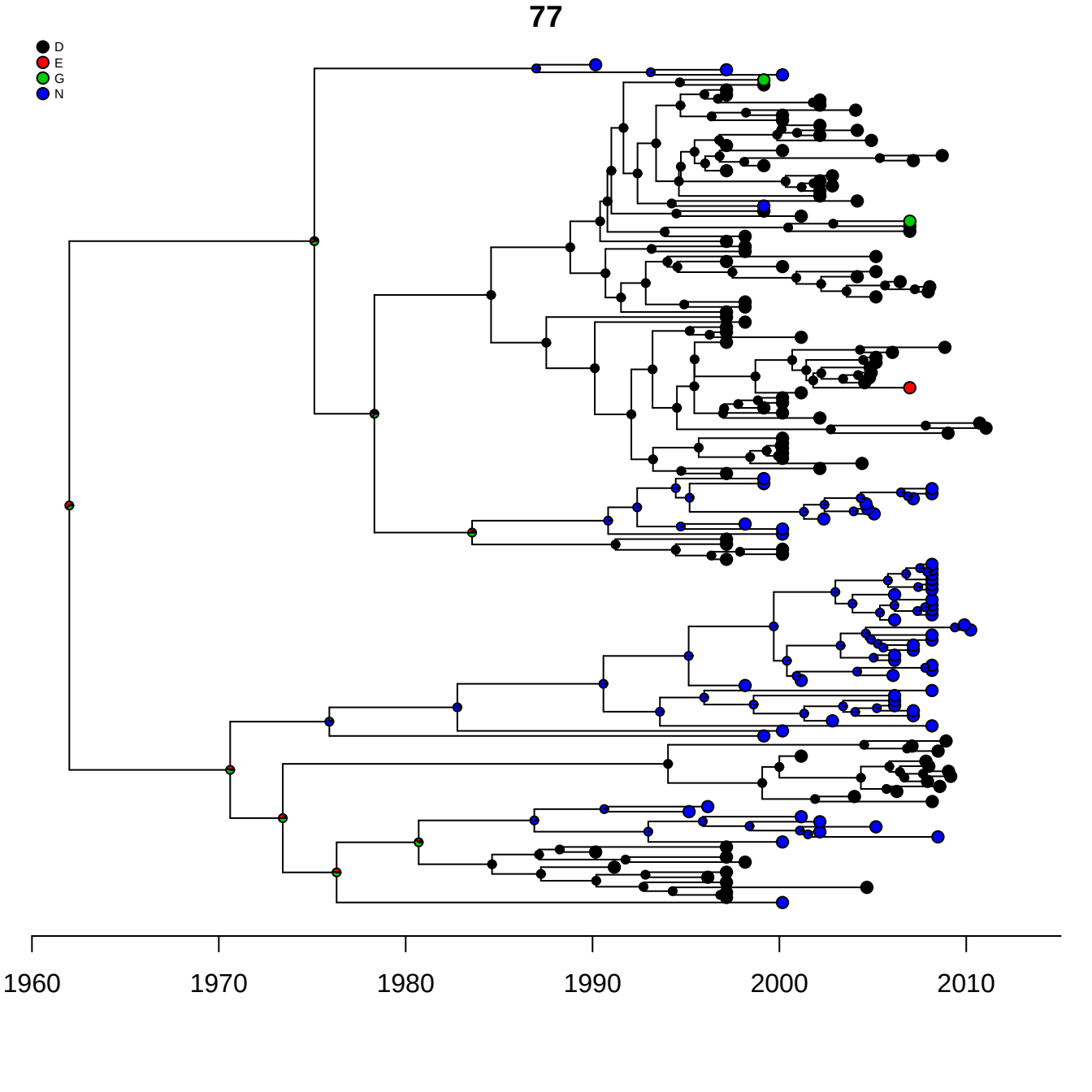

59
A
G
K
R
1960
1970
1980
1990
2000
2010

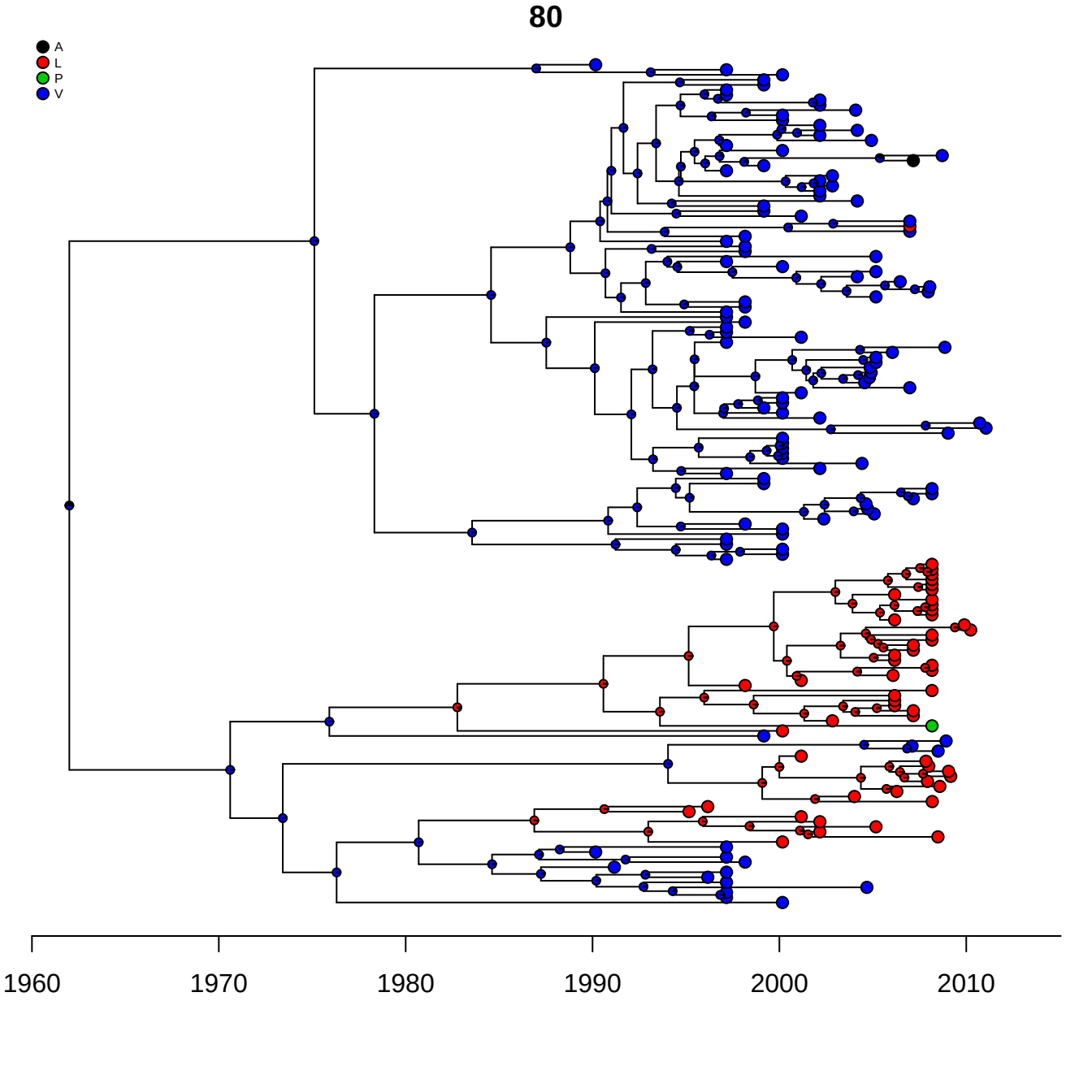

63
G
I
K
R
S
T
1960
1970
1980
1990
2000
2010

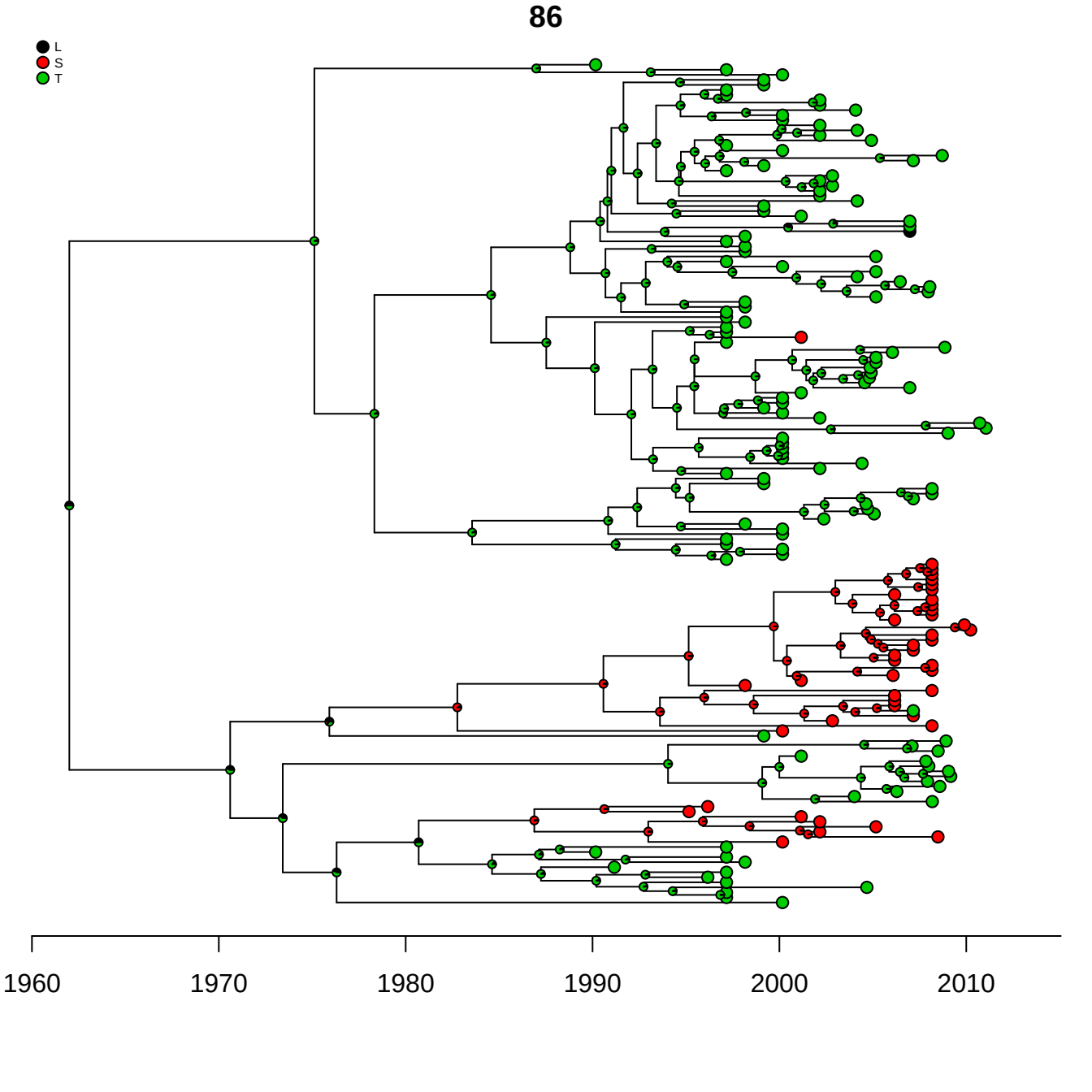

72
I
L
M
1960
1970
1980
1990
2000
2010

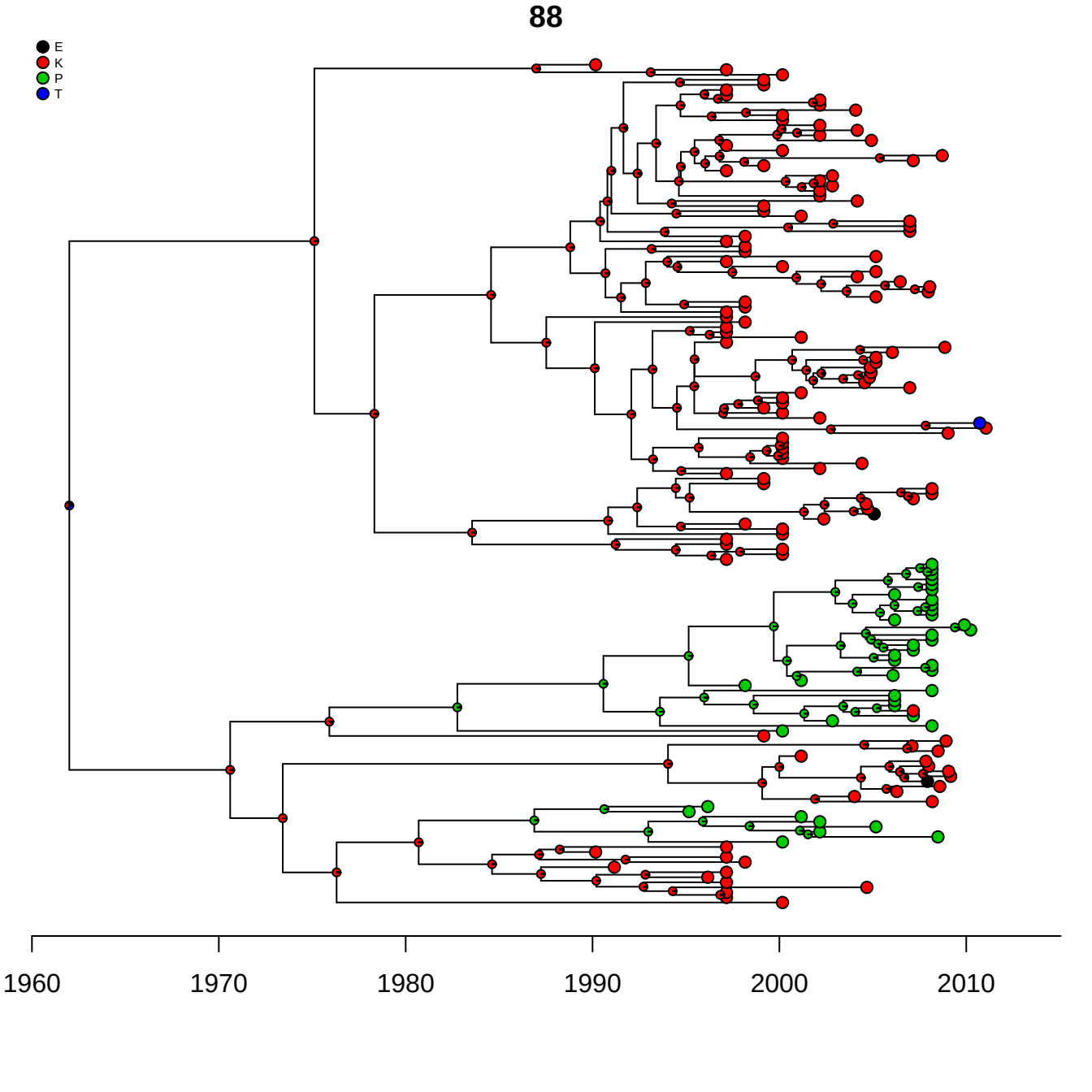

75
K
N
T
Y
1960
1970
1980
1990
2000
2010

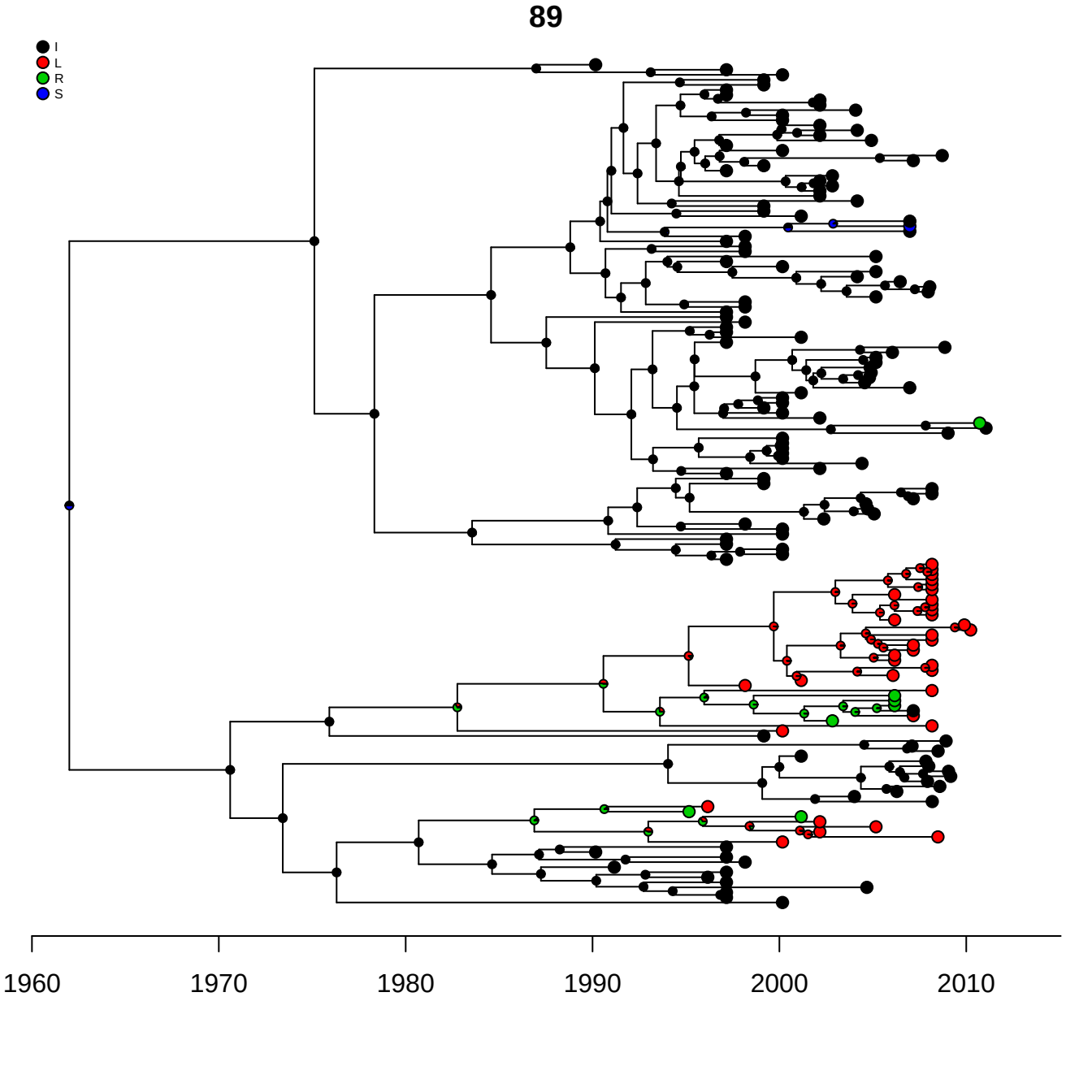

76
F
I
L
1960
1970
1980
1990
2000
2010

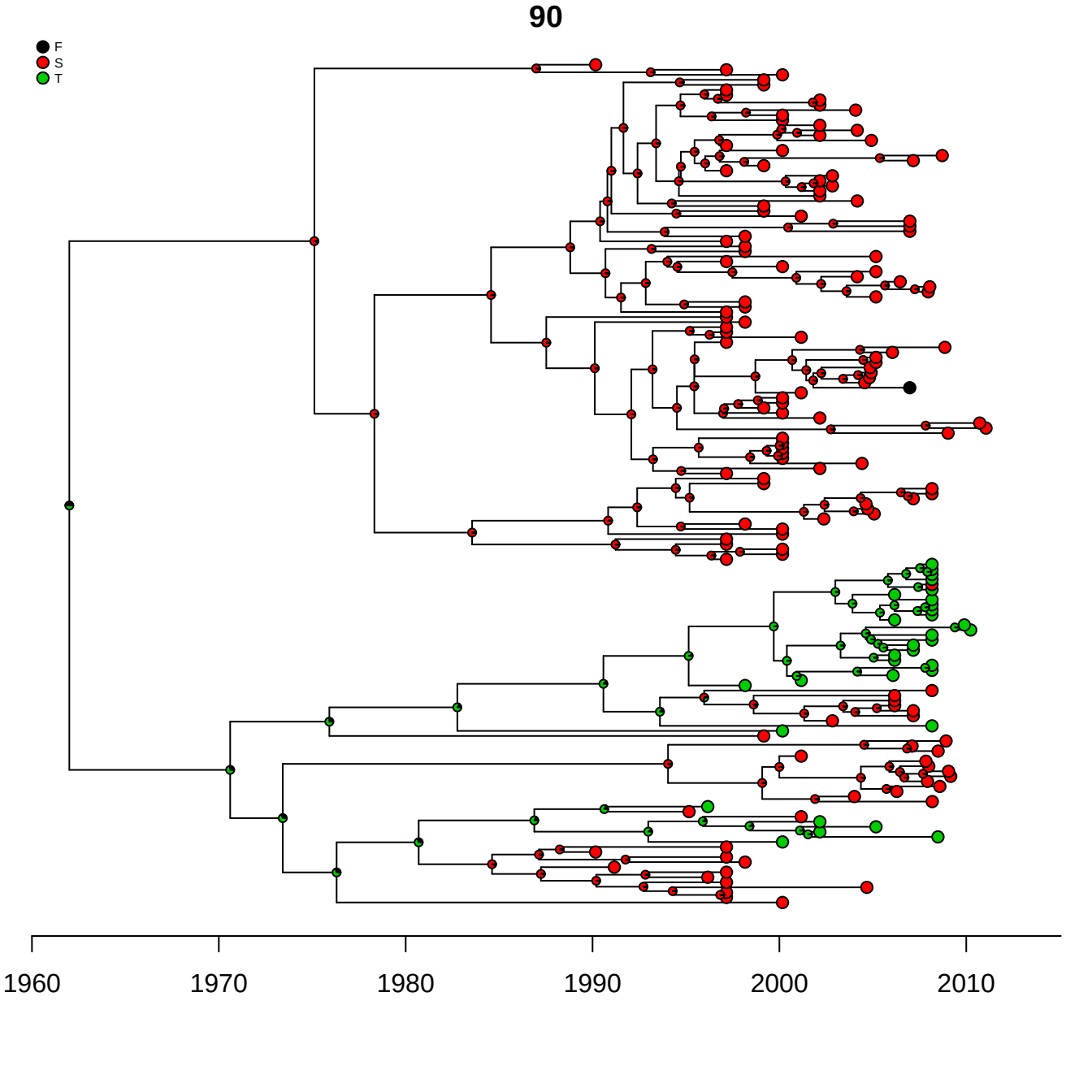

77
D
E
G
N
1960
1970
1980
1990
2000
2010

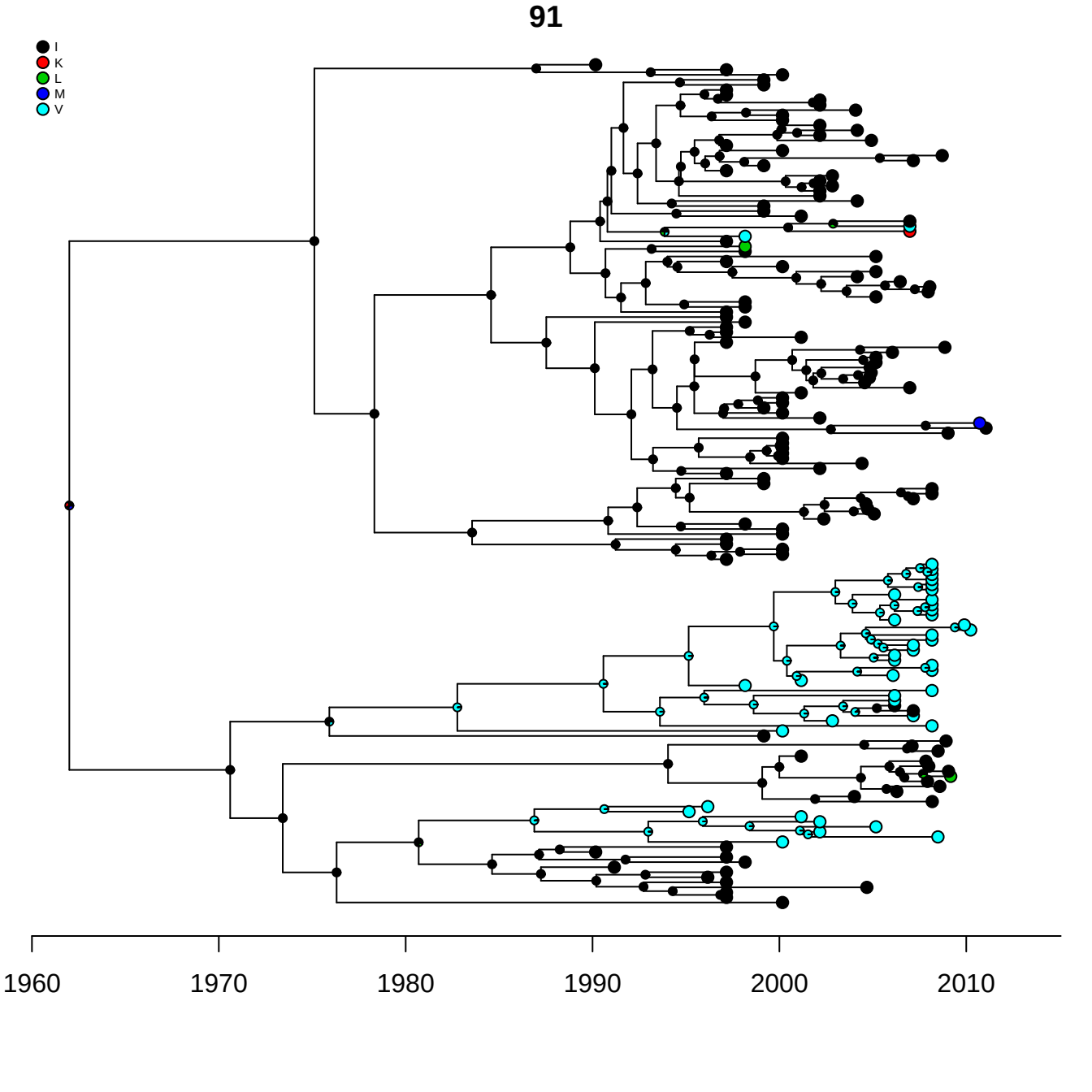

80
A
L
P
V
1960
1970
1980
1990
2000
2010

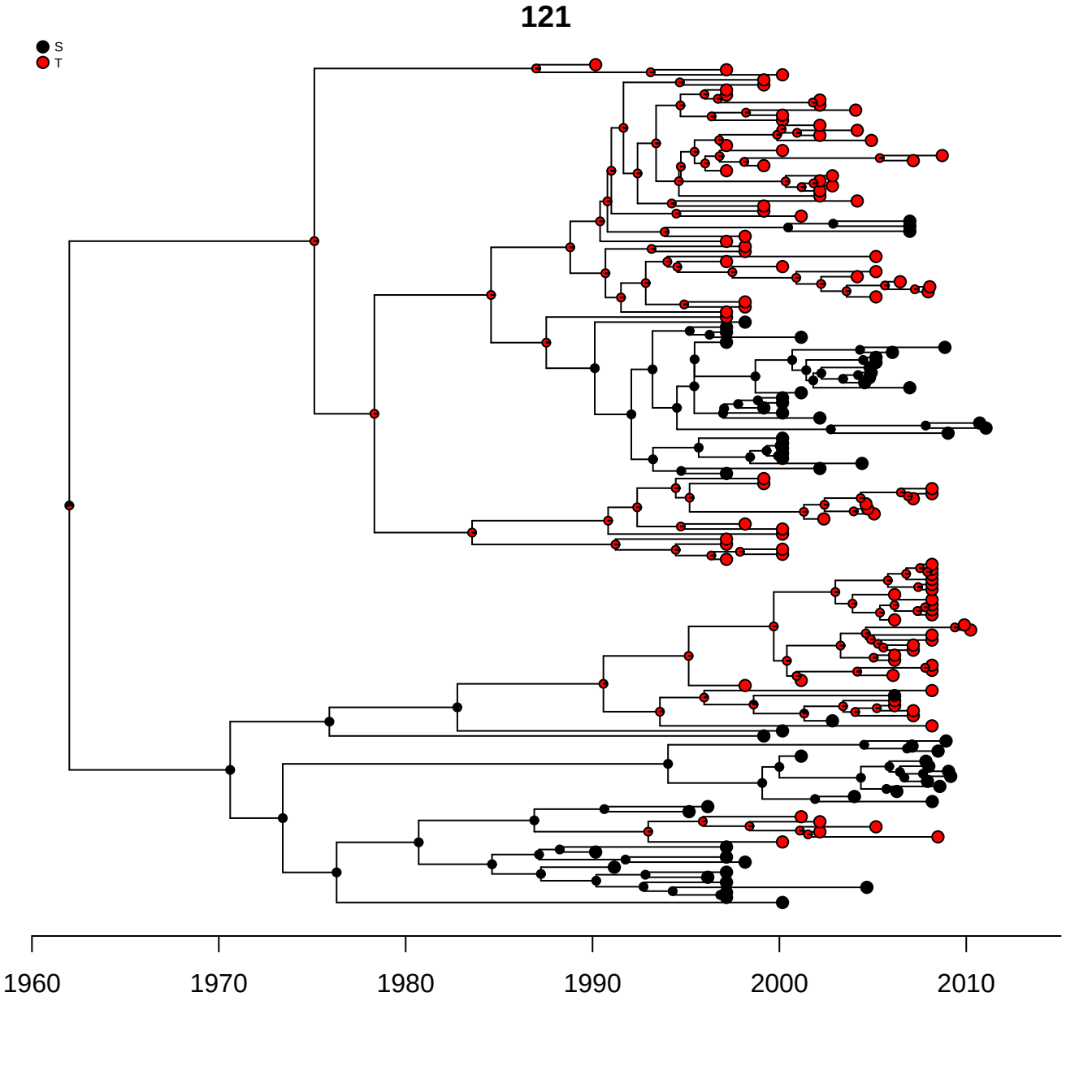

86
L
S
T
1960
1970
1980
1990
2000
2010

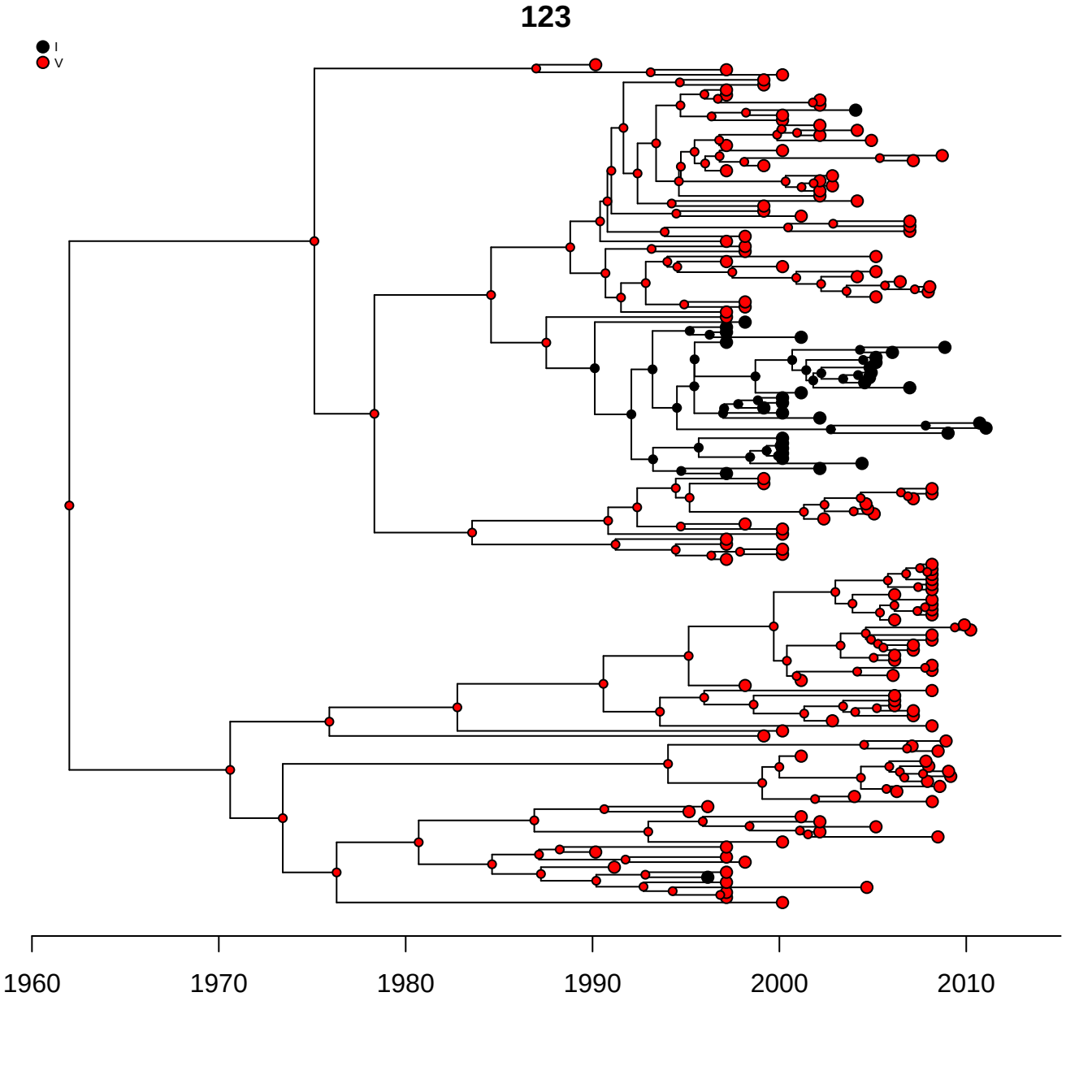

88
E
K
P
T
1960
1970
1980
1990
2000
2010

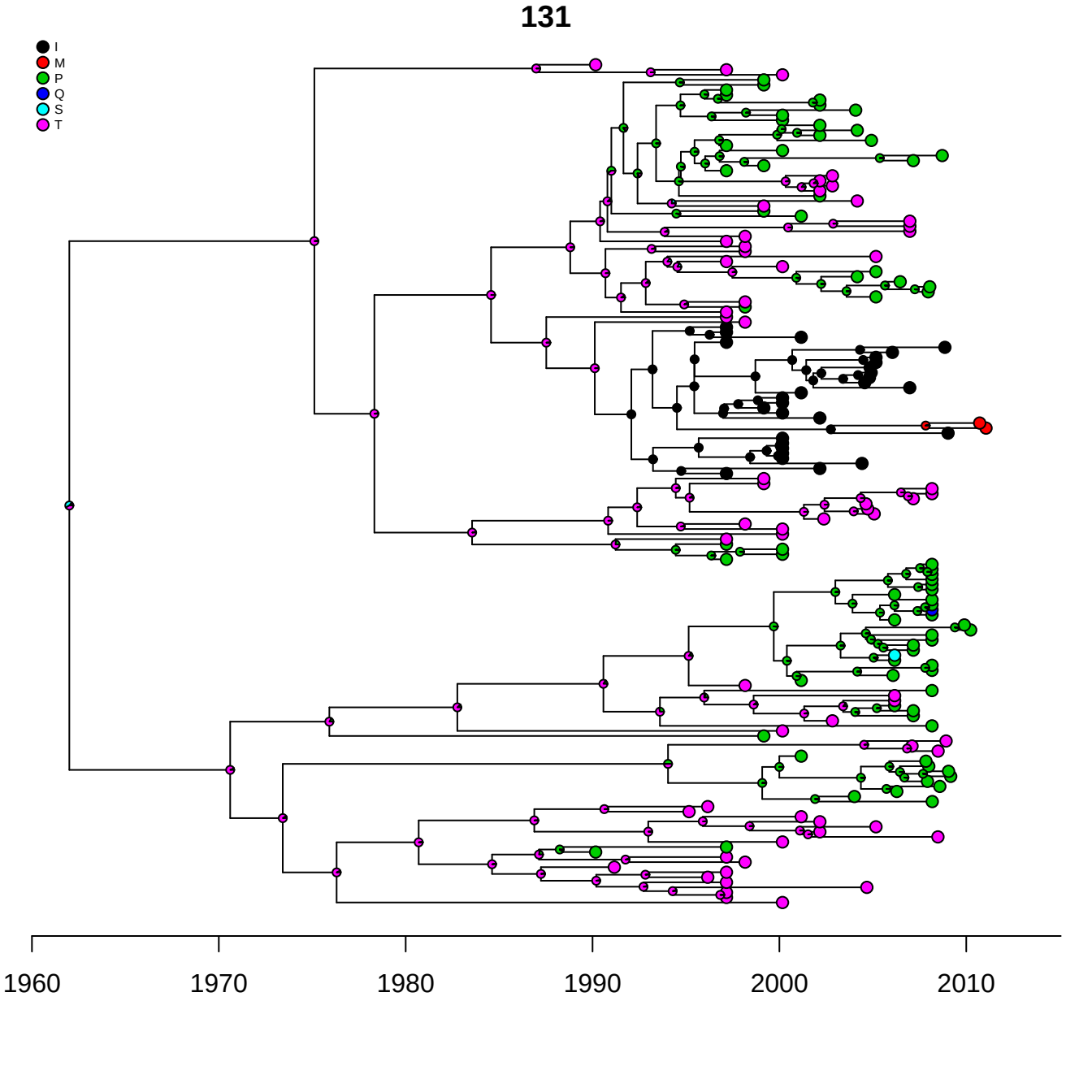

89
I
L
R
S
1960
1970
1980
1990
2000
2010

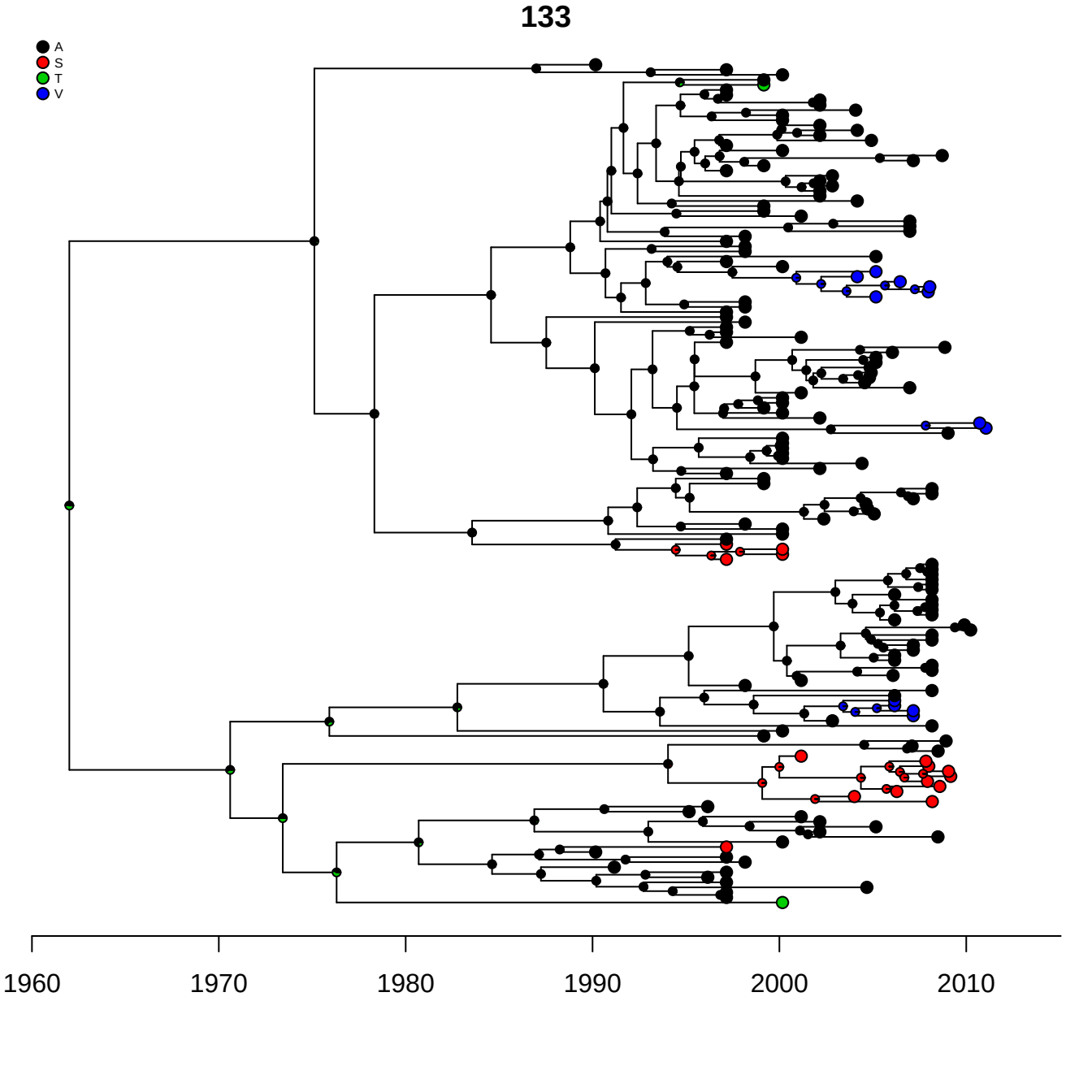

90
F
S
T
1960
1970
1980
1990
2000
2010

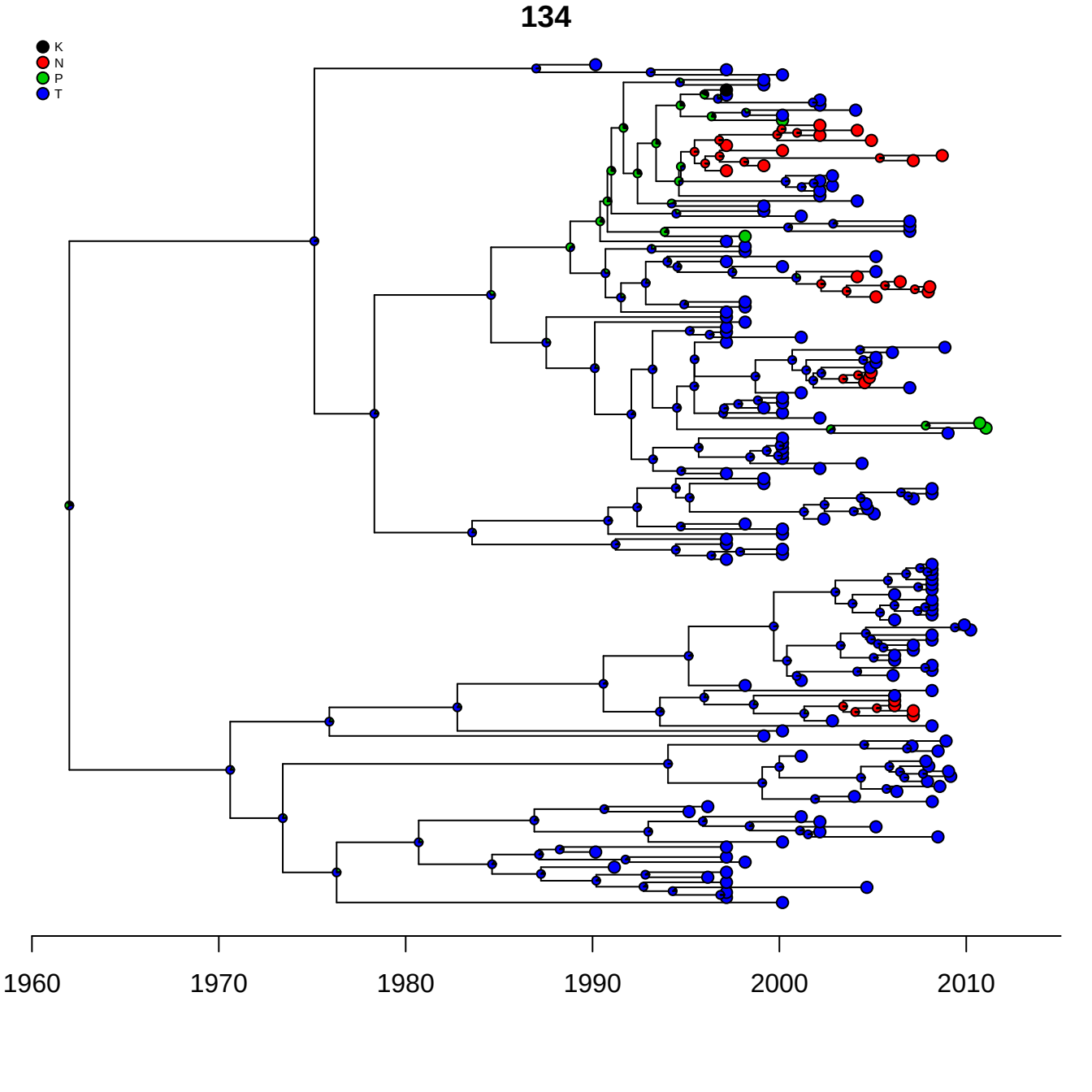

91
I
K
L
M
V
1960
1970
1980
1990
2000
2010

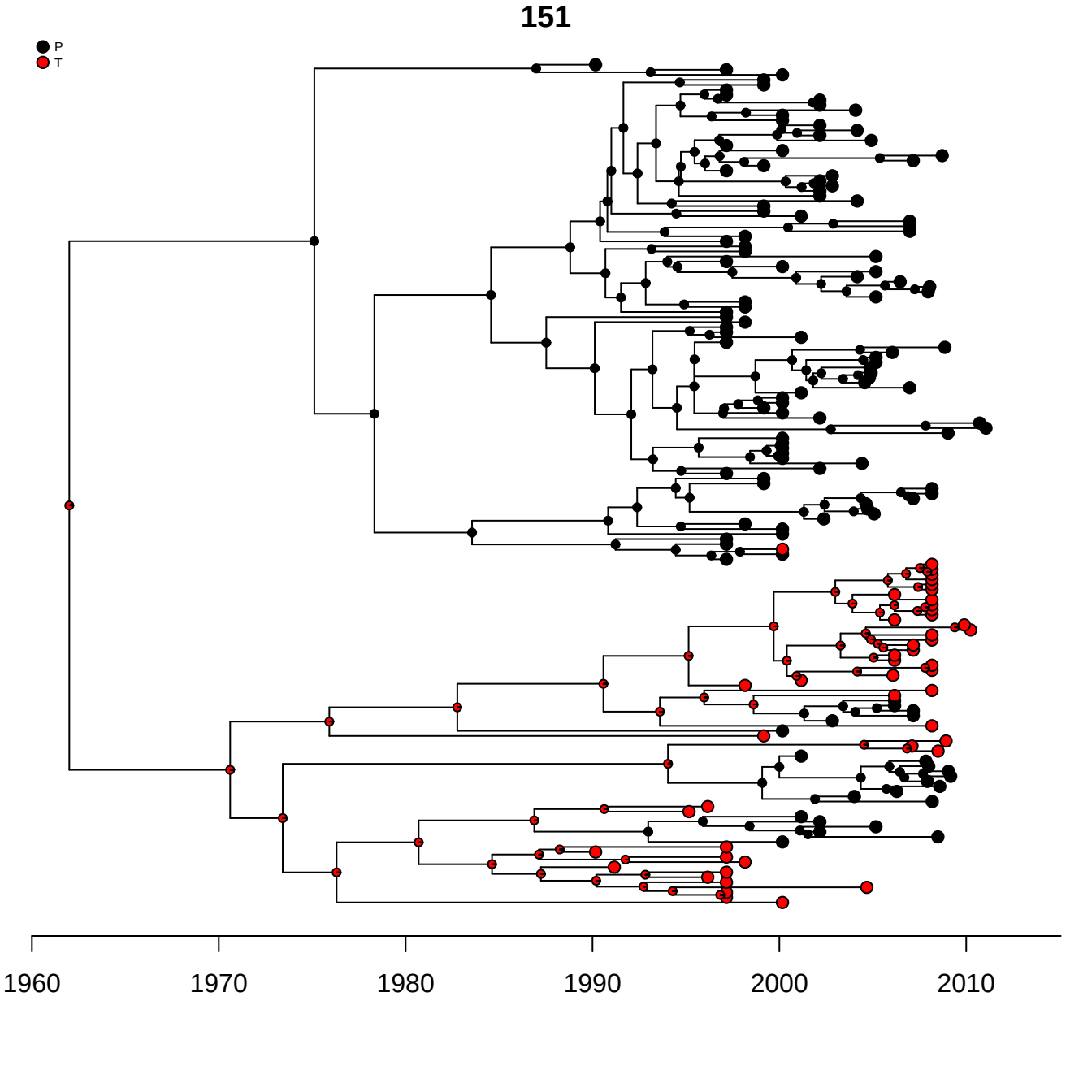

121
S
T
1960
1970
1980
1990
2000
2010

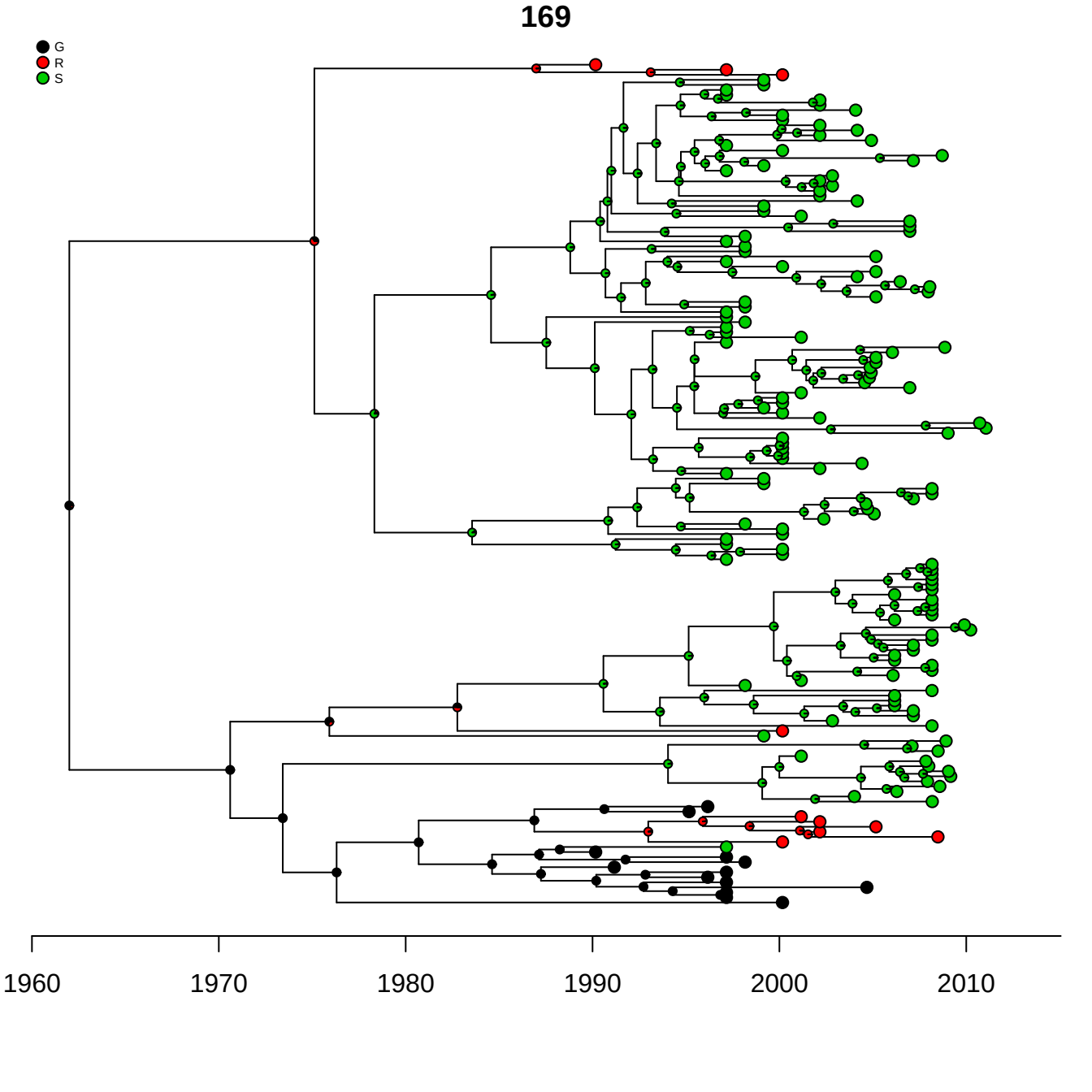

123
I
V
1960
1970
1980
1990
2000
2010

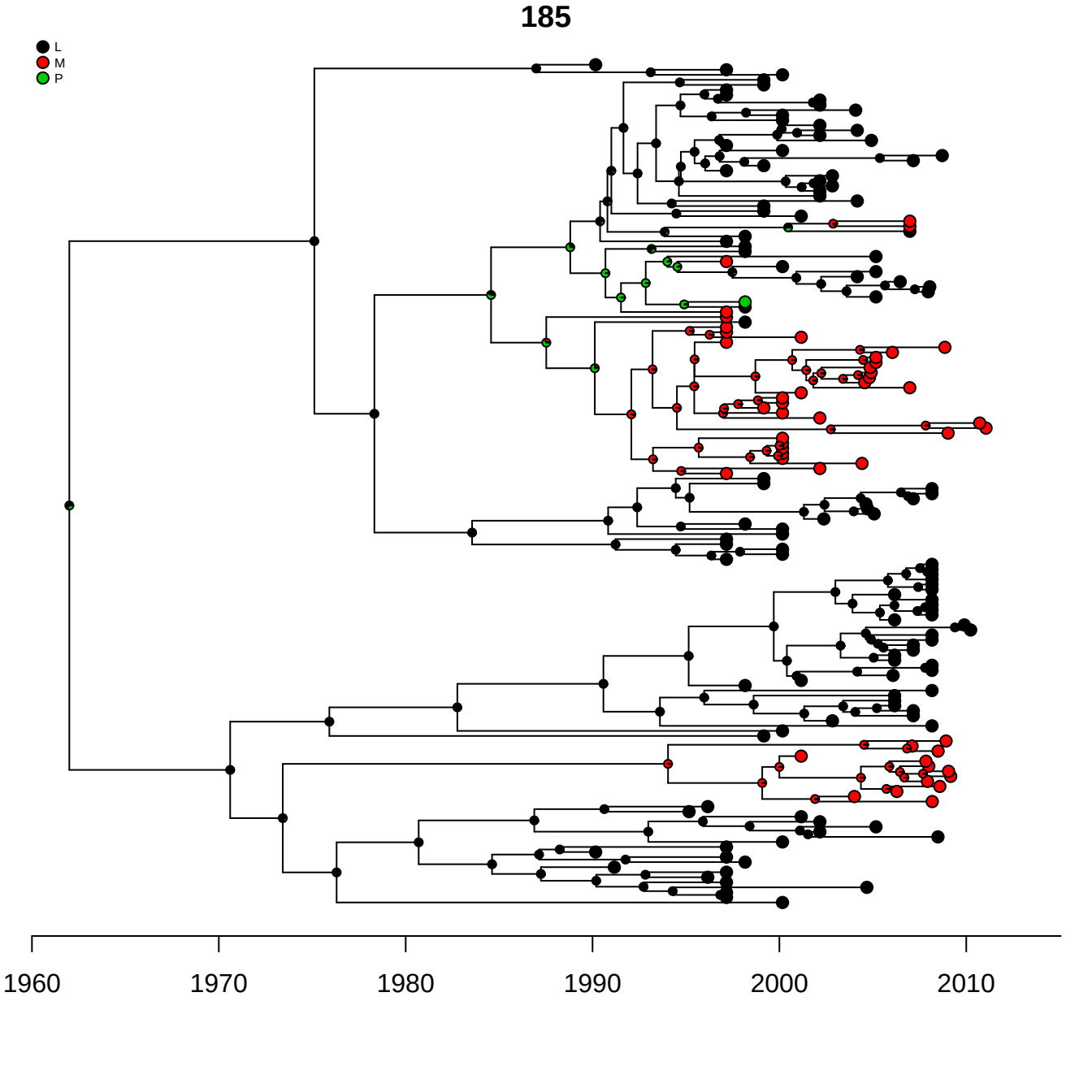

131
I
M
P
Q
S
T
1960
1970
1980
1990
2000
2010

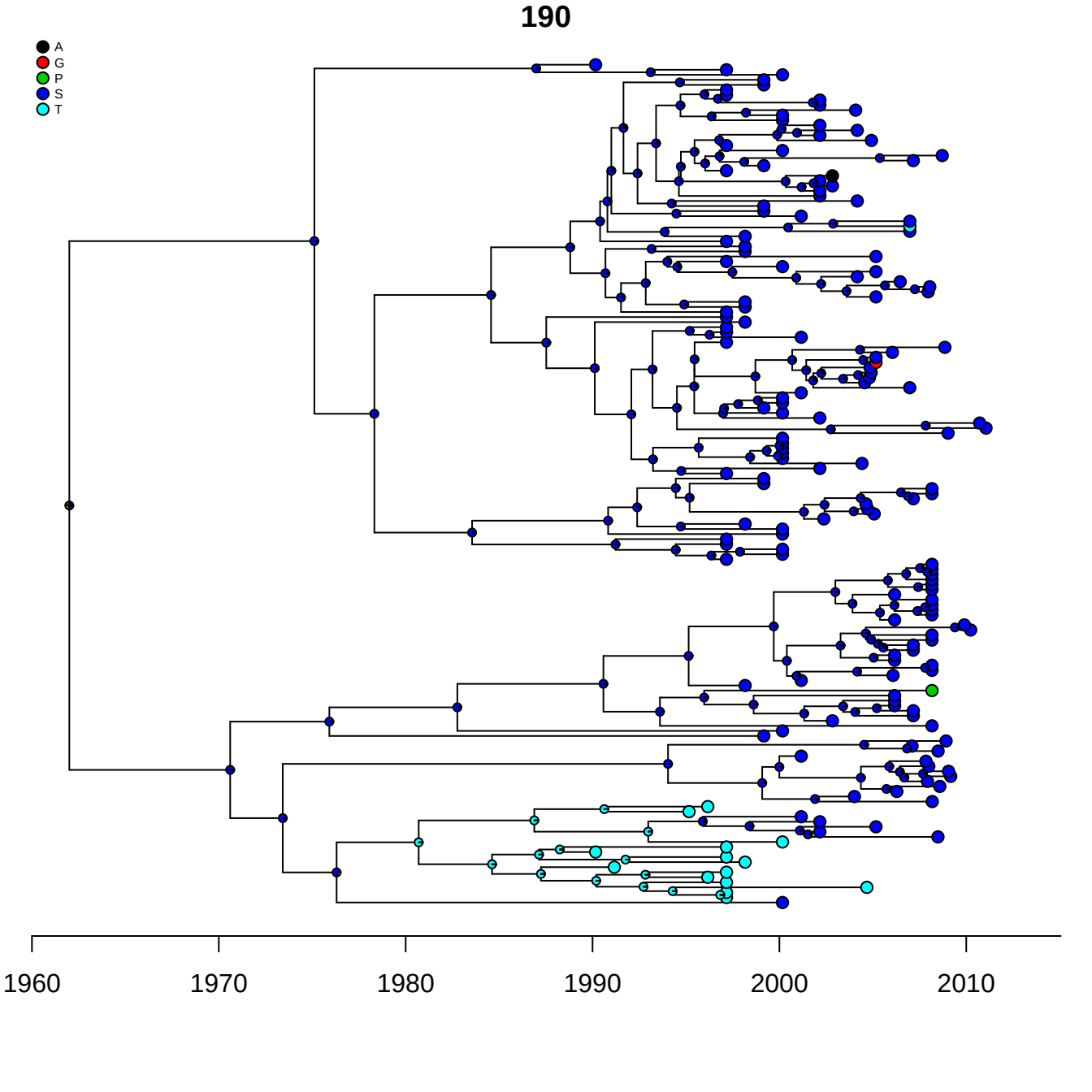

133
A
S
T
V
1960
1970
1980
1990
2000
2010

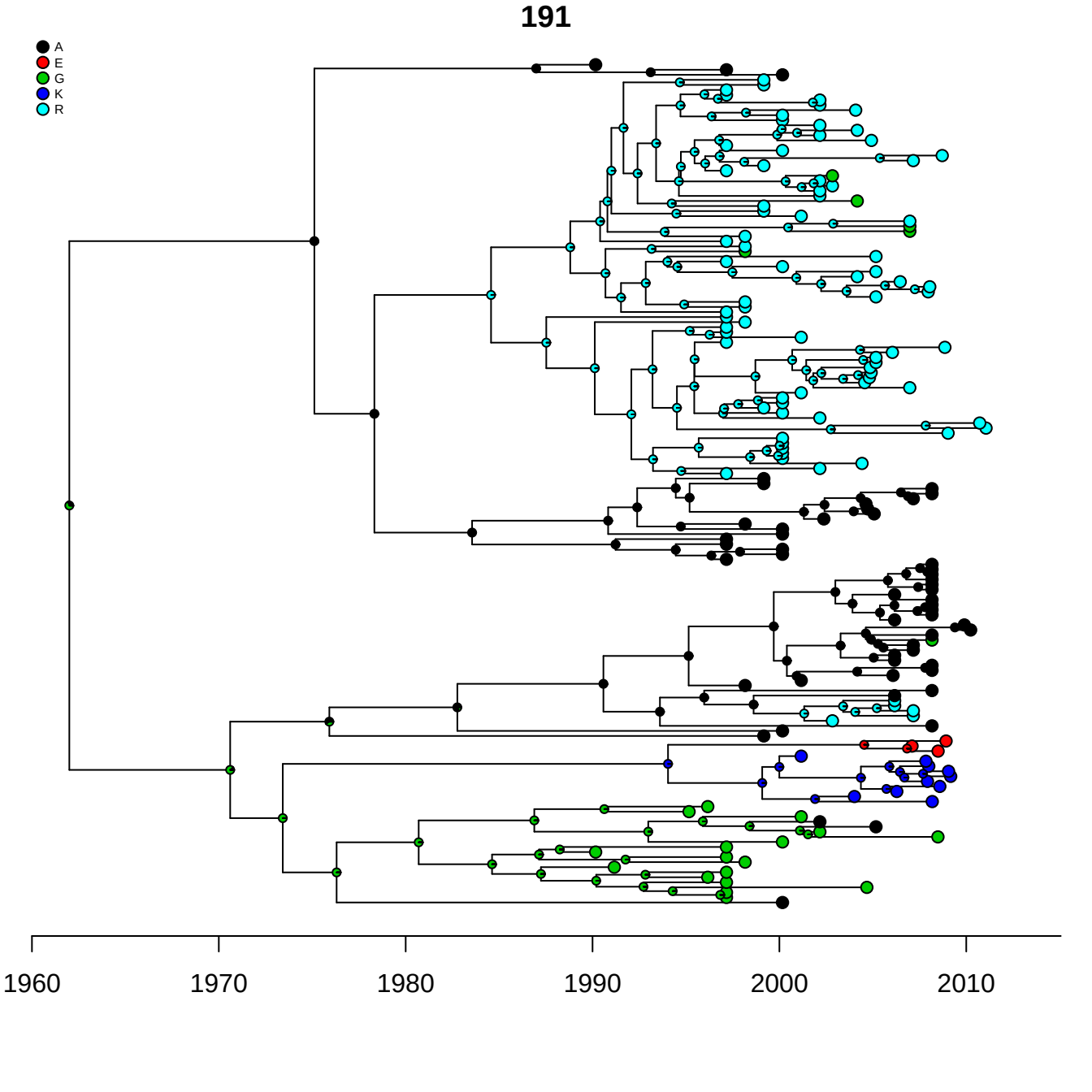

134
K
N
P
T
1960
1970
1980
1990
2000
2010

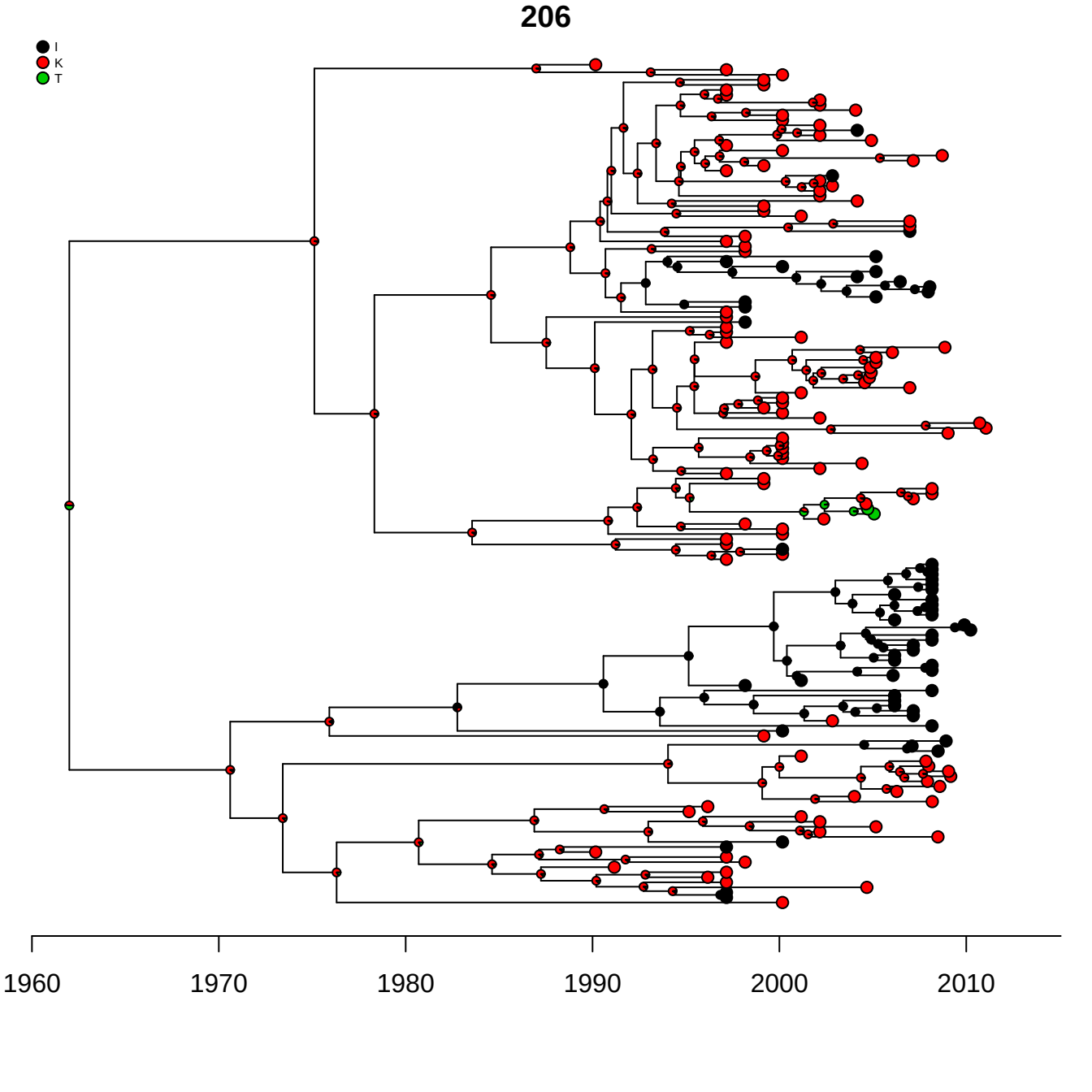

151
P
T
1960
1970
1980
1990
2000
2010

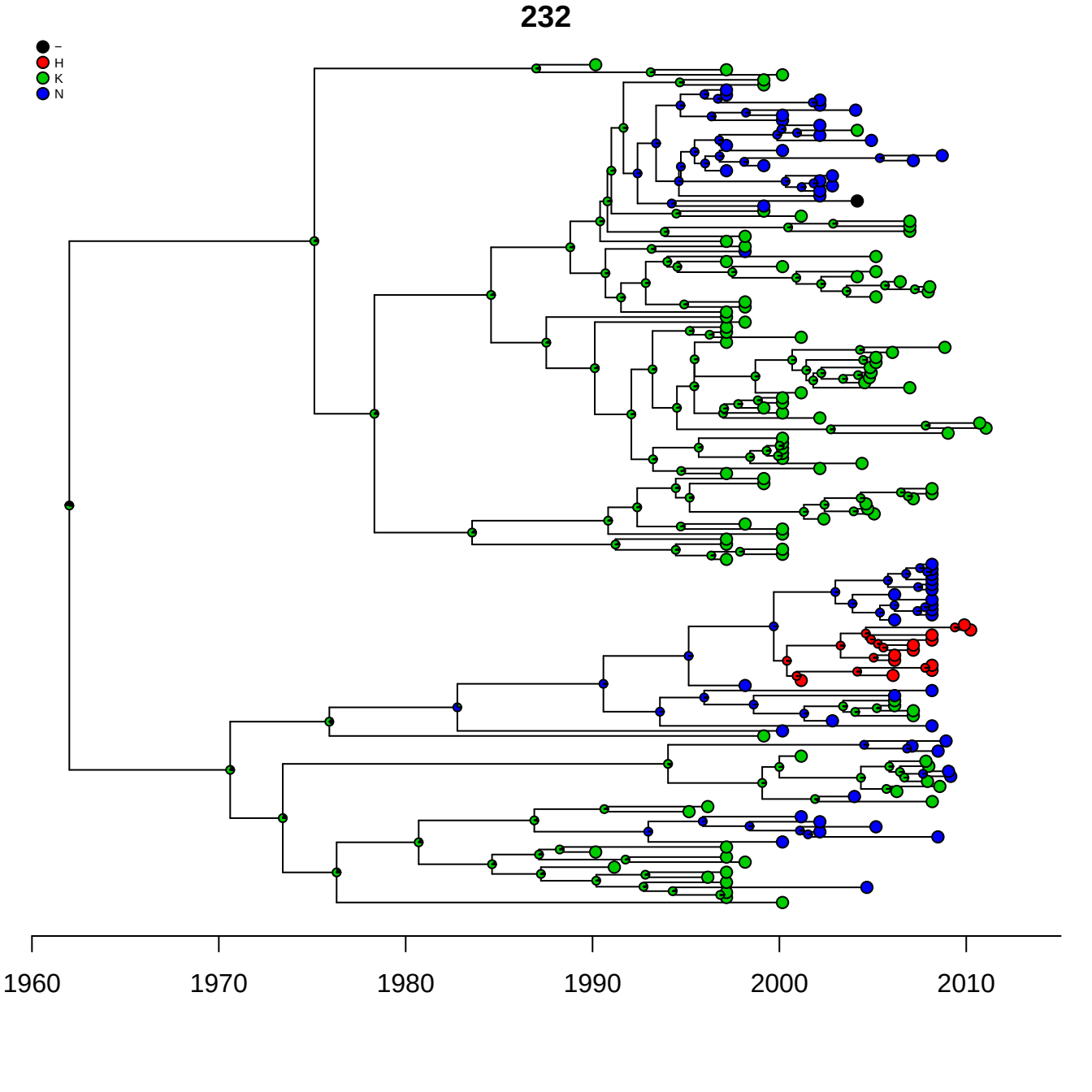

169
G
R
S
1960
1970
1980
1990
2000
2010

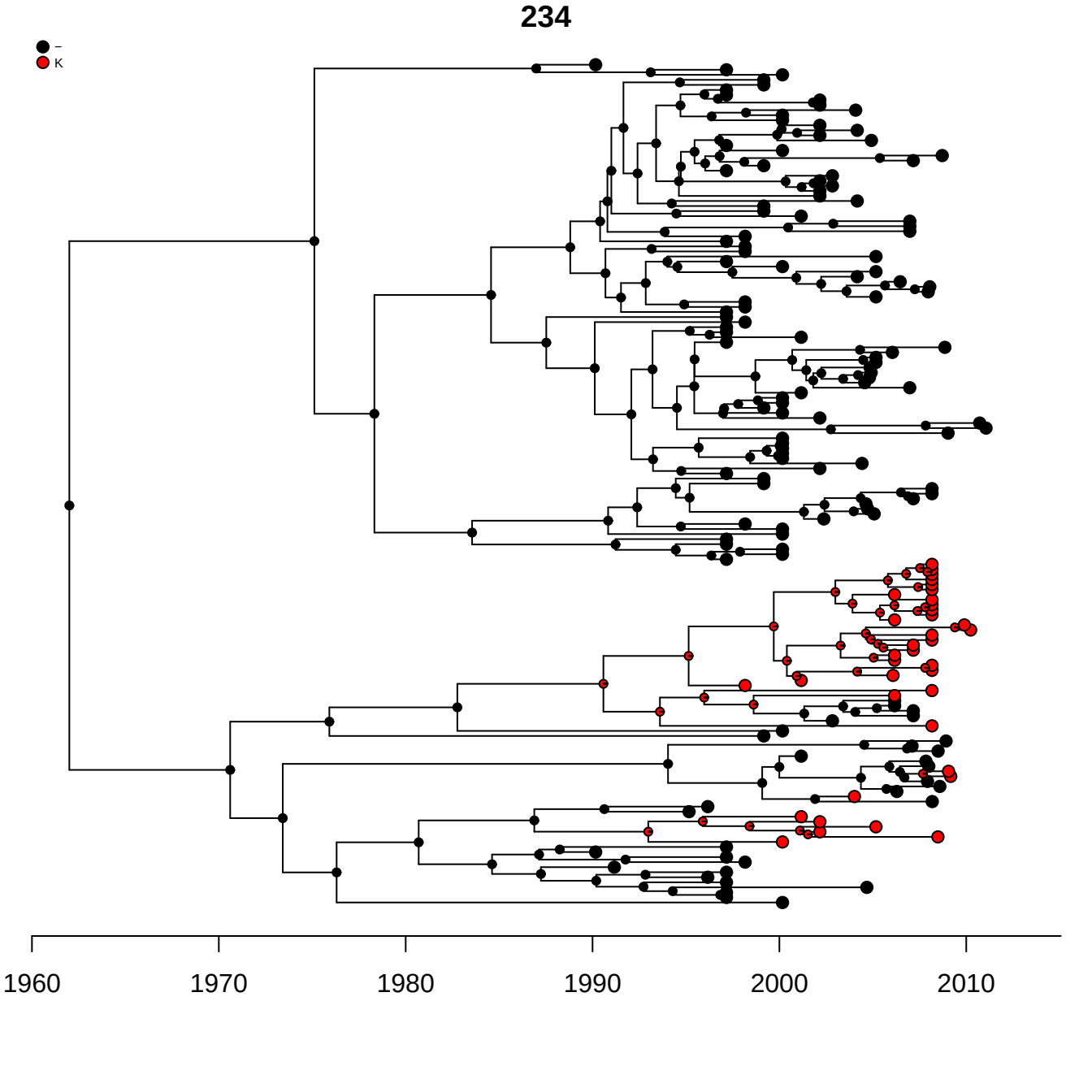

185
L
M
P
1960
1970
1980
1990
2000
2010


190
A
G
P
S
T
1960
1970
1980
1990
2000
2010


191
A
E
G
K
R
1960
1970
1980
1990
2000
2010

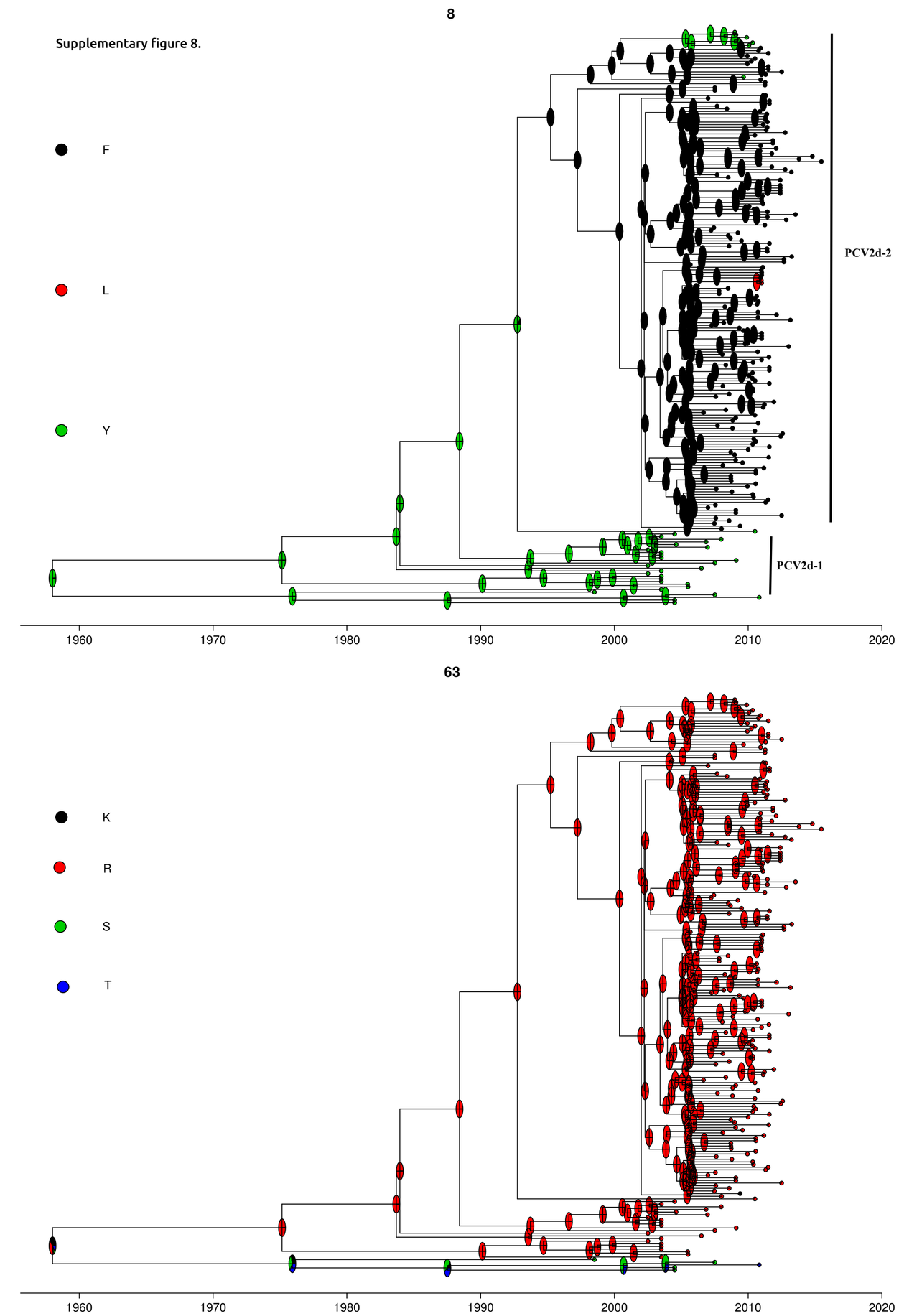

206
I
K
T
1960
1970
1980
1990
2000
2010

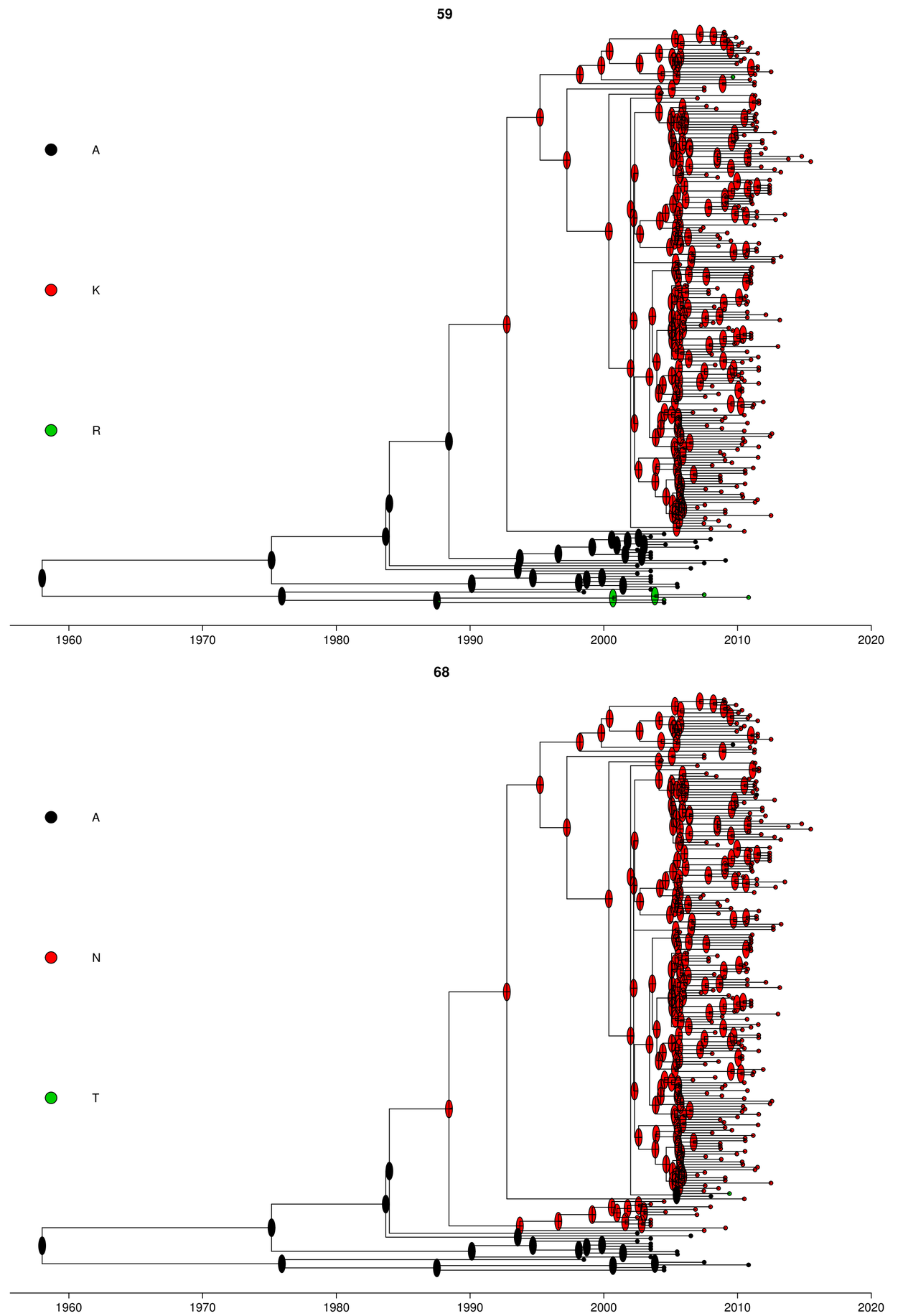

232
–
H
K
N
1960
1970
1980
1990
2000
2010

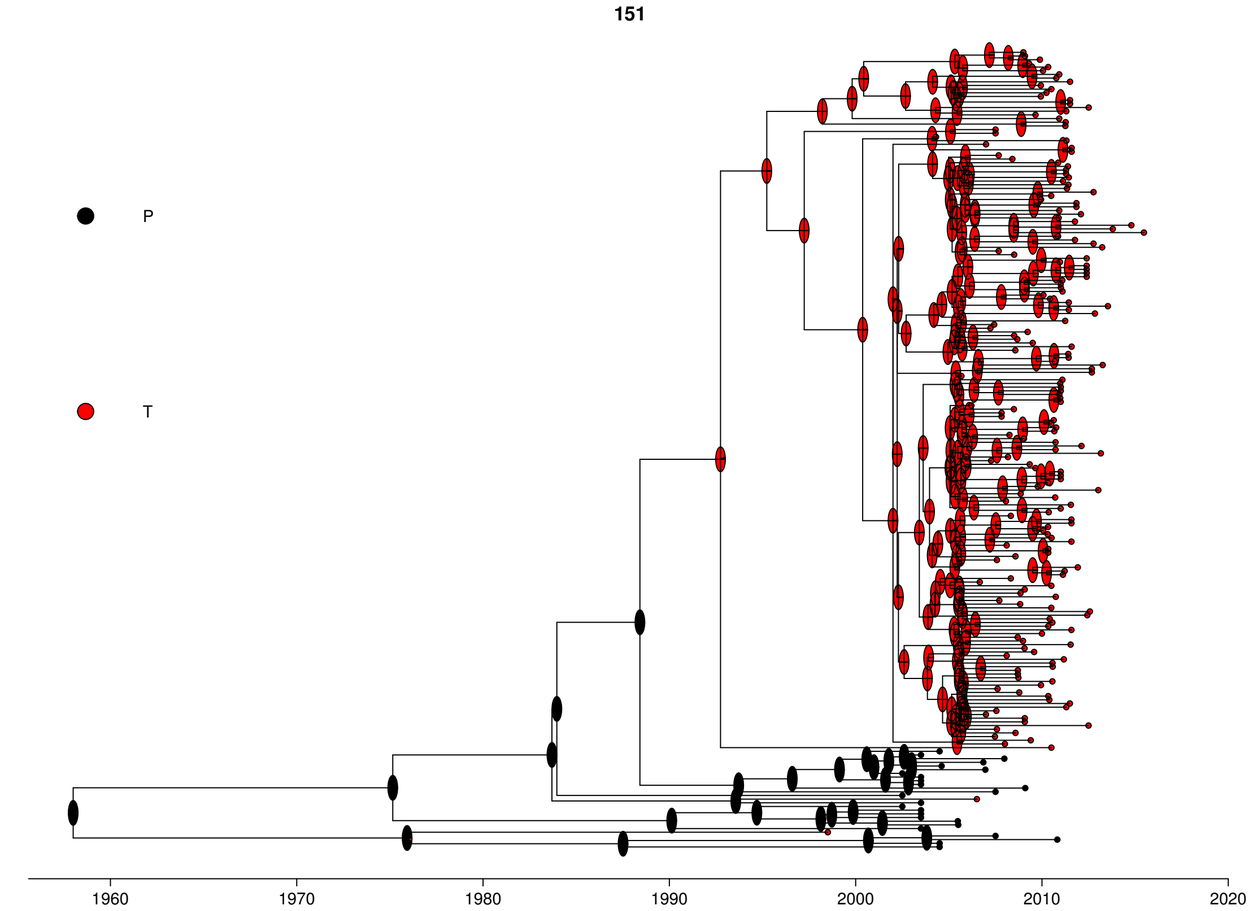

234
–
K
1960
1970
1980
1990
2000
2010

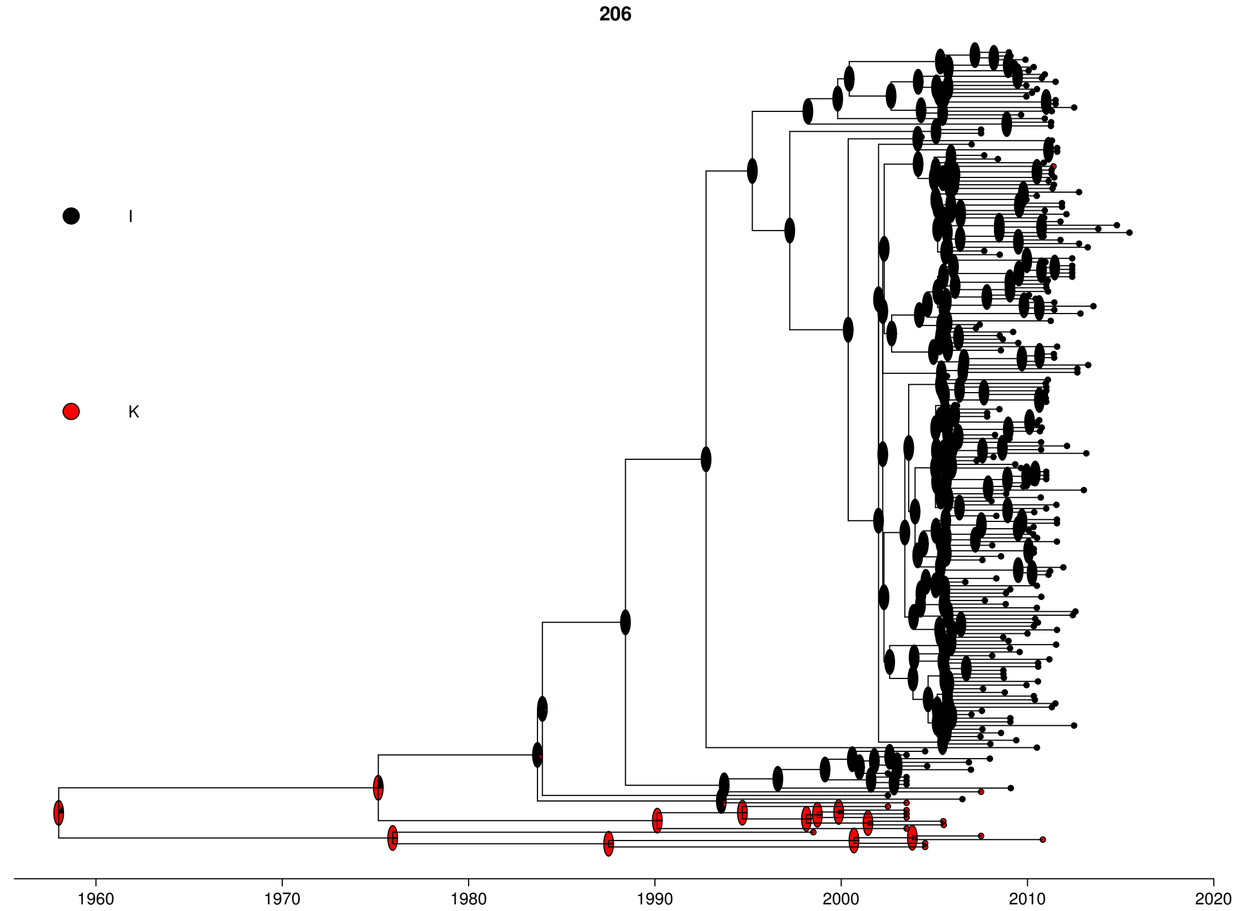
Supplementary figure 10. PCV2d maximum likelihood based reconstruction of amino- acid ancestral states over time. For clarity purpose, only a selection of significant variable positions are reported. The distinction between PCV2d clades 1 and 2, suggested by Xiao et al., 2015, is reported in the top left figure.

Supplementary figure 8.









	Asymmetric model		Symmetric model		LogMLE(Asym)-LogMLE(Sym)	
	LogMLE (PS)	LogMLE(SS)	LogMLE(PS)	LogMLE(SS)	LogBF(PS)	LogBF(SS)
PCV2a	-4537,42	-4536,55	-4541,91	-4542,42	**4,49**	**5,87**
PCV2b	-7157,73	-7162,17	-7156,19	-7160,63	-1,54	-1,54
PCV2d	-4585,64	-4586,24	-4581,35	-4581,99	-4,29	-4,25
CRF01	-6562,7	-6562,73	-6555,363	-6556,187	-7,337	-6,543
CRF02	-4970,6	-4970,87	-4936,1	-4936,43	-34,5	-34,44
CRF03	-5506,8	-5507,26	-5499,98	-5500,3	-6,82	-6,96

Supplementary Table 1. Logarithm of Marginal likelihood estimation (logMLE) calculated using path sampling (PS) and stepping-stone (SS) for symmetric and asymmetric spreading models. Logarithms of Bayesian Factor (LogBF) are also reported. A logBF > 3 (in bold) is traditionally considered as providing strong evidence in favor of the model with the highest ML.

			Chemical Composition		Polarity		Volume		Iso-electric point		Hydropathy	
	Codon	t	α1	p1	α2	p2	α3	p3	α4	p4	α5	p5
Rep	77	1.509	20.000	0.013	2.236	0.978	0.098	0.862	-20.000	0.000	11.205	0.016
	105	15.612	-3.807	0.203	5.370	0.070	4.138	0.001	5.461	0.198	-2.970	0.004
	106	6.612	1.290	1.000	19.976	0.023	0.892	1.000	-1.931	0.779	-0.007	0.946
	119	3.997	-1.125	1.000	0.341	1.000	20.000	0.049	-1.883	1.000	0.119	1.000
	139	0.519	8.908	0.073	-4.304	0.027	1.322	0.610	-1.327	0.625	0.300	0.794
	146	0.822	0.011	1.000	6.022	0.000	-1.054	0.855	0.619	1.000	-5.780	0.017
	150	2.039	-2.695	0.368	-0.028	1.000	-0.912	0.488	20.000	0.015	0.320	1.000
Rep'	77	1.212	20.000	0.051	3.185	1.000	-0.360	1.000	-7.267	0.008	0.943	1.000
	105	14.707	-1.890	0.746	2.687	0.427	3.606	0.001	1.970	0.415	-1.704	0.007
	106	6.558	0.830	1.000	19.967	0.014	0.685	1.000	-1.772	0.775	0.445	1.000
	164	2.443	0.478	1.000	-5.373	0.067	1.601	1.000	0.775	1.000	10.819	0.023
Cap	21	2.837	-4.080	0.036	1.551	1.000	1.376	0.820	2.157	0.123	0.801	0.769
	30	0.406	7.498	0.178	-11.692	0.013	-1.141	1.000	20.000	0.202	19.372	0.000
	46	4.595	-2.464	0.380	-3.349	0.241	3.871	0.155	2.881	0.001	6.514	0.028
	53	2.024	20.000	1.000	20.000	0.004	1.770	1.000	-1.211	1.000	-1.630	1.000
	59	2.005	0.487	0.652	1.753	0.389	-2.246	0.003	1.275	0.003	-0.013	1.000
	72	0.137	16.087	0.047	-6.836	0.144	-0.391	1.000	18.644	0.072	-3.415	0.145
	90	0.777	0.209	1.000	0.127	1.000	0.056	1.000	-7.525	0.003	6.737	0.616
	121	0.151	-2.961	0.286	-2.077	1.000	0.122	1.000	-1.588	1.000	20.000	0.000
	130	0.175	20.000	0.026	0.158	1.000	-2.041	0.373	-6.498	0.284	0.154	0.802
	131	1.121	-0.023	0.664	-1.233	0.688	-6.425	0.022	14.483	0.001	5.303	0.006
	133	0.560	-1.711	0.037	2.757	0.577	-0.491	1.000	-0.442	1.000	1.561	0.255
	134	1.355	-0.528	1.000	-3.242	1.000	18.474	0.007	0.565	1.000	-0.952	1.000
	151	1.268	0.675	1.000	0.992	1.000	-6.593	0.001	20.000	0.084	3.391	0.061
	169	1.384	1.095	0.013	-0.044	0.003	-1.591	0.000	-4.399	0.002	10.944	0.000
	190	0.948	-2.333	0.012	6.926	0.002	-1.151	0.362	1.022	0.494	0.942	0.526
	191	1.010	-1.365	0.049	0.756	1.000	0.018	1.000	0.017	1.000	0.551	1.000
	210	1.209	-3.039	1.000	-0.258	1.000	0.063	1.000	5.080	0.003	-0.725	1.000
	225	0.769	2.135	1.000	10.282	0.018	-1.019	0.645	0.413	0.764	-8.853	0.239
ORF3	29	0.484	-3.627	0.098	12.373	0.030	-0.101	0.772	2.149	1.000	-1.757	0.300
	36	4.471	-1.130	1.000	-0.766	1.000	-0.780	1.000	20.000	0.043	2.019	1.000
	41	0.544	-0.378	1.000	0.929	1.000	-1.531	0.010	0.009	1.000	-0.398	1.000
	47	0.075	-1.320	1.000	0.198	1.000	-1.904	0.030	-2.142	0.268	0.847	0.816
	57	1.834	-0.202	0.731	0.544	1.000	-0.253	0.332	-7.321	0.076	17.259	0.037
	89	1.716	-0.773	1.000	-3.220	0.037	0.596	0.805	-0.249	1.000	4.094	0.031
	103	6.065	-1.769	0.048	-1.766	0.000	1.258	0.074	-0.286	0.078	3.976	0.000

Supplementary Table 2. Summary of PRIME analysis. α and p-value are reported for each physico-chemical properties. α represent the site specific importance of the property. When α=0, selection is neutral with respect to that property. The higher is α (green)the higher is the purifying selection even on conservative changes. When α is negative(red), non-conservative changes are favored over conservative ones with regard to that proprieties.