Evolutionary history and phylogeography of *Sugarcane mosaic virus*

Ricardo I. Alcala-Briseno[1,2,*],

1 Genetic Engineering Department, Plant Biotechnology, Research Center and Advanced Studies (CINVESTAV), Irapuato, Guanajuato, Mexico.
2 Department of Plant Pathology, University of Florida, Gainesville, Florida, USA.

¤Current Address: Department of Plant Pathology, University of Florida, Gainesville, Florida, USA.
* ralcala@ufl.edu

Abstract

Sugarcane mosaic virus (SCMV) is one of the oldest reported virus (1919) and its distribution is on the tropical and subtropical areas of the planet. The SCMV is the causal agent of maize dwarf and sugarcane mosaic disease, and is able to infect sugarcane, maize and sorghum, among other grasses. Phylogenetic analysis and recombination patters have been described previously but there is no information about the geographic origin and phylogeography of the viral population. To determine the origin of SCMV we use TMRCA relaxed molecular clock implemented in BEAST, estimating that SCMV emerged about 1200 (+400) years ago. The phylogeographic analysis shows that SCMV is most likely to have a geographic origin ins the Southeast Asia.

Introduction

Sugarcane mosaic virus (SCMV) is a member of the genus *Potyvirus* in the *Potyviridae* family of plant viruses, the genomic structure in a single stranded (+) RNA, with a length of 9.6 kb and its encapsidated with aproximately 2,000 monomers of the coat protein (CP), forming a flexuous filaments of 750 nm in length and 10 nm in wide (Adams et al. 2005). The *Potyvirus* genus lacks of a regular cap (5'm7G), instead has at the 5UTR region a covalently linked viral protein (NIa-VPg) and a poly(A) tract at the 3'UTR. The genome is a monocistron enconding a large polyprotein and a small peptide (PIPO), is processed by ribosomal frameshift (+2). The polyprotein is processed by three proteases, two autoproteolytic proteins, protein 1 (P1), helper component and proteinase (HC-Pro) and the nuclear inclusion a (NIa-Pro) which cleavage the polyprotein releasing eight out of 10 proteins encoded by potiviruses (Urcuqui-Inchima et al,. 2001, Adams et al., 2005).

Sugarcane mosaic disease was first described in 1919 in Puerto Rico (Brandes 1919), as mosaics or mottles at the base of emerging leaves, which coalesce into narrow, continuous or broken streaks, plants also show symptoms of stunting. The host range of SCMV was described in the Poace family, first described in sugarcane the causal agent was able to infect maize, sorghum, rice, millet, crabgrass, foxtail and *Panicum*, recently has been described in other monocotyledons plants as *Setaria* sp. *Stenotaphrum* sp., *Eleusine* sp., *Maranta* sp. and *Musa textilis* ((Brandes, 1919; Williams and Alexander,

1965; Adams and Antoniw 2006; Ha et al., 2008; Harmon et al., 2015). SCMV is vectored by aphids in a non-persisten manner; the mainly known vector is *Myzus persicae* but *M. euphorbiae*, *Schizaphis graminum* and *Rhopalosiphum padi* can also transmit efficiently SCMV (Louie and Knoke, 1975).

The distribution of SCMV is ubiquitous in all subtropical and tropical areas where gramineous crops are cultivated, there are reports of SCMV over 25 countries. Brandes described the first anomalies in sugarcane crops from the Java island around 1882 and the disease was spread in the Americas by 1919 in Puerto Rico and reported in Argentina, Brazil, Peru and USA (Brandes, 1919; Abbot, 1929; Koike and Gillaspie, 1989). The sub-Saharan Africa reported the presence of SCMV by 1930's (Conjé, 2001). An outbreak in maize was reported in USA by 1963, reporting for the first time the maize dwarf virus disease, a cluster of SCMV isolates infecting maize (Louie and Knoke). Evidence of SCMV infections were later presented during the 1960s in sugarcane fields located in India, Thailand and Taiwan (Abbot and Stokes, 1966; Sharma et al., 2002), in maize in China (Chen et al., 2002), reports in maize, sorghum and sugarcane in Kenya and Australia (Teakle and Grylls, 1973; Louie and Darrah, 1980), Morocco (Fischer and Lockhart, 1974), Italy (Tosic et al., 1977), Cameroon, Pakistan and Iran (Gillaspie et al., 1978), Egypt, Japan and Colombia (Gillaspie and Mock, 1979), tropical Africa (Thottappilly et al., 1993), Germany (Oertel et al., 1997), Mexico (Delgadillo, 1987; Espejel et al., 2006) and Spain (Achon et al., 2007), in maize in Ethiopia and Rwanda (xxxx, xxxxx).Previous studies have been reported the phylogenetic relationship of SCMV, addressing two main characteristics: host and region. The group of isolates of sugarcane, and the group of isolates of maize which are divided in two continentals regions, the European and the Asian isolates.

However, besides all reports around the world, there is no studies reporting the origin and phylogeography of SCMV. In order to achieve the historical perspective and distribution pattern of SCMV, here we show the time of the most recent common ancestor with a relaxed molecular clock, inferring the center of origin of SCMV and its phylogeography spread and distribution using bayesian estimation and coalescent theory.

Materials and Methods

Data collection and sampling information.

A total 1030 sequences of Sugarcane mosaic virus were collected from the NCBI (April 2016), of which 138 sequences were full coat protein (CP) sequences and 37 whole genomes (WG). 798 partial CP and 26 non-CP sequences were not used in this study. Additionally, we include 17 sequences of SCMV isolated from maize, collected from Veracruz, Mexico in 2011, and a whole genome of SCMV isolate CAM6-1 (Achon 2007, kindly donated from Dr. Jean-Claude Girard and Dr. Michel Petershchmitt), fully sequenced by RACE-PCR (Clonetech, Mountain View, CA) and assembled using Geneious v9.1.3; samples included in this analysis were sequenced by the Sanger dideoxy method. The CP dataset consist in 151 isolates collected from different host sources as sugarcane (59 seq), maize (70 seqs), Abaca (1 seq), Maranta (1 seq), Musa (1 seq), Setaria (1 seq) and Stenotaphrum (1 seq) in 18 countries from five continents within tropical and subtropical areas since 1967 up to date. It was possible to collect metadata from each isolates, including collection date, host and location, some information was missing for some isolates (Suppl. Table 1). The 151 CP and 38 genome were codon aligned using MEGA 7 (Tamura et al., 2013) with muscle v3.8.31 w and manually edited with Geneious 9.1.3.

Maximum likelihood and temporal signal.

To determine the phylogenetic relationship of SCMV, the CP and WG were analzed with jModel test 2 (Darriba et al., 2012) which search for the best likelihood in relation to the genetic sequence. The best models of substitution for the CP and WG was TIM2 and GTR, respectevely, plus invariable sites and gamma parameter used in the maximum likelihood (ML) phylogenetic reconstruction using PhyML 3.0 (Guidon, et al., 2010). A statistic bootstraping test was done under 100 replication (Supp. Fig 1). The ML tree branches were collapsed using a perl script under 50% bootstrap support. The ML tree of the CP of SCMV was analyzed in Path-O-Gen (tree.bio.ed.ac.uk/software/pathogen) to evaluate the degree of temporal signal using the root-to-tip analysis, searching for a strong correlation coefficient ¿0.5 (Harkins, G. personal communication).

Recombination of SCMV.

In order to test for recombination, we analyze both datasets, the genomic sequence and the 151 CP of SCMV. The 38 genomes of SCMV were edited manually due to conflicting areas (CP mainly) difficult to align or with many gaps, doing these we reduce the noise for the recombination analysis. The RDP4 implementing RDP, GenConv, BootScan, MaxChi, SiScan and 3seq for linear sequence with default settings. The CP was analyzed only with RDP4 (Martin et al., 2015).

Bayesian evolutionary inference

We reconstructed time-calibrated phylogenetic history of SCMV using Bayesian estimation implemented in the package BEAST v2.4 (Bouckaert et al., 2014). We applied the Hasegawa-Kishino-Yano 85 (HKY) substitution model, also we partitioned the codon positions (+1 +2 +3) (Shapiro et al., 2006). We applied an uncorrelated relaxed molecular clock to adapt the rate variation among different lineages and to estimate for coalescent priors we use coalescent constant population lognormal distribution and coalescent bayesian skyline models (Drummond et al., 2005, Drummond et al. 2006). Three independent runs for 500 million generations for the MCMC, sampling every 1000 or 10,000, discarding the 10% of the chain burn-in. Beast results were resample using LogCombiner and Tarcer v1.6 (tree.bio.ed.ac.uk/software/tracer) was used to analyze the results. TreeAnnotator was used to generate the maximum clade credibility tree (MCC).

Discrete phylogeography of SCMV

In order to identify the ancestral locations and distribution of SCMV, I use a bayesian estimation discrete phylogeography analysis to determine the time distribution using BEASTv1.8.3 (Drummond et al., 2012). We applied the HKY 85 substitution model plus gamma and using a discrete trait symmetric substitution model with bayesian stochastic search variable selection (BSSVS). The discrete phylogeographic analysis use a continuous-time Markov chain (CTMC). The clock was set lognormal relaxed molecular clock model (rate 1.0) for taxa and strict molecular clock (rate 1.0) for location. The tree prior distribution for the discrete phylogeography was the coalescent model GMRF bayesian skyride, a flexible demographic model reconstructing the states at all ancestors. The MCMC chain was set at 10,000 and 100,000, both chains did not converge in tracer analysis, longer chains would be required to run for longer to have certainty in the phylogeographic analysis.

Results

Time divergence and diversity of SCMV

Sequence diversity is essential to calibrate a molecular clock, our exploratory analysis of root-to-tip divergence in order to detect temporal signal in these SCMV data set. The phylogenetic reconstruction of 151 CP isolates of SCMV (Fig. 1) from 18 countries all around the world was used to estimate the root-to-tip divergence following a standard protocol. Using Path-O-Gen, the linear regression on this data set was R^2=0.018 (Fig. 2) and a correlation coefficient of 0.136 using the sequences diversity and tip dates of the ML tree of SCMV. These results are showing a strong temporal signal for SCMV, meaning that SCMV isolates have enough temporal variation supporting the hypothesis of application of a molecular clock test.

After finding a strong evidence of temporal, we implemented the estimation rate using a strict and relaxed molecular clock models analyzed with two priors; the coalescent constant population and the bayesian skyline (citas, Drummond et al., 2005). The estimation rate was estimated for coalescent constant population $4.364e^{-4}$ (95% HPD: $1.9301e^{-4}$ - $6.9268e^{-4}$) and $6.0669e^{-4}$ (95% HPD: $2.7772e^{-4}$ - $9.7585e^{-4}$) substitutions per site per year for strict and relaxed molecular clock, respectively. In the other hand, the bayesian skyline estimations were $1.605e^{-4}$ (95% HPD: $9.0096e^{-5}$ - $2.3505e^{-4}$) and $8.509e^{-4}$ (95% HPD: $4.8885e^{-5}$ -1 $.2375e^{-3}$) (Table 1). The coefficient of variation for the strict and relax clock for the coalescent constant population and bayesian skyline was relatively similar but the evolutionary rate was quite divergent for the four methods compared. The time to the MRCA was positioned in about 2000 and 600 years ago, it depends of which of the four different methods was used, but for the bayesian skyline the relaxed molecular clock gave the most recent dates, similar to what it was previously reported. There is a very good demarcation by host, sugarcane and maize in three different clades, a group which I called the ancestral sugarcane lineage, which is at the bottom of the tree with isolates from China and Vietnam, which are close to the first site of description, the Java Island in 1889 (Fig. 3). A second clade of most recent common ancestor of a sugarcane lineage composed exclusively of sugarcane isolates from USA,China, Cameroon, Argentina, Pakistan, South Africa, India, and Iran. A third lineage of maize, with mainly maize isolates but also includes St. Augustine grass (Stenoraphrum secundatum), Musa, Setaria, Maranta, Abaca, arrowroot and sugarcane isolates. Five small clades are formed in the maize lineage, the oldest clade with two isolates from USA, one from Ohio 1965 from maize and an isolate of Florida in 2013 of St. Augustinegrass were at the bottom of maize lineage and this clade is formed with isolates from Mexico, Rwanda, Ethippia and Brazil. The following clade has isolates from Southeast Asia and isolates from maize, sugarcane and setaria. A small clade was formed from sugarcane, maize and maranta isolates from China and Vietnam and the two most recent groups, a clade of maize isolates from Mexico, Germany and China as ancestral isolates (KR611108 and KR611110) in this group. The last clade (upper part of the tree) which includes isolates of China, Germany, Spain and Argentina, most virus were isolated from maize and only one from sugarcane, in this clade a China isolate (KR611107) was rooting the clade.

Discrete phylogeography.

Assigning the location to the SCMV isolates I was able to assign a phylogeographic distribution in a MCC tree. There is no obvious trend in geographic isolation of SCMV, but there is a common ancestor, it could have been more than one introduction events that stablished different lineage populations in certain regions (see below). A MCC tree colored branched with locations (Fig. 4).The estimation of the discrete ancestral

Table 1. Comparison of coalescent model and molecular clock on the TMRCA estimation

Model	clock	Time for the MRCA[a] year	Evolutionary rate (μ) (substitution/site/year)	Coefficient of variation for (μ)
CCP[b]	strict	1150	$4.364e^{-4}$	0.9868
		(211-1778)	($1.9301e^{-4}$ - $6.9268e^{-4}$)	(0.8941-1.0827)
	relaxed	1074	$6.0669e^{-4}$	0.9475
		(99-1746)	($2.7772e^{-4}$ - $9.7585e^{-4}$)	(0.8546-1.0417)
BSL[c]	strict	446	$1.605e^{-4}$	N. A.[d]
		(390 b.c.-1102)	($9.0096e^{-5}$ - $2.3505e^{-4}$)	
	relaxed	1503	$8.509e^{-4}$	0.9513
		(1042-1824)	($4.8885e^{-5}$ -1 $.2375e^{-3}$)	(0.8624-1.0461)

Values between bracket represent the 95% HPD intervals.
[a]MRCA, most recent common ancestor.
[b]CCP, coalescent constant population.
[c]BSL, bayesian skyline.
[d]N.A. No available data.

location states was applied using a discrete trait substitution models with a BSSVS approximation. The phylogeographic analysis shows that most isolates have a common geographic origin from Southeast Asia, these clade is conformed with isolates from China and Vietnam sampled in 2004 and 2007 (FM9978** and D!92542**). There is a migration or migrations events of SCMV to the Americas, dating it most likely to 80 years ago; the two oldest isolates are from USA, from Ohio in 1965 (JX1883885) and possibly Louisiana 1970's (Southeastern US, SMU5735*), both isolates are 75% similar to each other (the virus demarcation species for the genera *Potyvirus* is 76% nt identity). Ancestral isolate of SCMV from the Southeastern US isolates was present in South America by the 60's and followed by eventual migrations to India, Iran, Cameroon, SouthAfrica, Pakistan, and Australia in the last 50 years. All these sugarcane isolates are 94% similar at nt level. Another introduction or early diversification of the isolate from the midwest of US, similar (by 88%) to the Florida isolate (KR261458) collected in 2013 from St. Augustingrass followed a series of dispersal events to Jalisco, Mexico (the Pacific area) and the Eastern African (Ethiopia and Rwanda) around 25 years ago.This clade is also distantly related (80 years ago) to the Brazilian isolates (sharing 91% nt similarity). About 80 years ago, in the other part of the world, another diversification expansion was spreading through Southeast Asia, with an expansion in host range that was commonly infected for SCMV, usually SCMV is infecting members of the Poace family, like sugarcane, maize or Setaria, but in this group another plant family was infected by SCMV, the family Musaceae. This previously described group possibly have a migration event in the last 50 years, isolates of SCMV diversified in Europe from isolates of Chinese origin (KR6111**) forming two distinct lineages, a mixture of isolates from Veracruz (Eastern) Mexico and German isolates and another with isolates from Germany, Spain and Argentina, both clades were mostly host specific to maize. The dispersal distribution agrees with the literature and genetic information for SCMV.

Spatiotemporal recombination.

In order to resemble the recombination history of SCMV I used the spatiotemporal reconstruction and genomic patterns of the recombination break points of the 39 isolates with genome sequence available of SCMV. The sequences were aligned (see methods) and manually edited, large gaps were eliminated in order to decrease the signal of false positive recombination sites. The most variable zone in the potyviruses is

in the coat protein, this region known as the hypervariable region located in the amino-terminal. An alignment of 39 sequences with 9,135 nucleotides was analyzed for possible recombination sites using the RDP, Gen Conv, Boot Scan, Max Chi, Siscan and 3 Seq analysis with the RDP4 package (Martin et al., xxxX) with standard parameters to search for recombination in lineal molecules of RNA. RDP showed a total of 41 recombination sites from 29 different isolates validated with at least one of the methods having a statistical P-value of > P= 0.5. The recombination sites were accepted only if different methods showed individually higher P values (see table 2). The recombination break point for isolates and countries is summarized in table 3. The evidence of recombination patterns of SCMV showed a geographic ancestry, sharing space and time to be able of recombine. There was two very strong signals of recombination in two isolates, one from Ethiopia (KP860936) possibly recombining with an isolate from China Shanxi-7 (SX-7, KP611111) and another from Rwanda (KF744391), and in the other hand, the Chinese isolate of Shangdong (SD, AY149118) with strong evidence of recombination with other two isolates from China, SX-7 and Hangzhou (HZ2, AJ297628). Both recombination events were in the position 5,000 nucleotides, around the cistron 6k2. The Iran isolate NRA (KT895080) has signal of recombination in a diffrerent position, from position 1250 to 8000 with the Ohio isolate as a major parent and as a minor parent the isolate from Cameroon (CAM6-1, acc. no. in process). The isolate HZ2 shows a second recombination break point from the position 200 to 2,000 nucleotides of the P1 cistron, contributing as a donor isolates were Beijing (Bj, AY042184) and the Shanxii isolate (SX, AY569692). A fifth recombination event was identified in the Mexican isolates, Jalisco (JAL-1, GU474635) and the VER-1 isolate (EU091075) as a major parent and as a minor parent the Ohio isolate, including the coat protein cistron of SCMV. Other recombination events have less statistical support and not very conclusive recombination break points using only RDP4(see table 3).

Table 2. Whole genome recombination analysis of *Sugarcane mosaic virus* statistical supports using RDP4

No	Rec	Major Par	Minor Par	RDP	GenConv	BootScan	MaxChi	Chimaera	SiScan	3seq
1	KP860936	KP611111	KF744391	4.1E-63	5.8E-95	1.3E-55	8.6E-30	7.0E-35	1.8E-32	3.4E-111
2	AY149118	KR611110	AJ297628	3.8E-30	1.7E-76	7.6E-35	2.6E-35	7.2E-23	2.3E-37	2.3E-47
3	KT895080	JX188385	pending	2.0E-15	-	4.5E-7	9.4E-19	7.2E-12	4.9E-31	6.2E-57
4	AJ297628	AY042184	AY569692	3.1E-16	2.8E-32	1.0E-2	1.4E-8	2.9E-9	5.8E-11	-
5	GU474635	EU091075	JX188385	5.1E-35	1.0E-45	3.6E-32	3.5E-10	4.2E-11	1.3E-10	1.9E-2
6	KR611111	AY042184	KR611107	1.2E-3	3.7E-25	1.3E-3	3.5E-6	1.7E-6	2.1E-4	3.7E-5
7	AY569692	Unknown	AY042184	5.2E-14	7.4E-40	-	3.8E-14	-	6.0e-19	-
8	KT895080	KF744390	AJ278405	3.8E-19	7.4E-32	1.5E-18	1.3E-14	4.8E-16	8.9E-18	1.6E-27
9	AY569692	Unknown	AY042184	1.1E-8	3.2E-20	-	1.2E-9	2.9E-5	5.9E-27	-
10	AY042184	AJ297628	AF494510	2.6E-8	2.9E-30	-	1.5E-6	6.9E-5	4.6E-5	1.6E-11
11	AY569692	KR611111	Unknown	1.4E-3	8.2E-12	-	9.4E-9	1.2E-2	8.4E-15	1.6E-24
12	JX237863	AJ278405	Unknown	1.6E-9	8.6E-17	4.4E-10	1.1E-5	2.6E-7	4.2E-9	-

Discussion

Sugarcane mosaic disease was first reported in the American continent by Brandes in 1919 and by the 1880's was known in Southeast Asia (Brandes, 1919). The time to MRCA estimated with bayesian estimation, position the origin of SCMV as old as 390 B.C. or the most recent origin at 1800's, undoubtedly there is a needed to calibrate better the molecular clock in order to reduce the HPD for the different methods used, coalescent constant population and bayesian skyline, but the skyline have a narrow HPD and is closer to the reported values. In order to improve accuracy of the molecular

Table 3. Recombination break points of *Sugarcane mosaic virus* using RDP4

No	Recombinant	Country	Major Parent	Country	Minor Parent	Country	Beg RBP	End RBP
1	KP860936	Ethiopia	KP6111111	China	KF744391	Rwanda	5323-5435	9058
2	AY149118	China	KR611110	China	AJ297628	China	4908-4994	9107
3	KT895080	Iran	JX188385	USA	pending	Cameroon	1225-1343	7899-8069
4	AJ297628	China	AY042184	China	AY569692	China	217	1447-2396
5	GU474635	Mexico	EU091075	Mexico	JX188385	USA	8203-8286	9096
6	KR611111	China	AY042184	China	KR611107	China	311	1316
7	AY569692	China	unknown	unknown	AY042184	China	1790	3040-3149
8	KT895080	Iran	KF744390	Rwanda	AJ278405	Australia	1261-8192	9121
9	AY569692	China	unknown	unknown	AY042184	China	6722	9063
10	AY042184	China	AJ297628	China	AF494510	China	3505-3604	4045-4298
11	AY569692	China	KR611111	China	unknown	unknown	1-3134	4849-5216
12	JX237863	Argentina	AJ278405	Australia	unknown	unknown	3913-9135	3913-9135

clock it is required to search for more parameters and priors to determine the best clock closest to reality for SCMV. However, there is strong temporal signal which can be exploited, allowing the exploration for better ways to estimate the emergence of SCMV. The phylogeographic analysis show that SCMV could be originated in the Southeastern region of Asia, Vietnam and Yunnan province in China are the ancestral SCMV isolates, the most southern isolates reported on the NCBI, closest to the Java Islands, possible center of origin of SCMV (Brandes 1919). A recent expansion of SCMV was documented in Puerto Rico, Argentina, Brazil, Peru and USA which is also complementary to the discrete phylogeography of SCMV, showing that two lineages, the Ohio-like (maize isolates) and Lousiana-like (sugarcane isolates) diversified in the Americas and spread all around the world (Brandes, 1919; Abbot, 1929; Koike and Gillaspie, 29 1989). The isolates of sugarcane from Southeastern U.S. migrated to South America, followed by introduction to the Subsaharan Africa, Middle East and Australia in the same range of time (Gillaspie et al., 1978, Gillaspie and Mock, 1979, Teakle and Grylls, 1973; Louie and Darrah, 1980, Thottappilly et al., 1993, Conj?e, 2001). This sugarcane group was first identified by Chen et al., 2002 and Alegria et al., 2003. The maize isolates from Ohio, was described for the first time as maize dwarf disease (Williams and Alexander, 1965), also closely related isolate of SCMV was identified in St. Augustinegrass in Florida in the fall of 2013 (Harmon et al., 2015). There was only one recombination event matching spatiotemporal dispersion of SCMV and was found in Jalisco, Mexico. There is a whole genome information for Jal-1, VER1 and Ohio isolate is available, the results by RDP4 show a recombination event (Table 2, recombination 5), this recombination is located by the CP, this Jal-1 recombinant is VER1 at the aminoterminal of the polyprotein and by the CP is the Ohio isolate. The Ohio isolate has an insertion of 14 aminoacids which makes this sequence particularly long in the CP, compared to the rest of isolates. There is a well defined migration of the Ohio isolate through the Pacific Mexico, Brazil, Rwanda and Ethiopia (Fig. 5). A synergism interaction have been reported in these African isolates, which are spreading with the *Maize chlorotich mottle virus* a *Machlomovirus* in the East Africa causing the maize lethal necrosis disease (Adams et al 2012, Wangai et al.,, 2012). Also, these Ohio-like isolates of Florida is causing an outbreak in St. Augustinegrass, a possible synergistic interaction could be happening with other viral organisms (Harmon et al., 2013, Alcalá-Briseño, manuscript in preparation). In Southeast Asia, was diversifiying in the possible center of origin of SCMV, two outlayers, isolates AY222743 and DQ925432, isolated from Abaca and Maranta are located as ancestral isolates of this maize lineage, this means that there is group diverging in the Musaceae family related

to SCMV, also reported as *Abaca mosaic virus* (AbaMV) (Gambley et al. 2004, Ha et al., 2008). Xu (et al., 2008) reported a group of Chinese isolates from noble sugarcane SCE/NSCE, which is positioned between the SCMV Musa isolates and the maize isolates (MZ:A) (see figure 4). The evidence of the first migration of SCMV isolates from maize originated from China in two possible introduction events one to Germany and other to Veracruz, Mexico in a range span of 40 and 25 years. In the other hand, Chinese ancestors of SCMV isolated from maize, were reported in Argentina, Germany and Spain, from a clade known as MZ:A (Xu et al. 2008; Gao et al., 2011). The geographic distribution of SCMV and the previous reported years gives a glimpse of the geographic and time migration/distribution of SCMV. There is no too many information of the phylogeography of SCMV, but definitely these preliminary studies, show in broad terms the time divergence to the MRCA, geographic origin and the possible pattern distribution of SCMV. More analysis need to be done in order to find out and confirm these partial results of these discrete phylogeographic bayesian analysis and more time to correlate the recombination events of SCMV in a possible geographic space and time. See S1 Table for isolate years, location and host.

Conclusions

- Sugarcane mosaic virus (SCMV) have strong temporal signal.
- Bayesian Skyline model and relaxed clock gives (so far) the best values for the TMRCA, positioning the appearance of SCMV about +500 years ago.
- The group of SCMV from Southeast Asia is the most ancestral (aSC).
- The phylogeography shows the possible distribution and divergence of SCMV.
- The spatiotemporal history shows possible patterns of recombination events.

Supporting Information

Fig. 1 Phylogenetic tree using the maximum likelihood for the CP of SCMV. ML was build in PhyML with the model TIM2+I+G with aLTR support.

Fig. 2 Root-to-tip analysis of the ML tree using Path-O-Gen to search temporal signal in the CP of SCMV.

Fig. 3 Representation of the SCMV history using the Maximum Clade Credibility tree generated by coalescent bayesian skyline with relaxed molecular clock generated in BEASTv2.4

Fig. 4 Estimation of the discrete ancestral location states of SCMV summarized by the Maximum Clade Credibility tree generated using the coalescent bayesian skyride with relaxed molecular clock generated in BEASTv1.8

Fig. 5 Spatiotemporal recombination and recombination break points generated in RDP4

Supp. Fig. 1 Maximum Likelihood tree of the whole genome of SCMV. ML was build in PhyML with the model TIM2+I+G with 100 bootstrap support.

S1 Table. Table with all sequences, annotations of Acc. No., hots, country, state, isolate, year and sequence.

References

- Abbot, E. V., and Stokes, I. E. 1966. A world survey of sugar cane mosaic virus strains. Sugar y Azucar. 61:27-29.
- Achon, M. A., Serrano, L., Alonso-Dueñas, N., and Porta, C. 2007. Complete genome sequences of Maize dwarf mosaic and Sugarcane mosaic virus isolates coinfecting maize in Spain. Archives of Virology. 152:2073-2078.
- Adams, I. P., Miano, D. W., Kinyua, Z. M., Wangai, A., Kimani, E., Phiri, N., et al. 2012. Use of next-generation sequencing for the identi cation and characterization of Maize chlorotic mottle virusand Sugarcane mosaic viruscausing maize lethal necrosis in Kenya. Plant Pathology. 62:741-749.
- Adams, M. J. 2006. DPVweb: a comprehensive database of plant and fungal virus genes and genomes. Nucleic Acids Research. 34:D382-D385.
- Adams, M. J., Antoniw, J. F., and Fauquet, C. M. 2004. Molecular criteria for genus and species discrimination within the family Potyviridae. Archives of Virology. 150:459-479.
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis ed. Andreas Prlic. PLoS Comput Biol. 10:e1003537-6.
- Chen, J, and Adams, M. J. 2001. A universal PCR primer to detect members of the Potyviridae and its use to examine the taxonomic status of several members of the family. Archives of Virology. 146:757-766.
- Chen, Jiong, Chen, Jianping, and Adams, M. J. 2002. Characterisation of potyviruses from sugarcane and maize in China. Archives of Virology. 147:1237-1246 Available at: http://link.springer.com/article/10.1007/s00705-001-0799-6.
- Cronjé, C. P. R. 2001. Virology in Sub-Saharan Africa:. eds. J Hughes and B O Odu. Ibada, Nigeria.: Proceedings of a Conference Organized by IITA.
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods. 9:772-772.
- Delgadillo, F. 1987. Identificación de la enfermedad de la necrosis letal del maíz en el estado de Guanajuato. Revista Mexicana de Fitopatologia. 5:21-26.
- Drummond, A. J. 2005. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Molecular Biology and Evolution. 22:1185-1192.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. 2006. Relaxed Phylogenetics and Dating with Con dence ed. David Penny. PLoS Biol. 4:e88-12.
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution. 29:1969-1973.

- Espejel, F., Jeffers, D., Noa-Carrazana, J. C., Ruiz-Castro, S., and Silva-Rosales, L. 2005a. Coat protein gene sequence of a Mexican isolate of Sugarcane mosaic virus and its infectivity in maize and sugarcane plants. Archives of Virology. 151:409-412.
- Fan, Z. F., Chen, H. Y., Liang, X. M., and Li, H. F. 2003. Complete sequence of the genomic RNA of the prevalent strain of a potyvirus infecting maize in China. Archives of Virology. 148:773-782.
- Fischer, H. U., and Lockhart, B. E. 1974. Identity of a strain of Sugarcane mosaic virus occurring in Morocco. Plant Disease reporter. 58:1121-1123.
- Frenkel, M. J., Jilka, J. M., McKern, N. M., Strike, P. M., Clark, J. M., Shukla, D. D., et al. 1991. Unexpected sequence diversity in the amino-terminal ends of the coat proteins of strains of sugarcane mosaic virus. J. Gen. Virol. 72:237- 242.
- Gambley, C. F., Thomas, J. E., Magnaye, L. V., and Herradura, L. 2004. Abaca mosaic virus: a distinct strain of Sugarcane mosaic virus. Austral. Plant Pathol. 33:475.
- Gascuel. 2009. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. :1-37.
- Gough, K. H., and Shukla, D. D. 1981. Coat protein of potyviruses. I. Comparison of the four Australian strains of sugarcane mosaic virus. Virology. 111:455-462.
- Grishsam, M. P. 1994. Strains of Sorghum Mosaic Virus Causing Sugarcane Mosaic in Louisiana. Plant Dis. 78:729.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology. 59:307- 321.
- Ha, C., Revill, P., Harding, R. M., Vu, M., and Dale, J. L. 2007. Identi cation and sequence analysis of potyviruses infecting crops in Vietnam. Archives of Virology. 153:45-60.
- Handley, J. A., Smith, G. R., Dale, J. L., and Harding, R. M. 1998. Sequence diversity in the coat protein coding region of twelve sugarcane mosaic potyvirus isolates from Australia, USA and South Africa. Archives of Virology. 143:1145- 1153.
- Harmon, P. F., Alcalá-Briseño, R. I., and Polston, J. E. 2015. Severe Symptoms of Mosaic and Necrosis in cv. Floratam St. Augustinegrass Associated with Sugarcane mosaic virusin Neighborhoods of St. Petersburg, FL. Plant Disease. 99:557-4.
- Koike, H., and Gillaspie, A. G., JR. 1989. Mosaic. In Diseases of Sugarcane, Elsevier, p. 301-322.
- Louie, R., and Darrah, L. L. 1980. Disease resistance and yield loss to Sugarcane mosaic virus in East African-adapted maize. Crop Science. 20:638-640.
- Louie, R., and Knoke, J. K. 1975. Strains of maize dwarf mosaic virus. Plant Disease reporter. 59:518-522.
- Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution. 1:vev003-vev003.

- Moradi, Z., Mehrvar, M., Nazi , E., and Zakiaghl, M. 2016. The complete genome sequences of two naturally occurring recombinant isolates of Sugarcane mosaic virus from Iran. Virus Genes. 52:270-280.
- Oertel, U., Schubert, J., and Fuchs, E. 1997. Sequence comparison of the 3? terminal parts of the RNA of four German isolates of sugarcane mosaic potyvirus (SCMV). Arch Virol. 142:675-687.
- Shapiro, B. 2005. Choosing Appropriate Substitution Models for the Phylogenetic Analysis of Protein-Coding Sequences. Molecular Biology and Evolution. 23:7-9.
- Sharma, P. N., Sharma, O. P., and Sharma, S. K. 2002. Diseases of Field Crops. Eds. V K Gupta and Y S Paul. New Delhi: Indus Publishing Company.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution. 30:2725-2729.
- Teakle, D. S., and Grylls, N. E. 1973. Australian Journal of Agricultural research. 24:465-477.
- Thottappilly, G., Bosqe-Pérez, N. A., and Rossel, H. W. 1993. Viruses and virus diseases of maize in tropical Africa. Plant Pathology. 42:494-509.
- Urcuqui-Inchima, S., Haenni, A. L., and Bernardi, F. 2001. Potyvirus proteins: a wealth of functions. Virus Research. 74:157-175. W, B. E. 1919. The mosaic disease of sugar cane and other grases. echnical Bulletin United States-Department of Agriculture.
- Wang, J.-G., Zheng, H.-Y., Chen, H.-R., Adams, M. J., and Chen, J.-P. 2009a. Molecular Diversities of Sugarcane mosaic virus and Sorghum mosaic virus Isolates from Yunnan Province, China. J Phytopath. 158:427-432.
- Wangai, A. W., Redinbaugh, M. G., Kinyua, Z. M., Miano, D. W., Leley, P. K., Kasina, M., et al. 2012. First Report of Maize chlorotic mottle virus and Maize Lethal Necrosis in Kenya. 96:1582- 1582.
- Williams, L. E., and Alexander, L. J. Maize dwarf mosaic, a new corn disease. Phytopathology. 55:802-804.
- Yang, Z. N., and Mirkov, T. E. 1997. Sequence and Relationships of Sugarcane Mosaic and Sorghum Mosaic Virus Strains and Development of RT-PCR-Based RFLPs for Strain Discrimination. Phytopathology. 87:932-939.