Determining the Pan-African sweetpotato virome:

understanding virus diversity, distribution and evolution and their impacts on sweetpotato production in Africa

Dina Gutierrez[1], Martine Zandjanakou-Tachin [2], Steffen Schulz[1], Segundo Fuentes[1], Maria Andrade[1], Douglas Miano[3], Joseph Ndunguru[4], Settumba Mukassa[5], Elizabeth Ngadze[6], Martin Chiona[7], Britta Kowalski[1], Zhangjun Fei[8], Shan Gao[8], Joseph Nii Lamptey[9], Edward Carrey[1],Jan Kreuze[1]

[1] International Potato Center
[2] Université d'Abomey-Calavi, Benin
[3] Kenya Agricultural Research Institute
[4] Mikocheni Agricultural Research Institute, Tanzania
[5] Makerere University, Uganda
[6] University of Zimbabwe
[7] National Agricultural Research Organization, Zambia
[8] Boyce Thomson Institute
[9] Crops Research Institute Ghana

Background

Food security remains a huge challenge for the millions of Africans dependent on agriculture for their subsistence. A low-level agricultural productivity and a high percentage of poor and undernourished people are common in Africa, particularly in sub-Saharan Africa (SSA). Sweetpotato, *Ipomoea batatas* (L.) Lam. (Family *Convolvulaceae*), is among the most important food crops in the world and an extremely important food crop for subsistence farmers in SSA. It is grown throughout the African continent and currently around 34.5% of global sweetpotato area is in Africa. SSA produces approximately 7 million tons of sweetpotato annually, only about 5% of global production. One major limitation in sweetpotato production is cultivar decline, mostly due to the cumulative effect of virus infection on this vegetatively propagated crop. Thus, viral diseases are considered a major limiting factor in sweetpotato production worldwide, and particularly in SSA. However, there is a widespread lack of basic information and understanding of virus populations throughout Africa, even though such basic information is required to manage the spread and impact of these viral diseases. This project will focus on evaluating a novel approach, deep sequencing and assembly of small RNAs from field-grown sweetpotato samples collected throughout Africa, to systematically and efficiently identify virus genome. Preliminary result from the surveys performed in Ghana and Mozambique are presented. In West-Africa the survey is expanded to include Cassava, Banana, Yams and Potato.

Methodology

- Conduct sample survey throughout Africa and generate an unbiased collection of field-grown sweetpotato samples.
- Develop a simplified procedure suitable for RNA processing, purification and storage, and small RNA library construction for sweetpotato.
- Determine continent-wide (Pan-African) sweetpotato virus genomes (viromes) by siRNA deep sequencing on geo-referenced samples.
- Develop computational methods to efficiently process and assemble siRNA sequences for sweetpotato virus genome identification.

Results

- Limited preliminary collections were performed in Ghana and Mozambique in 2010 and 2011 and submitted to siRNA sequencing by Illumina, between 250,000 and 3,000,000 siRNA sequence reads were obtained from each sample after cleaning out low quality reads
- Reads were assembled through a custom built pipeline combining de-novo assembly and alignment to known viruses, followed by identification using BLASTN and BLASTX
- Nearly all plants were infected by at least one virus, but multiple infections were most common. Only in a single sample from Mozambique no virus was detected
- SPFMV, SPVC, SPLCV and SPCSV were the most common viruses found in both Ghana and Mozambique, although strains were distinct
- SPPV was almost universally present and highly variable in sequence
- SPVG, SPV2 and SPMMV were only found in Mozambique, whereas SPCFV was frequently detected in Ghana, but only from one sample in Mozambique
- SPVCV was detected only in one sample from Ghana and SPCV was not identified in any of the samples.
- New potyvirus, alphasatelite, begomovirus, nanovirus-like and single positive stranded RNA virus of unknown classification were identified in some of the samples
- Mostly incomplete sequences were assembled, further optimization of the bioinformatics pipeline is required to assemble complete genomes automatically

Figure 1 Countries to be surveyed during 2012-2013.

Table 1 viruses identified in samples from Mozambique, x indicates virus was identified, xx indicates new variant <90% similarity to known sequences; new viruses: p=potyvirus, a=alphasatelite, n=nanovirus-like, b=begomovirus, s=single stranded RNA virus .

sample		Virus identified										
#	region	SPFMV	SPVC	SPVG	SPV2	SPMMV	SPCSV	SPCFV	SPLCV	SPPV	SPSMV	new
1	Angonia									X		
2	Angonia	X								X		
3	Angonia	X	X							X		
4	Angonia	X	X		X	X	X		X	X		
5	Angonia	X	X	X			X		X	X		p
7	Angonia	X					X		X	X		
8	Angonia	XX	XX	X						X		p
10	Angonia	X							X	X		
14	Gurue											
15	Gurue	XX	XX		X	XX	X			X	X	
17	Gurue	X	X		X	XX	X		X	X		
18	Gurue	XX	X	X	X	XX				X	X	p
19	Gurue	X					X			X	X	
20	Gurue	XX	X	X						X	X	p
21	Gurue	X										
22	Gurue	X		X			X		X		X	
24	Gurue									X		
25	Gurue	XX	X		X	XX	X			X	X	
27	Gurue	X	X		X	X	X			X	X	
28	Maputo	XX							X	X		
31	Maputo	XX			X	X	X			X		p
32	Maputo	XX	X	X					X	X	X	p
33	Maputo	X	X		X	X			X	X		
34	Maputo	X	XX			X	X		X	X		
35	Maputo	X					X			X		
36	Maputo	XX	XX	X		XX	XX			X		
37	Maputo	X	X	X					X	X		
39	Maputo	X	X				X			X		
41	Chockwe	XX	X	X		XX	X		X	X	X	
42	Chockwe	X	X	X		X	X		X	X		
44	Chockwe	XX	X	X		X		XX	X	X	X	
46	Chockwe	XX	X	X	X	XX	X		X	X		p,a,n
49	Chockwe	X				XX	X		X	X		b
51	Chockwe	X	X	X		XX	X		X	X		p,a,n
55	Chockwe	XX				X	X		X	X		b
66	TANZANIA	X					X		X	X		b

Table 2 viruses identified in samples from Ghana. See table 1 for explanations.

		Virus identified								
sample #	region	SPFMV	SPVC	SPCSV	SPCFV	SPLCV	SPVCV	SPPV	SPSMV	new
116	Baugonia					X		X		
23a	Bompro	X					X	X		
19a	Dompase	XX	XX	X		X				
30a	Dzogodze	XX				X		X		
104	Ejura	X				X		X	X	
102	Ejura	XX	X			X				
103	Ejura	X	XX	X				X		
105	Ejura	X		X		X		X	X	
11b	Esukyeano	X		X	XX	X		X		
12a	Esukyeano	X			X			X	X	
152b	Fumesua	X		X		X		X	X	
156a	Fumesua	X	X	X	X	X		X		
27a	Gomakarde	XX	X					X		
26a	Gomakarde	XX	X	X				X		p
25a	Gomakarde	X		X		X		X		
133	Kamboinse	X		X		X		X	X	
137	Kamboinse	X		X		X				a
140	Kamboinse	X		X		X		X	X	a
135	Kamboinse	X		X				X		
13a	Komenda	XX	XX	X	X			X		
14a	Komenda	X	X					X		
15a	Komenda	XX			XX			X		
39a	Kporkuve	X		X		X		X		
6a	Krobo Kwamu	XX	XX		XX			X		
7a	Krobo Kwamu	X	X		X	X		X		
8a	Krobo Kwamu	X	X		X	X		X		
9a	Krobo Kwamu	X	X		X	X		X		
4a	Krobo Kwamu	X		X	XX	X		X		
32a	Kudzordzi Korpe	XX	XX			X		X		
31a	Kudzordzi Korpe	X				X		X		
40b	Lume			X		X		X		
41a	Lume					X		X		s
43a	Lume	X				X		X		
119	Manchoro					X		X		
122	Manchoro					X			X	
123	Manchoro					X		X	X	
124	Nimbasinia					X		X		
126	Nimbasinia	XX			X	X		X		
113	Nyangua					X		X	X	
111	Nyangua	XX	X		X	X		X		
130	Tekuru					X		X		
34a	Vume	XX	XX	X		X		X		
36a	Vume	X						X		
35a	Vume	X				X		X		

References

1. Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., Simon, R. (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. **Virology** 388: 1-7