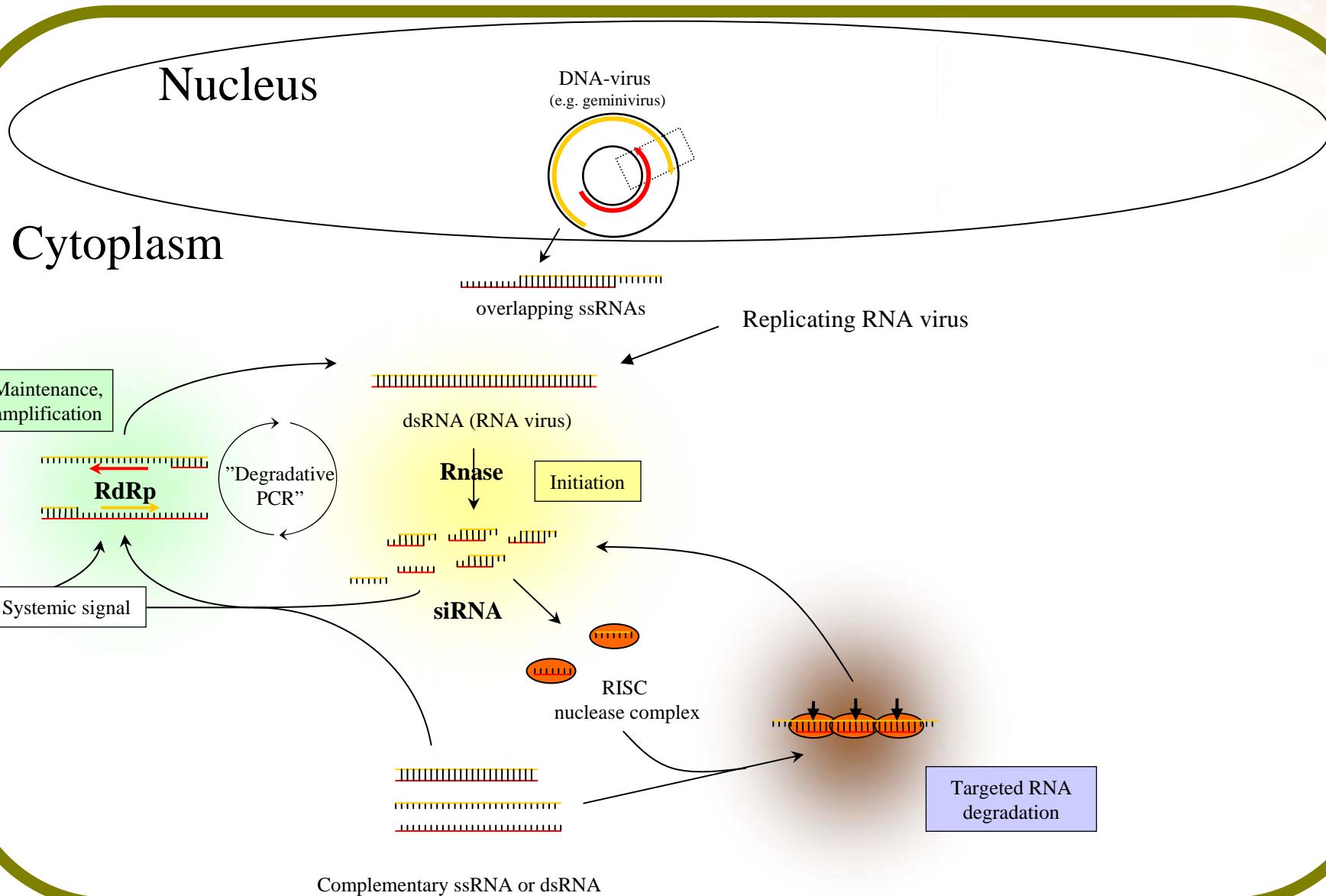


Deep Sequencing of Small RNA: a Generic Method for Diagnosis, Discovery and Sequencing of viruses

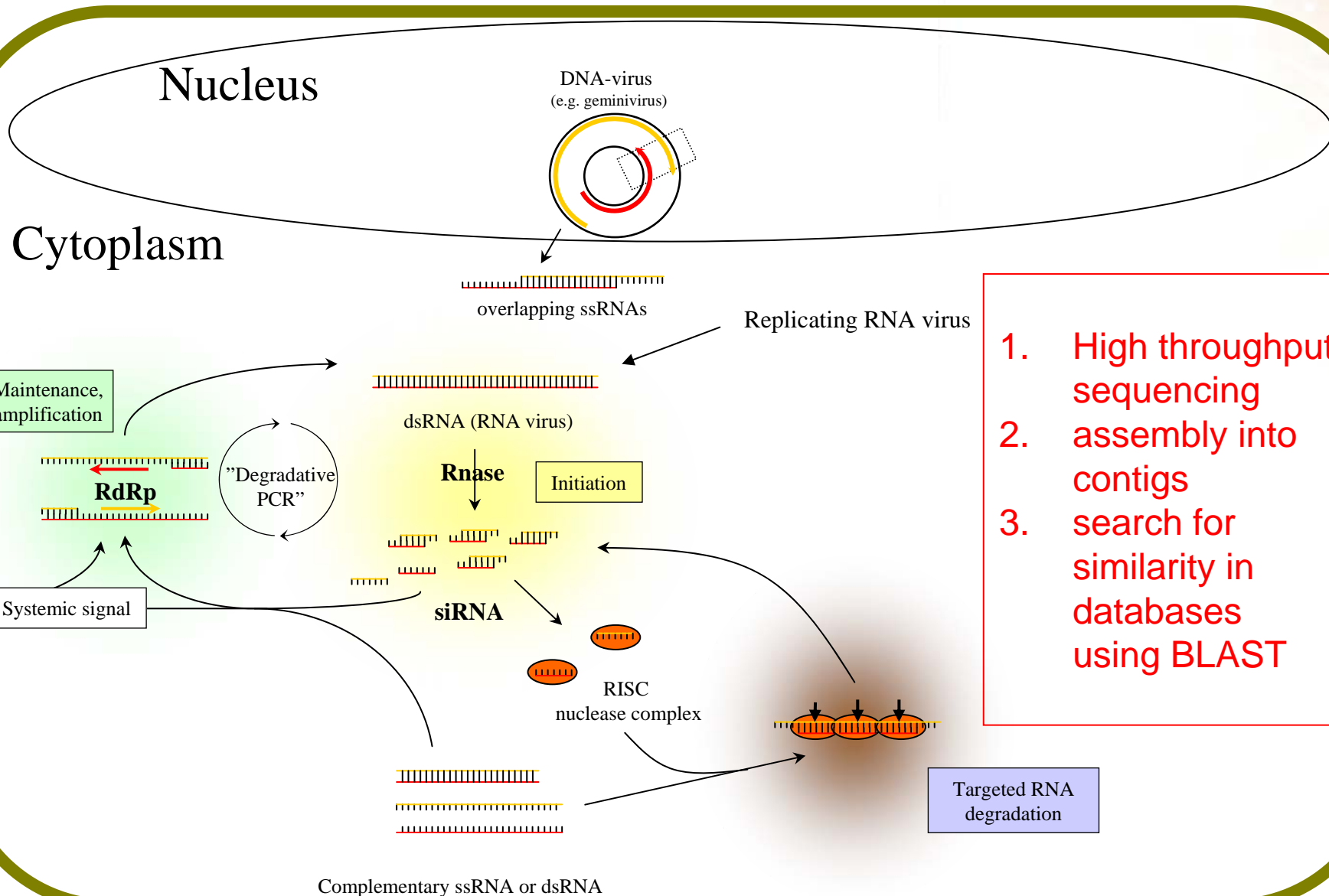
Jan F. Kreuze, Ana Perez, Milton Untiveros, Dora Quispe, Segundo Fuentes, Giovanna Muller, Ian Barker, Reinhard Simon, Claude Fauquet, Wilmer Cuellar



Anti-viral RNA silencing in eukaryotic organisms



Anti-viral RNA silencing in eukaryotic organisms



1. High throughput sequencing
2. assembly into contigs
3. search for similarity in databases using BLAST

Deep sequencing of small RNAs from virus infected sweetpotato

- *Sweetpotato feathery mottle virus* (SPFMV; genus *Potyvirus*, family *Potyviridae*), *Sweetpotato chlorotic stunt virus* (SPCSV; genus *Crinivirus*; family *Closteroviridae*) and SPFMV+SPCSV infected sweetpotato leaf samples
- Virus strains not sequenced
- Short sequence assemblers may be able to assemble viral siRNAs?

Assembly of small RNA sequences into virus specific contigs

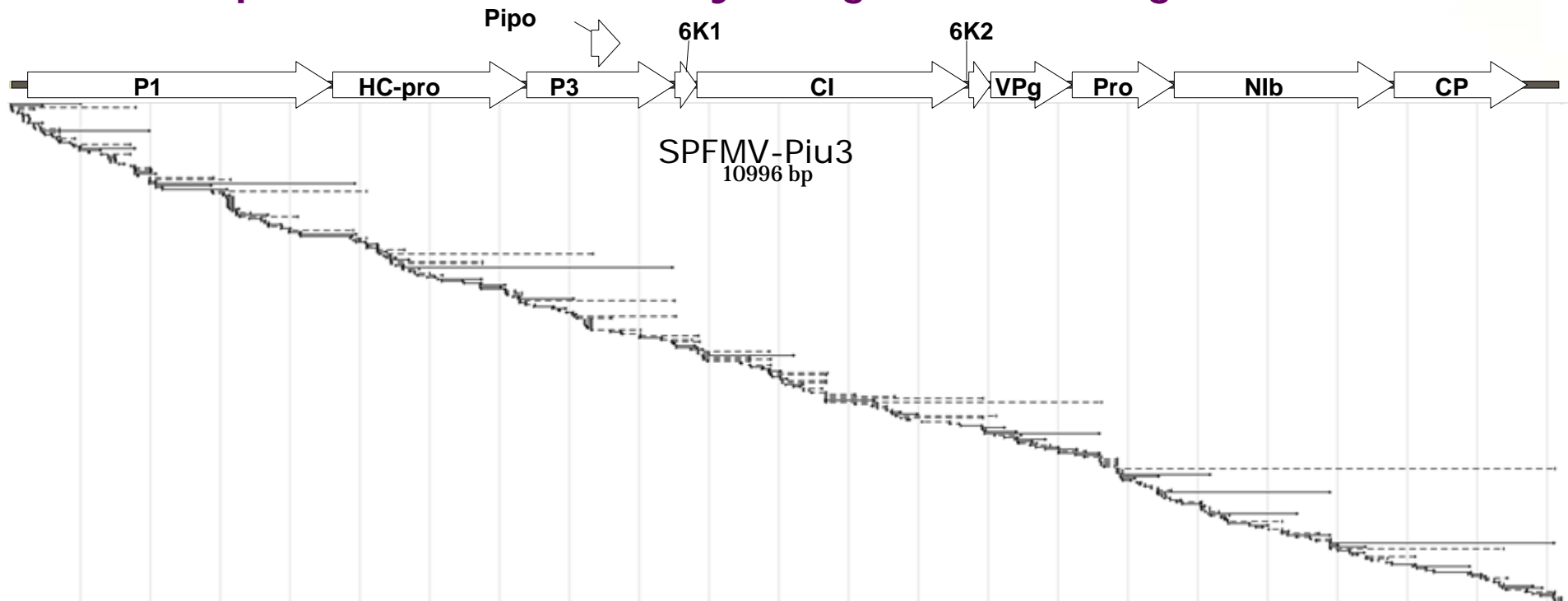
- De novo assembly of short reads (21-24 bp): SSAKE; VCAKE; Velvet
- Hundreds of contigs 50-3150
- Blast against database (virus sequences):

Number of contigs assembled by Velvet using 21–24 nt sRNA, or only 22 nt sRNA sequences, with virus specific hits as identified using Translated Nucleotide Blast (Blastx).

Plant infected with	siRNAs sequenced	Contigs identified	Contigs with Blastx hits 21–24 nt sRNA $k = 15$, $cov = 30^a$	Contig sizes 21–24 nt sRNA $k = 15$, $cov = 30^a$	Contigs with Blastx hits 21–24 nt sRNA $k = 15$, $cov = 3^b$	Contigs with Blastx hits only 22 nt sRNA $k = 15$, $cov = 3^b$
SPFMV	1,275,673	Total contigs	239		1633	431
		SPFMV	25	(≤ 949)	71	78
		SPCSV	0	–	0	0
		Badnavirus	38	(≤ 256)	62	55
		Mastrevirus	5	(≤ 210)	6	5
SPCSV	1,271,382	Total contigs	283		1675	285
		SPFMV	0	–	0	0
		SPCSV	10	(≤ 70)	64	12
		Badnavirus	44	(≤ 181)	63	44
		Mastrevirus	8	(≤ 260)	10	6
SPFMV + SPCSV	1,067,577	Total contigs	221		1363	581
		SPFMV	20	(≤ 1600)	43	51
		SPCSV	2	(≤ 81)	41	12
		Badnavirus	38	(≤ 266)	63	51
		Mastrevirus	5	(≤ 210)	8	5

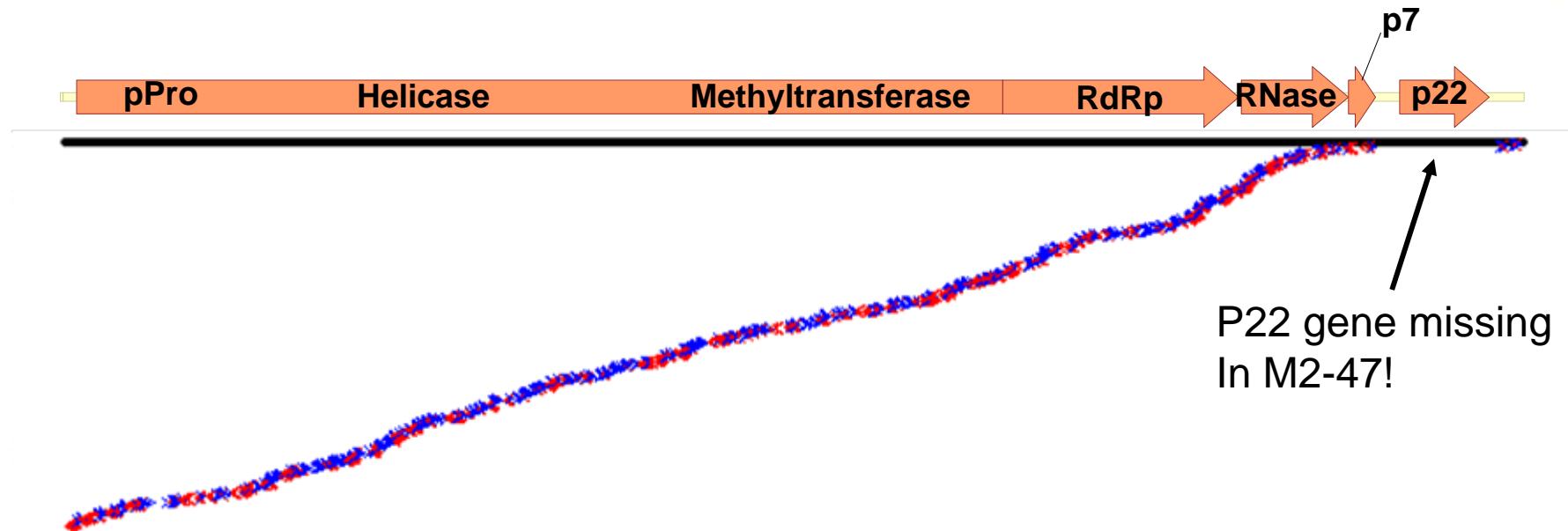
Sequencing by siRNA: complete nucleotide sequence of SPFMV-Piu3 (EA strain)

- siRNA contigs could be further assembled into bigger contigs
- Complete coverage of genome at average sequencing depth of 470x
- Sequence confirmed by Sanger: 99.8% agreement

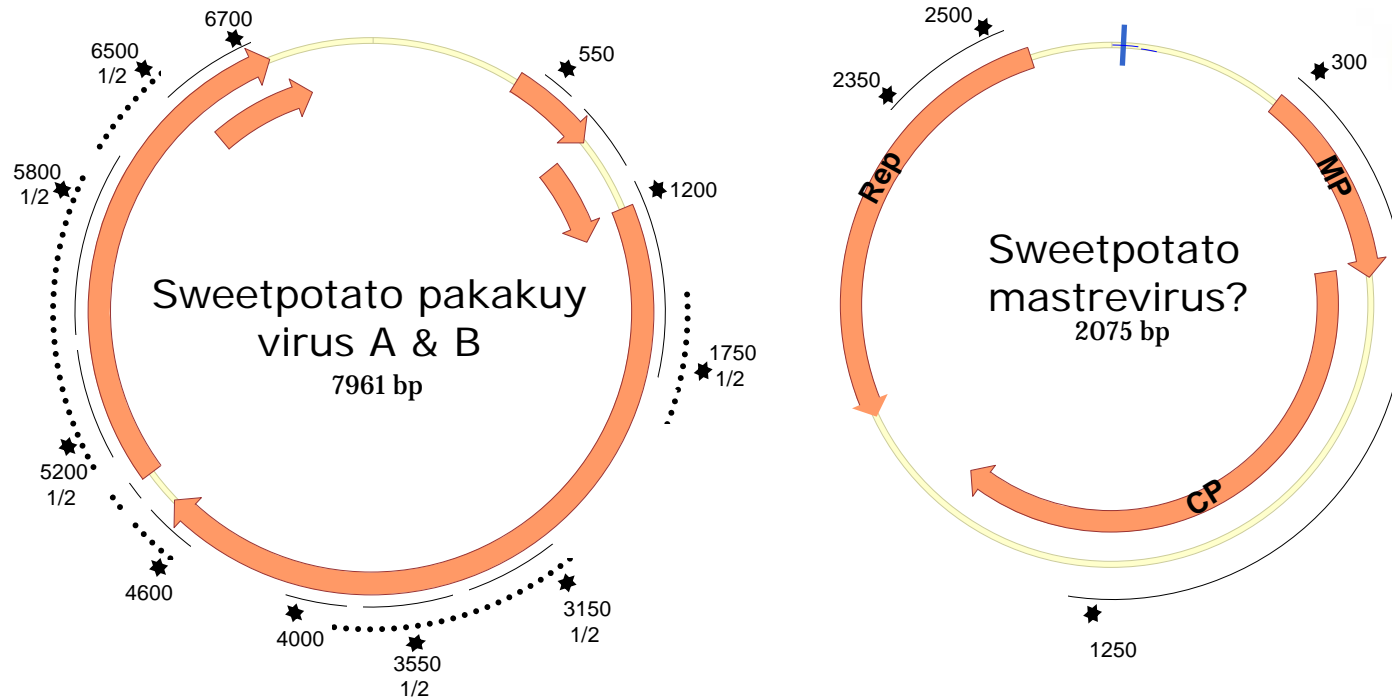


Sequencing by siRNA: Guide-strand assisted assembly of SPCSV

- Too few contigs to assemble SPCSV genome de novo
- Assembly using MAQ and sequenced strain as a guide could assemble >95% of genome
- Confirmed by Sanger sequencing

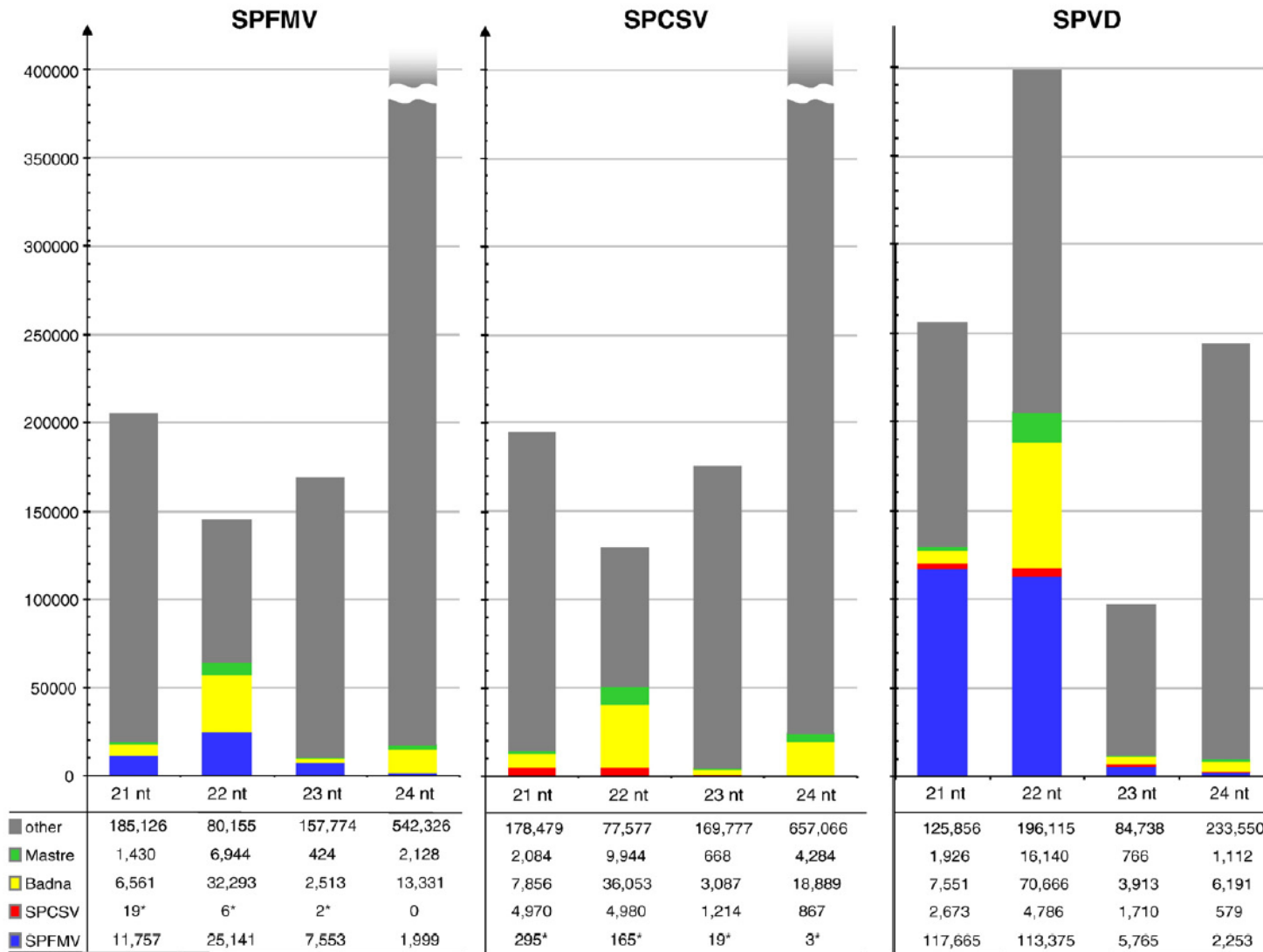


Sequencing by siRNA: ~ 50% genome of three novel viruses



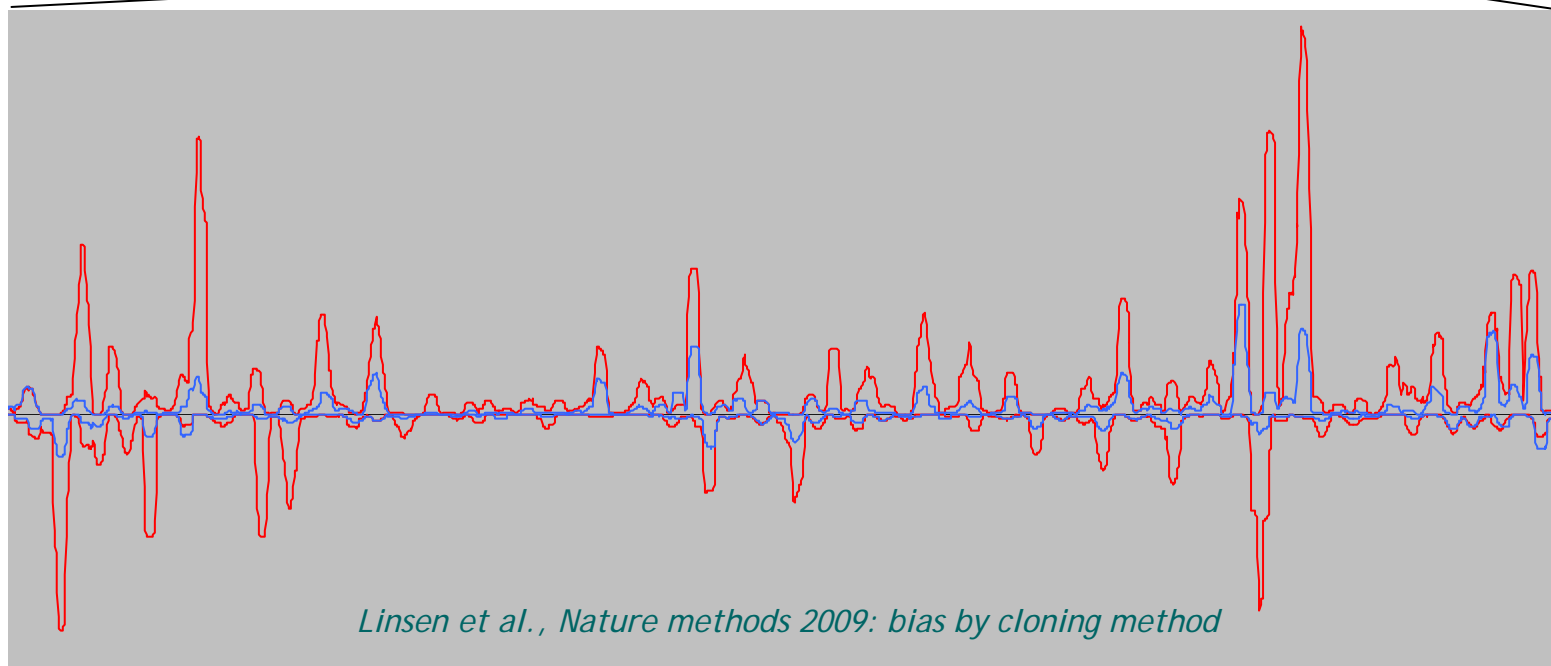
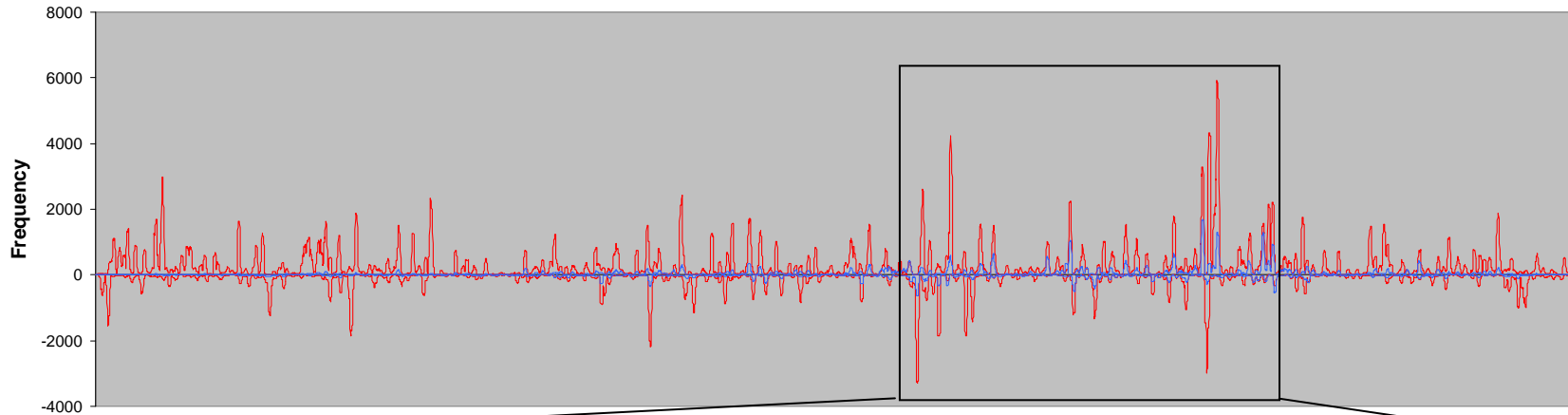
- Two pararetroviruses: Sweetpotato pakakuy virus A & B (genus *Badnavirus*; family *Caulimoviridae*)
- One single stranded DNA virus (genus ?; family *Geminiviridae*)
- Apparently symptomless, but can be transmitted by grafting to indicator host

Possibilities for improvement: most viral specific siRNAs are of 22nt size class



Possibilities for improvement: biased distribution across virus genome

SPFMV GAF1 & GAF3



Possibilities for improvement

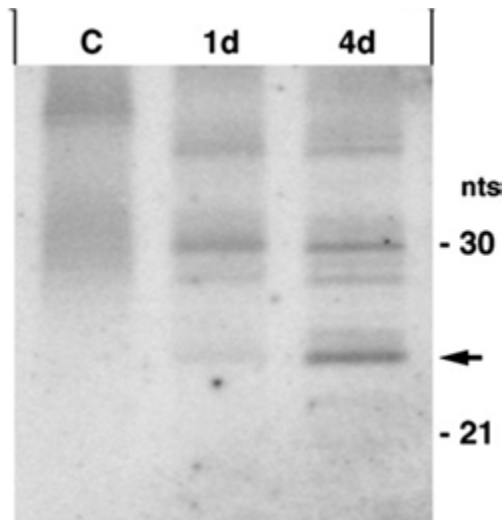
- Similar results obtained when using only 22nt siRNAs
- Simulation with random subsets indicate 120.000 21-24 nt reads required to identify SPCSV
- Only 30.000 22nt reads sufficient to recognize SPCSV
- siRNA cloning methods with less bias may improve coverage
- Bulking of samples may allow hundreds of samples to be analyzed in single lane of Illumina
- Improved extraction and sample preparation techniques needed

How broadly applicable is it: Results from other plant species

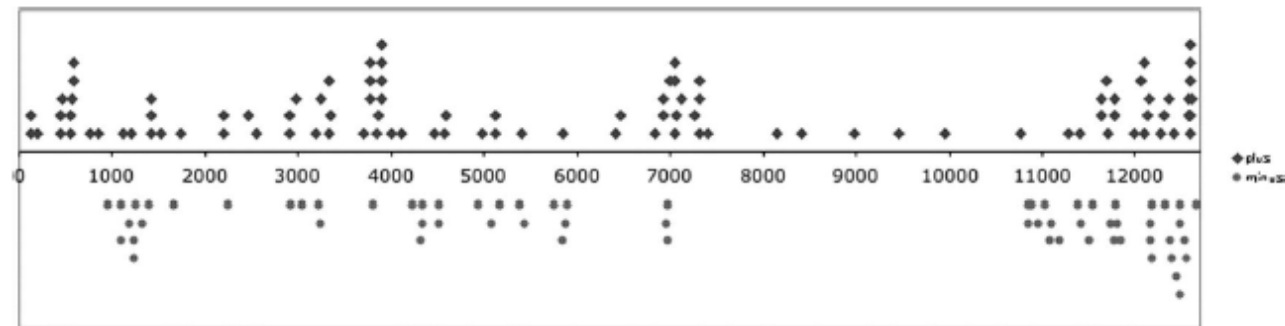
- Cassava infected with *Cassava brown streak virus* (genus *Ipomovirus*, family *Potyviridae*):
 - >97% genome sequence at 153x depth, confirmed by Sanger
 - Two additional viruses detected
- Potato infected with *Potato virus T* (Family: *Flexiviridae*):
 - >98.5% genome sequence at 32x depth
 - Cavemovirus sequences detected
- *Nicotiana benthamiana* infected with new potato disease of unknown etiology:
 - Identified as novel RNA virus

How broadly applicable is it: Other organisms

- Copious amounts of virus specific siRNAs reported from nematodes, fungi & insects:
 - Will probably work
- Abundant virus specific siRNAs not yet reported from virus infected mammalian cells!



Chotkowski et al., *Virology* 2008



Zhang et al., *J Virology* 2008

Conclusions

- Generic method for detection of viruses in plants
- RNA, DNA and reverse transcribing viruses
- Complete genomes can be assembled
- Sensitive: low titer symptomless infection & ELISA negative plants
- New viruses: no prior knowledge required
- Most likely also applicable to invertebrate animals & fungi
- Mammals?



Howard Buffett
Foundation

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