



Next generation diagnostics: from continental viromes to field level molecular diagnostics

Jan Kreuze



What is the International Potato Center (CIP)



A research-for-development with a focus on potato, sweetpotato and Andean roots and tubers.

**50
YEARS**

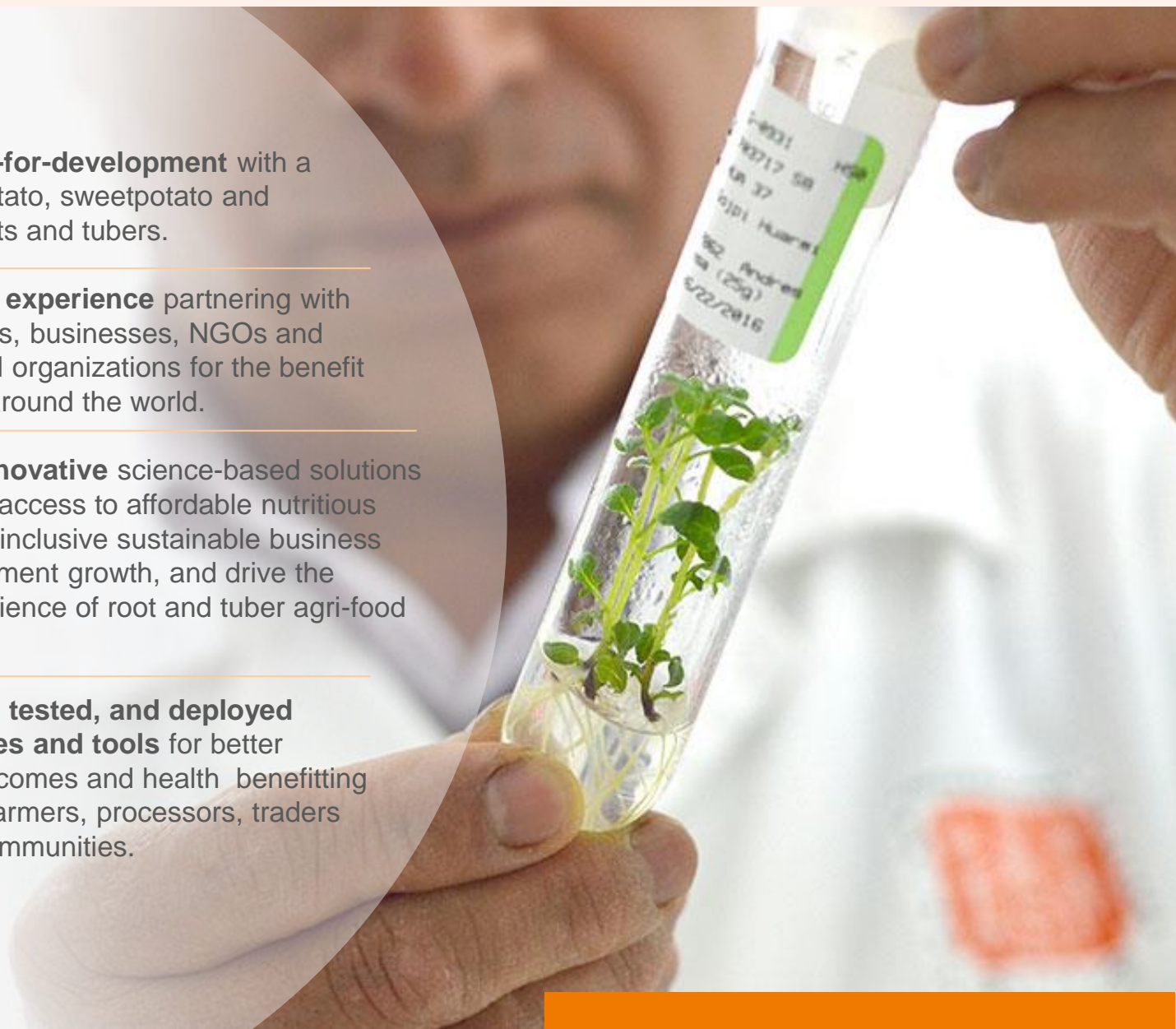
50 years of experience partnering with governments, businesses, NGOs and international organizations for the benefit of farmers around the world.



Delivers innovative science-based solutions to enhance access to affordable nutritious food, foster inclusive sustainable business and employment growth, and drive the climate resilience of root and tuber agri-food systems.

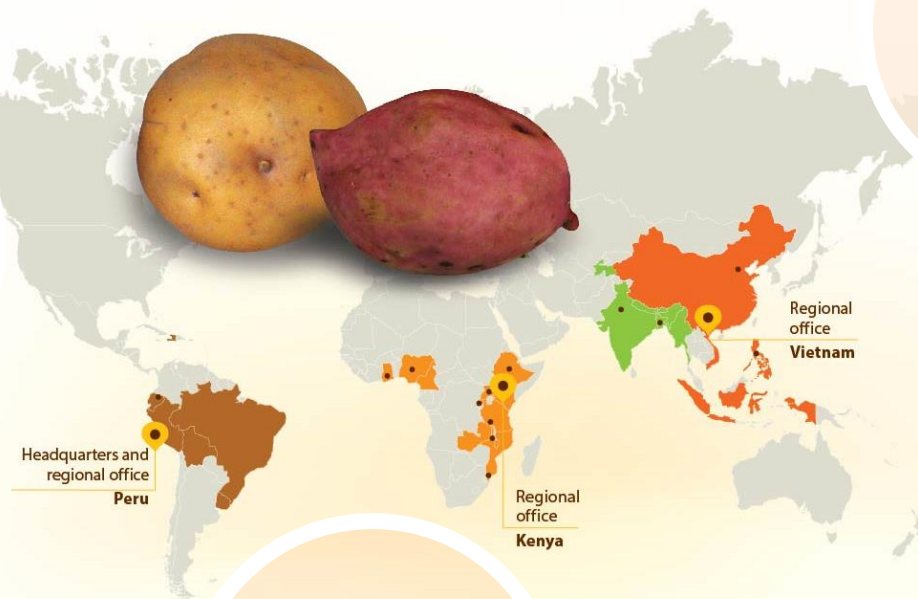


Developed, tested, and deployed technologies and tools for better harvests, incomes and health benefitting millions of farmers, processors, traders and their communities.



What is the International Potato Center (CIP)

Headquartered in Lima, Peru, CIP has a presence in 20+ countries in Africa, Asia and Latin America.



Proven experience scaling science innovations and approaches.

A CGIAR research centers: global research partnership for a food-secure future.

CIP: Tapping the potential of root and tuber crops

CIP institutional goals



Improved food and nutritional security through the introduction of healthier diets to vulnerable populations (women, young and displaced people).



Increased livelihood and employment opportunities and wide-scale development of inclusive value chains in rural and urban areas.



Sustainable intensification and diversification of agri-food systems through climate-resilient and biodiverse agriculture.

Aligned to the Sustainable Development Goals

SDGS

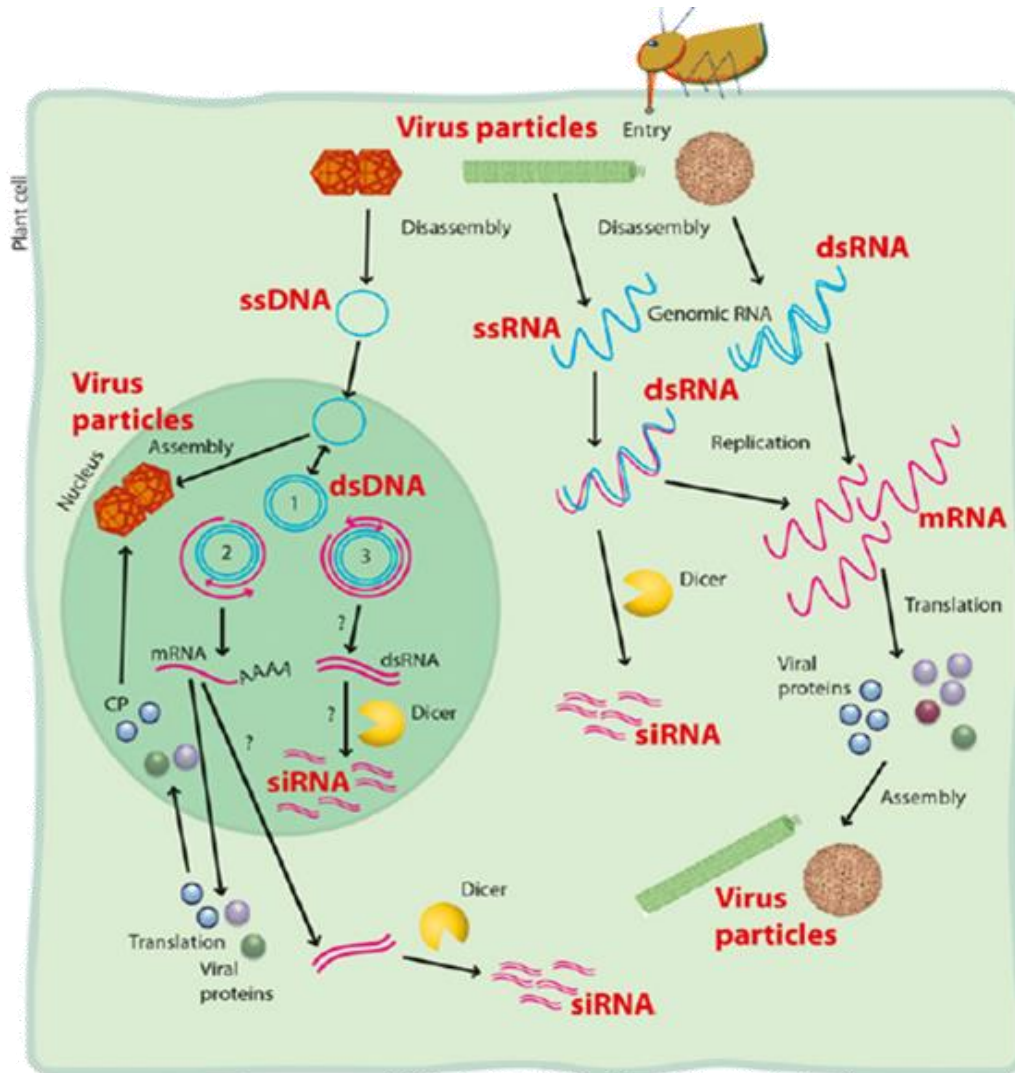




One CGIAR

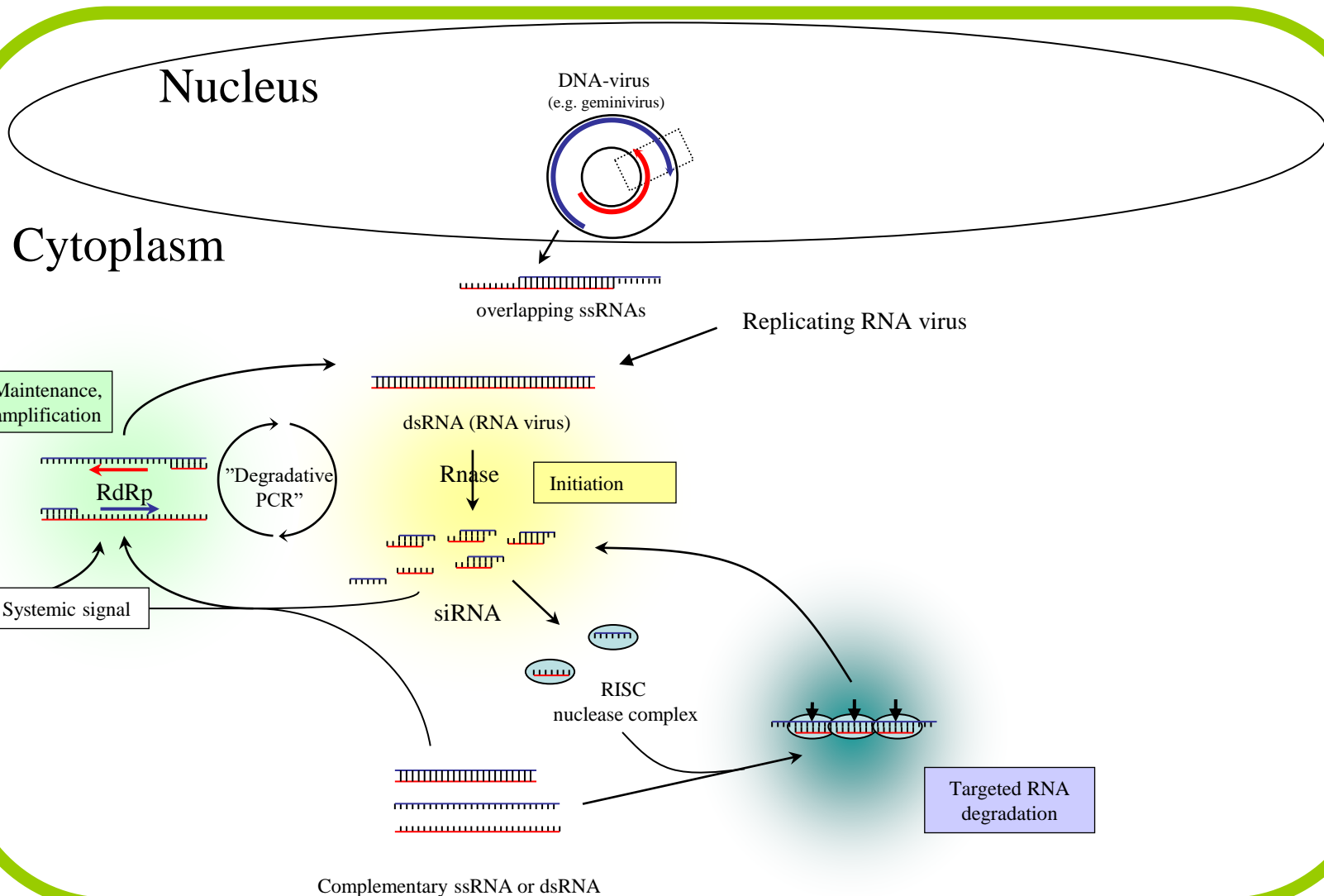


How to use high throughput sequencing?

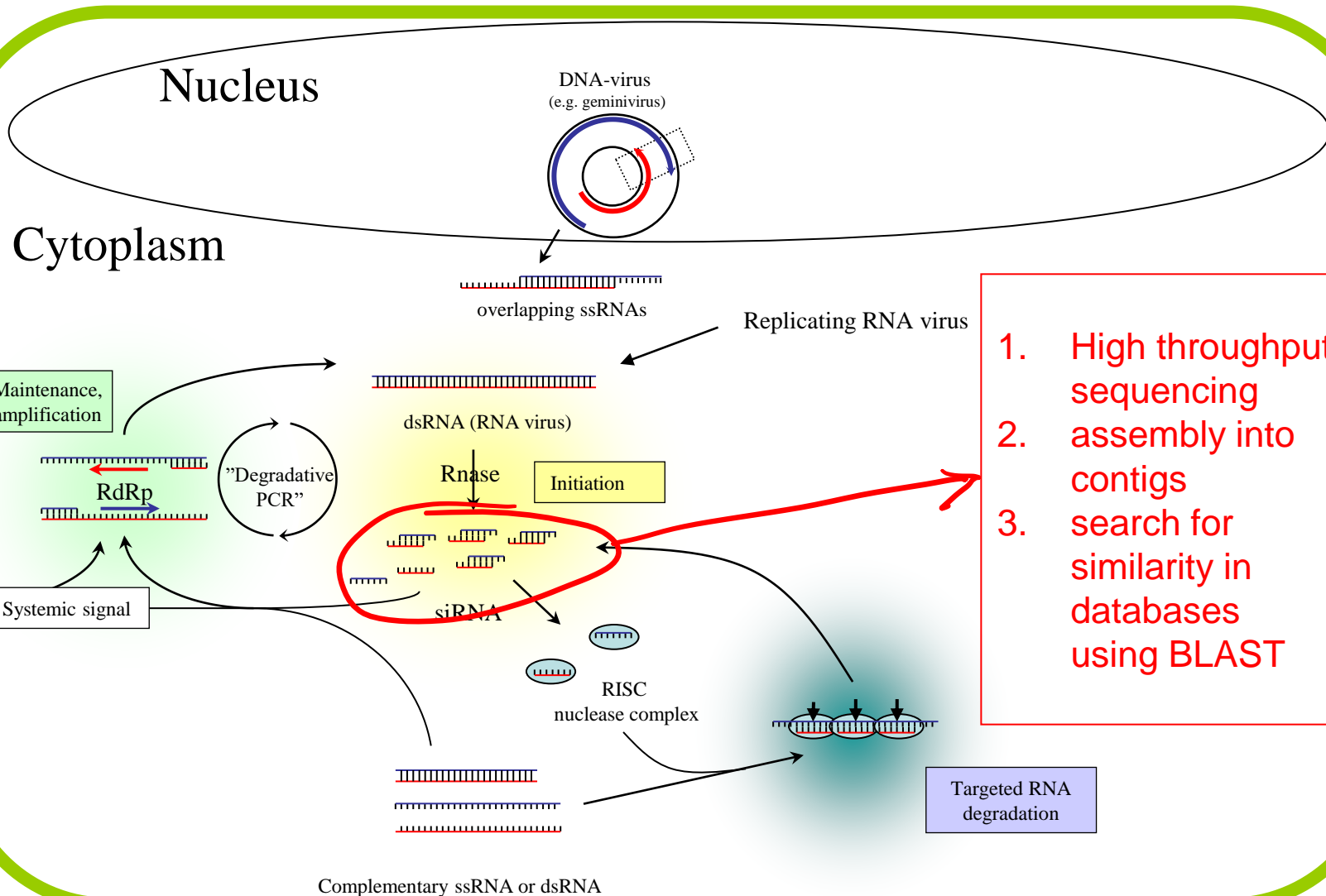


- ✓ Total DNA or RNA or ribosome depleted **RNA**
- ✓ Virion-associated nucleic acid (VANA) from **virus particle**
- ✓ Double stranded RNA (**dsRNA**)
- ✓ Virus-derived small interfering RNAs (**siRNA**)

Anti-viral RNA silencing in eukaryotic organisms



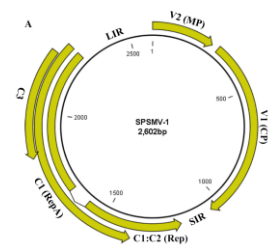
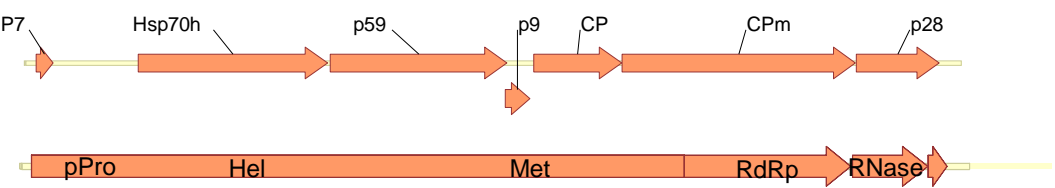
Anti-viral RNA silencing in eukaryotic organisms



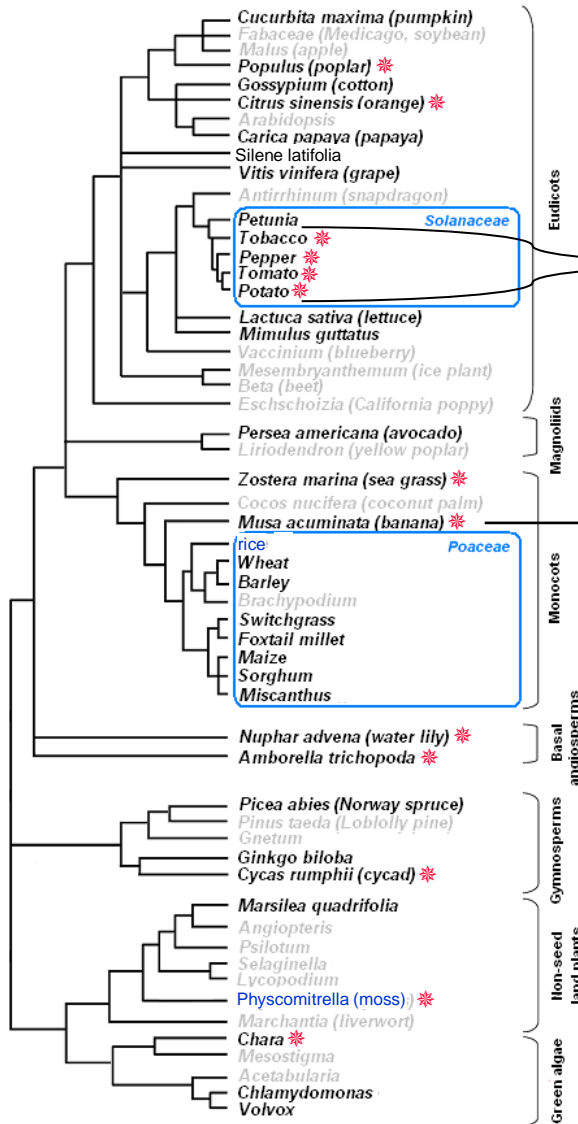
1. High throughput sequencing
2. assembly into contigs
3. search for similarity in databases using BLAST

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



Other plant species

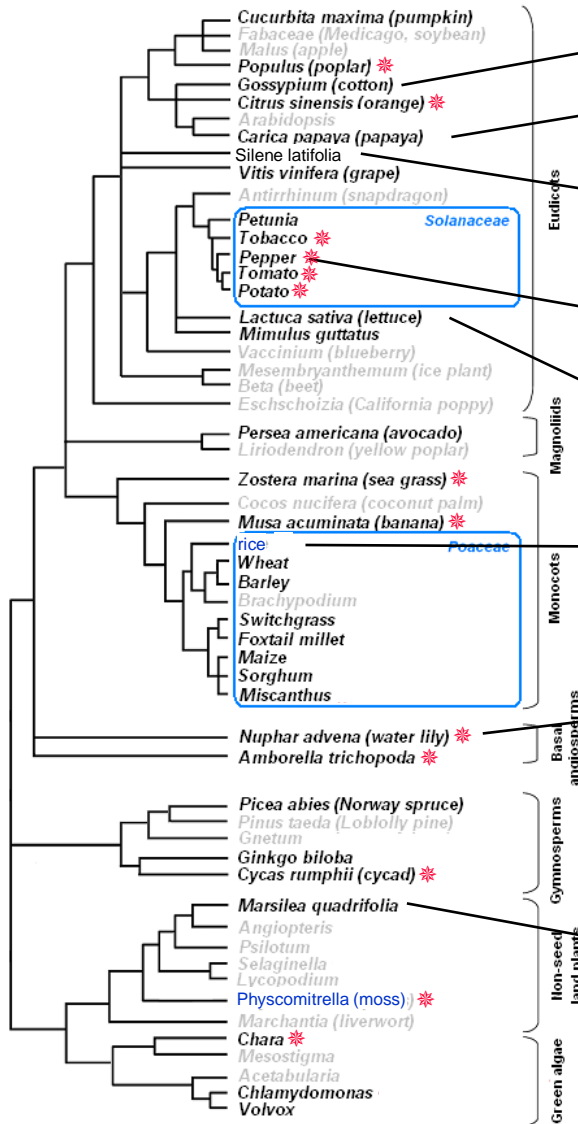


Comparative sequencing of plant small RNAs
<http://smallrna.udel.edu/>

TVCV (integrated)

Integrated badnaviruses

* similarity to caulimoviridae



✓ new totivirus

✓ Papaya ring spot virus: 99.6% coverage



✓ new partitivirus
✓ new totivirus

✓ Bell pepper endornavirus (Endornavirus)

✓ Lettuce big-vein associated virus (Varicosavirus)

✓ Mirafiore lettuce virus (Ophiovirus)

✓ Rice ragged stunt virus (Oryzavirus; Reoviridae)

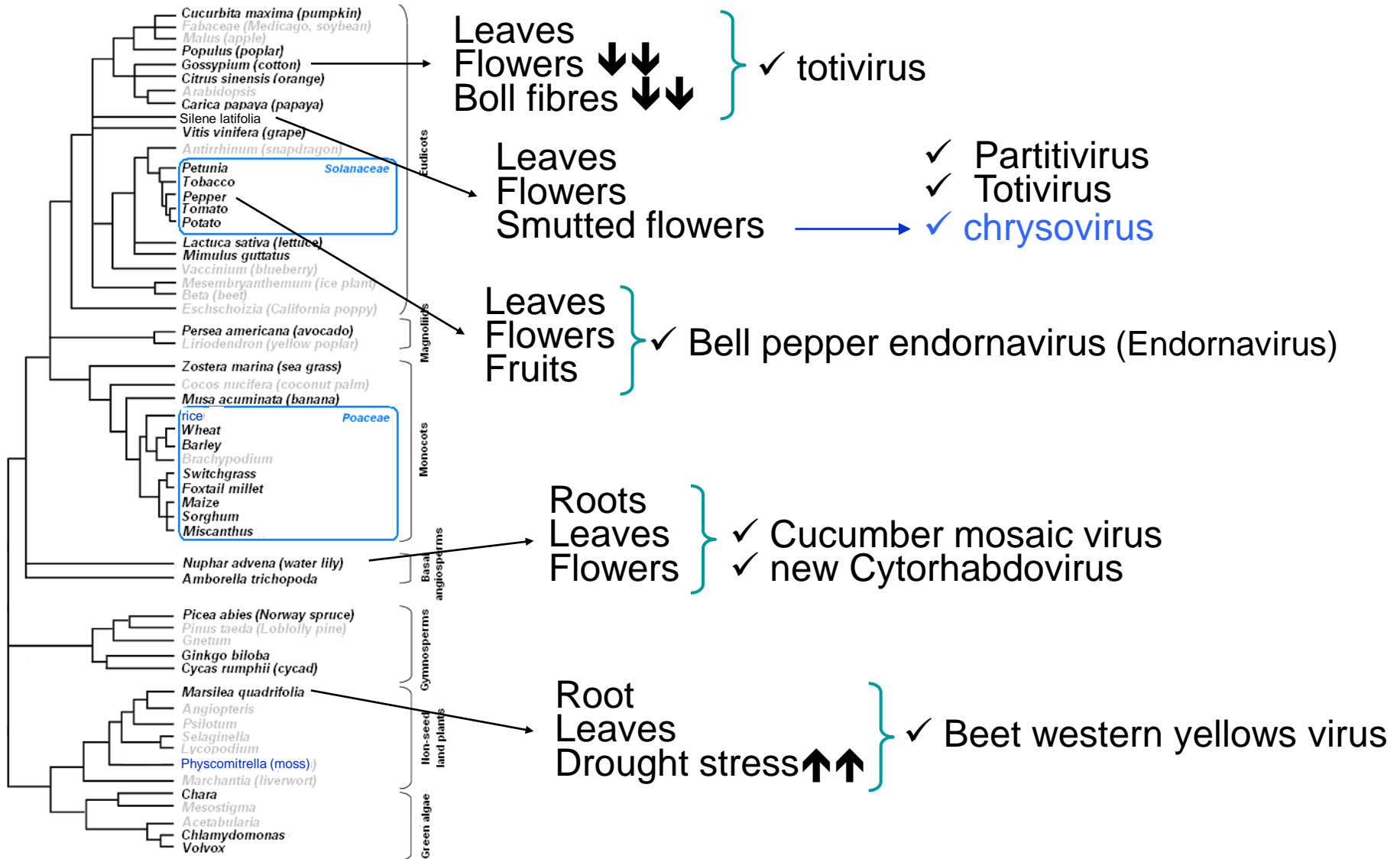


✓ Cucumber mosaic virus
✓ new Cytorhabdovirus

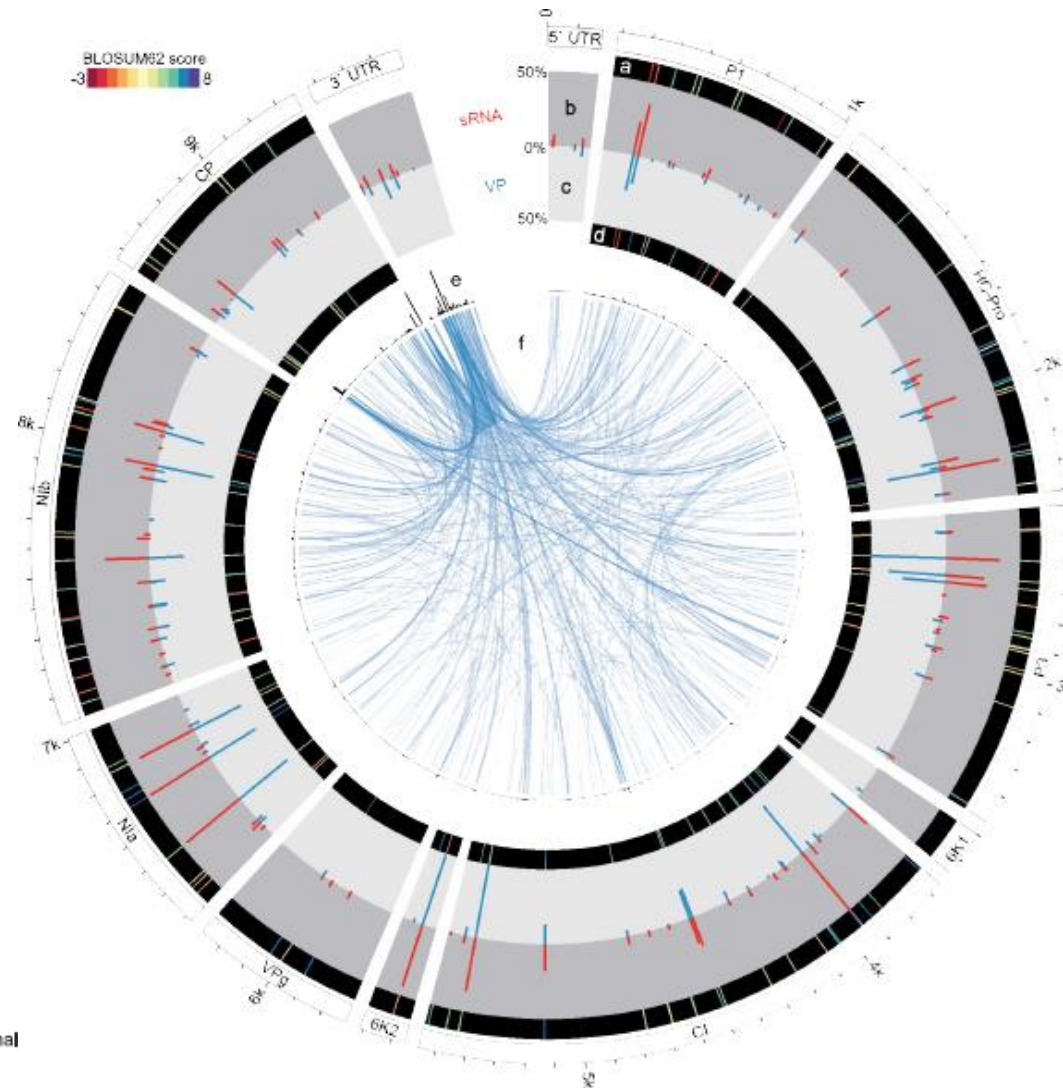
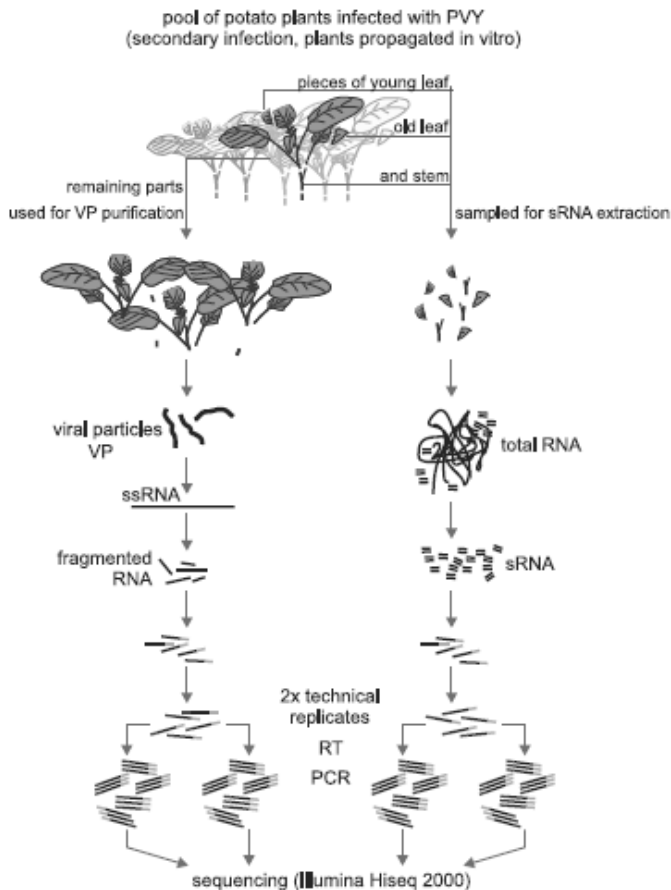


✓ Beet western yellows virus

* similarity to Caulimoviridae



SNP landscape highly similar between siRNA and RNA from virus particles



1

Consensus genome
sequence reconstruction

2

Analysis of mutational cloud

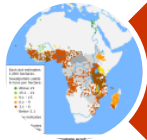
3

Recombination anal

The African sweetpotato virome



Conduct unbiased geo-referenced sample collection of field-grown sweetpotatoes throughout Africa



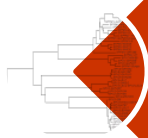
Simplified procedure for RNA processing, purification and storage, and small RNA library construction for sweetpotato.



Computational methods to efficiently process and assemble siRNA into complete viral genomes (virome) by siRNA deep sequencing on samples.

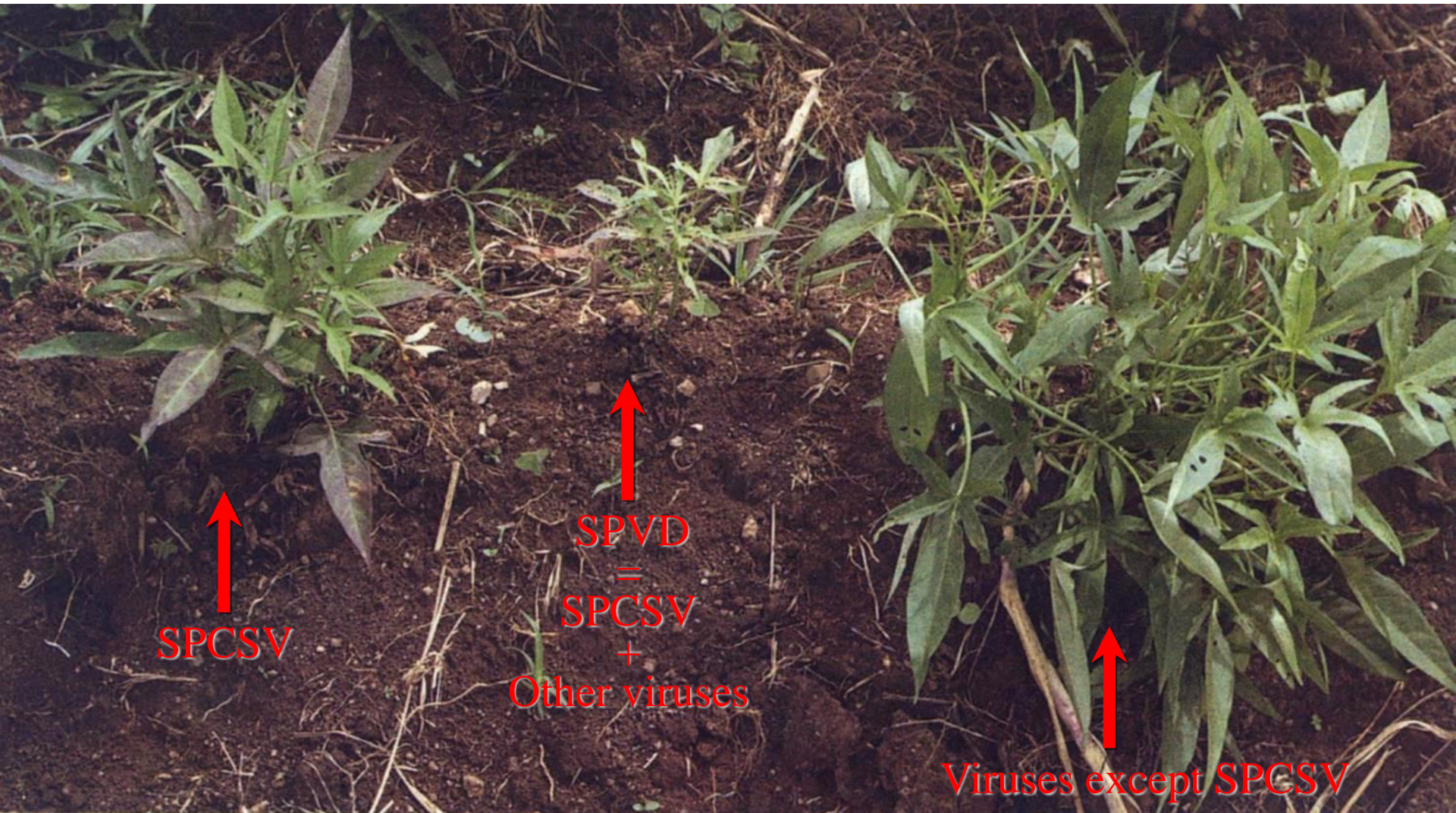


Assembly of information into database

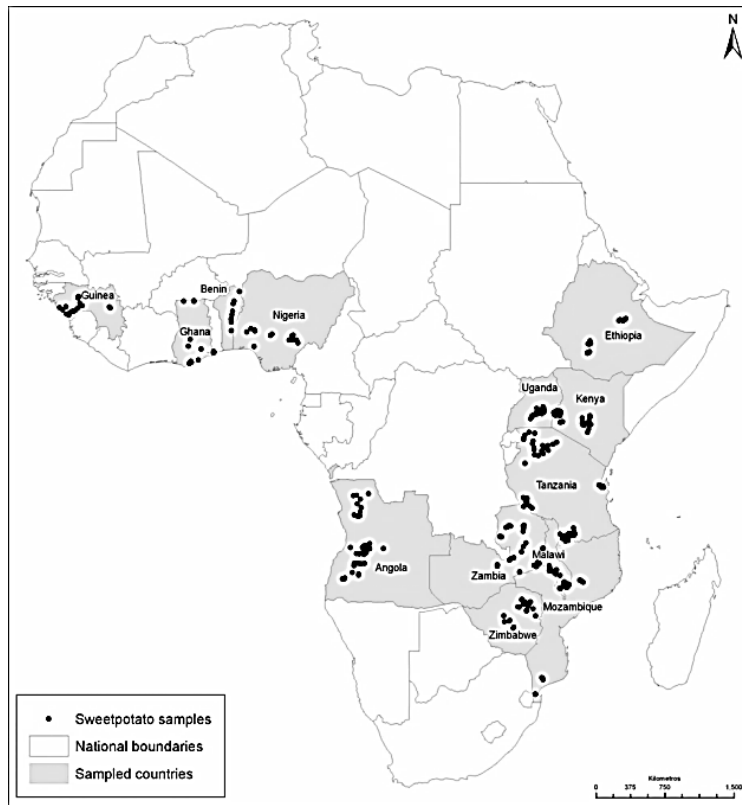


Analyze data

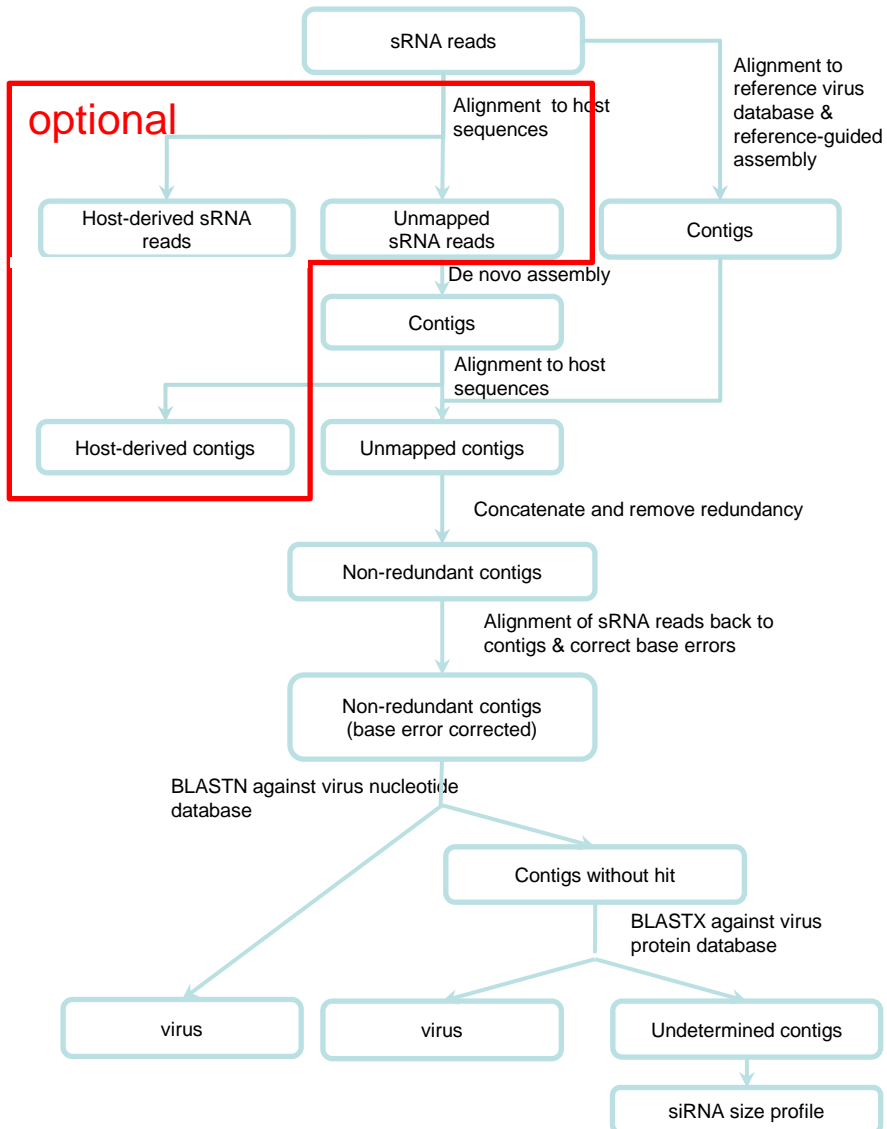
Viruses of sweetpotato



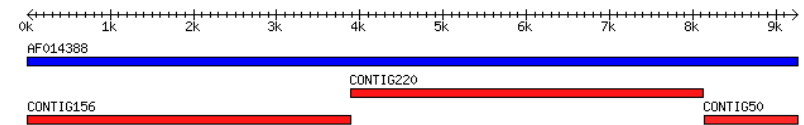
~2000 samples and bioinformatics pipeline for virus identification using siRNA assembly and genome subtraction



Small RNA sequence analysis pipeline: VirusDetect_v1.6



Reference	Length	Coverage (%)	#contig	Depth	Depth (Norm)	%Identity	%Iden Max	%Iden Min	Genus	Description
FJ150422	4806	4806 (100)	11	675.9	18.5	94.64	96.00	84.82	NA	Drosophila A virus isolate HD, complete genome.
GQ342962	3260	3257 (99.9)	3	524.3	14.4	98.11	98.75	97.03	NA	Drosophila melanogaster birnavirus SW-2009a strain DBV segment A, complete sequence.
GQ342963	3014	3014 (100)	1	854.8	23.4	98.47	98.47	98.47	NA	Drosophila melanogaster birnavirus SW-2009a strain DBV segment B, complete sequence.
KF947078	13534	1773 (13.1)	27	35.2	1.0	99.10	100	97.14	NA	Spodoptera frugiperda rhabdovirus isolate Sf, complete genome.
GQ257737	12333	12333 (100)	6	661.2	18.1	96.87	99.29	91.67	NA	Nora virus isolate Umea 2007, complete genome.
M32779	2225	2112 (94.9)	7	20.4	0.6	99.57	100	97.69	alphabaculovirus	Autographa californica nucleopolyhedrovirus insertion element IFP2.2 genomic sequence.
EF690537	3107	3089 (99.4)	5	654.2	17.9	94.82	96.59	94.12	alphanodavirus	Flock house virus isolate TNCL segment RNA1 protein A mRNA, complete cds.
EF690538	1383	1383 (100)	1	378.8	10.4	94.44	94.44	94.44	alphanodavirus	Flock house virus isolate TNCL segment RNA2 protein alpha mRNA, complete cds.
AF014388	9264	9244 (99.8)	3	1096.2	30.1	98.03	98.21	96.88	cripavirus	Drosophila C virus strain EB, complete genome.



Order	Query ID	Query Start	Query End	Subjt Start	Subjt End	Identity	E value	Strand
1	CONTIG50	1	1121	8144	9264	1086/1121(96%)	0.0	1

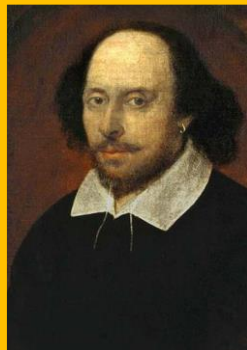
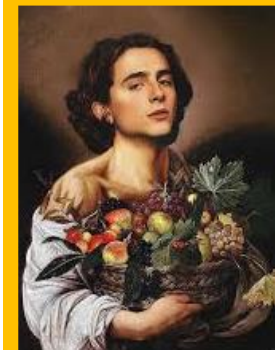
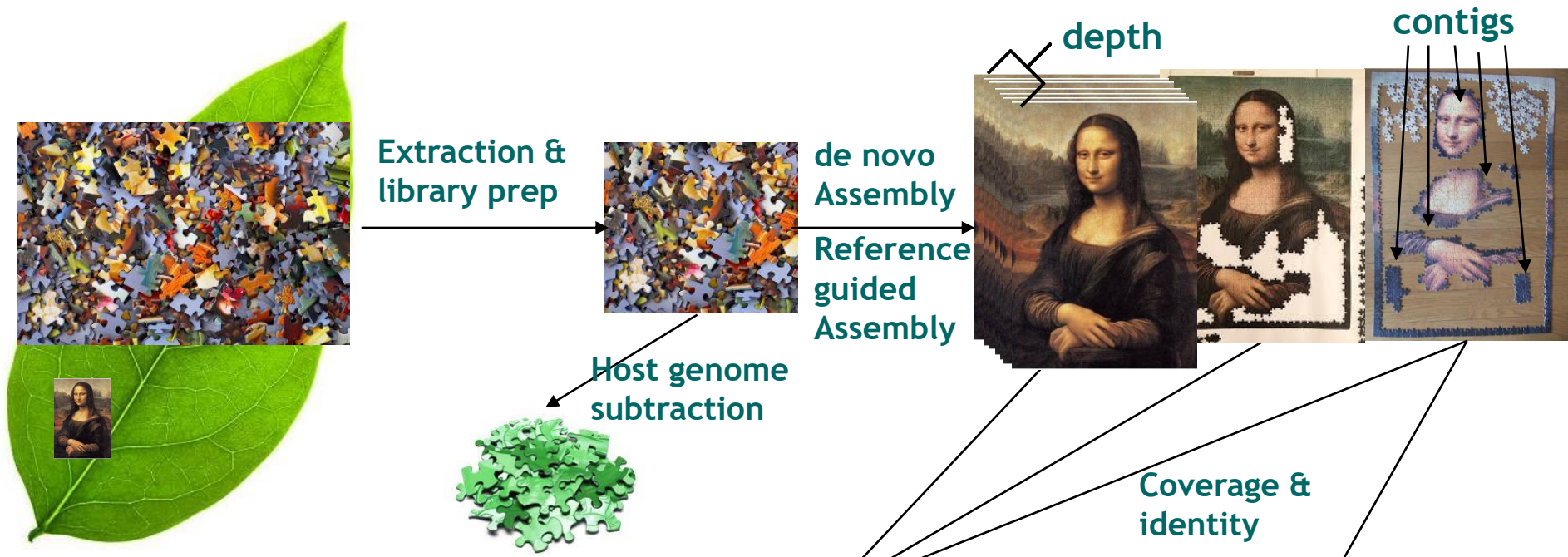
Alignment:

```

Query: 1   tgaggttgaatatgctcgtattgttgcgaagtaatgggtgaagatgtagctattcaaaag 60
          |||
Sbjct: 8144 tgaggttgaagatgctcgtattgttgcgaagtaatgggtgaagattagctattcaaaag 8203

Query: 61   aaacgatgctcaacatgggtttcatccaatgacctagacactcataagattgactcaaa 120
          |||
Sbjct: 8204 aaacgatgctcaacatgggtttcatccaatgacctagacactcataagattgactcaaa 8263
  
```

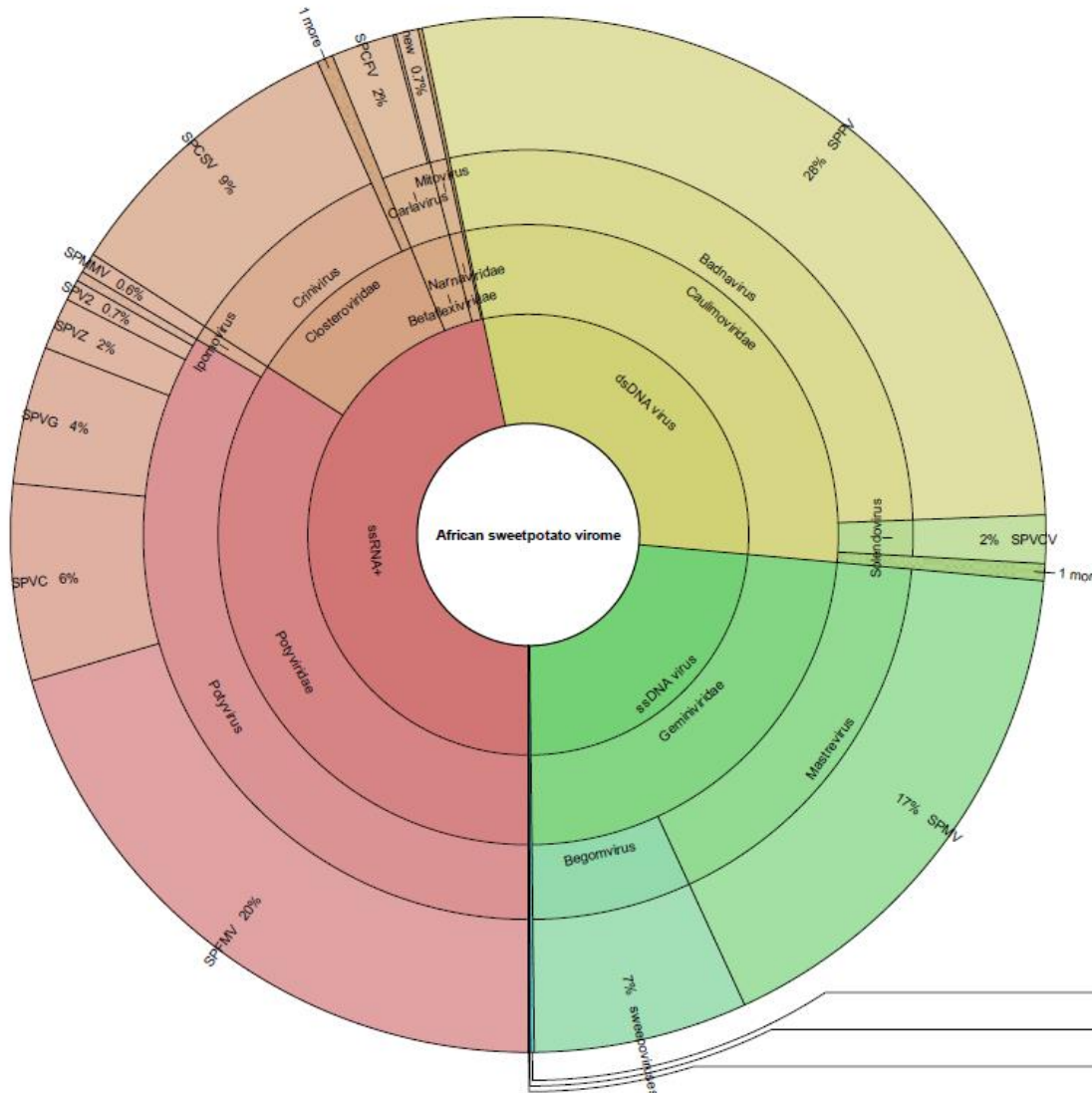
sRSA: a didactic explanation



Reference database

Results: the viruses

3193 viruses from 1168 samples



New viruses:
 Potyviruses: 3
 Ampeloviruses: 3
 Mitoviruses: many
 Nepovirus
 Soymovirus
 Emaravirus
 Begomoviruses
 Alphasatellites



Badnaviruses of Sweet Potato: Symptomless Coinhabitants on a Global Scale

Jan F. Kreuze^{1*}, Ana Perez², Marco Galvez Gargurevich¹ and Wilmer J. Cuellar¹

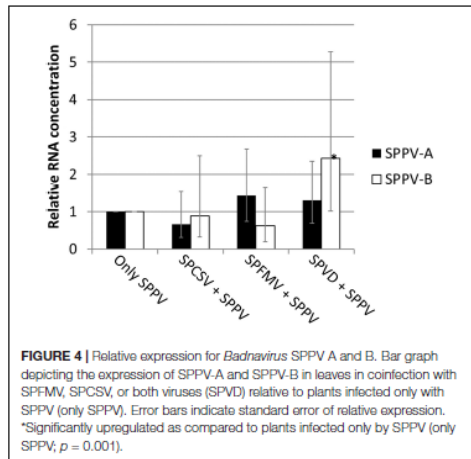
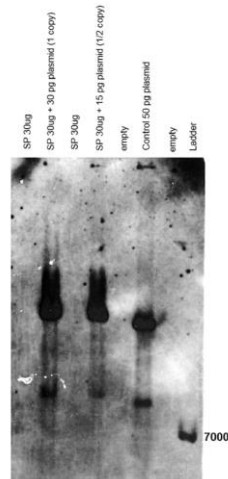
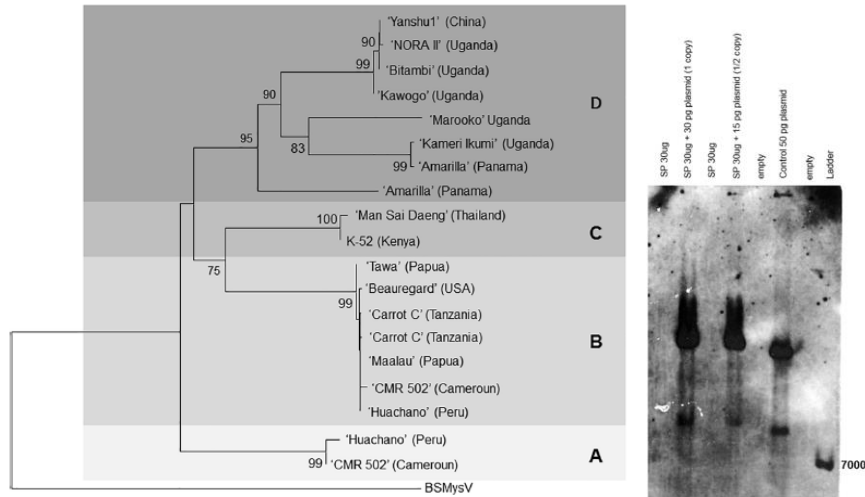
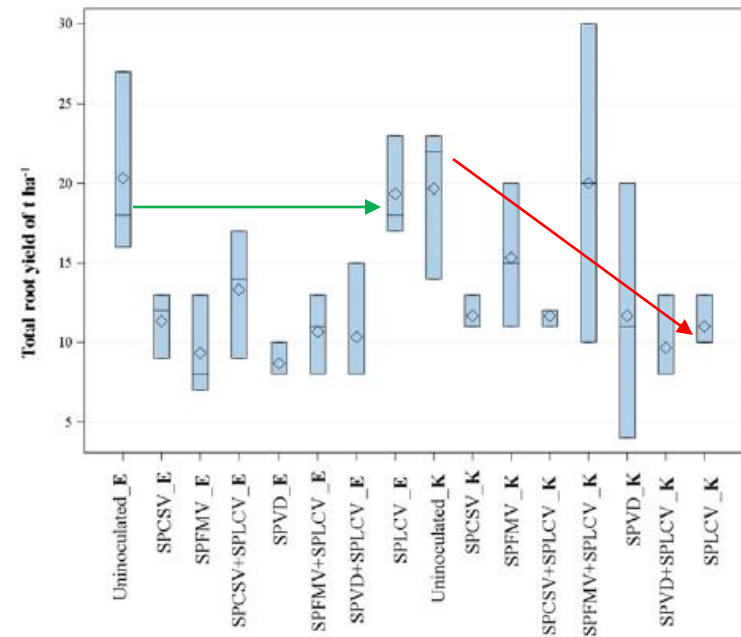


FIGURE 4 | Relative expression for *Badnavirus* SPPV A and B. Bar graph depicting the expression of SPPV-A and SPPV-B in leaves in coinfection with SPFMV, SPCSV, or both viruses (SPVD) relative to plants infected only with SPPV (only SPPV). Error bars indicate standard error of relative expression. *Significantly upregulated as compared to plants infected only by SPPV (only SPPV; $p = 0.001$).

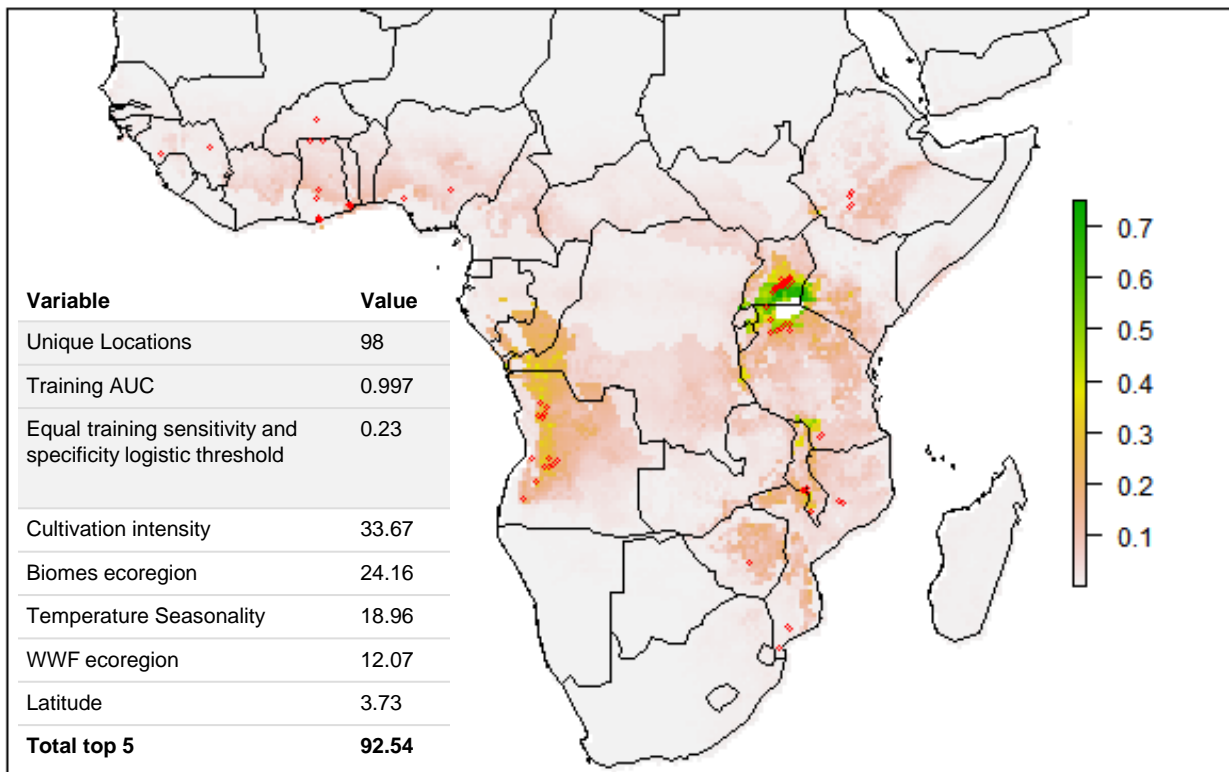
Storage Root Yield of Sweetpotato as Influenced by *Sweetpotato leaf curl virus* and Its Interaction With *Sweetpotato feathery mottle virus* and *Sweetpotato chlorotic stunt virus* in Kenya

Bramwel W. Wanjala,^{1,2} Elijah M. Ateka,² Douglas W. Miano,³ Jan W. Low,¹ and Jan F. Kreuze^{4,†}

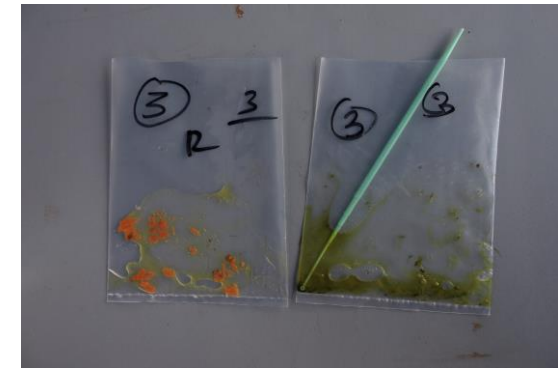
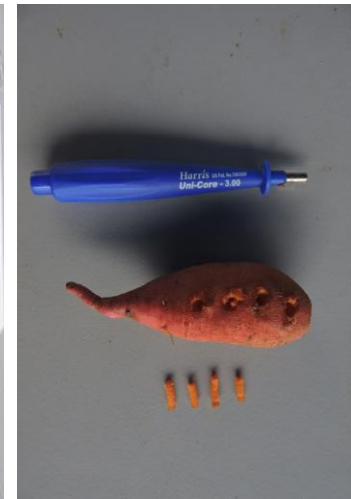
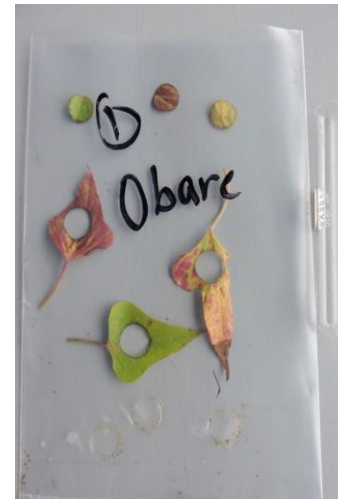
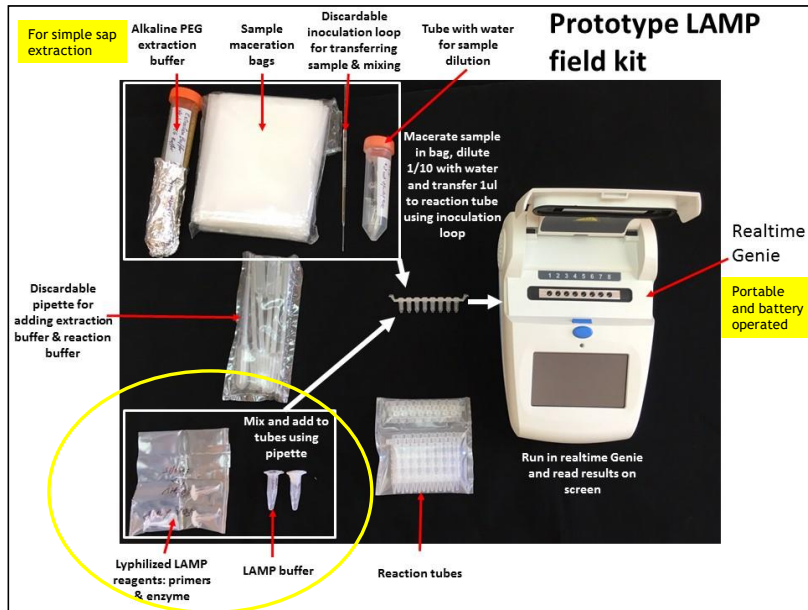


Can we make predictions about likely occurrence beyond evaluated areas?

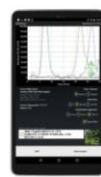
Example for begomoviruses: Niche analysis using MaxEnt modeling and bioclimatic-ecosystem variables



Field LAMP kit



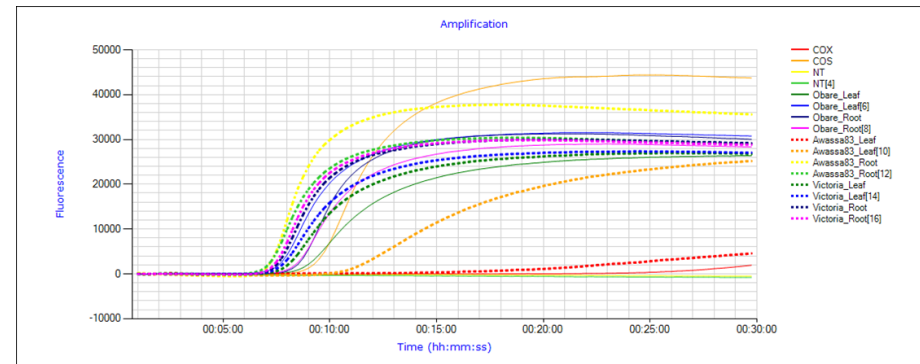
Lyophilized reagents



BioRanger

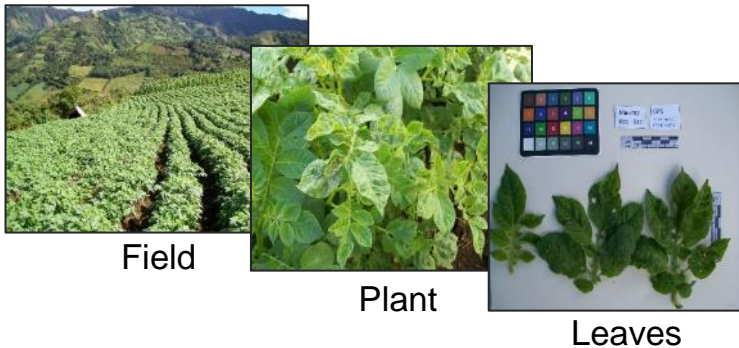


Easy to interpret



The Peruvian potato virome


- Sample collection:
994 geo-referenced potato leaf samples (Cajamarca, Huanuco, Junin, Huancavelica, Cusco)



Peruvian potato virome - Web platform

Pathogens


potpathodiv.org/vflist?fid=CuscoParuroYaurisque



Home
Map
R. solanacearum
P. infestans
Virome
Admin
About

Baseline to monitor changes in the future

District:	Yaurisque
Locality:	
Latitude:	-13.68939°
Longitude:	-71.95618°
Altitude:	3106 m



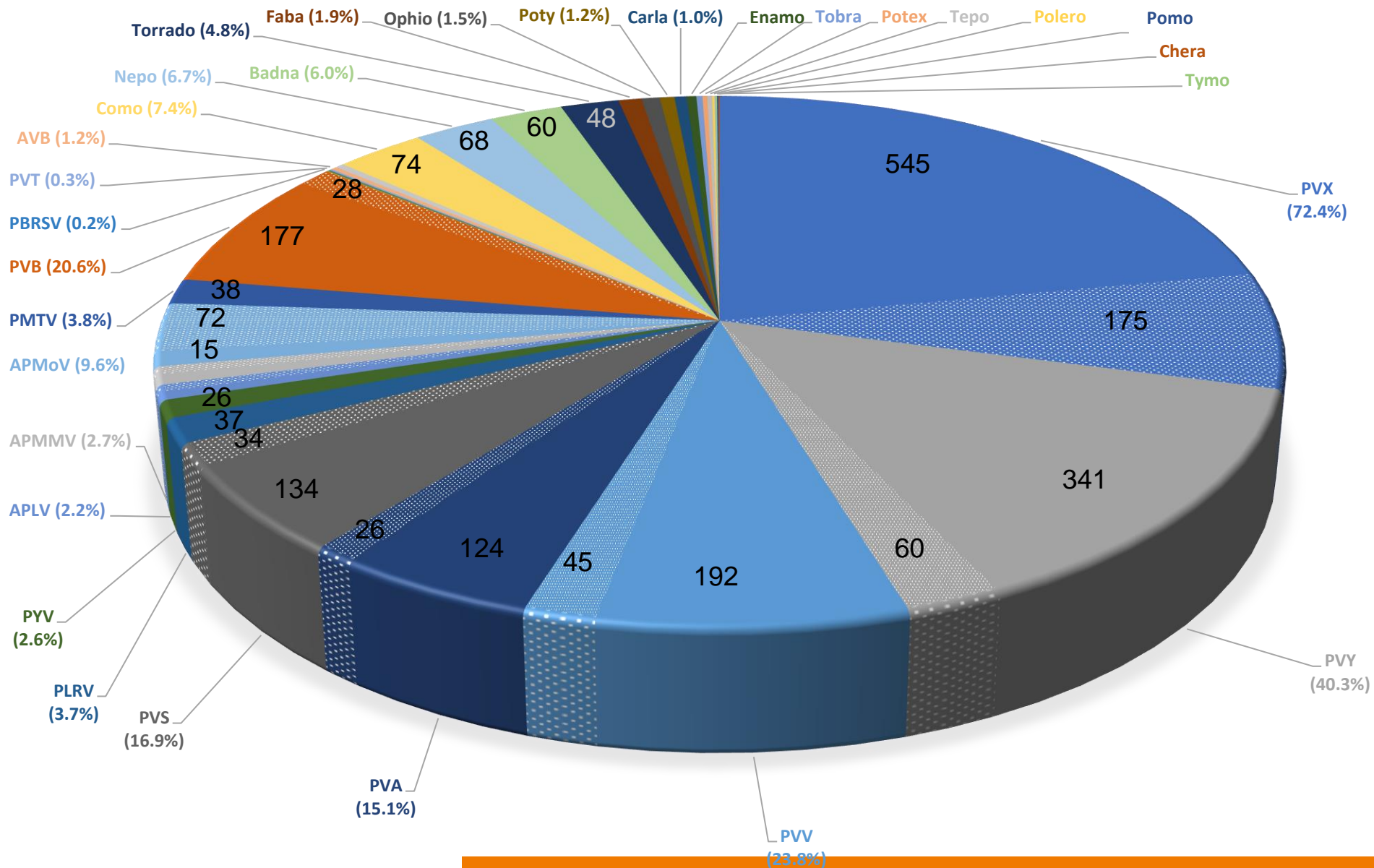
Samples in field Cus

Show 10 entries

Sample	Date	Pathogen Host	Cultivar	Sequenced
Cus-061	2016	Potato	Cica	1
Cus-062	2016	Potato	Cica	1

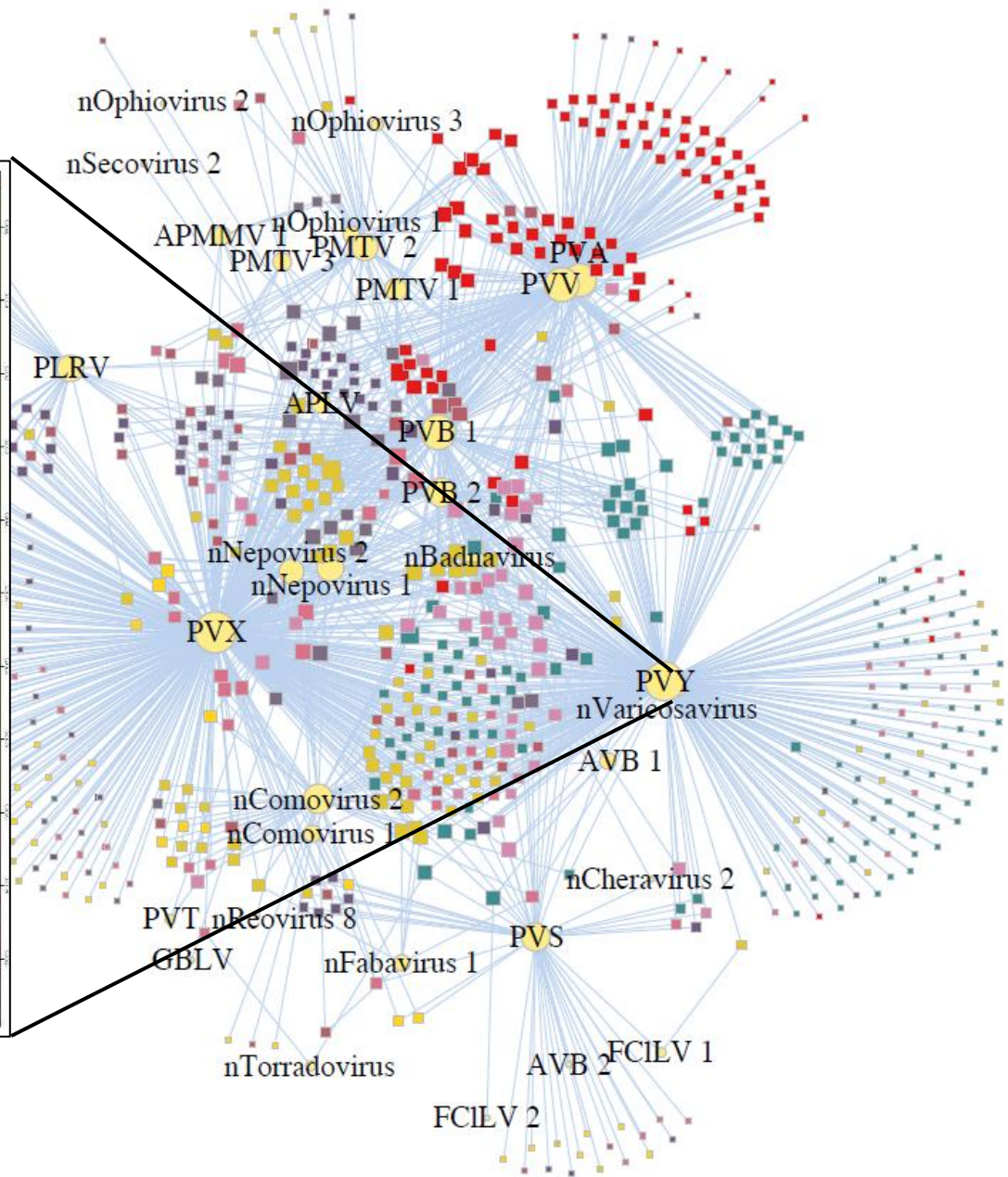
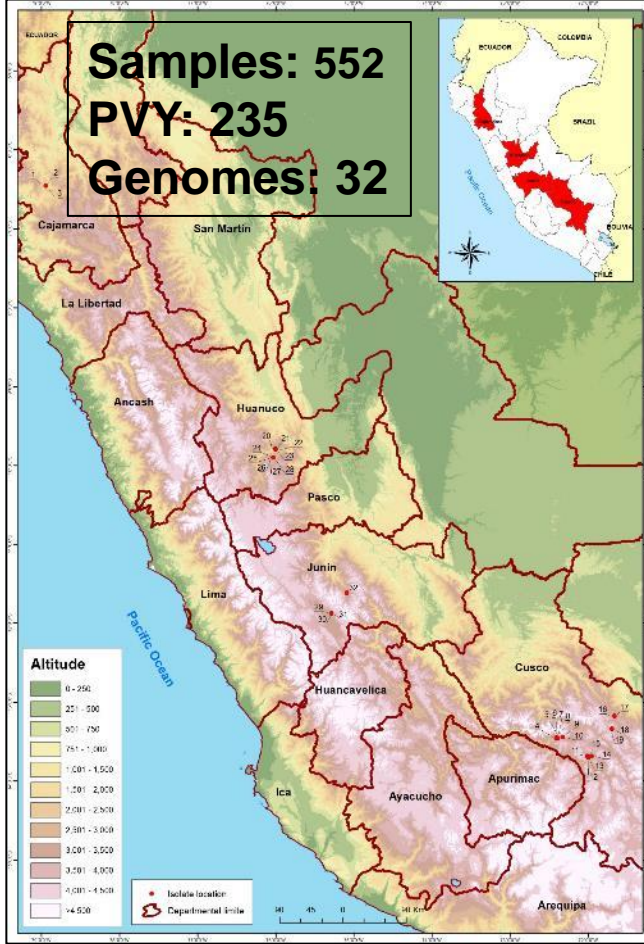
ENG 12:15 PM
INTL 5/21/2018

General results

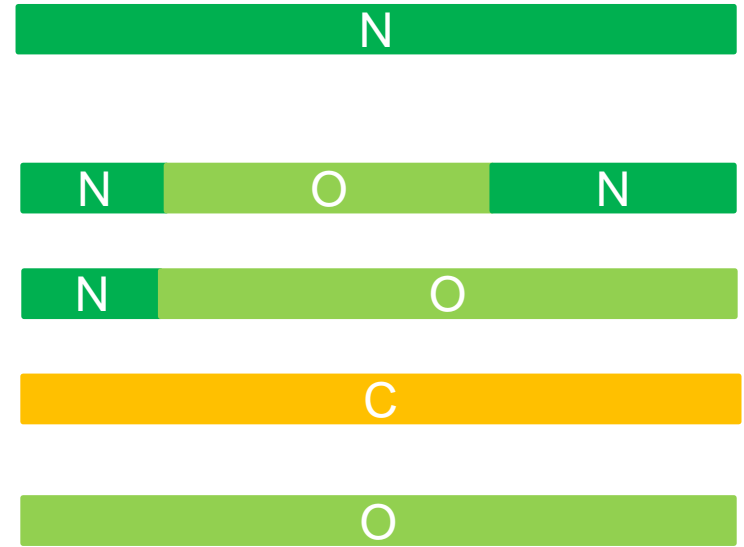
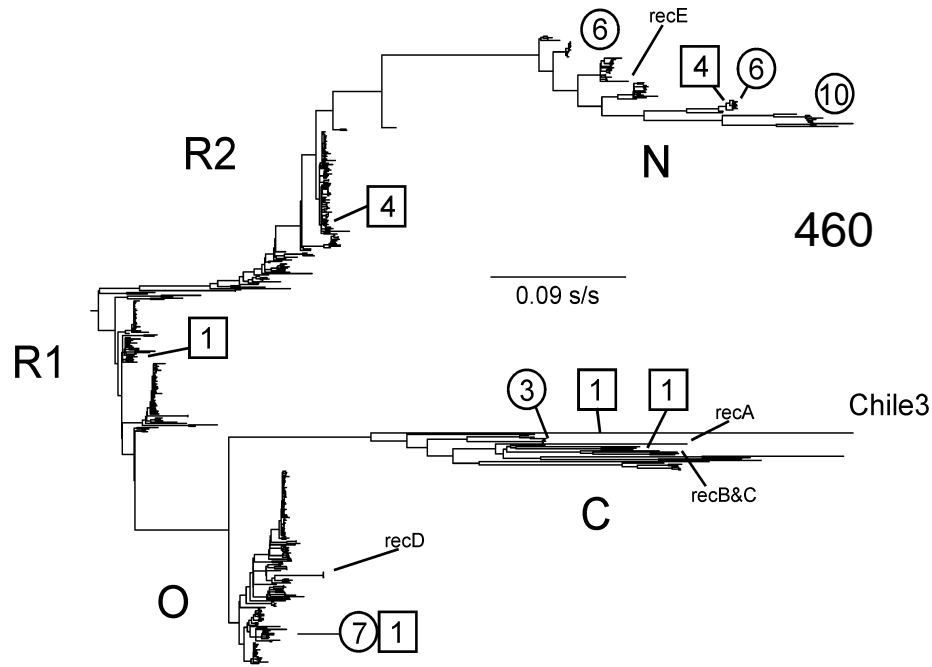


- virus
- Apurimac
- Cajamarca
- Cusco

Samples: 552
PVY: 235
Genomes: 32

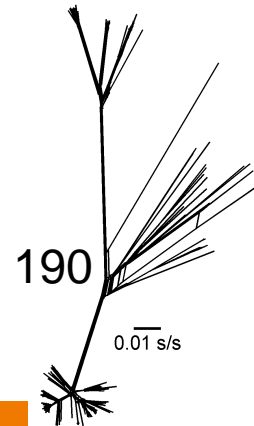
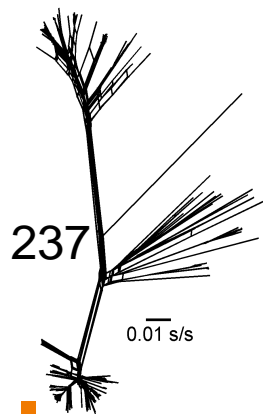
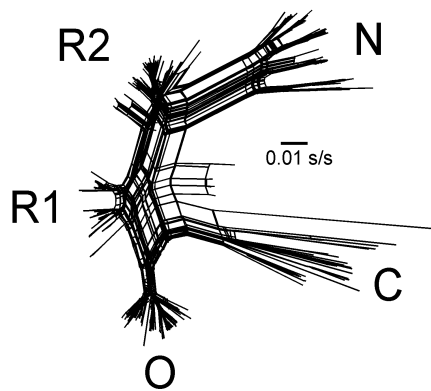


Phylogeny

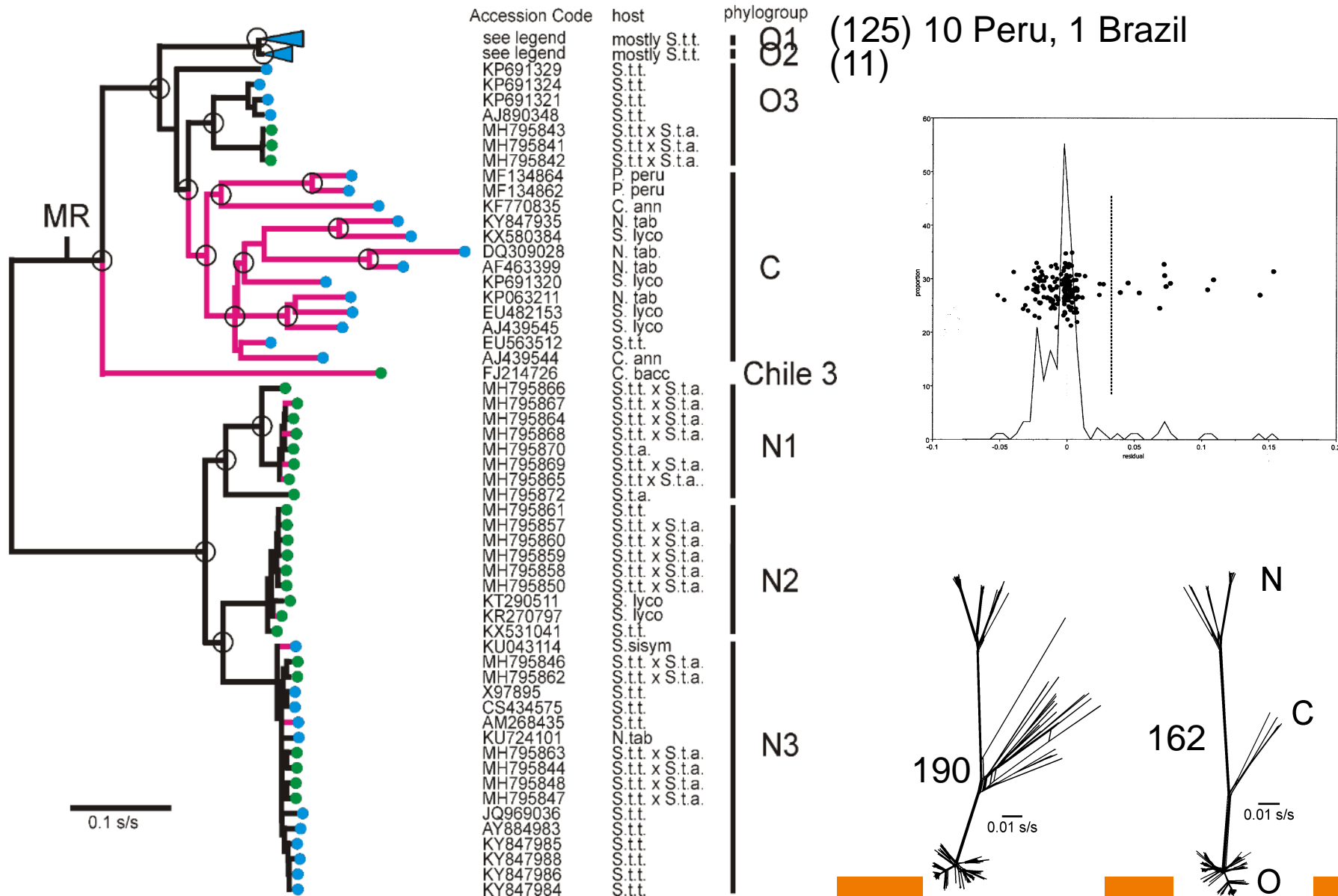


R1&2 removed

RDP-filtered



Identifying temporal signals using TempEst



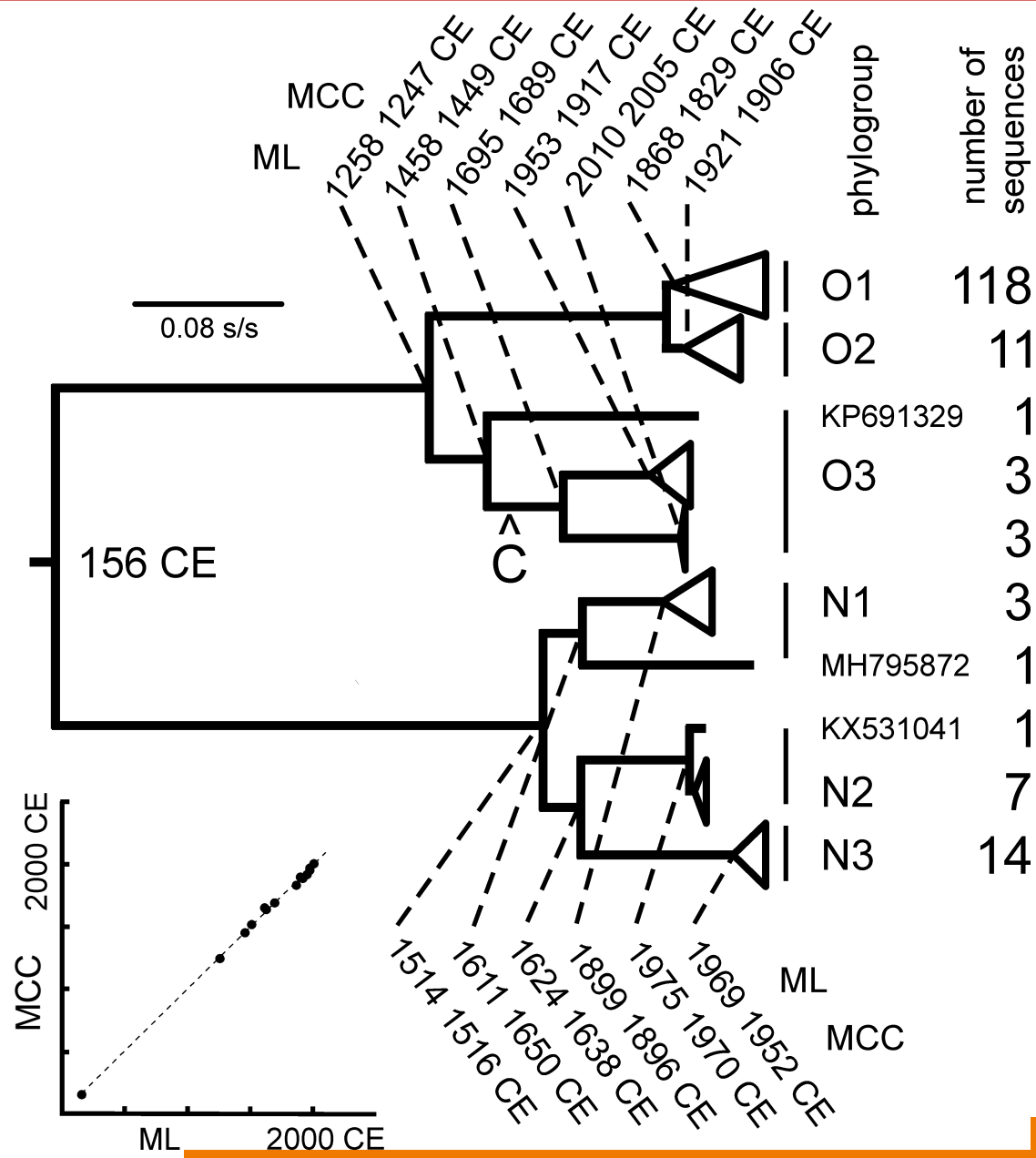
Timescale analysis of ORF sequences using



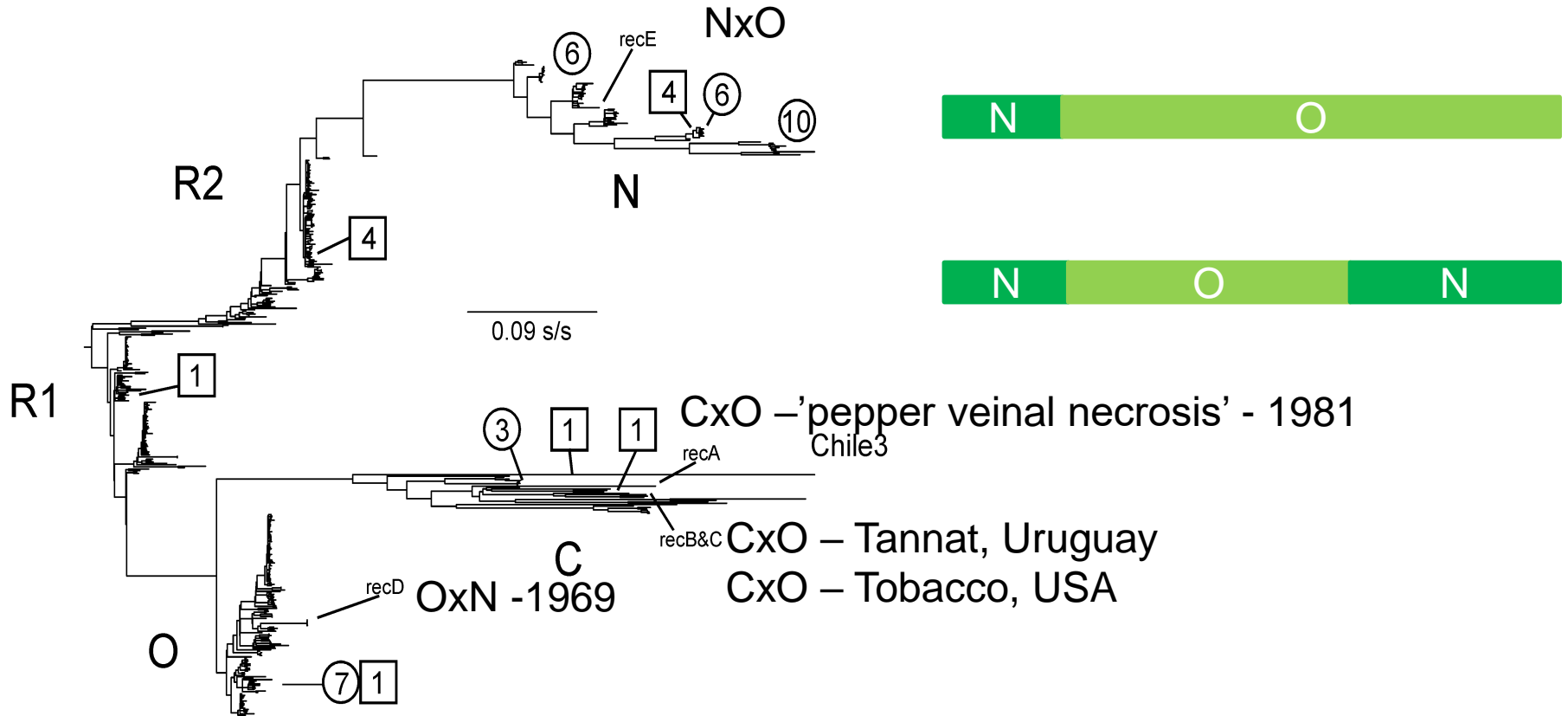
Evaluated twelve combinations of substitution models, three clock and four population growth models

Parameter	Constant population size	Expansion growth	Exponential growth	Bayesian skyline plot
Demographic model	Constant population size	Expansion growth	Exponential growth	Bayesian skyline plot
Path sampling (BF)	141.31	99.47	107.51	<u>179.87</u>
Stepping-stone sampling (BF)	152.94	<u>260.15</u>	110.3	185.96
TMRCA (95% CI)	1873 (1090-2608)	1841 (1157-2622)	1663 (1071-2384)	1879 (1192-2659)
TMRCA effective sample size	239	238	247	261
Substitution rate (nt/site/year)	9.66×10^{-5} (7.10×10^{-5} – 1.23×10^{-4})	9.30×10^{-5} (6.79×10^{-5} – 1.18×10^{-4})	9.89×10^{-5} (7.40×10^{-5} – 1.25×10^{-4})	9.16×10^{-5} (6.90×10^{-5} – 1.15×10^{-4})

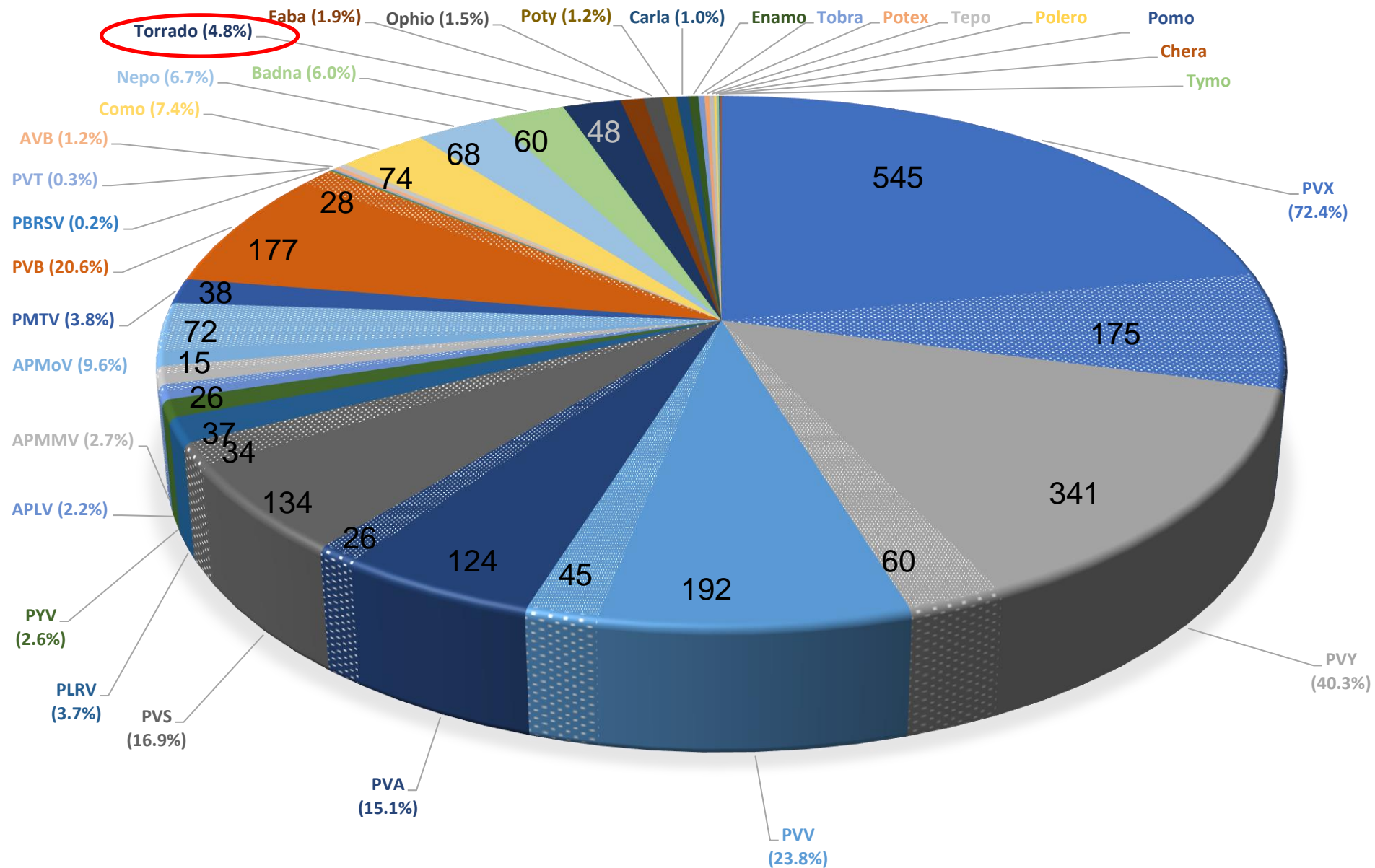
Estimated TMRCA of significant nodes



The recombinants



Are any of the new viruses of relevance?

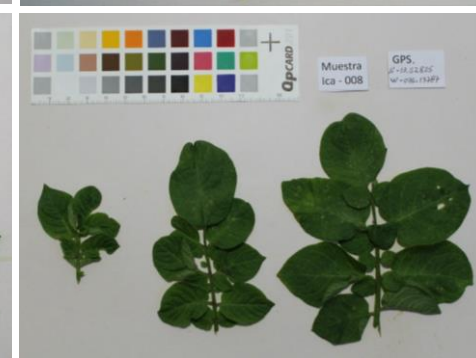
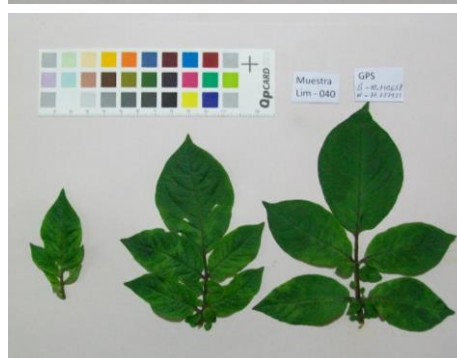
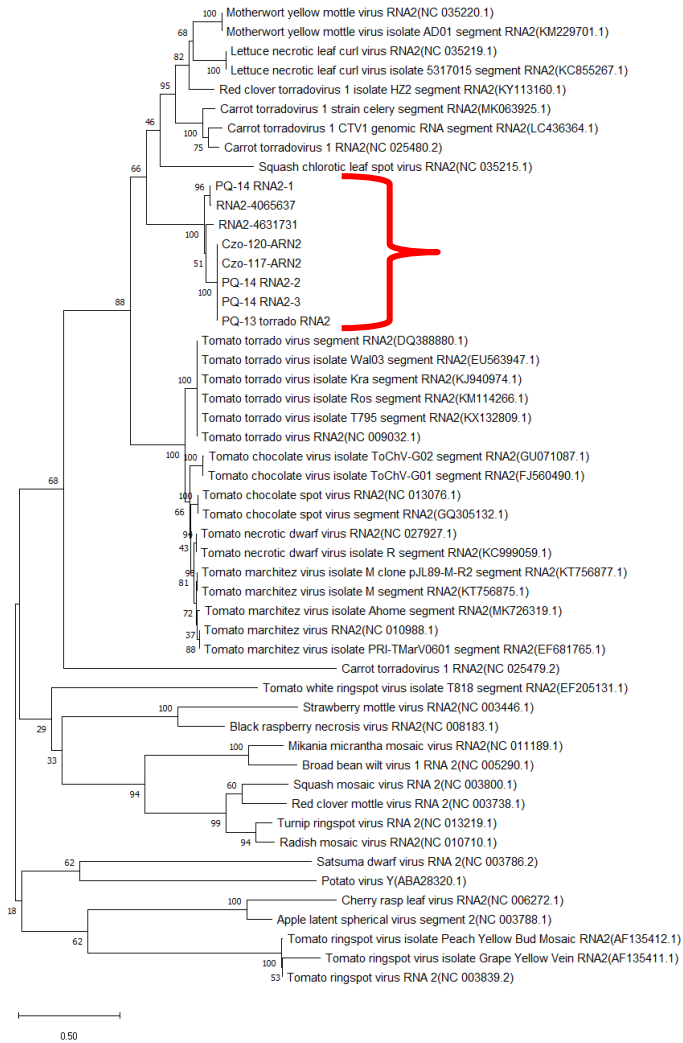


Rugose stunting disease of potato: the history of the emergence and decline of a viral disease

- Observed since 1990 in Arequipa, Moquegua and Tacna and studied as SB26 and SB29
- Isometric virus, transmitted by transmitted by brown leafhoppers (*Russelliana solanicola*) and generated losses of between 35-85% depending on the variety
- Farmers changed their planting dates to manage the disease or quit potato altogether and it is no longer considered a problem
- It was never determined what type of virus it was



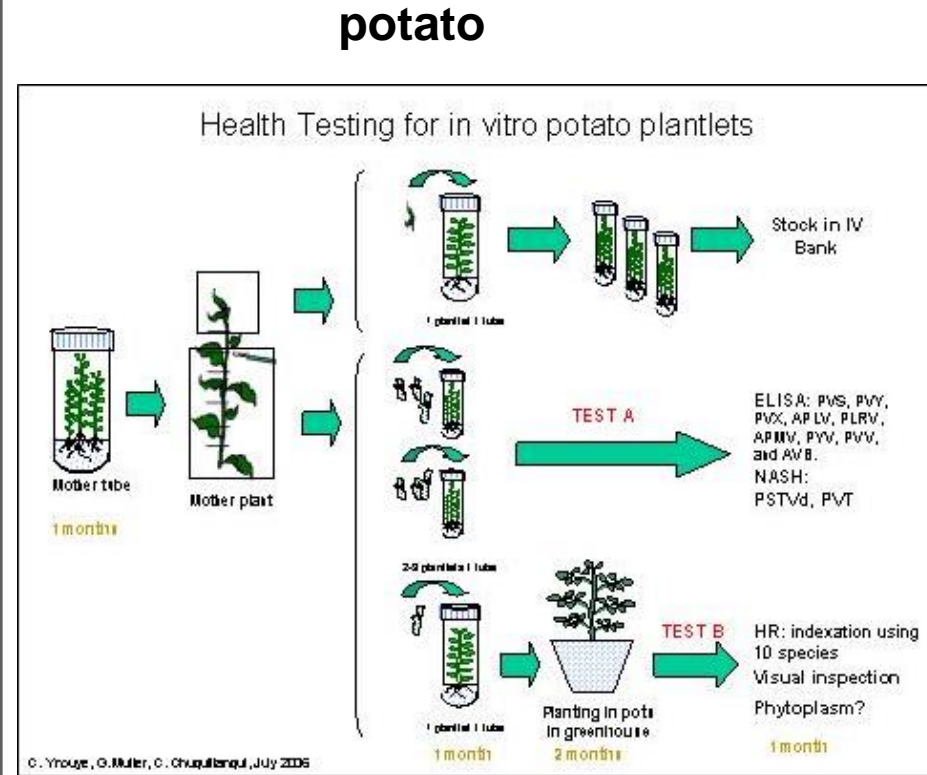
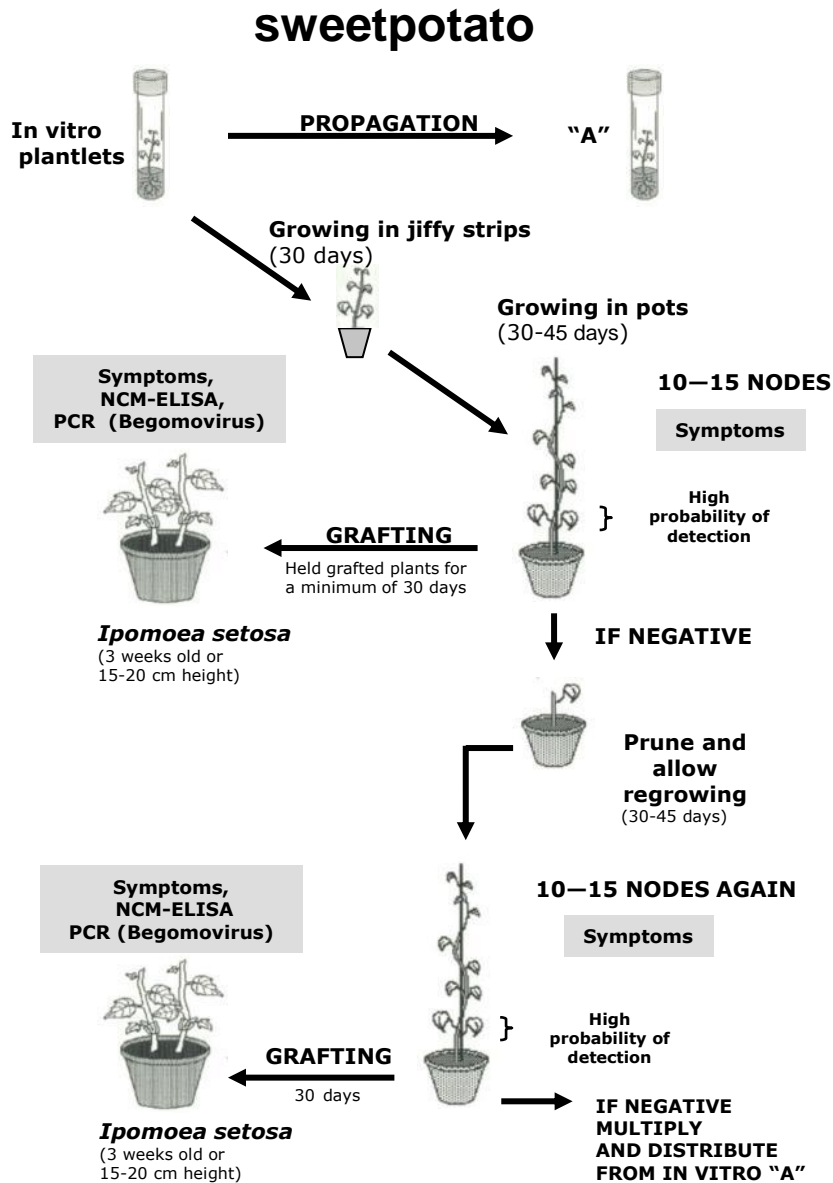
Potato rugose stunting virus is a torradovirus and present in low frequencies in potatoes throughout Peru



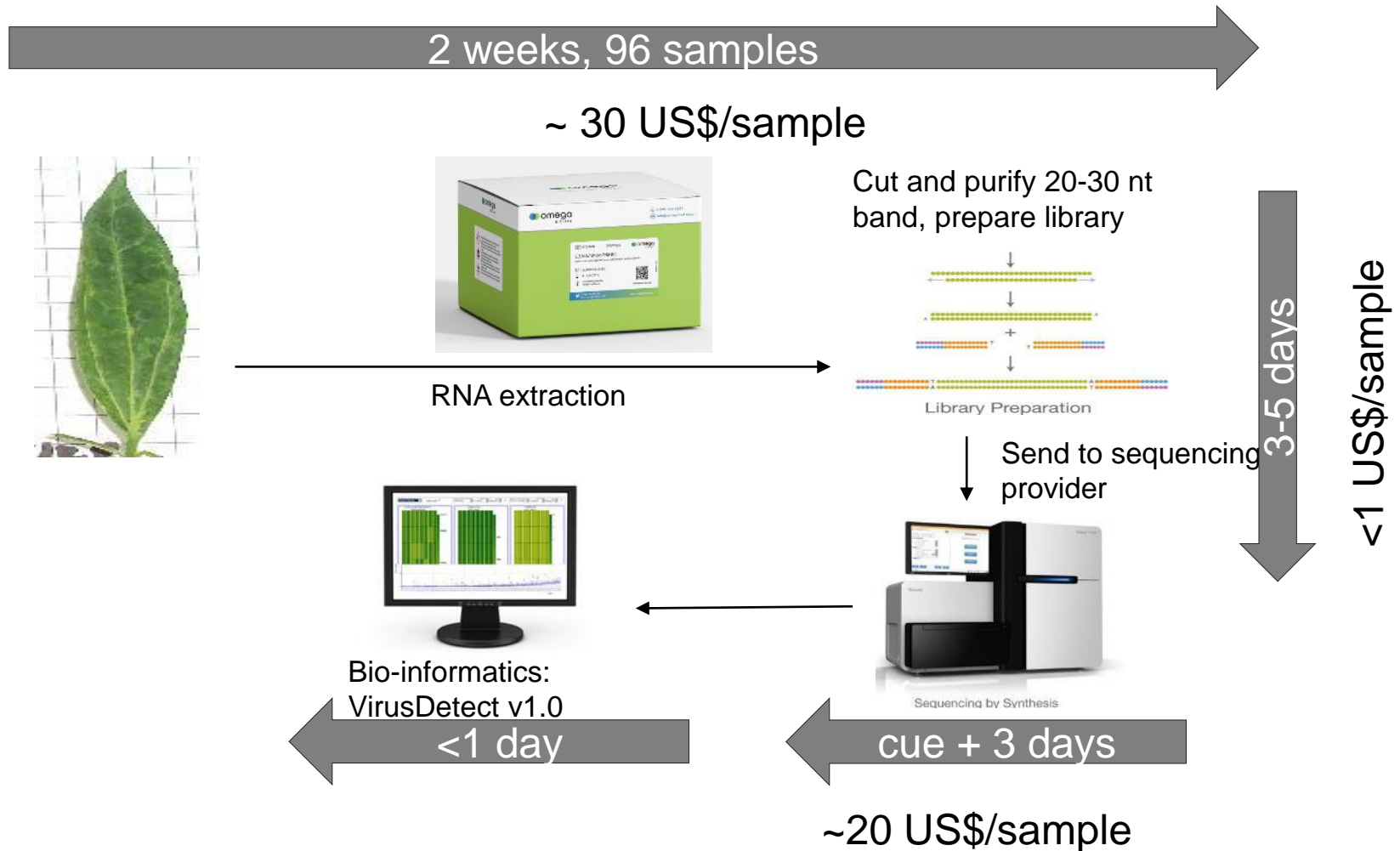
Can it be used for Diagnostics?

- ❖ **Technical challenges**
 - Laboratory protocol?
 - Bioinformatic algorithms?
- ❖ **Performance evaluation**
 - Sensitivity variability?
 - Specificity?
 - Reproducibility?
 - Repeatability?
- ❖ **Routine analysis**
 - Contamination ?

Current indexing process



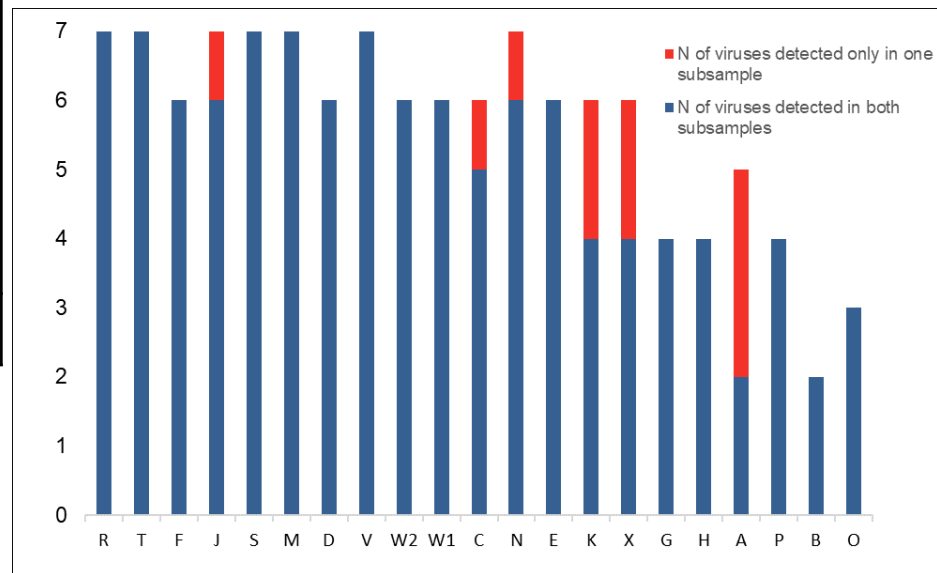
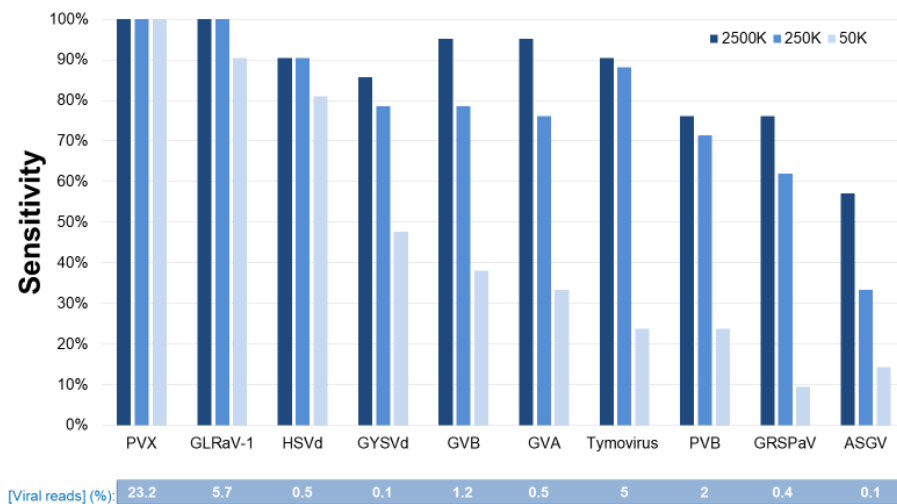
sRSA for virus indexing: How is it done, what does it cost, how long does it take



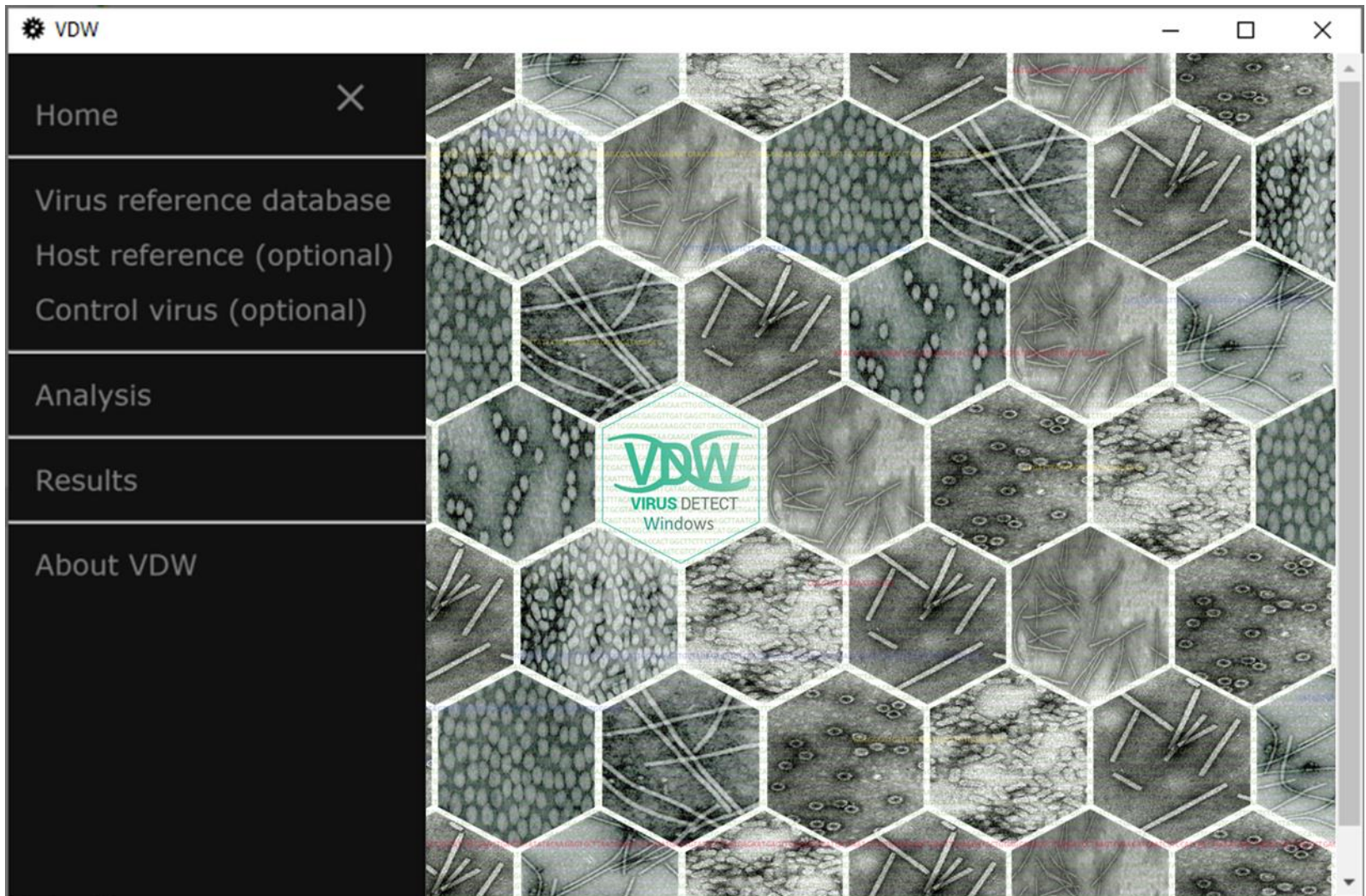
Lab ID	Building the contigs						BLAST comparison		
	Software	Extension step	Minimal contig length (nt)	k-mer range	Removal redundant reads	Host filtering	Method	Database	Cut off used
A	Velvet/Oases then Seqman	N	100	13-21	Y	Y	MegaBLAST + BLASTN	GenBank (nr)	None
B	CLC Genomics Workbench	Y	60	Variable	N	N	BLASTX + BLASTN	GenBank (viruses and viroids)	e-value of 10 ⁻³
C	CLC Genomics Workbench	N	30	16-19	N	N	BLASTN	Local databases of complete virus and viroid genomes mined from Genbank	None
D	Velvet/Assembly Assembler	Y	38	(9-)11-25	N	Y	BLASTX + BLASTN	GenBank (viruses and viroids)	None
E	CLC Genomics Workbench	N	50	15-21	N	N	BLASTX + BLASTN	GenBank (nr + nt)	None
F	Velvet	Y	26	13-17	N	N	BLASTX + BLASTN	GenBank (nr)	e-value of 10 ⁻² Homology cut off >80%, 100% coverage for known viruses identification
G	Velvet	N	N	17	N	N	BLASTN +TBLASTX of non-ID contigs	Local databases (viruses and viroids; ribosomal RNA; host) and GenBank (nr) for TBLASTX	Sequencing depth >5
H	Velvet/Assembly Assembler	N	21	7-21	N	N	BLASTX + BLASTN	Genbank (nr)	None
J	Velvet	Y (Only the 2,5M depth)	40	13-15-17	N	N	BLASTN/X/P	GenBank (nr)	Default parameters, e-value 10 ⁻¹
K	Mapping and then de novo assembly with Velvet	N	31	13-15-17	Y	N	MegaBLAST / BLASTN	GenBank (viruses and viroids)	e-value 10 ⁻¹⁰ Homology >95%
M	Velvet	N	29	15-17	N	N	BLASTX + BLASTN	GenBank (viruses and viroids)	e-value 10 ⁻⁴ (BLASTX) e-value 10 ⁻⁶ (BLASTN)
N	Velvet and BWA backtrack	N	50	11-19	Y	N	BLASTX + BLASTN	GenBank (viruses and viroids)	Bit score >= 30
O	CLC Genomics Workbench	N	50	16-19	N	N	BLASTN	GenBank (nt)	e-value 10 ⁻³ Homology >85%
P	Velvet/Assembly Assembler	N	21	9-31	N	N	BLASTX + BLASTN	GenBank (nr + nt)	e-value 10 ⁻³
R	Mapping against refseqDB, and de novo assembly, both in CLC Genomics Workbench	Y	21	12	N	Y	BLASTX + BLASTN	GenBank (nr)	None
S	VirusDetect	N	40	9-19	N	Y	BLASTX + BLASTN	GenBank (nt + nr)	Sequencing depth >5; reference genome coverage >10%; e-value 10 ⁻⁵
T	Blasting raw reads against Genbank (nt) and de novo assembly with CLC Genomics Workbench	N	50	17	N	N	BLASTN	Genbank (nt) + Refseq virus and viroids	e-value 10 ⁻⁴
V	AByss	N	16	16	N	N	BLASTN	GenBank (nt)	e-value 10 ⁻⁵
W1	CLC Genomics Workbench	N	60	14	N	N	BLASTX	GenBank (viruses and viroids)	e-value 10 ⁻³
W2	CLC Genomics Workbench	N	60	Variable	N	N	BLASTX	GenBank (viruses and viroids)	e-value 10 ⁻³
X	CLC Genomics Workbench	Y	21	17	N	N	BLASTN + BLASTX	GenBank virus + viroid	e-value 10 ⁻³

Bio-informatics pipelines: How consistent are they?

LAB ID	SENSITIVITY				FALSE DISCOVERY RATE		
	2,500,000	250,000	50,000	AVERAGE	RATE		
					2,500,000	250,000	50,000
A	90%	53%	10%	51%	0%	0%	0%
B	80%	35%	30%	48%	0%	0%	0%
C	80%	71%	60%	70%	0%	0%	0%
D	100%	82%	50%	77%	17%	7%	9%
E	80%	82%	30%	64%	0%	0%	0%
F	100%	88%	80%	89%	0%	0%	0%
G	100%	53%	20%	58%	0%	0%	0%
H	70%	65%	30%	55%	0%	0%	0%
J	100%	94%	70%	88%	0%	0%	9%
K	90%	71%	40%	67%	0%	0%	0%
M	90%	94%	50%	78%	0%	6%	18%
N	90%	82%	30%	67%	0%	0%	0%
O	40%	41%	20%	34%	0%	0%	0%
P	70%	59%	20%	50%	0%	0%	0%
R	100%	100%	100%	100%	9%	6%	9%
S	100%	100%	50%	83%	0%	0%	0%
T	100%	100%	90%	97%	0%	0%	0%
V	80%	88%	60%	76%	0%	0%	0%
W1	90%	82%	40%	71%	0%	0%	0%
W2	90%	82%	60%	77%	0%	0%	0%
X	80%	71%	30%	60%	0%	8%	27%
AVERAGE	87%	76%	46%	70%	GLOBAL FDR RATE: 1.9%		



VirusDetect-Windows





The International Potato Center (known by its Spanish acronym CIP) is a research-for-development organization with a focus on potato, sweetpotato, and Andean roots and tubers. CIP is dedicated to delivering sustainable science-based solutions to the pressing world issues of hunger, poverty, gender equity, climate change and the preservation of our Earth's fragile biodiversity and natural resources.

www.cipotato.org



CIP is a member of CGIAR

CGIAR is a global agriculture research partnership for a food secure future. Its science is carried out by the 15 research centers who are members of the CGIAR Consortium in collaboration with hundreds of partner organizations.

www.cgiar.org