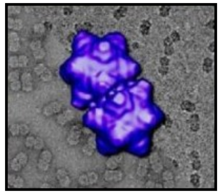
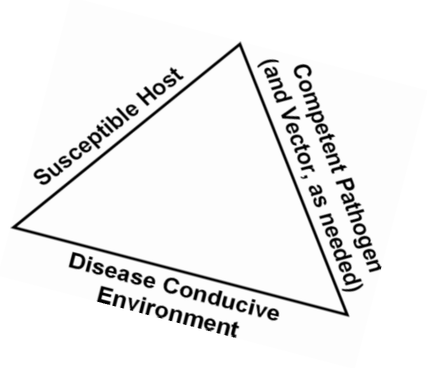


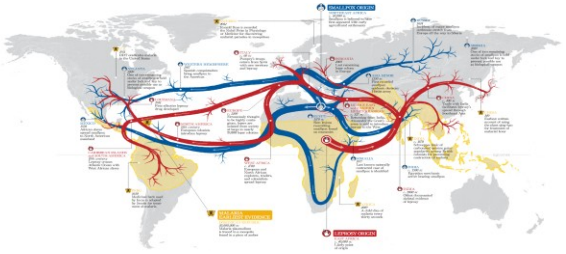
Genomic pathology and molecular diagnostics for tracking the status of a rapidly evolving geminiviruses in cotton-vegetable systems

Judith K. Brown
School of Plant Sciences
University of Arizona, Tucson AZ USA



Susceptible Host
Competent Pathogen (and Vector, as needed)
Disease Conducive Environment

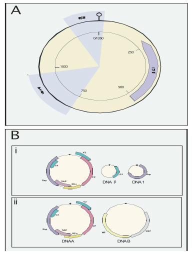
7th ACRDN
Nagpur, Maharashtra
Sept 15-17, 2017



Distribution of cotton-infecting begomoviruses: whitefly-transmitted ssDNA genomes

mosaics

New World: low diversity ***Cotton leaf crumple virus***:
AZ, CA, TX, Mexico, Guatemala, Caribbean, Brazil (93-100%)
(Idris and Brown, in prep)



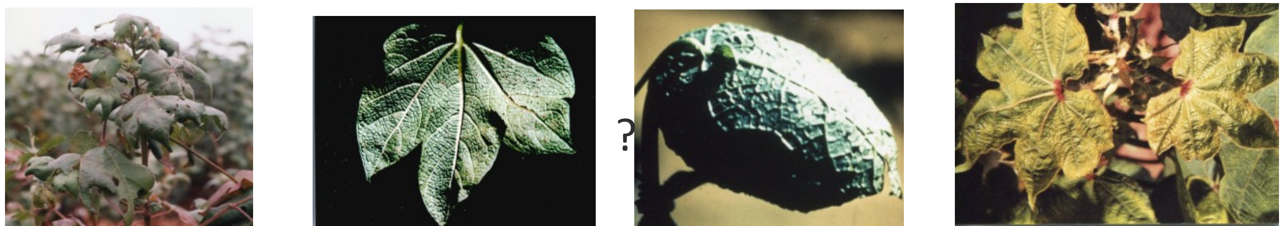
Asia: extreme diversity
Cotton leaf curl virus
complex (8+ species?)
Often-same beta, different alphasats

Leaf curl &
enations

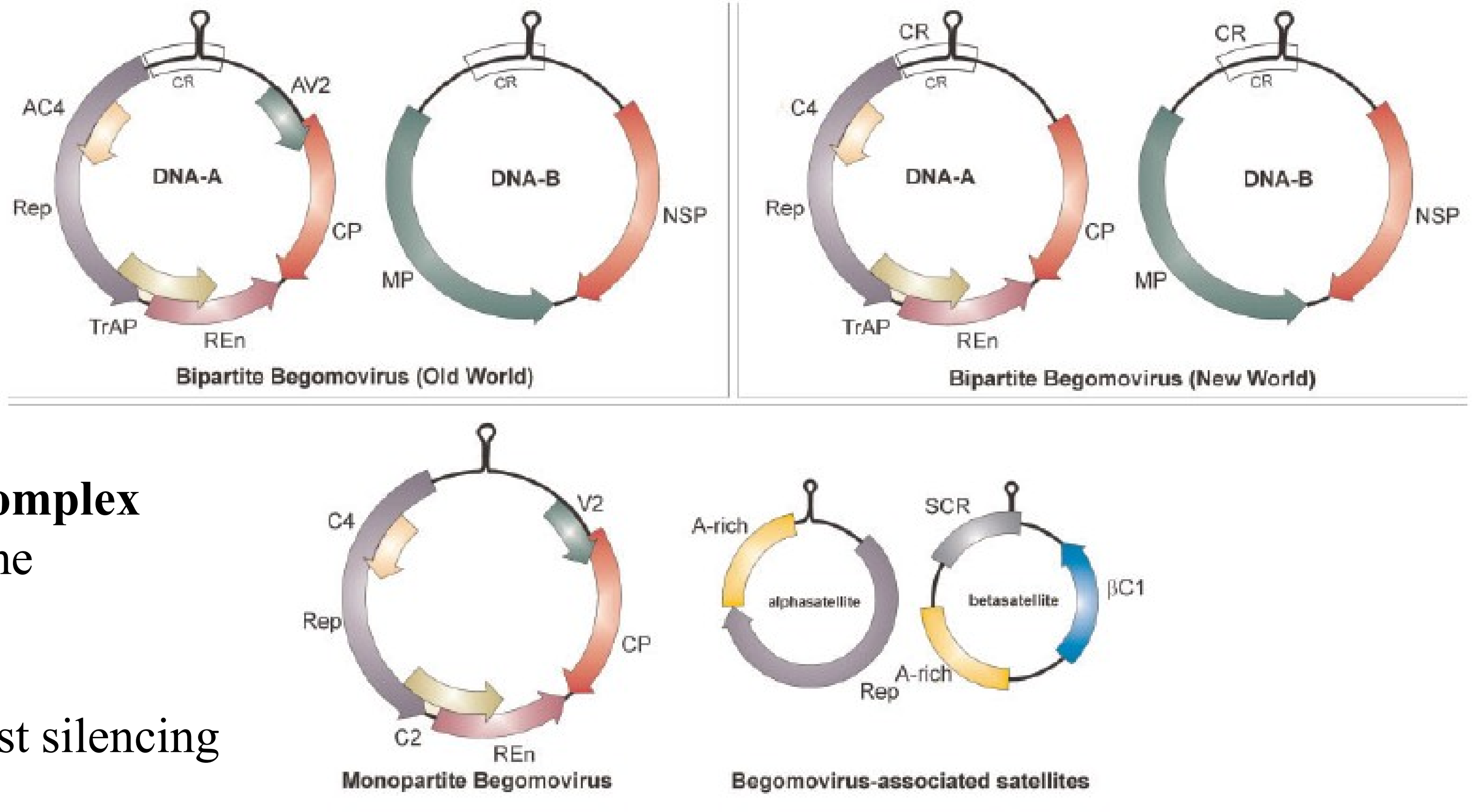
Africa, Arabian Peninsula

Cotton leaf curl Gezira virus: moderate diversity
Sudan cotton, Sida (Idris & Brown)
Burkina Faso okra (Tiendrébéogo, 2010)
Arabian Peninsula: CLCuGV-okra
TomatoL TYLCV-OM recomb (Asia-Oman) with
Gezira alphasatellite (CLCuGA)
Diverse betasats 88-98%, alphasats 89%

Leaf curl
Vein thickening



Genome organizations - Genus, *Begomovirus*: ssDNA



Cotton leaf
crumple virus

USA, Mexico
Central America

Cotton leaf curl complex
Monopartite genome
–'helper virus'
Betasatellite
– suppressor of host silencing
Alphasatellite
-modulation of virulence ?

Sattar et al., 2013. J Gen Virol 94: 695-710.

Cotton leaf curl disease pandemic timeline

1967

Sporadic isolated
infections of cotton

First report CLCuD
Pakistan Punjab

1988

Introduction of high
yielding
cotton variety (S12)
Jassid control,
but whitefly susceptible

CLCuMuV + CLCuMB
first epidemic

$5 billion losses

1992-1997

Introduction of resistant varieties 1998-99
Widespread cultivation
Narrow germplasm base

1997-2001

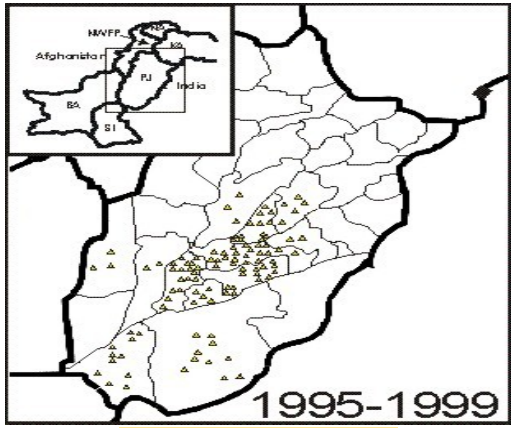
second epidemic:
Recombinant CLuKoV-Bur
+ CLCuMB

Resistance
Breaking 2000-2003

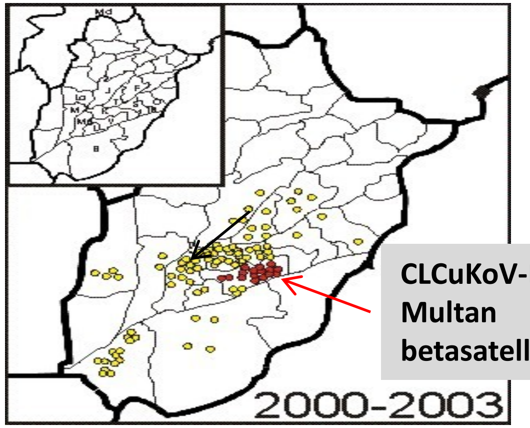
(Amrao et al., 2010)

2004-2009
Spread to all cotton areas of
Pakistan; Identified in northern India

2017



Multan betasat
+ CLCuMV

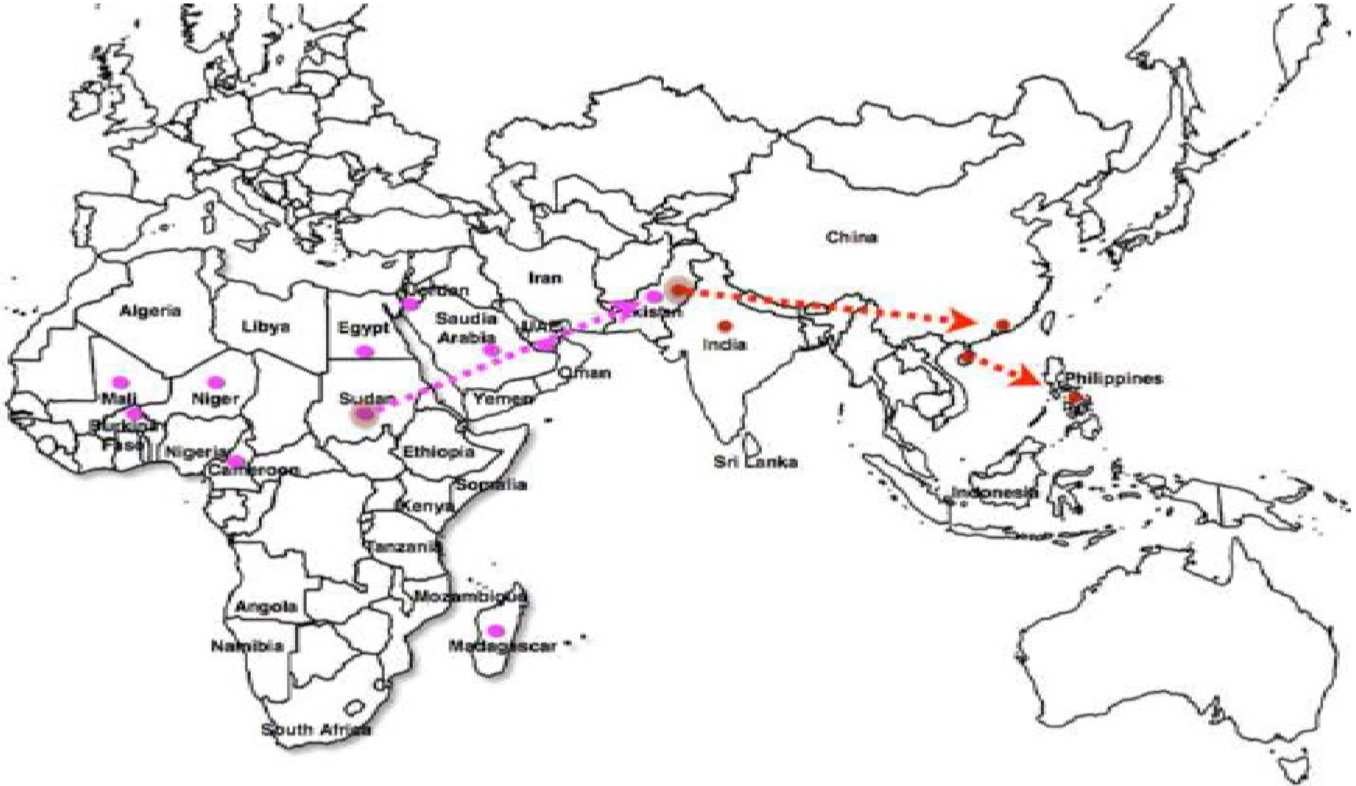


CLCuKoV-Bur +
Multan
betasatellite

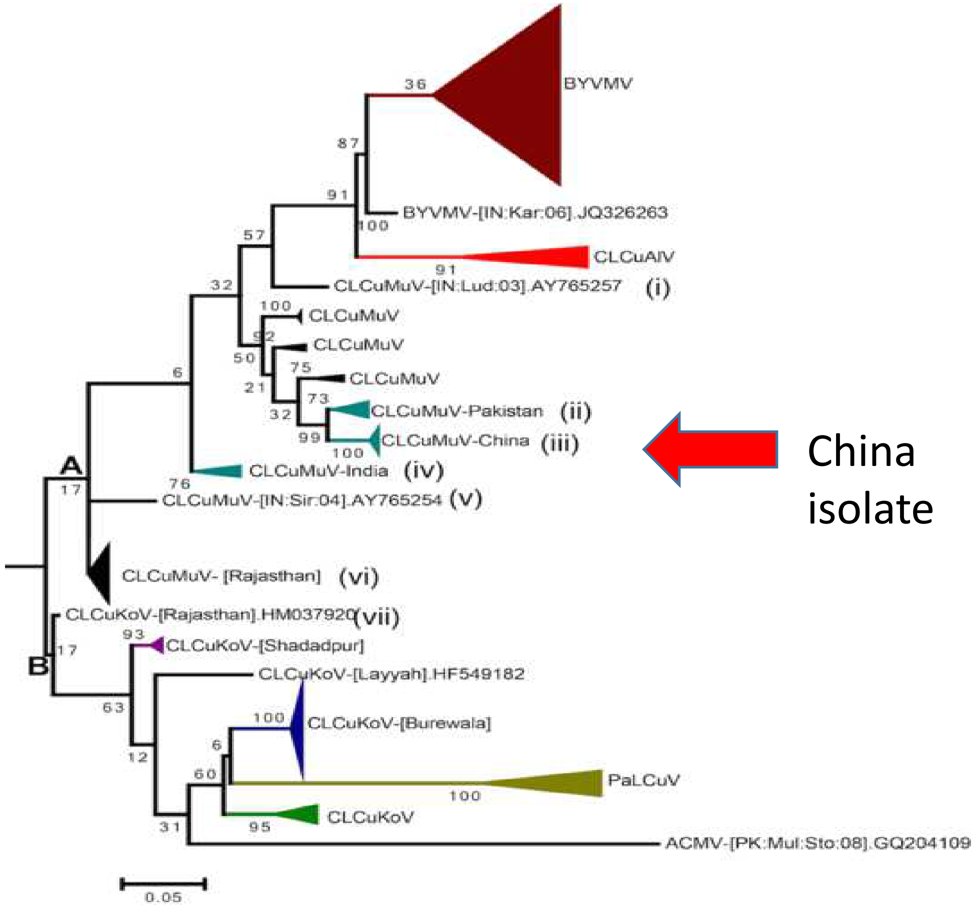
Initial ‘red’ zone of outbreak
of resistance-breaking recombinant

(modified; Courtesy R. Briddon)

Cotton leaf curl Multan virus spread to China *via* ornamentals spp. 2008 Philippines – Hibiscus 2014



Cotton leaf curl Gezira virus from Africa introduced into Pakistan 2011



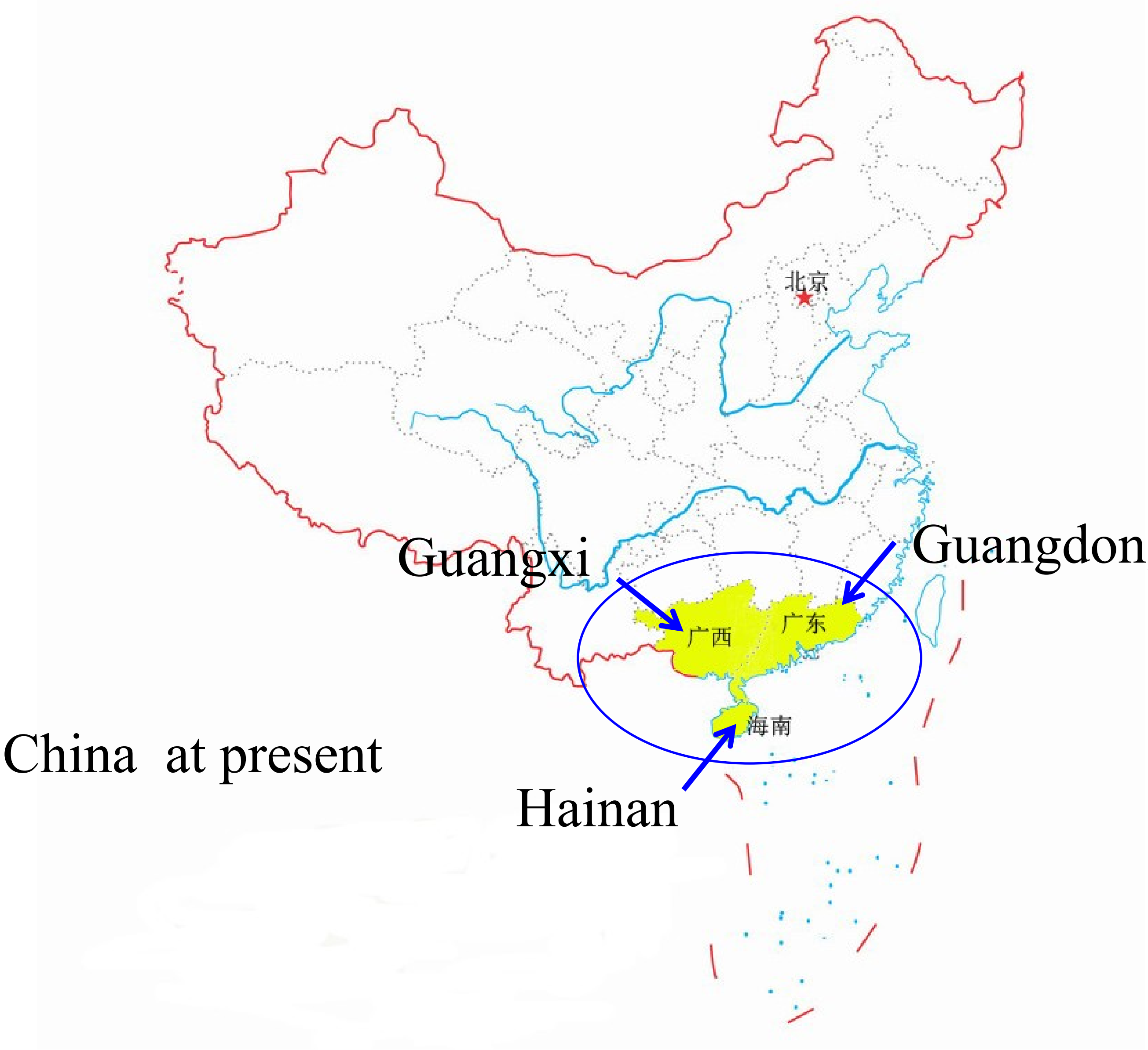
China isolate

Saleem et al 2016

Begomoviruses of cotton in China

Dr. Zifu He
Plant Protection Institute
Guangdong Academy of Agricultural Sciences
Guangzhou, China

CLCuMV distribution in China-2012



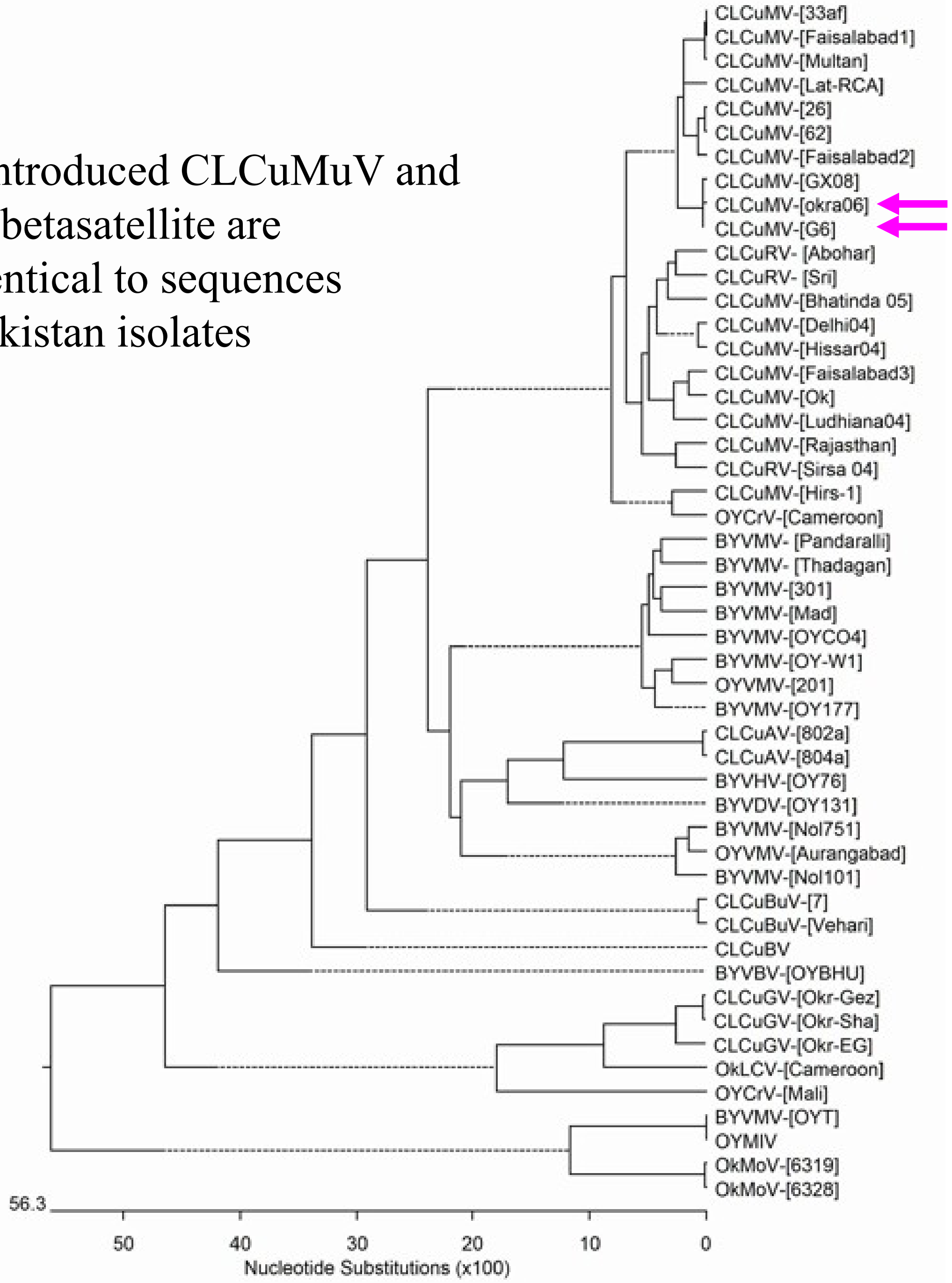
Guangxi

Guangdong

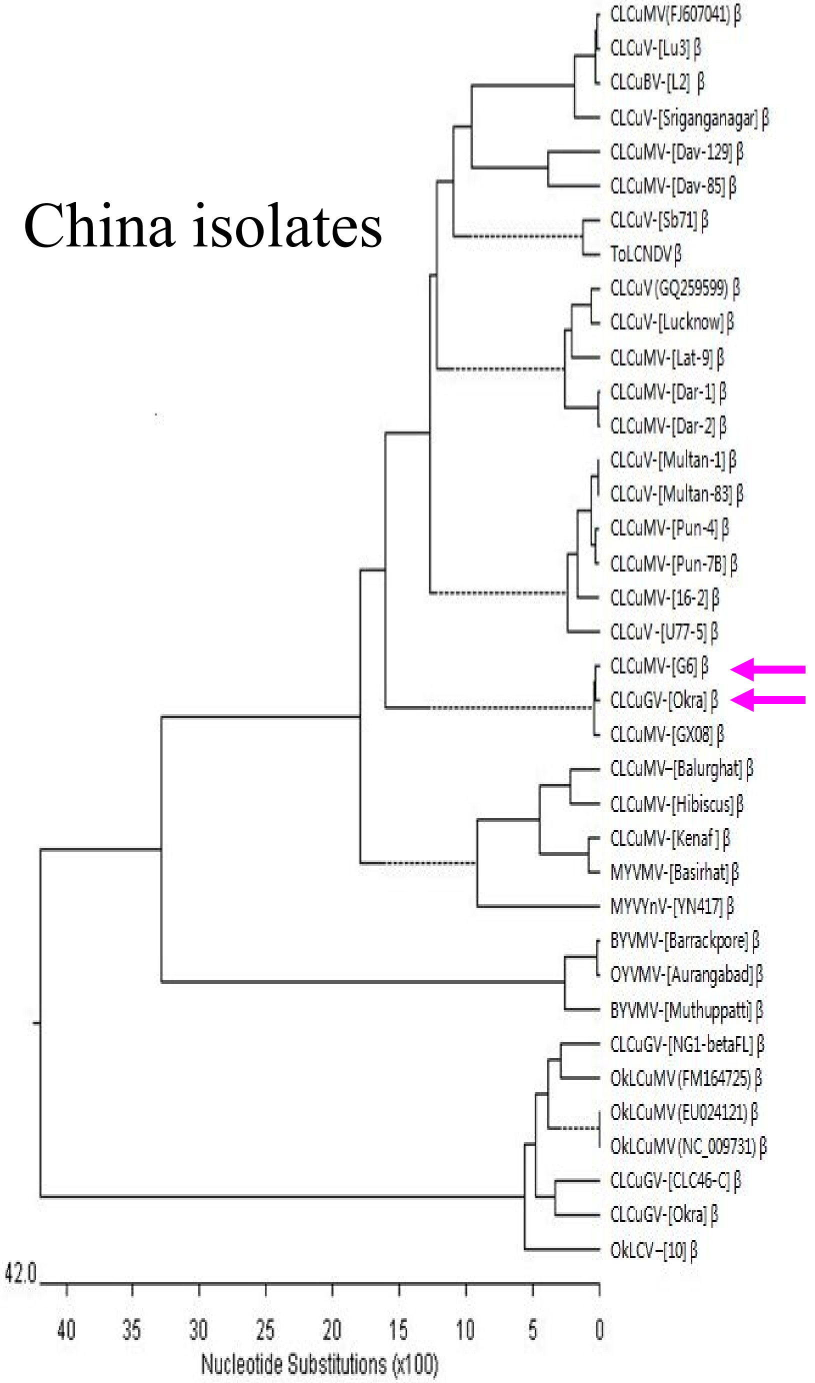
Hainan

Only in South China at present

China-introduced CLCuMuV and ClCuM betasatellite are 99% identical to sequences from Pakistan isolates



China isolates



CLCuMV in China - multiple isolates/hosts

--Nucleotide identities of DNA-A and betasatellite of 13 isolates from Guangdong, Guangxi and Hainan provinces of China, shared at >99%

--Nucleotide identities of DNA-A and beta satellite of isolates from four hosts (*Hibiscus rosa-sinensis*, cotton, okra and *Malvaviscus arboreus*) in China more than 99%

--Suggests one or few introductions at about the same time

--Virus has not diversified since the introduction

--Possibility for recombination with endemic viruses = new variants, spread ?

Tracking the viruses …using
molecular and genomic pathology tools
to guide diagnostics development
Epidemiological studies
Population analysis

Genomic Surveillance

•**Classical epidemiological contexts promote development of tools for early pathogen/pest detection** – until recently, have faced major constraints – needed to know 'what' we are looking for when direct methods of discovery were unavailable

•'**Genomic surveillance**' has become a tractable method for detection of infectious disease agents /vectors – Next-generation sequencing platforms

•Growing awareness that virus diversity drives/promotes evolution of plant hosts and insect vector

•**Goals of genomic surveillance**:
-identify host-determinant markers that inform ability to trace the 'origin of transmission' in space /temporally
-implement markers to guide prevention and/or arrested spread at or near source.
-allow actions to be taken at- source, prior to extensive spread
-keeping good long term records is an essential resource for making epidemiological inferences that can guide management

Field samples – spatial distribution and natural host range 2011-15

Analysis included 452 geminiviruses and 476 betasatellites sequences for frequency and distribution.

‘Historic sequence data’ – temporal, spatial distribution

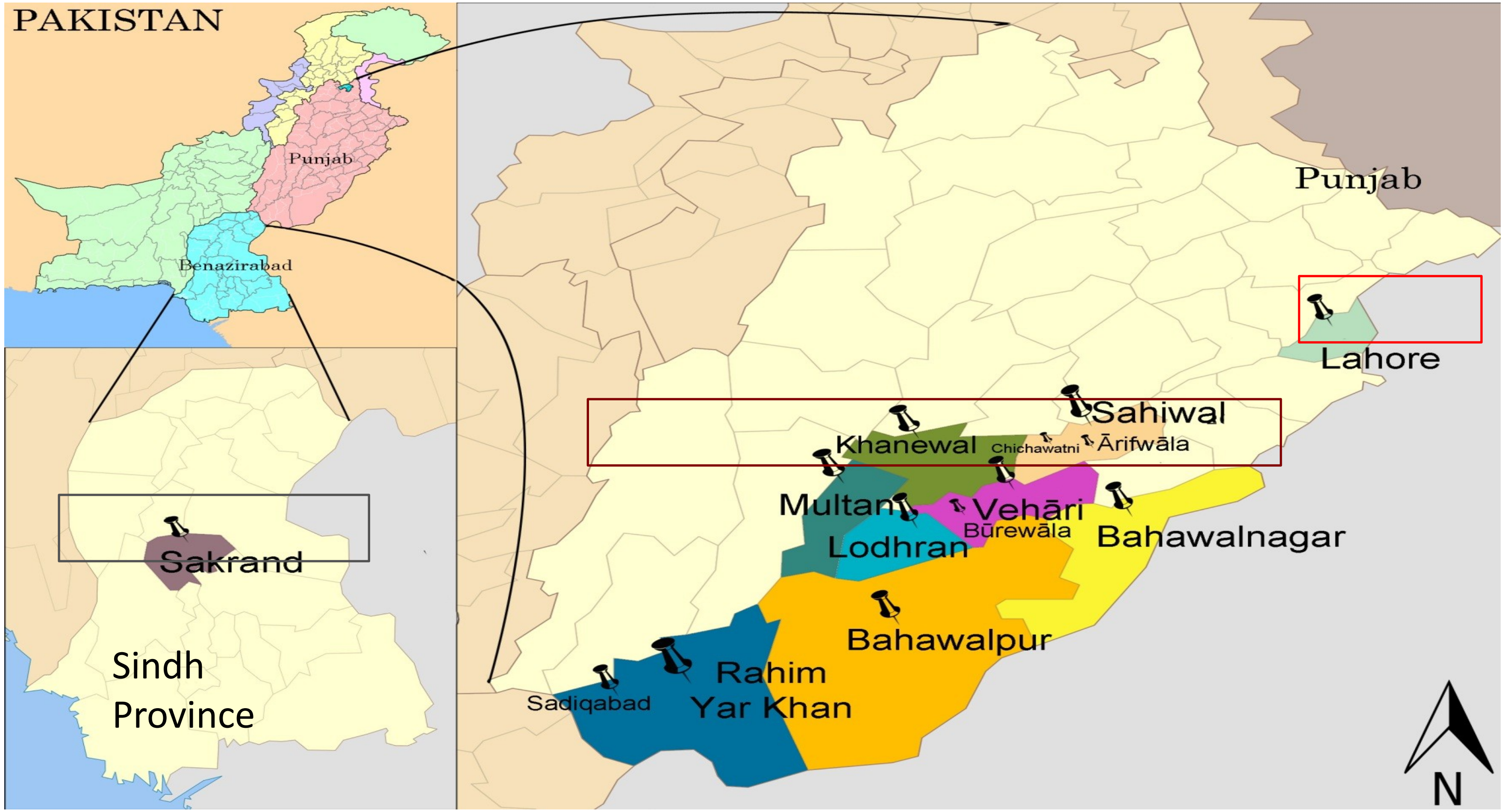
Downloaded all publicly available CLCuD-associated begomovirus and besatellite sequences in GenBank. Eliminated haplotypes (>1 nt).

Sample collection sites – cotton, vegetables, wild spp

Cotton –predominantly grown in the **Punjab** (tropical, wet) and **Sindh** Provinces.

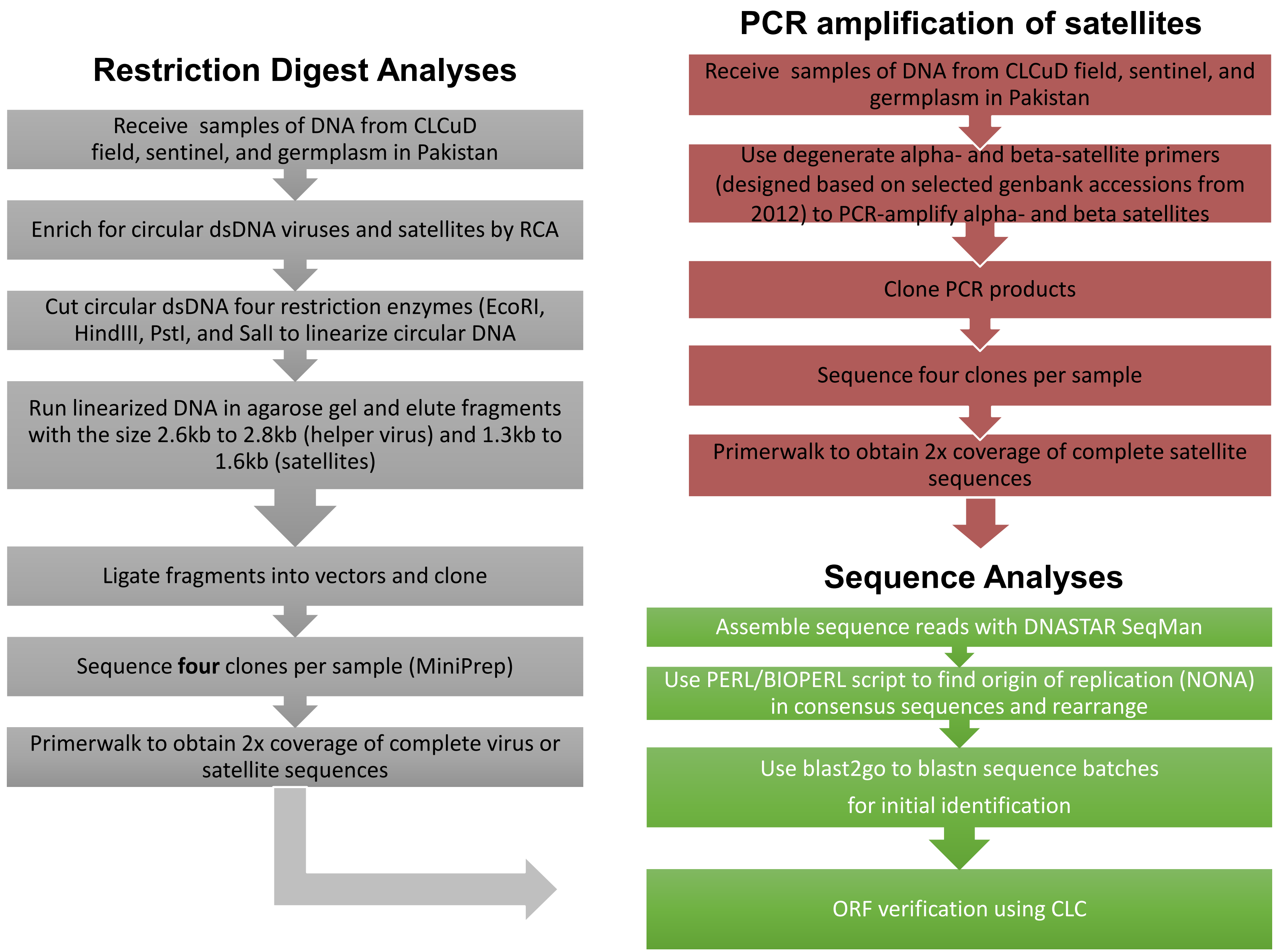
Districts leading in cotton production:
southern Punjab: Rahim Yar Khan, Bahawalpur, Bahawal nagar, Multan, Dera Ghazi Khan, Muzaffargarh, Vehari, Khanewal.

-Sindh Sanghar and Sakrand important producion districts.



In the Province of **Lahore** sugarcane and wheat, and some vegetables predominate; urban interface brings ornamentals and vegetable gardens.

Pipeline: 2 approaches rely on *a priori* knowledge of what is ‘known’ to be present



Restriction Digest Analyses

- Receive samples of DNA from CLCuD field, sentinel, and germplasm in Pakistan
- Enrich for circular dsDNA viruses and satellites by RCA
- Cut circular dsDNA four restriction enzymes (EcoRI, HindIII, PstI, and SalI to linearize circular DNA
- Run linearized DNA in agarose gel and elute fragments with the size 2.6kb to 2.8kb (helper virus) and 1.3kb to 1.6kb (satellites)
- Ligate fragments into vectors and clone
- Sequence **four** clones per sample (MiniPrep)
- Primerwalk to obtain 2x coverage of complete virus or satellite sequences

PCR amplification of satellites

- Receive samples of DNA from CLCuD field, sentinel, and germplasm in Pakistan
- Use degenerate alpha- and beta-satellite primers (designed based on selected genbank accessions from 2012) to PCR-amplify alpha- and beta satellites
- Clone PCR products
- Sequence four clones per sample
- Primerwalk to obtain 2x coverage of complete satellite sequences

Sequence Analyses

- Assemble sequence reads with DNASTAR SeqMan
- Use PERL/BIOPERL script to find origin of replication (NONA) in consensus sequences and rearrange
- Use blast2go to blastn sequence batches for initial identification
- ORF verification using CLC

Illumina pipeline – Next Gen Sequencing = ‘discovery’ – no ‘*a priori*’ knowledge
Proof of principle – negative detection by both ’*a priori*’ approaches

1

- **Field samples from pipeline selected for illumina sequencing field collections 2015**
- **no helper virus seqs from PCR pipeline (only beta/alphas)**

2

- **Plant samples -DNA amplified by RCA to enrich for circular dsDNA viruses and satellites.**
- **RCA products - Illumina HiSeq 2500 sequencing**

3

- **Illumina data assembled *de novo***
- **DNASTAR SeqMan NGen 12.0**

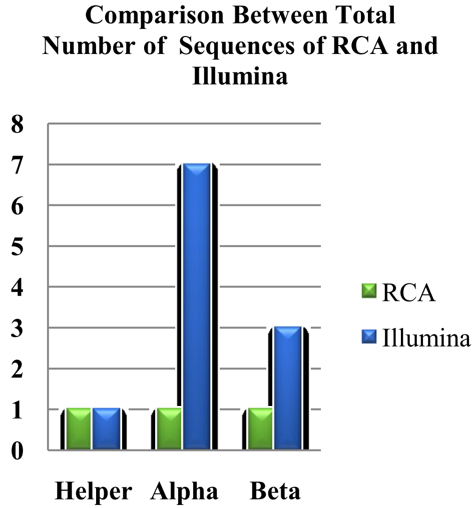
4

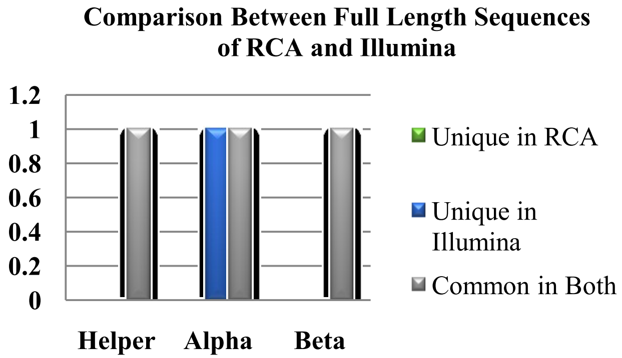
- **Assembled output annotated *in silico w/* Blast2GO 2.8. Sequences identified as begomoviruses/satellites, rearranged, first nt in *ori***

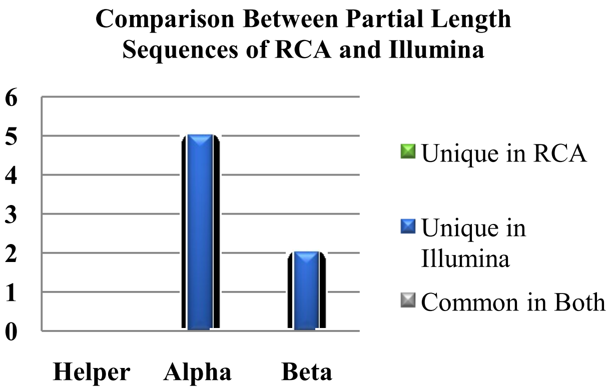
Sample 127:

Host plant: Cotton

Location: Rahim Yar Khan, Punjab


Comparison Between Total Number of Sequences of RCA and Illumina


Comparison Between Full Length Sequences of RCA and Illumina


Comparison Between Partial Length Sequences of RCA and Illumina

Illumina confirmed 1 full length helper, 1 full length alpha, and 1 full length beta that were sequenced by RCA. In addition, Illumina sequenced 5 new partial alphas, 1 new full length alpha, and 2 new partial betas.

Blast Results

From RCA Pipeline

	Sequence Description	HA	EV	S	SL	AL
FL	*Cotton leaf curl Burewala virus complete genome*	FR750321	0	99	2763	2766
FL	*Cotton leaf curl Multan betasatellite*	EU384601	0	99	1561	763
FL	*Gossypium darwinii symptomless alphasatellite*	EU384646	0	97	1353	1373

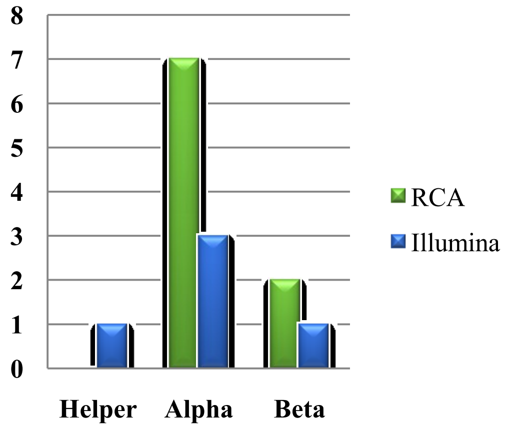
Illumina Sequencing

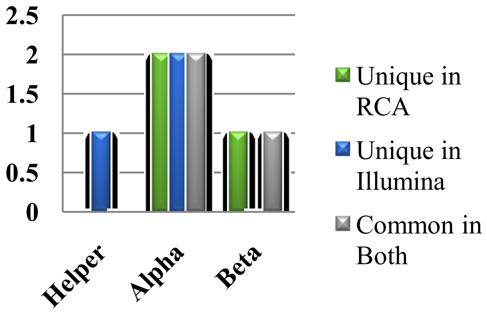
	Best Hit	MS	TS	QC	EV	I	ACC
FL	*Gossypium darwinii symptomless alphasatellite*	2264	3110	96	0	96	EU384646.1
P	*Gossypium darwinii symptomless alphasatellite*	232	1003	91	9.00E-57	92	EU384651.1
P	*Gossypium darwinii symptomless alphasatellite isolate*	271	1143	93	1.00E-68	95	EU384651.1
P	*Gossypium darwinii symptomless alphasatellite isolate*	279	625	91	5.00E-71	98	EU384651.1
P	*Cotton leaf curl Burewala alphasatellite*	208	1050	91	1.00E-49	99	FN658729.1
P	*Gossypium darwinii symptomless alphasatellite isolate*	217	487	82	1.00E-52	90	EU384650.1
FL	*Cotton leaf curl Multan betasatellite*	2179	4580	97	0	94	HG000665.1
FL	*Croton yellow vein mosaic alphasatellite*	1460	2968	95	0	84	FN658711.1
P	*Cotton leaf curl virus-associated DNA beta isolate*	324	676	64	2.00E-84	92	AF534188.1
P	*Cotton leaf curl Multan betasatellite*	273	859	90	2.00E-69	87	HE602954.1
FL	*Cotton leaf curl Burewala virus complete genome*	4228	5400	100	0	97	FR750321.1

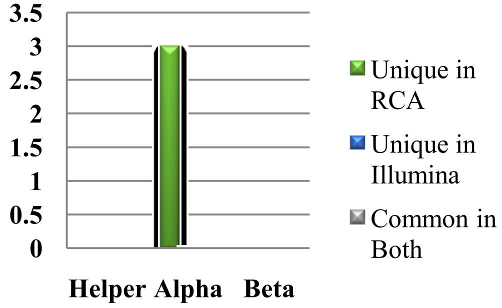
P= Partial sequence, **FL**= Full Length
HA=Hit Accession, **EV**= E-Value, **S**=Similarity (%), **SL**= Sequence Length, **AL**= Alignment length
MS= Maximum Score, **TS**= Total Score, **QC**= Query Cover (%), **EV**= E-Value, I= Identity (%), **ACC**= Accession Number

Sample 80

Host plant: Cotton
Location: Faisalabad (NIBGE), Punjab, Pakistan


Comparison Between Total Number of Sequences of RCA and Illumina


Comparison Between Full Length Sequences of RCA and Illumina


Comparison Between Partial Length Sequences of RCA and Illumina

Illumina sequencing:

1 new full length helper and 2 new full length alphas.

Confirmed 2 full length alphas and 1 beta sequenced by RCA.

3 partial alphas and 1 full length beta unconfirmed by Illumina.

Blast Results

From RCA Pipeline

	Sequence Description	HA	EV	S	SL	AL
FL	*Cotton leaf curl Multan betasatellite isolate*	JF502389	0	96	1338	1303
FL	*Cotton leaf curl virus-associated DNA beta*	AM084380	0	97	1334	1302
FL	*Gossypium darwinii symptomless alphasatellite*	FR877533	0	97	1367	1375
P	*Nanovirus-like particle rep gene for replication associated protein*	AJ512954	4.00E-113	94	628	265
P	*Cotton leaf curl Burewala alphasatellite*	HQ316180	0	98	1007	591
FL	*Gossypium davidsonii symptomless alphasatellite*	EU384652	0	93	1354	639
FL	*Croton yellow vein mosaic alphasatellite*	FN658711	0	87	1347	1363
FL	*Cotton leaf curl Burewala alphasatellite*	FN658728	0	98	1359	1366
P	*Cotton leaf curl Burewala alphasatellite*	HQ728354	2.68E-43	99	140	104

Illumina Sequencing

	Best Hit	MS	TS	QC	EV	I	ACC
FL	*Xanthium symptomless alphasatellite*	2378	2754	94	0	98	HF547408
FL	*Cotton leaf curl Burewala alphasatellite*	2394	2730	91	0	99	FN658728
FL	*Croton yellow vein mosaic alphasatellite*	1696	1841	85	0	88	FN658711
FL	*Cotton leaf curl Multan betasatellite*	2307	3988	99	0	96	HF567943
FL	*Cotton leaf curl Burewala virus segment A*	4091	4091	99	0	99	JF416947
FL	*Mesta yellow vein mosaic virus-associated alphasatellite*	2123	4180	100	0	95	JX183090

P= Partial sequence, **FL**= Full Length
HA=Hit Accession, **EV**= E-Value, **S**=Similarity (%), **SL**= Sequence Length, **AL**= Alignment length
MS= Maximum Score, **TS**= Total Score, **QC**= Query Cover (%), **EV**= E-Value, I= Identity (%), **ACC**= Accession Number

1

- **Mine variability of existing GB sequences:** download 'helper viruses' (2340); β- satellite (1334) sequences /GenBank database (2015)
- **Align with full-length DNA sequences from 'field' pipeline** samples (2011 – 2013) (MUSCLE; CLC Viewer v7.5)

2

- **Haplotype analyses @ 100% identity to exclude multiple identical genotypes:** 1630 haplotypes for helpers and 1246 for β satellites (FaBOX)
- **Phylogenetic analyses (Neighbor joining, MEGA 6)**

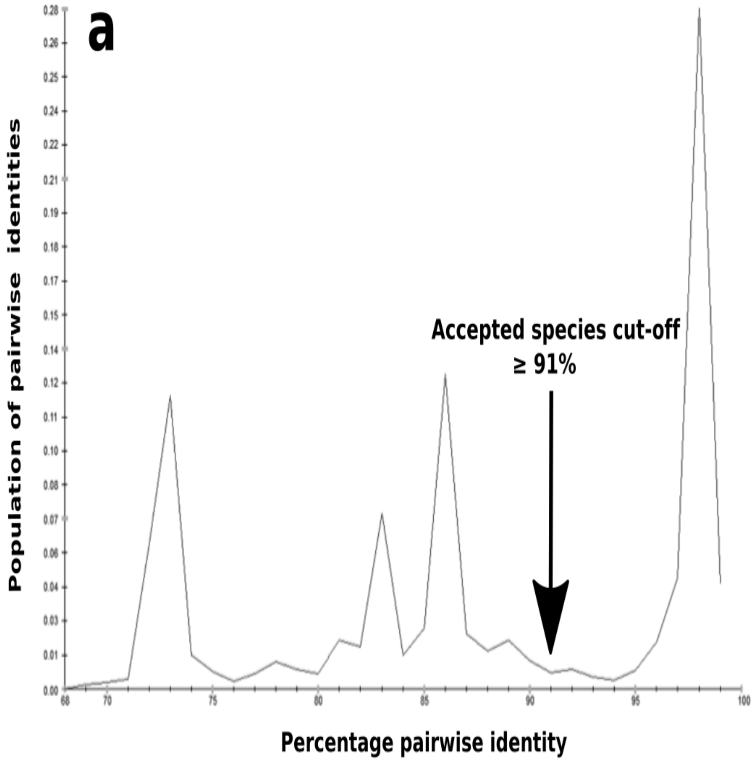
3

- **Species Demarcation Tool (SDT v1.2) pairwise distance analysis**
- Species threshold based on ICTV value at 91% helpers, 78% β

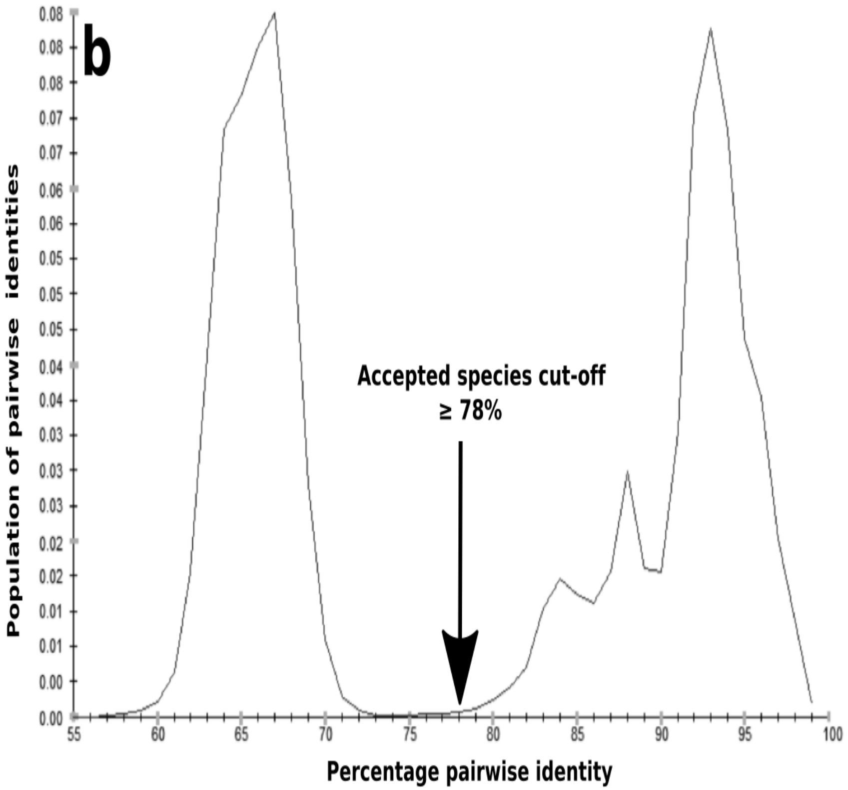
4

- **Primer design using 'groups' defined by SDT analyses**
- **PCR for testing primers** for specificity on each group and on other groups using cloned inserts for representative viruses and betasats identified in environmental surveys

Sorted by pairwise distances: species cut-off



Begomoviruses

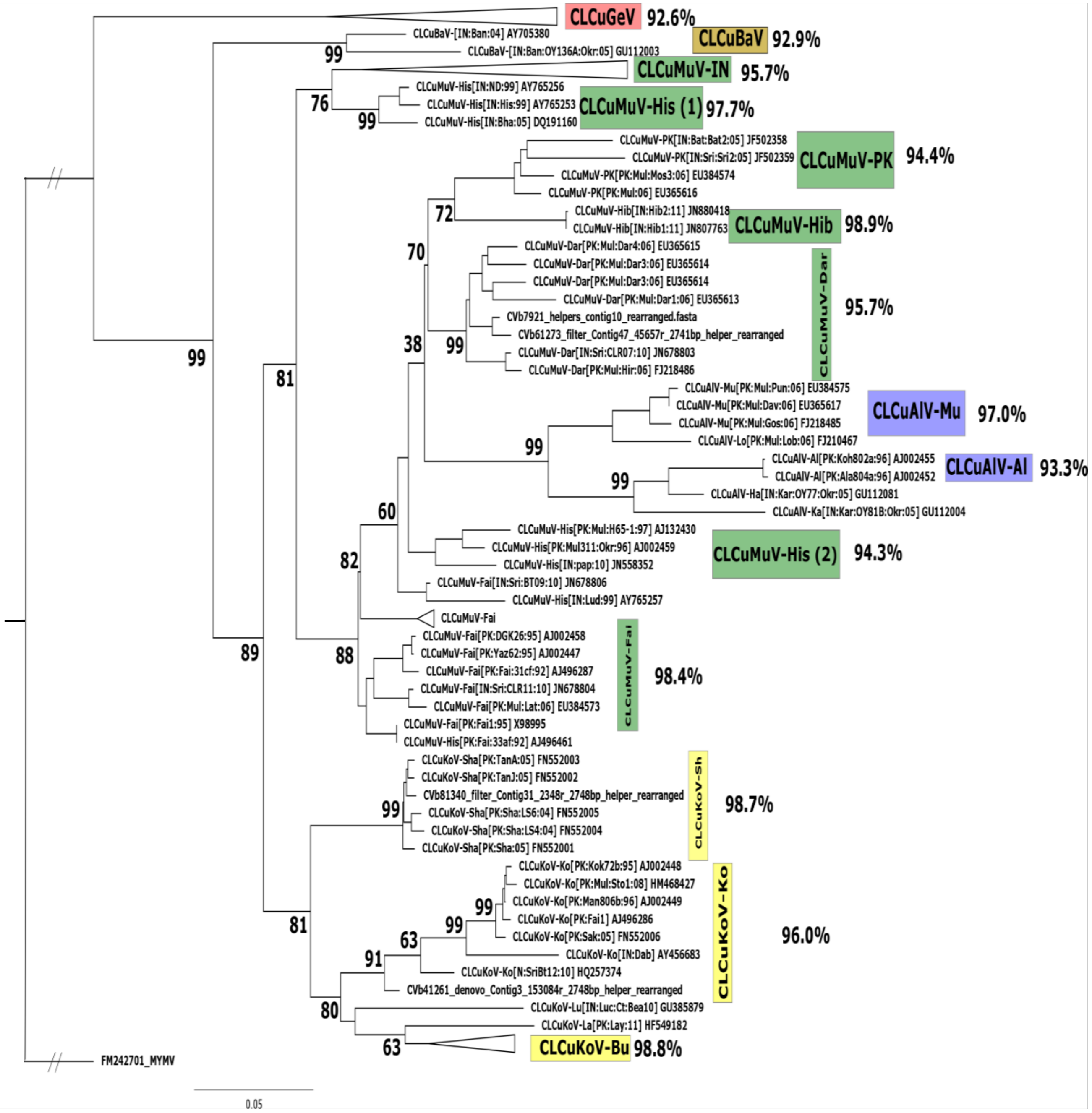


Betasatellites

Phylogenic analysis

Five 'core' leaf curl 'Helper' begomovirus species

… in agreement with pairwise distance analysis (SDT) at <91% species cutoff



Cotton leaf curl Alabad
Cotton leaf curl Banglore
Cotton leaf curl Gezira
Cotton leaf curl Kokhran
Cotton leaf curl Multan

Analyzed 223 and 228 begomoviral genome sequences determined in this study and available in the NCBI-GenBank database, respectively.

Phylogeny betasatellites: four 'core' molecules; agreement with four pairwise distance groups at 78% nt identity cutoff

Four core cotton
leaf curl betasatellites

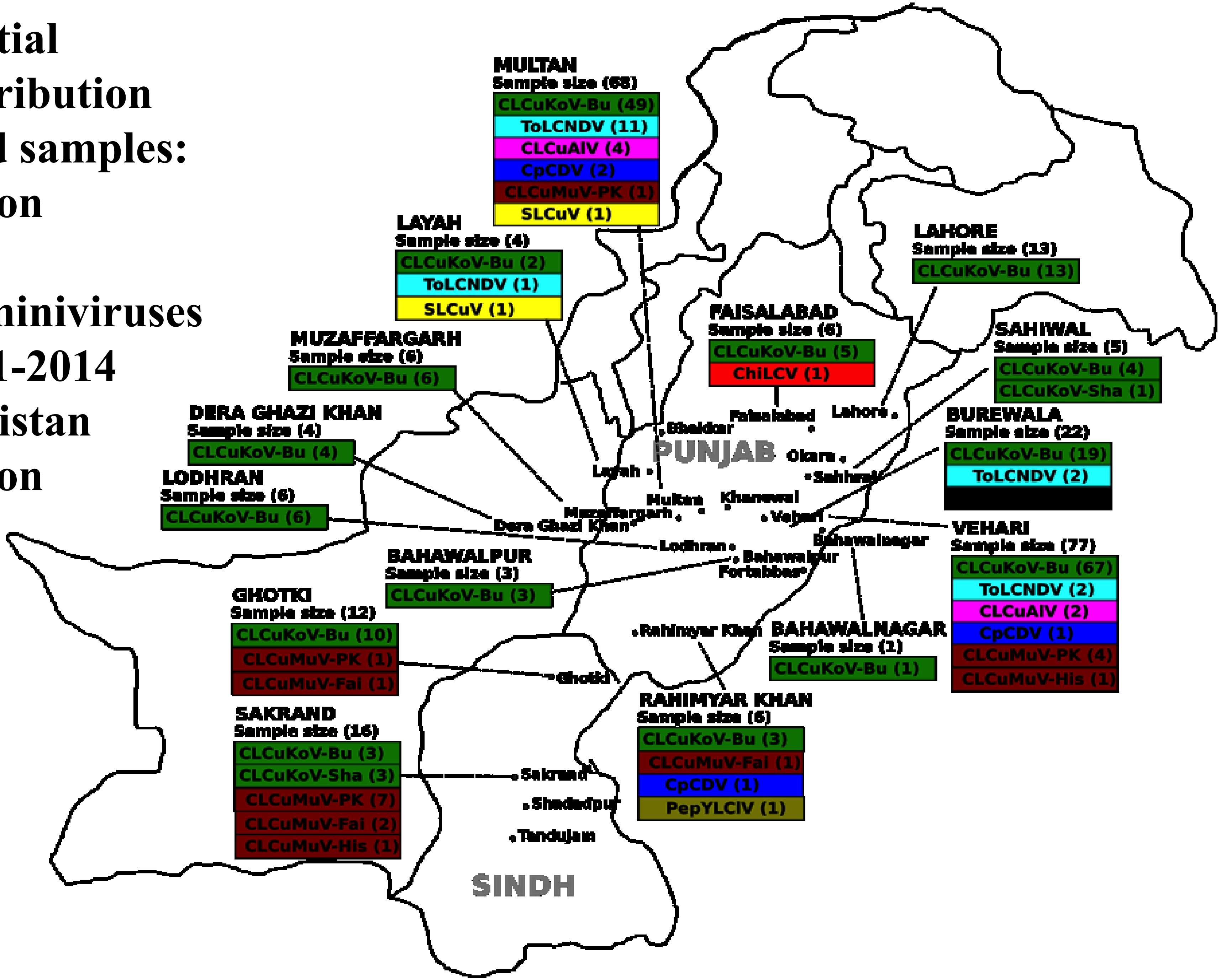
Maximum likelihood tree of betasatellites associated with cotton leaf curl disease:

- **Chili leaf curl beta**
- **Cotton leaf curl Gezira beta**
- **Cotton leaf curl Multan beta**
- **Okra leaf curl beta**

Analysis included 468 and 506 betasatellite DNA sequences determined in this study, and available in the GenBank database, respectively.

Spatial distribution field samples: cotton

Geminiviruses 2011-2014 Pakistan cotton



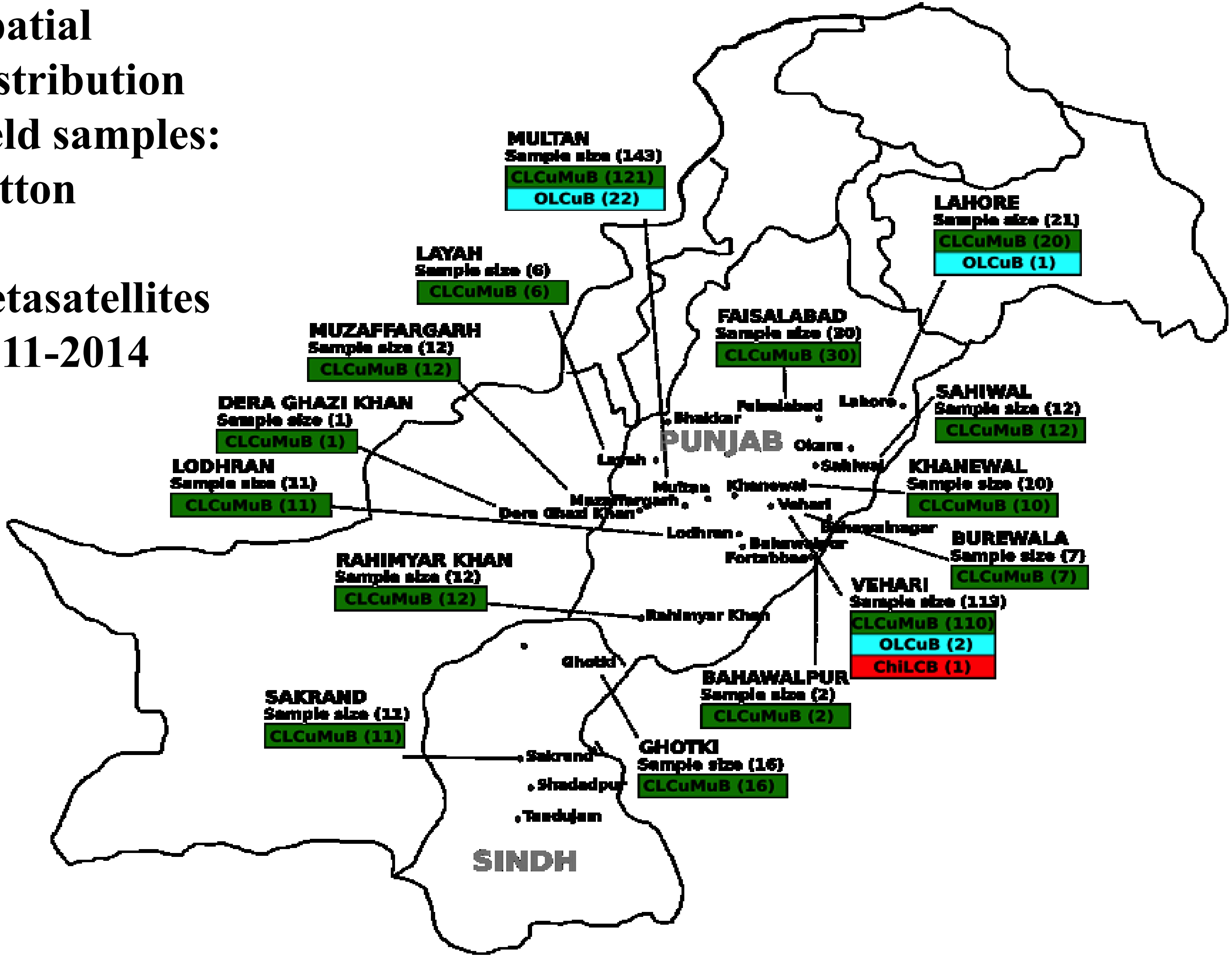
Distribution and abundance of **cotton-infecting geminiviruses** in **cotton** 2011-2015 in Pakistan.

Sequences were obtained using Sanger and/or Illumina DNA sequencing platforms.

All available full-length, and some partial length sequences, at >1500 nucleotides in length, were used for the analysis.

Spatial distribution field samples: cotton

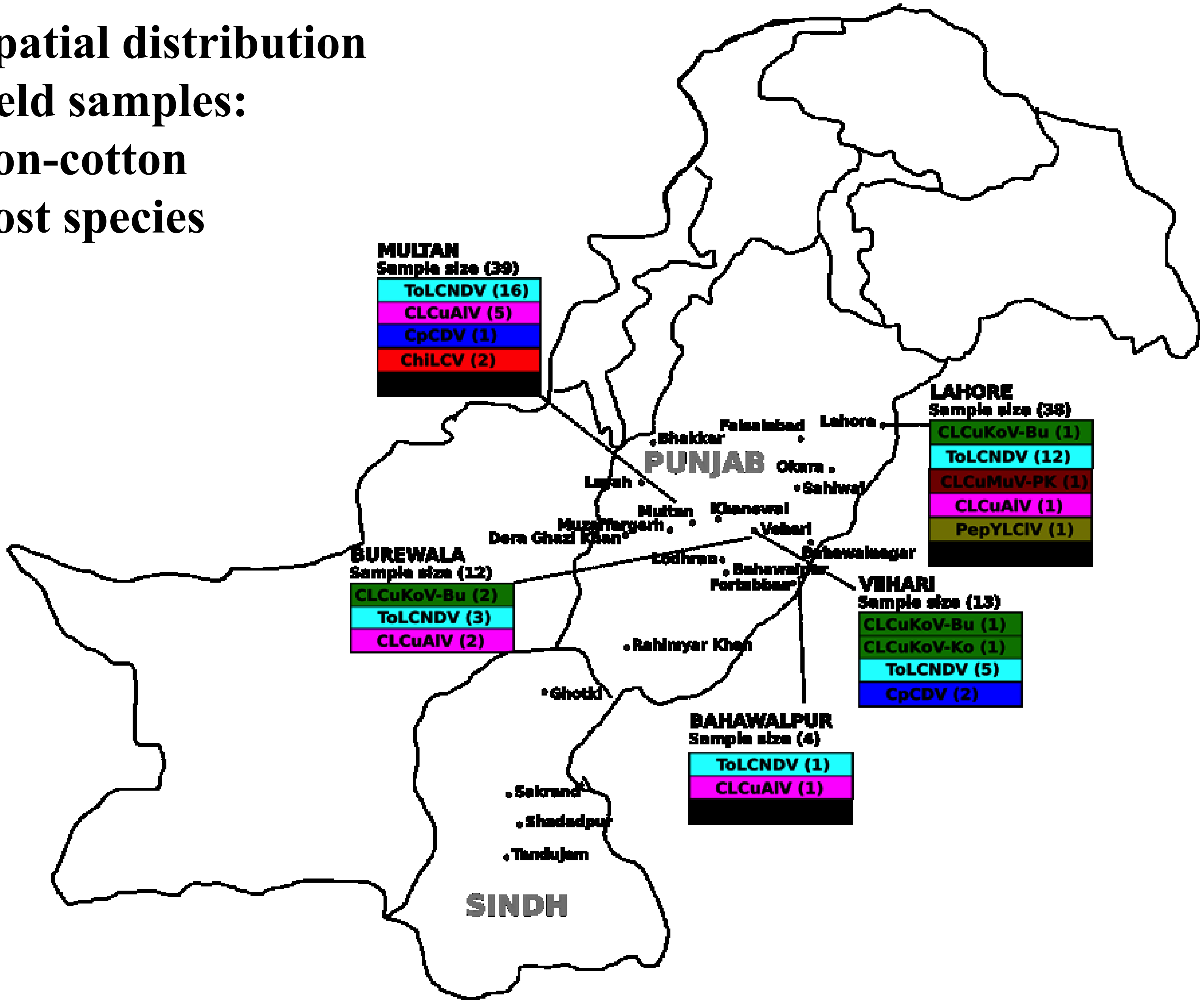
Betasatellites 2011-2014



Distribution and abundance of **betasatellites** associated with leaf curl disease of **cotton** during 2011-2014 in Pakistan.

Full-length DNA sequences of betasatellites were obtained using Sanger and/or Illumina sequencing platforms

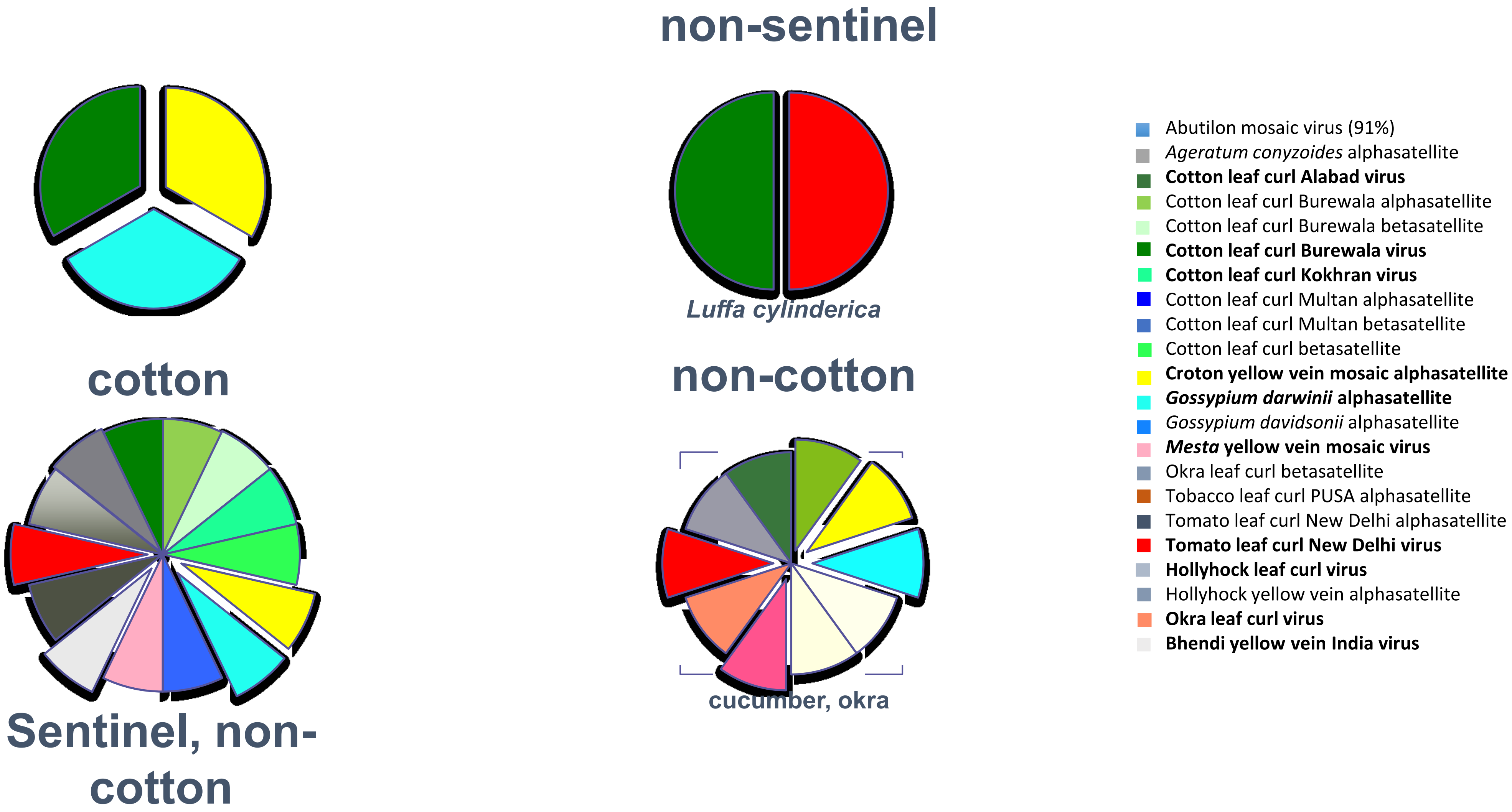
Spatial distribution field samples: non-cotton host species



Distribution and prevalence by location of leaf curl disease-associated geminiviruses identified in '**cultivated and non-cultivated cotton**' host species in **China, India, and Pakistan**.

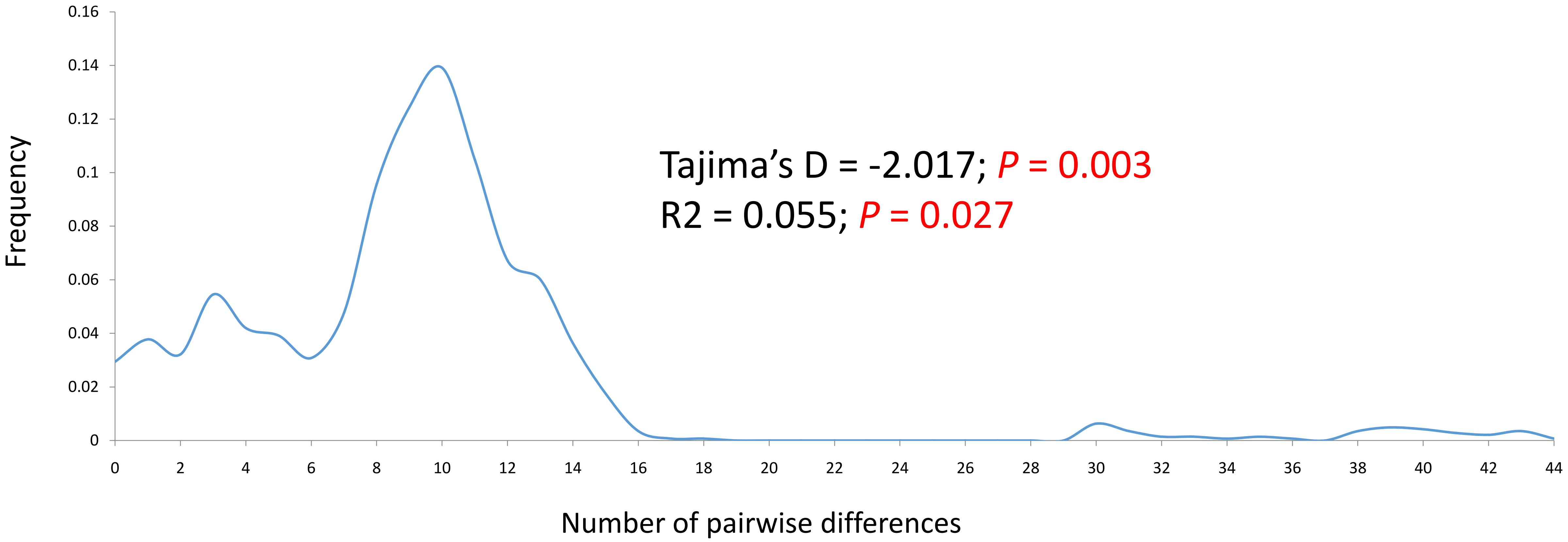
Analysis includes all available full-length and partial genome of >1500 nucleotides in length determined herein, and all full-length genome sequences available in NCBI-GenBank database.

Begomoviruses & satellites in Burewala – greatest diversity in 'proposed' outbreak site



Location of initial population expansion !

Pairwise differences test for population expansion
CLCuKoV-Bu



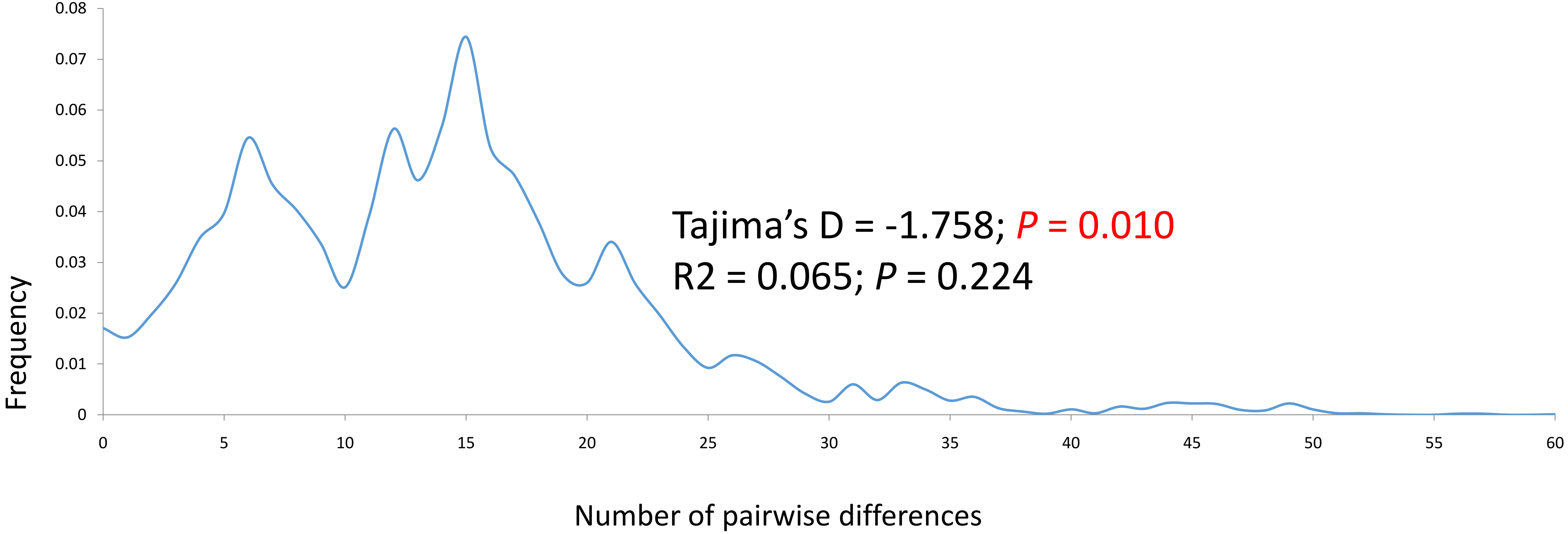
Tajima’s D = -2.017; *P* = 0.003

R2 = 0.055; *P* = 0.027

Results: **Strong evidence for population expansion** in the CLCuMV-Bur “**panmictic**” population

Tajima’s test indicates virus is in **mutation-drift equilibrium** (‘under selection’)

Pairwise differences test for population expansion betasatellite CLCuMB



Tajima’s D = -1.758; *P* = 0.010

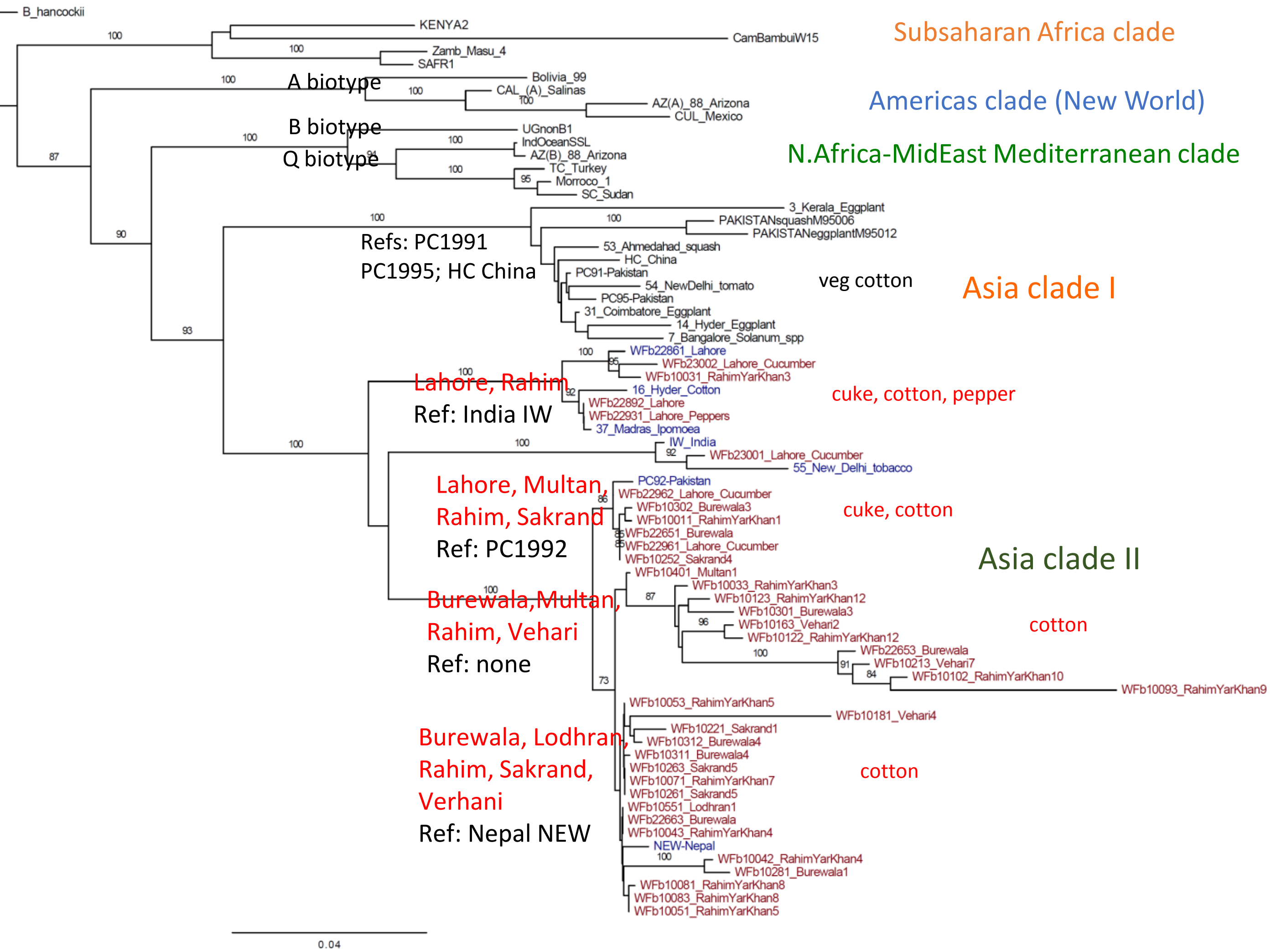
R2 = 0.065; *P* = 0.224

Results: **No** strong evidence of **recent population expansion =** satellite has been present and stable as is, for some time

Significant negative Tajima’s D reflects ‘**purifying**’**, or positive selection**

= single CLCuMB supported by many different helper genomes/species – promiscuity potentially facilities host jumps; supports mixed infections, increases chances of recombination

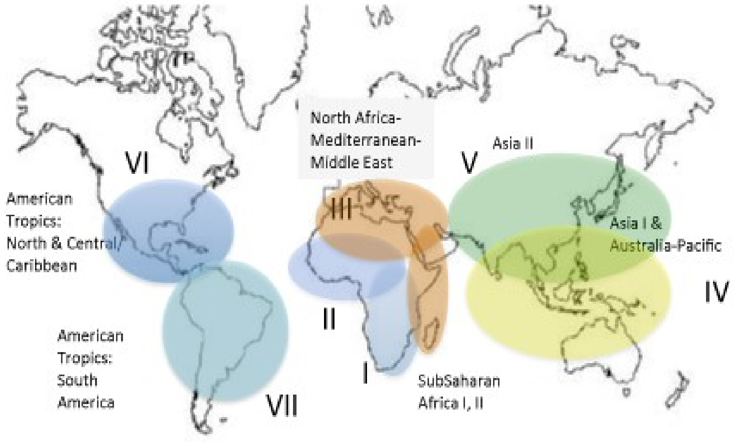
Role of whitefly vector in CLuKoV-Bu spread?



NJ tree

Bootstrap ≥70

- All 2013 haplotypes are members of the Asia II clade
- Previous sampling 1992-1995 showed more haplotype diversity –Asia I, II



Connectors indicate districts sampled; charts show proportion of genetic group abundance in non-cotton (inner) and cotton (outer) hosts.

-Lahore is only locality with all 3 haplotypes (not cotton-area).

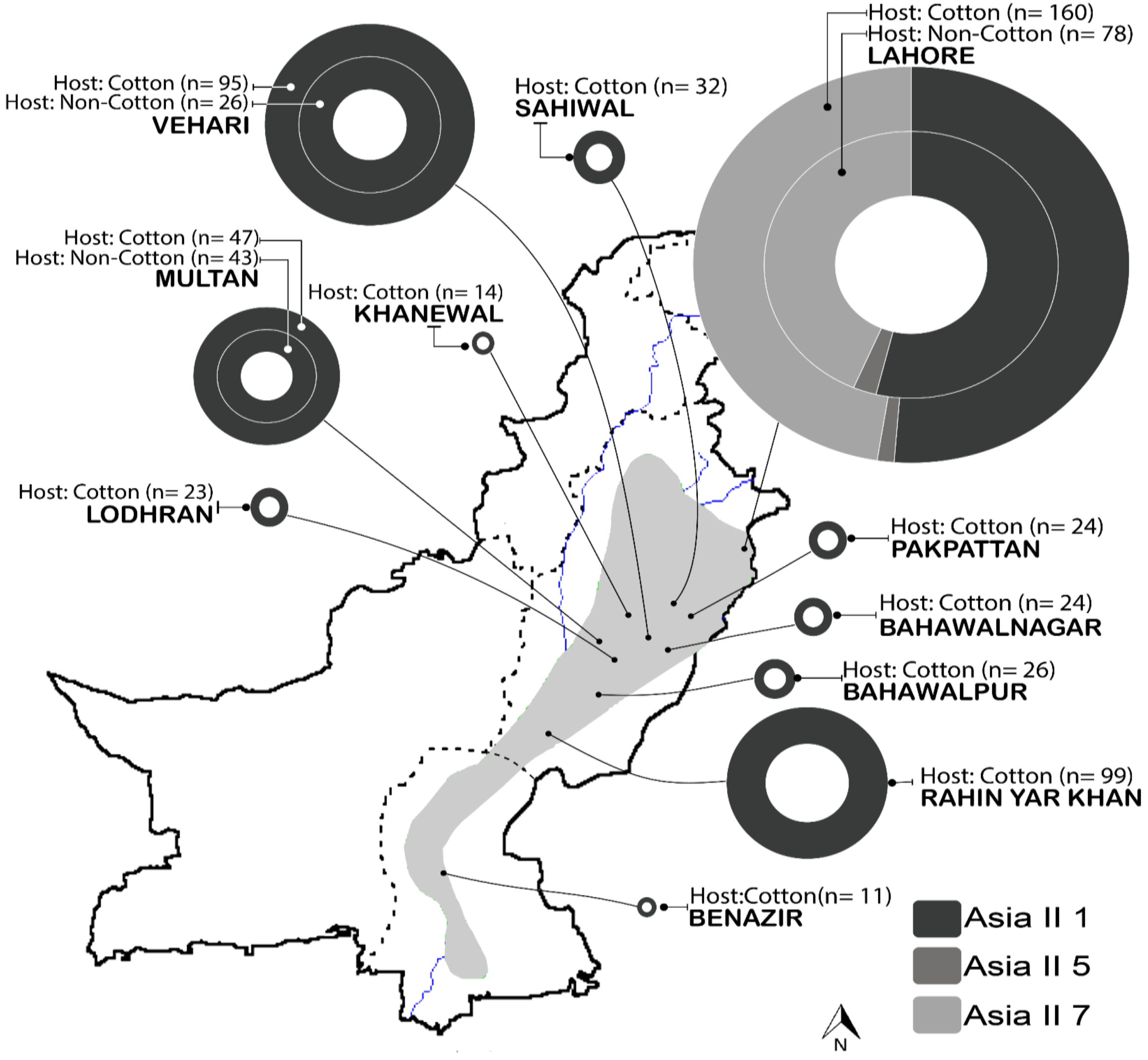
Three main haplotype distributions in Pakistan:

ASIA II-1 is broadly distributed throughout cotton, all crops

Asia II 5 and Asia II-7 found in Lahore, a non-cotton area.

Asia II-7 – minor representation: cotton, pepper, cucumber, tomato, okra and squash

Map of Pakistan showing cotton-growing locations in light-grey



Mismatch distribution:
-predominant haplotype **Asia II 1 is undergoing genetic expansions**

Tajima test revealed Asia II-1 not under mutation-drift equilibrium
-evolving under purifying selection

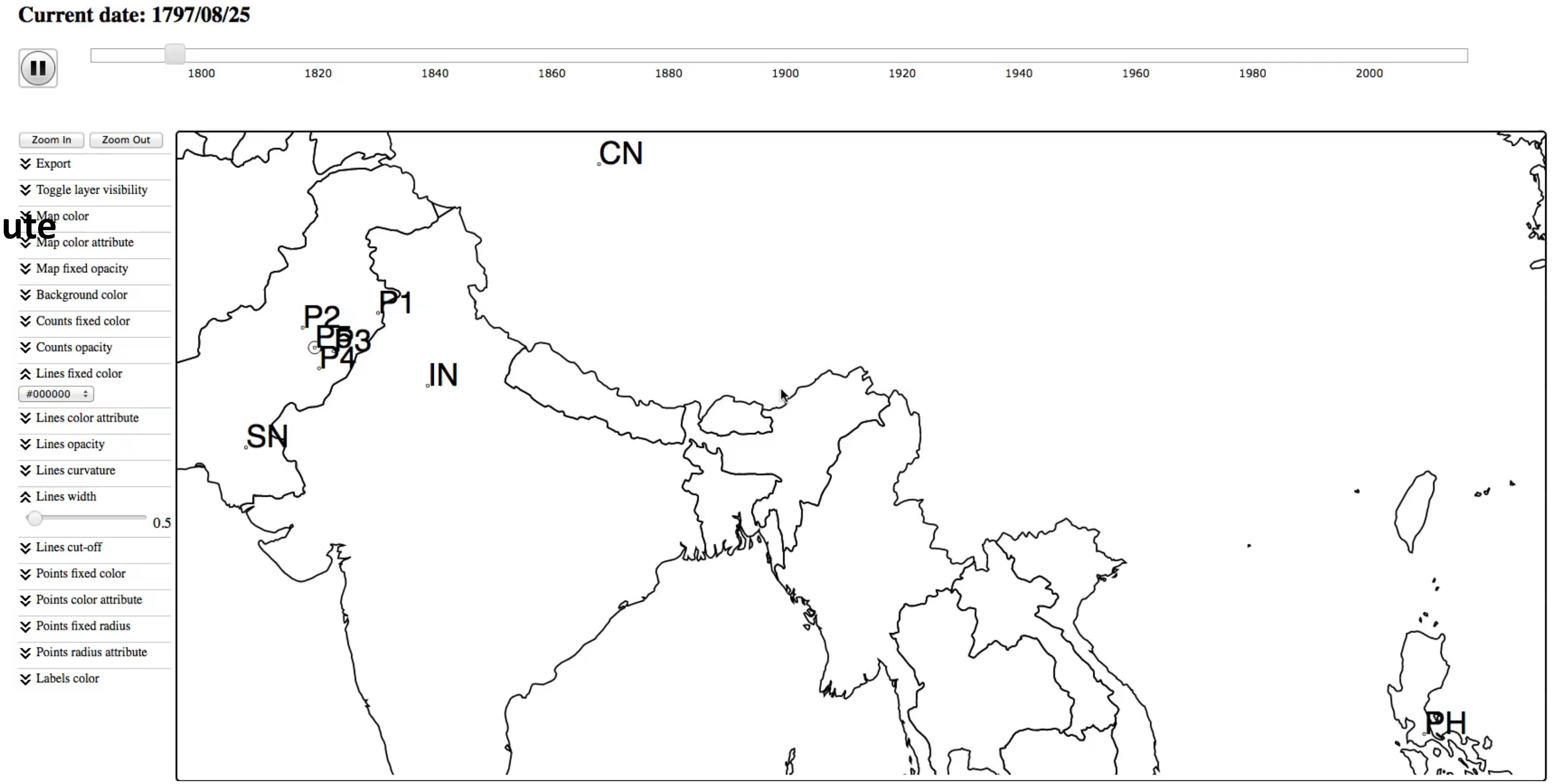
Selection?

- Resistance-breaking virus (single cv)
- Virus-vector interactions: virus driving WF expansion/or WF driving virus expansion
- Monoculture cotton?
- Climate effects-expansion?

‘*SPREAD*’: Cotton leaf curl viruses and Indian sub-continent

Lines color attribute

—	P1	(Punjab 1)
—	P2	(Punjab 2)
—	P3	(Punjab 3)
—	P4	(Punjab 4)
—	P5	(Punjab 5)
—	IN	(India)
—	PH	(Philippines)
—	SN	(Sindh)
—	CN	(China)



Current date: 1797/08/25

Lines fixed color: #000000

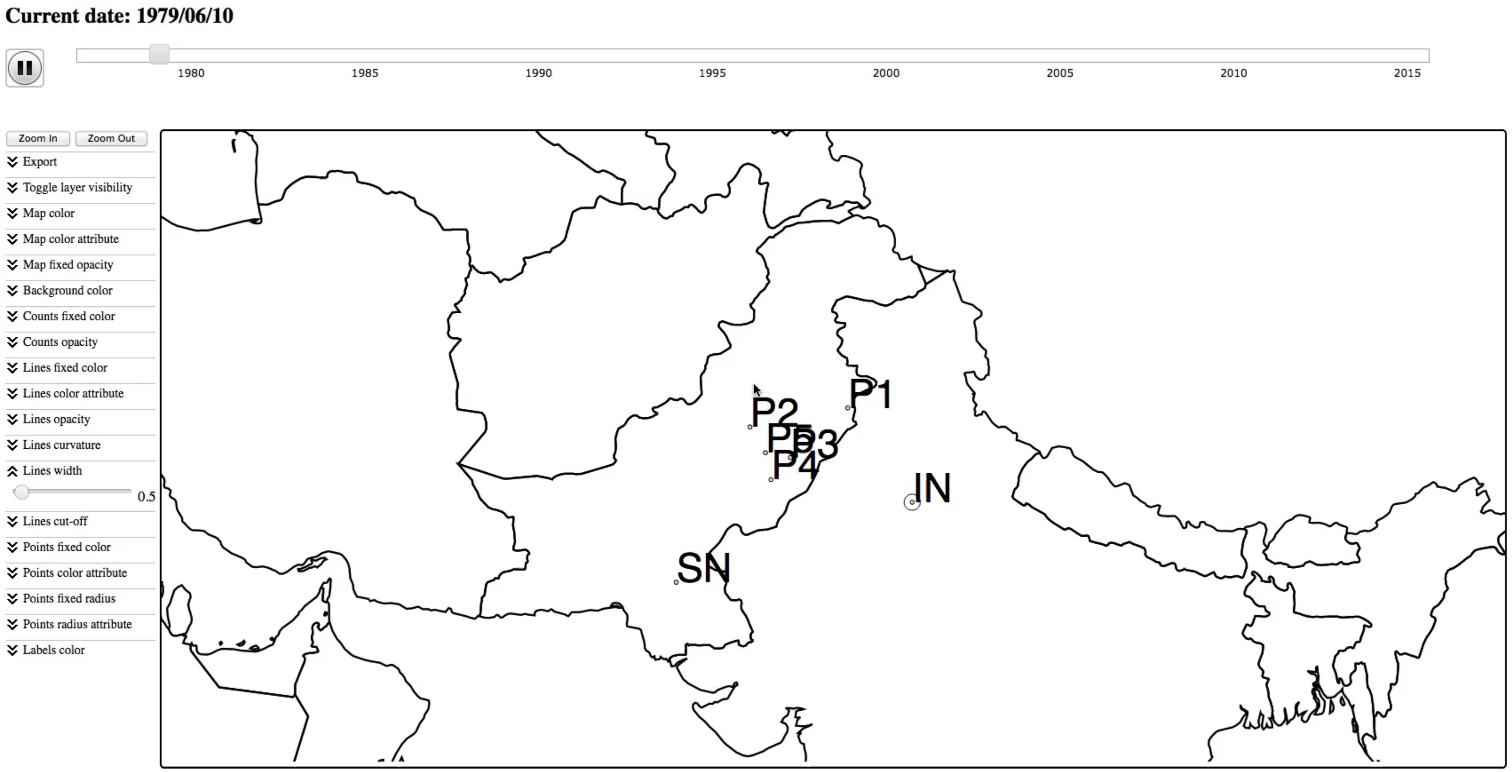
Lines width: 0.5

Full length genomes: *Cotton leaf curl Kokhran virus* (CLCuKoV) (78-GenBank; 212 this study), *Cotton leaf curl Multan virus* (CLCuMuV) (109 GB; 10-this study), *Cotton leaf curl Alabad virus* (CLCuAlV) (8-GB); and *Cotton leaf curl Banglore virus* (CLuCBaV) (2-GB)..

Punjab 1 = Lahore, Faisalabad, Sahiwal **Punjab 2** = Bhakar, Layah, Dera Gazhi Khan; **Punjab 3** = Burewala, Vehari, Sahiwal; **Punjab 4** = Bahawalpur, Bahawalnagar, **Punjab 5** = Multan, Muzafargarh, Khanewal, Rahimyar Khan; Sindh = Sakrand, Ghotki; **India** = Harayana, Punjab and Rajasthan states.

Results: Cotton leaf curl viruses (specifically Cotton leaf curl Multan virus) probably evolved in central Punjab (Pakistan) and entered into India in 1960s and has dispersed back and forth between India and Pakistan since then. In around 2000 CLCuKoV emerged.. In about 2004 CLCuMuV was dispersed form India to China ,and in about 2011 it was transported on plants (Hibiscus) from China to the Philippines.

‘*SPREAD*’: *Cotton leaf curl Kokhran virus*-Burewala strain



Lines color attribute

— P1 (Punjab 1)
— P2 (Punjab 2)
— P3 (Punjab 3)
— P4 (Punjab 4)
— P5 (Punjab 5)
— IN (India)
— SN (Sindh)

Full length genome sequences: 276 CLCuKoV-Bu (64 sequences from GenBank and 212 from this study).

Punjab 1 = Lahore, Faisalabad, Sahiwal **Punjab 2** = Bhakar, Layah, Dera Gazhi Khan; **Punjab 3** = Burewala, Vehari, Sahiwal; **Punjab 4** = Bahawalpur, Bahawalnagar, **Punjab 5** = Multan, Muzafargarh, Khanewal, Rahimyar Khan; Sindh = Sakrand, Ghotki; **India** = Harayana, Punjab and Rajasthan states.

Results: *Cotton leaf curl Kokhran virus*-Burewala (CLCuKoV-Bu) originated in India and dispersed by whiteflies to central Punjab in Pakistan) around 2000 from where it was dispersed by whiteflies to other parts of Punjab. In ~2012 CLCuKoV-Bu was dispersed by whiteflies to Sindh. Continued to disperse back and forth between India and Pakistan.

Summary

- *Cotton leaf curl Multan virus*, the species associated with the initial epidemic occurring during the 1990s, was detected primarily in southern provinces – Punjab and Sindh.
- CLCuMuV, *Cotton leaf curl Alabad virus,* and CLCuKoV-Kokhran strain, also detected in cotton during Multan epidemic, were present in cotton and non-cotton hosts in locations where 'Multan' 'susceptible' cotton varieties were planted.
- Geminiviruses previously found in non-cotton species, unexpectedly, were detected in cotton: *Okra enation leaf curl virus*, *Tomato leaf curl New Delhi virus*, *Squash leaf curl virus,* and *Chickpea chlorotic dwarf virus*. (intraspecific recombination!).
- CLCuMB) was most prevalent betasatellite in cotton and non-cotton hosts. Important factor in spread – 'stable component of the complex'
- Whitefly *B. tabaci* Asia II-1 mitotype shows evidence of population expansion; may have occurred coincident with the emergence and spread of CLCKoV-Bu in 2000.
- Host-virus-whitefly interactions could explain enhanced virus spread and Asia II-1 upsurgence, resulting in displacement of previously prevalent *B. tabaci* mitotypes.
- Great potential for CLCuMuV to spread from the Pacific to Australia, across to HI and Guam, then US west coast, Mexico, Central America
- Has CLCuKoV-Bu already spread *via* ornamental species/whitefly but remains undetected?

Acknowledgements:

Dr. M. Ilyas

Dr. Zia Ur-Rehman

Sofia Avelar, M.S. student

Funding:

USDA Cotton Improvement project-Pakistan

Cotton Incorporated, USA

Thank you

Objectives/Approaches

Spatial and temporal distribution analysis of geminivirus-betasatellites associated with the cotton leaf curl disease in Pakistan 2011-2015

Approaches:

- Molecular analysis: RCA, PCR, restriction digestion/cloning pipeline
- Genomic pathology: Illumina ‘discovery tool’: 2015 begomovirus-negative samples
- Analyses: 2011-2015 and all Genbank db ‘leaf curl’ genome and betasatellite sequences

- Phylogenetic analysis: relationships, diversity within and between clades
- Changes in prevalence and location: dynamics in time and space
- SNPS analysis: assessment of single nucleotide changes - mutations
- BEAST and ‘SPREAD’: epidemiological study

Molecular diagnostics for the leaf curl complex

(validation, in progress)

Primer and probe design for isothermal PCR

1

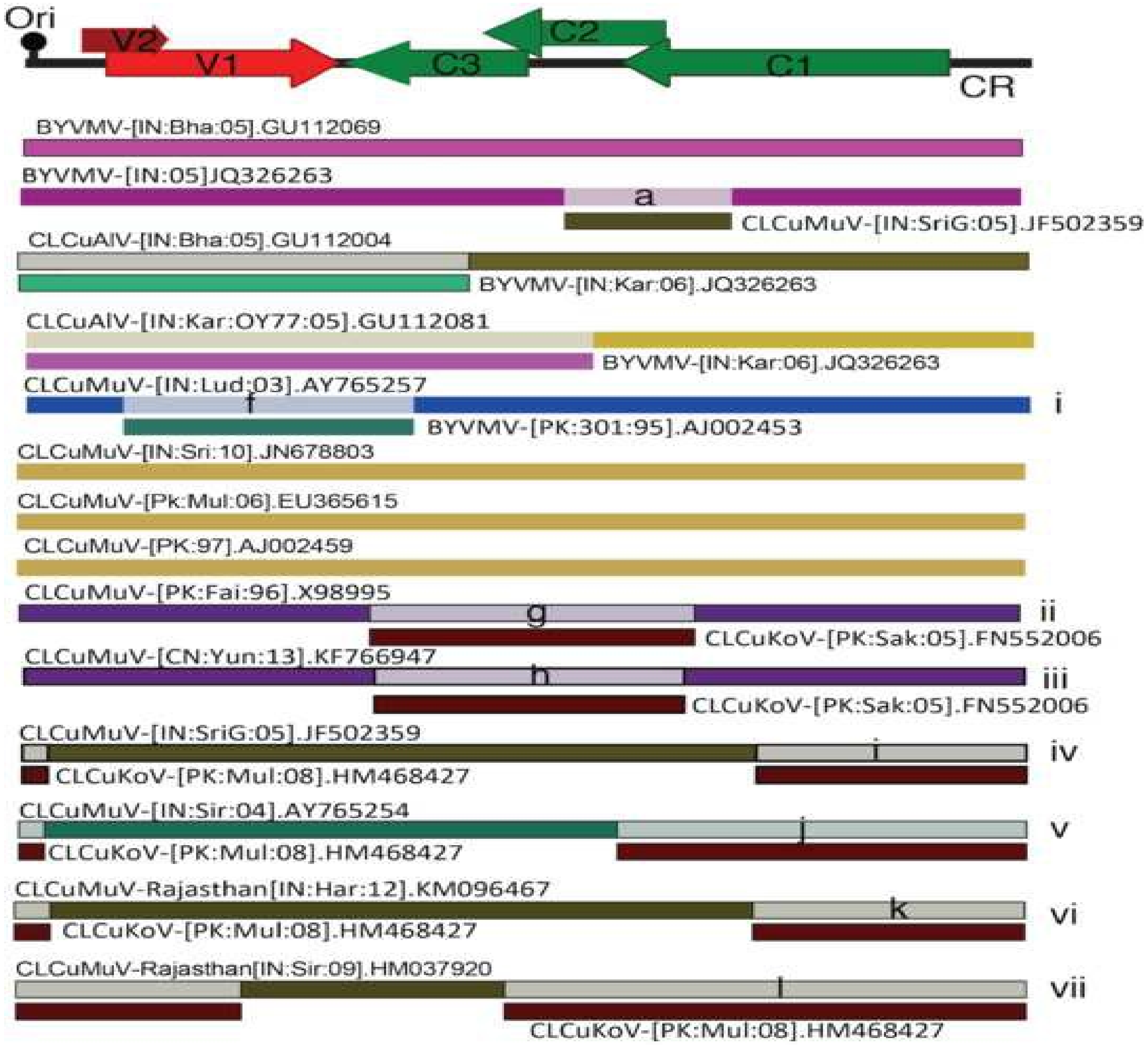
- Primer/probe design using ‘groups’ defined by SDT analyses

2

- Conventional PCR for testing primers for specificity on each group and on other groups using cloned inserts for representative viruses and betasatellites identified in environmental surveys

3

- Isothermal PCR using AmplifyRP® Acceler8™ Discovery for testing primers for specificity on each group and on other groups using cloned inserts for representative viruses and betasatellites identified in environmental surveys

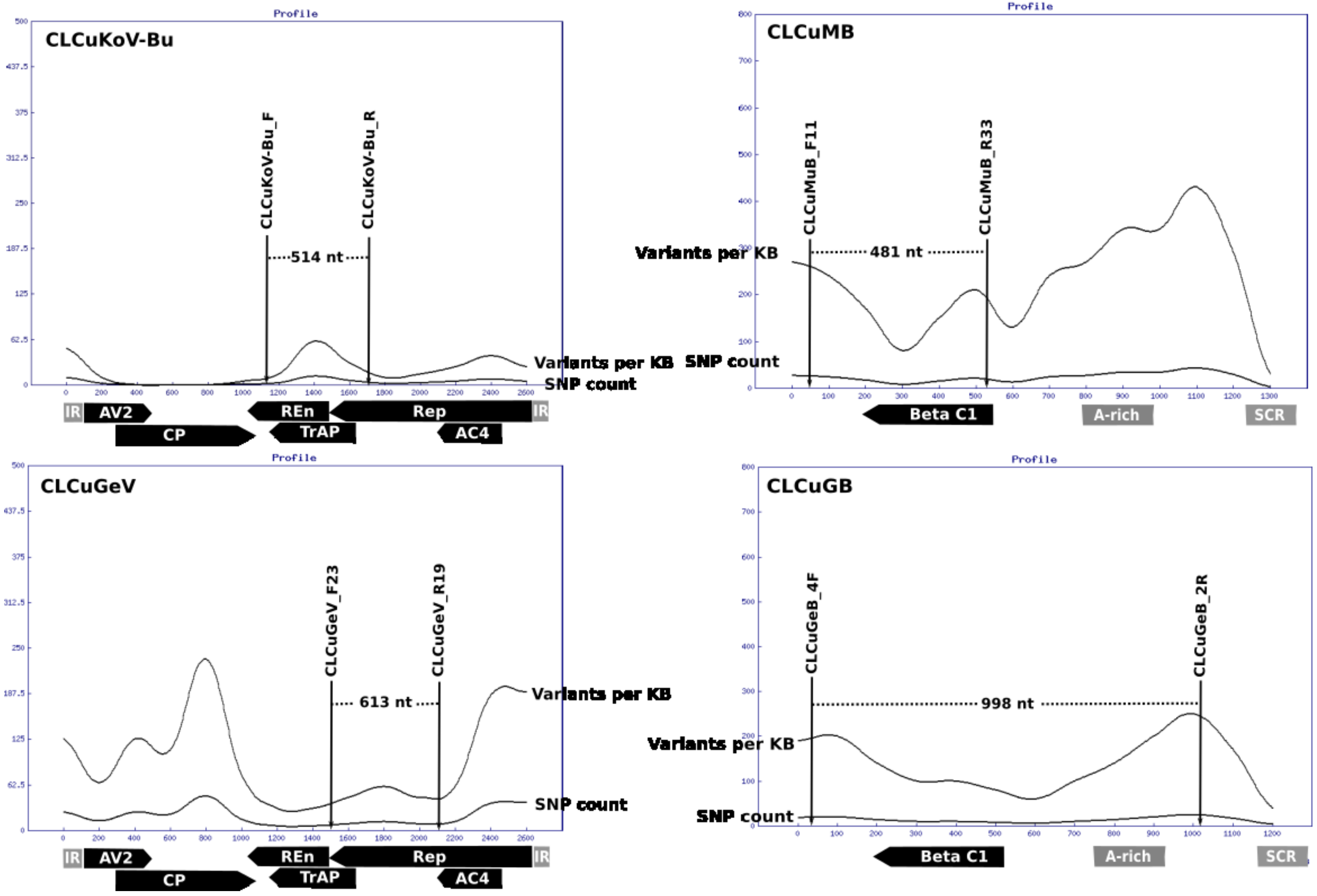


Considerations for primer design:

Extensive diversity among cotton leaf curl virus genomes

1. high genetic recombination

Single nucleotide polymorphisms (mutations)



Begomoviruses

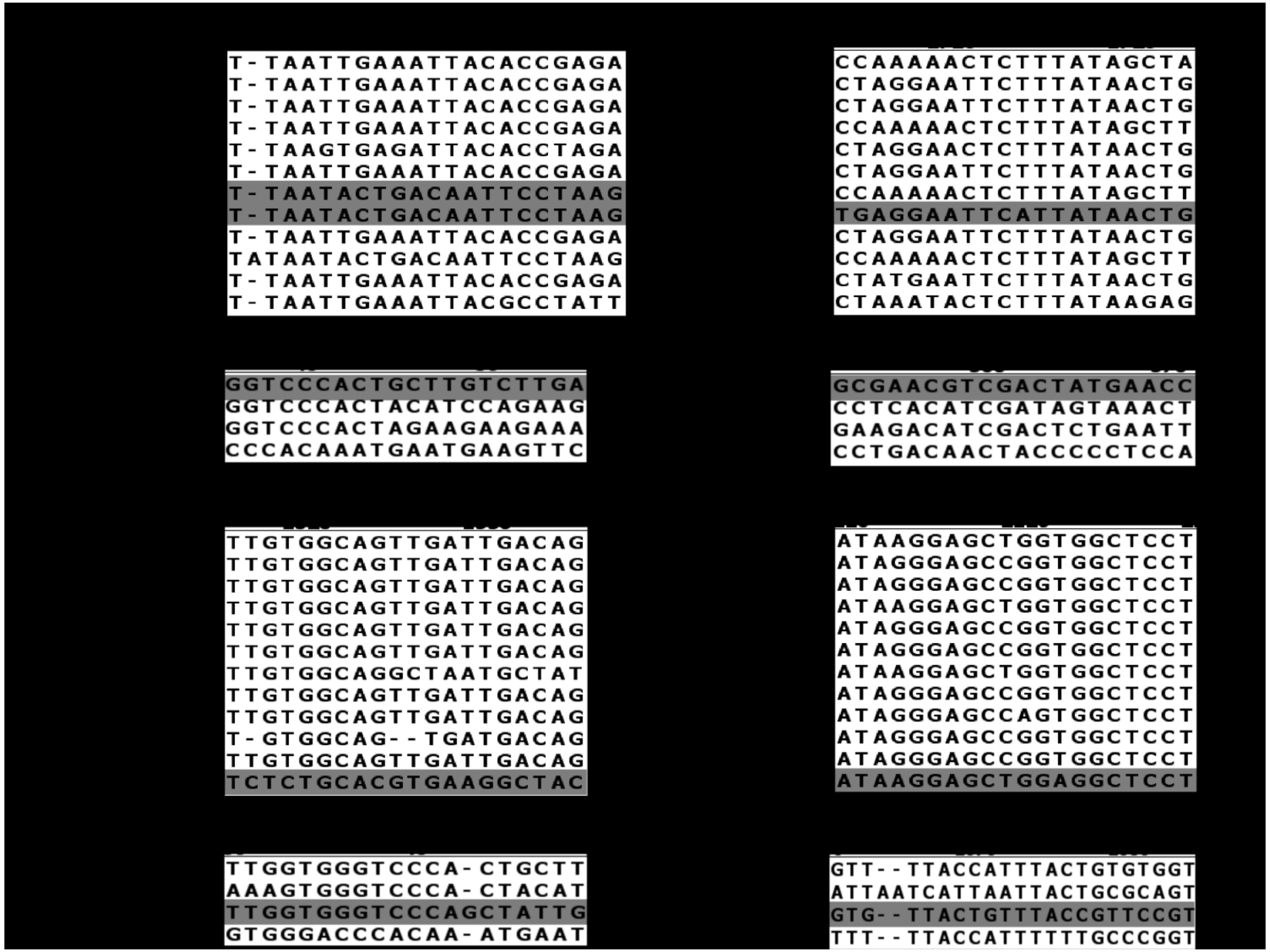
Betasatellites

SNPs analysis avoid regions of high diversification

Single nucleotide polymorphisms & number of variants per genome sequence for *Cotton leaf curl Gezira virus* and *Cotton leaf curl Kokhran virus-Burewala* and the most commonly associated betasatellite, Cotton leaf curl Gezira and Cotton leaf curl Multan beta, respectively.

Predicted begomoviral proteins are shown as AV2, coat protein (CP1), replication enhancer (REn), transcriptional activator (TrAP), replication associated protein (Rep), and AC4.

The predicted protein encoded by the betasatellite is beta C1. The non-coding region of begomoviruses is the intergenic region (IR); for betasatellites, the A-rich region (A-rich) and satellite conserved region (SCR).



```
T-TAATTGAAATTACACCGAGA
T-TAATTGAAATTACACCGAGA
T-TAATTGAAATTACACCGAGA
T-TAATTGAAATTACACCGAGA
T-TAAGTGAGATTACACCTAGA
T-TAATTGAAATTACACCGAGA
T-TAATACTGACAATTCCTAAG
T-TAATACTGACAATTCCTAAG
T-TAATTGAAATTACACCGAGA
TATAATACTGACAATTCCTAAG
T-TAATTGAAATTACACCGAGA
T-TAATTGAAATTACGCCTATT
```

```
CCAAAAACTCTTTATAGCTA
CTAGGAATTCTTTATAACTG
CTAGGAATTCTTTATAACTG
CCAAAAACTCTTTATAGCTT
CTAGGAACTCTTTATAACTG
CTAGGAATTCTTTATAACTG
CCAAAAACTCTTTATAGCTT
TGAGGAATTCATTATAACTG
CTAGGAATTCTTTATAACTG
CCAAAAACTCTTTATAGCTT
CTATGAATTCTTTATAACTG
CTAAATACTCTTTATAAGAG
```

```
GGTCCCACTGCTTGTCTTGA
GGTCCCACTACATCCAGAAG
GGTCCCACTAGAAGAAGAAA
CCCACAAATGAATGAAGTTC
```

```
GCGAACGTCGACTATGAACC
CCTCACATCGATAGTAAACT
GAAGACATCGACTCTGAATT
CCTGACAACTACCCCCTCCA
```

```
TTGTGGCAGTTGATTGACAG
TTGTGGCAGTTGATTGACAG
TTGTGGCAGTTGATTGACAG
TTGTGGCAGTTGATTGACAG
TTGTGGCAGTTGATTGACAG
TTGTGGCAGTTGATTGACAG
TTGTGGCAGGCTAATGCTAT
TTGTGGCAGTTGATTGACAG
TTGTGGCAGTTGATTGACAG
T-GTGGCAG--TGATGACAG
TTGTGGCAGTTGATTGACAG
TCTCTGCACGTGAAGGCTAC
```

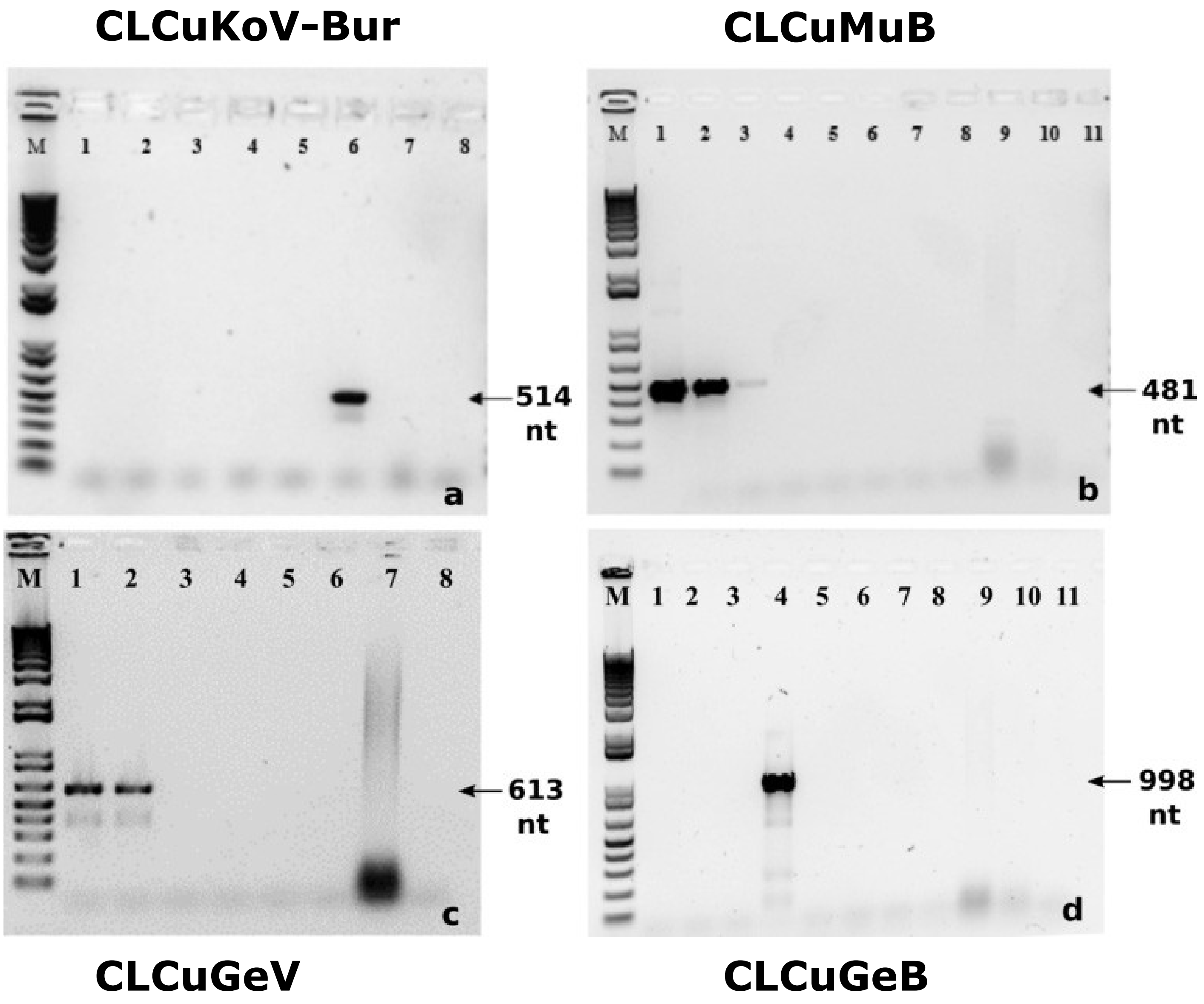
```
ATAAGGAGCTGGTGGCTCCT
ATAGGGAGCCGGTGGCTCCT
ATAGGGAGCCGGTGGCTCCT
ATAAGGAGCTGGTGGCTCCT
ATAGGGAGCCGGTGGCTCCT
ATAGGGAGCCGGTGGCTCCT
ATAAGGAGCTGGTGGCTCCT
ATAGGGAGCCGGTGGCTCCT
ATAGGGAGCCAGTGGCTCCT
ATAGGGAGCCGGTGGCTCCT
ATAGGGAGCCGGTGGCTCCT
ATAAGGAGCTGGAGGCTCCT
```

```
TTGGTGGGTCCCA-CTGCTT
AAAGTGGGTCCCA-CTACAT
TTGGTGGGTCCCAGCTATTG
GTGGGACCCACAA-ATGAAT
```

```
GTT--TTACCATTTACTGTGGT
ATTAATCATTAATTACTGCGCAGT
GTG--TTACTGTTTACCGTTCCGT
TTT--TTACCATTTTTTGCCCGGT
```

Example of alignments used for primer design: sequences that are 100% conserved are highlighted, and the primer coordinates are shown in round brackets

PCR amplification with specific primers

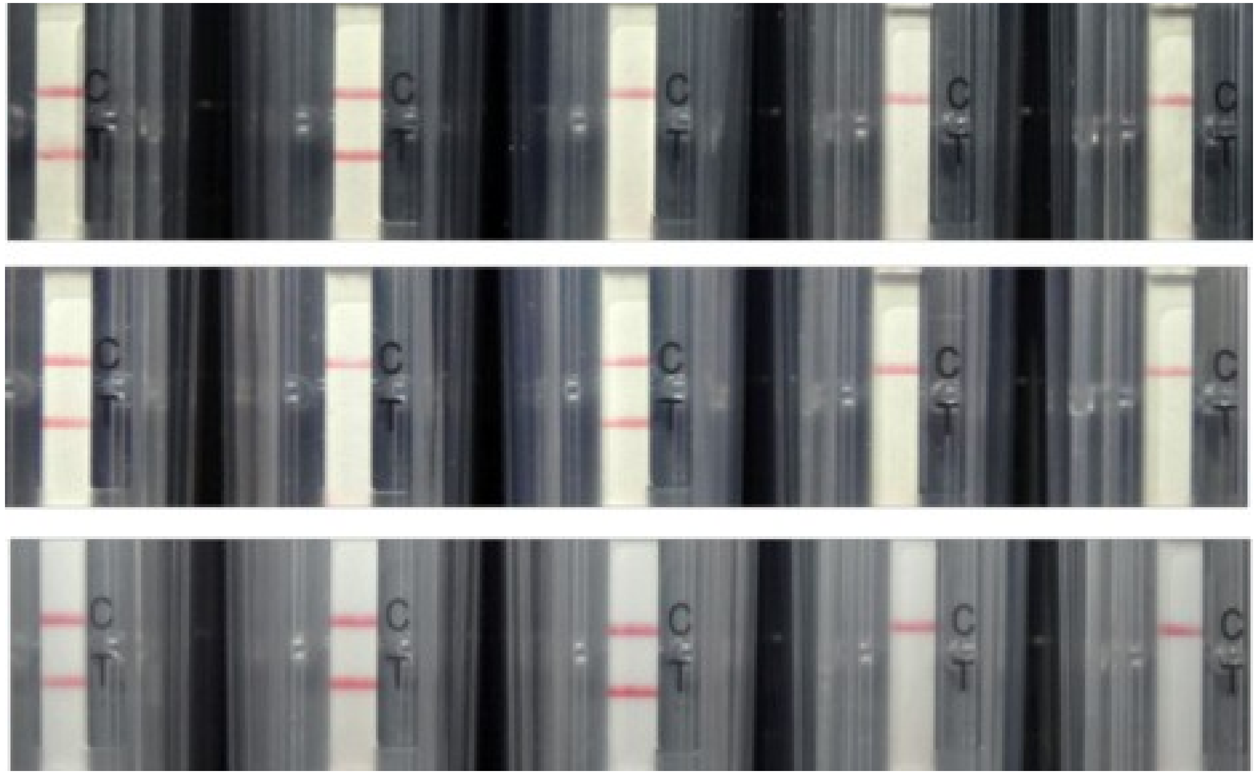


A, C
Lane 1, CLCuGeV (AF260241)
Lane 2, CLCuGeV (AY036007)
Lane 3, CLCuMuV (China)
Lane 4, BYVV
Lane 5 CpCDV
Lane 6, CLCuKoV-Bur
Lane 7, healthy cotton
Lane 8 water negative control.

B, D
Lane M 1 Kb Plus DNA ladder
Lane 1 CLCuMuB (China)
Lane 2 CLCuMuB (Pakistan)
Lane 3 CLCuMuB (kenaf)
Lane 4 CLCuGeB’
Lane 5 OLCuB
Lane 6 OLCuB
Lane 7 BYVB
Lane 8 ChLCuB
Lane 9 healthy cotton
Lane 10 healthy tomato
Lane 11 water negative control

AmplifyRP Acceler8 Discovery Platform

Templates : 1, plasmid positive control; 2, cotton infected with CLCuMuV and CLCuMuB; 3, cotton infected with CLCuKoV and CLCuMuB; 4 healthy cotton; and 5, GEB3 control. Plant extracts were obtained by macerating leaves with GEB3.



Control line
Test line

CLCuMuV primers/probe

Control line
Test line

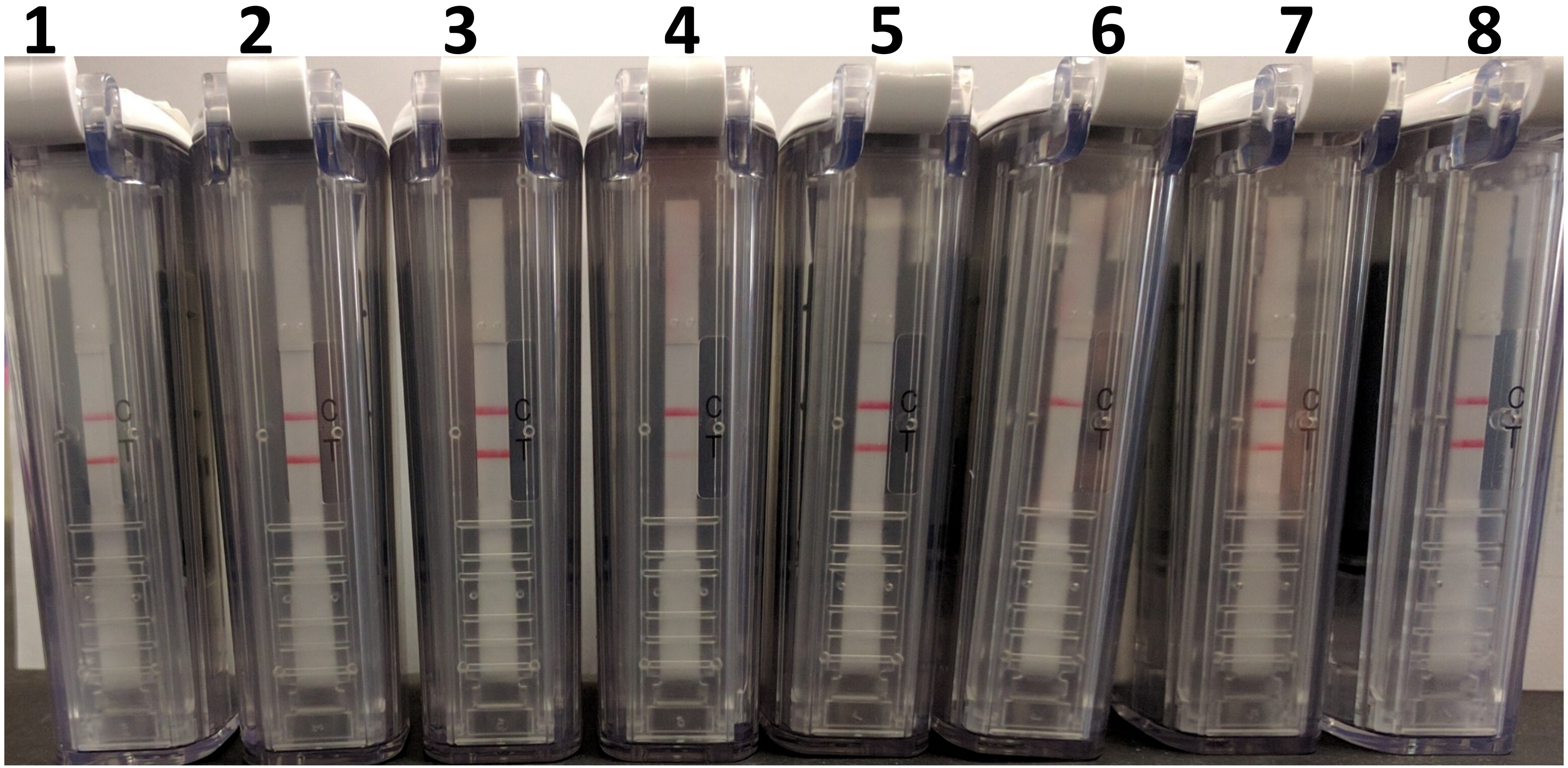
CLCuKoV primers/probe

Control line
Test line

CLCuMuB primers/probe

Isothermal primers/probe for broad-spectrum betasatellite detection
AmplifyRP® Acceler8™ Discovery using 0.5 ng cloned betasatellites

		Forward		Reverse		Probe
Cotton leaf curl Multan Betasatellite	MuB-F2 (30mer)	ATTATGGGTCTGTTTGTTGTTGGTGATATG	MuB-R2 (33mer)	GAAGAGATCGAGATAGAAGATATAGTTCACAG	MuB-P2 (47mer)	ACGGTTCGATTACATCCATTCCCAATATCTHTGGGTTTTCAAGTAC



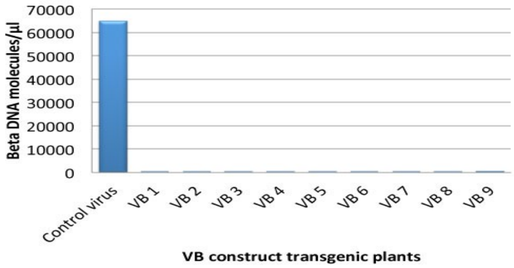
1. 1246 CLCuMB
2. 946 CLCuMB
3. 738 CLCuMB
4. 948 CLCuGB
5. 1349 ChLCB
6. 1250 OLCuB
7. 1251 BYVB
8. 476 CLCuGB

Broad detection primer-probe

Caveats: high variability, recombination; PCR only detects what you look for

Transgenic resistance in cotton: direct transformation of embryo with hairpins (RNAi)

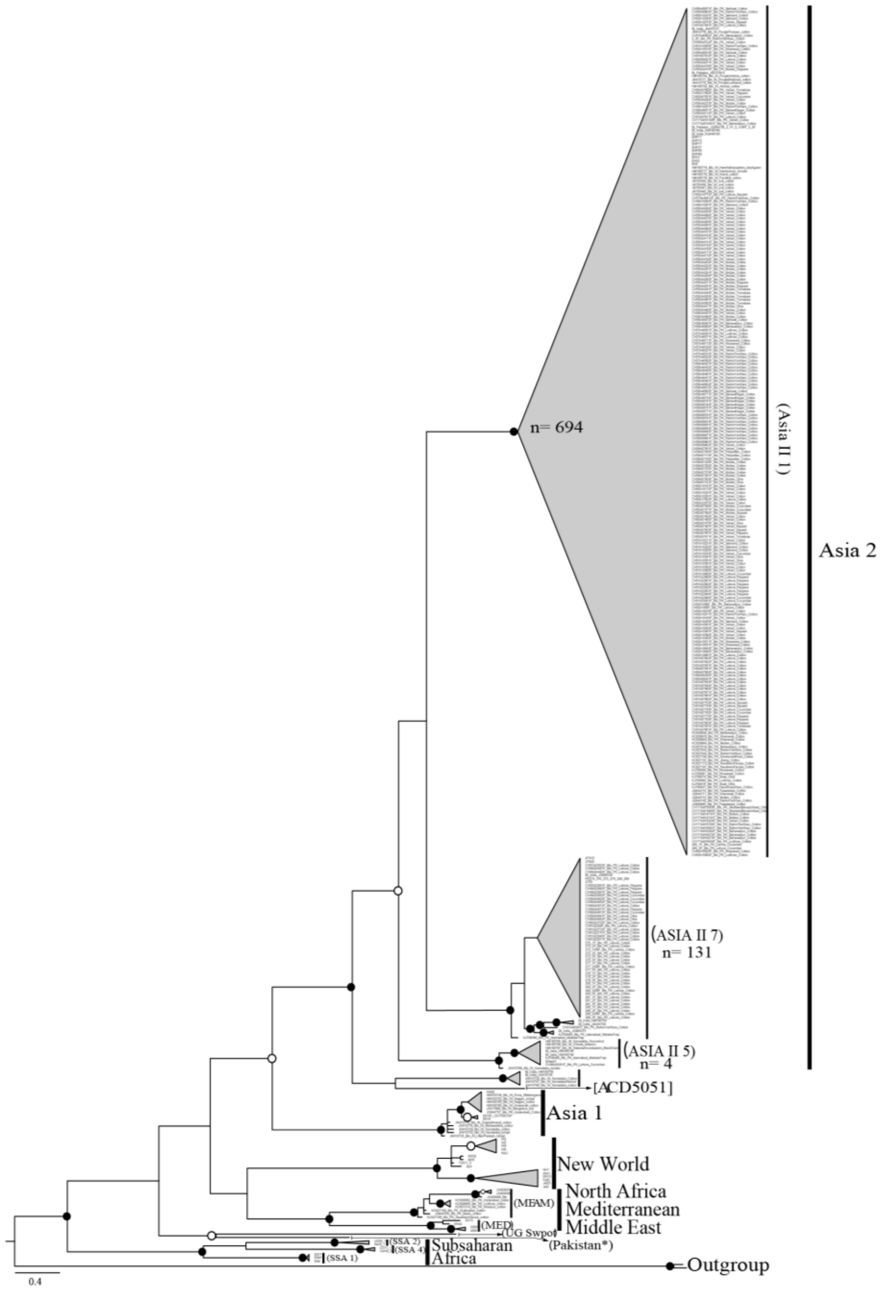
Field-test CEMB: whitefly inoculation June 2014



Joint Collaboration: IAGS and CEMB, U-Punjab, NIBGE, U-AZ

Confirmed insertions: FISH, Southern blot, molecular methods

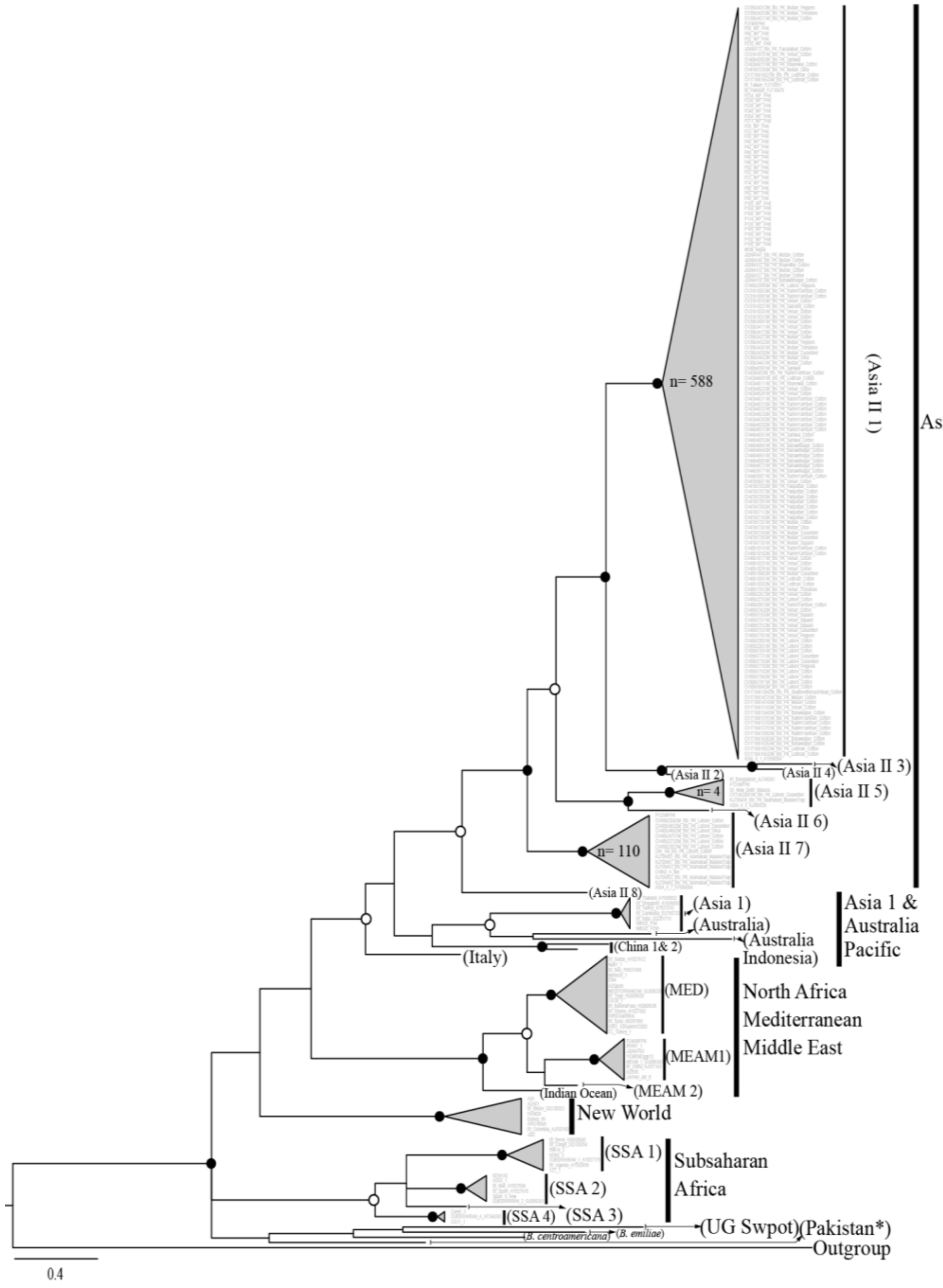
*Constructs 8,9, others: virus and whitefly protection



Bayesian inference of the *B. tabaci* s.s.g. phylogeny based on the 5′-fragment. Nomenclature in parentheses corresponds to ‘putative’ species cutoffs at 3.5% divergence (Dinsdale et al. 2010).

The bold lines correspond to the seven major geographical clades (Brown (2010). Closed circles: ≥99% posterior probability, opened circles: 95% - 98% posterior probability.

The outgroup sequence is of the greenhouse whitefly *Trialeurodes vaporariorum* (Westwood).



Bayesian inference of the *B. tabaci* s.s.g. phylogeny based on the **3’-end of the COI gene**. Nomenclature in parentheses corresponding to ‘putative’ species of Dinsdale et al. (2010), and bolded names correspond to the major geographical clades from Brown (2010).

Closed circles show ≥99% posterior probability, opened circles show 95% - 98% posterior probabilities; * a previously undetected haplotype variant from Pakistan (Ashfaq et al. (2014).

The outgroup is the banded-winged whitefly *Trialeurodes abutilonea* (Haldeman).

Example: Isothermal Primers/Probe for Cotton leaf curl 'Burewala' virus using AmplifyRP® Acceler8™ Discovery

After 10 minutes

Experiment Control line

Test line

CLCuBuV CLCuGV CLCuMuV

Positive detection and specific for CLCuBV

After 20 minutes

Experiment Control line

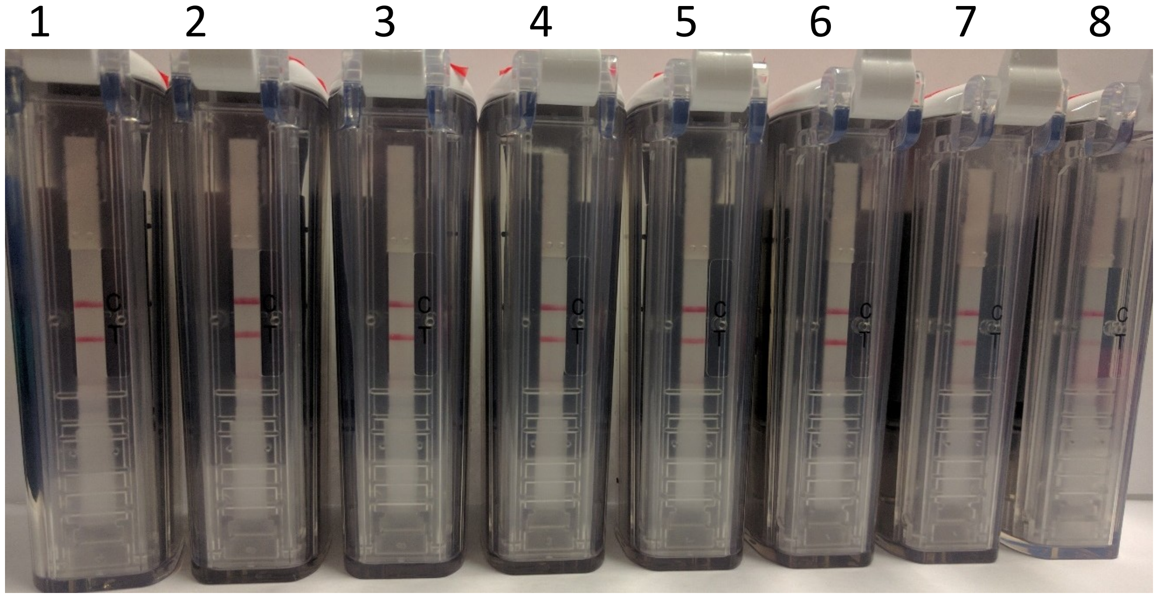
Test line

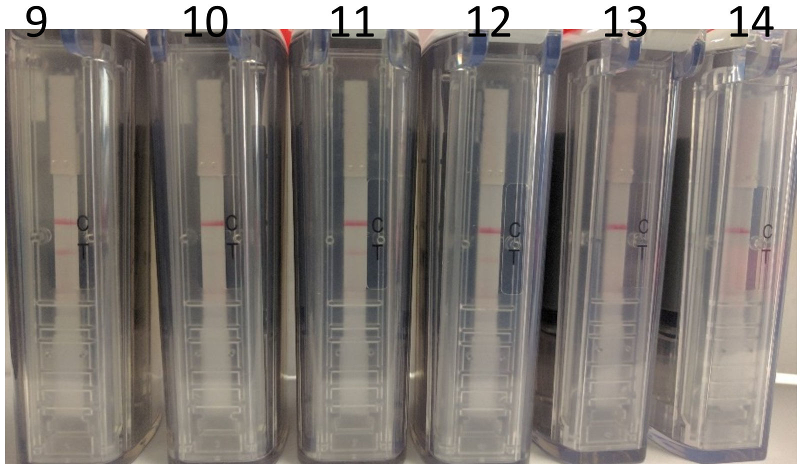
CLCuBuV CLCuGV CLCuMuV

		Forward		Reverse		Probe
Cotton leaf curl Burewala virus	BuV-F6 (34mer)	CCATAGAGTTGGTAAGAGATTTTGTGTTAAGTC	BuV-R5 (33mer)	ACTGGGCTCATTATCAAACATATTAAATACCTC	BuV-P6 (50mer)	TCACACGAATAGTGTGATGTTTTTCTTGGTTAGAGATCGTAGACCTGTTG

Isothermal Primers/Probe for broad detection of helper virus complex
AmplifyRP® Acceler8™ Discovery 0.5 ng cloned viral genomes

		Forward		Reverse		Probe
Cotton leaf curl Multan Virus	MuV-F1 (32mer)	GTCAAGAAGTTTGTCAGAGTTAACAATTATGT	MuV-R2 (33mer)	AATATATCCTAATCTTCAACGTAGCATAAACAG	MuV-P2 (51mer)	GCAGGAAAATACGAGAATCATACGGAAAATGCGTTAATGCTTTATATGGC



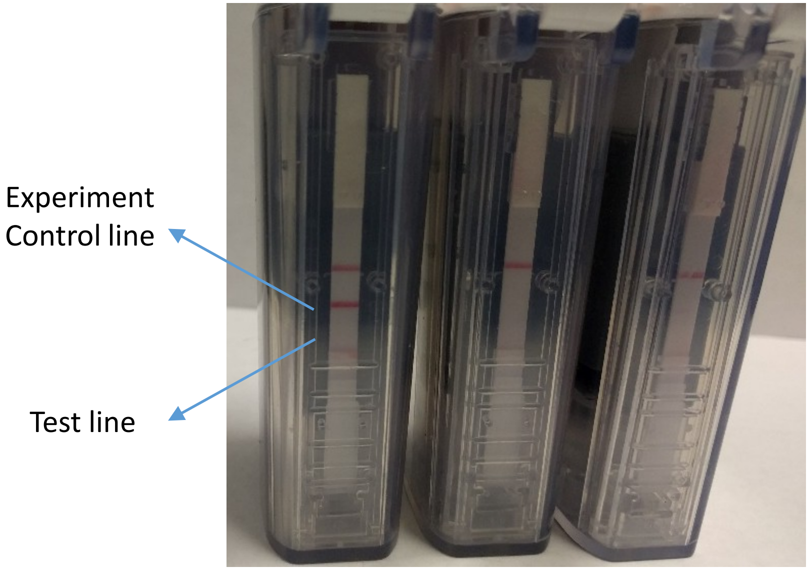


1. 737 CLCuMuV
2. 1213 ToLCNDV-B
3. 1188 CLCuRaV
4. 1195 ToLCNDV-A
5. 1187 CLCuBuV
6. 1203 BYVV
7. 475 CLCuGeV
8. 1348 CpCDV
9. 401 TYLCV
10. 551 SLCV-A
11. 712 HoYVMV
12. 947 CLCuKoV
13. 1189 OELCuV
14. 552 SLCV-B

Cotton leaf curl complex

Example of Results: Isothermal primers/probe for Cotton leaf curl Multan betasatellite using AmplifyRP® Acceler8™ Discovery

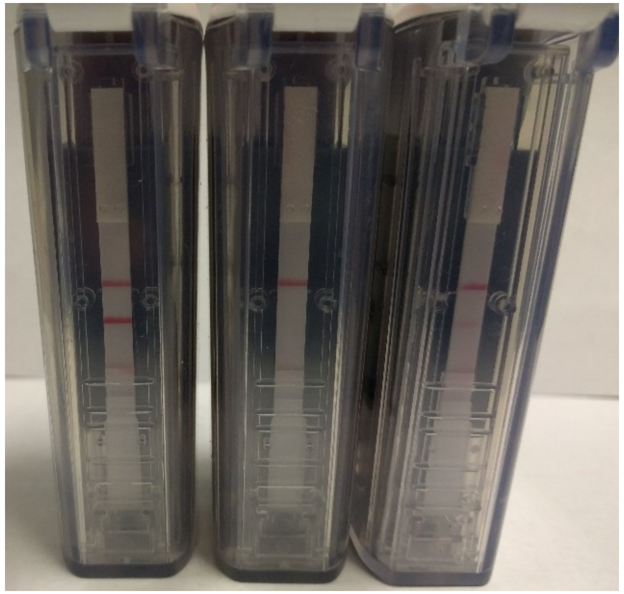
After 10 minutes



CLCuMB CLCuGB OLCuB

Positive detection and specific for CLCuMV

After 20 minutes



CLCuMB CLCuGB OLCuB

		Forward		Reverse		Probe
Cotton leaf curl Multan Betasatellite	MuB-F1 32mer	GTCAAGAAGTTTGTCAG AGTTAACAATTATGT	MuB-R1 33mer	GTCATAAAAATATATCCTA ATCTTCAACGTAGC	MuB-P1 50mer	TACGAGAATCATACGGAAAATGCGT TAATGCTTTATATGGCTTGTACCCA