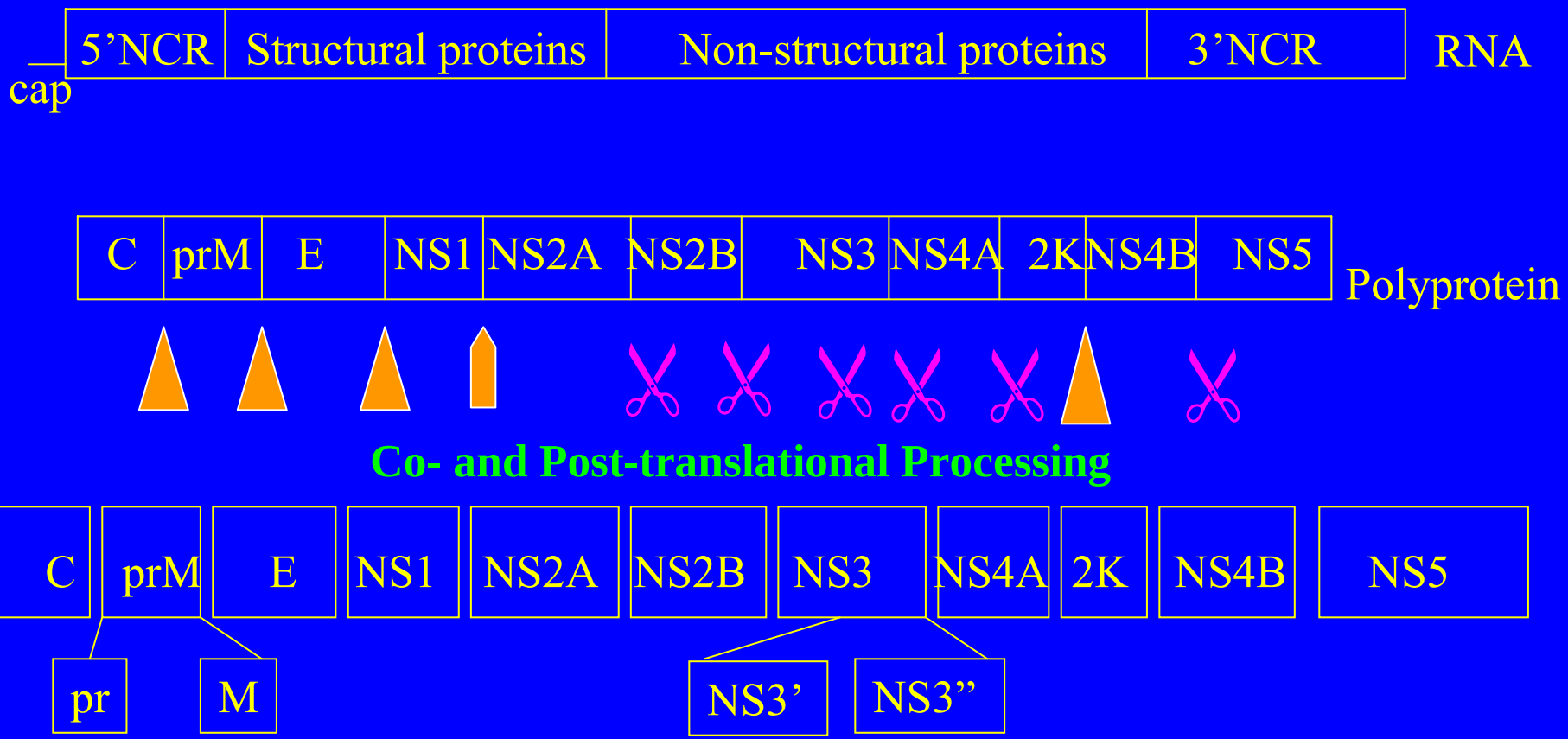
Genetic variation in West Nile virus

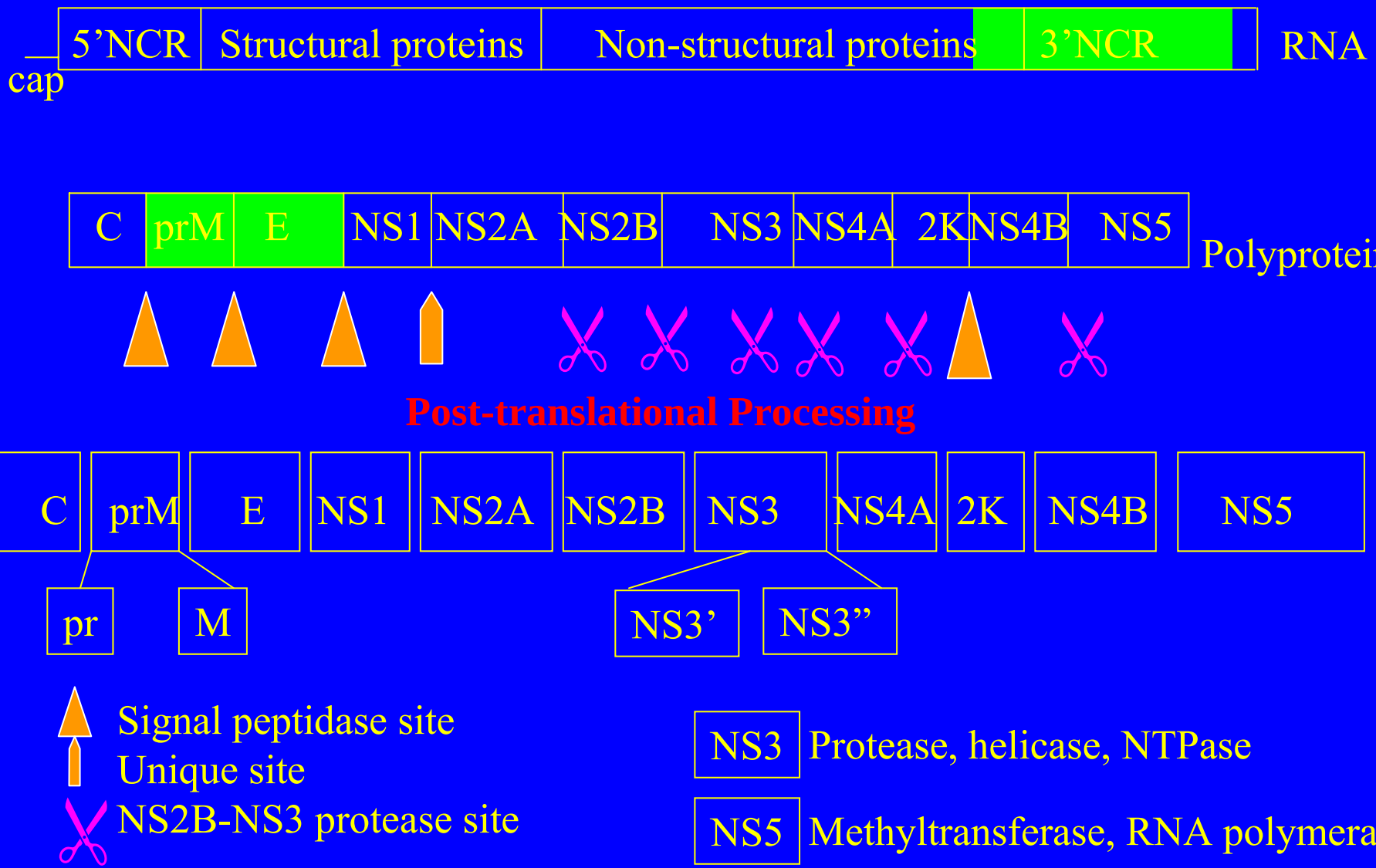
Alan D. T. Barrett

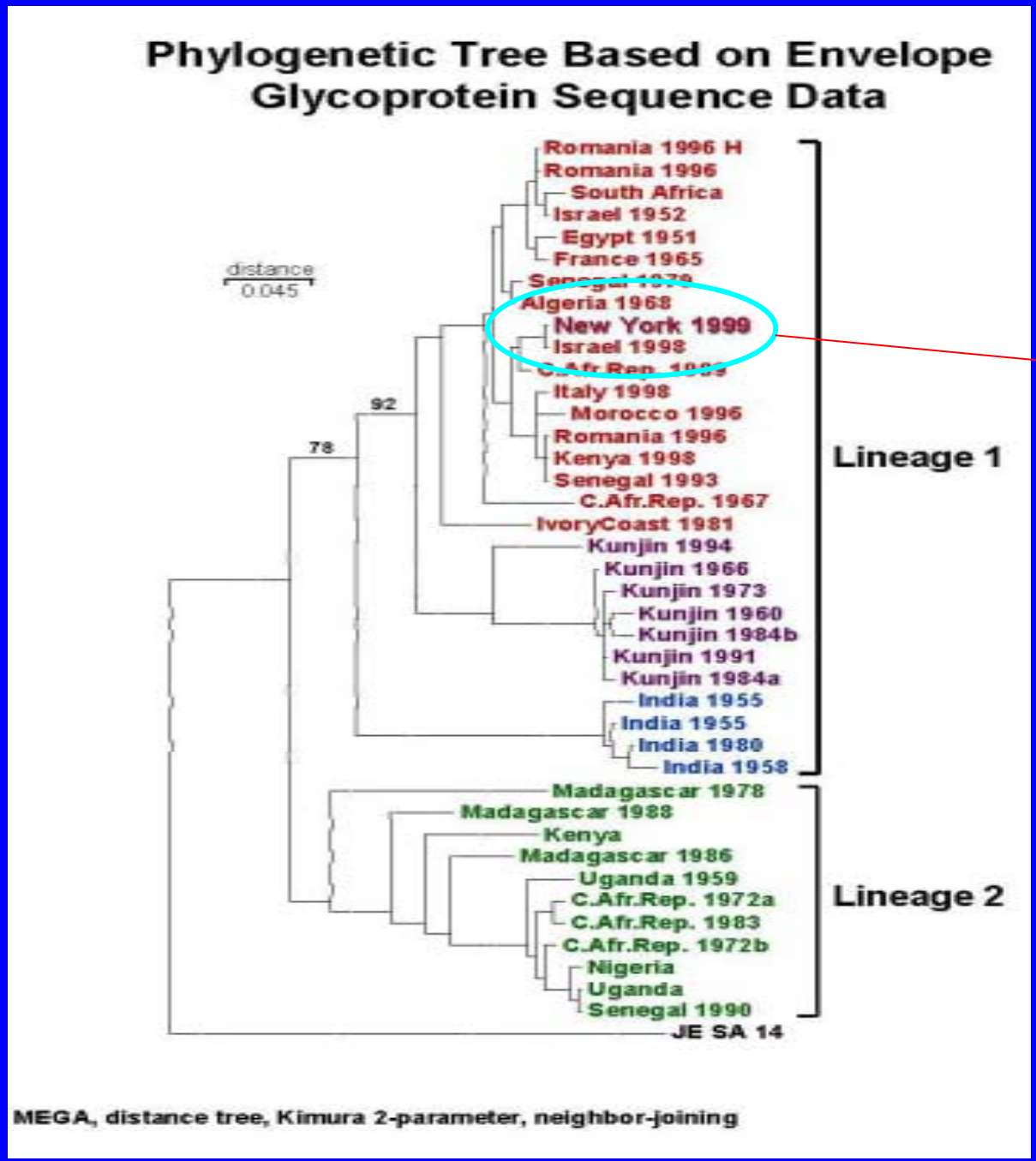
Department of Pathology,
Sealy Center for Vaccine Development,
Center for Biodefense and Emerging Infectious Diseases,
University of Texas Medical Branch at Galveston

Flavivirus Genome

- ss (+) RNA genome → mRNA
- Approximately 11 kb
- 5'-m^7GpppAmp cap
- Lacks 3'-polyA tail
- Codes for
  - 3 structural proteins
    - Capsid (C), membrane (prM/M), envelope (E)
  - 7 non-structural proteins
    - NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5


5'NCR
Structural proteins
Non-structural proteins
3'NCR
cap
RNA
C
prM
E
NS1
NS2A
NS2B
NS3
NS4A
2K
NS4B
NS5
Polyprotein
Co- and Post-translational Processing
pr
M
NS3'
NS3"
Signal peptidase site
Unique site
NS2B-NS3 protease site
NS3
Protease, helicase, NTPase
NS5
Methyltransferase, RNA polymerase


cap
5'NCR
Structural proteins
Non-structural proteins
3'NCR
RNA
C
prM
E
NS1
NS2A
NS2B
NS3
NS4A
2K
NS4B
NS5
Polyprotein
Post-translational Processing
C
prM
E
NS1
NS2A
NS2B
NS3
NS4A
2K
NS4B
NS5
pr
M
NS3'
NS3"
Signal peptidase site
Unique site
NS2B-NS3 protease site
NS3
Protease, helicase, NTPase
NS5
Methyltransferase, RNA polymerase



Lanciotti et al. 1999. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern U.S. [Science 286:2333-337.]

≥ 99.8% nucleotide homology with Israel 1998 over complete genome

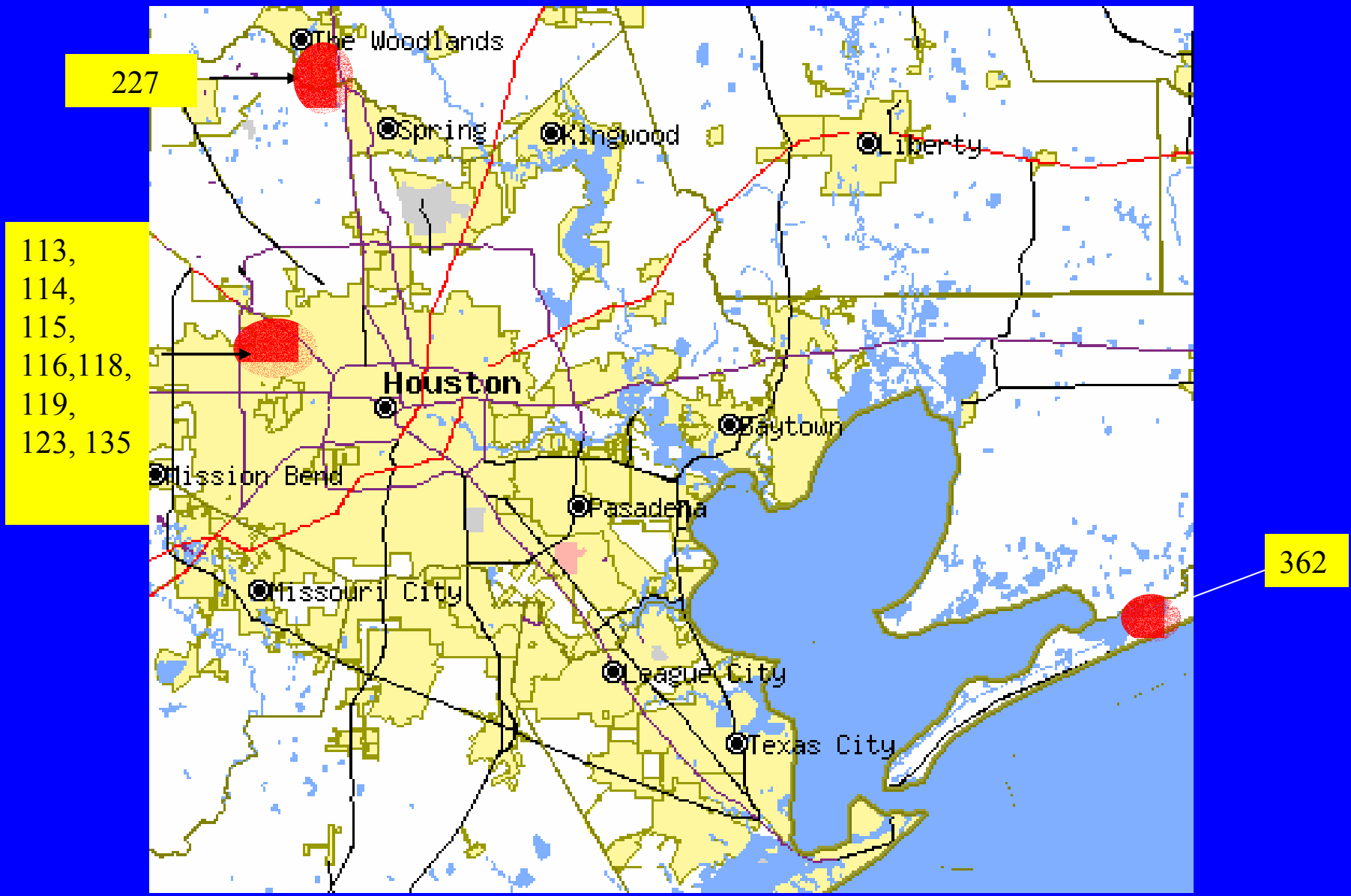
Lanciotti et al. 2002. Virology

Phylogenetic studies of North American WN viruses

- Anderson et al., PNAS 2001 and Ebel et al., EID 2001
  - Partial nucleotide sequences (921 and 1,503 respectively) with 3 nucleotide mutations and 2 amino acid substitutions
- Lanciotti et al., Virology 2002
  - Complete genomes of 1999 and 2000 isolates ≥99.8% identical by nucleotides and ≥99.9% amino acids to WN-NY99
- Huang et al., EID 2002
  - Complete genome of human isolate from 2001: ≥ 99.7% nucleotide homology and ≥99.8% amino acid homology
  - **No sequence analyses of WNV isolates from 2002 had been reported and none were reported outside of the northeastern seaboard.**

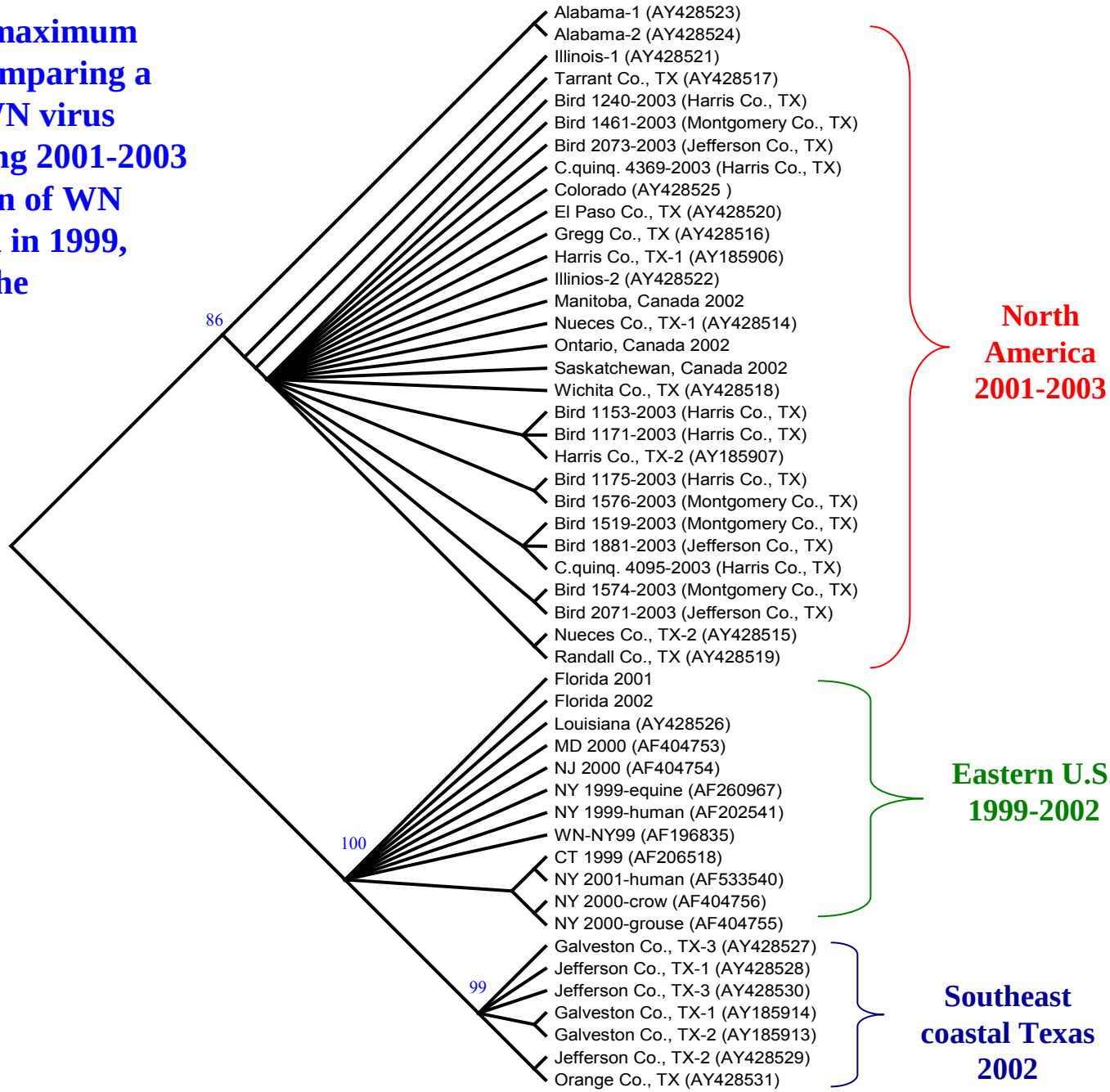
WN virus in Texas, 2002
(Beasley et al., 2003)

- First isolated by Bob Tesh (UTMB), Ray Parsons (Harris County) and Pushker Raj (TDH) in Houston, June 2002.
- Two "genetic variants"; one in Harris County and one on Bolivar Peninsula
- Harris County: 0.18% nucleotide divergence from New York 1999 strain 382-99.
- Bolivar Peninsula: 0.35% nucleotide divergence from New York 1999 strain 382-99.
- Harris County and Bolivar Peninsula isolates differ by 0.5%.

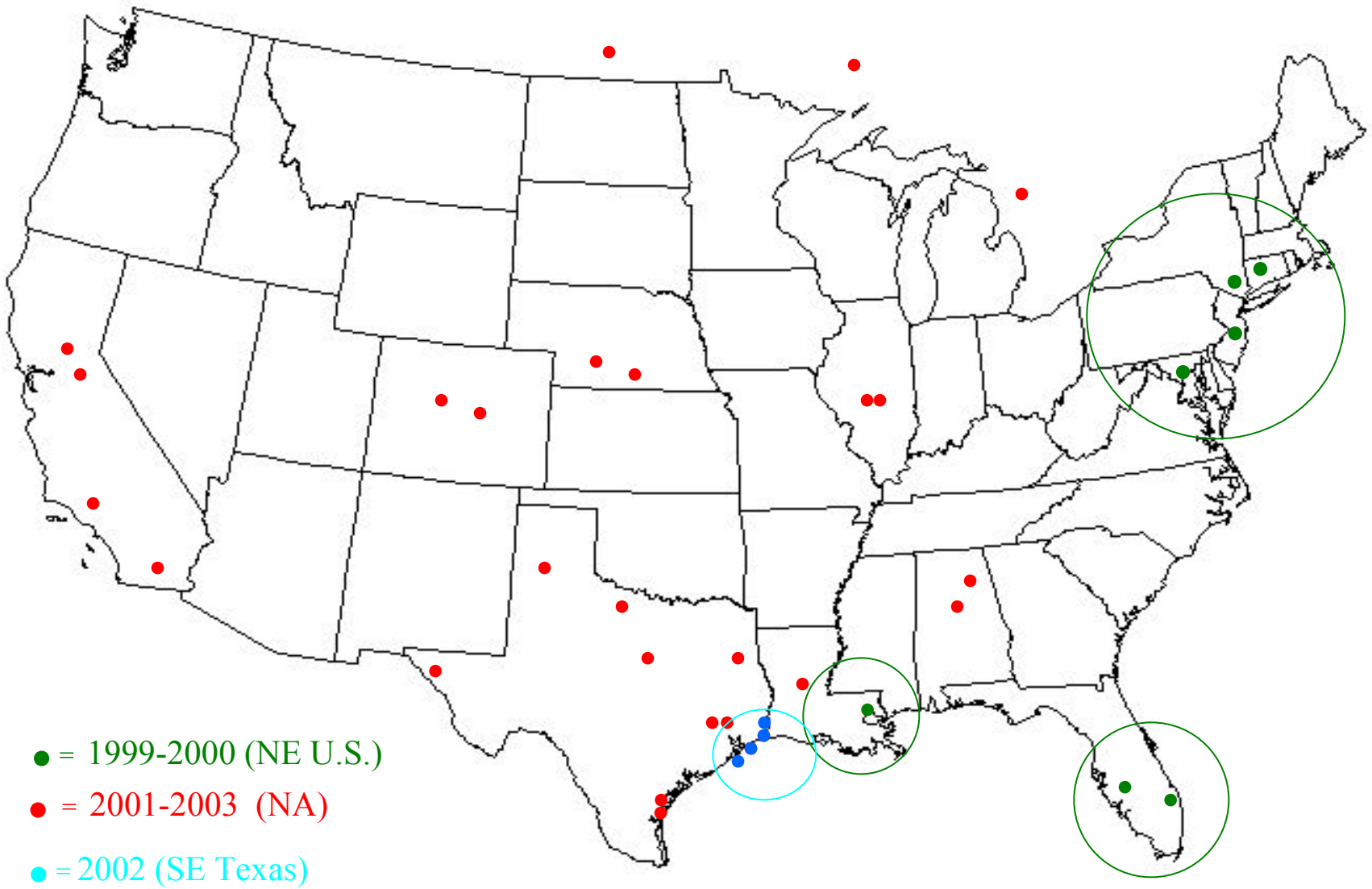

227
113,
114,
115,
116,118,
119,
123, 135
362
The Woodlands
Spring
Kingwood
Liberty
Houston
Baytown
Mission Bend
Pasadena
Missouri City
League City
Texas City

Cladogram based on maximum parsimony analysis comparing a 2004-nt sequence of WN virus isolates collected during 2001-2003 to a homologous region of WN virus isolates collected in 1999, 2000, and 2001 from the northeastern U.S.

(Davis et al., 2003)



Genetic distribution of WNV isolates, 1999-2003



SE coastal Texas strain
(Granwehr et al., 2004)

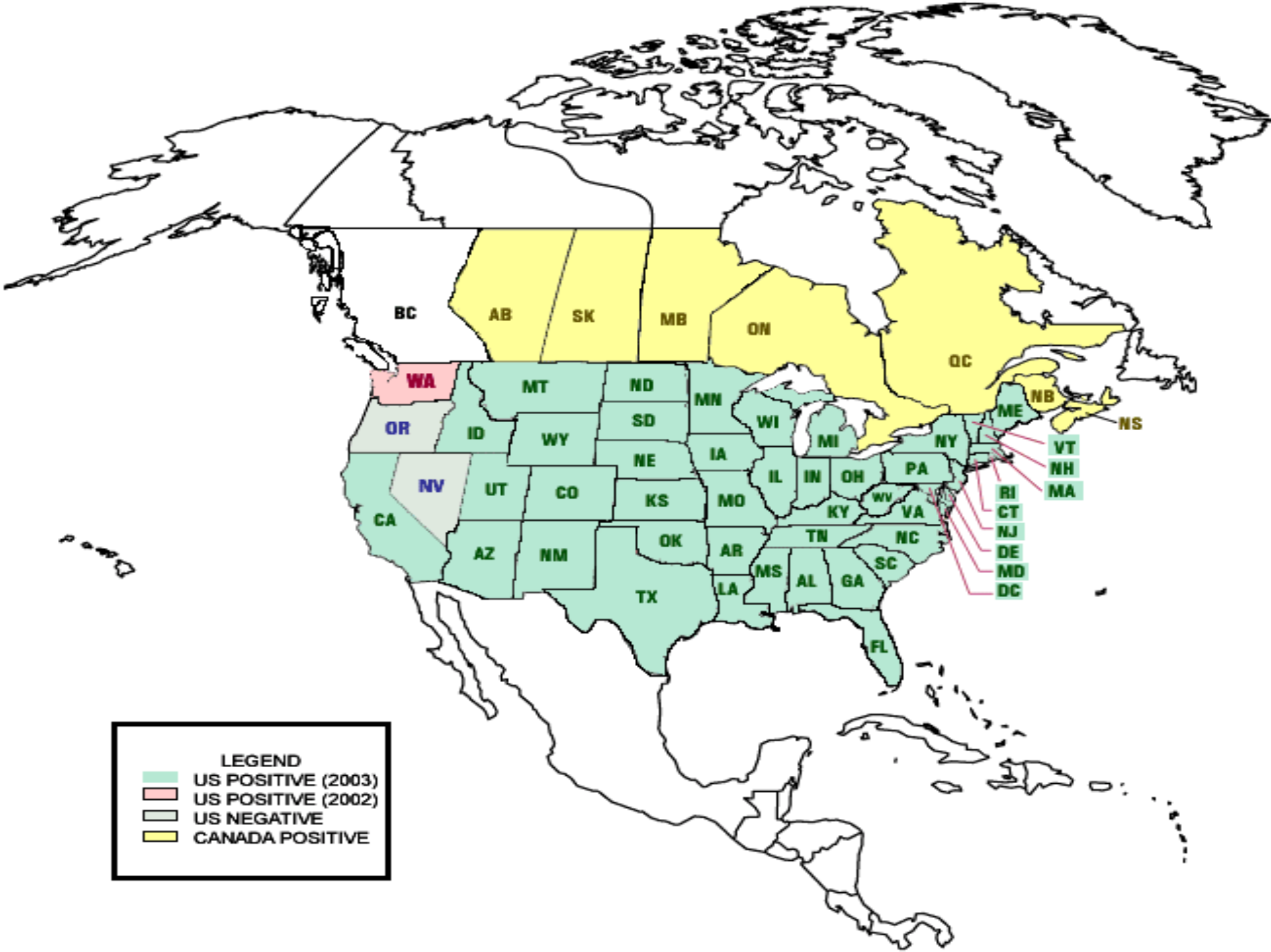
Residue	SE coastal Texas	Lineage I	Lineage II
E-76	Ala	Thr	Thr
NS1-94	Gly	Glu	Asn
NS2A-90	Met	Met	Lys
NS2A-138	Ile	Val	Val
NS4B-173	Ile	Val	Val
NS5-526	Ile	Thr	Thr

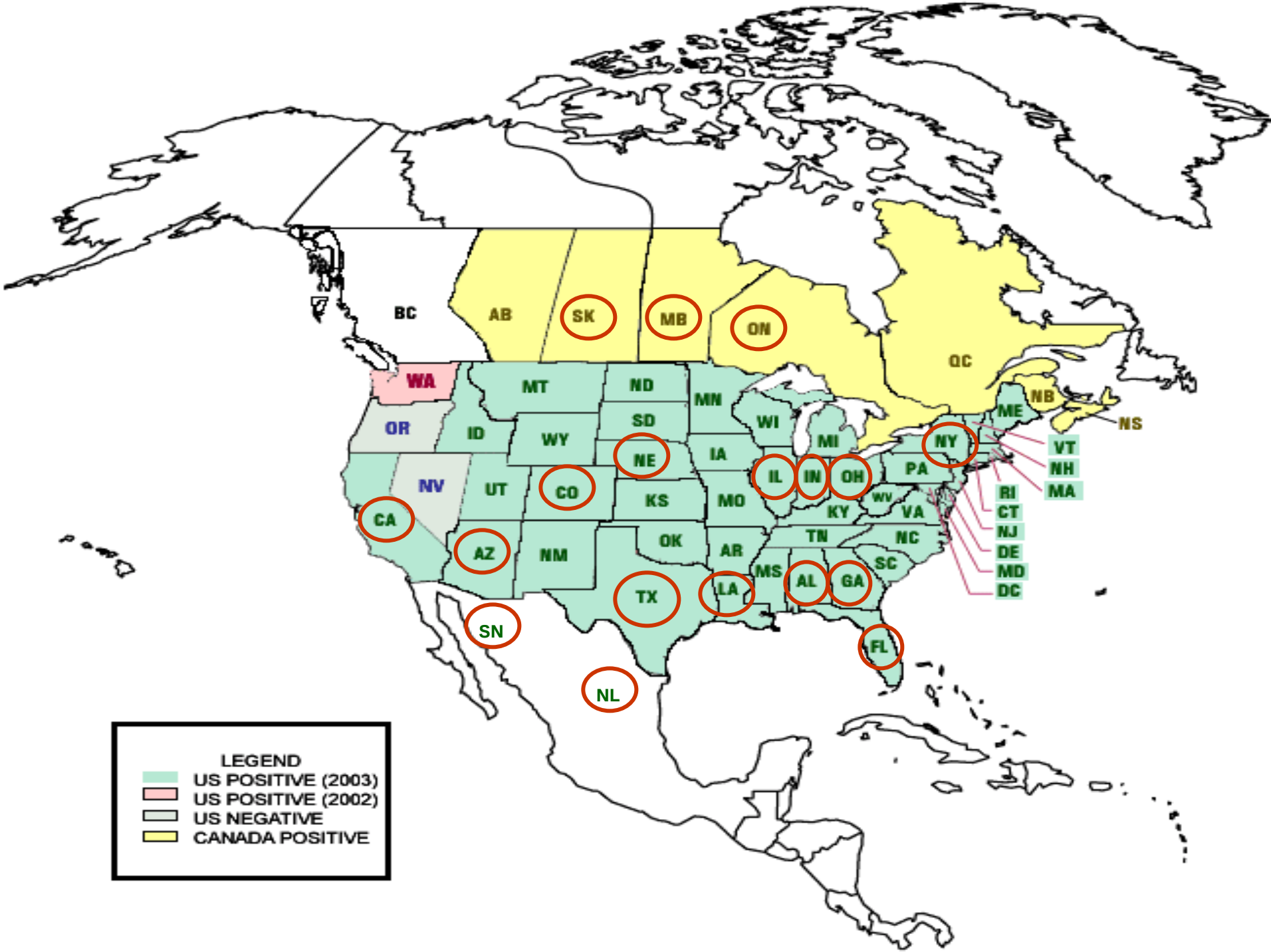
Overview of sequence studies

- Multiple introductions into TX during summer of 2002 and geographical clustering of variants (coastal vs. state-wide variants)
    - Long-distance spread by birds with limited migration patterns (ie. circum-Gulf route)
    - Localized spread by birds/mosquitoes over shorter-distance
- Identified nucleotide mutations at 31 positions (8 in prM, 23 in E) not previously observed in North American WN virus strains (no changes shared with NE United States strains)
- 7 amino acid substitutions (2 in prM, 5 in E)
- Revealed emergence of two distinct variants: 2001-2002 North America variant and coastal SE Texas variant
- Most divergent strains to date: individual isolates with 7 nucleotide mutations and 3 amino acid substitutions.
- Maximum nucleotide divergence from WN-NY99 was 0.35% (average of 0.18%)

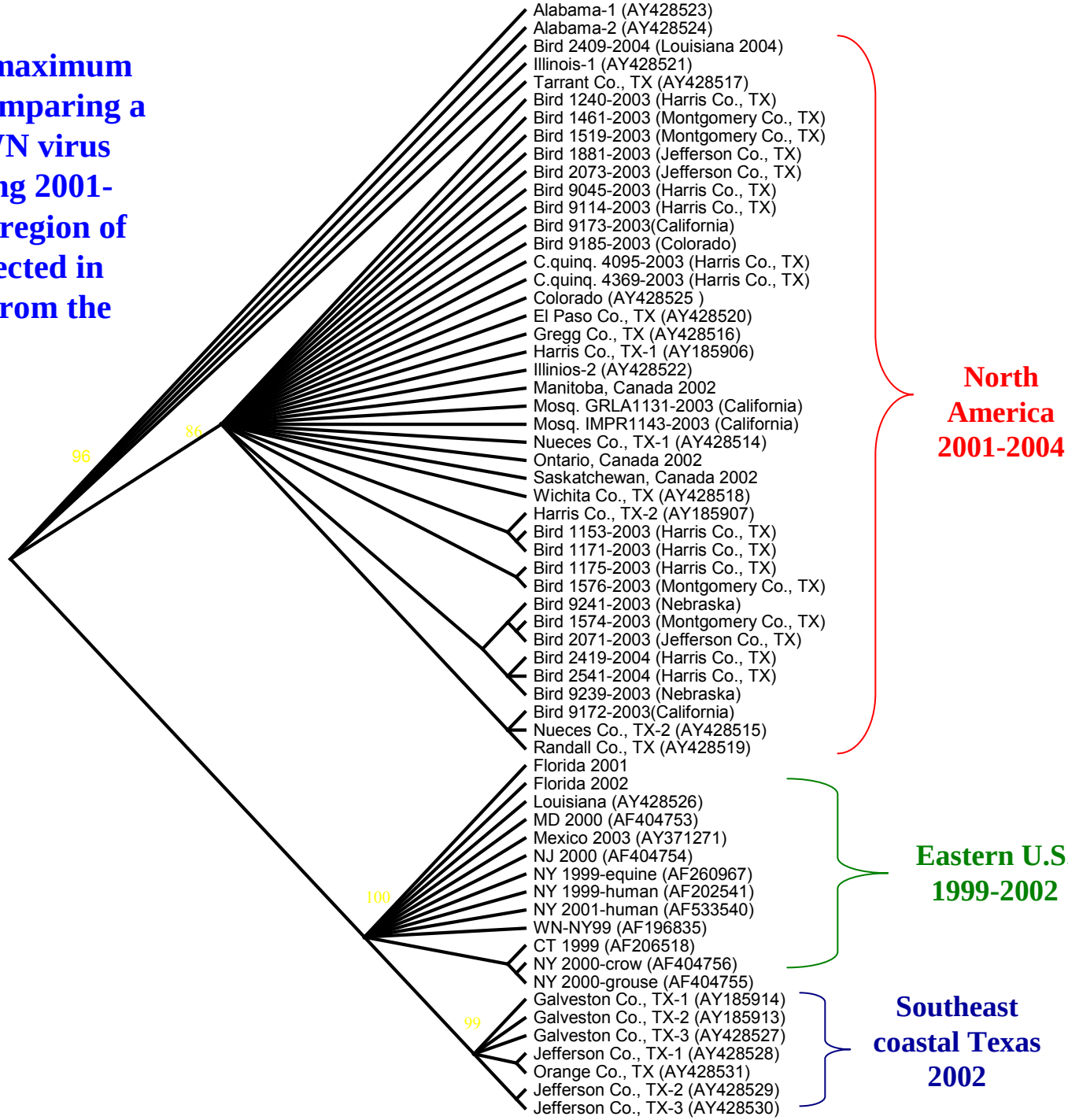
Evolution of WN virus in the United States

- Evidence suggests limited evolution between 1999 and 2002.
- Majority of nucleotide changes are transitions (U ←→ C).
- Strain isolated in 1999 and 2000 in NY, NJ, CT and MD differ ≤ 0.35% from prototype New York strain 382-99.
- SE coastal Texas isolates are the most divergent to date (0.55%).
- Distinct populations of virus evolving?
- No strong selection → genetic drift?

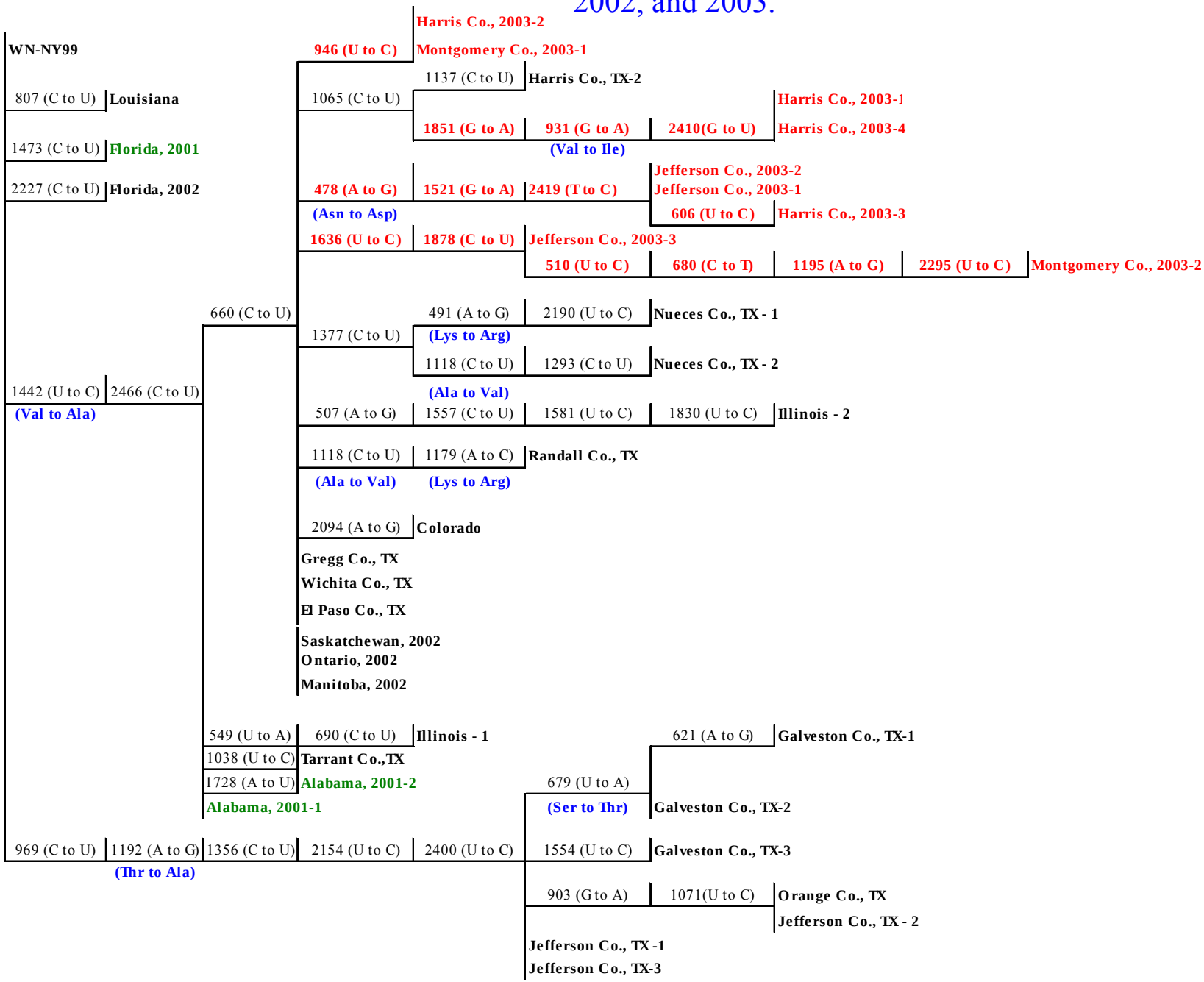

BC
AB
SK
MB
ON
QC
NB
NS
WA
OR
CA
NV
ID
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
ME
VT
NH
MA
RI
CT
NJ
DE
MD
DC
LEGEND
US POSITIVE (2003)
US POSITIVE (2002)
US NEGATIVE
CANADA POSITIVE


BC
AB
SK
MB
ON
QC
NB
NS
ME
VT
NH
MA
RI
CT
NJ
DE
MD
DC
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
NY
PA
VA
NC
SC
FL
SN
NL
LEGEND
US POSITIVE (2003)
US POSITIVE (2002)
US NEGATIVE
CANADA POSITIVE

Cladogram based on maximum parsimony analysis comparing a 2004-nt sequence of WN virus isolates collected during 2001-2004 to a homologous region of WN virus isolates collected in 1999, 2000, and 2001 from the northeastern U.S.



Phylogram comparing a 2004-nt sequence of WN-NY99 with 35 WN virus isolates collected during 2001, 2002, and 2003.



Bayesian analysis of prM and E genes of 115 North American WNV isolates. The following criterion were used in analysis:

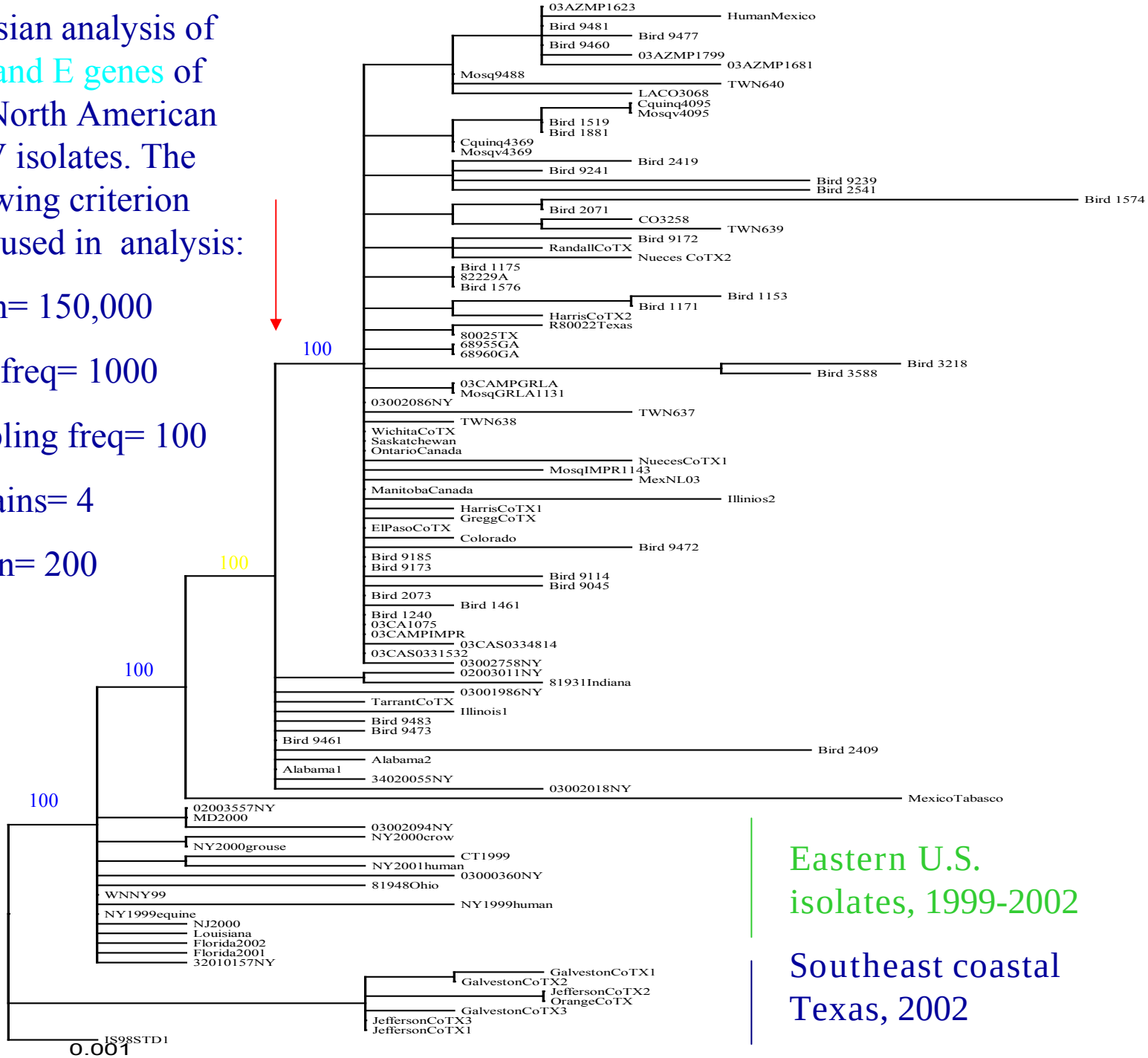
N gen= 150,000

Print freq= 1000

Sampling freq= 100

N chains= 4

burnin= 200



North America genotype

Eastern U.S. isolates, 1999-2002

Southeast coastal Texas, 2002

Bayesian analysis of prM and E genes of 115 North American WNV isolates. The following criterion were used in analysis:

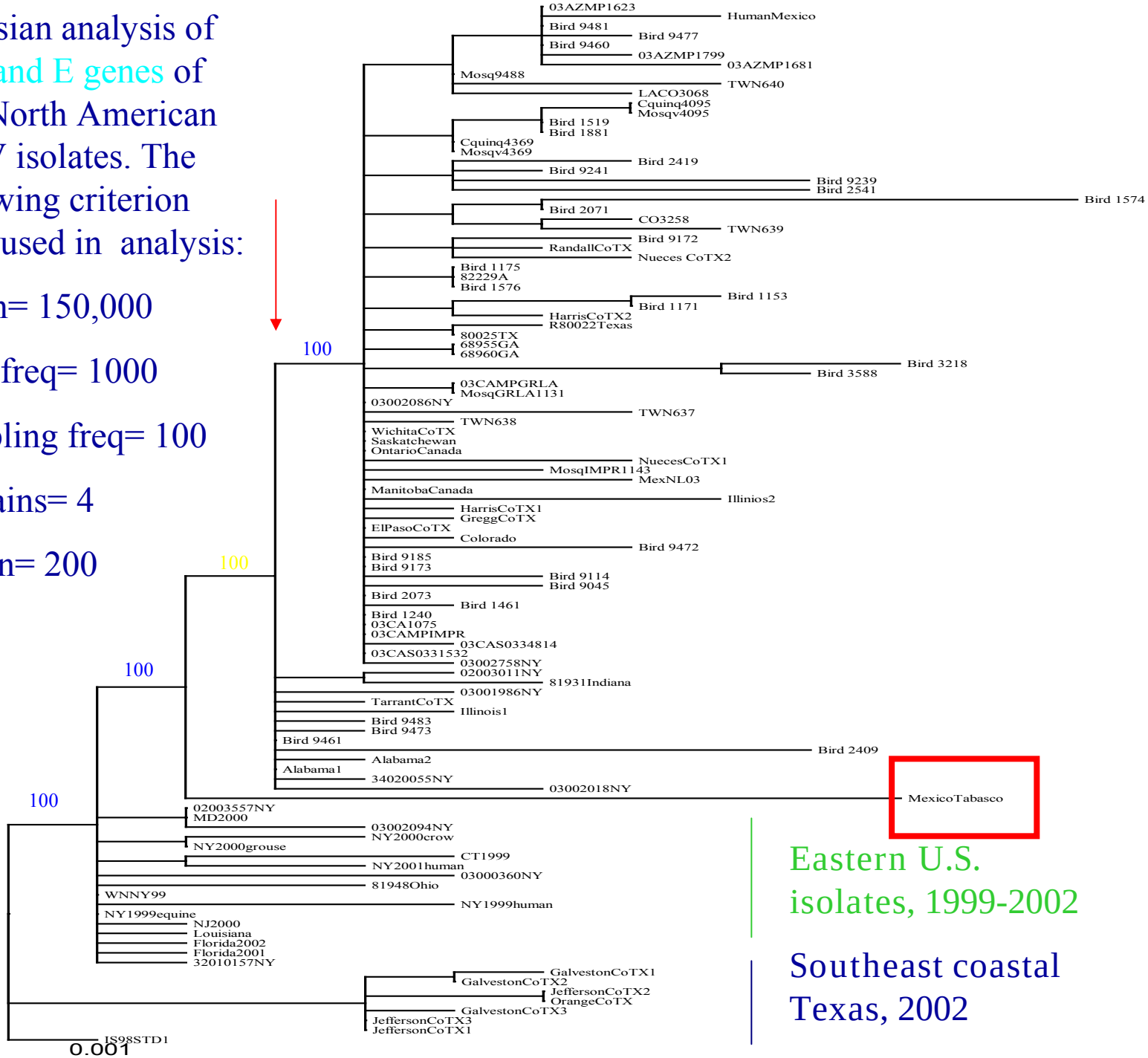
N gen= 150,000

Print freq= 1000

Sampling freq= 100

N chains= 4

burnin= 200

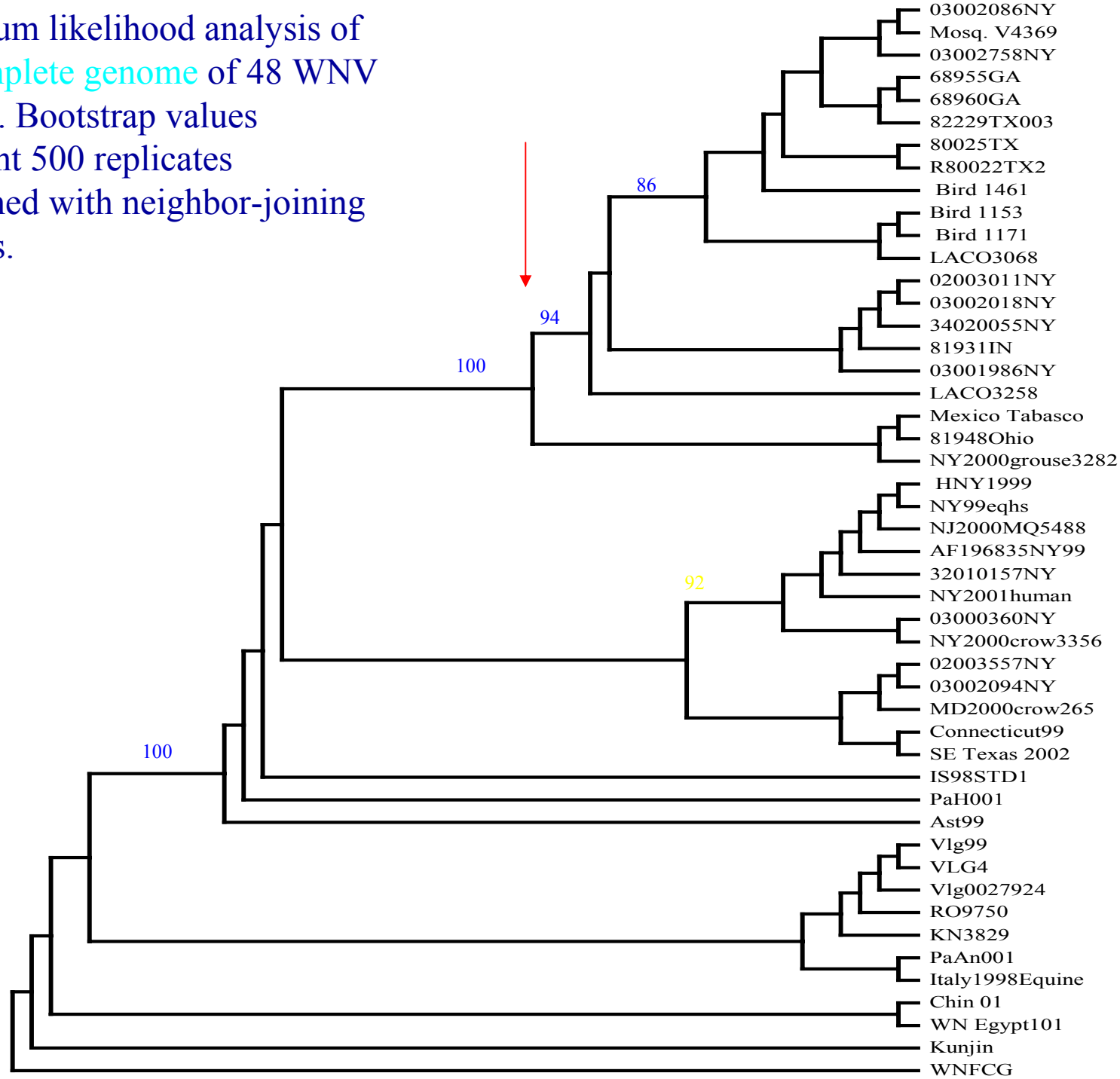


North America genotype

Eastern U.S. isolates, 1999-2002

Southeast coastal Texas, 2002

Maximum likelihood analysis of the complete genome of 48 WNV isolates. Bootstrap values represent 500 replicates performed with neighbor-joining analysis.

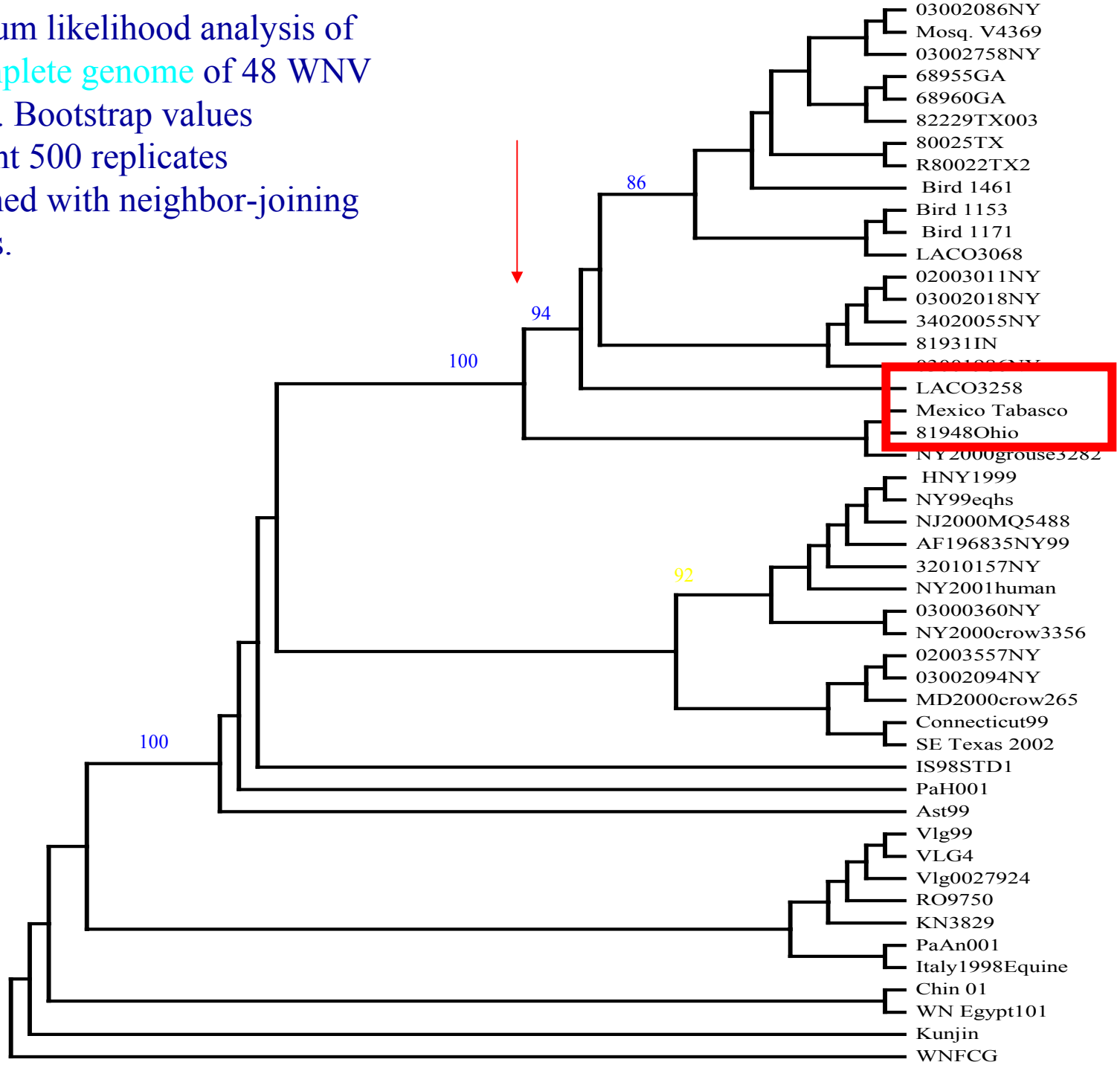


North America genotype

Eastern U.S. isolates, 1999-2002

Old World WNV isolates

Maximum likelihood analysis of the complete genome of 48 WNV isolates. Bootstrap values represent 500 replicates performed with neighbor-joining analysis.



West Nile in the Americas

- United States
- Canada
- Mexico
- Dominican Republic
- El Salvador
- Jamaica

Genomic sequence of TM-171 Mex03 isolate

(Beasley et al., 2004)

Isolate from dead raven at wildlife reserve in Villahermosa, Tabasco.

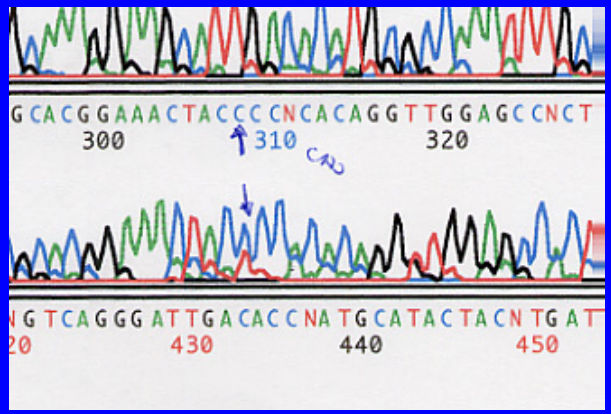
RNA and, subsequently, Vero cell passaged virus sent to UTMB.

46 nucleotide differences (0.42%) from NY99; 4 amino acid differences:

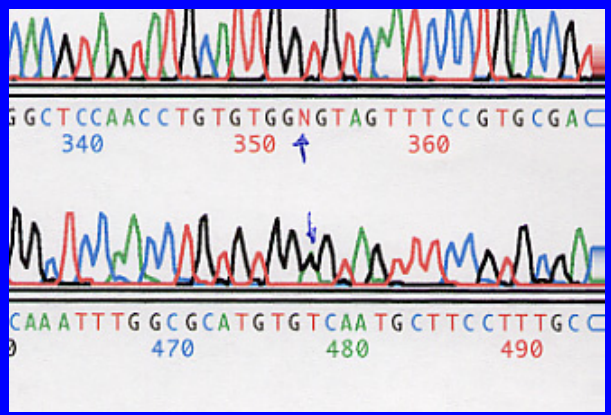
prM/M-141	Ile → Thr
E-156	Ser → Pro* (loss of glycosylation motif)
NS4B-245	Ile → Val*
NS5-898	Thr → Ile*

Mixed sequence at the E glycosylation site of Mex03 isolate.

Forward



Reverse



Consensus sequences:

NY99 AAC TAC TCC ACA (NYST)
MEX AAC TAC CCC ACA (NYPT)

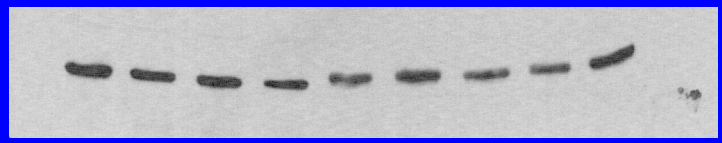
Cloned Mex03 PCR product and sequenced 5 clones:

- three clones "CCC" – Pro
- two clones "TCC" – Ser

Plaque purification of E glycosylation variants of Mex03.

- Variants selected by two rounds of plaque purification in Vero cells.

- Screened by Western blot (diff. in apparent mol. wt.)



- Selection of variants confirmed by nucleotide sequencing.

- gly+ variants grew faster (~24hrs earlier to harvest) and to ~10-fold higher titer, and had slightly larger plaque size.

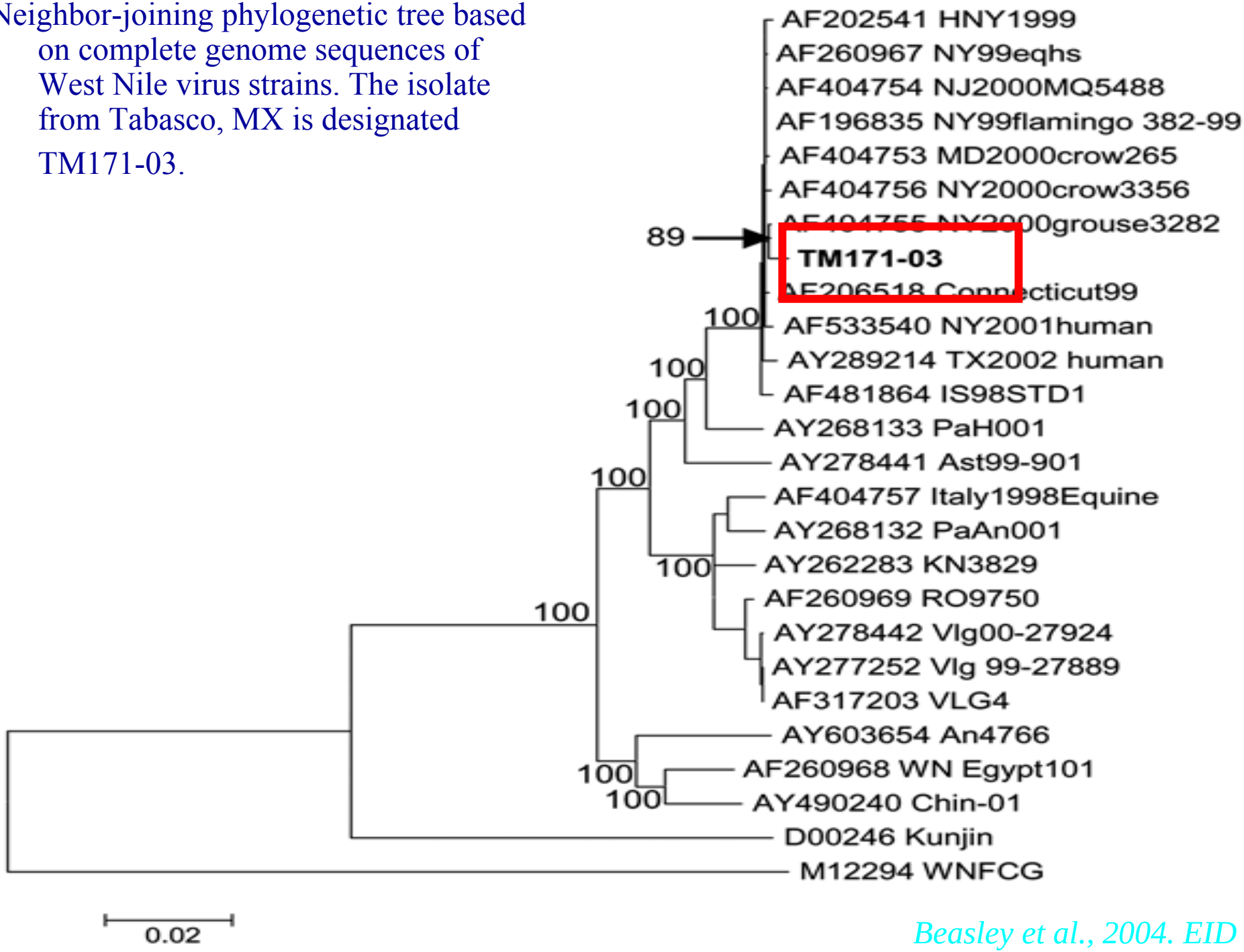
Mouse virulence of Mex03 glycosylation variants.

Virus	E154-156	i.p. LD_{50} (pfu)	A.S.T. ± s.d. (days)	PD_{50} (pfu)
Mex03	mixed	0.5	8.2 ± 1.8	n.d.
Mex03-pp1	NYP	>1000	--	5.0
Mex03-pp2	NYP	794	10.3 ± 0.6	0.8
Mex03-pp5	NYS	3.2	8.5 ± 0.8	0.5
Mex03-pp6	NYS	2.0	9.2 ± 1.0	2.0

n.d. – not determined

All strains contained prM-141 I→T; NS4B-245 I→V; NS5-898 T→I

Neighbor-joining phylogenetic tree based on complete genome sequences of West Nile virus strains. The isolate from Tabasco, MX is designated TM171-03.



Beasley et al., 2004. EID

Conclusions

- Evidence suggests limited, but continuing, divergence from isolates collected in 1999 and 2000
- Emergence of genetically distinct variants
    - Dominant North America variant since at least 2002
    - Coastal SE Texas variant in 2002 only → become extinct?
- Dominant variant has emerged throughout the majority of North America with >90% of isolates collected in 2002 and after belonging to the dominant clade
  - Emergence of novel genotype corresponds with displacement or extinction of earlier genotypes
- Subclades are readily illustrated depending on year of isolation and collection location
  - Higher degree of nucleotide identity within individual states and during the same transmission season
- Mexico 2003 isolate is most divergent from the Americas with 9 nt and 2 aa differences from New York 1999 isolates and has lost E protein glycosylation site.
- Genetic mutations could eventually lead to phenotypic changes in viral antigenicity or associated virulence (mouse model shows no attenuation in either variant)?

Phenotypic variants – Texas 2003

(Davis et al., 2004)

Nucleotide sequencing and phenotypic comparisons of 29 WNV isolates collected by Harris Co. Mosquito Control Division in and around Houston, TX between 9 May – 8 Sept. 2003:

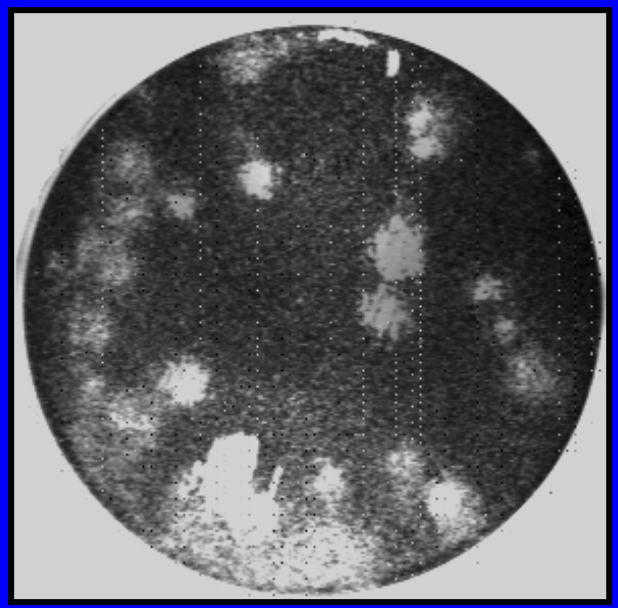
- 17 from dead birds
- 12 from mosquito pools (*Culex spp.*)

Viruses isolated in Vero cells, amplified by one additional passage and plaque titrated on Vero cells.

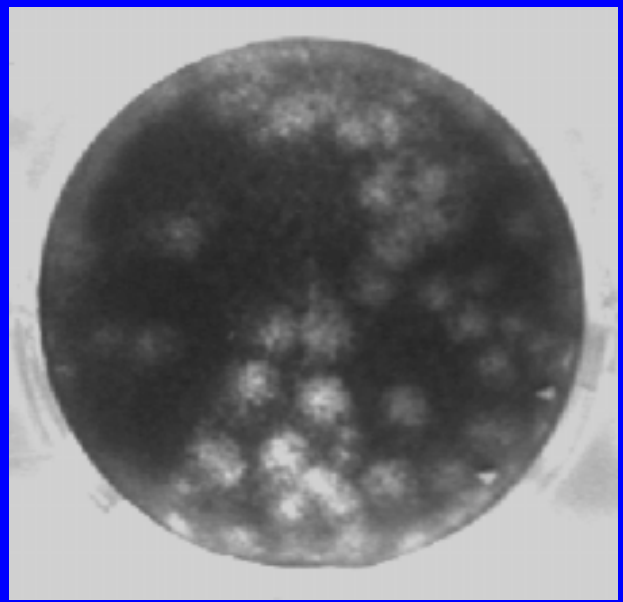
6 strains (5 bird, 1 mosq.) with small plaque (SP) morphology

Plaque morphology of WNV isolates

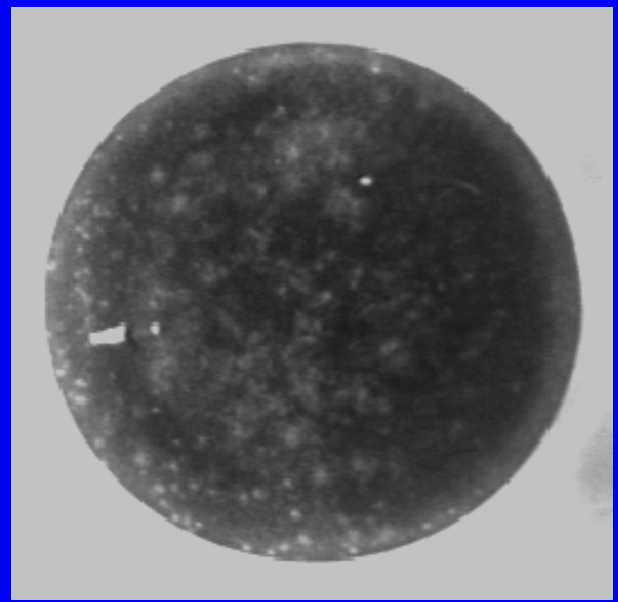
WN-NY99 (382-99)



WNV 2002



WNV 2003 sp



Mean plaque size ≥ 1.5mm

Mean plaque size < 1.0mm

In Vero cells, 72 hours post-infection.

Stained with crystal violet.

Phenotypic comparisons of LP/SP WNV strains

Small plaque phenotype often a marker of attenuation of flaviviruses.

Other phenotypic comparisons:

- temperature sensitivity at 39.5°C
- mouse neuroinvasion

  - i.p. inoculation of 3-4 week old Swiss Webster mice

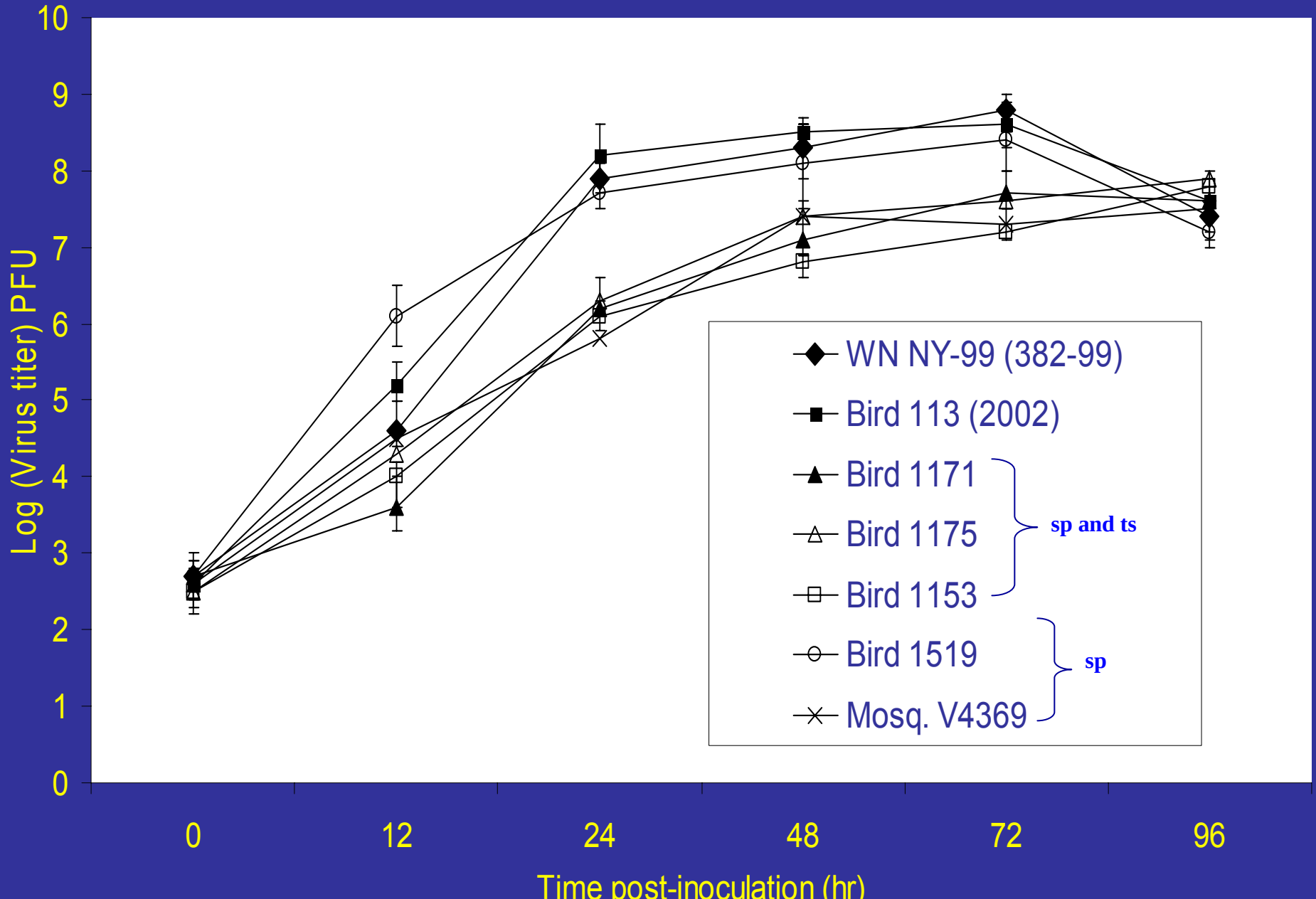
Phenotypic comparisons of LP/SP WNV strains (cont.)

Strain	Source	Plaque	TS index*	ip LD_{50} (pfu)	AST ± sd	ic LD_{50} (pfu)
382-99	Flamingo	L	-0.5	0.8	7.2±0.6	0.1
TWN93 (02)	Bird	L	0.3	0.5	8.0±1.0	0.1
TWN301 (03)	Bird	L	-0.2	0.6	7.0±1.0	0.2
TWN305 (03)	Bird	S	0.3	51,000	9.0±4.0	0.1
TWN274 (03)	Bird	S	-2.7	23,000	9.5±1.0	0.3
TWN382 (03)	Mosq.	S	-1.8	645,000	8.3±3.0	0.1
TWN269 (03)	Bird	S	-2.7	2,000	9.7±3.3	nd

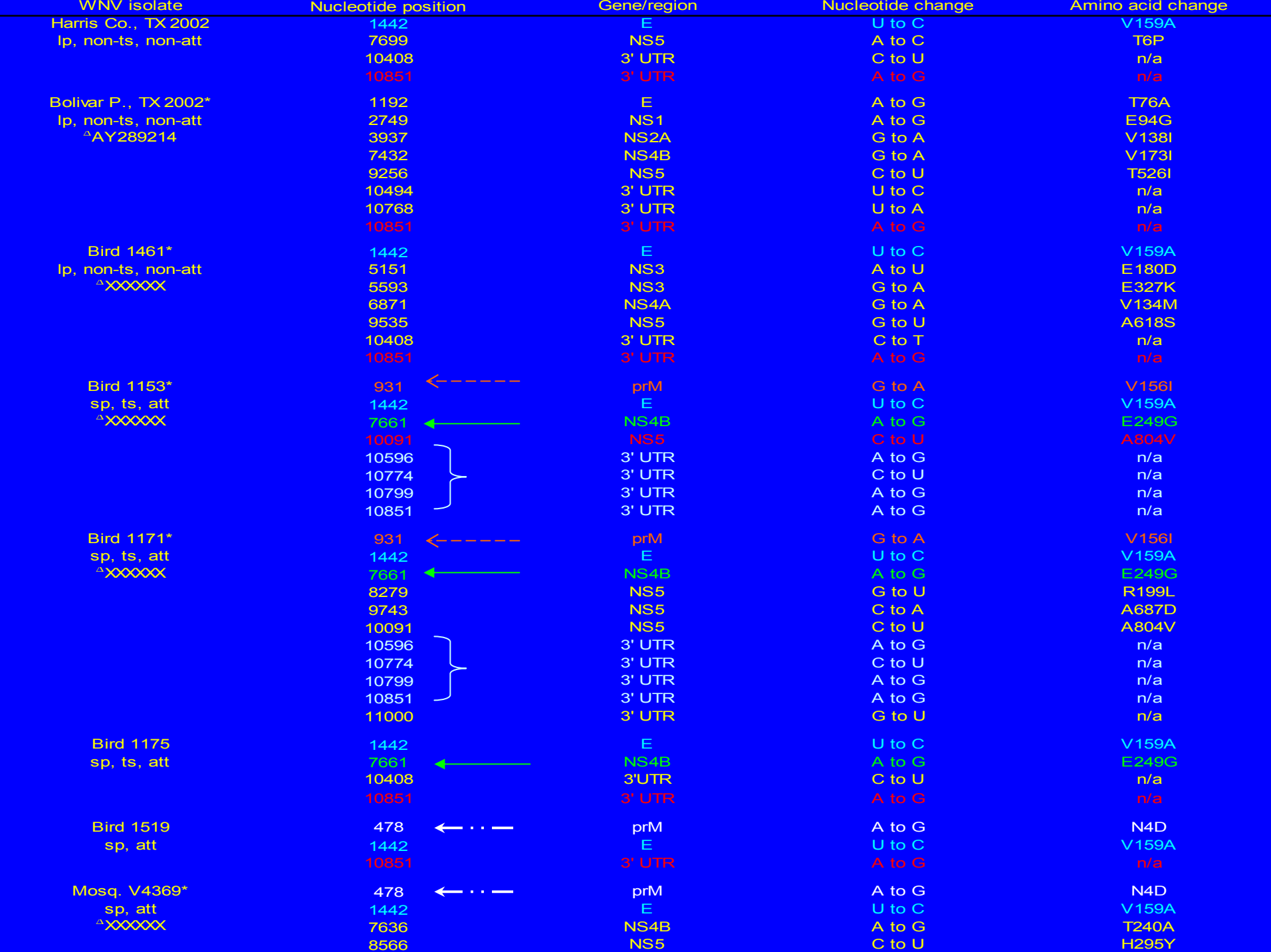
*TS index is $\log_{10}$ difference in virus titer at 39.5°C compared with 37°C; negative values indicate decreased titers.

nd – not determined; L = large, S = small.

Viral growth curve in Vero cells of 2003 WNV isolates in comparison to isolates from 1999 and 2002



WNV isolate	Nucleotide position	Gene/region	Nucleotide change	Amino acid change
Harris Co., TX 2002	1442	E	U to C	V159A
lp, non-ts, non-att	7699	NS5	A to C	T6P
	10408	3' UTR	C to U	n/a
	10851	3' UTR	A to G	n/a
Bolivar P., TX 2002*	1192	E	A to G	T76A
lp, non-ts, non-att	2749	NS1	A to G	E94G
ΔAY289214	3937	NS2A	G to A	V138I
	7432	NS4B	G to A	V173I
	9256	NS5	C to U	T526I
	10494	3' UTR	U to C	n/a
	10768	3' UTR	U to A	n/a
	10851	3' UTR	A to G	n/a
Bird 1461*	1442	E	U to C	V159A
lp, non-ts, non-att	5151	NS3	A to U	E180D
ΔXXXXXX	5593	NS3	G to A	E327K
	6871	NS4A	G to A	V134M
	9535	NS5	G to U	A618S
	10408	3' UTR	C to T	n/a
	10851	3' UTR	A to G	n/a
Bird 1153*	931	prM	G to A	V156I
sp, ts, att	1442	E	U to C	V159A
ΔXXXXXX	7661	NS4B	A to G	E249G
	10091	NS5	C to U	A804V
	10596	3' UTR	A to G	n/a
	10774	3' UTR	C to U	n/a
	10799	3' UTR	A to G	n/a
	10851	3' UTR	A to G	n/a
Bird 1171*	931	prM	G to A	V156I
sp, ts, att	1442	E	U to C	V159A
ΔXXXXXX	7661	NS4B	A to G	E249G
	8279	NS5	G to U	R199L
	9743	NS5	C to A	A687D
	10091	NS5	C to U	A804V
	10596	3' UTR	A to G	n/a
	10774	3' UTR	C to U	n/a
	10799	3' UTR	A to G	n/a
	10851	3' UTR	A to G	n/a
	11000	3' UTR	G to U	n/a
Bird 1175	1442	E	U to C	V159A
sp, ts, att	7661	NS4B	A to G	E249G
	10408	3'UTR	C to U	n/a
	10851	3' UTR	A to G	n/a
Bird 1519	478	prM	A to G	N4D
sp, att	1442	E	U to C	V159A
	10851	3' UTR	A to G	n/a
Mosq. V4369*	478	prM	A to G	N4D
sp, att	1442	E	U to C	V159A
ΔXXXXXX	7636	NS4B	A to G	T240A
	8566	NS5	C to U	H295Y


WNV isolate
Nucleotide position
Gene/region
Nucleotide change
Amino acid change
Harris Co., TX 2002
lp, non-ts, non-att
1442 E U to C V159A
7699 NS5 A to C T6P
10408 3' UTR C to U n/a
10851 3' UTR A to G n/a
Bolivar P., TX 2002*
lp, non-ts, non-att
ΔAY289214
1192 E A to G T76A
2749 NS1 A to G E94G
3937 NS2A G to A V138I
7432 NS4B G to A V173I
9256 NS5 C to U T526I
10494 3' UTR U to C n/a
10768 3' UTR U to A n/a
10851 3' UTR A to G n/a
Bird 1461*
lp, non-ts, non-att
ΔXXXXXX
1442 E U to C V159A
5151 NS3 A to U E180D
5593 NS3 G to A E327K
6871 NS4A G to A V134M
9535 NS5 G to U A618S
10408 3' UTR C to T n/a
10851 3' UTR A to G n/a
Bird 1153*
sp, ts, att
ΔXXXXXX
931 prM G to A V156I
1442 E U to C V159A
7661 NS4B A to G E249G
10091 NS5 C to U A804V
10596 3' UTR A to G n/a
10774 3' UTR C to U n/a
10799 3' UTR A to G n/a
10851 3' UTR A to G n/a
Bird 1171*
sp, ts, att
ΔXXXXXX
931 prM G to A V156I
1442 E U to C V159A
7661 NS4B A to G E249G
8279 NS5 G to U R199L
9743 NS5 C to A A687D
10091 NS5 C to U A804V
10596 3' UTR A to G n/a
10774 3' UTR C to U n/a
10799 3' UTR A to G n/a
10851 3' UTR A to G n/a
11000 3' UTR G to U n/a
Bird 1175
sp, ts, att
1442 E U to C V159A
7661 NS4B A to G E249G
10408 3'UTR C to U n/a
10851 3' UTR A to G n/a
Bird 1519
sp, att
478 prM A to G N4D
1442 E U to C V159A
10851 3' UTR A to G n/a
Mosq. V4369*
sp, att
ΔXXXXXX
478 prM A to G N4D
1442 E U to C V159A
7636 NS4B A to G T240A
8566 NS5 C to U H295Y

Conclusions

- First evidence of phenotypic variation in North American West Nile virus
- Attenuation of mouse neuroinvasiveness may not be indicative of attenuation in birds, horses, humans
- No indication that attenuated viruses persist in nature
  - No more SP/mouse attenuated isolates identified → another extinct lineage?
- Sequencing and reverse genetics studies can be used to identify molecular determinants of phenotypic variation

Genetic and phenotypic variation in New York, 2002
Ebel et al. 2004. AJTMH

- New genotype emerged in New York in 2002 (55% of all isolates in 2002 and 85% of all isolates in 2003)
- Genetically homogeneous to Texas isolates from 2002-2003 (prM and E)
- In vitro growth studies (C6/36 and Vero) showed no significant differences
- In vivo mosquito transmission studies (*Culex pipiens*)
  - Significantly higher proportion of mosquitoes became infected (infectious virus in bodies) and developed disseminated infections (infectious virus in legs) following feeding with dominant genotype compared to WN-NY99
  - Significantly higher proportion of mosquitoes were able to transmit (infectious virus in salivary secretions) at days 5 and 7 post-feed with dominant genotype
- Displacement of WN-NY99 genotype by dominant genotype may be due to differences in mosquito transmission efficiency (reduction of EIP with dominant genotype)

Acknowledgements

Barrett lab

David Beasley
Mike Holbrook
Todd Davis
Amber Engel
Li Li
Monica McArthur
Melissa Whiteman
Jason Wicker
Shuliu Zhang
Bruno Granwehr

Other UTMB

Bob Tesh + lab

Steve Higgs + lab

Scott Weaver + lab
Jose Estrada-Franco

Funding

CDC; State of Texas Advanced Research Program; Clayton Foundation; McLaughlin Foundation

Collaborators

- Harris County Mosquito Control Division
- Texas Department of Health
- CDC, Division of Vector-Borne Infectious Diseases
- Illinois Natural History Survey
- University of Alabama, Birmingham
- Louisiana Dept. of Health
- Florida Dept. of Health
- Colorado State University
- Nebraska Dept. of Health
- University of California, Davis
- New York State Dept. of Health
- Health Canada
- Mexico – Dept. de Salud Publico