Fighting the Tuberculosis Pandemic using Machine Learning

Kristin P. Bennett

Rensselaer Polytechnic Institute

TB-Insight Team

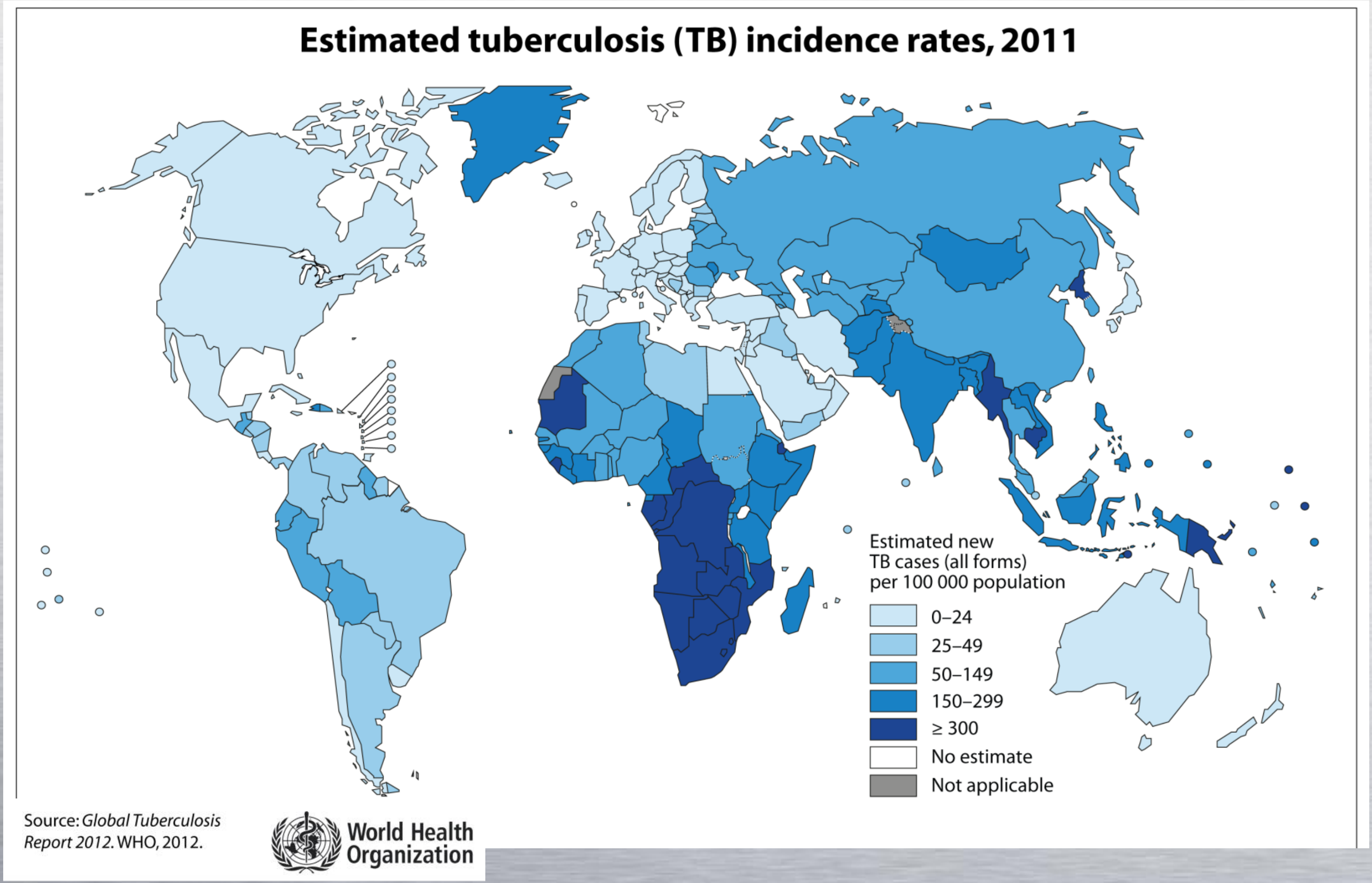
TB insight

Tuberculosis Tracking and Control

http://www.cs.rpi.edu/~bennek/tbinsight/index.html

This work is supported by NIH R01LM009731.

1/3 of World Latently Infected with TB



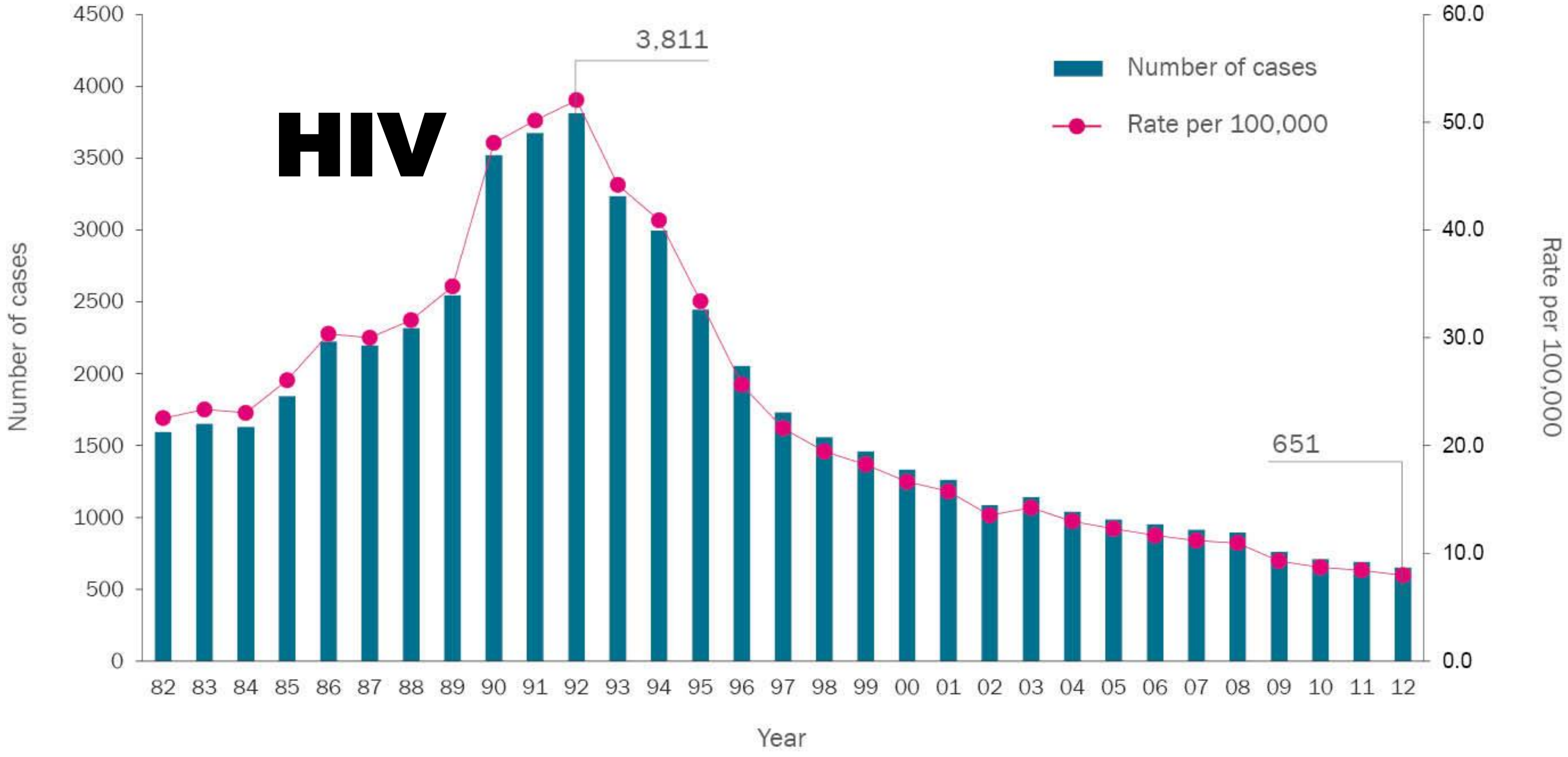
Estimated tuberculosis (TB) incidence rates, 2011

Estimated new TB cases (all forms) per 100 000 population
0–24
25–49
50–149
150–299
≥ 300
No estimate
Not applicable
World Health Organization

Source: *Global Tuberculosis Report 2012.* WHO, 2012.

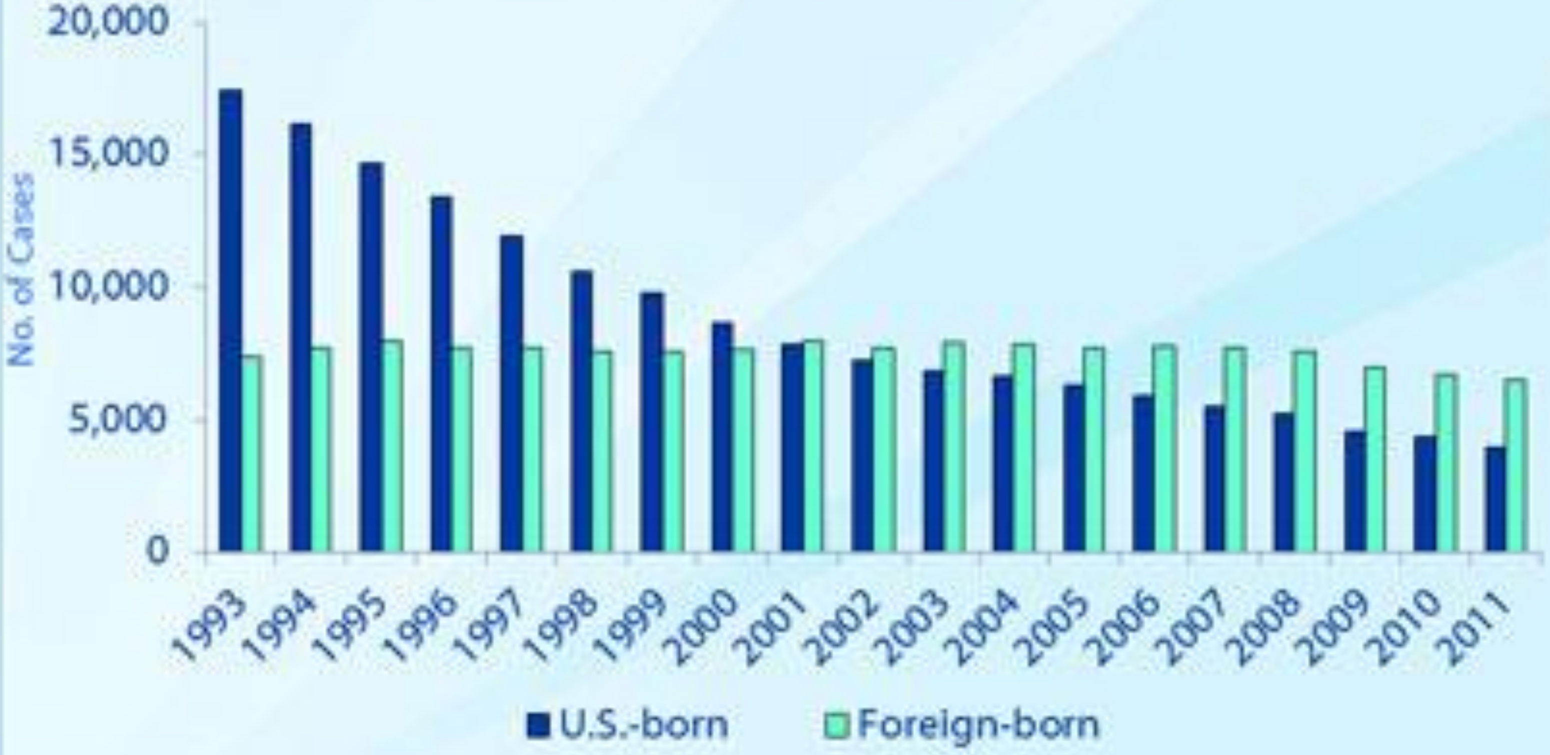
2.4 Million Deaths per Year

Tuberculosis cases and rates, New York City, 1982-2012



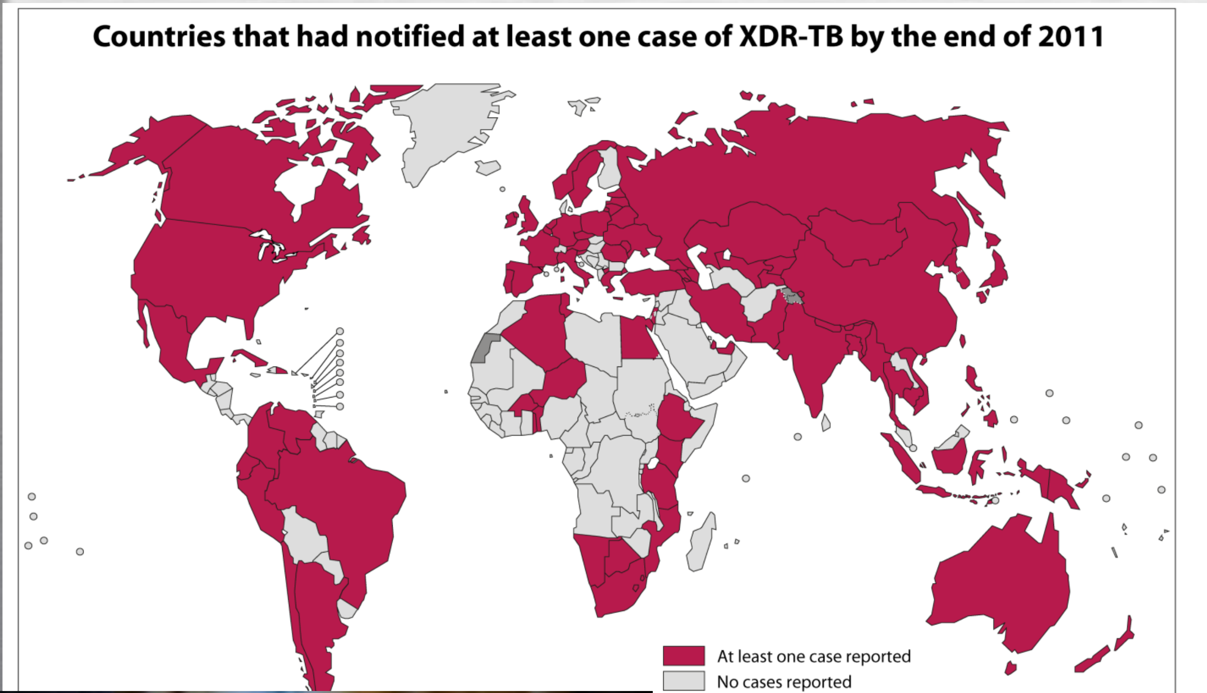
1. Rates are based on official Census data.

Number of TB Cases in U.S.-born vs. Foreign-born Persons United States, 1993–2011*



*Updated as of June 25, 2012.

Drug Resistance Threat



Countries that had notified at least one case of XDR-TB by the end of 2011

sion of
f any
oundaries.
be full agreement.

Source: *Global Tuberculosis Report 2012.* WHO, 2012.

World Health Organization

- Susceptible
- Drug Resistant
- MDR-TB - Multi-Drug Resistant
- XDR-TB - Extremely-Drug Resistant
- TDR-TB?- Totally Drug Resistant

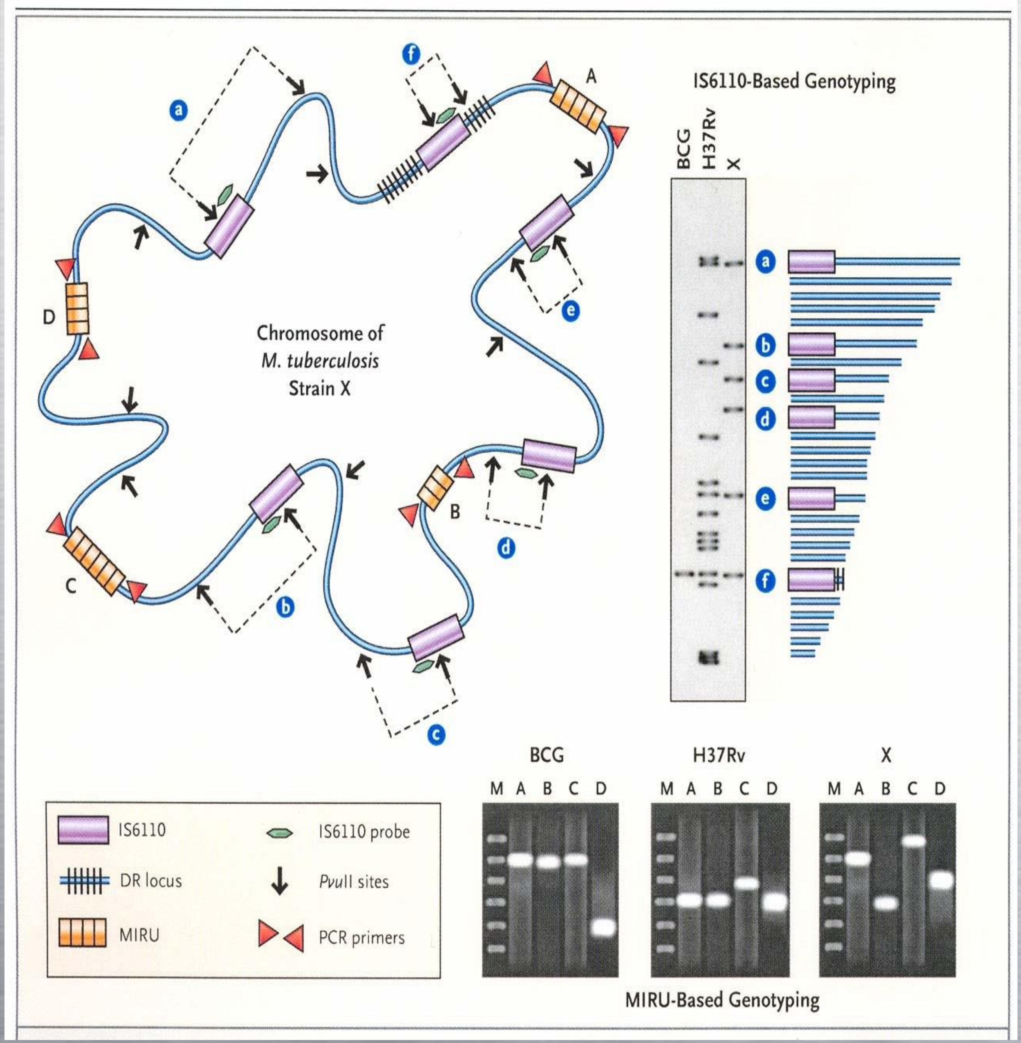
Modern TB Control

TB Controller: Find source(s) of infection in order to identify people who need treatment and stop future transmission.

Tools: **Contact Investigation**

DNA Fingerprints of TB bacteria

Two or More DNA fingerprints gathered for every TB Patient in USA



- Spoligotype

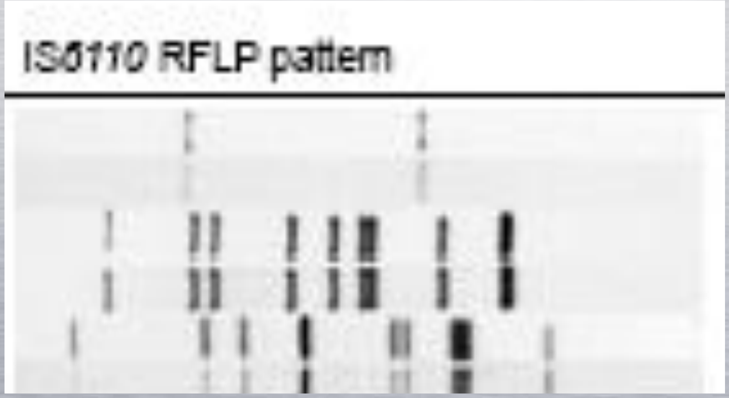
1111111111111110000000000000000000000000000

- MIRU

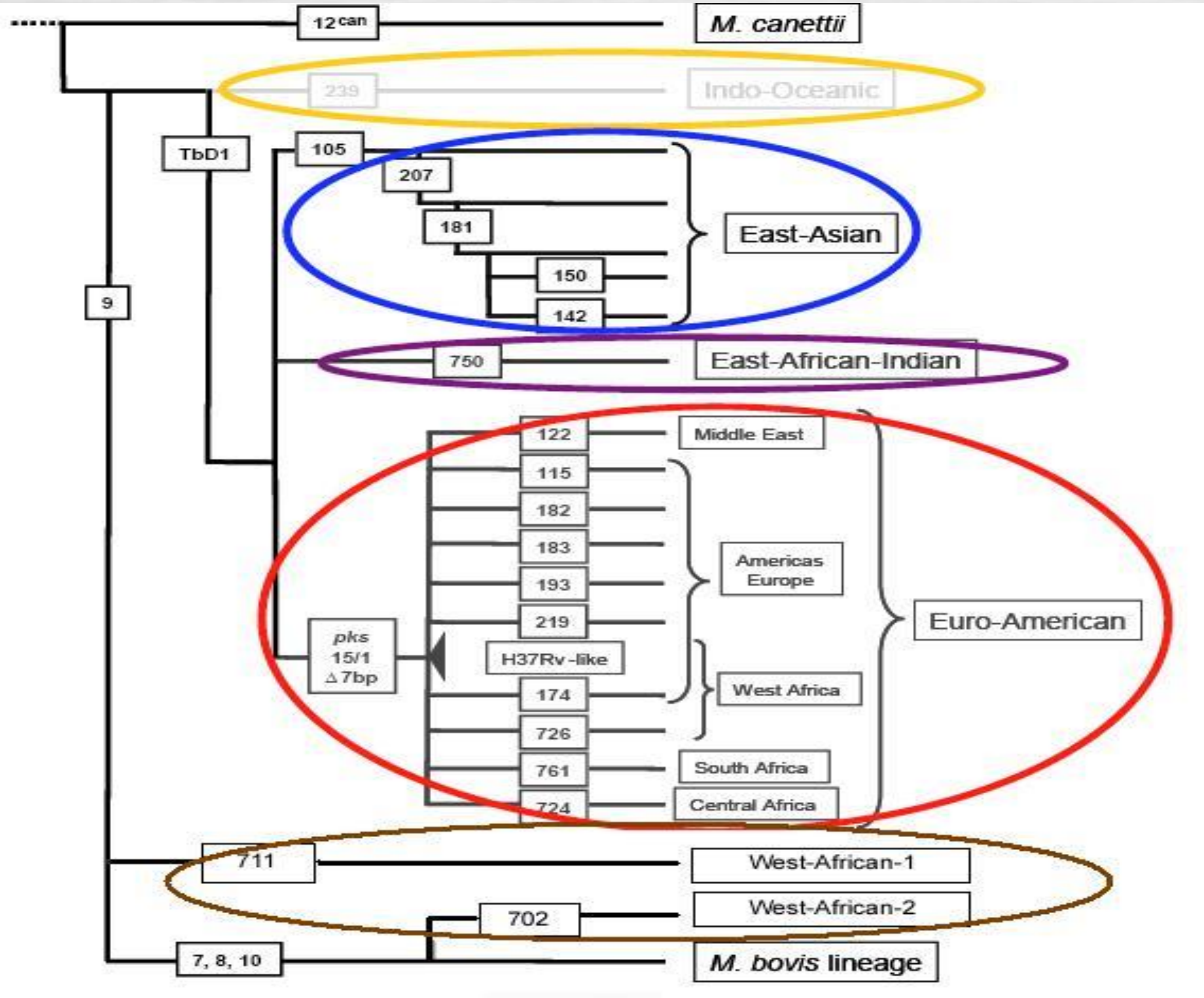
  22531b153321

- RFLP

  BW90



Major Phylogeographic Lineages of the MTBC

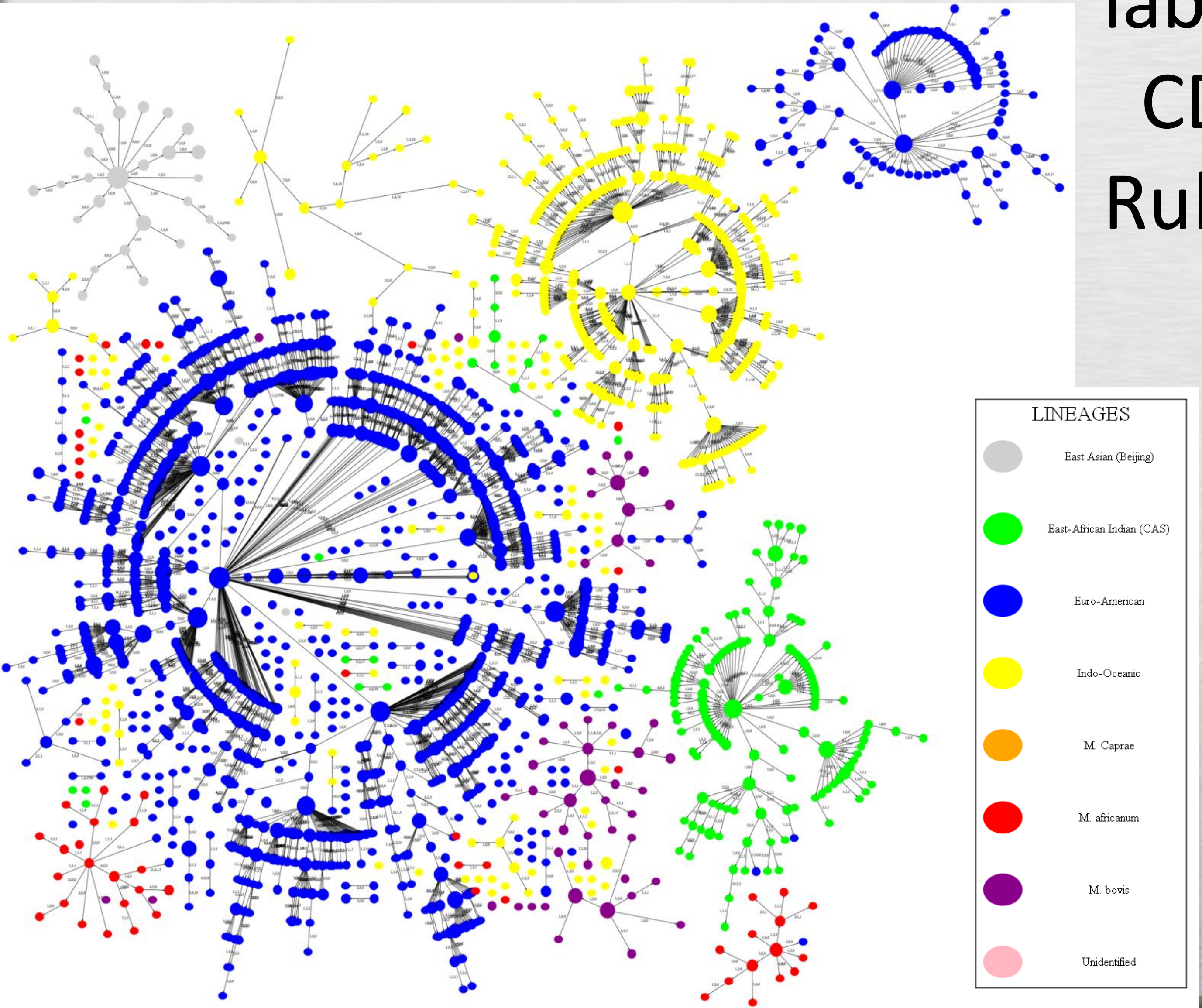


Determined by RD's

Predictable by Spoligotype

[Gagneux S et al. PNAS 2006]

Spoligotype Genetic Diversity 37K Patients in US – 2004-2008



Spoligoforests labeled by CDC Expert Rules

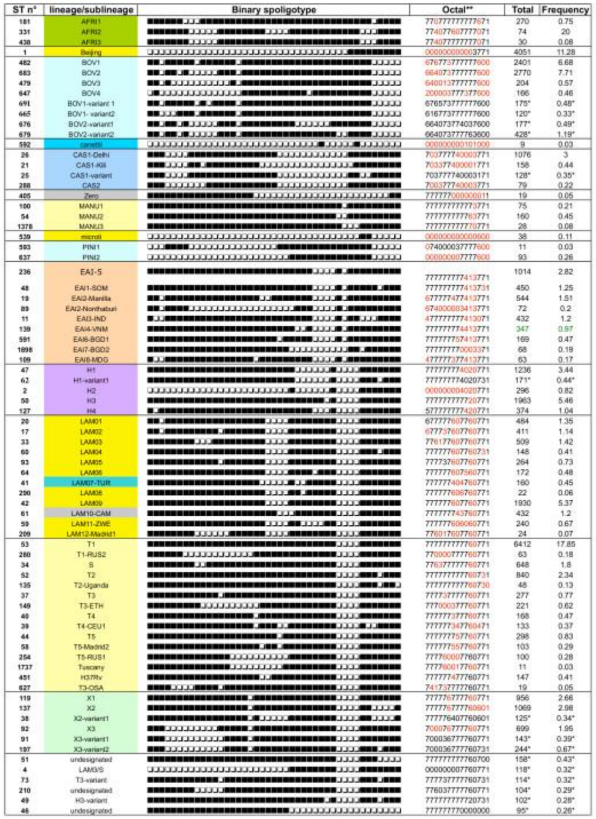
Classification Models

TB-Lineage

- Rule-based model: 2012.
- Bayesian Network: 2010.

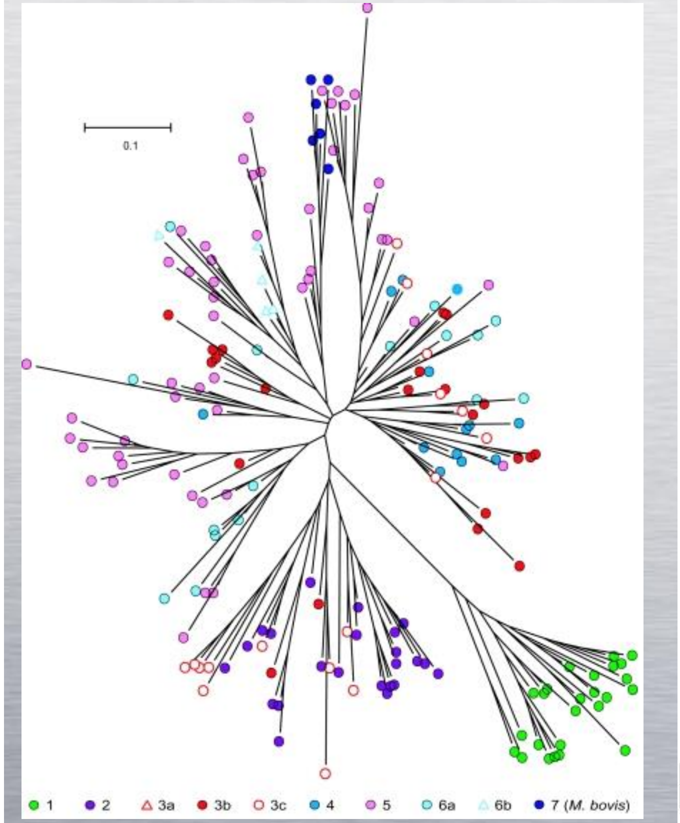
Sublineages-varying opinions

62 sublineages based on spoligotype signatures
[Brudey et al BMC Microbiol. 2006]

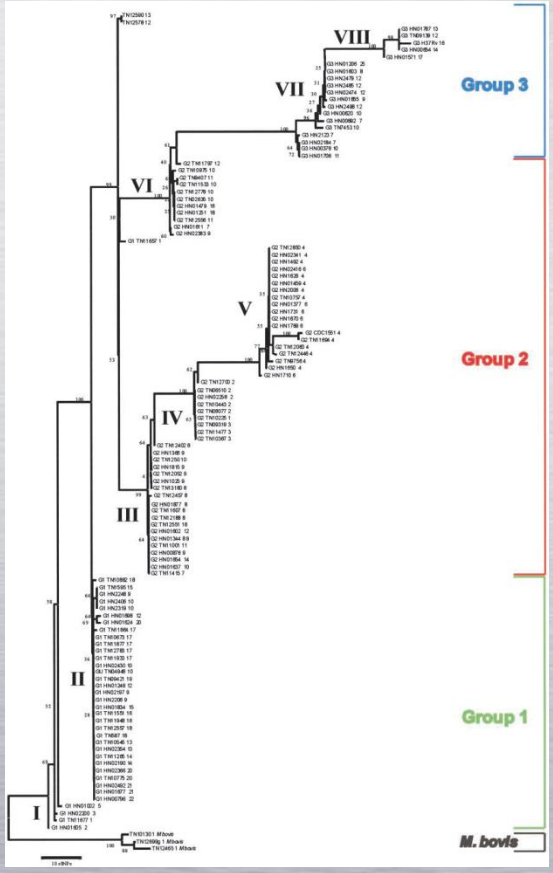


ST n°	lineage/sublineage	Octal**	Total	Frequency
181	AFRI1	770777777777671	270	0.75
331	AFRI2	774077607777071	74	20
438	AFRI3	774077777777071	30	0.08
1	Beijing	000000000003771	4051	11.28
482	BOV1	676773777777600	2401	6.68
683	BOV2	664073777777600	2770	7.71
479	BOV3	640013777777600	204	0.57
647	BOV4	200003777377600	166	0.46
691	BOV1-variant 1	676573777777600	175*	0.48*
665	BOV1- variant2	616773777777600	120*	0.33*
676	BOV2-variant1	664073774037600	177*	0.49*
679	BOV2-variant2	664073777763600	428*	1.19*
592	canettii	000000000101000	9	0.03
26	CAS1-Delhi	703777740003771	1076	3
21	CAS1-Kili	703377400001771	158	0.44
25	CAS1-variant	703777740003171	128*	0.35*
288	CAS2	700377740003771	79	0.22
405	Zero	777777000000011	19	0.05
100	MANU1	777777777773771	75	0.21
54	MANU2	777777777763771	160	0.45
1378	MANU3	777777777770771	28	0.08
539	microti	000000000000600	38	0.11
593	PINI1	074000037777600	11	0.03
637	PINI2	000000007777600	93	0.26
236	EAI-5	777777777413771	1014	2.82
48	EAI1-SOM	777777777413731	450	1.25
19	EAI2-Manilla	677777477413771	544	1.51
89	EAI2-Nonthaburi	674000003413771	72	0.2
11	EAI3-IND	477777777413071	432	1.2
139	EAI4-VNM	777777774413771	347	0.97
591	EAI6-BGD1	777777757413771	169	0.47
1898	EAI7-BGD2	777777770003371	68	0.19
109	EAI8-MDG	477777377413771	63	0.17
47	H1	777777774020771	1236	3.44
62	H1-variant1	777777774020731	171*	0.44*
2	H2	000000004020771	296	0.82
50	H3	777777777720771	1963	5.46
127	H4	577777777420771	374	1.04
20	LAM01	677777607760771	484	1.35
17	LAM02	677737607760771	411	1.14
33	LAM03	776177607760771	509	1.42
60	LAM04	777777607760731	148	0.41
93	LAM05	777737607760771	264	0.73
64	LAM06	777777607560771	172	0.48
41	LAM07-TUR	777777404760771	160	0.45
290	LAM08	777777606760771	22	0.06
42	LAM09	777777607760771	1930	5.37
61	LAM10-CAM	777777743760771	432	1.2
59	LAM11-ZWE	777777606060771	240	0.67
209	LAM12-Madrid1	776017607760771	24	0.07
53	T1	777777777760771	6412	17.85
280	T1-RUS2	770000777760771	63	0.18
34	S	776377777760771	648	1.8
52	T2	777777777760731	840	2.34
135	T2-Uganda	777777777760730	48	0.13
37	T3	777737777760771	277	0.77
149	T3-ETH	777000377760771	221	0.62
40	T4	777777377760771	168	0.47
39	T4-CEU1	777777347760471	133	0.37
44	T5	777777757760771	298	0.83
58	T5-Madrid2	777777557760771	103	0.29
254	T5-RUS1	777760007760771	100	0.28
1737	Tuscany	777760017760771	11	0.03
451	H37Rv	777777477760771	147	0.41
627	T3-OSA	741737777760771	19	0.05
119	X1	777776777760771	956	2.66
137	X2	777776777760601	1069	2.98
38	X2-variant1	777776407760601	125*	0.34*
92	X3	700076777760771	699	1.95
91	X3-variant1	700036777760771	143*	0.39*
197	X3-variant2	700036777760731	244*	0.67*
51	undesignated	777777777760700	158*	0.43*
4	LAM3/S	000000007760771	118*	0.32*
73	T3-variant	777737777760731	114*	0.32*
210	undesignated	776037777760771	104*	0.29*
49	H3-variant	777777777720731	102*	0.28*
46	undesignated	777777770000000	95*	0.26*

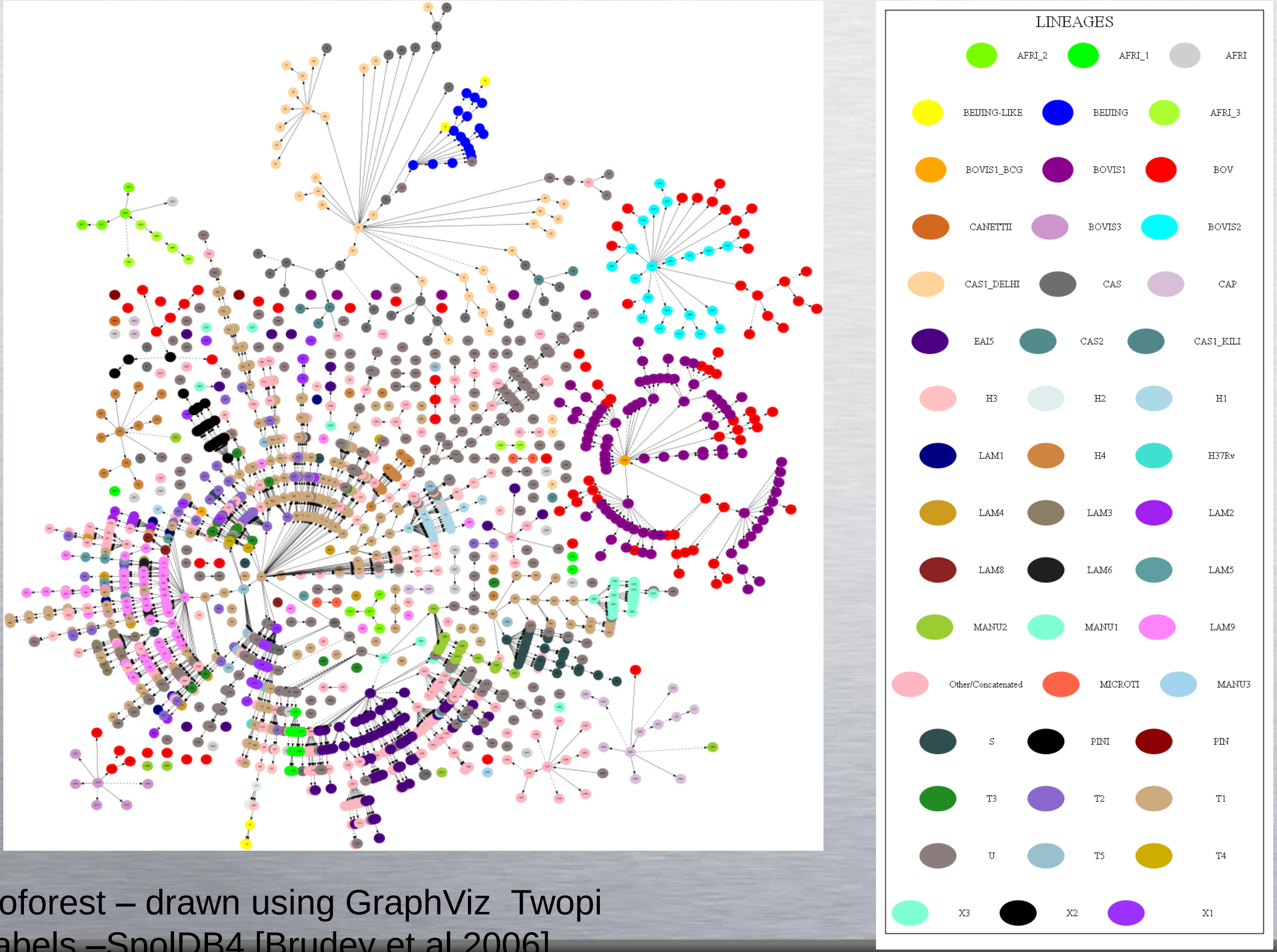
10 groups based on 212 single nucleotide polymorphism (SNP) markers
[Filliol J. Bacteriol 2006]



8 groups based on 230 sSNPs
[Gutacker Genetics 2002]



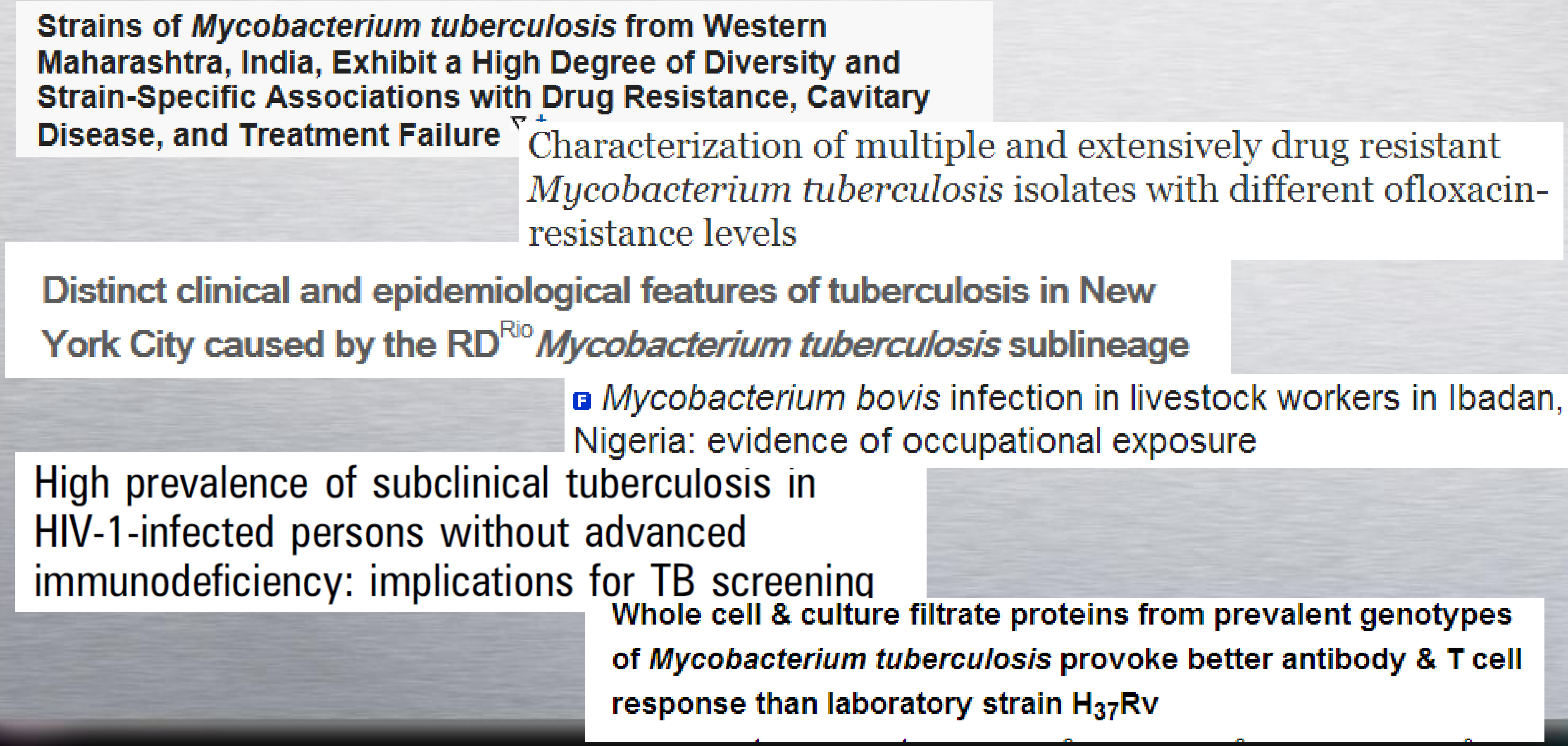
What’s the story with sublineages?



Spoligoforest – drawn using GraphViz Twopi
Data labels –SpolDB4 [Brudey et al 2006]

SPOTCLUST: Hidden-Parent Bayesian Network for Spoligotypes (Vitol et al, 2006)

Standard resource for **tuberculosis sublineage identification** used in over 96 publications.



Strains of *Mycobacterium tuberculosis* from Western Maharashtra, India, Exhibit a High Degree of Diversity and Strain-Specific Associations with Drug Resistance, Cavitary Disease, and Treatment Failure

Characterization of multiple and extensively drug resistant *Mycobacterium tuberculosis* isolates with different ofloxacin-resistance levels

Distinct clinical and epidemiological features of tuberculosis in New York City caused by the RD^{Rio} *Mycobacterium tuberculosis* sublineage

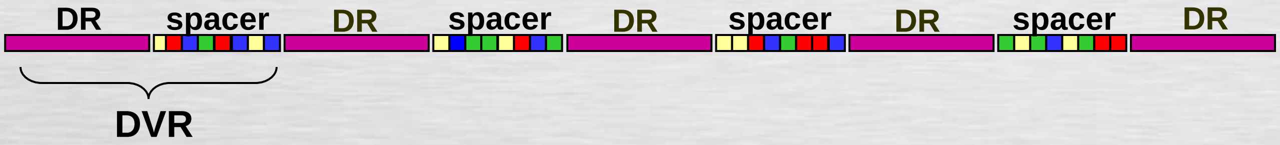
F *Mycobacterium bovis* infection in livestock workers in Ibadan, Nigeria: evidence of occupational exposure

High prevalence of subclinical tuberculosis in HIV-1-infected persons without advanced immunodeficiency: implications for TB screening

Whole cell & culture filtrate proteins from prevalent genotypes of *Mycobacterium tuberculosis* provoke better antibody & T cell response than laboratory strain $H_{37}Rv$

Spoligotyping

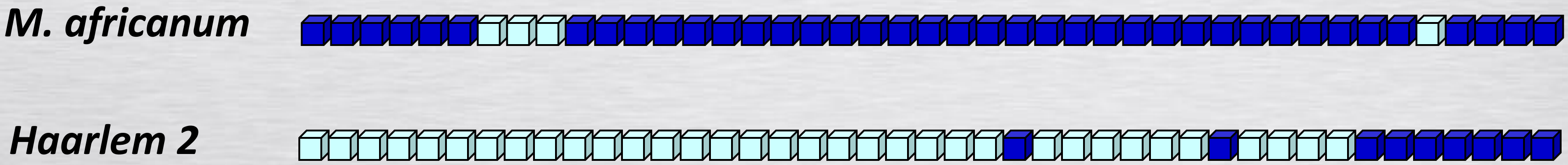
- Direct repeats (DR) separated by variable spacers



- Contiguous on chromosome, order well conserved
- Forty three spacers used
- Presence of a spacer is detected: 1- present (■), 0 - absent (□)

Strain	Binary description of spoligotypes
M. tuberculosis Beijing	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■■
M. bovis	■■□■■■■■□■■■■■■□■■■■■■■■■■■■■■■■■■■■■■□□□□□
	□■□■■■■■□■■■■■■□■■■■■■■■■■■■■■■■■■■■■■□□□□□
Indo-Oceanic	■■□■■■■■□■■■■■■□■■■■■■■■■■■■■■■■■■■■■■□□■□□

Rule-Based Method:
(Sub)Lineage Visual Rules

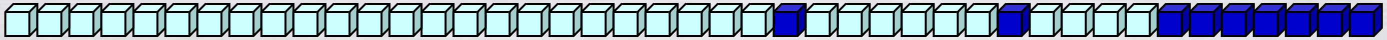


- Determined by human experts.
- Ill-defined.
- Incomplete.
- Frequently ambiguous.
- No precedence.
- May or may not correspond to actual evolutionary groups.

First Try – Naïve Bayes

M. tuberculosis Haarlem2 Family

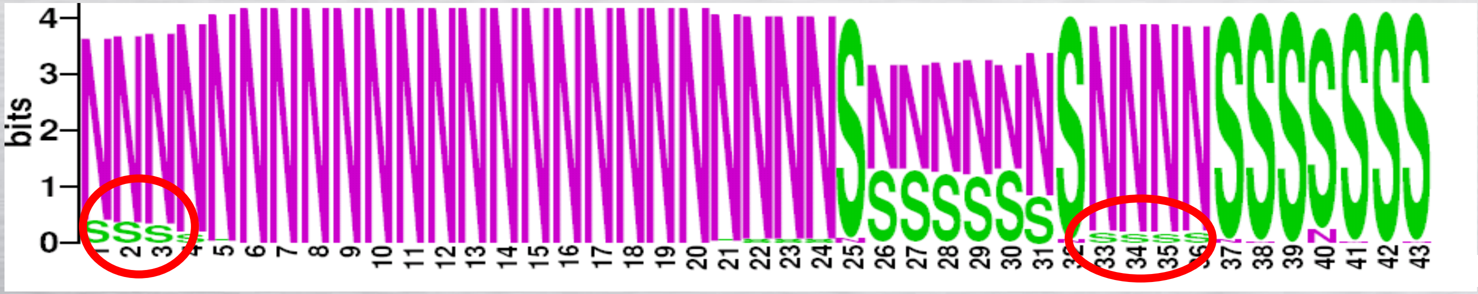
- Prototype = probabilities



S - spacer present

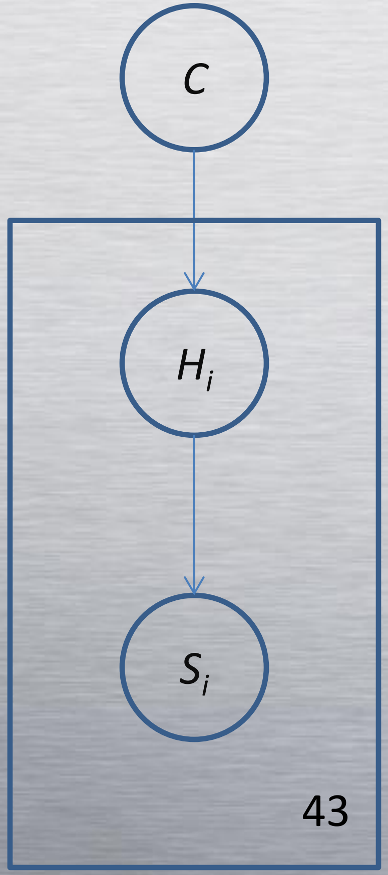
N - spacer absent

- Bernoulli Mixture Model



- Biology is wrong!!!

Unsupervised Hidden Parent Multivariate Bernoulli Mixture Model



Model child spacer *S*, given unobserved parent spacer *H*

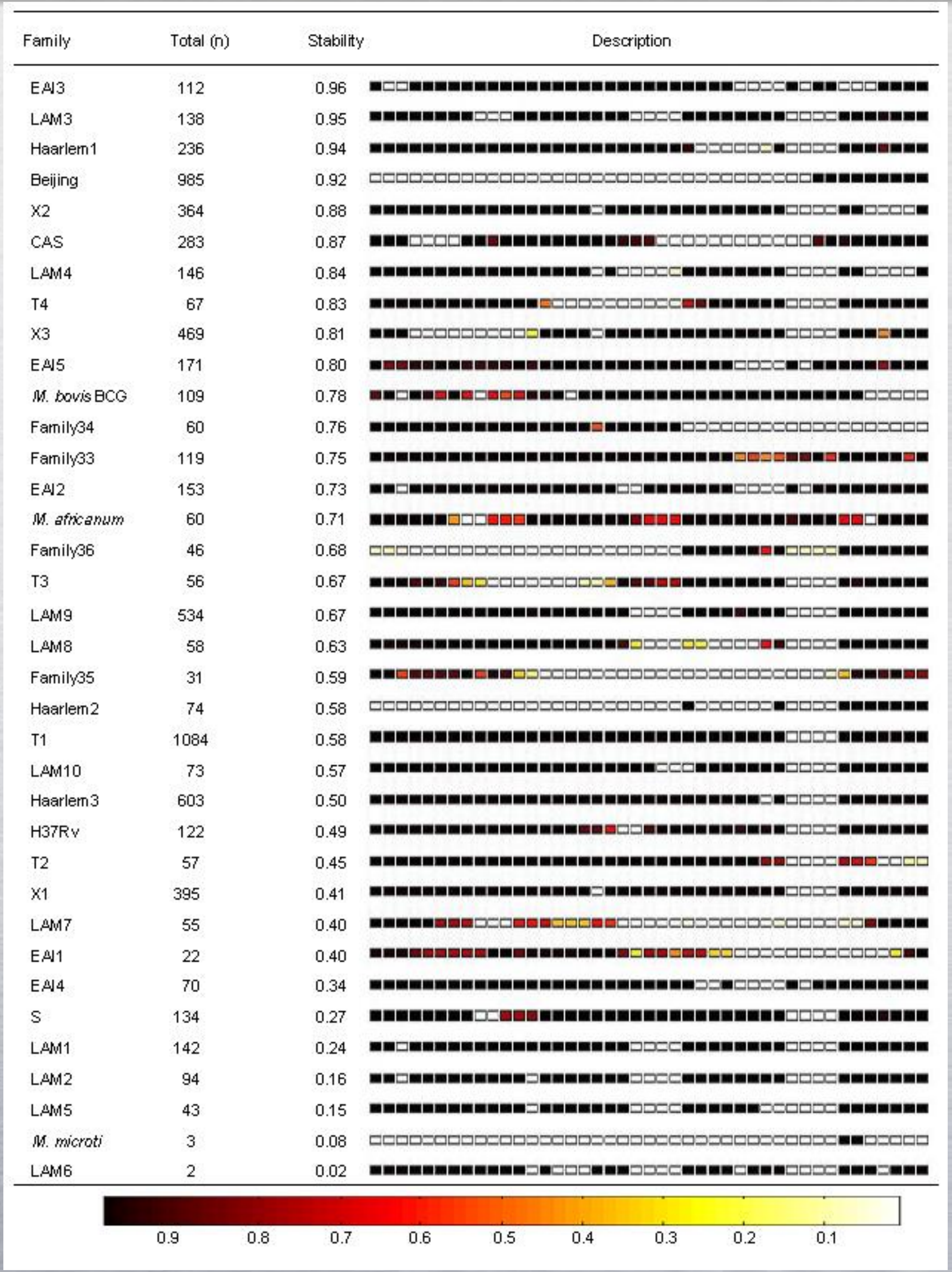
- With very high probability child matches parent
- Children are much more likely to lose spacer than gain
  - $P(S=1|H=1) = 0.99$, $P(S=0|H=1) = 0.01$
  - $P(S=0|H=0) = 1\text{-}1e\text{-}7$ $P(S=1|H=0) = 1e\text{-}7$

SPOTCLUST (2006)

- Unsupervised except

34 SPOLDB2 Visual Rules used to initialize clusters

- Trained using 535 spoligotypes
- Number of sublineages (36) picked by MCCV



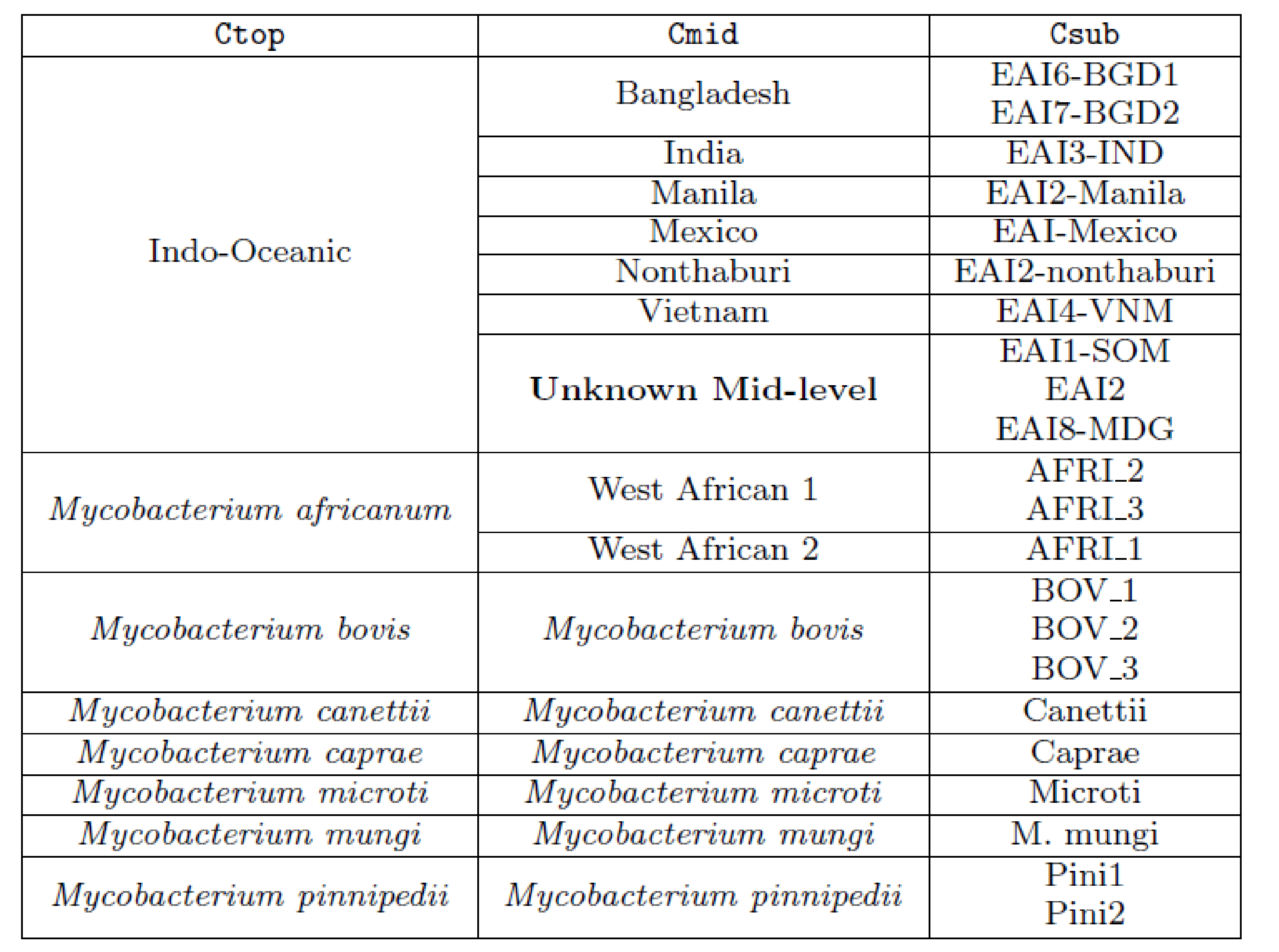
Family	Total (n)	Stability	Description
EAI3	112	0.96	
LAM3	138	0.95	
Haarlem1	236	0.94	
Beijing	985	0.92	
X2	364	0.88	
CAS	283	0.87	
LAM4	146	0.84	
T4	67	0.83	
X3	469	0.81	
EAI5	171	0.80	
M. bovis BCG	109	0.78	
Family34	60	0.76	
Family33	119	0.75	
EAI2	153	0.73	
M. africanum	60	0.71	
Family36	46	0.68	
T3	56	0.67	
LAM9	534	0.67	
LAM8	58	0.63	
Family35	31	0.59	
Haarlem2	74	0.58	
T1	1084	0.58	
LAM10	73	0.57	
Haarlem3	603	0.50	
H37Rv	122	0.49	
T2	57	0.45	
X1	395	0.41	
LAM7	55	0.40	
EAI1	22	0.40	
EAI4	70	0.34	
S	134	0.27	
LAM1	142	0.24	
LAM2	94	0.16	
LAM5	43	0.15	
M. microti	3	0.08	
LAM6	2	0.02	

New Challenges

- More data: 119,684 isolates from US CDC, NYDOH, NY State DOH, and Institut Pasteur de Guadeloupe, MIRUVNTRPlus
- More types of DNA fingerprints –

  Spoligotypes and MIRU
- More proposed sublineages (70?)
- Putative labels from multiple experts
- Missing Data

Who’s Right?



Ctop	Cmid	Csub
Indo-Oceanic	Bangladesh	EAI6-BGD1 EAI7-BGD2
	India	EAI3-IND
	Manila	EAI2-Manila
	Mexico	EAI-Mexico
	Nonthaburi	EAI2-nonthaburi
	Vietnam	EAI4-VNM
	Unknown Mid-level	EAI1-SOM EAI2 EAI8-MDG
Mycobacterium africanum	West African 1	AFRI_2 AFRI_3
	West African 2	AFRI_1
Mycobacterium bovis	*Mycobacterium bovis*	BOV_1 BOV_2 BOV_3
Mycobacterium canettii	*Mycobacterium canettii*	Canettii
Mycobacterium caprae	*Mycobacterium caprae*	Caprae
Mycobacterium microti	*Mycobacterium microti*	Microti
Mycobacterium mungi	*Mycobacterium mungi*	M. mungi
Mycobacterium pinnipedii	*Mycobacterium pinnipedii*	Pini1 Pini2

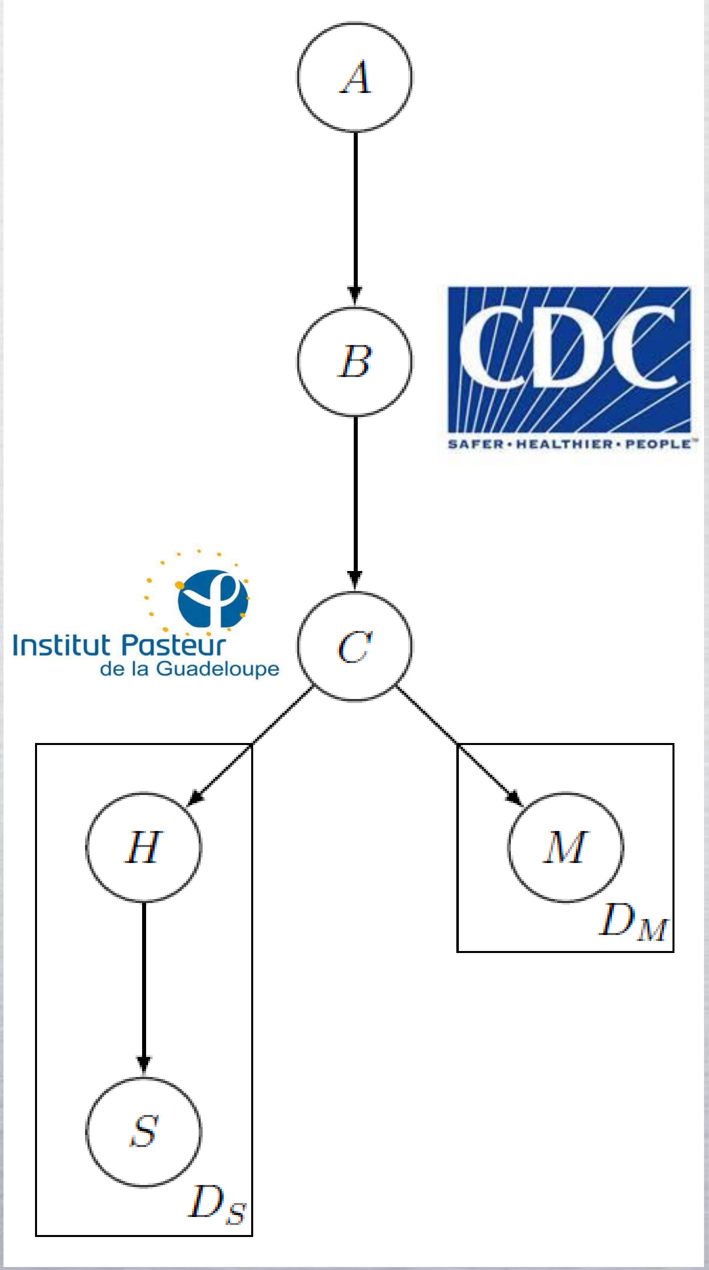
Semi-supervised Hierarchical Lineage Model

A: 12 Major Lineages

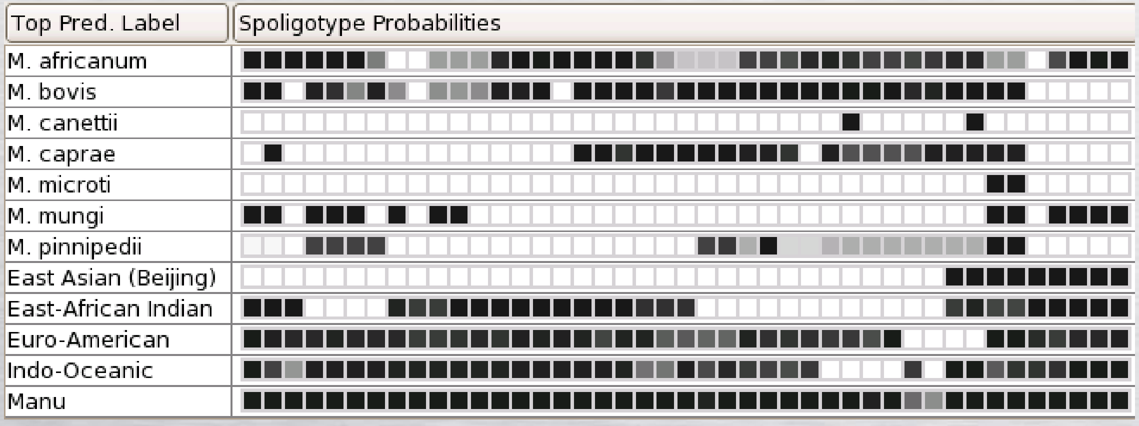
B: 22 Mid-level Lineages

C: 70 + 9 Sub-lineages

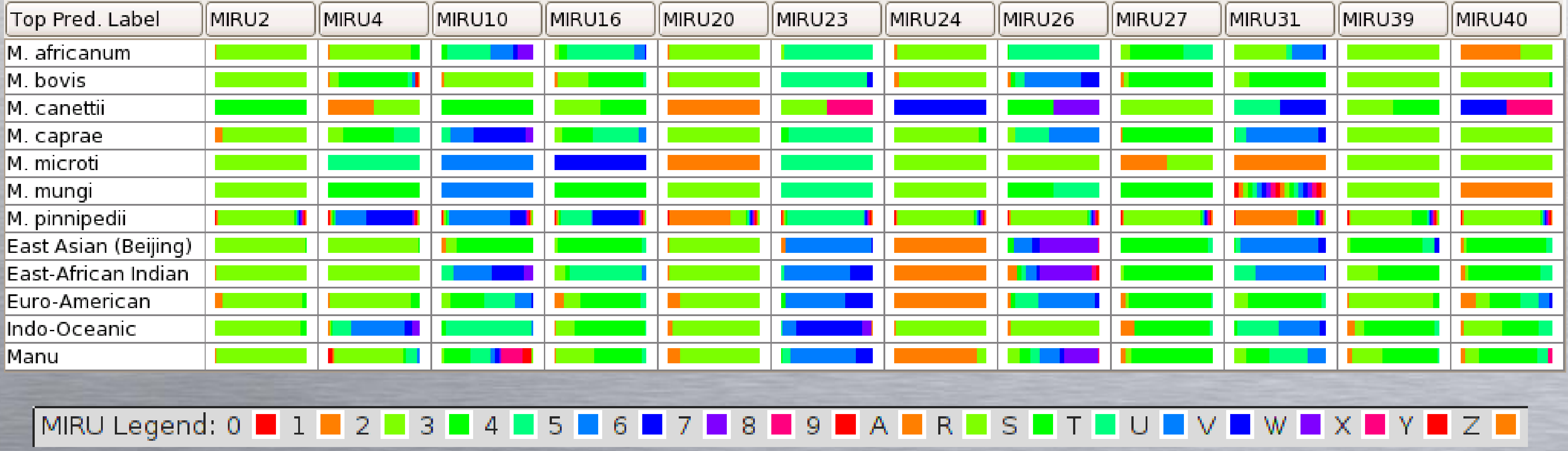
- Estimated 92% Cross-validated Accuracy



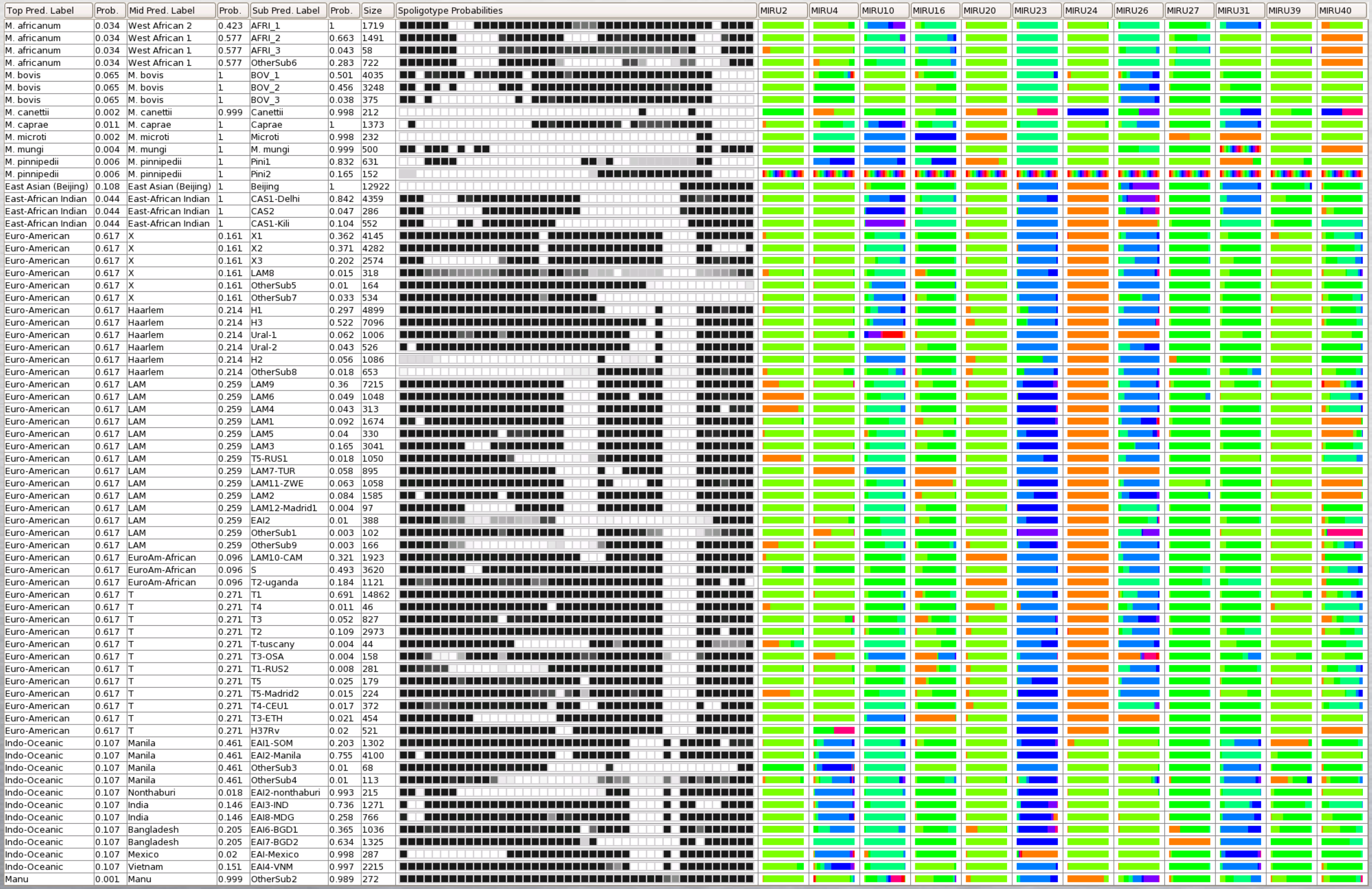
Major Lineage Results



- Balanced Classification Rate about 98%
- No changes in major lineages
- MANU Modern?

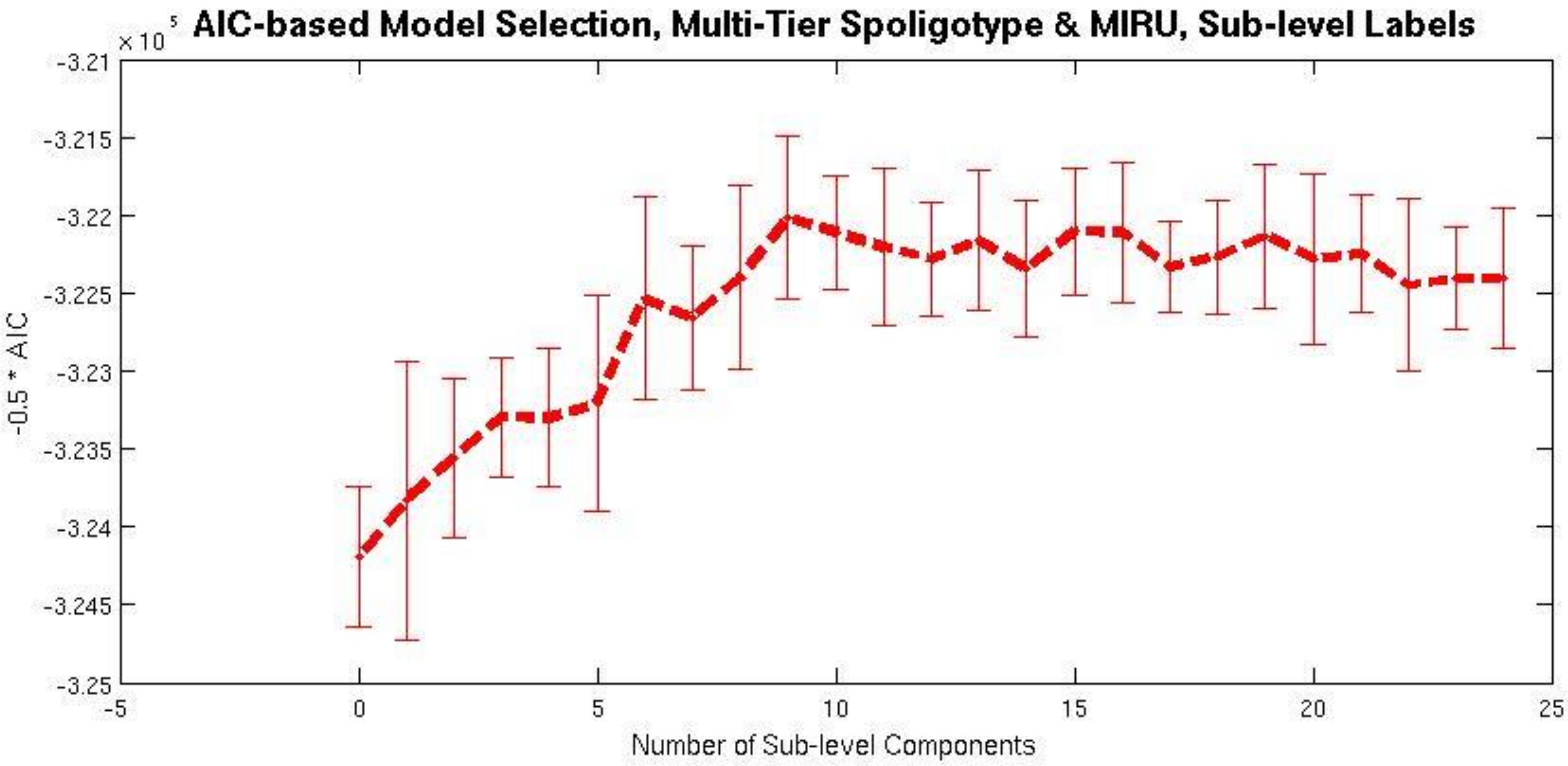


Sub-level Results



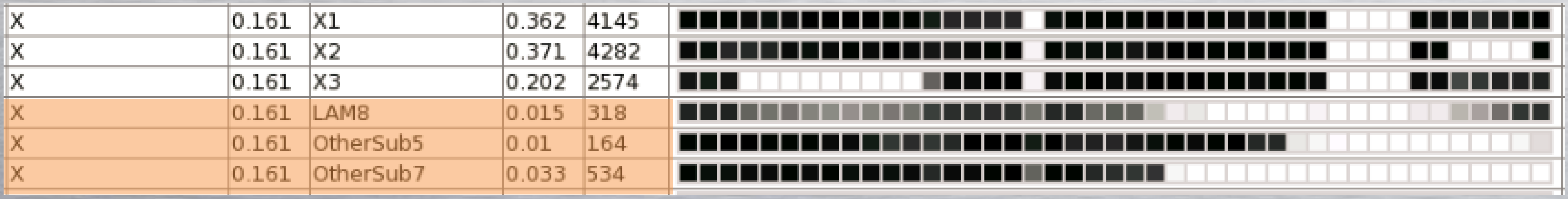
Top Pred. Label	Prob.	Mid Pred. Label	Prob.	Sub Pred. Label	Prob.	Size	Spoligotype Probabilities	MIRU2	MIRU4	MIRU10	MIRU16	MIRU20	MIRU23	MIRU24	MIRU26	MIRU27	MIRU31	MIRU39	MIRU40
M. africanum	0.034	West African 2	0.423	AFRI_1	1	1719													
M. africanum	0.034	West African 1	0.577	AFRI_2	0.663	1491													
M. africanum	0.034	West African 1	0.577	AFRI_3	0.043	58													
M. africanum	0.034	West African 1	0.577	OtherSub6	0.283	722													
M. bovis	0.065	M. bovis	1	BOV_1	0.501	4035													
M. bovis	0.065	M. bovis	1	BOV_2	0.456	3248													
M. bovis	0.065	M. bovis	1	BOV_3	0.038	375													
M. canettii	0.002	M. canettii	0.999	Canettii	0.998	212													
M. caprae	0.011	M. caprae	1	Caprae	1	1373													
M. microti	0.002	M. microti	1	Microti	0.998	232													
M. mungi	0.004	M. mungi	1	M. mungi	0.999	500													
M. pinnipedii	0.006	M. pinnipedii	1	Pini1	0.832	631													
M. pinnipedii	0.006	M. pinnipedii	1	Pini2	0.165	152													
East Asian (Beijing)	0.108	East Asian (Beijing)	1	Beijing	1	12922													
East-African Indian	0.044	East-African Indian	1	CAS1-Delhi	0.842	4359													
East-African Indian	0.044	East-African Indian	1	CAS2	0.047	286													
East-African Indian	0.044	East-African Indian	1	CAS1-Kili	0.104	552													
Euro-American	0.617	X	0.161	X1	0.362	4145													
Euro-American	0.617	X	0.161	X2	0.371	4282													
Euro-American	0.617	X	0.161	X3	0.202	2574													
Euro-American	0.617	X	0.161	LAM8	0.015	318													
Euro-American	0.617	X	0.161	OtherSub5	0.01	164													
Euro-American	0.617	X	0.161	OtherSub7	0.033	534													
Euro-American	0.617	Haarlem	0.214	H1	0.297	4899													
Euro-American	0.617	Haarlem	0.214	H3	0.522	7096													
Euro-American	0.617	Haarlem	0.214	Ural-1	0.062	1006													
Euro-American	0.617	Haarlem	0.214	Ural-2	0.043	526													
Euro-American	0.617	Haarlem	0.214	H2	0.056	1086													
Euro-American	0.617	Haarlem	0.214	OtherSub8	0.018	653													
Euro-American	0.617	LAM	0.259	LAM9	0.36	7215													
Euro-American	0.617	LAM	0.259	LAM6	0.049	1048													
Euro-American	0.617	LAM	0.259	LAM4	0.043	313													
Euro-American	0.617	LAM	0.259	LAM1	0.092	1674													
Euro-American	0.617	LAM	0.259	LAM5	0.04	330													
Euro-American	0.617	LAM	0.259	LAM3	0.165	3041													
Euro-American	0.617	LAM	0.259	T5-RUS1	0.018	1050													
Euro-American	0.617	LAM	0.259	LAM7-TUR	0.058	895													
Euro-American	0.617	LAM	0.259	LAM11-ZWE	0.063	1058													
Euro-American	0.617	LAM	0.259	LAM2	0.084	1585													
Euro-American	0.617	LAM	0.259	LAM12-Madrid1	0.004	97													
Euro-American	0.617	LAM	0.259	EAI2	0.01	388													
Euro-American	0.617	LAM	0.259	OtherSub1	0.003	102													
Euro-American	0.617	LAM	0.259	OtherSub9	0.003	166													
Euro-American	0.617	EuroAm-African	0.096	LAM10-CAM	0.321	1923													
Euro-American	0.617	EuroAm-African	0.096	S	0.493	3620													
Euro-American	0.617	EuroAm-African	0.096	T2-uganda	0.184	1121													
Euro-American	0.617	T	0.271	T1	0.691	14862													
Euro-American	0.617	T	0.271	T4	0.011	46													
Euro-American	0.617	T	0.271	T3	0.052	827													
Euro-American	0.617	T	0.271	T2	0.109	2973													
Euro-American	0.617	T	0.271	T-tuscany	0.004	44													
Euro-American	0.617	T	0.271	T3-OSA	0.004	158													
Euro-American	0.617	T	0.271	T1-RUS2	0.008	281													
Euro-American	0.617	T	0.271	T5	0.025	179													
Euro-American	0.617	T	0.271	T5-Madrid2	0.015	224													
Euro-American	0.617	T	0.271	T4-CEU1	0.017	372													
Euro-American	0.617	T	0.271	T3-ETH	0.021	454													
Euro-American	0.617	T	0.271	H37Rv	0.02	521													
Indo-Oceanic	0.107	Manila	0.461	EAI1-SOM	0.203	1302													
Indo-Oceanic	0.107	Manila	0.461	EAI2-Manila	0.755	4100													
Indo-Oceanic	0.107	Manila	0.461	OtherSub3	0.01	68													
Indo-Oceanic	0.107	Manila	0.461	OtherSub4	0.01	113													
Indo-Oceanic	0.107	Nonthaburi	0.018	EAI2-nonthaburi	0.993	215													
Indo-Oceanic	0.107	India	0.146	EAI3-IND	0.736	1271													
Indo-Oceanic	0.107	India	0.146	EAI8-MDG	0.258	766													
Indo-Oceanic	0.107	Bangladesh	0.205	EAI6-BGD1	0.365	1036													
Indo-Oceanic	0.107	Bangladesh	0.205	EAI7-BGD2	0.634	1325													
Indo-Oceanic	0.107	Mexico	0.02	EAI-Mexico	0.998	287													
Indo-Oceanic	0.107	Vietnam	0.151	EAI4-VNM	0.997	2215													
Manu	0.001	Manu	0.999	OtherSub2	0.989	272													

Model Adds Sublineages but not Mid or Major Lineages



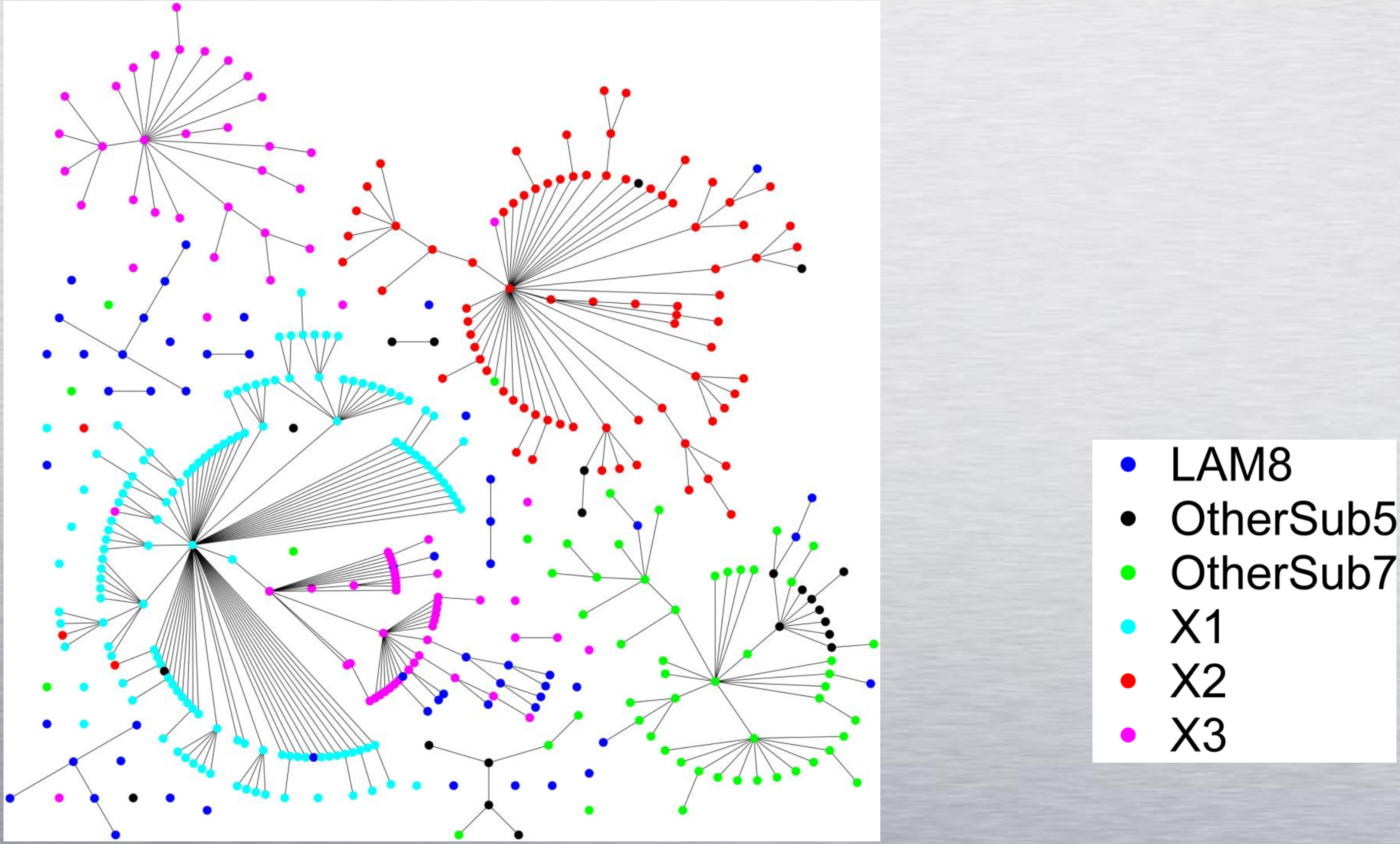
Ten New Putative Sublineages

- Discovers new sublineages and assigns them mid-level lineage
- New lineage characterized by ‘’long’’ deletions frequently covering “typical” lineage deletions
- Covers many previously unlabeled isolates.



X	0.161	X1	0.362	4145
X	0.161	X2	0.371	4282
X	0.161	X3	0.202	2574
X	0.161	LAM8	0.015	318
X	0.161	OtherSub5	0.01	164
X	0.161	OtherSub7	0.033	534

Visualizing X Lineage Spoligoforest



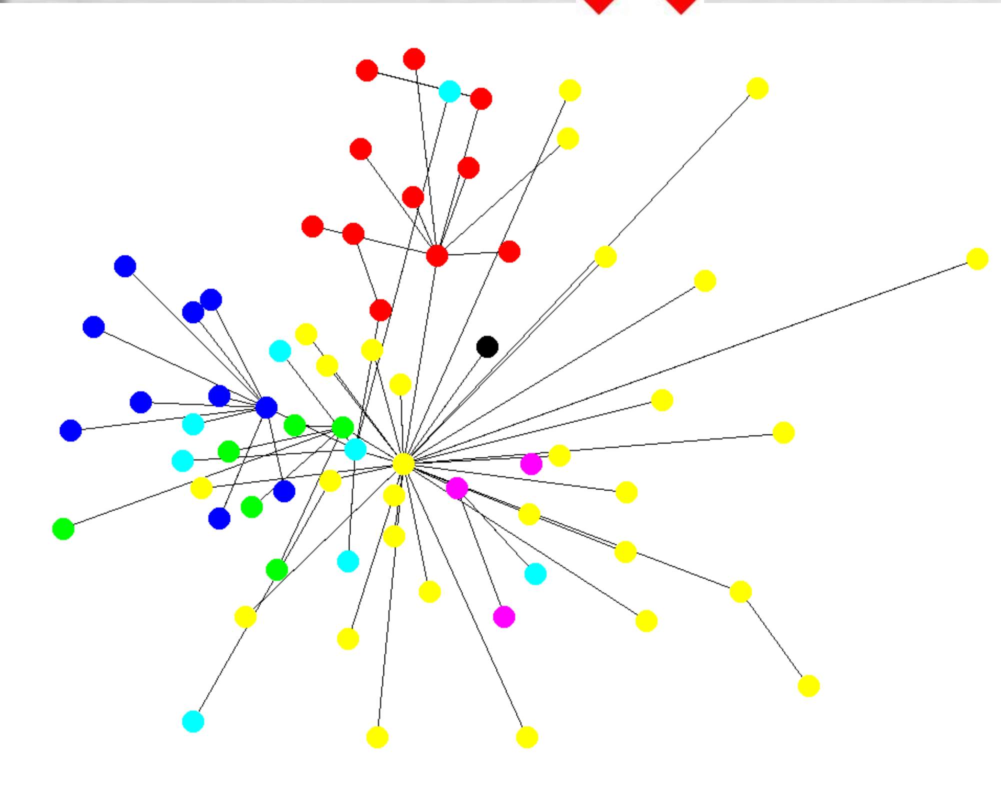
Graphvis Two-PI

Multi-Objective Embedding Methods

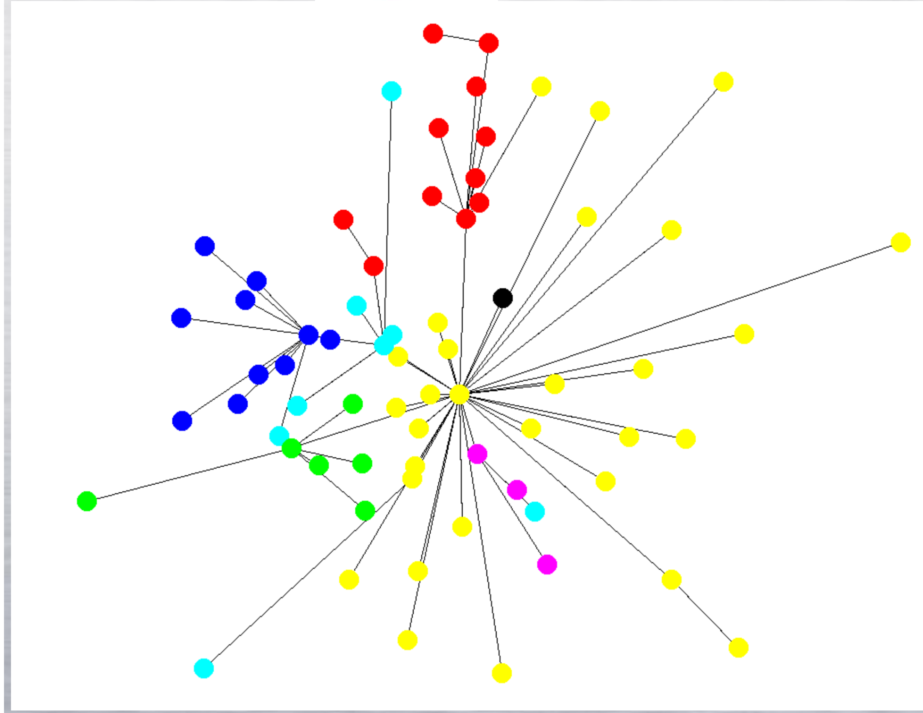
Good visualizations

minimize *stress* and *edge crossings*

✗



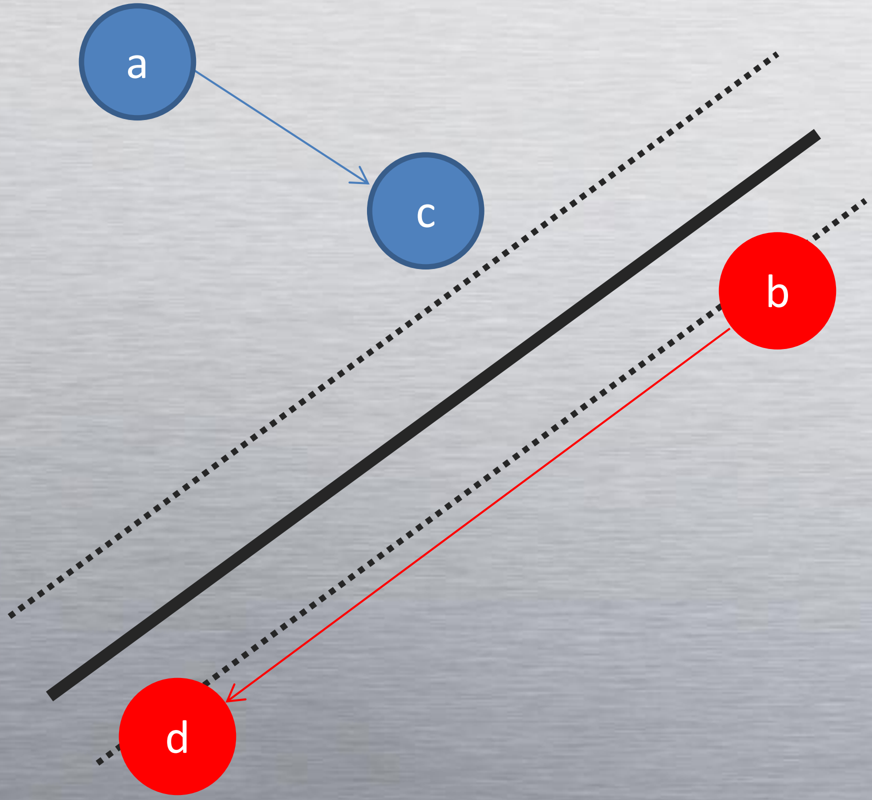
✓



Edge Crossing Constraints as Classification

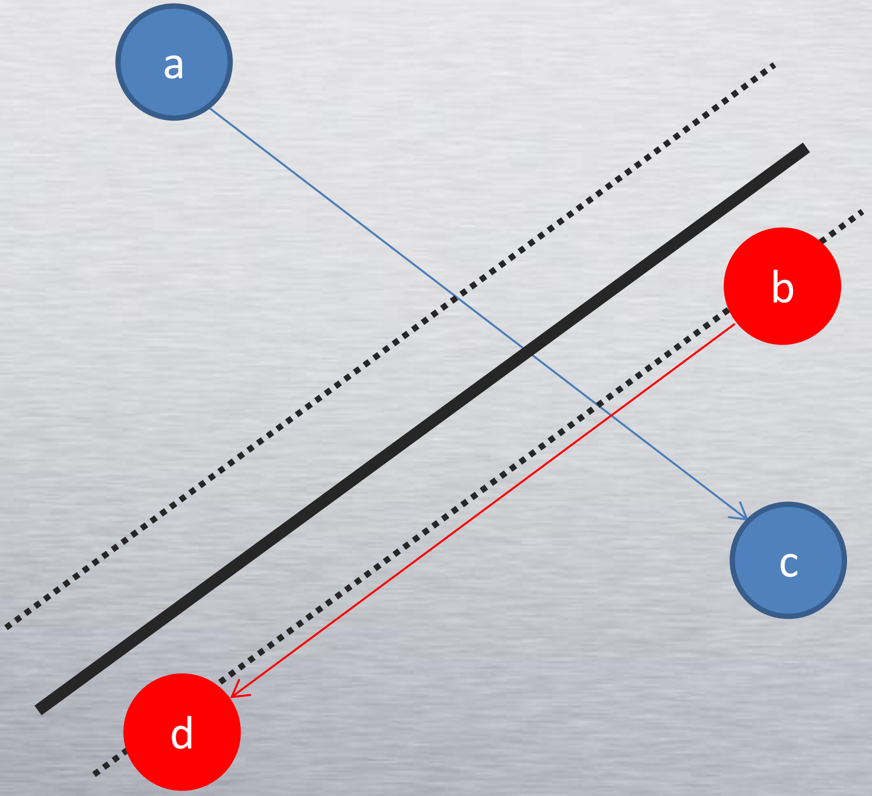
No Crossing =

Correct Classification ✓



Crossing =

Incorrect Crossing ✗



Multi-Objective Graph Embedding

Add classifier for each potential edge/node crossing parameterized by *U*

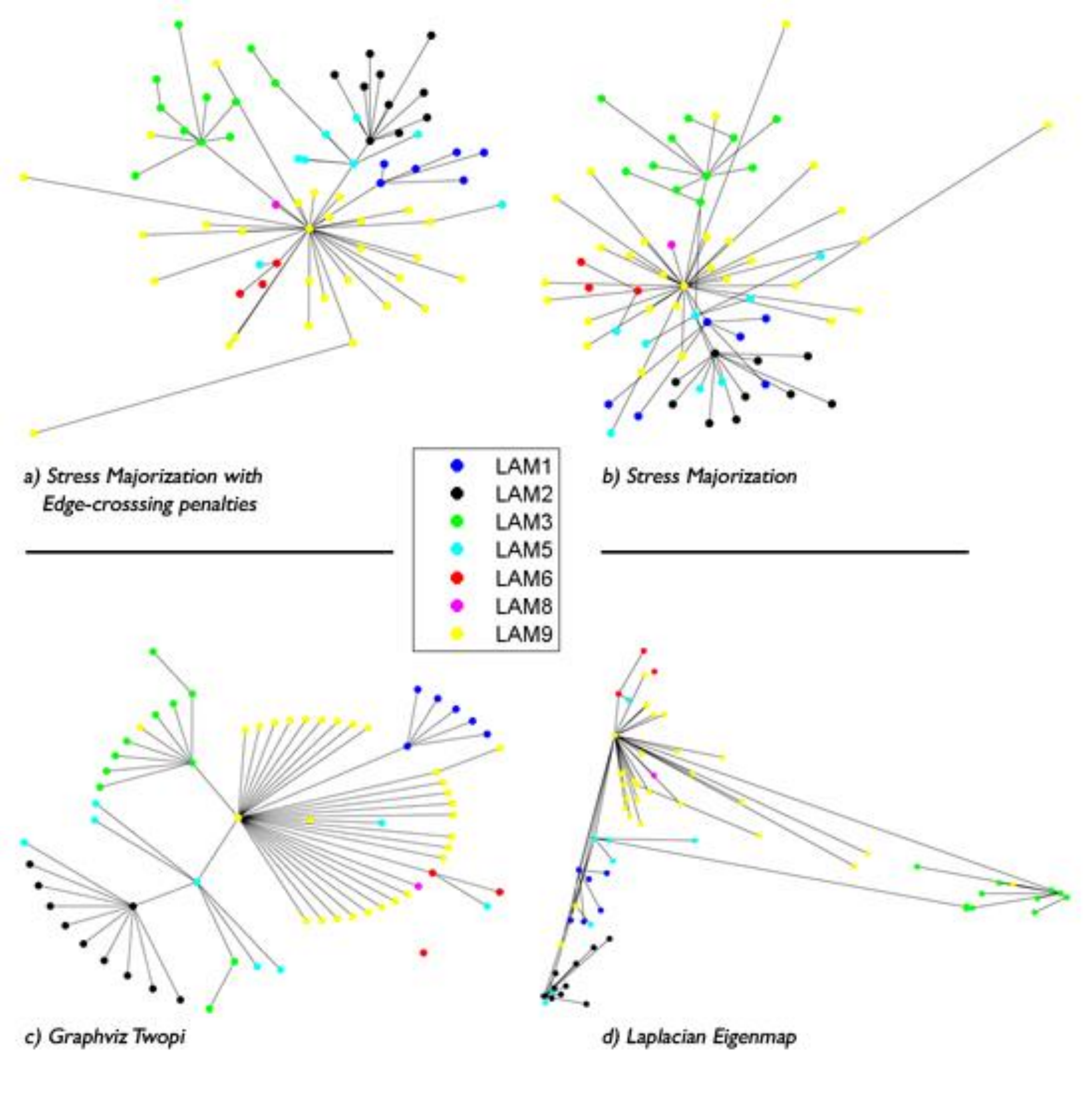
embedding error on *X* (MDS stress)

+ misclassification error based on *U* (SVM max margin)

$$\boxed{\min_{X,u} Stress(X) + \sum_{i=1}^{m} \rho_i [\| (-A_i(X)u_i + 1)_+ \|_1 + \| (B_i(X)u_i + 1)_+ \|_1^1]}$$

Optimize by an alternating algorithm on *X*, *U* using scalable classification and embedding algorithms

Comparison: LAM sublineages



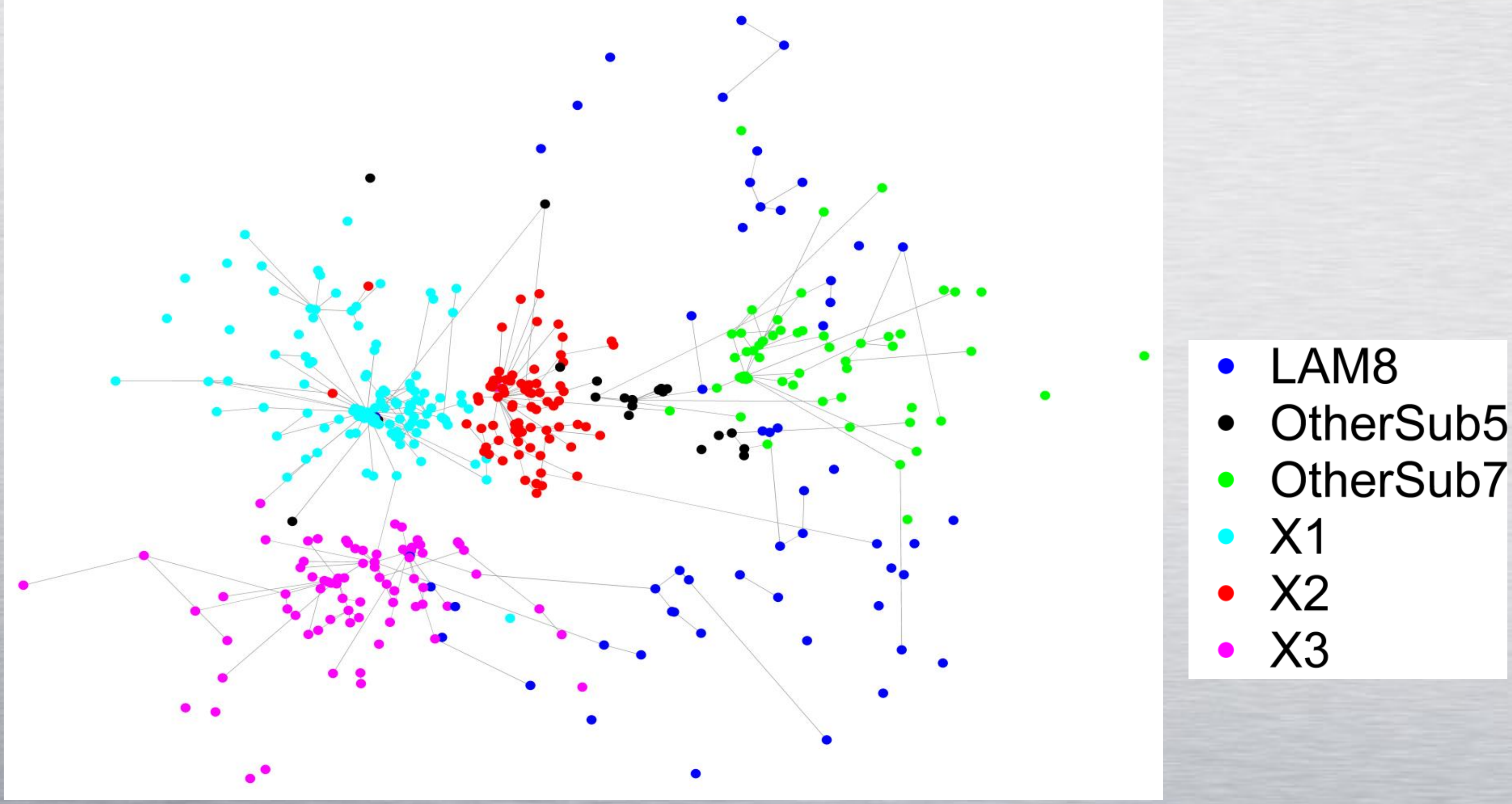
a) *Stress Majorization with Edge-crosssing penalties*

b) *Stress Majorization*

c) *Graphviz Twopi*

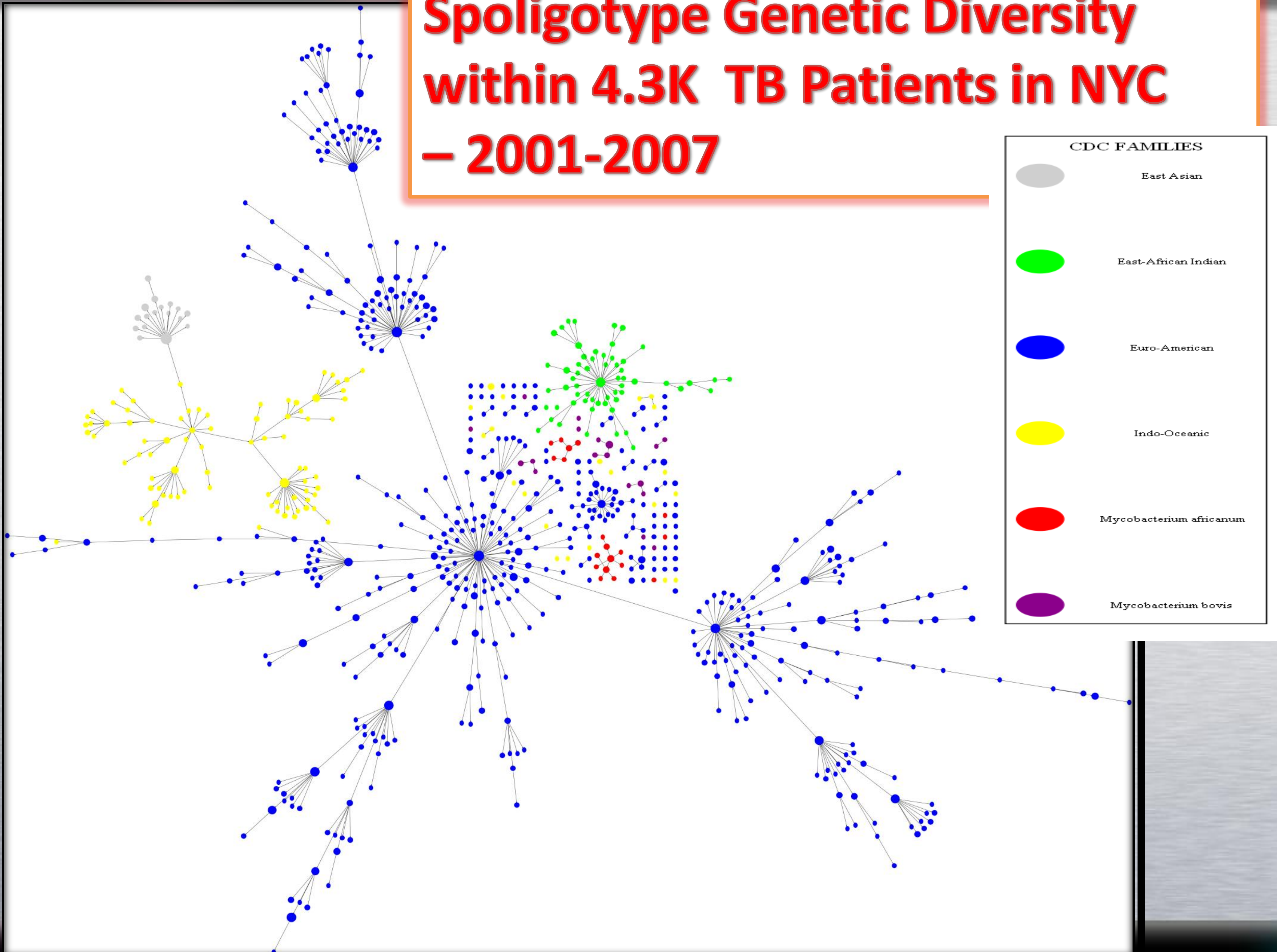
d) *Laplacian Eigenmap*

Visualizing X Lineage adding MIRU and Spoligotype distances



Shabbeer et al, 2012 Multiobjective Embedding, MDS +SVM

Spoligotype Genetic Diversity within 4.3K TB Patients in NYC – 2001-2007



Host-Pathogen Graphs

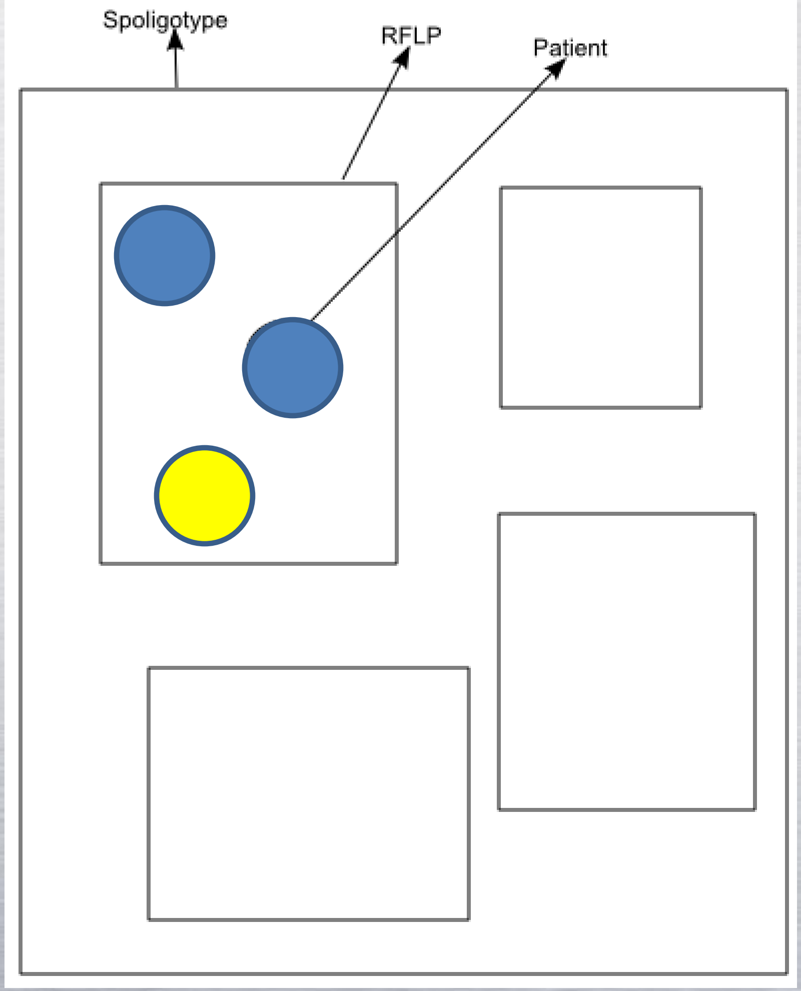
Patients = Circles
DNA Fingerprint = Box

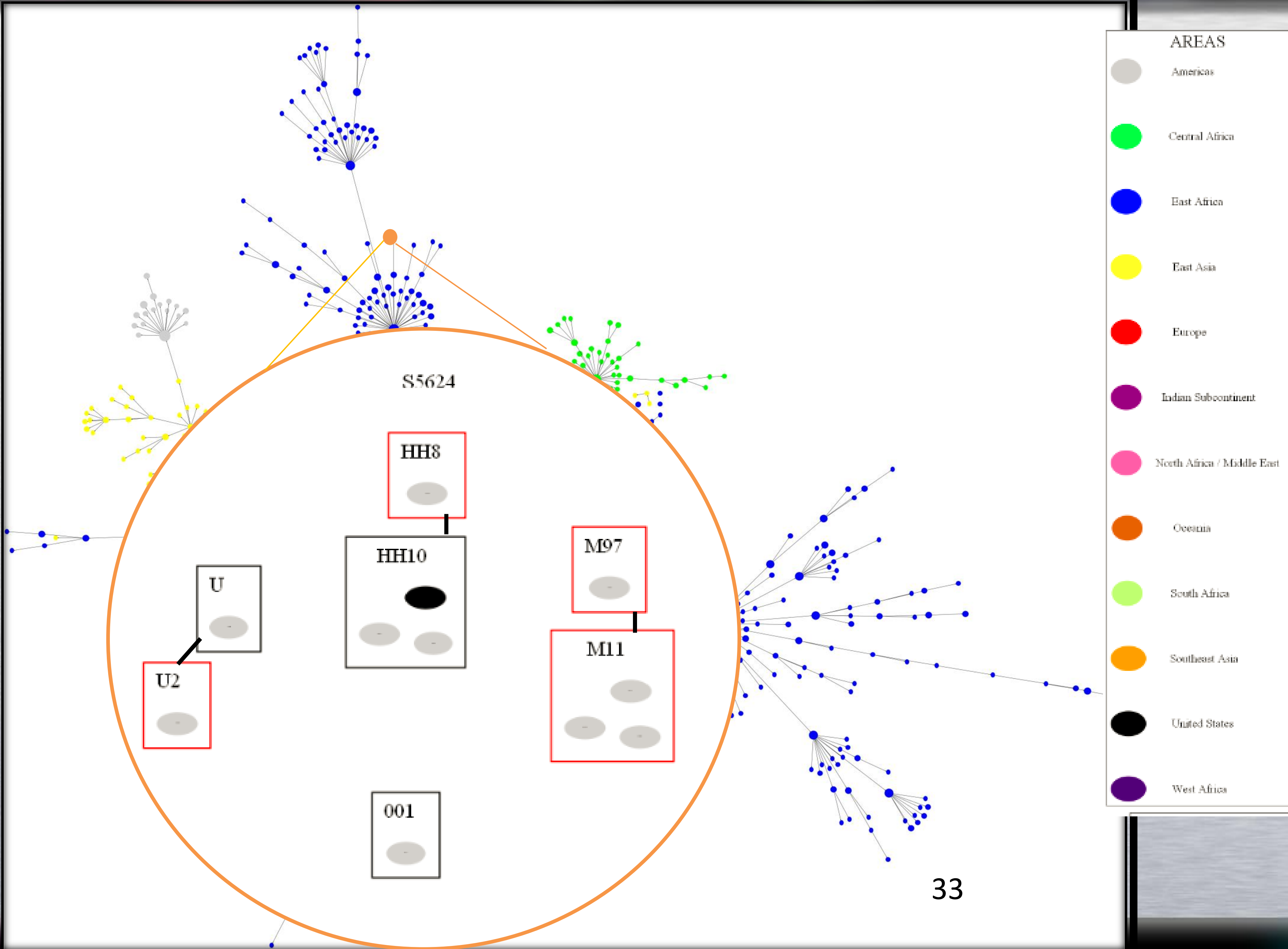
Boxes nested to indicate multiple DNA fingerprints

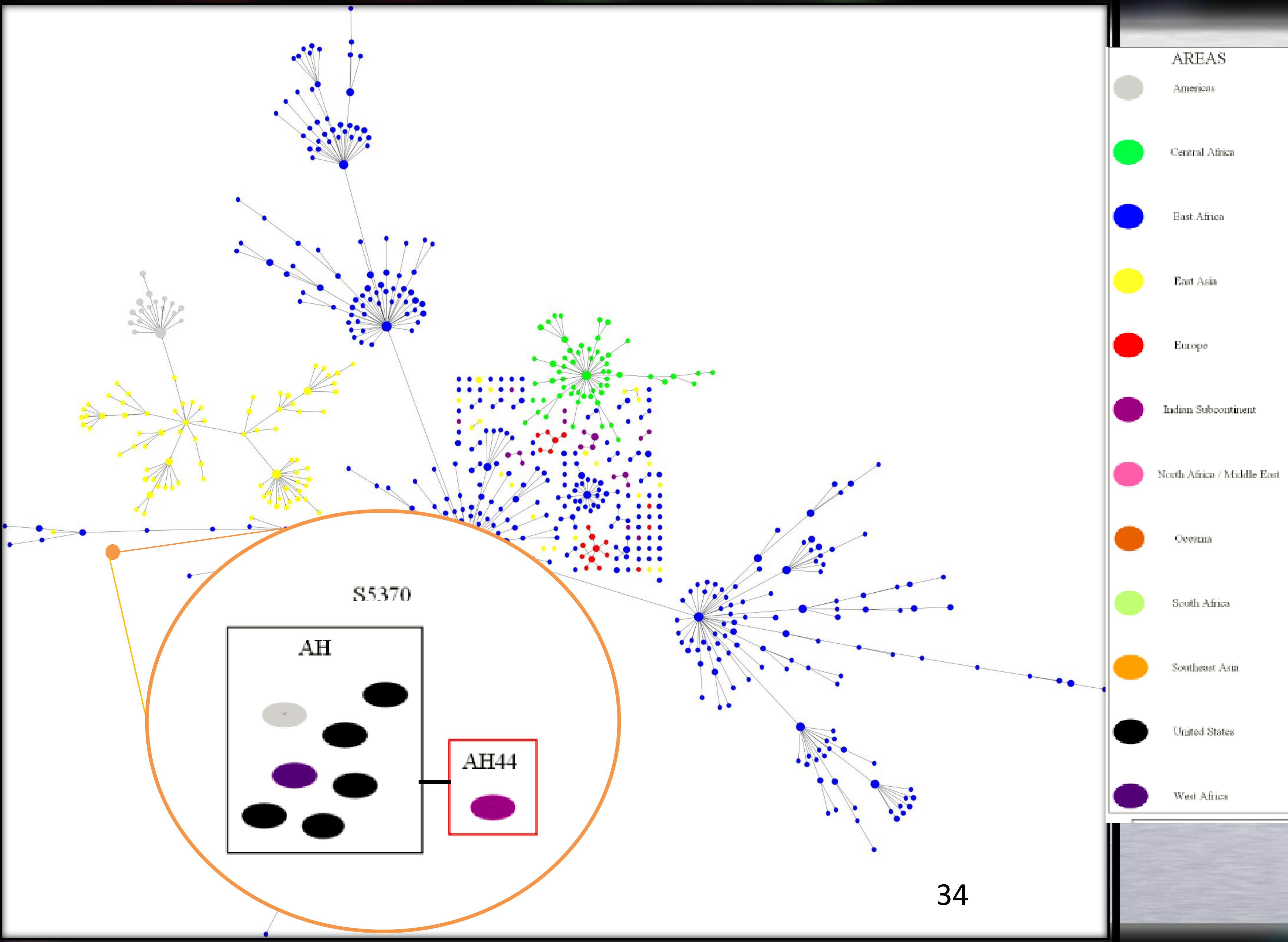
- Other box = spoligotypes
- Inner box = RFLP.

Color by Patient property = Region of Birth

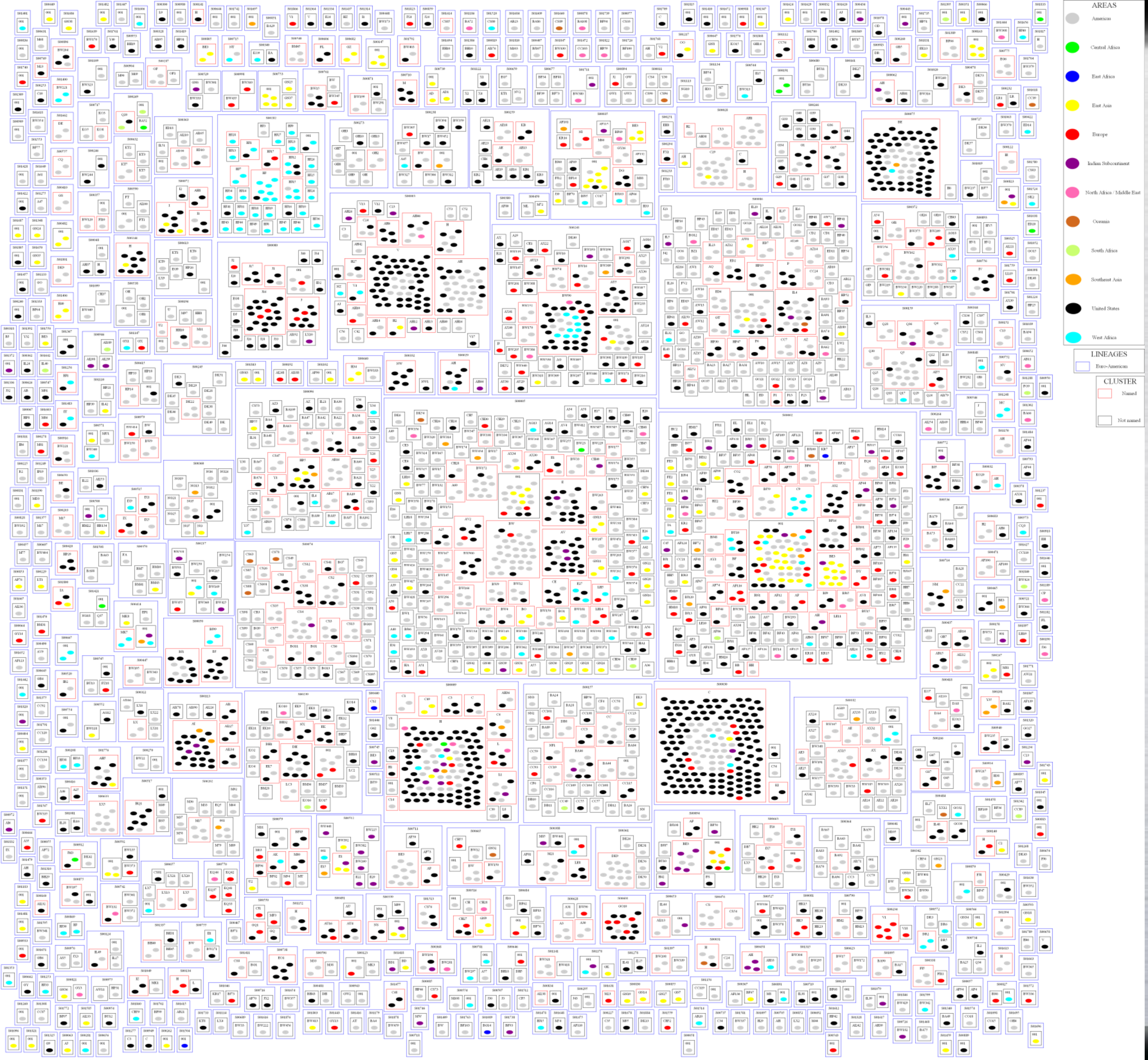
Split by lineage




S5624
HH8
HH10
M97
U
U2
M11
001
AREAS
Americas
Central Africa
East Africa
East Asia
Europe
Indian Subcontinent
North Africa / Middle East
Oceania
South Africa
Southeast Asia
United States
West Africa

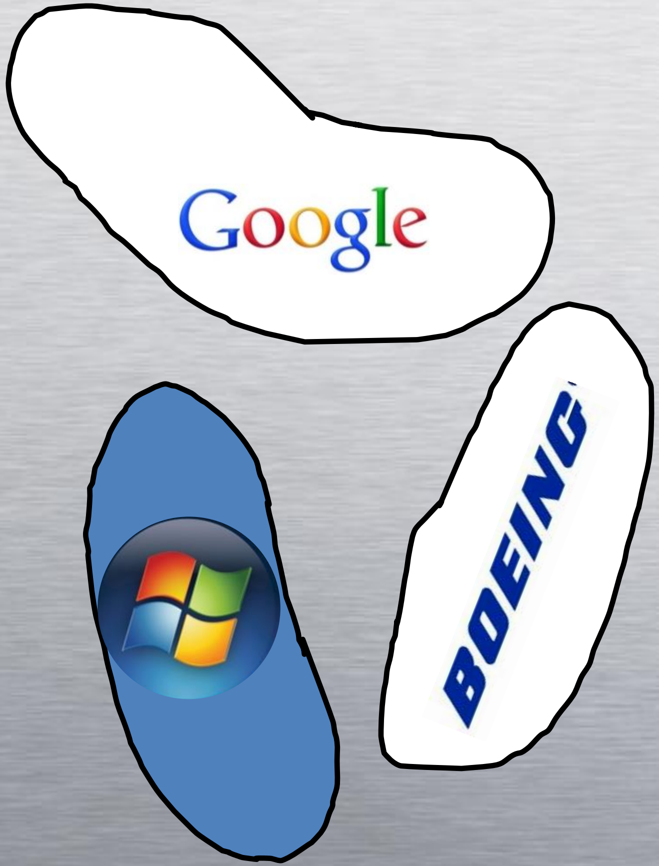

S5370
AH
AH44
AREAS
Americas
Central Africa
East Africa
East Asia
Europe
Indian Subcontinent
North Africa / Middle East
Oceania
South Africa
Southeast Asia
United States
West Africa

Euro American



Disease as Stock Market

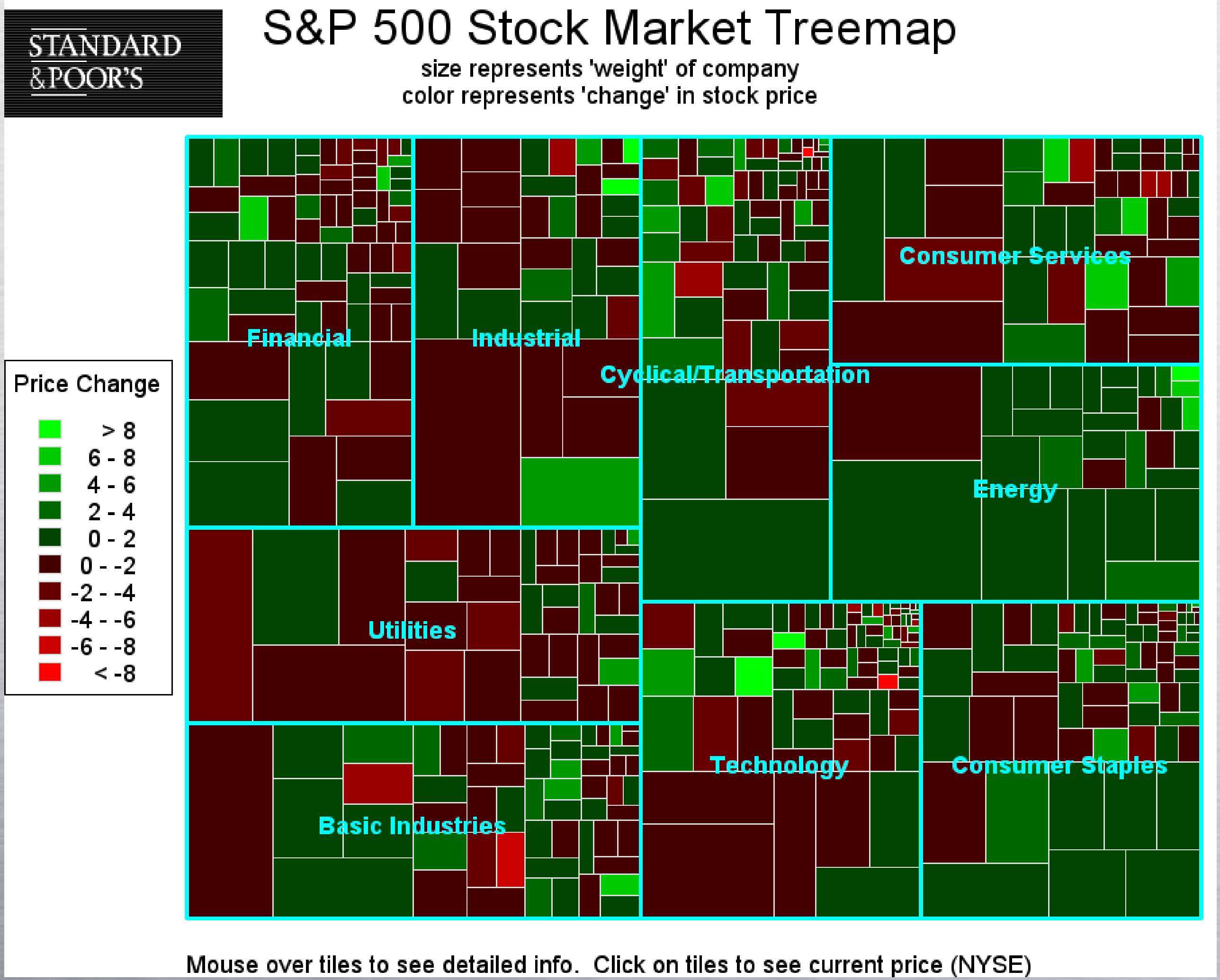
Companies=Bacteria



CDC
UNITED STATES
FEDERAL
TRADE
COMMISSION
BUILDING
1000
800
600
400
8
9
11
1
3
5
7

Buyers = Patients

Stock Market Tree Map



S&P 500 Stock Market Treemap

size represents 'weight' of company
color represents 'change' in stock price

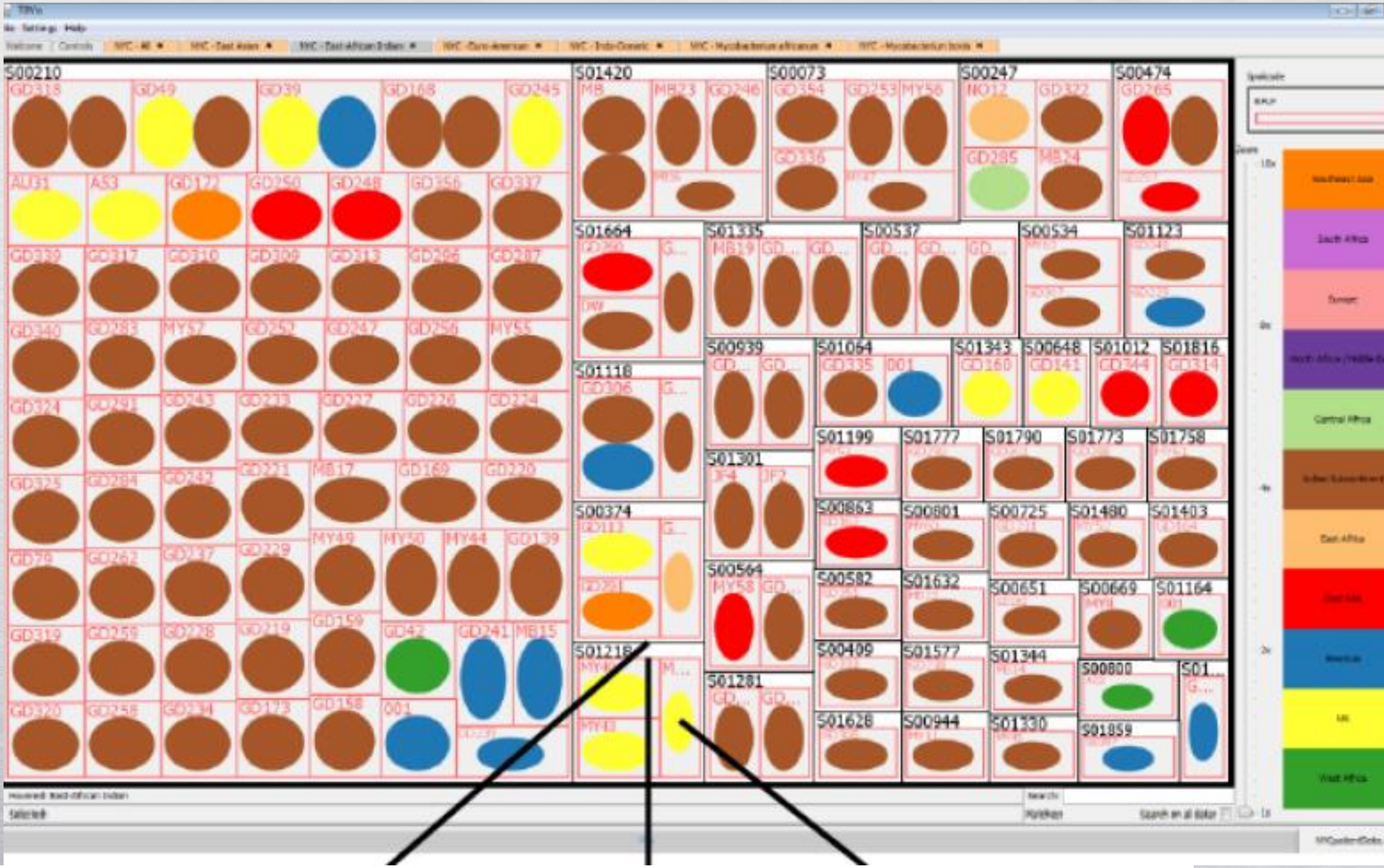
Price Change

🟩	> 8
🟩	6 - 8
🟩	4 - 6
🟩	2 - 4
🟩	0 - 2
🟥	0 - -2
🟥	-2 - -4
🟥	-4 - -6
🟥	-6 - -8
🟥	< -8

Mouse over tiles to see detailed info. Click on tiles to see current price (NYSE)

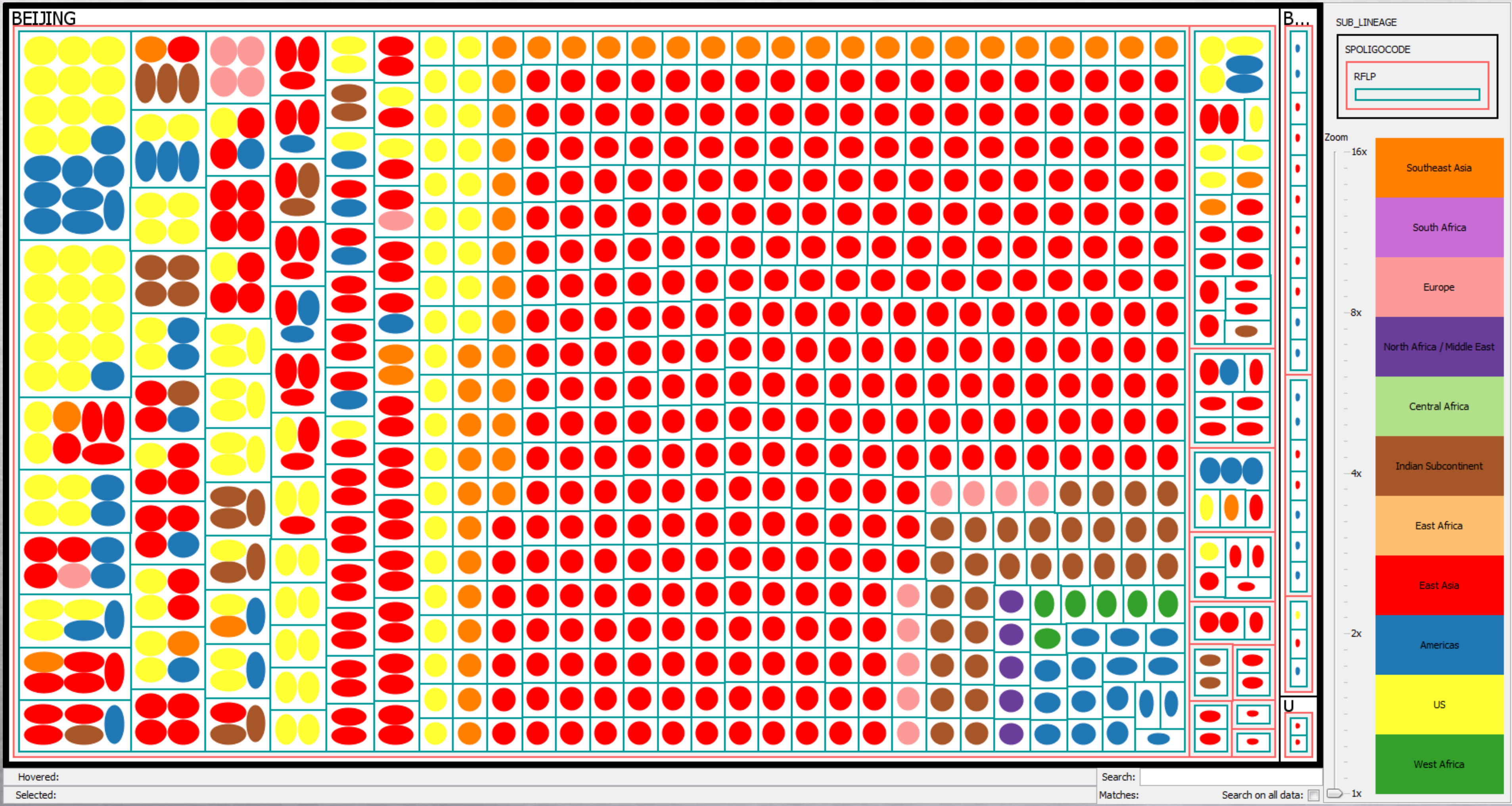
http://robslink.com/SAS/democd9/sp500.htm

Split by Lineage - East African Indian



Patient	ID	Biomarker 1	Biomarker 2	TB continent
Patient 1	105	S00669	MY8	Indian Subcontinent
Patient 2	2443	S00210	GD139	Indian Subcontinent
Patient 3	2452	S00210	MY44	Indian Subcontinent
Patient 4	2487	S00247	NO12	East Africa

NYC - East Asian



Hovered:

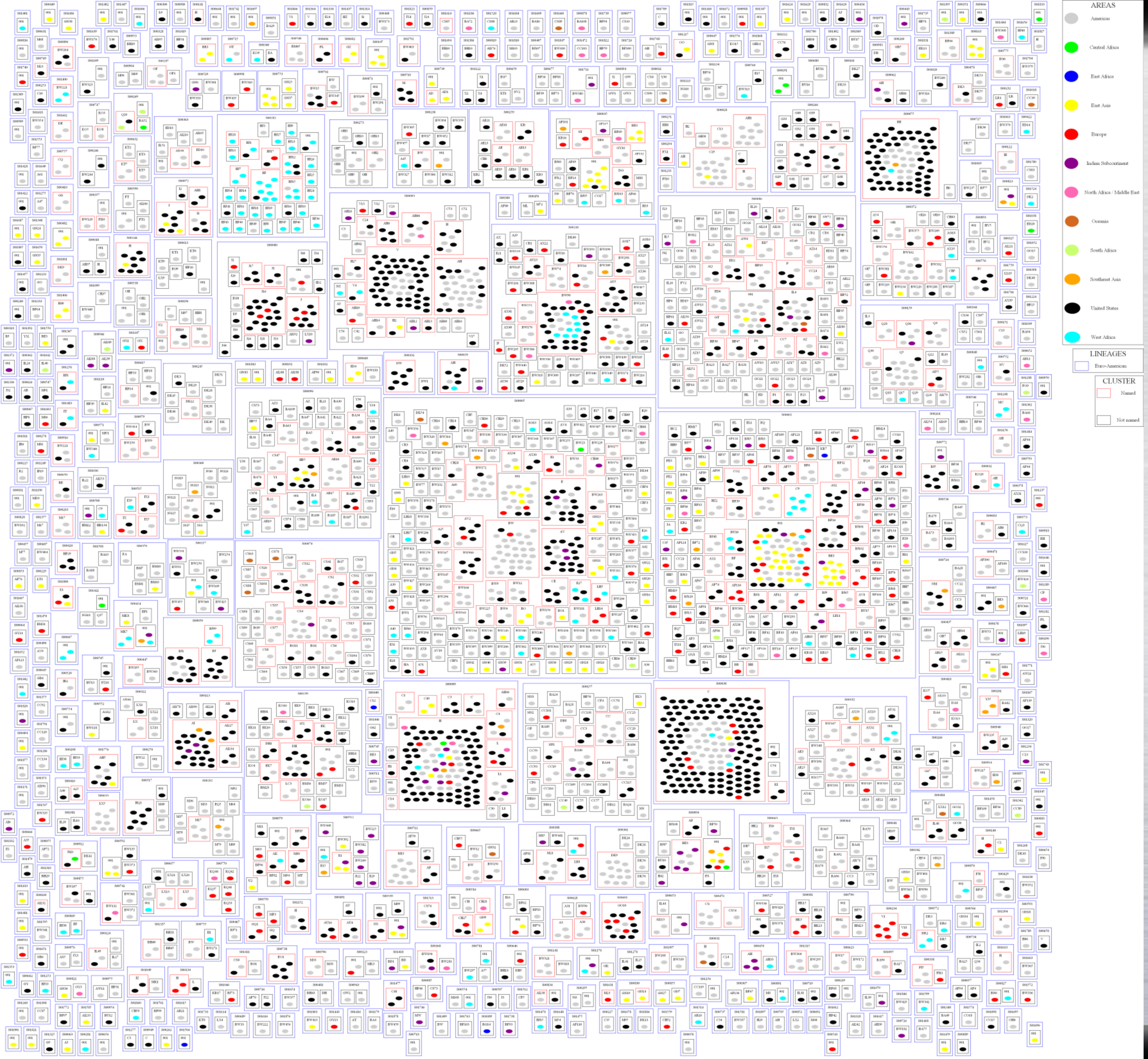
Selected:

Search:

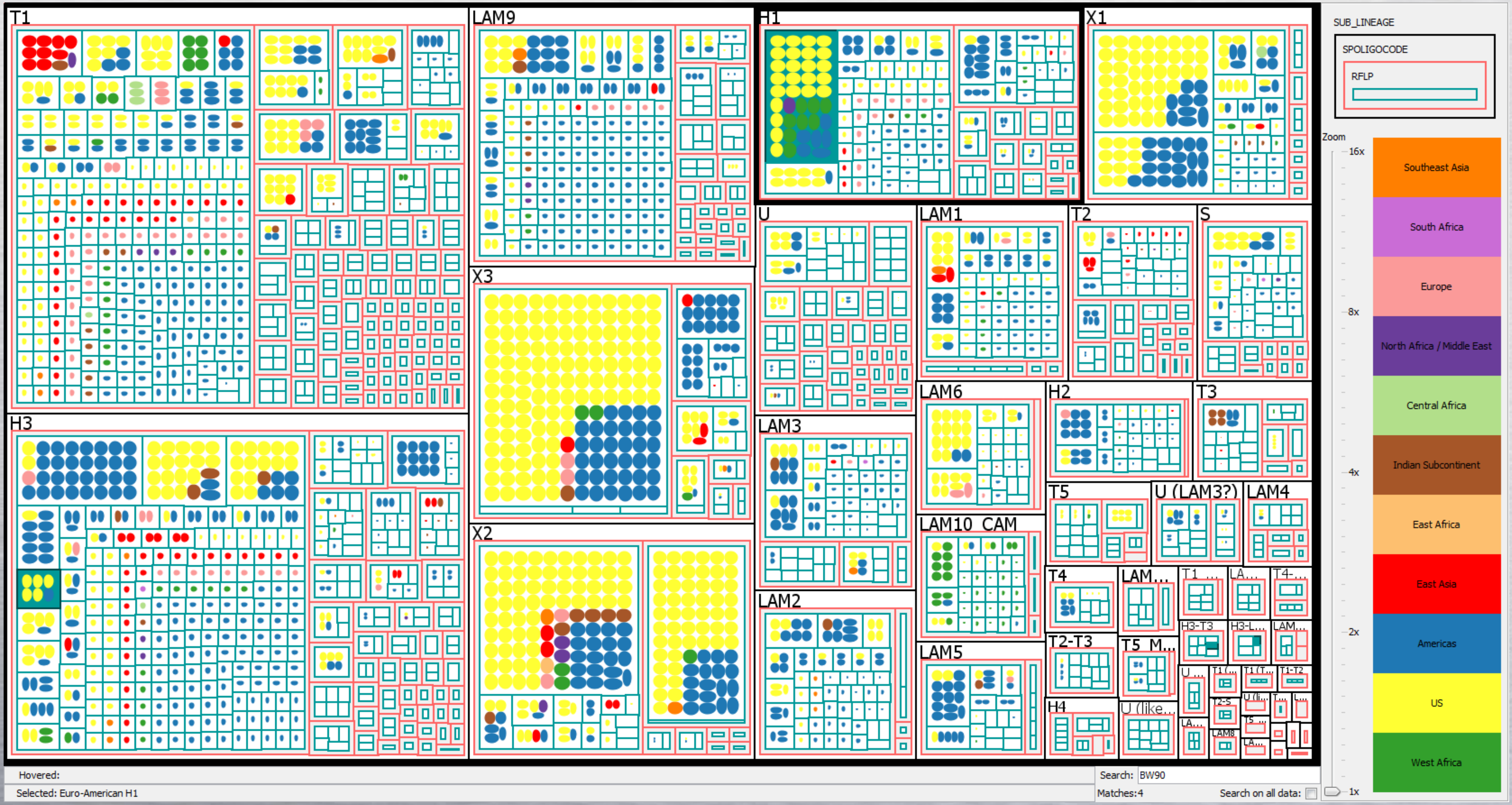
Matches:

Search on all data: □

Euro American



NYC - Euro-American



Hovered:

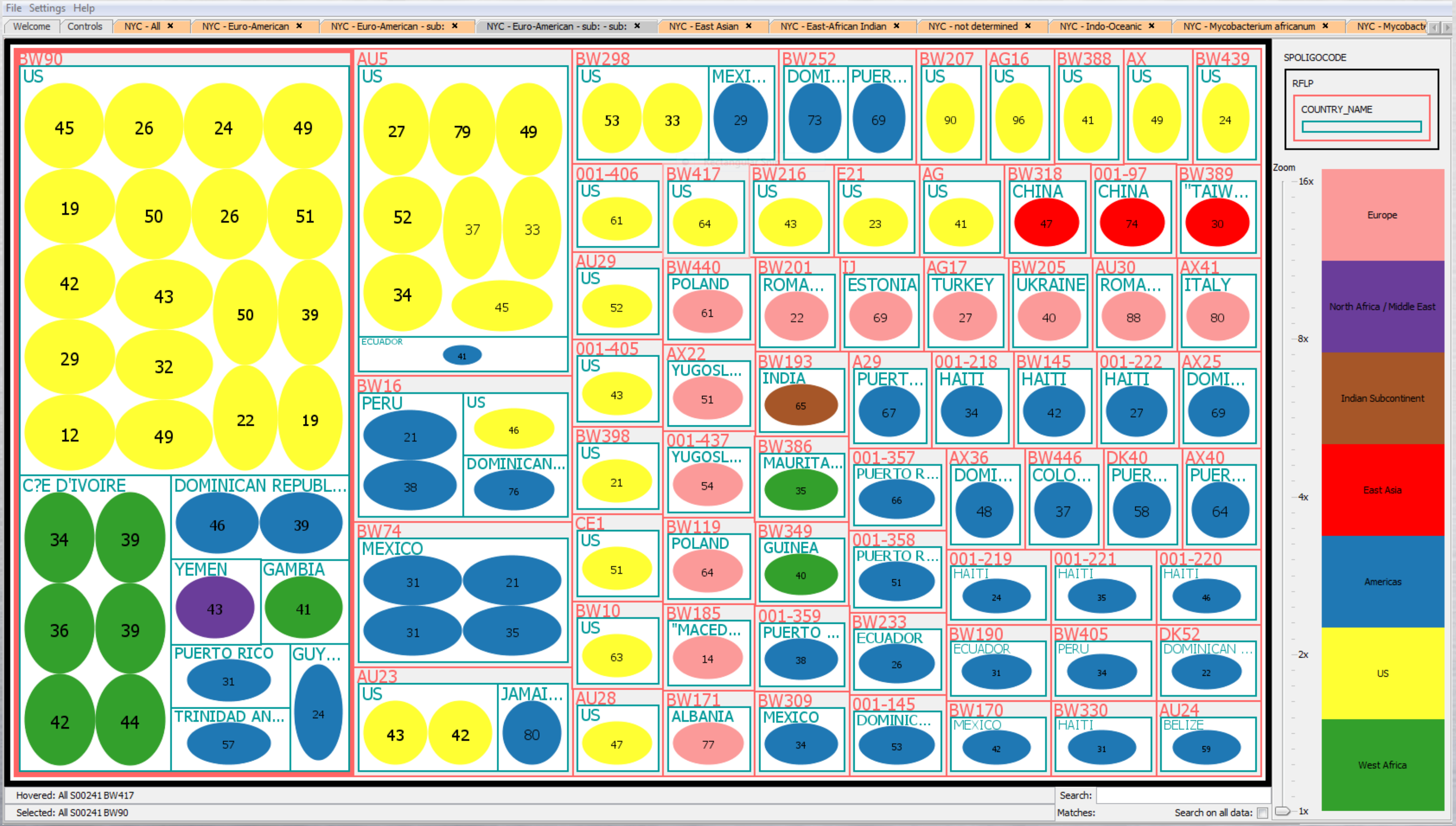
Selected: Euro-American H1

Search: BW90

Matches:4

Search on all data: □

Spoligotype – S00241
BW90 with Patient Age

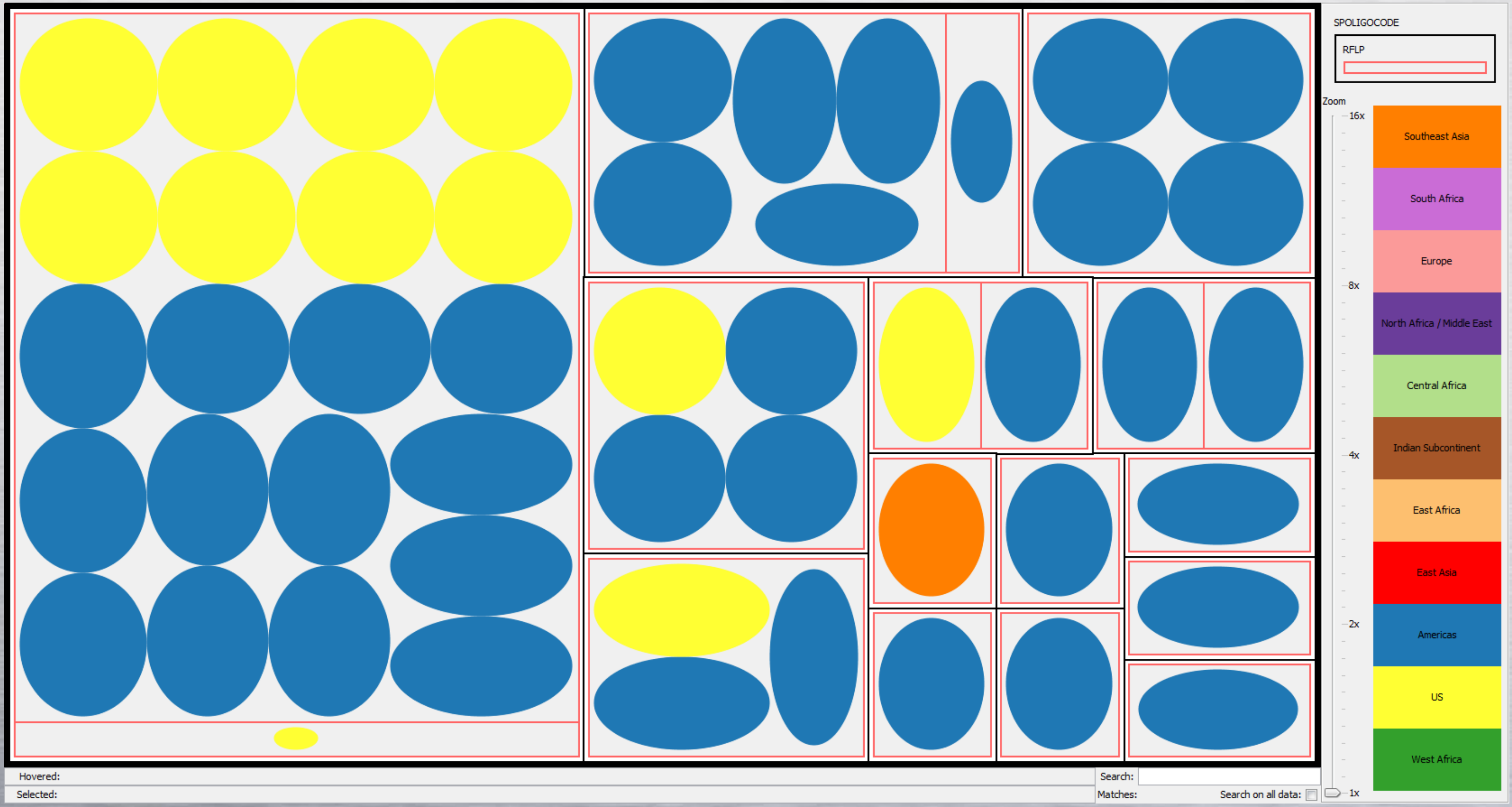


Hovered: All S00241 BW417

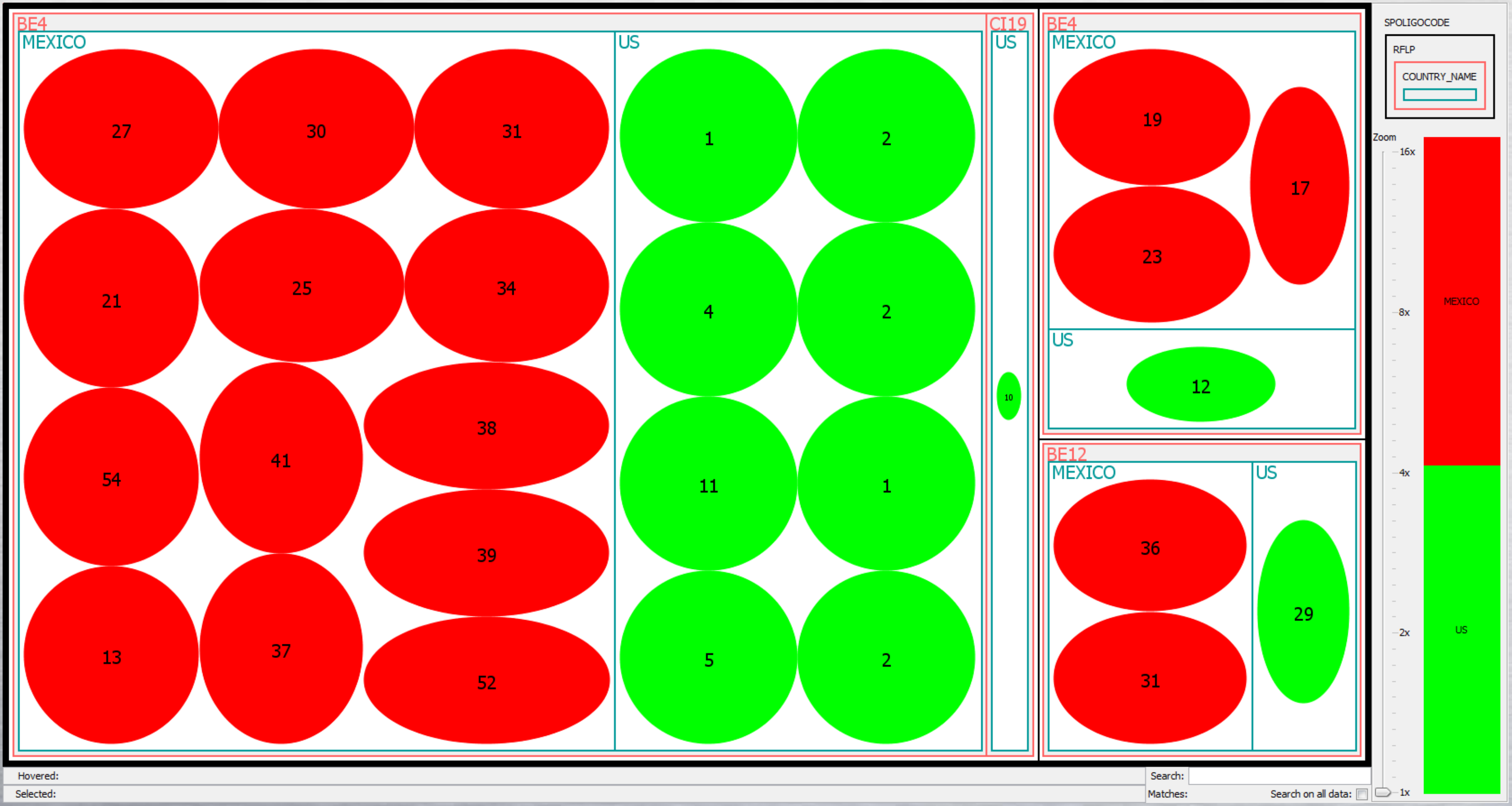
Selected: All S00241 BW90

Ongoing transmission in the US that became resistant to INH antibiotic

NYC - *M. bovis*

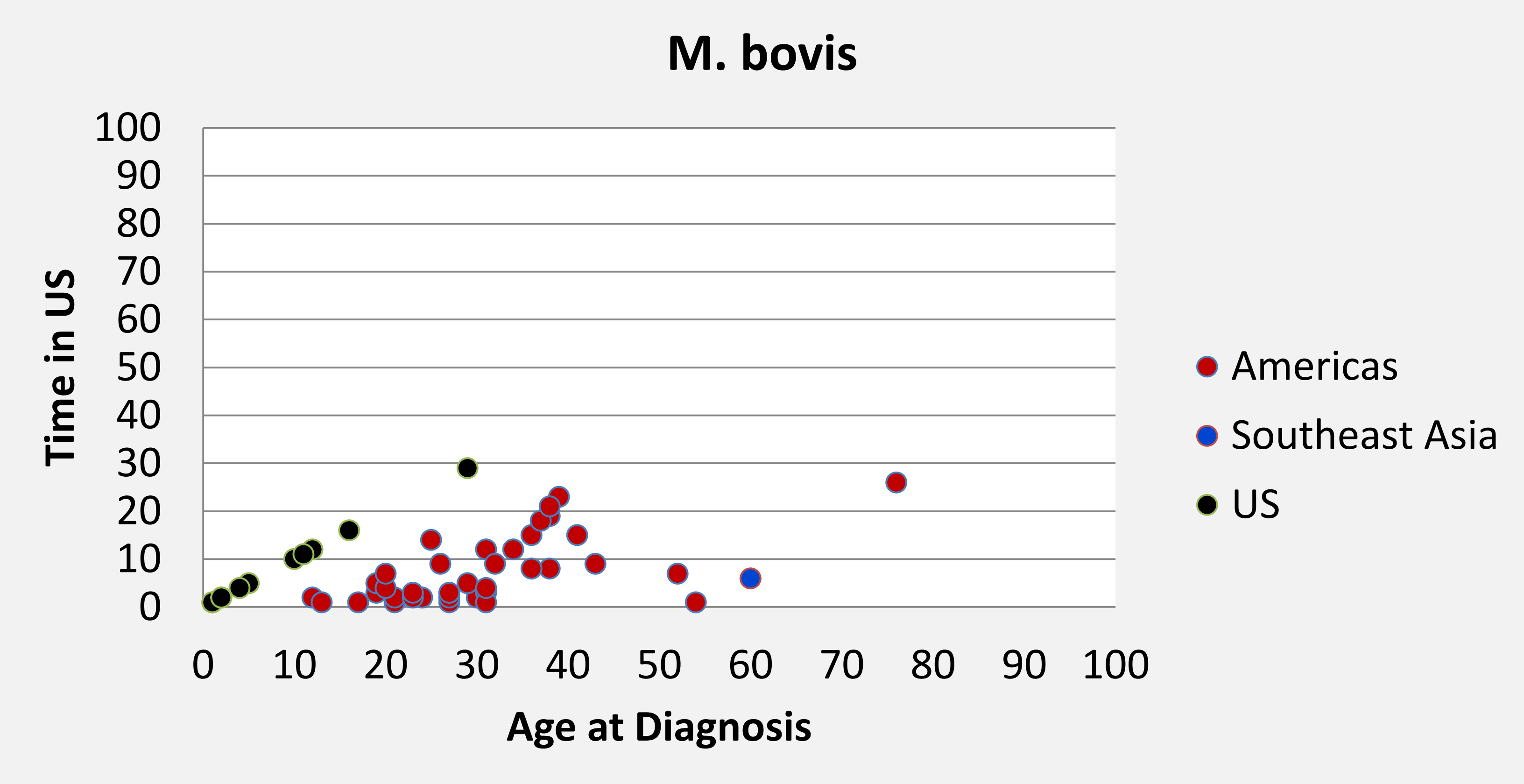

SPOLIGOCODE
RFLP
Zoom
16x
8x
4x
2x
1x
Southeast Asia
South Africa
Europe
North Africa / Middle East
Central Africa
Indian Subcontinent
East Africa
East Asia
Americas
US
West Africa
Hovered:
Selected:
Search:
Matches:
Search on all data: □

M. Bovis with Age of Patient/Country



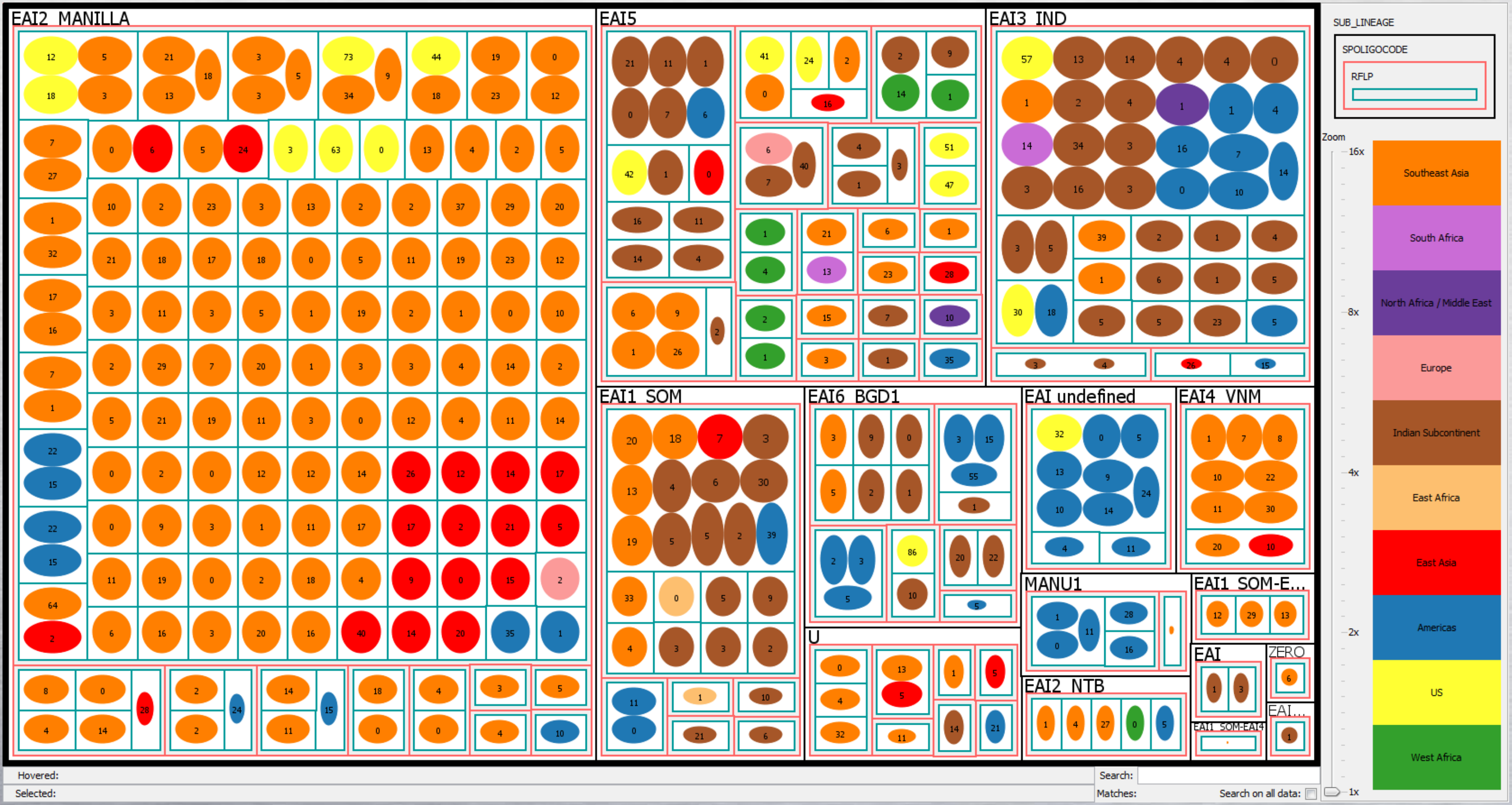
Extremely Young: Outbreak in US-born children of Mexican Parents likely due to unpasteurized cheese.

NYC *M. bovis* (2001-2007)



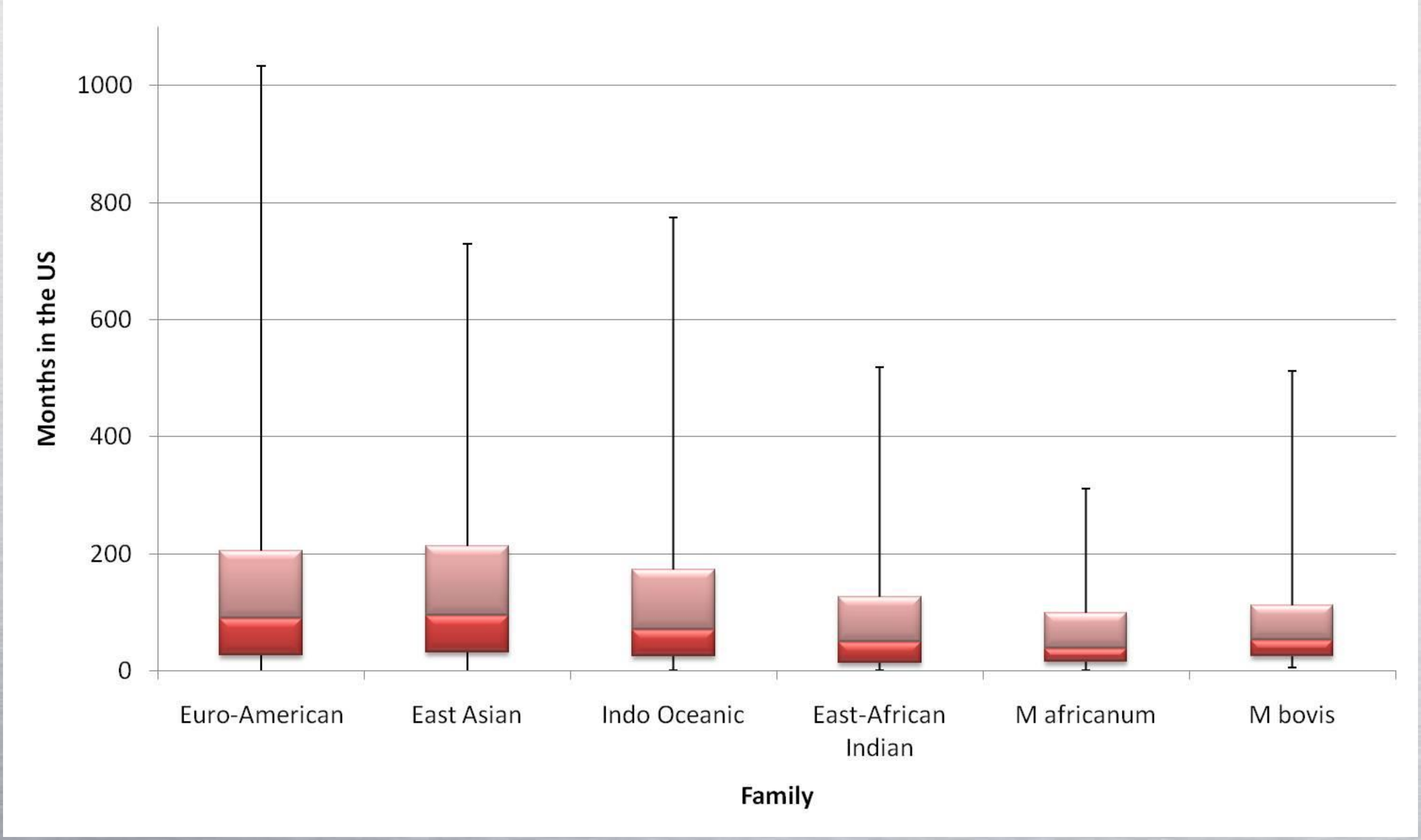
- Extra pulmonary *M. bovis* strikes
  - Mexican Immigrants
  - US-born children of Mexican Immigrants
- Hypothesized caused: Unpasturized cheese

Indo-Oceanic with Time in US



Large clusters with few US patients and no found epi-links.

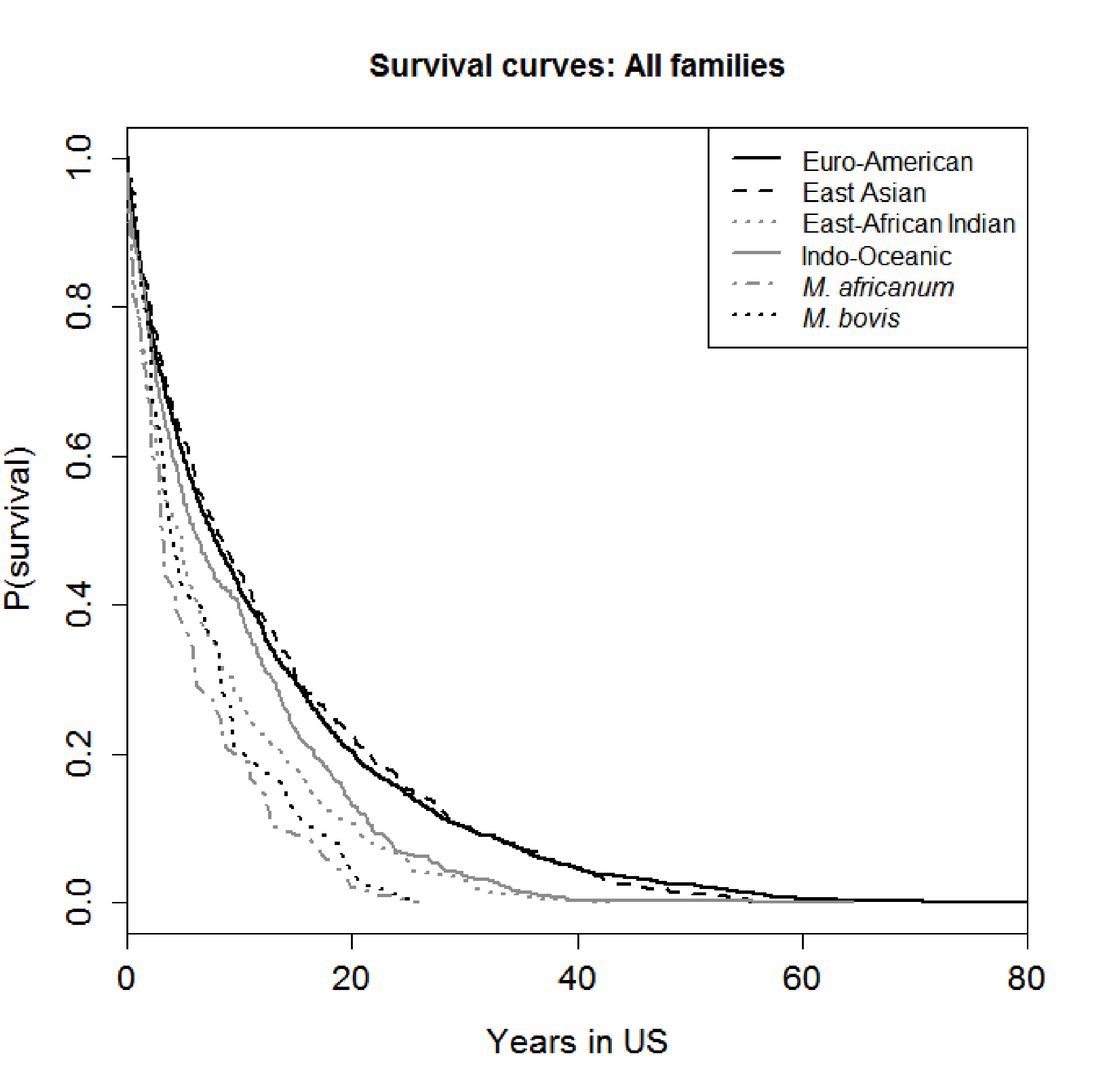
Indo-Oceanic Anomaly



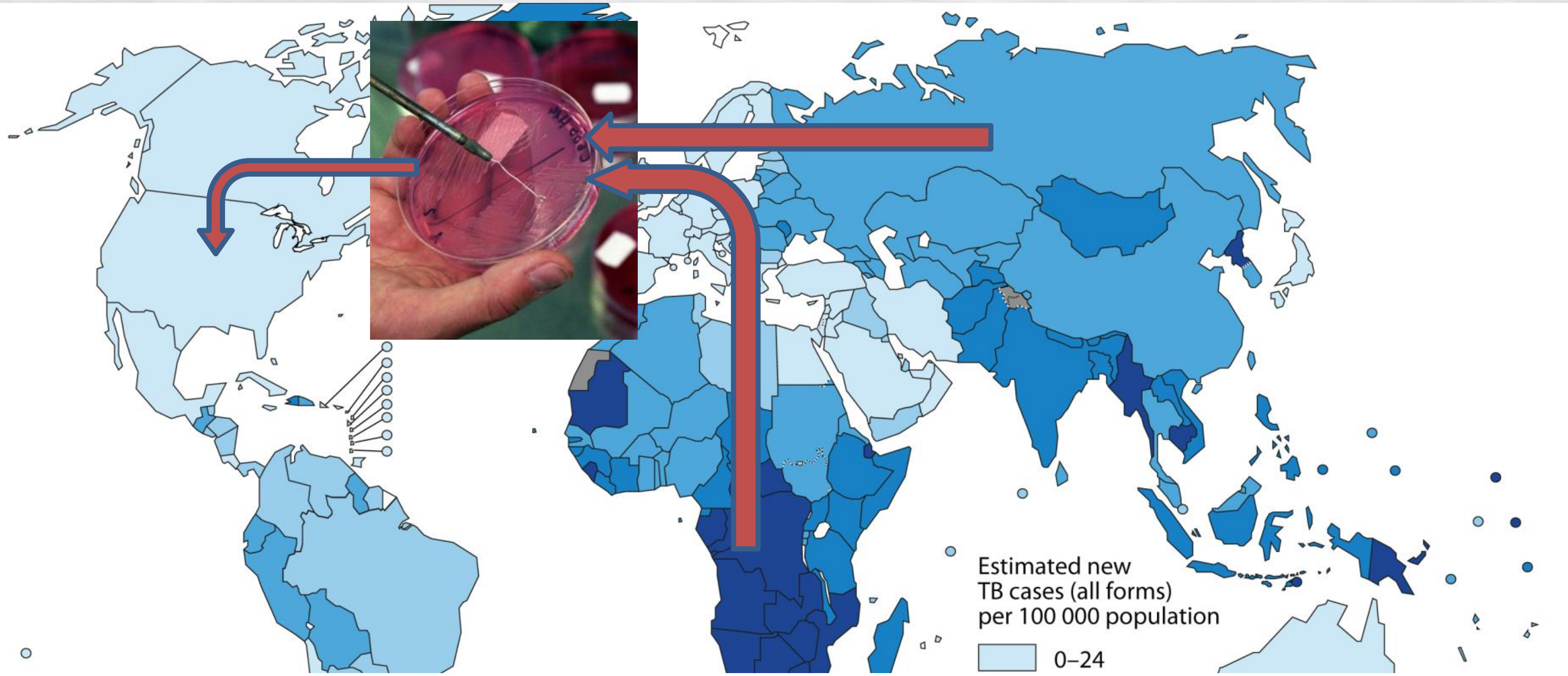
Hypothesized cause: IO strains have longer latency phenotype

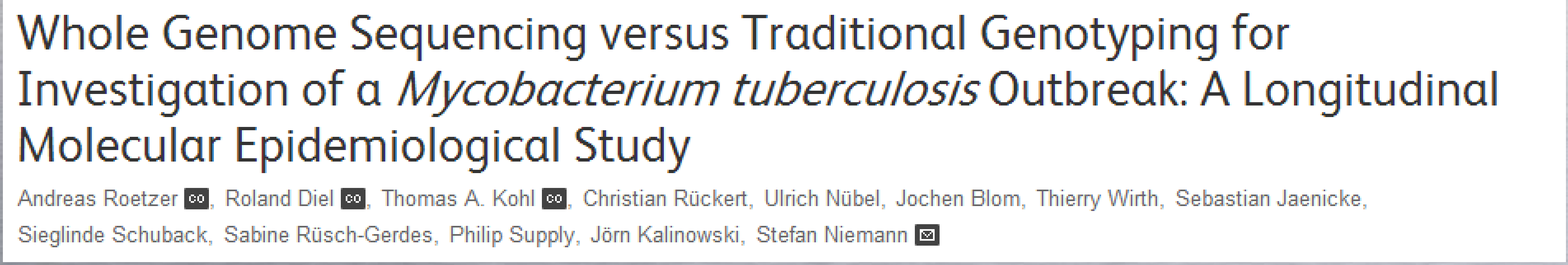
Survival curves for all lineages

Proportion of cases not yet activated



Surveillance Data can reveal novel phenotypes and genotypes





Whole Genome Sequencing versus Traditional Genotyping for Investigation of a *Mycobacterium tuberculosis* Outbreak: A Longitudinal Molecular Epidemiological Study

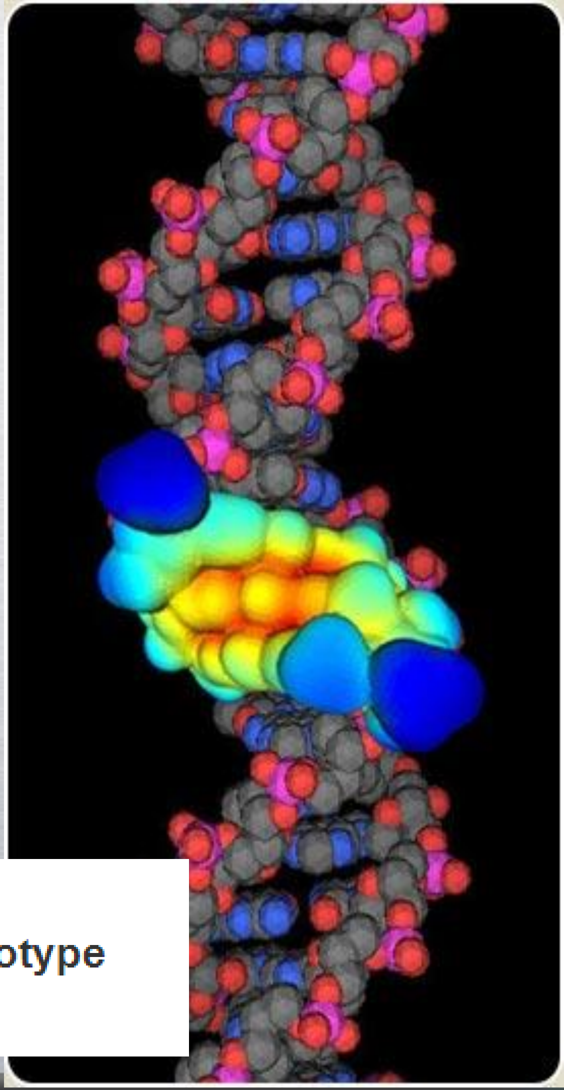
Andreas Roetzer, Roland Diel, Thomas A. Kohl, Christian Rückert, Ulrich Nübel, Jochen Blom, Thierry Wirth, Sebastian Jaenicke, Sieglinde Schuback, Sabine Rüsch-Gerdes, Philip Supply, Jörn Kalinowski, Stefan Niemann

Challenges of Disease Control using Molecular Epidemiology

- **Coupling human intelligence and analytics to help public health care workers control disease**
- **Informing local decisions with global data**
- **Allocating scarce control resources effectively by predicting disease dynamics**
- **Incorporating rapidly evolving data**
  - Contact Investigations
  - New biomarkers for pathogen/host
  - Electronic Medical Records
  - Social media
- **Getting the biology right**
- **Preserving Privacy**

Personalized Medicine based on Host and Pathogen DNA

- Discovery of Host/Pathogen Coadaptation
- Control and treatment efforts guided by host and pathogen DNA
- Better models for drug development, etc

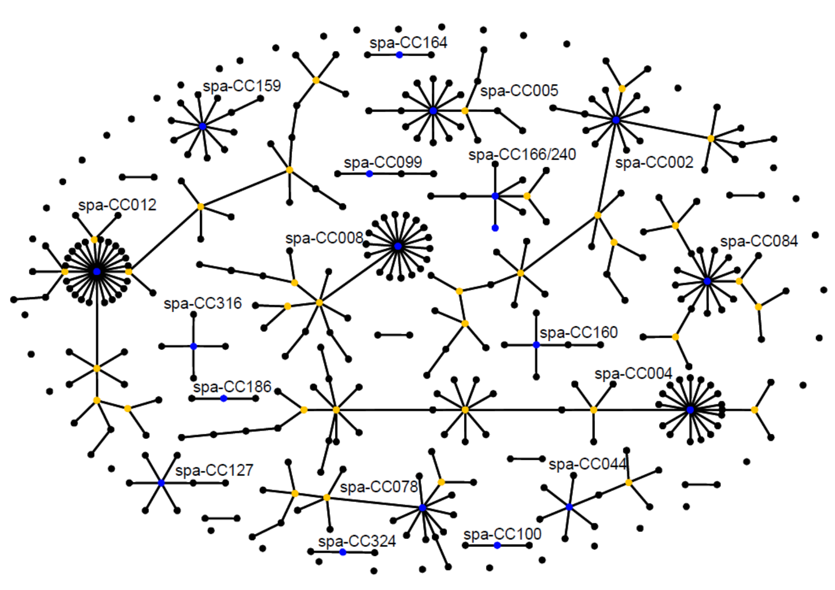


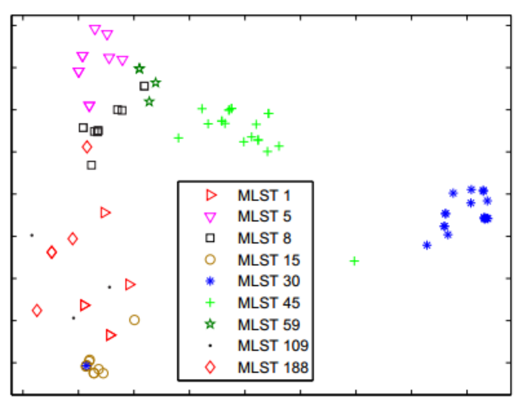
Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype

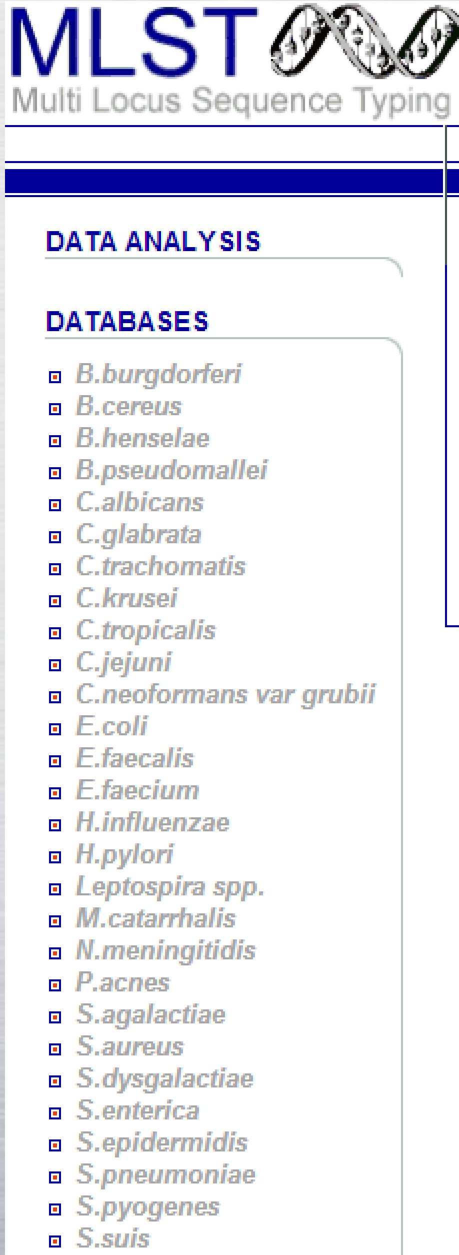
Thorax, 2013

Other Diseases/Pathogens

MRSA







MLST
Multi Locus Sequence Typing

DATA ANALYSIS

DATABASES

- B.burgdorferi
- B.cereus
- B.henselae
- B.pseudomallei
- C.albicans
- C.glabrata
- C.trachomatis
- C.krusei
- C.tropicalis
- C.jejuni
- C.neoformans var grubii
- E.coli
- E.faecalis
- E.faecium
- H.influenzae
- H.pylori
- Leptospira spp.
- M.catarrhalis
- N.meningitidis
- P.acnes
- S.agalactiae
- S.aureus
- S.dysgalactiae
- S.enterica
- S.epidermidis
- S.pneumoniae
- S.pyogenes
- S.suis
- V.vulnificus

Mellmann *et al. BMC Microbiology* 2007, CDC, Agius et al, *IEEE/ACM Trans. on Comp. Bio.*, 2007.

TB-Insight Project Team
Supported by NIH R01-LMN009731

RPI:

- Professors: K. Bennett and B. Yener
- Graduate Research Assistants:

  Amina Shabbeer, **James Blondin**, **Inna Vitol,** Janani Ranganathan, Srivatsan Raghavan, Cagri Ozcaglar, Chris Gatti
- Postdoc: **Minoo Aminian**
- Undergraduate Researchers: Eric Dubois, Kane Hadley, Michael

During this talk approximately
998 people developed active TB
276 people died of TB

Lauren Cowan, Jeff Driscoll, *US Centers for Disease Control*

Natalia Kurepina, Barry Kreiswirth, *Public Health Research Institute*

Nalin Rastogi, Phillip Supply *Institut Pasteur*

Vincent Escuyer, New York State Department of Health

Shama Ahuja, Bianca Perri, Jeanne Sullivan

New York City Department of Health

TB insight
Tuberculosis Tracking and Control