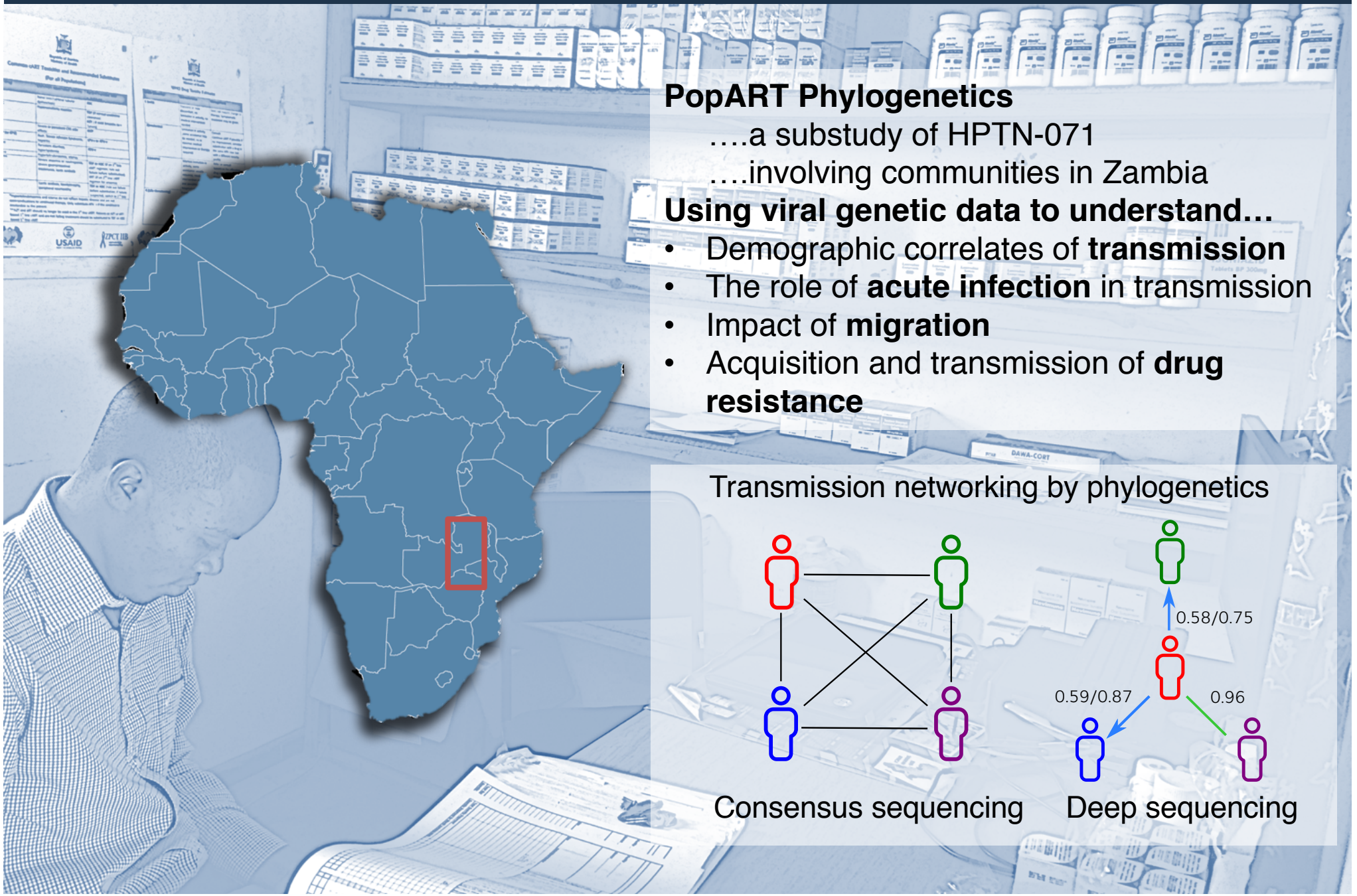
Using NGS data to identify HIV drug resistance mutations

webinar: David Bonsall

PopART Phylogenetics



PopART Phylogenetics

….a substudy of HPTN-071

….involving communities in Zambia

Using viral genetic data to understand…

- Demographic correlates of **transmission**
- The role of **acute infection** in transmission
- Impact of **migration**
- Acquisition and transmission of **drug resistance**

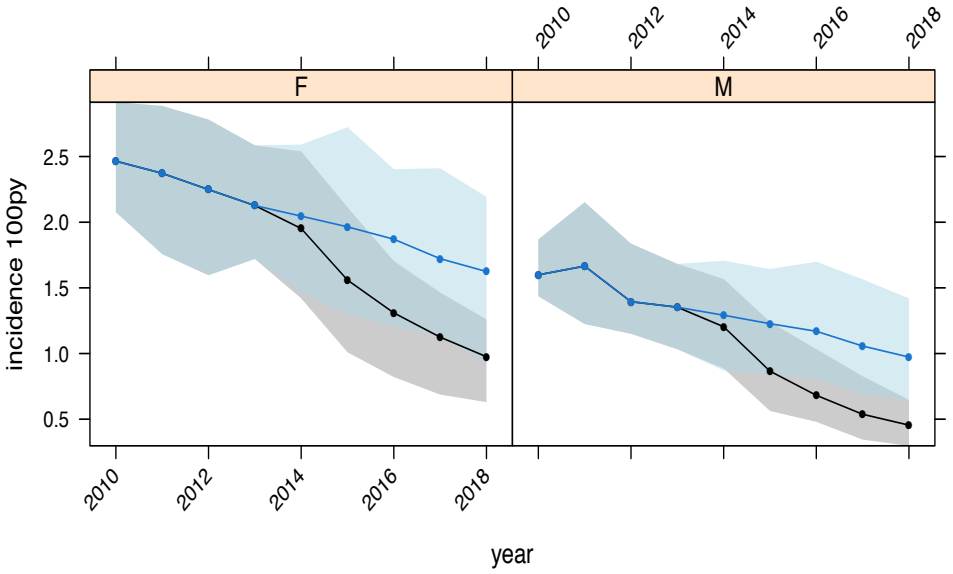
Transmission networking by phylogenetics

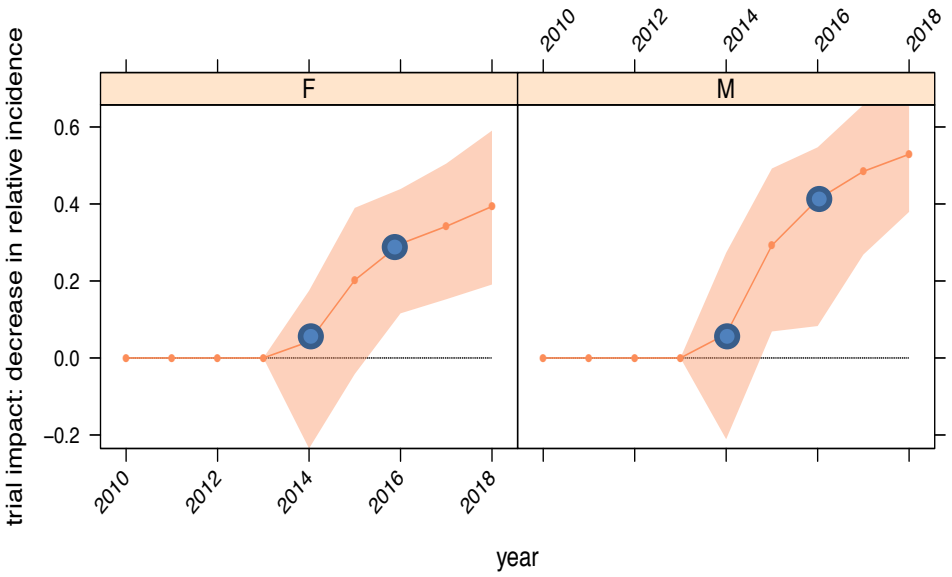
Consensus sequencing

Deep sequencing

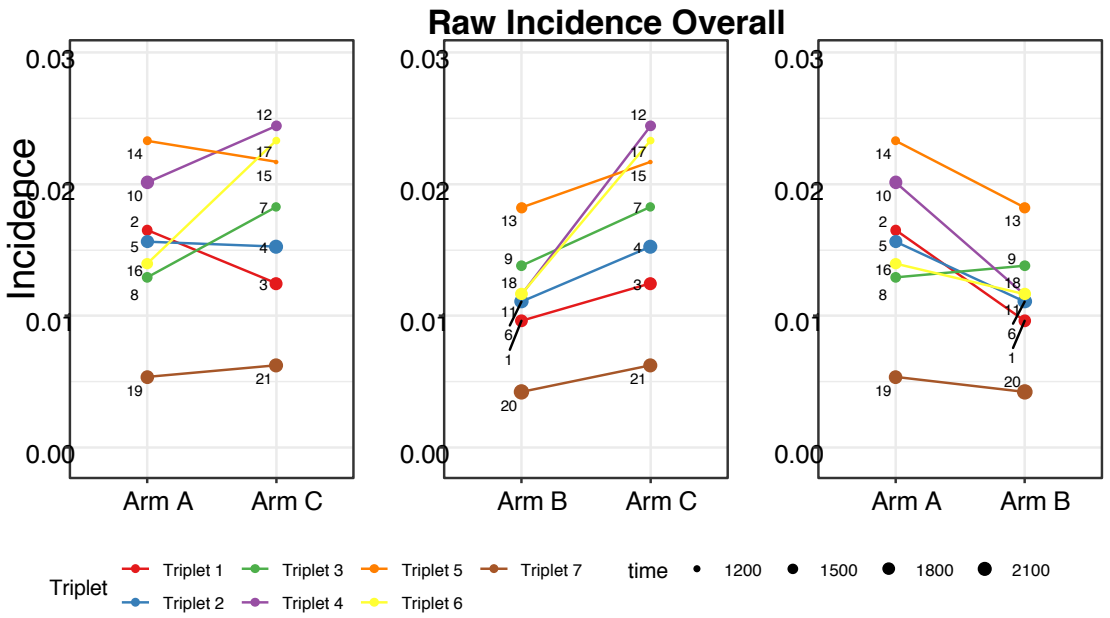
Primary outcome

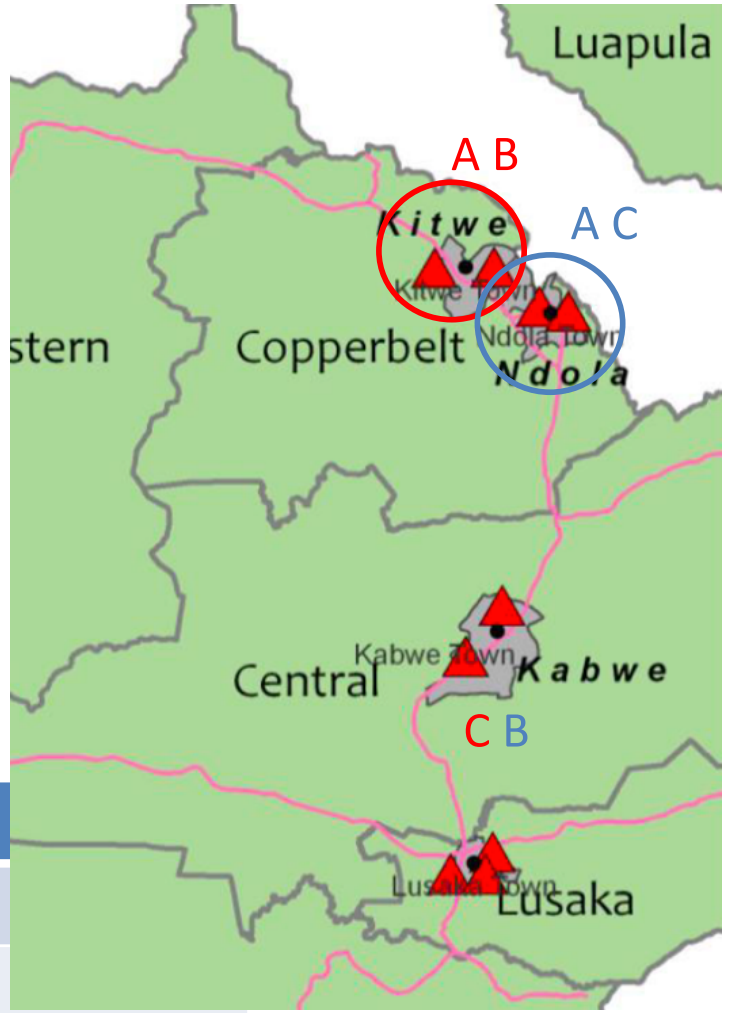
- Relative (reduction) in incidence between arm A and arm C communities observed between PC 12 and PC 36





Difference between A and B?





Triplet	A	B	C	
1	2*	1*	3**	
2	5*	6**	4*	
5	14	13	15	Metro
7	19	20	21	Winelands

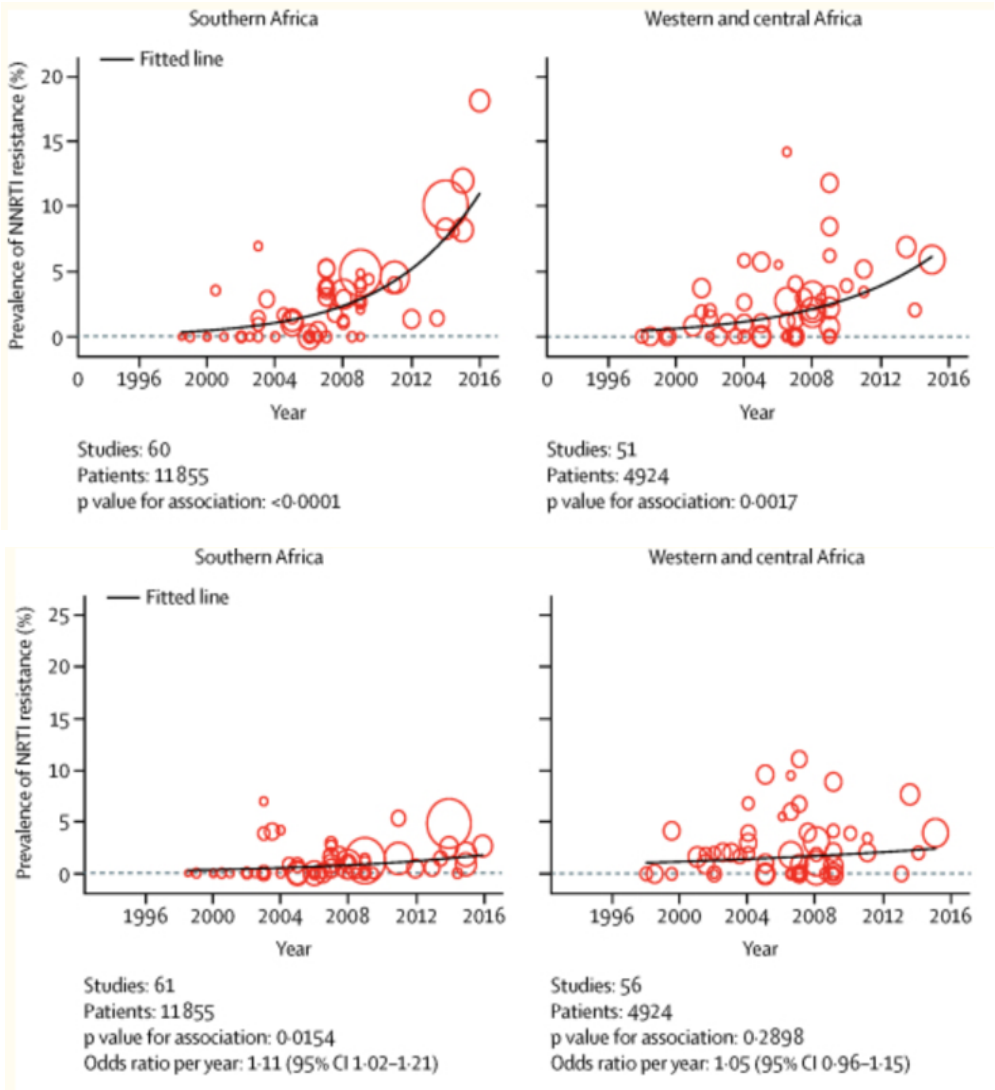
* Copperbelt, ** Central

PopART HPTN 071

- Drug Resistance in PopART

  - What is the prevalence of drug resistance within PopART at baseline?
  - To what extent did drug resistance limit effectiveness of treatment-based prevention of new infections?
  - Did drug resistance contribute to heterogeneity observed in the trials primary outcome?
  - (How did PopART impact incidence of drug resistant infections?)

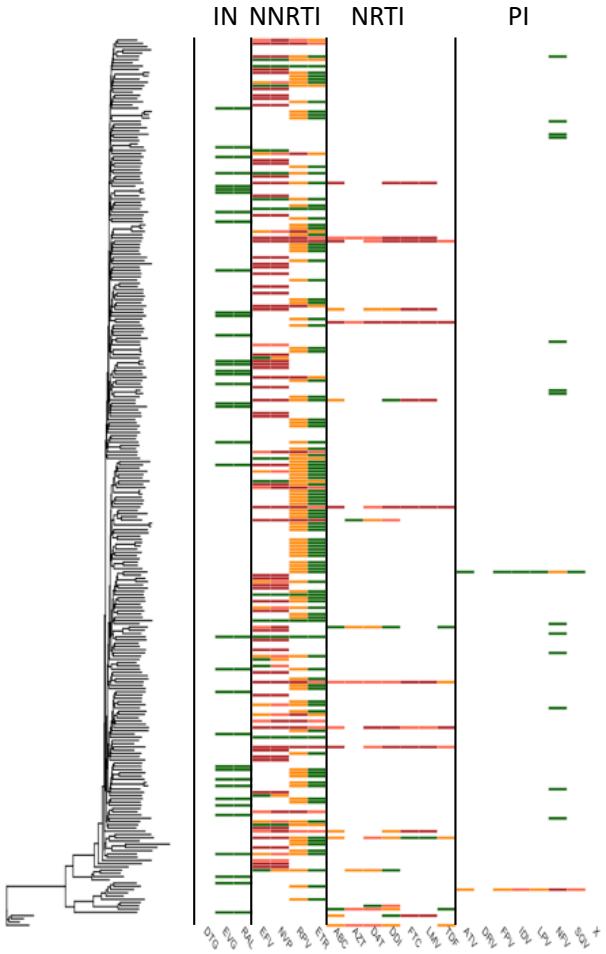
HIV Drug Resistance – An evolving problem



Gupta RK, Lancet Infect Dis. 2018

Phylogenetics of drug resistance: Populations

POPART (Zambia – mostly clade C)

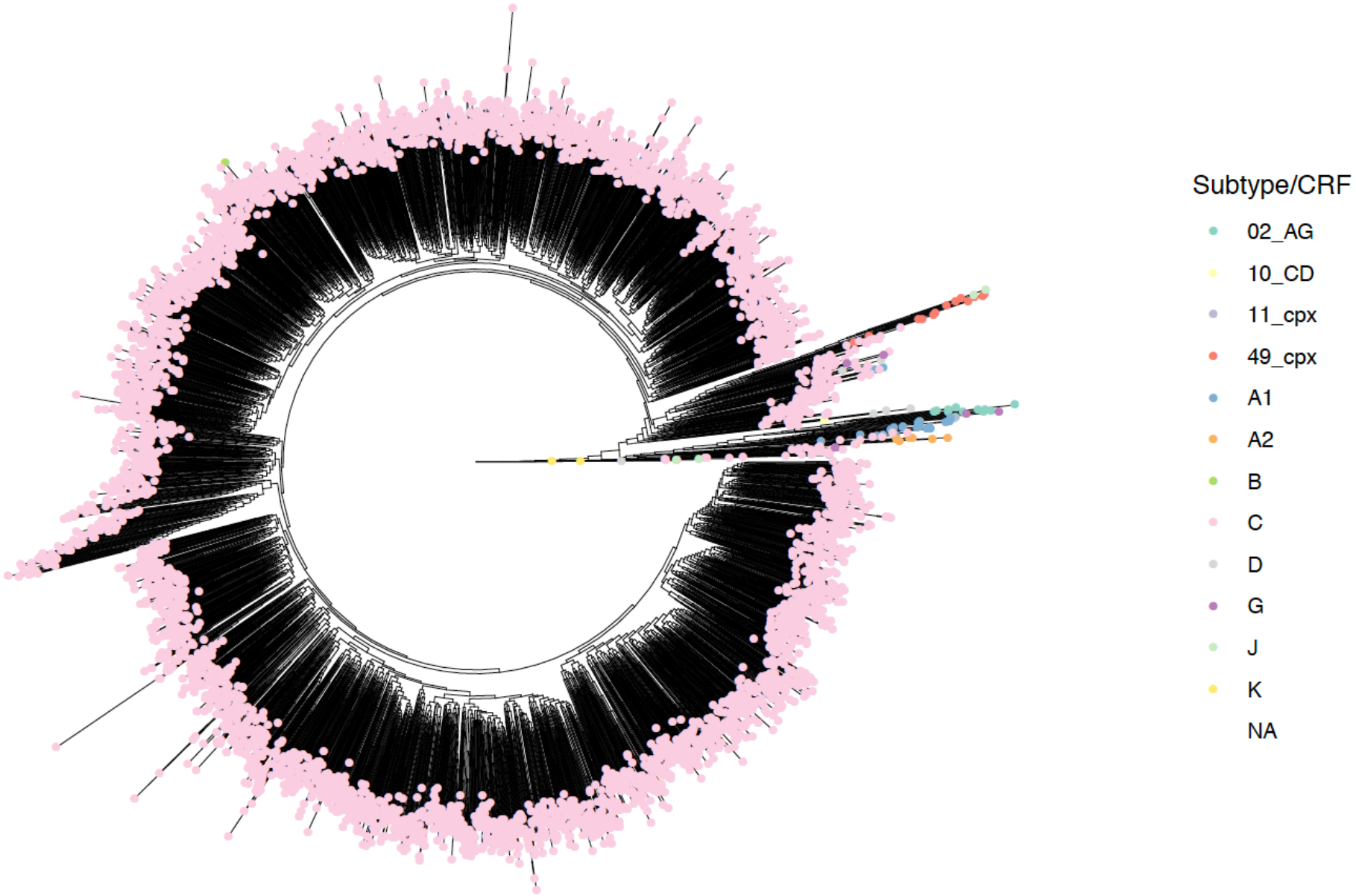


BEEHIVE (Europe – mostly clade B)

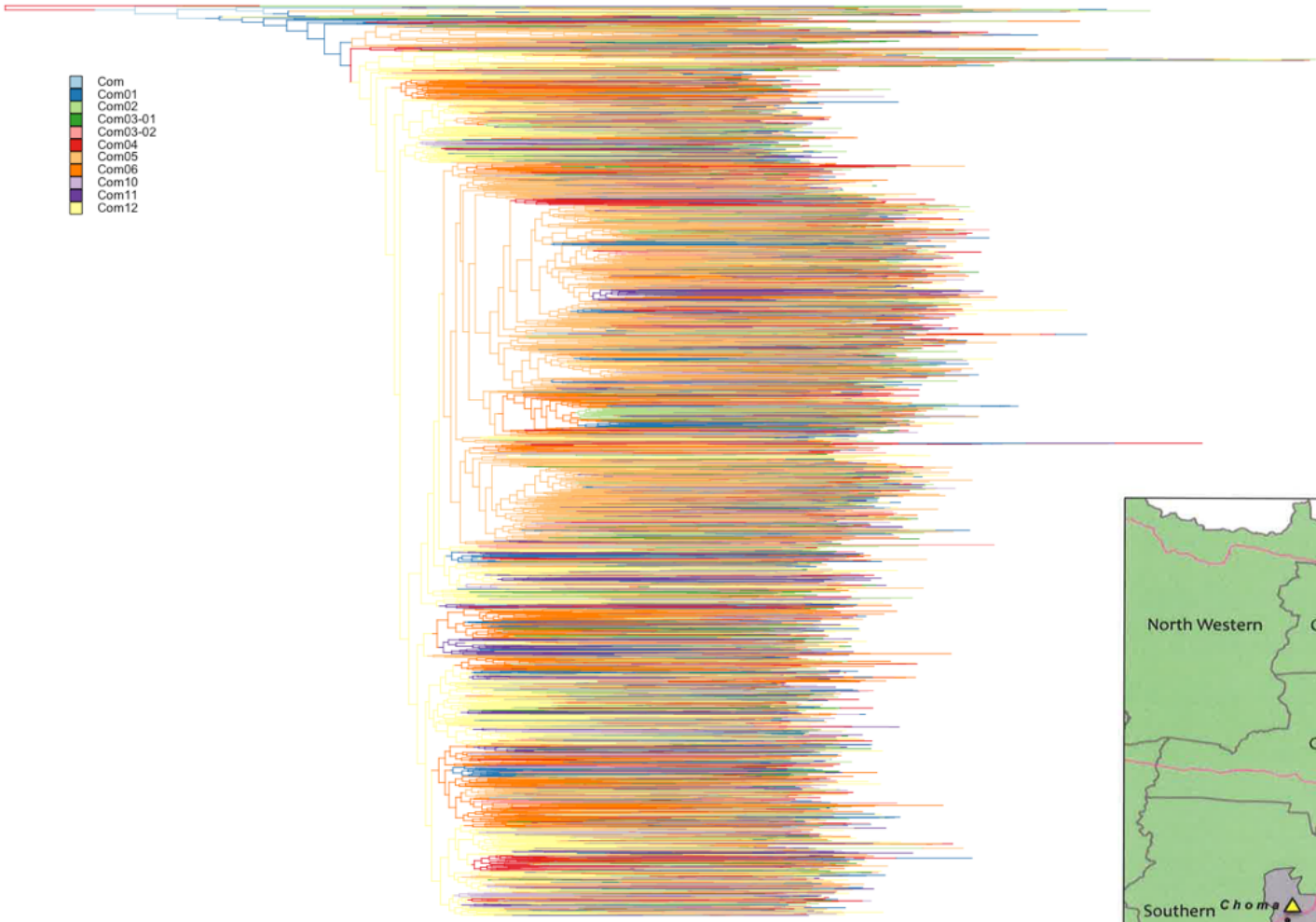
Beyond PopART…

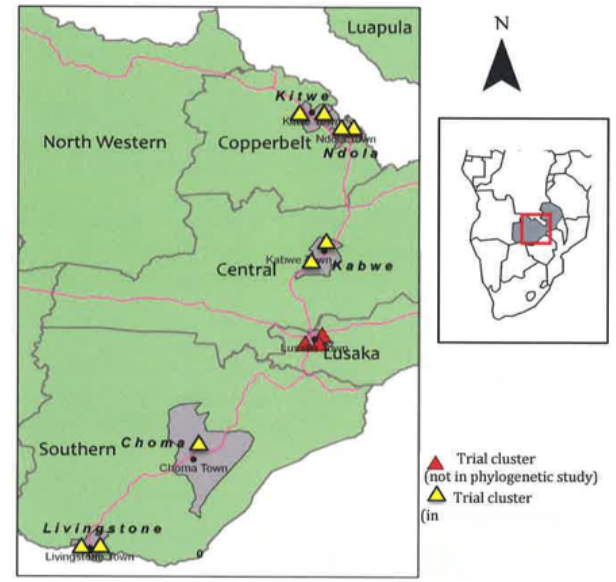
- Refining the drug resistance forecast.
  - Rates of transmission
  - Rates of reversion
  - Rates *de novo* selection of DR
  - Role of low frequency drug resistance mutations
  - Importance of mutational load
  - Role multi-drug combinations

- What does DR mean for TasP programmes?
- Is there a clinical role for drug resistance testing in Africa?
  - Do we need it? Can we predict drug resistance without a test?
- How do we use reistance information? Better empiric therapy vs. targeted therapy?

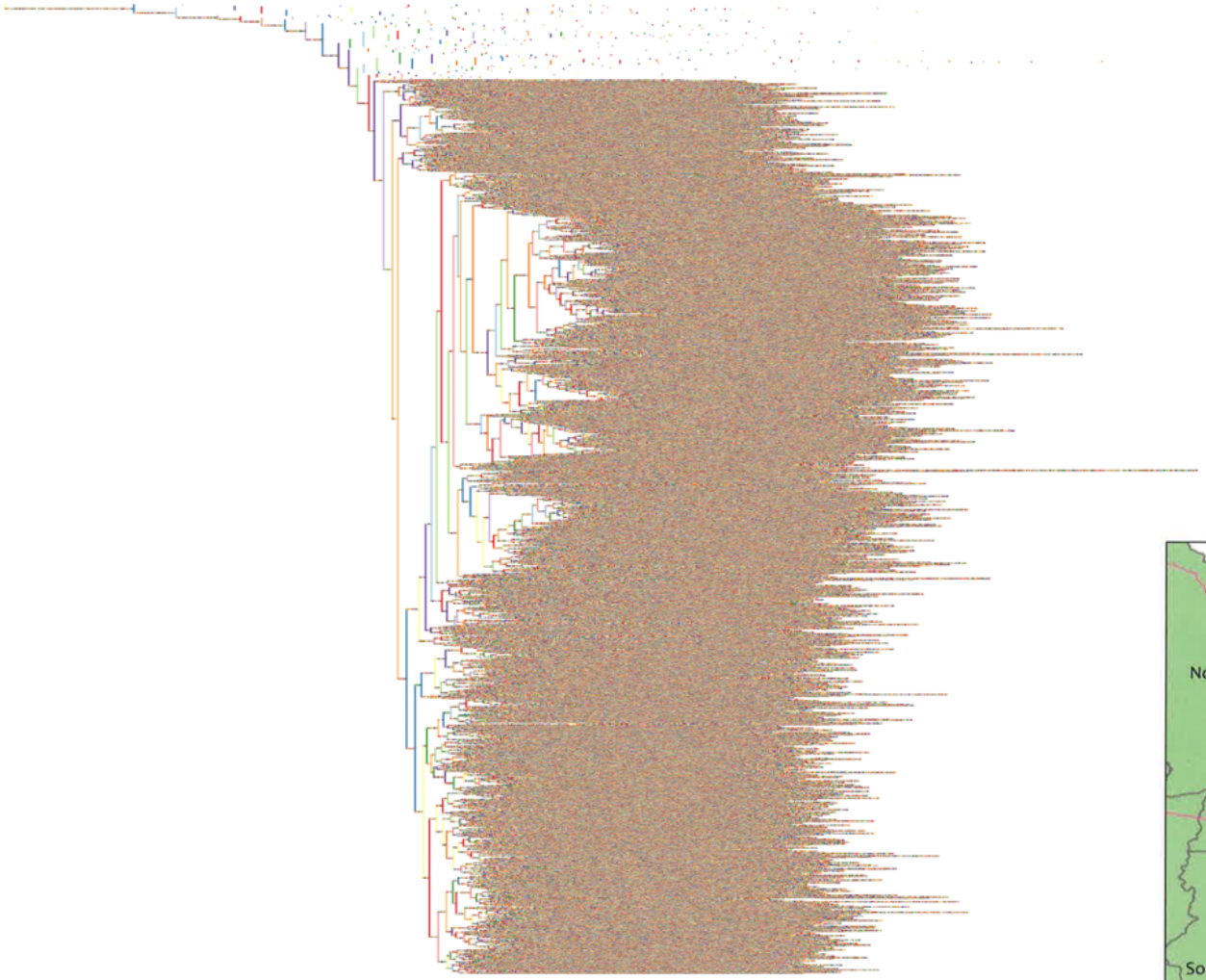

Subtype/CRF
02_AG
10_CD
11_cpx
49_cpx
A1
A2
B
C
D
G
J
K
NA

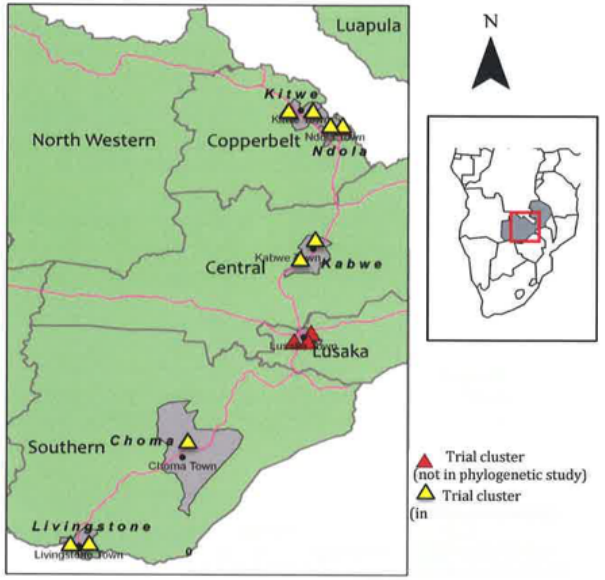
Are strains of HIV structured by community?



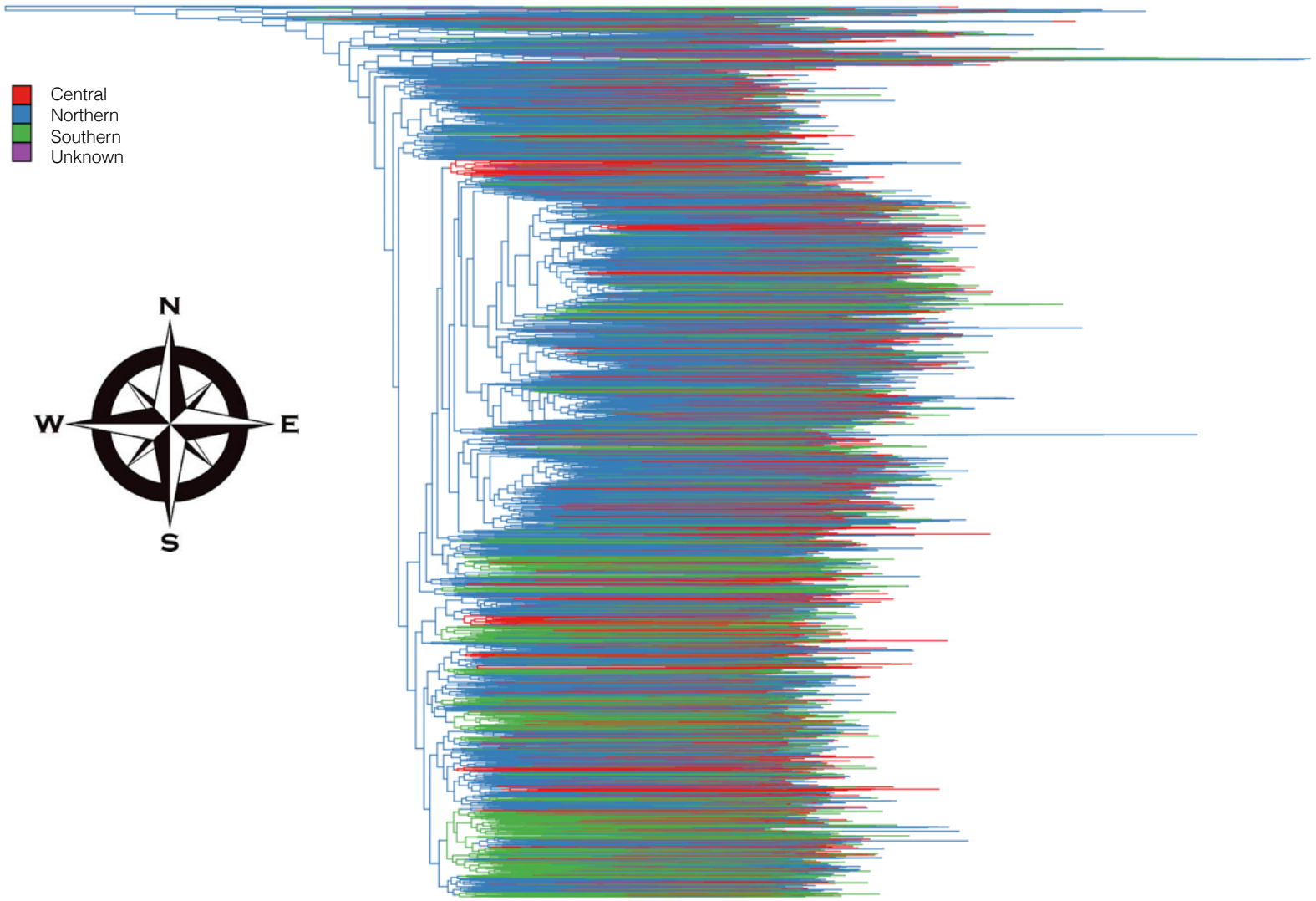


Are there strains of HIV structured by community?





Are strains of HIV structured by community?

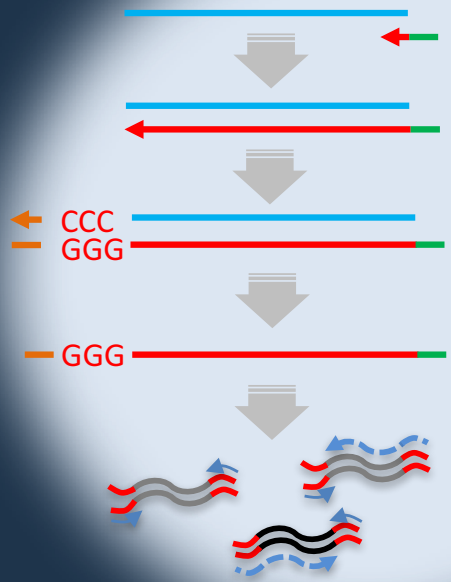


Nucleic Extraction

EasyMag (Biomerieux)

Library prep: SMARTer (TakaraBio)

Adapter attachment
without ligation
Automated
384 - well
format



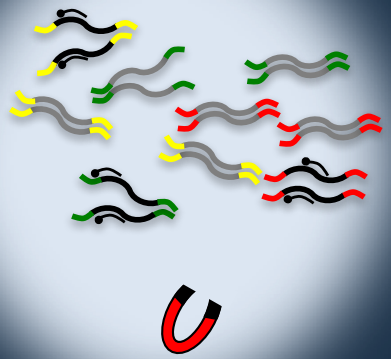
384 samples
per wk
£30 / sample

Minimal PCR

+ Size selection of larger
fragments

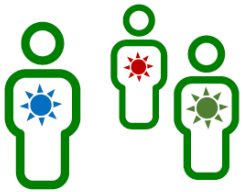
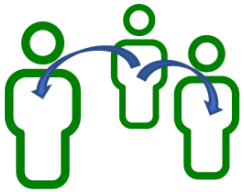
Sequencing

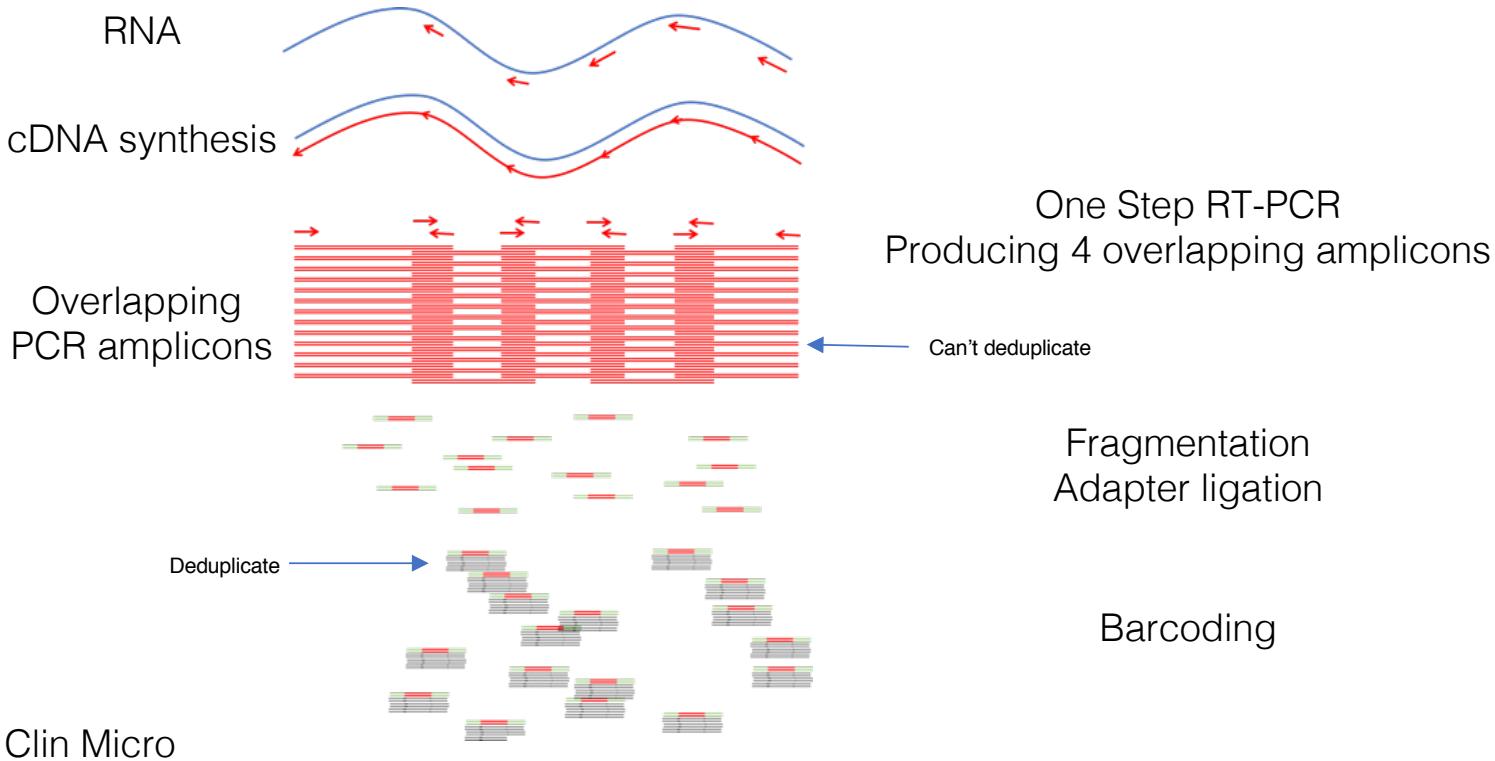
Illumina



Enrichment

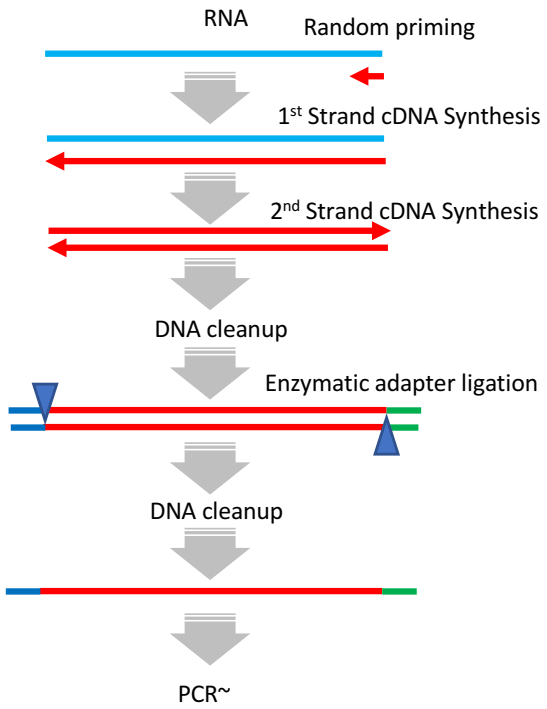
Custom oligonucleotides
designed to capture the
full HIV diversity

Measurement:	Viral load	Genotype	Transmission network	Drug resistance levels
				
Sequencing design:	Quantitative standards	Unbiased probe capture	Minimal PCR, Fragment size selection	Quantitative sequencing, Optimisation for low viral loads
Analysis design:	PCR duplicate removal	Accurate mapping, Consensus calling	Ancestral host-state reconstruction	Haplotype calling, HIVdb



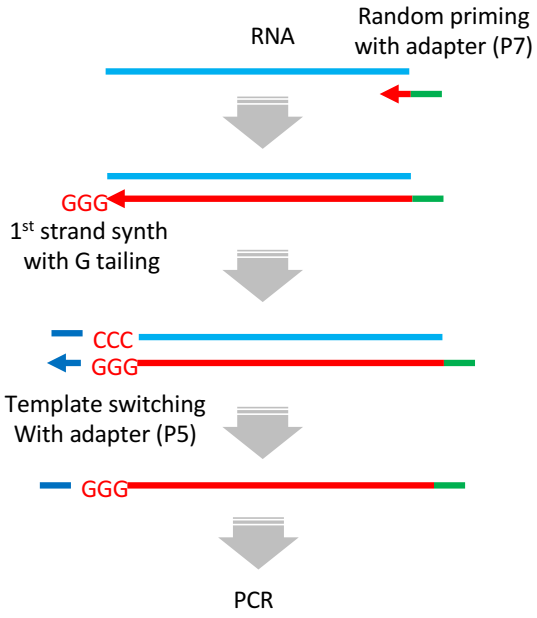
Astrid Gall et al. 2012. J Clin Micro

Ligation

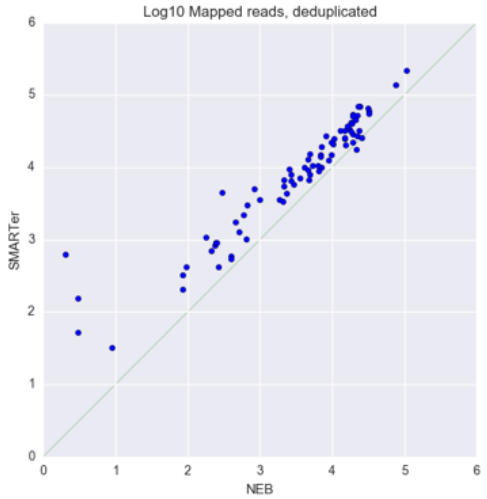


SMARTer (TakaraBio)

Switching mechanism at the 5' end of the RNA template

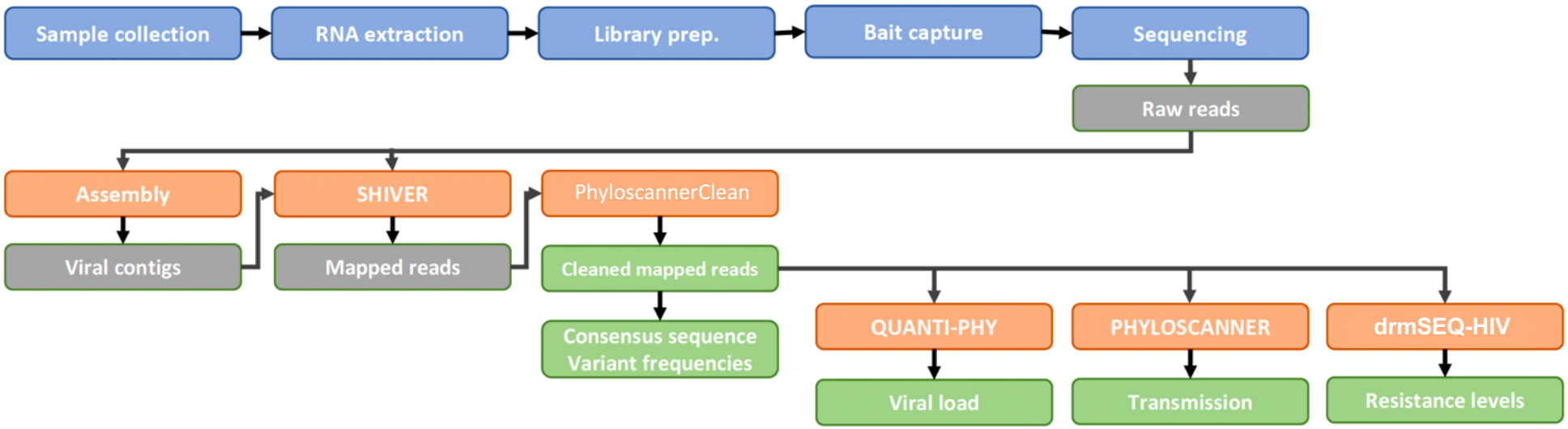


Template switching
L. Petalidis 2003. Nucleic Acids Research



SMARTer

- Faster (4 hours vs 2 days)
- Fewer steps pre-PCR
- Efficient additional of sequencing adapter
- Greater yields of unique sequences
- Fewer PCR cycles
- Greater library complexity
- Longer inserts


Sample collection
RNA extraction
Library prep.
Bait capture
Sequencing
Raw reads
Assembly
SHIVER
PhyloscannerClean
Viral contigs
Mapped reads
Cleaned mapped reads
Consensus sequence
Variant frequencies
QUANTI-PHY
PHYLOSCANNER
drmSEQ-HIV
Viral load
Transmission
Resistance levels

drmSEQ

Rapid HIV drug resistance typing for quantitative NGS data.

- Codon aware alignment of paired-end short read data.
- Matching clade reference mapping (clade C to clade C)
- Derives mutation coordinates relative to HXB2
- Scores resistance according to HIVdb (Stanford)
    - ...or a custom database.
- **Quantitative summary** of read data filtering on…
    - Absolute read count threshold
    - % prevalence threshold

- Reports **linked mutations**

- **FAST**
    - Laptop: 6000 reads in 2 mins.
    - HPC: All of PopART HC data (34 million reads) in 10 hours

- **FLEXIBLE (experimental)**
    - Whole genome characterization possible
    - (Co-receptor predictions, CTL escape, envelope glycosylation)

drmSEQ – Features

- Sequencing
  - Base-calling …SHIVER trims low quality bases and primers
  - Read length …veSEQ optimized to maximize long fragments
  - Contamination …*phyloscanner*
  - APOBEC hypermutations …drmSEQ
- Bioinformatics
  - Reference mapping …matched clade mapping, alignment score filtering
  - Pol vs. WGS …whole genome with drmSEQ (requires additional database)
  - Indel management ...handled by drmSEQ
- HIV variant calling… calls variants from reads NOT consensus sequences
- Database + Algorithm …HIVdb (Stanford) – Flexible, can be changed
- Minority variants…after cleaning, can apply minimum thresholds for coverage and % frequency
- Haplotyping…possible within limits of read length (250 – 500 nt)

Input 1:
Reads (fasta)

```
>Read1
ATCGACTACGCACACTACGACTAC
>Read2
AGCATCAGCATCAGCATCAGAACAAC
>Read(n)
```

Input 2:
Codon-aware multiple alignment

Reads	.....	TTC	---	CAA	CAA	GGG	GAG	GCC
Clade C	.....	TTC F	CCA P	CAA Q	--- -	GGG G	GAG E	GCC A
Clade B	.....	TTC F	CCA P	CAA Q	CAA Q	GGG G	AAG K	GCC A
Clade D	.....	TTC F	CCA P	GAA E	--- -	GGG G	GAG E	GCC A
Clade F	.....	TTC F	CCC P	CAA Q	--- -	GGG G	GAG E	GCC A
HXB2 Ref	.....	TTC F	CTA L	CAA Q	--- -	GGG G	AAG K	GCC A

Input 3:
HIVdb (Stanford)

		Drug				
Rule	**Gene**	**DOR**	**EFV**	**ETR**	**NVP**	**RPV**
...	...	...	...	...	...	...
A98G	**RT**	15	15	10	30	15
L100I	**RT**	15	60	30	60	60
L100V	**RT**	10	30	10	30	15
K101E	**RT**	15	15	15	30	45
K101H	**RT**	0	10	10	15	10
...	...	...	...	...	...	...

Step 1
Local alignment of reads to subtype matched consensus (DNA->Protein)

Step 2
Codon positioning (relative to HXB2

Step 3a
Read annotation with DRMs

Step 3b
Quantification of mutations and combinations of mutations

Step 4
Summarise overall resistance (apply quantification filters: eg 10%, minimum coverage of 2)

Step 5 (optional)
Integration with Phyloscanner

Output 1
Per Read Resistance Score

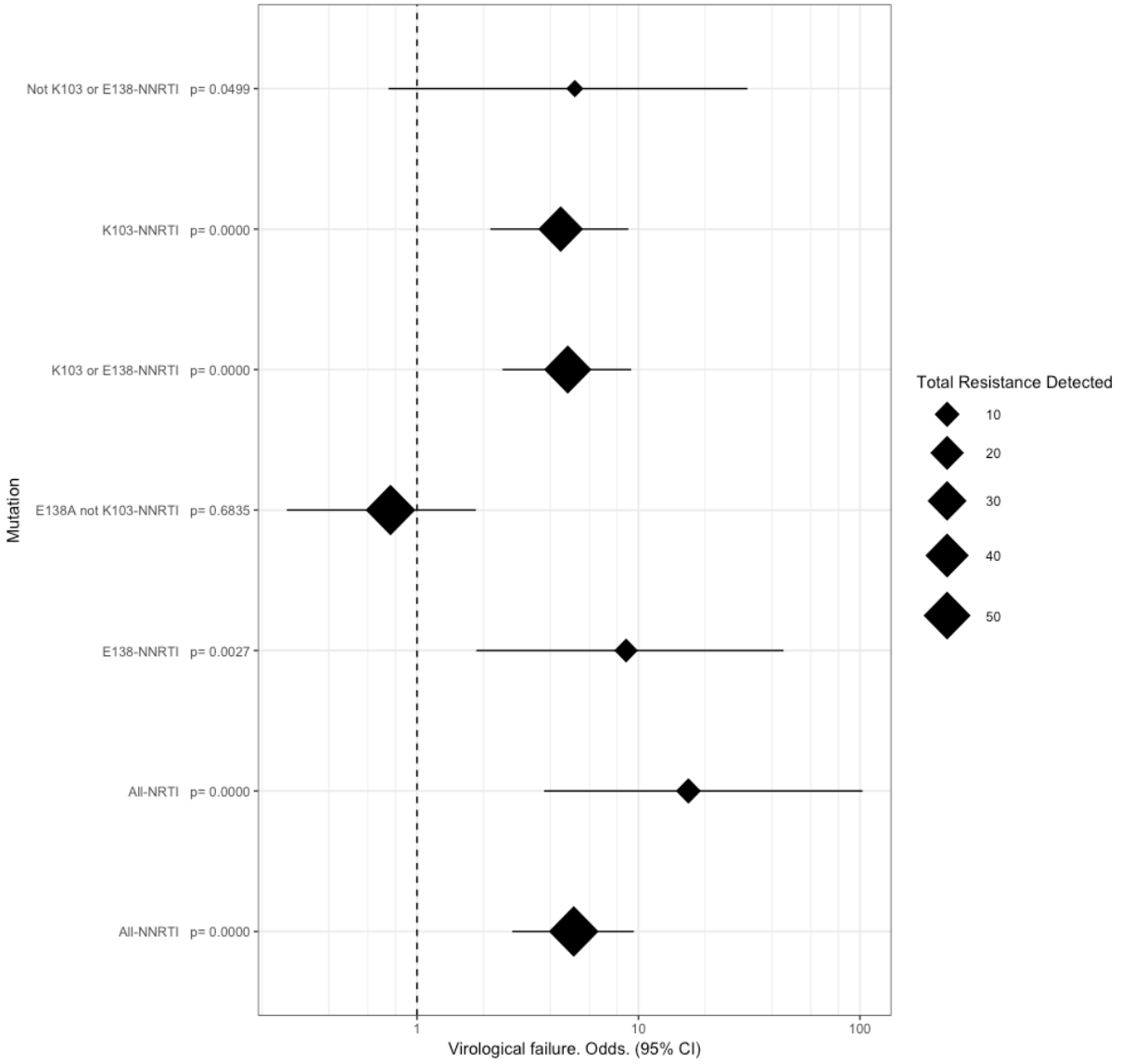
readID	DOR	EFV	ETR	NVP	RPV	ABC	AZT	D4T	DDI	FTC	Mutation
Read1	0	0	0	0	0	0	0	0	0	0	
Read2	2	4	2	4	3	4	2	4	4	0	rtL74I+rtV75S+rtL100V
Read3	2	4	2	4	3	4	2	4	4	0	rtL74I+rtV75S+rtL100V
Read4	2	4	2	4	3	4	2	4	4	0	rtL74I+rtV75S+rtL100V
Read5	3	4	3	4	3	0	0	0	0	0	rtG190A
Read6	3	4	3	4	3	0	0	0	0	0	rtG190A
Read7	3	4	3	4	3	0	0	0	0	0	rtG190A
Read8	0	0	0	0	0	4	0	4	4	4	rtK65R

Output 2
Resistance summary reads

readID	DOR	EFV	ETR	NVP	RPV	ABC	AZT	D4T	DDI	FTC
Summary	3	4	3	4	3	4	2	4	4	0

Mutations	rtG190	DRM	3	43 %	
	rtK65R	DRM	1	14 %	(Threshold not met)
	rtL74I	DRM	3	43 %	
	rtV75S	DRM	3	43 %	
	rtL100V	DRM	3	43 %	
	No Mutation	DRM	1	14 %	
Combinations	rtL74I+rtV75S+rtL100V		3	-	

Impact of DR on viral suppression



PopART HPTN 071

Low frequency drug-resistant mutations
Do they matter?

No association with treatment failure

O Peuchant *et al*., AIDS 2008;
M Balduin *et al*., JCV 2009;
MR Jakobsen *et al*., (CID 2010;
KJ Metzner *et al*., JCV 2011;
JD Stekler *et al*., PLoS One 2011;
P Messiaen *et al*., Virology 2012;
V Bansode *et al*., BMC Infect Dis 2013;
VF Boltz *et al*., JID 2014;
S Mohamed *et al*., JMV 2014;
KJ Metzner *et al*., AIDS 2014;
U Neogi *et al*., AIDS 2014;
F Nicot *et al*., JCV 2015;
A Zoufaly *et al*., JAC 2015;
M Casadella *et al*., AIDS 2015;
DP Porter *et al*., Viruses 2015;
V Van Eygen *et al*., JMV 2016;
ML Mzingwane *et al*., Virol J 2016;
DS Clutter *et al*., JID 2017;
O Epaulard *et al*., JCV 2017;
S Raymond *et al*., CID 2018;

Association with treatment failure

JA Johnson *et al*., PLoS Med 2008;
KJ Metzner *et al*., CID 2009;
BB Simen *et al*.,JID 2009;
AM Geretti *et al*., JAIDS 2009;
R Paredes *et al*., JID 2010;
DD Goodman *et al*., AIDS 2011;
JZ Li *et al*., JAMA 2011;
M Pingen *et al*., HIV Med 2012;
A Cozzi-Lepri *et al*., JAC 2015;
S Avila-Rios *et al*., Lancet HIV 2016;
SC Inzaule *et al*., JAC 2018

Credit to Karin Metzner for literature search

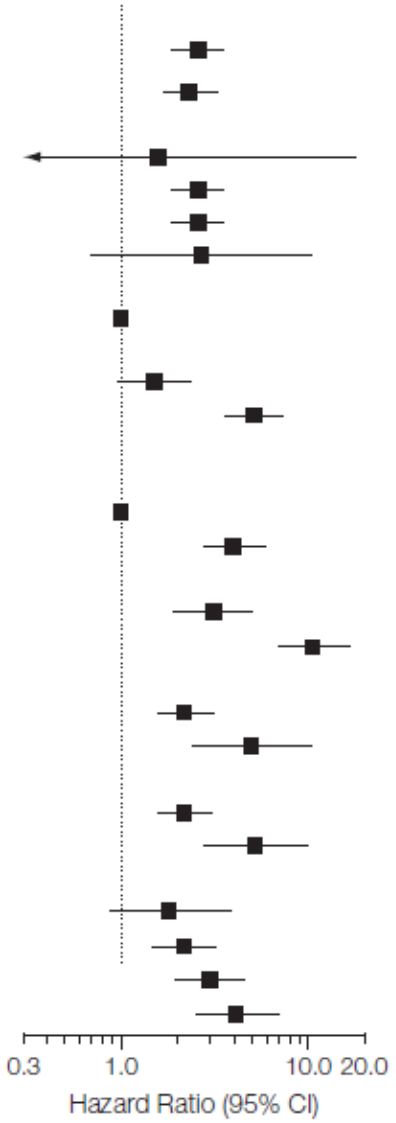
Mutational load may be more important (few studies)

1% K103N, VL 1000 = 10 mutants per ml

1% K103N, VL 10^5 = 1000 mutants per ml

Mutational load + Increased potency from mutation combinations (no studies?)

Group vs No Minority Variant[c]	Virologic Failure, No.[a]		Total No.		HR (95% CI)[b]
	Minority Variant	No Minority Variant	Minority Variant	No Minority Variant	
Any minority variant	111	208	187	798	2.6 (1.9-3.5)
Any minority variant (multivariate)[d]	98	138	152	617	2.3 (1.7-3.3)
Minority variant type					
NRTI-resistant	2	104	9	202	1.6 (0.1-17.7)
NNRTI-resistant	109	210	178	807	2.6 (1.9-3.5)
Efavirenz-resistant	105	188	165	723	2.6 (1.9-3.5)
Nevirapine-resistant	4	22	13	84	2.7 (0.7-10.3)
Minority variant and adherence					
No minority variant and any adherence					1 [Reference]
Any minority variant					
Adherence ≥95%	35	138	73	617	1.5 (0.98-2.3)
Adherence <95%	63	138	79	617	5.1 (3.6-7.2)
No minority variant					
Adherence ≥95%					1 [Reference]
Adherence <95%		43	231	386	4.0 (2.8-5.8)
Any minority variant					
Adherence ≥95%	35	43	73	386	3.1 (1.9-5.0)
Adherence <95%	63	43	79	386	10.6 (6.9-16.4)
Minority variant, %					
<1	91	209	154	781	2.2 (1.6-3.1)
≥1	18	209	30	781	5.0 (2.4-10.3)
<0.5	86	107	143	654	2.2 (1.6-3.0)
≥0.5	14	107	32	654	5.2 (2.8-9.8)
Minority variant copies, No.					
1-9	8	148	15	720	1.8 (0.9-3.8)
10-99	41	148	71	720	2.2 (1.5-3.2)
100-999	35	148	55	720	3.0 (2.0-4.5)
≥1000	20	148	38	720	4.1 (2.5-6.8)



JZ LI *et al*., JAMA 2011

J Antimicrob Chemother 2015; **70**: 930–940
doi:10.1093/jac/dku426 Advance Access publication 21 October 2014

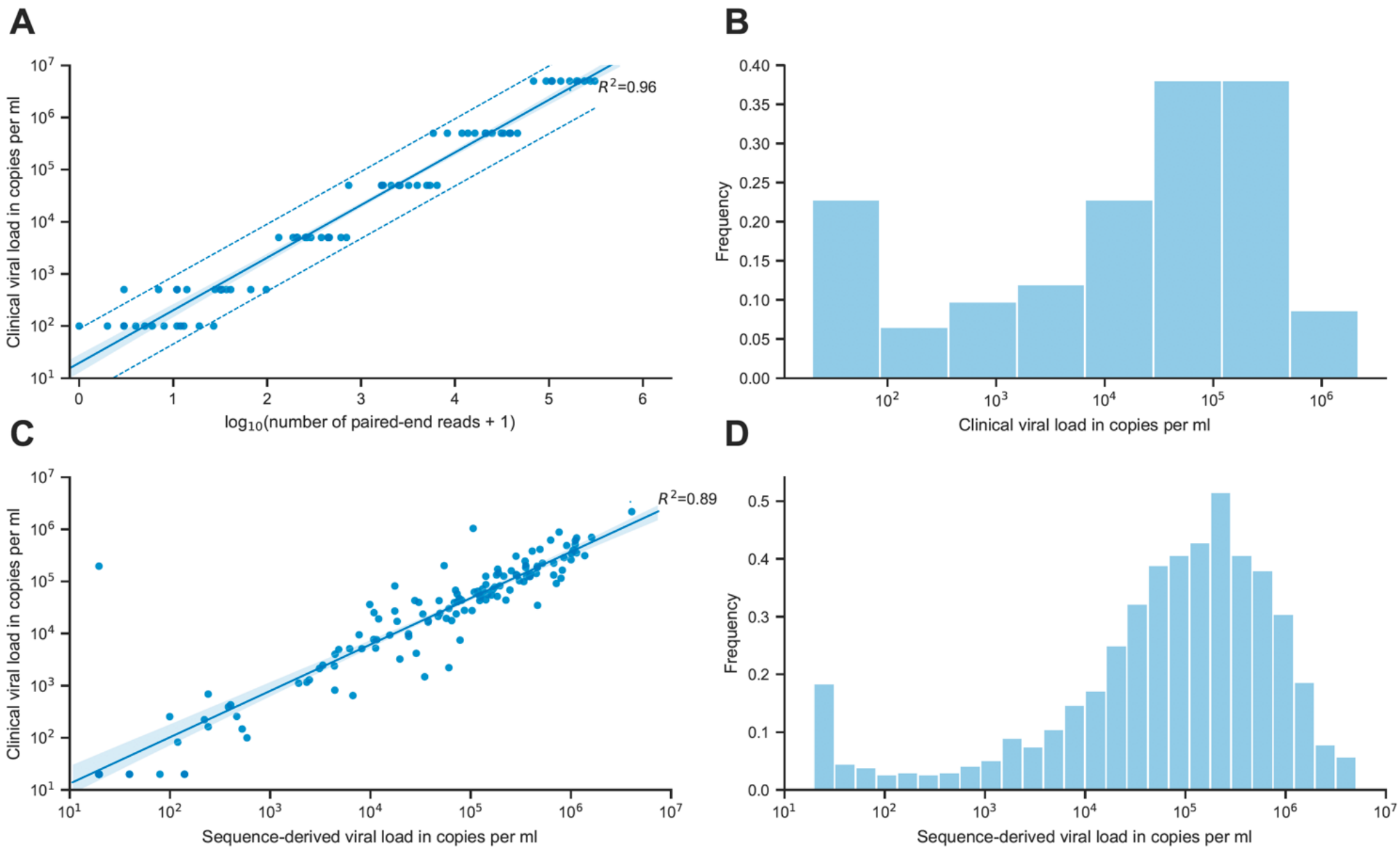
Journal of Antimicrobial Chemotherapy

Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: a multicohort European case–control study using centralized ultrasensitive 454 pyrosequencing

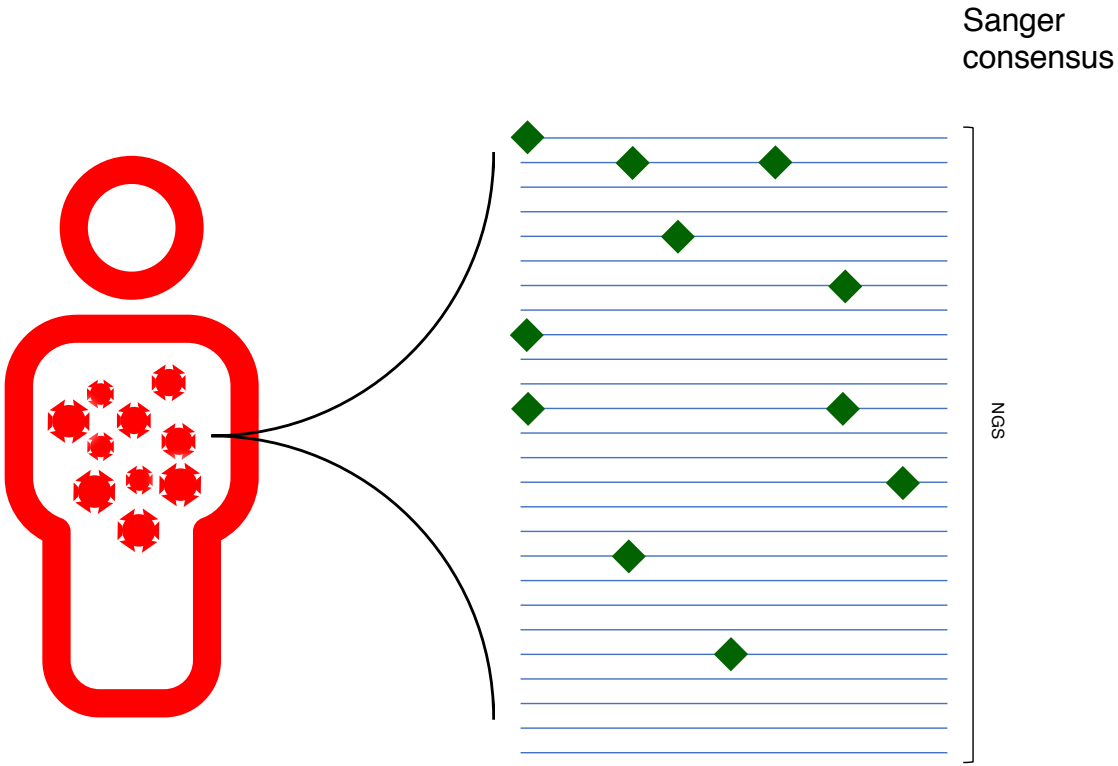
Alessandro Cozzi-Lepri[1]†, Marc Noguera-Julian[2]†, Francesca Di Giallonardo[3], Rob Schuurman[4], Martin Däumer[5], Sue Aitken[4], Francesca Ceccherini-Silberstein[6], Antonella D'Arminio Monforte[7], Anna Maria Geretti[8], Clare L. Booth[9], Rolf Kaiser[10], Claudia Michalik[11,12], Klaus Jansen[11], Bernard Masquelier[13], Pantxika Bellecave[13], Roger D. Kouyos[3], Erika Castro[14], Hansjakob Furrer[15], Anna Schultze[1], Huldrych F. Günthard[3], Francoise Brun-Vezinet[16], Roger Paredes[2]† and Karin J. Metzner[3]*† on behalf of the CHAIN Minority HIV-1 Variants Working Group‡

Table 2. Factors associated with virological failure

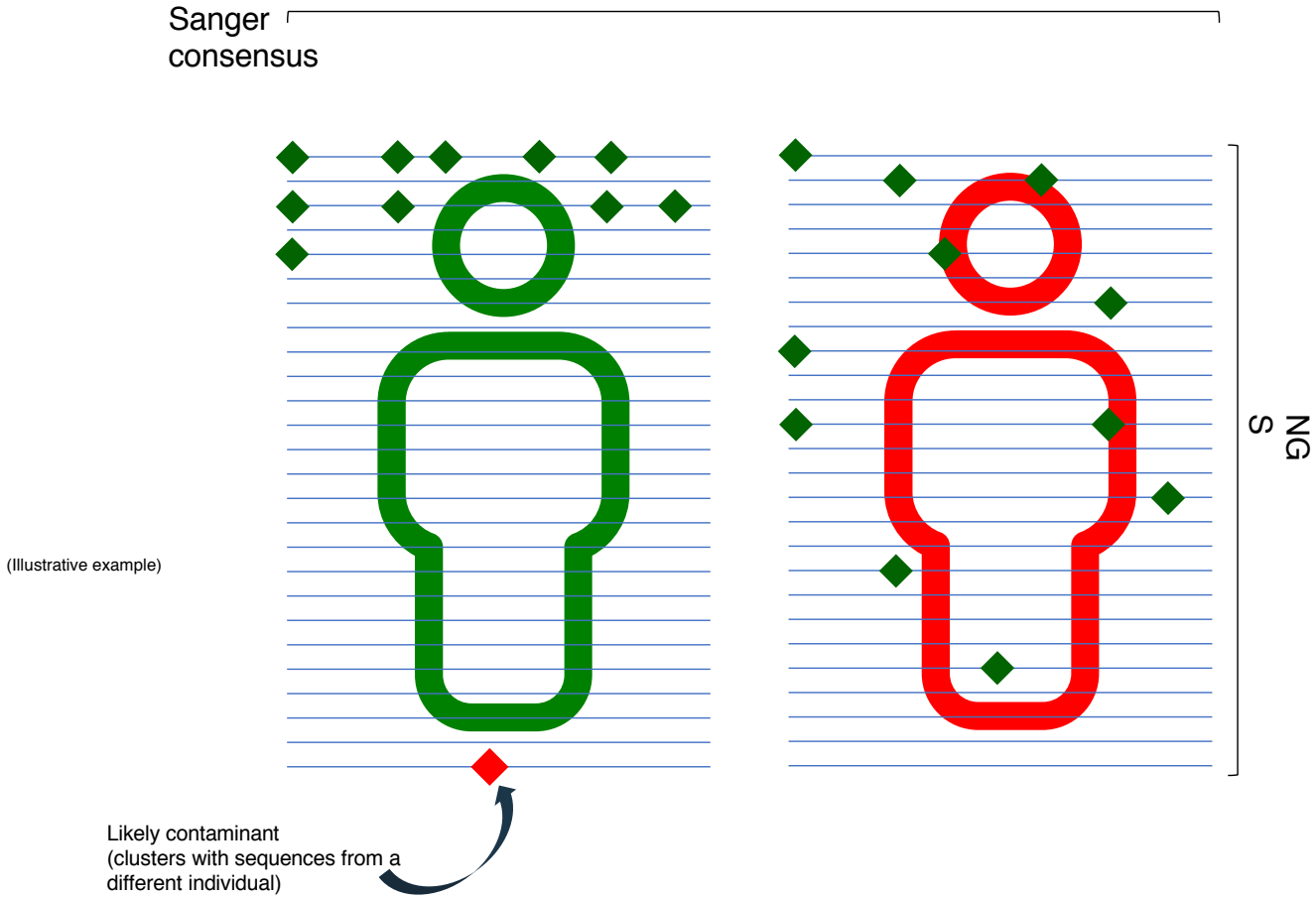
			ORs of viral rebound >200 RNA copies/mL plasma			
	Cases, *N*=76	Controls, *N*=184	unadjusted OR (95% CI)	*P*	adjusted[a] OR (95% CI)	*P*


A
Clinical viral load in copies per ml
log10(number of paired-end reads + 1)
R2=0.96
B
Frequency
Clinical viral load in copies per ml
C
Clinical viral load in copies per ml
Sequence-derived viral load in copies per ml
R2=0.89
D
Frequency
Sequence-derived viral load in copies per ml

Interpreting Drug Resistance from NGS data



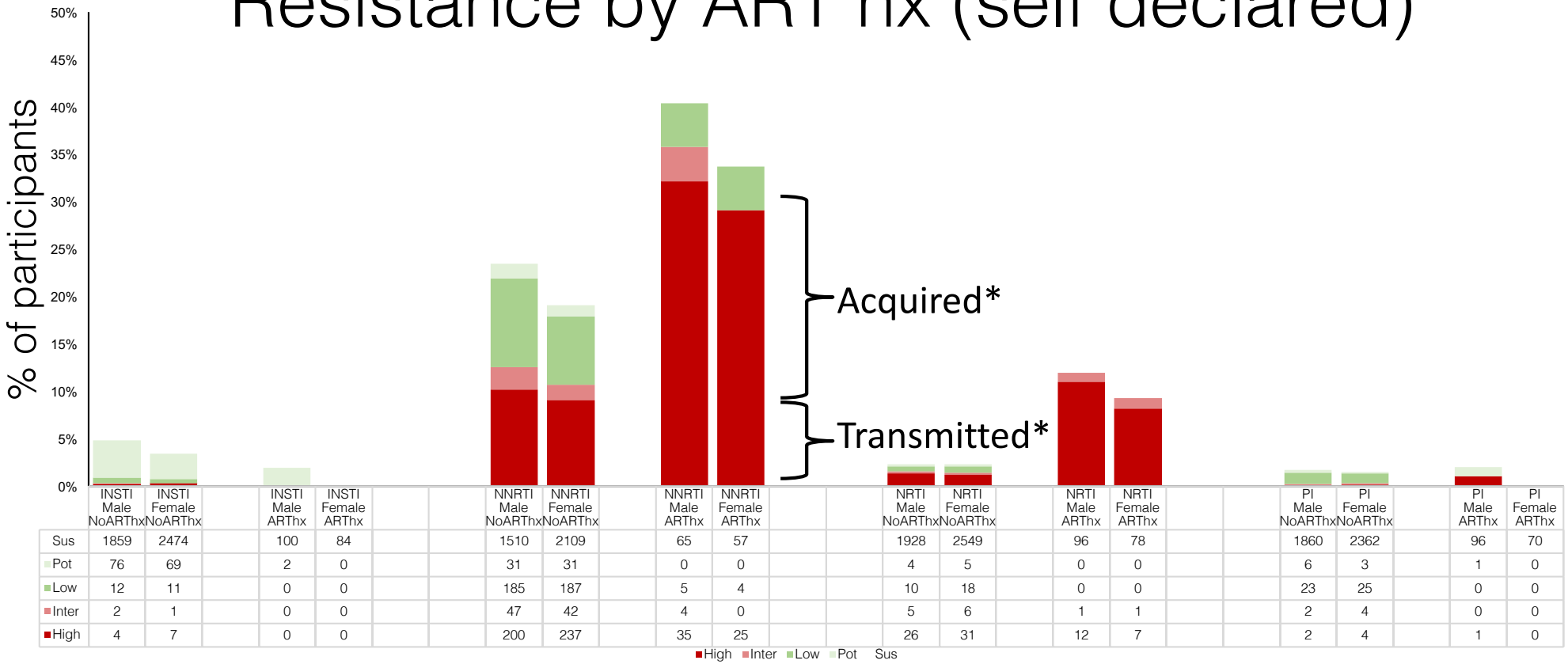
Interpreting Drug Resistance from NGS data



Transmission of Drug resistance (Work in Progress)

HIV drug resistance can be acquired and transmitted

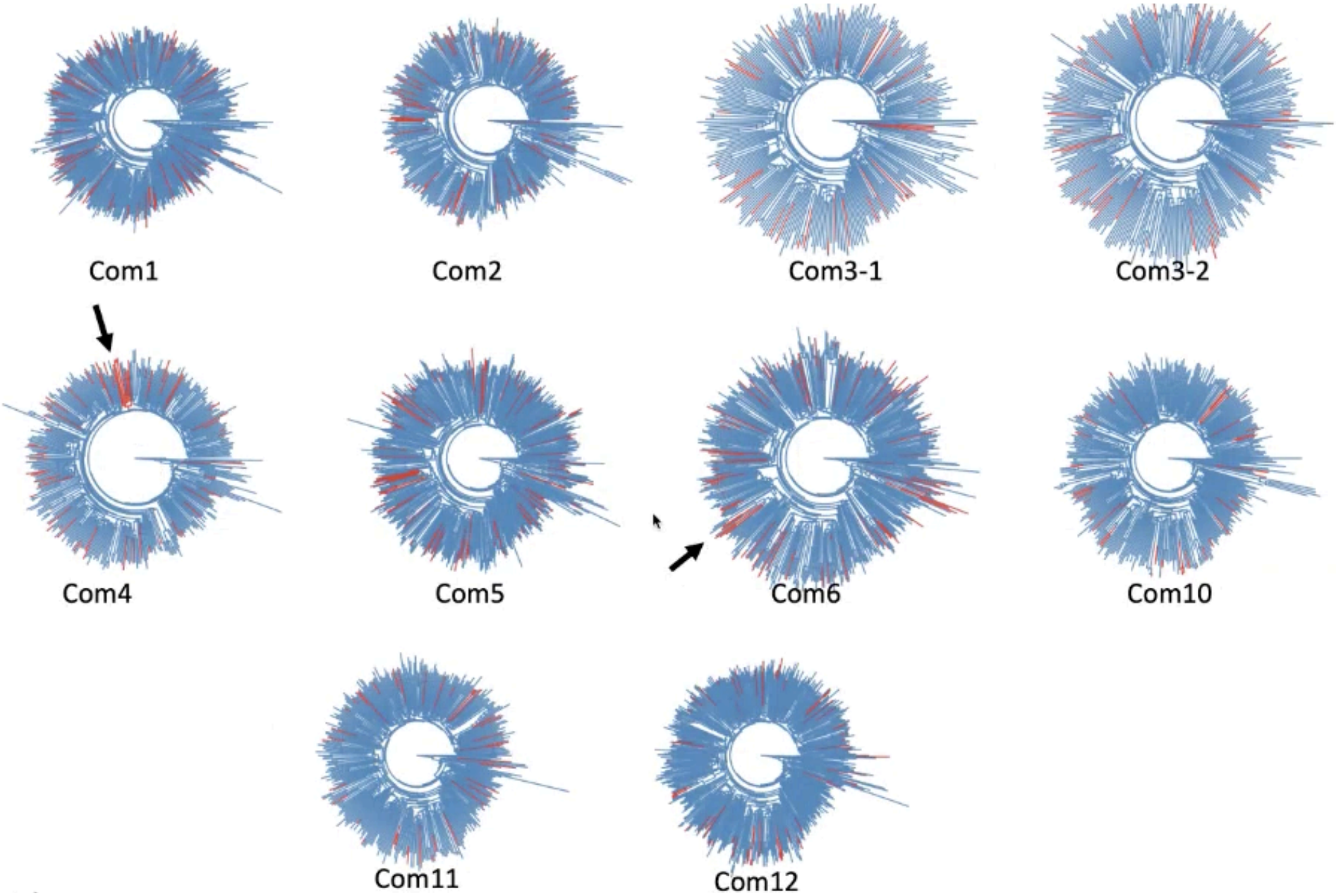
Resistance by ART hx (self declared)



	INSTI Male NoARThx	INSTI Female NoARThx	INSTI Male ARThx	INSTI Female ARThx	NNRTI Male NoARThx	NNRTI Female NoARThx	NNRTI Male ARThx	NNRTI Female ARThx	NRTI Male NoARThx	NRTI Female NoARThx	NRTI Male ARThx	NRTI Female ARThx	PI Male NoARThx	PI Female NoARThx	PI Male ARThx	PI Female ARThx
Sus	1859	2474	100	84	1510	2109	65	57	1928	2549	96	78	1860	2362	96	70
Pot	76	69	2	0	31	31	0	0	4	5	0	0	6	3	1	0
Low	12	11	0	0	185	187	5	4	10	18	0	0	23	25	0	0
Inter	2	1	0	0	47	42	4	0	5	6	1	1	2	4	0	0
High	4	7	0	0	200	237	35	25	26	31	12	7	2	4	1	0

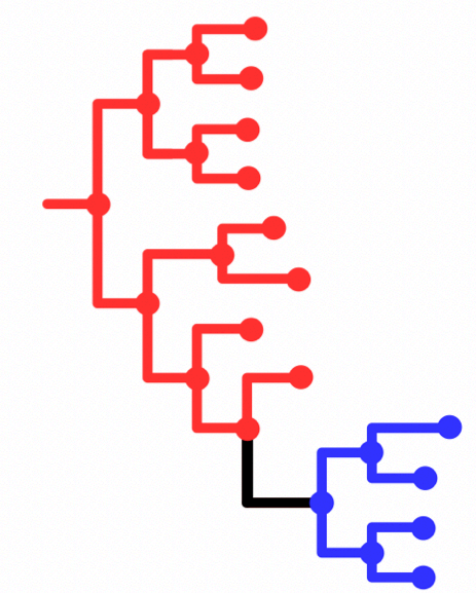
*Self reporting may not be accurate

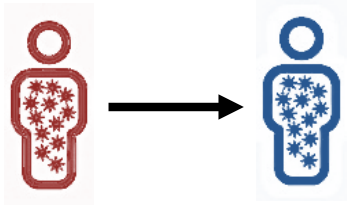
Ancestral State Reconstruction (Discrete state = NNRTI Resistance)



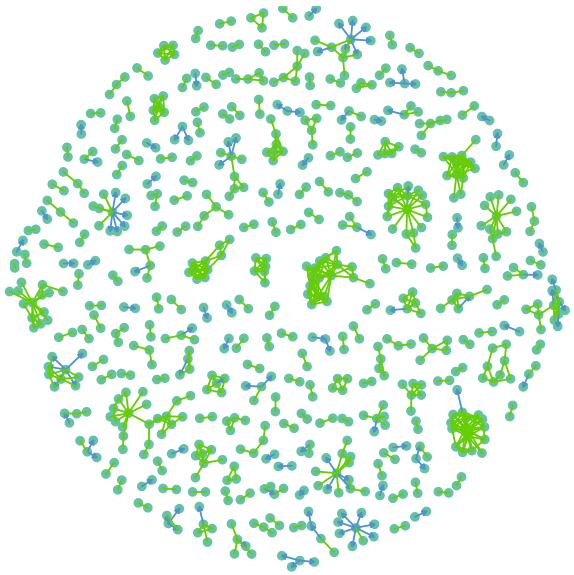
Phylogenetic networks

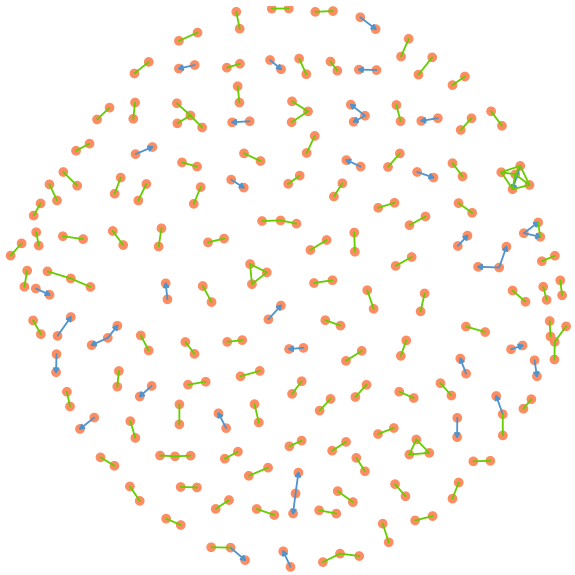
Identify linked partnerships….

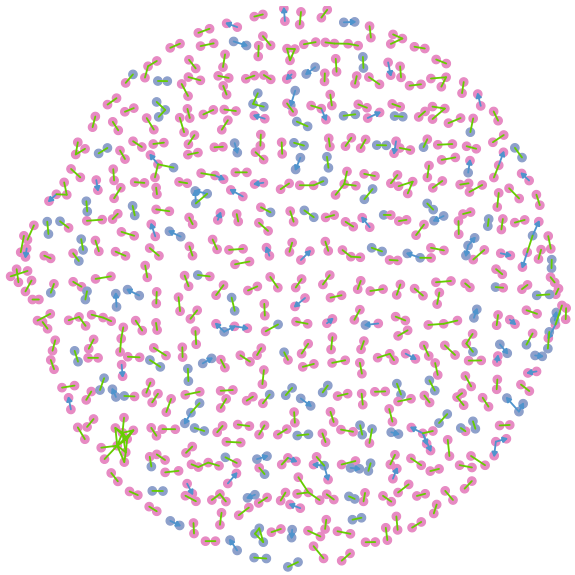




…for all individuals….







Summary

The Oxford HIV Sequencing Pipeline:

High resolution molecular epidemiology
Viral load
Drug resistance

1 technician:
384 samples per week
Consumable cost: $40-$45/sample

What are the potential benefits of ROUTINE implementation of quantitative HIV sequencing in LMIC:

Drug resistance?
Resource provisioning?
Reaching the undiagnosed and untreated?
Economics?
The 90:90:90 goals and “the missing 27%”

Challenges:

Setting-up in LMIC: Portable (low-throughput, fast) vs. Centralised (Highthroughput, slow)
Who? Indiscriminate vs targeted sampling. Communities vs. individuals
Efficient and ethical and infrastructures for data handling

Acknowledgement

Christophe Fraser
Tanya Golubchik
Matthew Hall
Chris Wymant
Lucie Abeler-Dorner
Mariateresa de Cesare
Rafael Sauter
Will Probert

The Beehive investigators
The PopART investigators (HPTN 071)
The PopART Phylogenetics investigators (HPTN 071-2)
PANGEA Consortium members

Africa Health Research Institute
Deenan Pillay
Steven Kemp
Ravindra Gupta
Anne Derache

Zambart Project
Barry Kosloff
Mohammed Limbada
Ab Schaap
Helen Ayles

Wellcome Centre for Human Genetics
Rory Bowden
George MacIntyre-Cockett

Peter Medawar Building for Pathogen Research
Ellie Barnes
Paul Klenerman
Anthony Brown
Azim Ansari

The HPTN 071 Study Team, led by:

Dr. Richard Hayes
Dr. Sarah Fidler
Dr. Helen Ayles
Dr. Nulda Beyers
Dr. Peter Bock

Government Agencies:

PEPFAR Implementing Partners:

Summary

The Oxford HIV Sequencing Pipeline:

High resolution molecular epidemiology
Viral load
Drug resistance

} 1 technician:
384 samples per week
Consumable cost: $40-$45/sample

What are the potential benefits of ROUTINE implementation of quantitative HIV sequencing in LMIC:

Drug resistance?
Resource provisioning?
Reaching the undiagnosed and untreated?
Economics?
The 90:90:90 goals and “the missing 27%”

Challenges:

Setting-up in LMIC: Portable (low-throughput, fast) vs. Centralised (Highthroughput, slow)
Who? Indiscriminate vs targeted sampling. Communities vs. individuals
Efficient and ethical infrastructures for data handling

Acknowledgement

Christophe Fraser
Tanya Golubchik
Matthew Hall
Chris Wymant
Lucie Abeler-Dorner
Mariateresa de Cesare
Will Probert
Rafael Sauter

The Beehive investigators
The PopART investigators (HPTN 071)
The PopART Phylogenetics investigators (HPTN 071-2)
PANGEA Consortium members

Africa Health Research Institute
Deenan Pillay
Steven Kemp
Ravindra Gupta
Anne Derache

Zambart Project
Barry Kosloff
Mohammed Limbada
Ab Schaap
Helen Ayles

Wellcome Centre for Human Genetics
Rory Bowden
George MacIntyre-Cockett

Peter Medawar Building for Pathogen Research
Ellie Barnes
Paul Klenerman
Anthony Brown
Azim Ansari

The HPTN 071 Study Team, led by:

Dr. Richard Hayes
Dr. Sarah Fidler
Dr. Helen Ayles
Dr. Nulda Beyers
Dr. Peter Bock

Government Agencies:

PEPFAR Implementing Partners: