Optimal strategies for familial searching

Maarten Kruijver[1], Ronald Meester[1], Klaas Slooten[1,2]
1) VU University, Amsterdam
2) Netherlands Forensic Institute, The Hague

30 May, 2014

Overview

Familial searching is the process of looking for close relatives of an offender in a DNA database

- The familial searching process
- Generating the candidate list
- Simulation studies
- Conclusions
- Questions

The familial searching process

Case profile
- Case profile does not match any database profile

Database LRs
- Compute LR for kinship (full siblings or parent/offspring) with all database members
- Today, only full siblings

Candidate list
- Top *k*: select the top *k* LRs, for some *k*
- Fixed LR: select DB members with LR > *t*, for some threshold *t*
- Profile-dependent: select DB members with LR > t_α, s.t. P(LR > t_α) = α for true sibs

Eliminate false positives
- Lineage markers (Y-STR typing) can eliminate many false positives
- Typing additional loci can also eliminate false positives

Final candidate list
- The final candidate list is investigated tactically by the authorities

Generating the candidate list

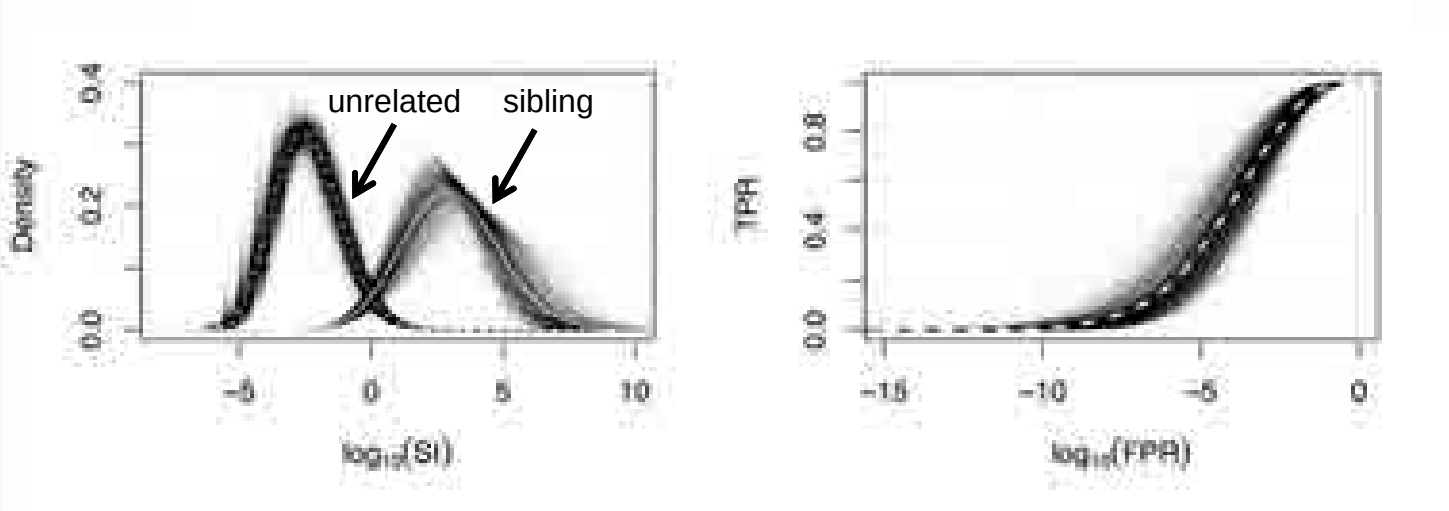
- Trade-off between workload (eliminating false positives) and probability of detection (PoD)

Workload and PoD per case are driven by:

- Case profile (rare alleles or common alleles?);
- Search strategy and tuning parameters (k, t, α);
- Database size (N).

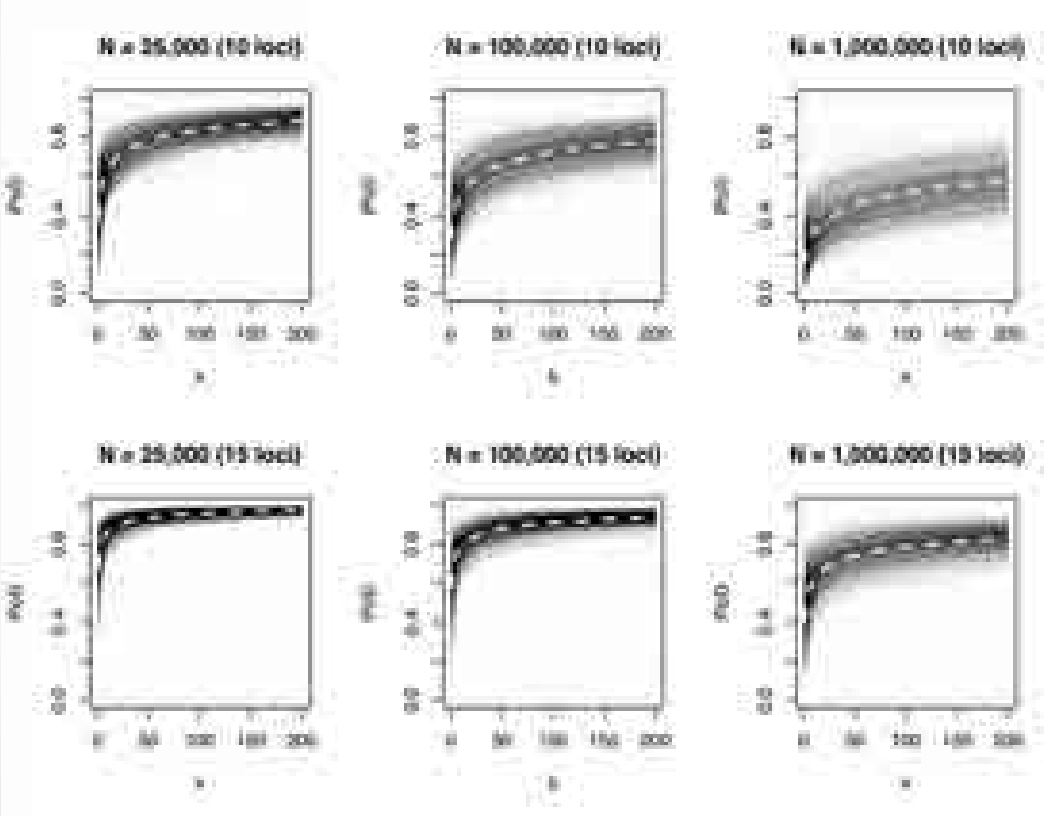
Case profile

- For 1,000 simulated SGMplus profiles, the SI-distribution is obtained with respect to a true full sibling and an unrelated person



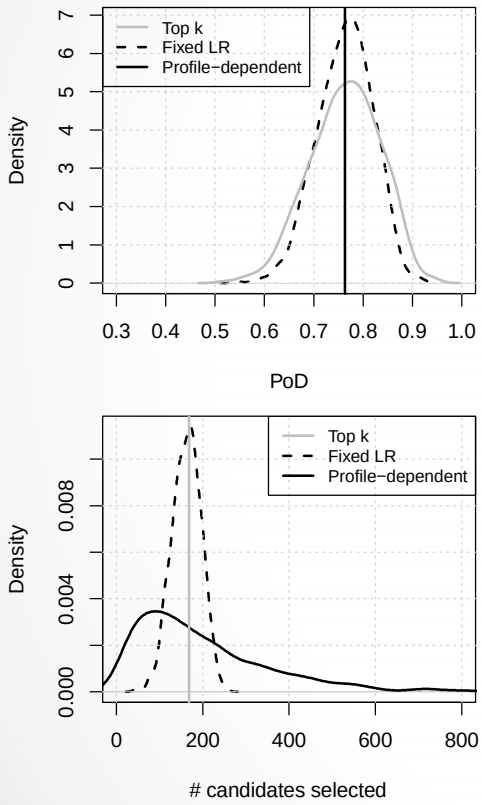
- Distribution differs a lot between case profiles. Large effect on TPR and FPR.
- Variation is caused by rarity of the profile. Profiles with rare alleles are especially amenable to familial searching.
- Effect on search strategies is investigated next

Simulation: top *k* strategy (1,000 profiles)



- Large variation in PoD for fixed *k*
- Increasing *k* gives quickly diminishing returns in terms of PoD
- Using 15 instead of 10 loci makes it possible to increase DB size ~10 times, while retaining the PoD

The fixed-LR strategy is optimal



- Fixed-LR strategy is most efficient in the long-run: lowest FPR for given TPR
- How many more false positives with top k or profile-dependent threshold?
- Take top 168 strategy as point of reference in fixed DB (N=100,000). Tuning parameters (t, α) such that the average PoD coincides with top 168
- Fixing workload is cheap; fixing PoD is not

Conclusions

- Workload and PoD per case depend on case profile
- A fixed-LR threshold is optimal in the long run
- Fixing workload is cheap; fixing PoD not
- Results are easily reproduced using the R-package DNAprofiles, freely available from CRAN

Questions?

Netherlands Forensic Institute
Ministry of Security and Justice

Familial Searching on Complex Mixtures

Athens, ESWG of the ISFG

30 May 2014

Klaas Slooten
Netherlands Forensic Institute /
VU University Amsterdam

28 May 2014

Familial Searching

When we do not find a trace donor in a database, we look for his relatives instead.

For a single source trace, this means that we evaluate the LR for hypotheses

- Hp: database member is related to trace donor
- Hd: database member is unrelated to trace donor

For a given type of relatedness. In practice, only parent-child and sibling relationships are considered. LR's are called PI and SI respectively.

Legislation

Familial Searching is carried out regularly in the following jurisdictions:

- UK
- California
- New Zealand
- The Netherlands

Dutch database: almost 200,000 profiles.

Oldest half of them typed with SGMPlus (10 loci) and the remainder typed with NGM (=SGMPlus + 5 other loci)

Feasibility of a FS

Can be expressed in several ways:

1. Probability for a relative to rank in the top-n for various n
2. Probability for a relative to exceed a LR-threshold (True positive rate)
3. Probability for an unrelated individual to exceed a LR-threshold (False Positive Rate)

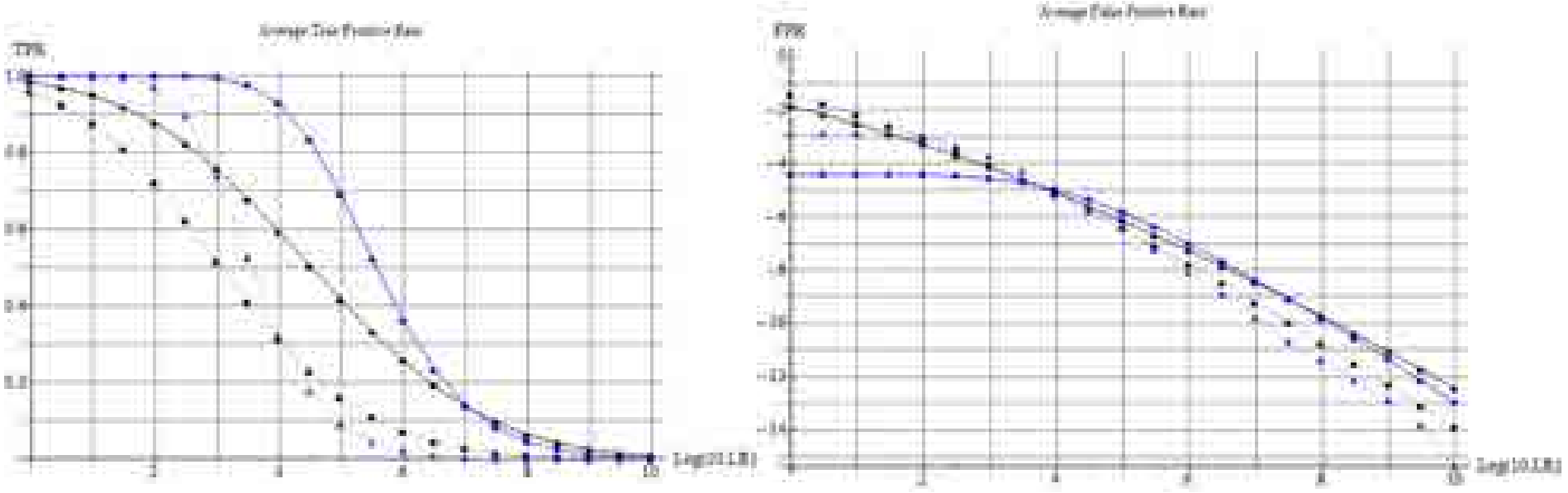
Ranking is illustrative but depends on the database size.

TPR and FPR can be used to choose a strategy, for any database.

Moreover, it is known that selecting candidate relatives is most efficient when this is done with a LR threshold.

Best scenario: single source full profile

In this case, FPR and TPR for sibs are as follows, on 10 and 15 loci:



Mixtures?

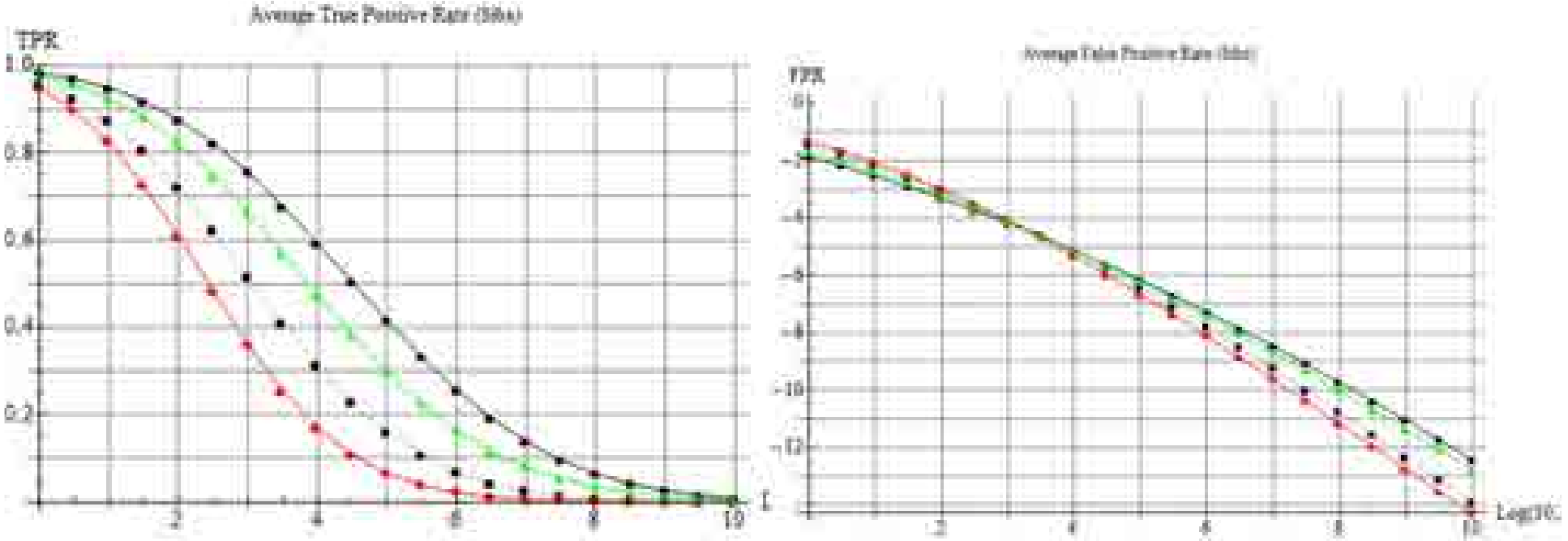
In a mixture, the profile of the PoI is mixed up with others.

Is it still feasible to find relatives of the donors?

Chung, Fung, Hu (2010) and Chung, Fung (2013) considered FS on 2p-mixtures ("victim"+"offender") in the absence of dropout and with the victim profile known.

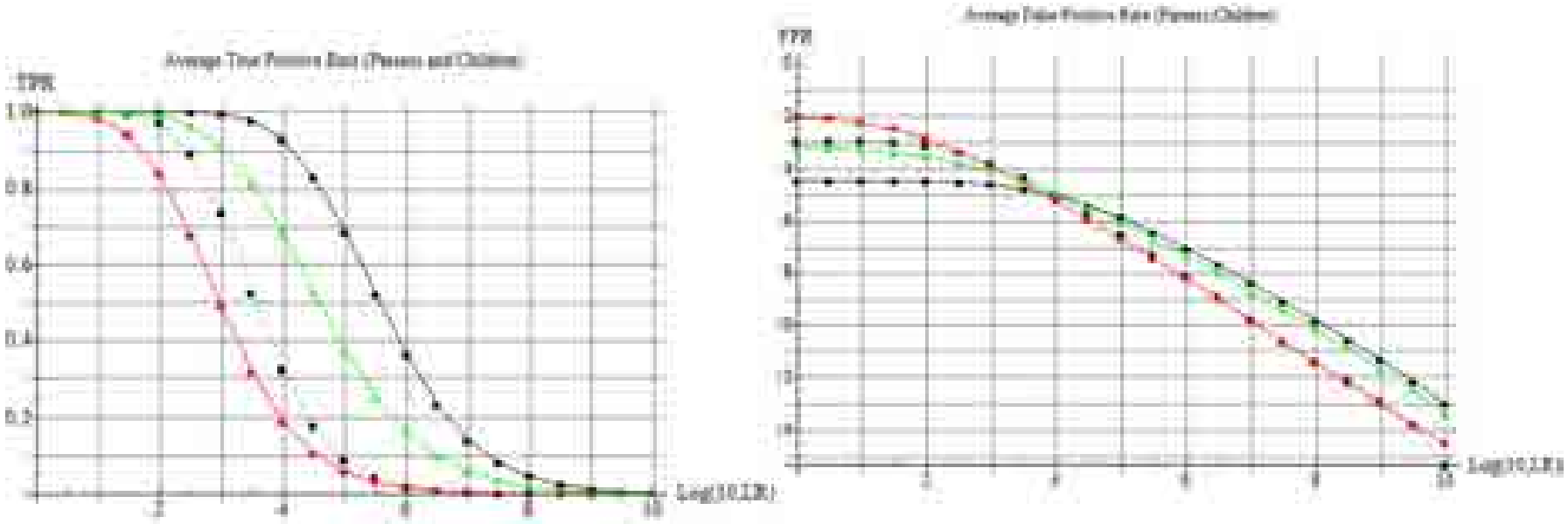
In that case, the offender profile can be deduced on some loci, and restricted to a few possibilities on the others, and it turns out that a FS is still quite effective.

TPR and FPR, 15 loci, sibs, mixtures no dropout



Black: full: single source NGM, dotted: ss SGMPlus
Green: Victim Known, Red: No donor known

TPR and FPR, 15 loci, parents and children



Black: full: single source NGM, dotted: ss SGMPlus
Green: Victim Known, Red: No donor known

Results 2p with allelic dropout

We have written a program, called MixKin, for calculations with mixtures and relatedness, taking dropout and drop-in into account.

Dropout probabilities can be chosen per contributor to the mixture.

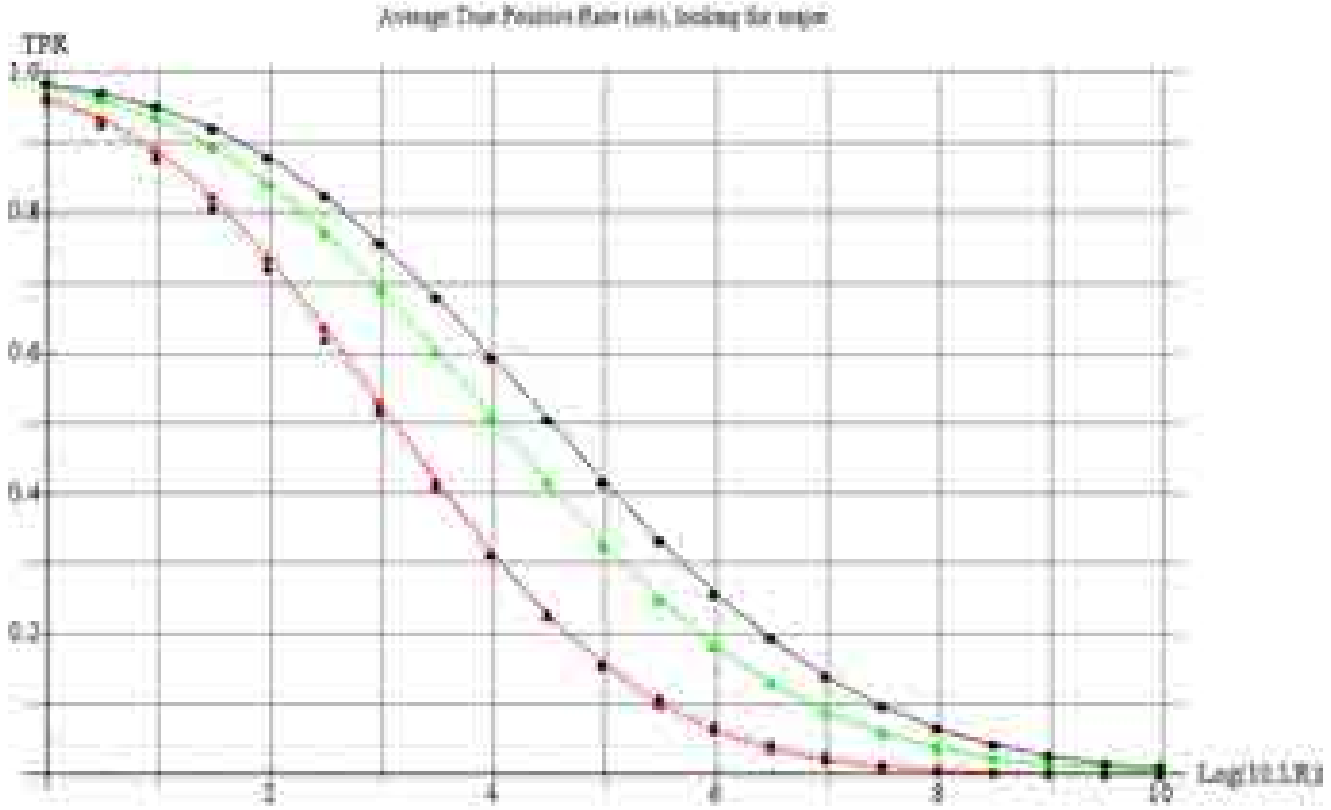
Eg

Hp: brother of PoI with dropout d1 and random person with dropout d2 are the contributors

Hd: random persons with dropout d1 and d2.

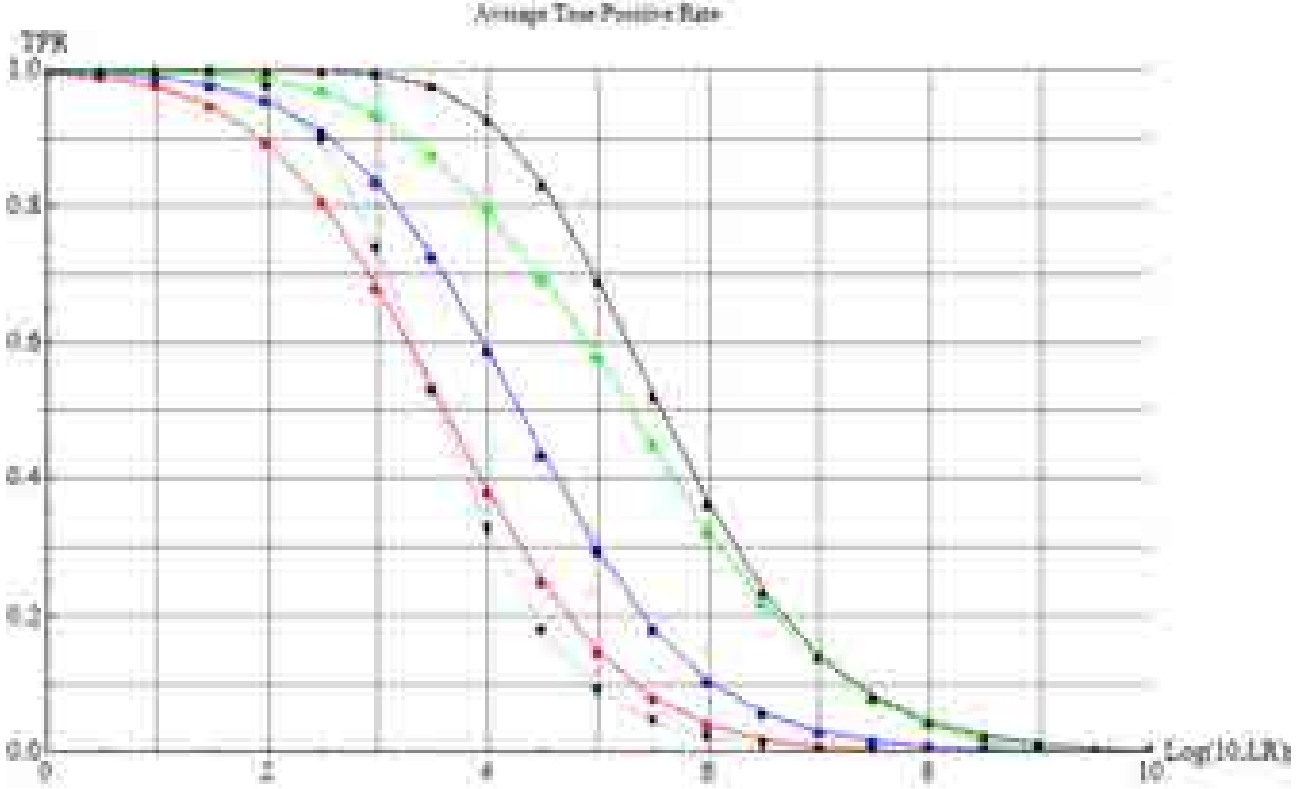
Or conditioned on a “victim”. We’ll show results for sibs and NGM, 2p and 3p. Assume three replicates for the mixture.

Different dropout: (0.05, 0.3) look for major



Black: full: single source NGM, dotted: ss SGMPlus; Green: Victim Known, Red: No donor known. (FPR almost same for all)

3p dropout (0.1,0.4,0.7),PI, major donor



Black: ss NGM (full) ss SGMPlus (dotted)
Red: three unknowns, blue: 0.4 donor known, green: 0.4 and 0.7 donors known

Conclusions

- Familial Searching is regularly done on single source traces using 10 loci, but not on mixtures
- However, NGM on mixtures can often outperform SGMplus on single source traces
- So there is no reason not to carry out a Familial Search on mixtures

- Mixtures with dropout can perform better than mixtures without! At least if one uses replicates of the mixture. In that case, a mixture with donors with different dropout can be deconvoluted more easily than one without dropout, or with equal dropout.

- Even 3p mixtures are not hopeless, especially with one or two donors known.

- Questions? K.slooten@nfi.minvenj.nl

DEVELOPMENT OF AN IN-HOUSE DNA PROFILE DATABASING/MATCHING SOFTWARE SUITE

HUSEYIN SEVAY

Department of Information Systems Engineering
Near East University, Nicosia, Cyprus

OVERVIEW

- About ...
- Problem
- Goals and Motivation
- Tools and Advantages
- Proposed Software Architecture
- Current status
- Conclusions

ABOUT ...

- Cyprus has a missing persons project (MPP)
- Expertise in DNA forensic analysis will continue to be increasingly important in years to come in Cyprus, well beyond the MPP
- Collaboration with Turkish Cypriot DNA Lab (TCDL) in conducting academic research
- Main theme:
  - Local capacity building
  - Contribution to our communities in Cyprus
  - Sharing our experiences

WHAT IS OUT THERE

- A wide variety of both commercial and public-domain software packages for DNA
  - Profile databasing
  - Searching (paternity, kinship, familial, direct)
  - Workflow management/laboratory information management
  - ... other related tasks

SOFTWARE PACKAGES

- DNA-View, Qualitype® Abetter LIMS, Qualitype GeneProof®2, GeneMapper® ID/X, Converge™ (DNA Workflow, Direct Match, Familial Match, Kinship and Paternity), Familias, FamLink/X, GenePop, Arlequin, PowerStats
- ... etc ...

PROBLEM

- For small-scale labs, commercial software could be *cost-prohibitive*
- Different labs/projects have different needs
- Different software packages, different interfaces, different learning curves
- Data format conversion is a big hassle
- Not everyone is a programming expert
- Error-prone manual handling of data

GOALS

- To develop a web-based, secure, user-friendly all-in-one software suite
- Automate main DNA profile databasing and matching tasks for a small lab
- Minimize human error
- Handle Y-STR, autosomal STR, mtDNA data, and other systems in the future
- **Leverage new technologies**
  - Programming languages, DB systems, etc.

MOTIVATION

- Response to local needs such as the CMP project and/or TCDL in Cyprus
- Need for databasing, FRS profiles for multiple systems (Y-STR, mtDNA, autosomal), familial search with skeletal/dental data vs FRS data, population statistics, etc.
- Free/open source spirit
- Platform independence

PYTHON

Wikipedia says:

- Conceived in late 1980s
- Started to be implemented late 1989
- Python 2.0 release: October 2000
- Python 3.0 release: December 2008

WHAT PYTHON OFFERS

- High-level type-free programming
- High-level built-in data structures
  - Lists, tuples, dictionaries
- Lots and lots of libraries
  - csv, xlwt/xlrd, ConfigParser, email, PyQT, SymPy, PIL, HTTPLib2, Matplotlib, ReportLab, NumPy, SciPy, etc.

FROM SCHEME TO PYTHON

WISDOM AND WONDER

SCIENCE AND ART

Home *About Me* *Hacks* *Archive*

Why MIT switched from Scheme to Python

> Costanza asked Sussman why MIT had switched away from Scheme for their introductory programming course, 6.001. This was a gem. He said that the reason that happened was because engineering in 1980 was not what it was in the mid-90s or in 2000. In 1980, good programmers spent a lot of time thinking, and then produced spare code that they thought should work. Code ran close to the metal, even Scheme — it was understandable all the way down. Like a resistor, where you could read the bands and know the power rating and the tolerance and the resistance and V=IR and that's all there was to know. 6.001 had been conceived to teach engineers how to take small parts that they understood entirely and use simple techniques to compose them into larger things that do what you want.
>
> But programming now isn't so much like that, said Sussman. Nowadays you muck around with incomprehensible or nonexistent man pages for software you don't know who wrote. You have to do basic science on your libraries to see how they work, trying out different inputs and seeing how the code reacts. This is a fundamentally different job, and it needed a different course.
>
> So the good thing about the new 6.001 was that it was robot-centered — you had to program a little robot to move around. And robots are not like resistors, behaving according to ideal functions. Wheels slip, the environment changes, etc — you have to build in robustness to the system, in a different way than the one SICP discusses.
>
> And why Python, then? Well, said Sussman, it probably just had a library already implemented for the robotics interface, that was all.

(via wingolog)

This was written by Grant. Posted on Tuesday, March 24, 2009, at 12:29. Filed under Link. Tagged Learning, philosophy, Programming, Python, Scheme, Teaching. Bookmark the permalink. Follow comments here with the RSS feed. Post a comment or leave a trackback.

MongoDB

Wikipedia says:

- Development began in 2007
- Became a stand-alone product in 2009
- Production ready since March 2010

WHAT MongoDB OFFERS

- Open source, *document-oriented* No-SQL DB
- No tables and rows as in SQL — No relations!
- Instead schema-free documents in Binary JavaScript Object Notation (BSON)
- No special DB language to *create*, *read*, *update*, and *delete* (CRUD)
- Instead a programmatic approach (function calls) to CRUD
- Powerful scale-up and replication features

AUTOBAHN

- Open-source real-time framework
- Provides implementations of the **WebSocket** protocol and Web Application Messaging Protocol (**WAMP**)
- WebSockets allow bidirectional real-time messaging on the Web
- WAMP provides *asynchronous* Remote Procedure Calls and Publish & Subscribe on top of WebSocket protocol

ADVANTAGES 1/2

- STR & mtDNA data is symbolic
- Mapping: locus ➛ maternal/paternal alleles
- JavaScript objects, Python dictionaries, and JSON/BSON data format are compatible
- An example:

```
{
  D21S11: ["25", "34.2"],
  CSF1PO: ["8", "9"],
  FGA:    ["20", "22"],
}
```

JavaScript

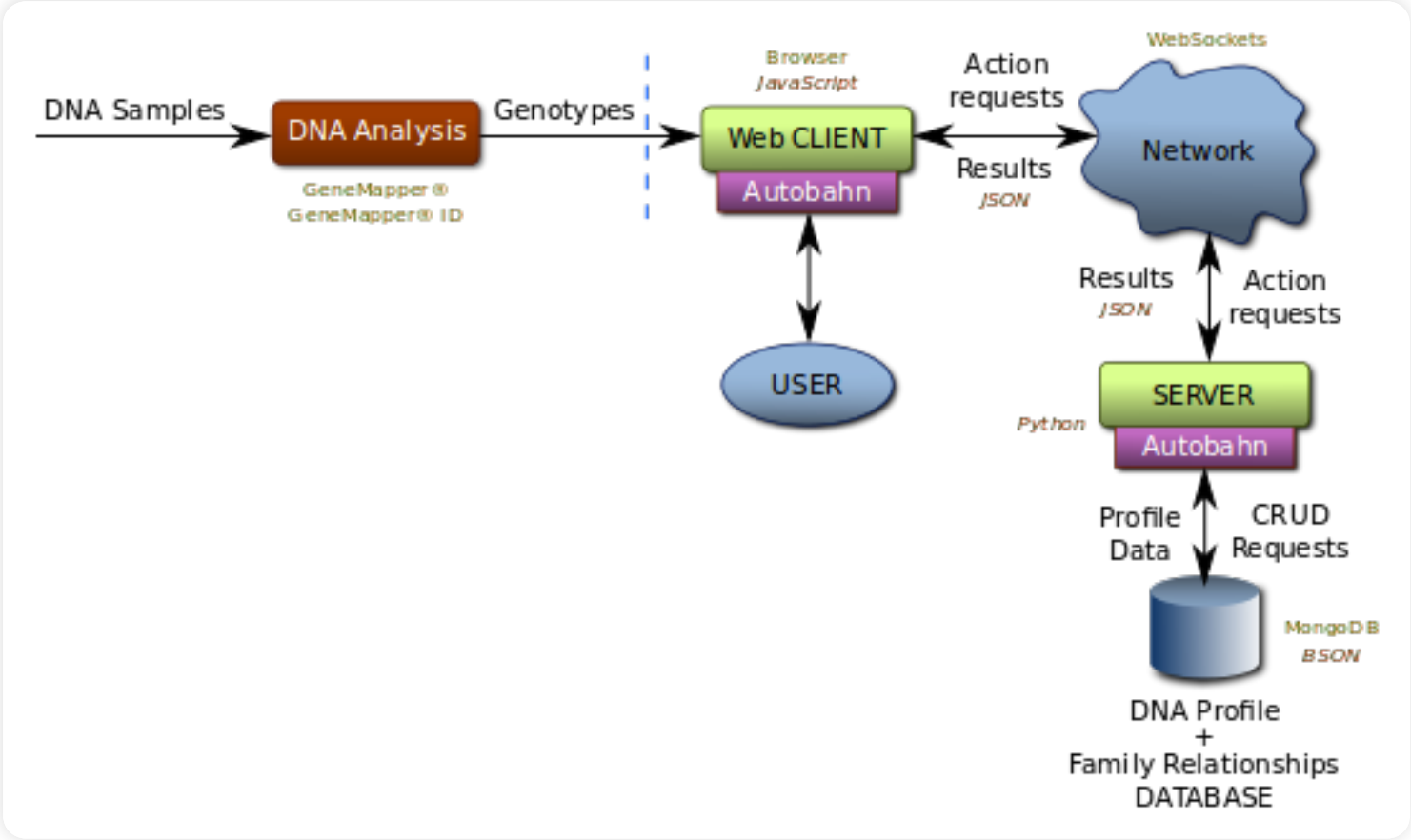
```
{
  "D21S11": ["25", "34.2"],
  "CSF1PO": ["8", "9"],
  "FGA":    ["20", "22"],
}
```

Python

ADVANTAGES 2/2

- Can implement symbolic math (SymPy)
- MongoDB server-executed custom JavaScript query capability makes direct partial matches possible
  - 70%, 80%, 100%, etc.

PROPOSED ARCHITECTURE



CURRENT STATUS

- A proof of concept for proposed architecture
- Direct percentage match system
- Computation of allele frequencies/key statistical parameters of forensic interest
- Automatic computation of k-values, F/Θ ...
- Investigation of a symbolic computation model for DNA search statistics (SymPy)

CONCLUSIONS

- Critical for small-scaled labs to have an all-in-one solution
- Web offers great potential for visualization
- Working progress ...
- Leveraging of new technologies for current and future needs
- Creation of a cost-effecive solution

ACKNOWLEDGMENTS

- Dr Cemal Gürkan, TCDL
- Salim Jibrin, MSc student

THANKS

hsevay@gmail.com

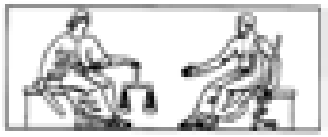
Faculty of Health and Medical Sciences

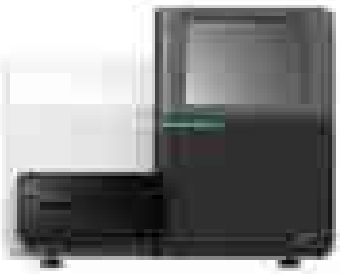
STR sequencing on second generation sequencing platforms

Eszter Rockenbauer, MSc, PhD

Section of Forensic Genetics
Department of Forensic Medicine
Faculty of Health and Medical Sciences
University of Copenhagen
Denmark

SGS in Copenhagen





GS Junior 454 Sequencing system (2009)

- mtDNA sequencing
- STR sequencing

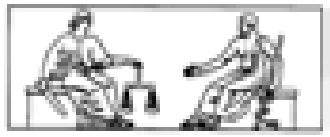
MiSeq (2013)

- mtDNA sequencing
- miRNA
- X chromosome haplogroups etc …
- Illumina SGS Forensic kit >200 markers

Ion PGM™ (2013)

- Ion Torrent™ HID SNP assay (169plex)
- Ion Torrent™ STR 10plex assay

STR sequencing using SGS



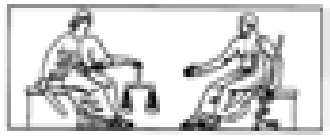
Strategy on the GS Junior*

- Amplicon sequencing with Multiplex IDentifiers (MIDs) in the PCR primer sequence
- Quantify and pool amplicons into a library
- emPCR and sequencing

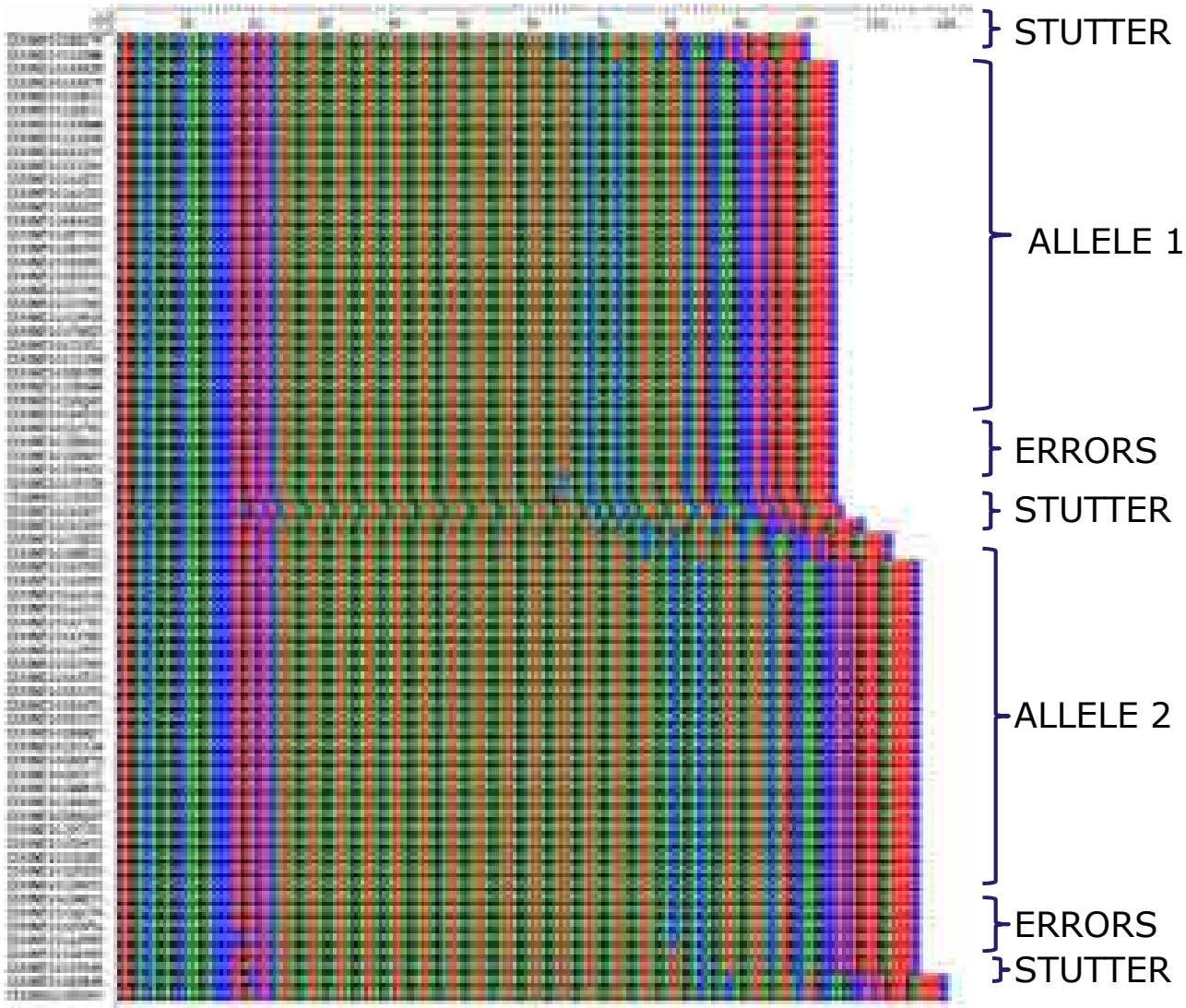
Data analysis

- Filter by MID
- Filter by STR
- At least one primer binding region must be present
- Both STR ends must be present
- Trim the reads according to flanking sequence
- Alignment in BioEdit (Ibis Biosciences)

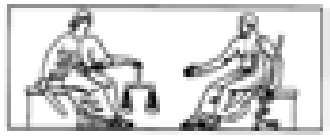
*Fordyce et al., Biotechniques 51 (2011) 127-133



STR sequencing using SGS



STR sequencing using SGS



SGS based allele frequency database

- 394 sequenced alleles in unrelated Danes*
- 256 sequenced alleles in unrelated Somalis
- D3S1358, D12S391, D21S11

No. of alleles in total

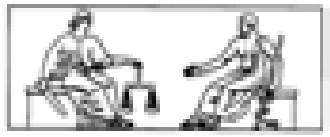
	PCR-CE	SGS
D3S1358	9	20
D12S391	19	69
D21S11	17	49
Total	45	138

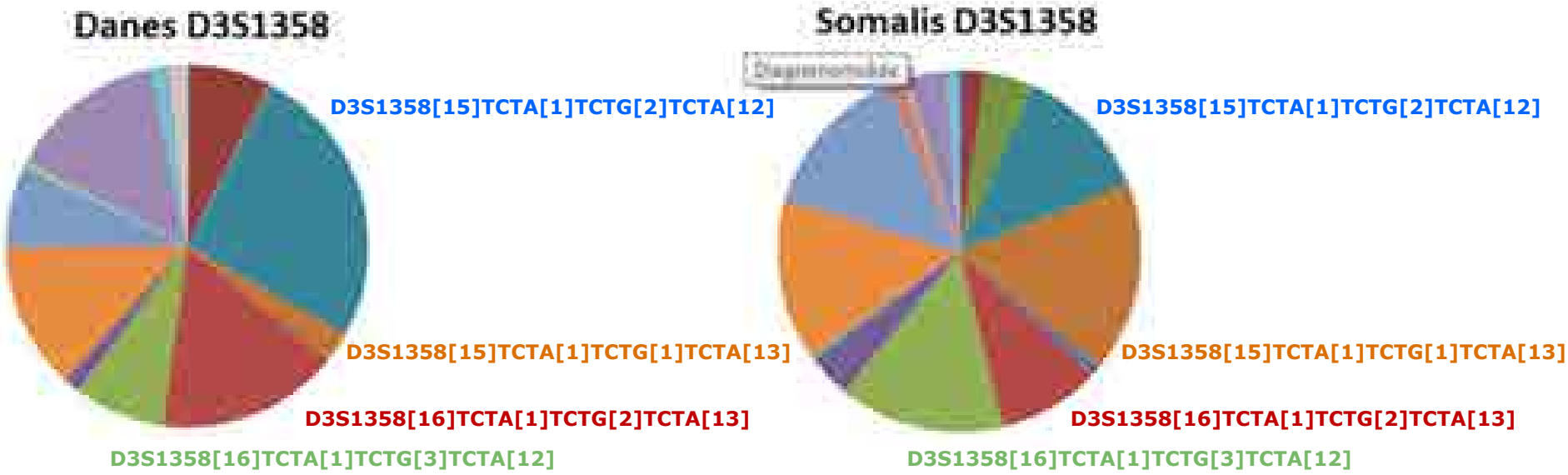
No. of alleles only seen in one group

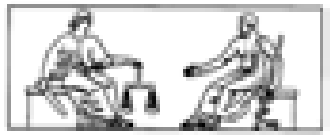
	Danes	Somalis
D21S11	4	3
D12S391	30	16
D3S1358	15	20
Total	49	39

*Gelardi, Rockenbauer et al., in press

STR sequencing using SGS







STR sequencing using SGS

- STR loci: D3S1358, D12S391, D21S11

Forensic statistics – Danes*

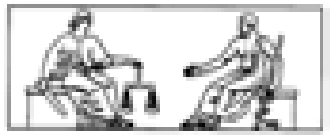
	PCR-CE	SGS
Power of discrimination	0.9999	0.999995
Paternity exclusion power	97.1	99.2
PI (trio)	59.2	415.0
PI (duo)	16.1	82.4

Forensic statistics - Somalis

	PCR-CE	SGS
Power of discrimination	0.9997	0.999998
Paternity exclusion power	94.0	99.6
PI (trio)	41.1	544.0
PI (duo)	12.4	103.0

*Gelardi, Rockenbauer et al., in press

STR sequencing using SGS



STR nomenclature*

- New guidelines are needed
- Locus name used in forensic genetics
- Length of repeat/length of repeat unit
- Sequence details
- Polymorphisms in the flanking region (SNPs and indels)

D12S391[21]AGAT[11]AGAC[9]AGAT[1]
D12S391[21]AGAT[11]AGAC[10]
D12S391[21]AGAT[12]AGAC[8]AGAT[1]
D12S391[21]AGAT[12]AGAC[9]
D12S391[21]AGAT[13]AGAC[7]AGAT[1]
D12S391[21]AGAT[13]AGAC[8]
D12S391[21]AGAT[13]GGAC[1]AGAC[7]
D12S391[21]AGAT[14]AGAC[6]AGAT[1]
D12S391[21]AG**G**T[1]AGAT[12]AGAC[9]

*Gelardi, Rockenbauer et al., in press

STR sequencing using the Ion PGM

Ion Torrent™ STR 10plex assay

- 10plex STR kit with forensic core STRs designed for the Ion PGM
- Based on the CODIS system
  - AMEL
  - CSF1PO
  - D16S539
  - D3S1358
  - D5S818
  - D7S820
  - D8S1179
  - TH01
  - TPOX
  - VWA

Concordance study (10 Danes)*

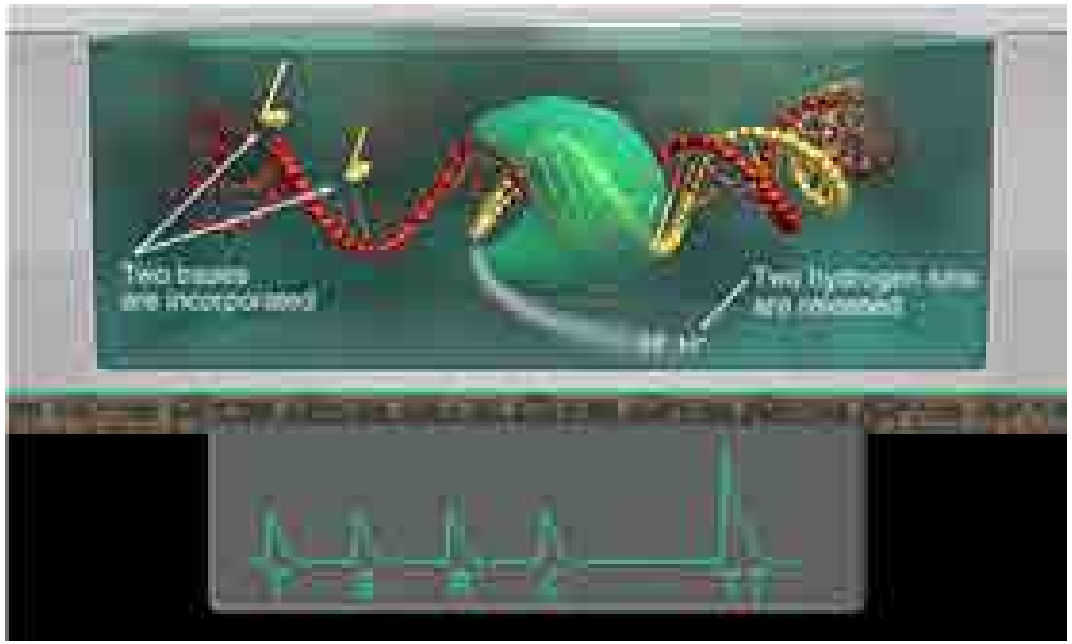
- Complete concordance with AmpFlSTR® IDFL Plus PCR Amplification Kit

*Fordyce et al., in prep.

STR sequencing using the Ion PGM

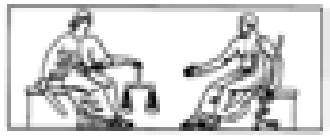
The Ion Torrent STR 10plex

- Amplification with multiplex PCR
- Amplicon sizes 70-170 bp
- Ligation of barcoded adaptors
- Quantification, emPCR and sequencing*



*Semi-conductor technology (Rothberg et al. 2011)

STR sequencing using the Ion PGM



Data analysis with the Torrent Suite v4.0.2 software

HID_STR_Genotyper plugin:

- Profile summary with allele calls
- Coverage plots
- Locus data (only STR sequence!)

Flanking sequences

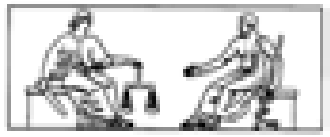
- Available from raw data (BAM-files)

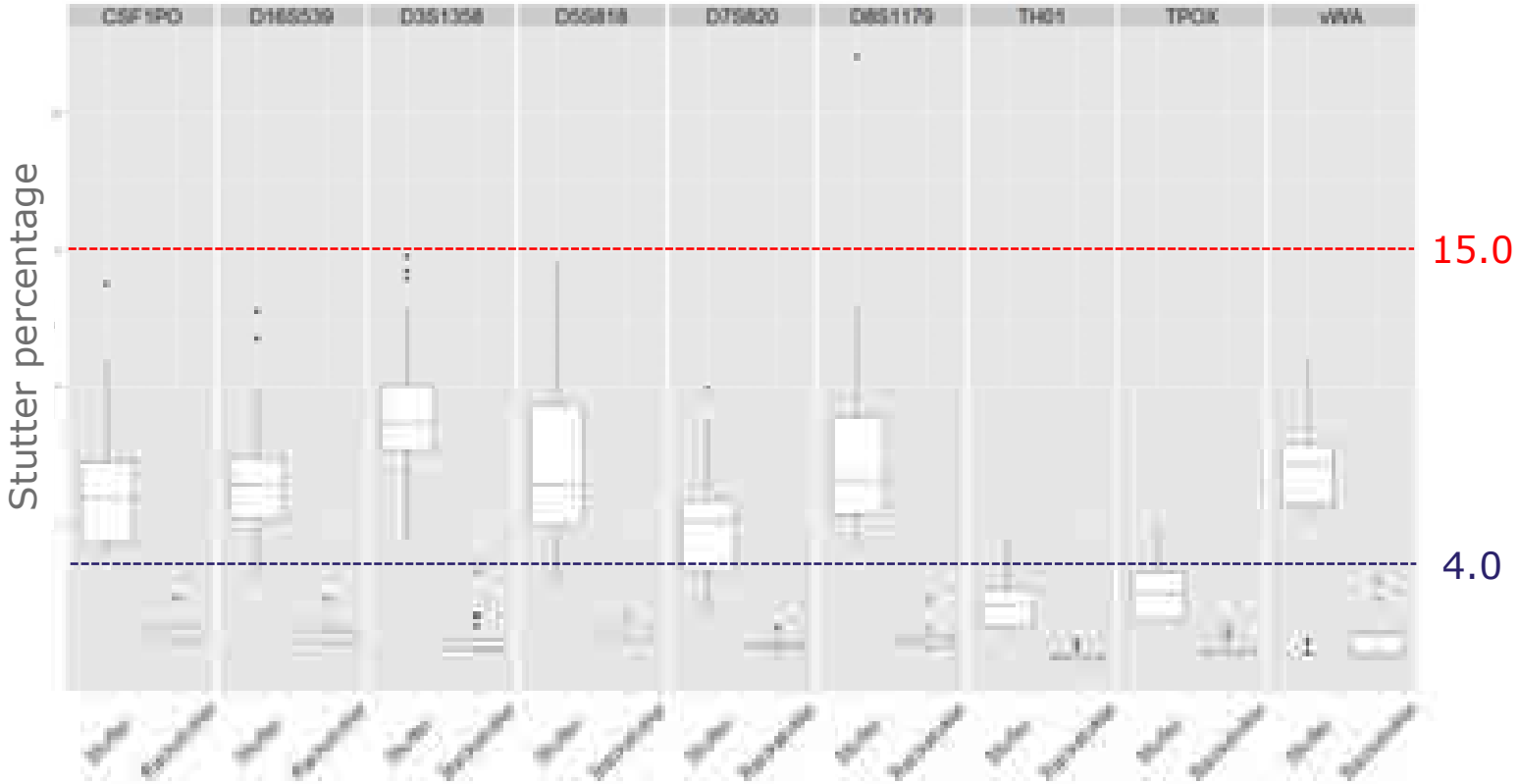
Exportable text files with additional data

- Genotype file and histograms

*Fordyce et al., in prep.

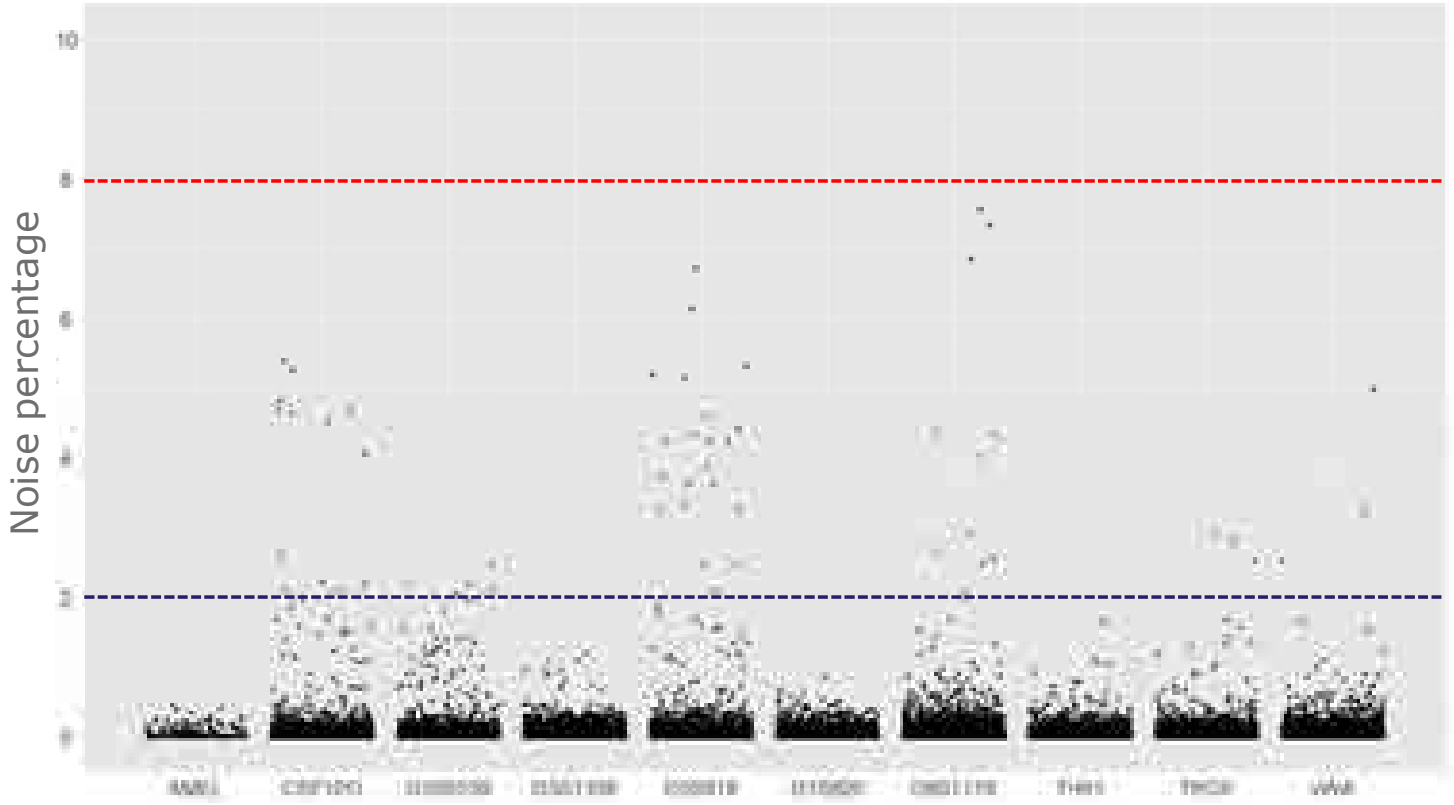
STR sequencing using the Ion PGM





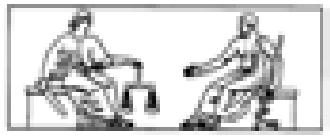
Fordyce et al., in prep.

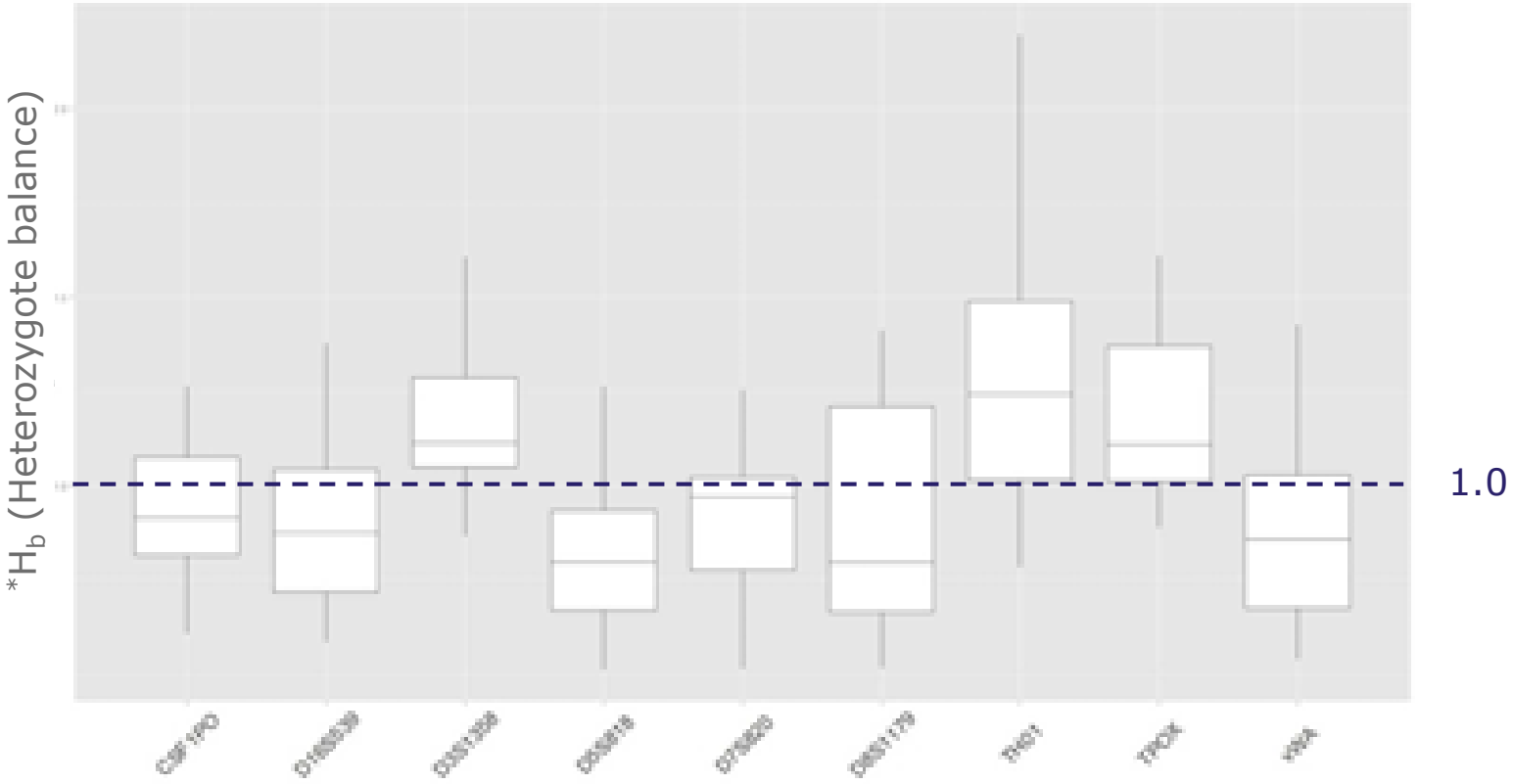
STR sequencing using the Ion PGM



Fordyce et al., in prep.

STR sequencing using the Ion PGM

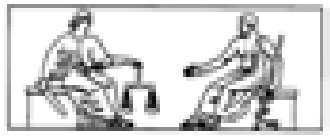




*H_b = HMW/LMW

Allele balance

Fordyce et al., in prep.



Acknowledgements

Professor and Director of the Institute:
Niels Morling

Postdoc:
Sarah Fordyce

Forensic geneticists:
Claus Børsting
Helle S. Mogensen
Lena Poulsen

Students:
Chiara Gelardi
Johanna Manninen
Sigrun Dalsgaard

Thank you for your attention

RÄTTSMEDICINALVERKET
NATIONAL BOARD OF FORENSIC MEDICINE

www.rmv.se

Evaluation of genetic testing using NGS of sudden cardiac death genes in sudden unexplained deaths

Henrik Grèen
Associate Professor in Pharmacogenetics
Department of Forensic Genetics and Forensic Toxicology
National Board of Forensic Medicine
Sweden

Genetics in Sudden Unexpected Deaths

Using standard forensic autopsy routines it is often difficult to determine the causes of Sudden Cardiac Deaths (excluding coronary artery diseases).

- In some cases no structural abnormalities can be detected: “autopsy negative” or “sudden unexplained death” (SUD).

- Macroscopically and/or microscopically visible myocardial abnormalities can be seen in some case

Studies have shown that sudden unexplained death to a high extent is caused by genetic cardiac disorders[1,2]

1 Wisten, A., et al., Mutation analysis of cases of sudden unexplained death, 15 years after death: prompt genetic evaluation after resuscitation can save future lives. Resuscitation, 2012. 83(10): p. 1229-34.
2 Skinner, J.R., et al., Prospective, population-based long QT molecular autopsy study of postmortem negative sudden death in 1 to 40 year olds. Heart rhythm : the official journal of the Heart Rhythm Society, 2011. 8(3): p. 412-9.

Genetics in Sudden Cardiac Deaths – Our approach

Screening of multiple genes linked to SCD associated diseases.

- Large panel for several diseases / syndromes
- Limit the bioinformatics to the genes of the specific sample

Prediction of genetic variants.

- How can we help the forensic pathologist?
- How do we interpret new variants?

RÄTTSMEDICINALVERKET

www.rmv.se

Large gene panel for forensic use

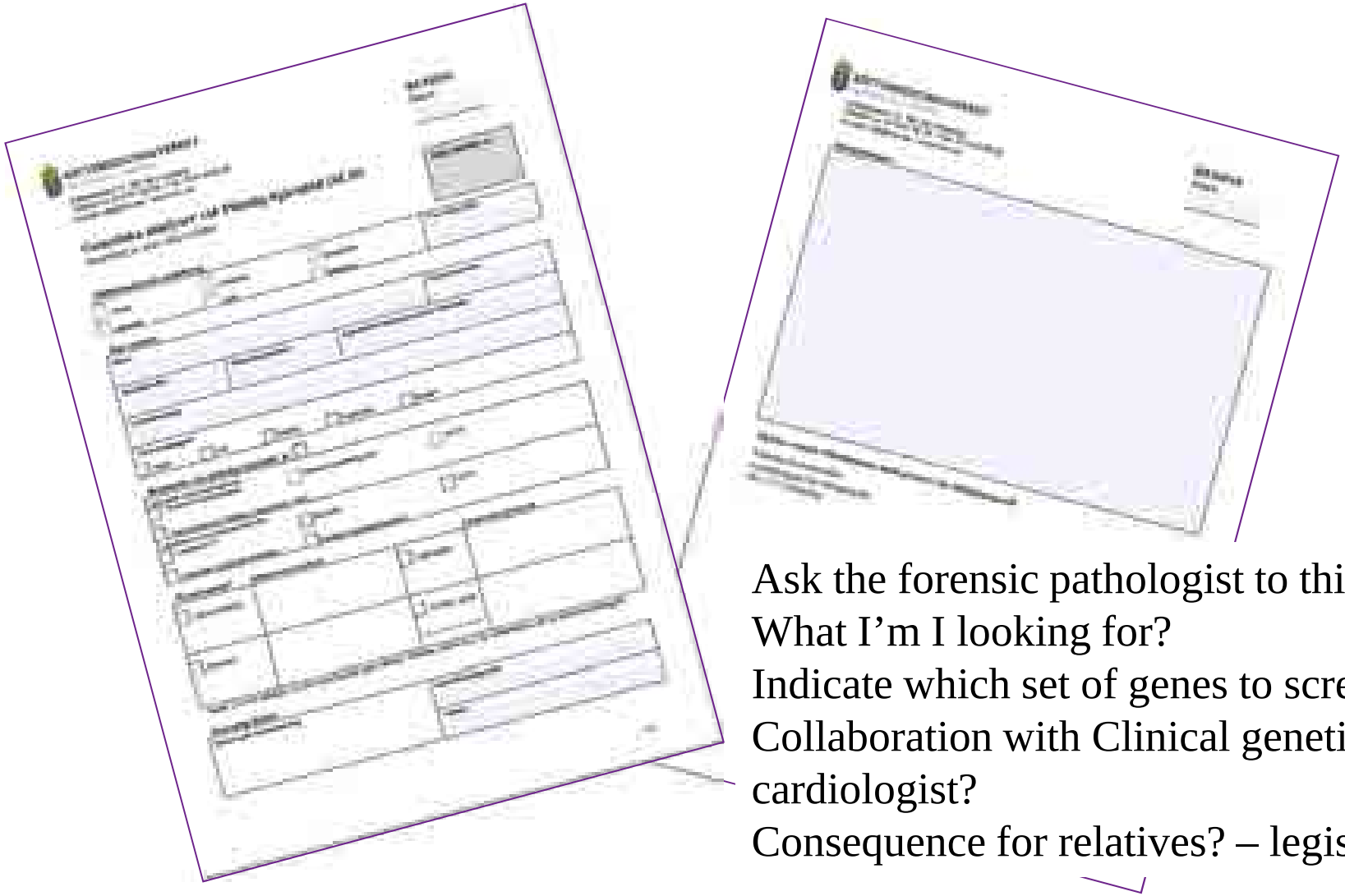
Analyzed by several certified labs in the world for and literature references for disease associated genes

Genes divided into:

- High and low priority genes

Long QT Syndrome	Short QT sydrome	Brugada syndrome	CPVT	Idiopathic Ventricular Fibrillation	Hypertrophic Cardiomyopathy	Dilated Cardiomyopathy	Marfan Syndrome/ Loeys-Dietz Syndrome/ Familial Thoracic Aortic Aneurysms and Dissections	ARVC	Familjär hyperkolesterolemi
AKAP9	CACNA1C	**CACNA1C**		CACNA1C	**ACTC1**	ABCC9	**ACTA 2**	**DES**	APOb
ANK2	CACNA2D1	**CACNB2**	**CASQ2**	CACNA2D1	ACTN2	**ACTC1**	**COL3A1**	**DSC2**	LDLR
CACNA1C	CACNB2	**GPD1L**	**KCNJ2**	CACNB2	CAV3	**ACTN2**	COL5A1	**DSG2**	LDLRAP1
CAV3	**KCNH2**	**KCNE3**	**RYR2**	KCNJ8	**CSRP3**	ANKRD1	COL5A2	**DSP**	PCSK9
KCNE1	**KCNJ2**	**SCN1B**			FHL1	**BAG3**	**FBN1**	**JUP**	
KCNE2	**KCNQ1**	**SCN3B**			**GLA**	CSRP3	**TGFBR1**	**PKP2**	
KCNH2		**SCN5A**			**LAMP2**	**DES**	**TGFBR2**	RYR2	
KCNJ2					**MYH6**	**DMD**		TGFB3	
KCNJ5					**MYH7**	**DSG2**		**TMEM43**	
KCNQ1					**MYL2**	EMD			
SCN4B					**MYL3**	**LDB3**			
SCN5A					**PLN**	**LMNA**			
SNTA1					**PRKAG2**	**MYBPC3**			
					TCAP	MYH6			
					TNNC1	**MYH7**			
					TNNI3	NEXN			
					TNNT2	**RBM20**			
					TPM1	**SCN5A**			
					TTN	***TAZ***			
					TTR	TCAP			
						TMPO			
						TNNC1			
						TNNI3			
						TNNT2			
						TPM1			
						TTN			
						VCL			
		CACNA2D1	ANK2		ANKRD1	CAV3	FBN2	TTN	
					BAG3	DSC2	MYH11		
		KCNJ8			LDB3	DSP	*MYLK*		
					MYBPC3	FHL1	SMAD3		
					MYLK2	FHL2			
					MYOZ2	GLA			
					NEXN	LAMP2			
					RYR2	MYL2			
					VCL	MYL3			
						NEBL			
						PKP2			
						PLN			
						PRKAG2			
						TTR			

Referral



Ask the forensic pathologist to think of:
What I'm I looking for?
Indicate which set of genes to screen for.
Collaboration with Clinical geneticist/ cardiologist?
Consequence for relatives? – legislation?

Sample material, target enrichment and sequencing

- Blood from the heart, femoral vein (Forensic toxicology) or heart muscle
- QC prior to target enrichment
  - Nanodrop: 260/230 (>1.4), 260/280 (>1.7) – ratios
  - Tapestation: – “minor smearing” below 2.5 kb
  - Qubit for quantification
- HaloPlex target enrichment and amplification for sequence capture
- MiSeq for massive parallel sequencing using of sudden cardiac genes 10 bp flanking

MiSeq

4-6 samples multiplexed
MiSeq Reagent Kit v2, 300 cycles or
MiSeq Reagent Kit v3, 300 cycles
Cluster density 1000 000 - 1400 000 clusters/mm^2.

Bioinformatic pipeline

Variant classification

- >99% have a coverage of >20x
- Genetic variants with an allele freq <0.02 in dbSNP is checked for pathogenicity in dbSNP and the public domain of HGMD – Human Gene Mutation Database.

- Genetic variants with an allele freq <0.001 in dbSNP or exome variant server is determined for predicted pathogenicity using conservation, PROVEAN, SIFT, MutationTaster and recently also polyphen.
  - Only high priority genes!

- Splice and stop always reported - if not common or close to 3'-end (50 bp – depending on gene size).
  - Both high and low priority genes.

Results from the genotyping

We have gotten referrals from all six Dept of Forensic Medicine in Sweden. About 1 referral each week.

In total 41 referrals so far:

- 32 finished and 9 on-going.
- Age range 0-41 years.

Results (of the 32)

- 6 individuals with pathogenic variants
- 8 individuals with probably pathogenic variants (predicted)
- Including genes: LDB3, KCNH2, KCNQ1, TNNC1 and others
- Associated with Long QT, ARVC, HCM, etc

Conclusions

- NGS works excellently for screening for genetic variants. Even in ”not too decomposed” material.

- Genetic information in SUD cases can be useful in getting a basis for considering SCD.

Acknowledgement

Anna Gréen and Cecilia Gunnarsson

Department of Clinical Genetics and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden.

Jonny Kumlin and Henrik Gréen

Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden.

Peter Krantz Department of Clinical Sciences, Malmö, Lund University
Department of Forensic Medicine, Lund, Sweden

The forensic pathologists at the Departments of Forensic Medicine in Sweden!

Thank you for listening!

Population Genetics of 17 Y-STR Markers in Turkish Cypriots from Cyprus

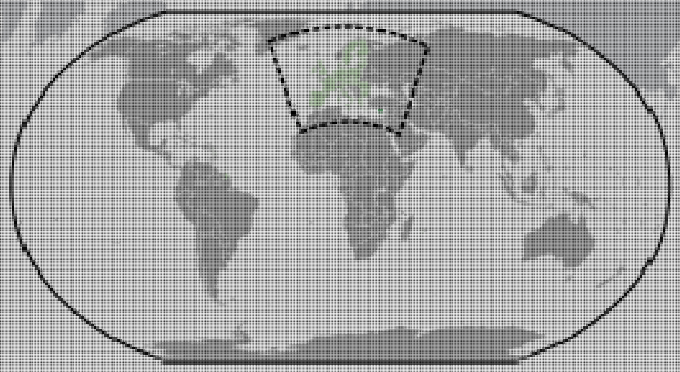
Cemal Gürkan, Ph.D.

Turkish Cypriot DNA Laboratory (TCDL)

***ISFG ESWG Meeting in Athens,* May 30, 2014**

Outline

1. Cyprus & Turkish Cypriots
2. Committee on Missing Persons in Cyprus
2. Turkish Cypriot DNA Laboratory
3. Y-STR Genetics of Turkish Cypriots
5. Acknowledgements



Cyprus

- Third largest island in the Mediterranean Sea
- Situated at the crossroads of civilizations
- Earliest human activity dates back to 12,000 years ago
- Rich cultural heritage and ethnic composition:

  Greek Cypriots, Turkish Cypriots, Maronites, Armenians & Latins

  * Independent since 1960

  * Divided since 1974

  * Member of E.U. since 2004

Taken from Wikipedia @, 2014

Turkish Cypriots

- First historical appearance upon Ottoman conquest of Cyprus (1571)
- Individuals trace their origins to:
  - *(a) Ottoman officers & soldiers stationed in Cyprus*
  - *(b) Settlers from around the Ottoman Empire*
  - *(c) Forced exiles from the Karaman region in Anatolia*
  - *(d) Converts from Greek Cypriots & Latins*
  - *A combination of all with unknown actual proportions*
- Today, the second largest ethnic community on the island
- Traditionally lived all over the island
- Largely moved to/concentrated in North Cyprus since 1974
- 2006 census by the Turkish Cypriot authorities suggest:
  - *(a) de jure population: 256,644*
  - *(b) born in Cyprus: 148,542*
  - *(c) both parents born in Cyprus: 120,007*

Committee on Missing Persons (CMP) in Cyprus

- Established in 1981 as a bi-communal body
- 1 TC Member, 1 GC Member & 3rd member selected by ICRC & appointed by UN SG
- Recovery, identification & return to families the remains of Cypriots that went missing during the inter-communal fighting of 1963/1964 and the events of 1974

493 Turkish Cypriots & 1508 Greek Cypriots

http://www.cmp-cyprus.org/

The CMP Project & Current Contributors

Current Contributors:

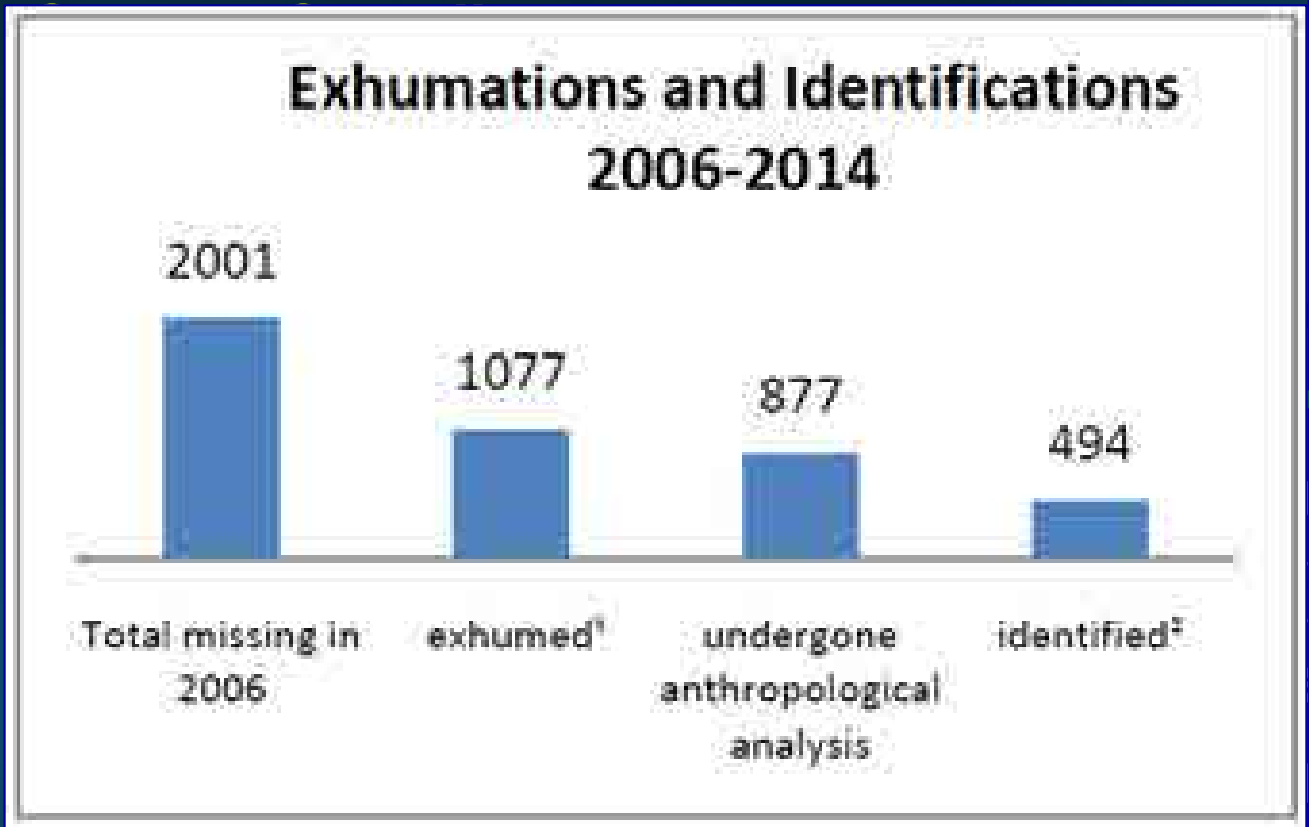
> CMP Bi-communal Forensic Team
(*local archaelogists, anthropologists & geneticists*)

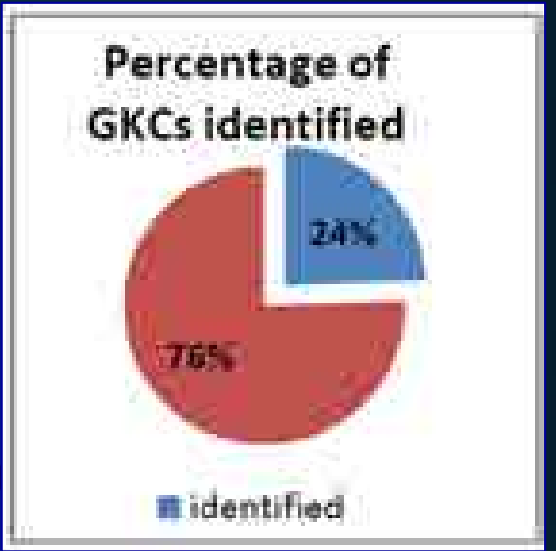
> Turkish Cypriot DNA Laboratory (TCDL)
(*TC FRS DNA typing & TC pop studies*)

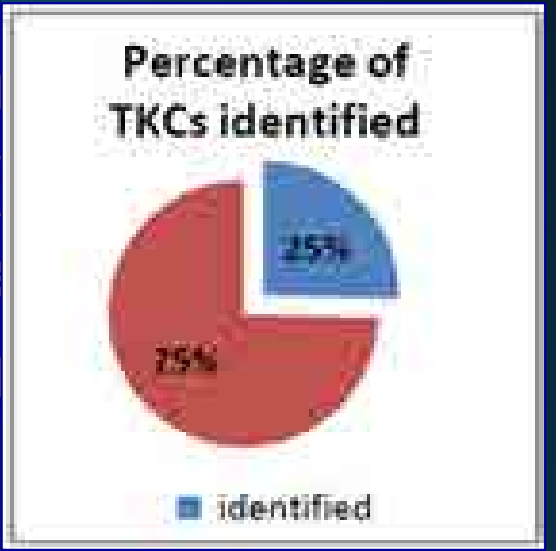
> Greek Cypriot DNA Laboratory (CING LabFoG)
(*GC FRS DNA typing & GC pop studies*)

> International Commission on Missing Persons (ICMP)
(*skeletal/dental DNA typing & matching w/ FRS data*)

The CMP Project & Current Contributors







Figures taken from the CMP website (30 April 2014)

Turkish Cypriot DNA Laboratory (TCDL)

- Set up in late 2005 to contribute to the CMP Project
- Operates under the CMP Turkish Cypriot Member Office
- TC FRS Sample Bank of 1,100+ relatives
- All nSTR/ySTR profiling at TCDL since the summer of 2012

  (*Initial nSTR/ySTR/mtDNA analyses were @ CING LabFoG*)
- Also responsible for conducting TC population sudies
- TC Population Sample Bank of 600+ volunteers

~ 600/120,000 (~0.5% coverage of all TC's w/ Cypriot parents)

Turkish Cypriot Y-STR Population Study

1. Sample collection from 253 healthy, unrelated volunteers along with informed consent
2. Volunteers aged 18 & above, with at least their father also born in Cyprus

   (*i.e., all are descendents from the pre-1974 TC population*)
3. Each Y-STR haplotype is assigned to a traditional geographic location in Cyprus

Turkish Cypriot Y-STR Population Study

Findings recently published in

Forensic Science International: Genetics 10 (2014) e1-e3.

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Forensic Population Genetics – Short Communication

Population genetics of 17 Y-STR markers in Turkish Cypriots from Cyprus☆

K. Terali[a], T. Zorlu[b], O. Bulbul[b], C. Gurkan[a,*]

[a] Turkish Cypriot DNA Laboratory, Committee on Missing Persons in Cyprus Turkish Cypriot Member Office, Nicosia, North Cyprus
[b] Institute of Forensic Sciences, Istanbul University, Istanbul, Turkey

First ever Y-STR dataset from Cyprus in the literature

Turkish Cypriot Y-STR Population Study

Major Findings - I: *TC haplotype / gene diversity*

1. Among 253 haplotypes (Ht's),

   229 different Ht's (Discrimination Capacity, DC: 90.51%)

   206 unique (81.42%), 22 in duplicate & 1 in triplicate

2. 7 out of the 22 Ht pairs in duplicate, 1 pair in the triplicate also shared the same traditional georgraphic origin

   (*i.e. apparently not closely related, but shared paternal lineage*)

3. Most common Ht (*1.19%*) never been reported before

   (*@YHRD, ~56,000 Ht's as of 22 August 2013*)

4. Calculated haplotype diversity (HD) is 0.9992 (exc. Dys385a/b)

5. Calculated av. gene diversity (GD) is 0.6429 (exc. Dys385a/b)

Turkish Cypriot Y-STR Population Study

Major Findings - II: *Allelic variants observed*

1. No locus duplications/null alleles were observed
2. 43 allelic variants based on observed fragment sizes
   (a) 25x Dys458*.2 variants (*in ~10% of all Ht's*)
   > *associated with the Y-haplogroup J1 (Middle East)*
   > 8.2% in N. Greece, 16.8% in Cukurova, TR, 4.5% in Italy
   (b) various other allelic variants @ 8 different loci
   > 1 novel never reported before: Dys438*9.4
   > 1 variant reported once before in Brasil: Dys458*14.3
   > Both fully characterized by sequencing @ NIST

Allelic variants are powerful tools for pop genetics/forensics

Turkish Cypriot Y-STR Population Study

Major Findings - III: *AMOVA using YHRD online tool*

Table S3. Pairwise genetic distance matrix based on the Φ_{st} values between the Turkish Cypriot population and neighboring/distant populations

Population	AuAbo	TrCAr	PsArb	Lithu	GrNrt	TrSEA	TwPwn	TrCkr	Egypt	Italy	TrCyp
AuAbo	-	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TrCAr	0.2101	-	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PsArb	0.1527	0.1771	-	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Lithu	0.1025	0.2860	0.2212	-	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GrNrt	0.1272	0.1033	0.1268	0.1399	-	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TrSEA	0.0996	0.0427	0.0562	0.1180	0.0324	-	0.0000	**0.1127**	0.0004	0.0000	**0.0776**
TwPwn	0.3663	0.4990	0.4691	0.4621	0.4273	0.4382	-	0.0000	0.0000	0.0000	0.0000
TrCkr	0.0702	0.0326	0.0897	0.0945	0.0255	0.0086	0.3735	-	0.0000	0.0000	0.0002
Egypt	0.1352	0.1438	0.0491	0.1566	0.0458	0.0246	0.4200	0.0441	-	0.0000	0.0000
Italy	0.0444	0.1686	0.1194	0.1186	0.0712	0.0514	0.3708	0.0277	0.0673	-	0.0000
TrCyp	0.1203	0.0492	0.0845	0.1507	0.0189	0.0064	0.4118	0.0150	0.0229	0.0644	-

P values are shown above and the Φst values below the diagonal.

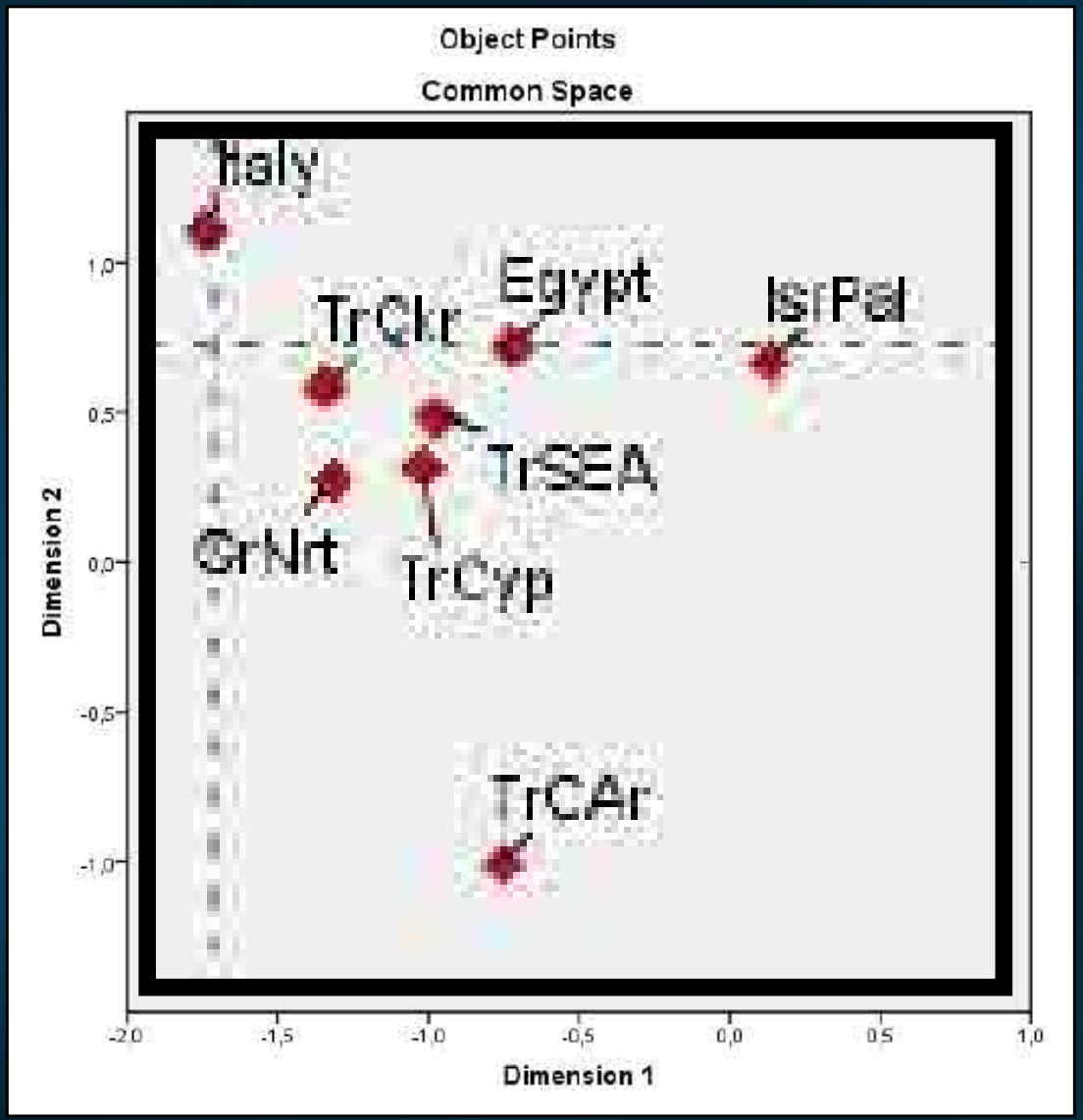
P values in bold and italics represent statistically insignificant differences before and after applying the Bonferroni correction, respectively

The level of significance is $p<0.01$ and $p<0.00018$ before and after the Bonferroni correction, respectively

Populations for comparisons largely selected based on the availability of similar datasets from nearby populations

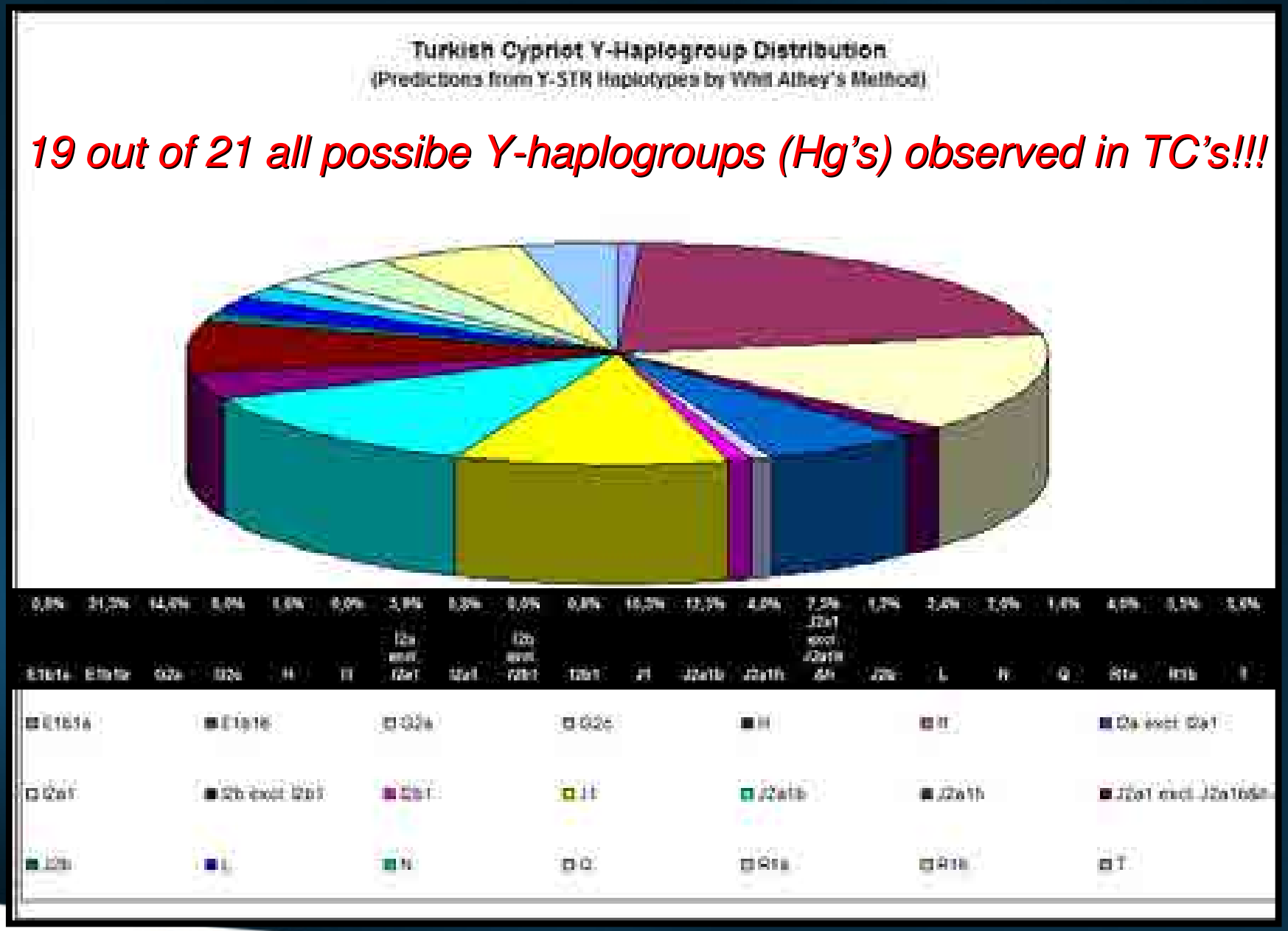
Turkish Cypriot Y-STR Population Study

Major Findings - IV: *2-D MDS plot of AMOVA results*



Follow-Up Studies on the Y-STR data

Y-haplogroup predictions based on the Y-STR data (*n* = 253)



ACKNOWLEDGEMENTS

1. ISFG ESWG Athens Meeting Organizers

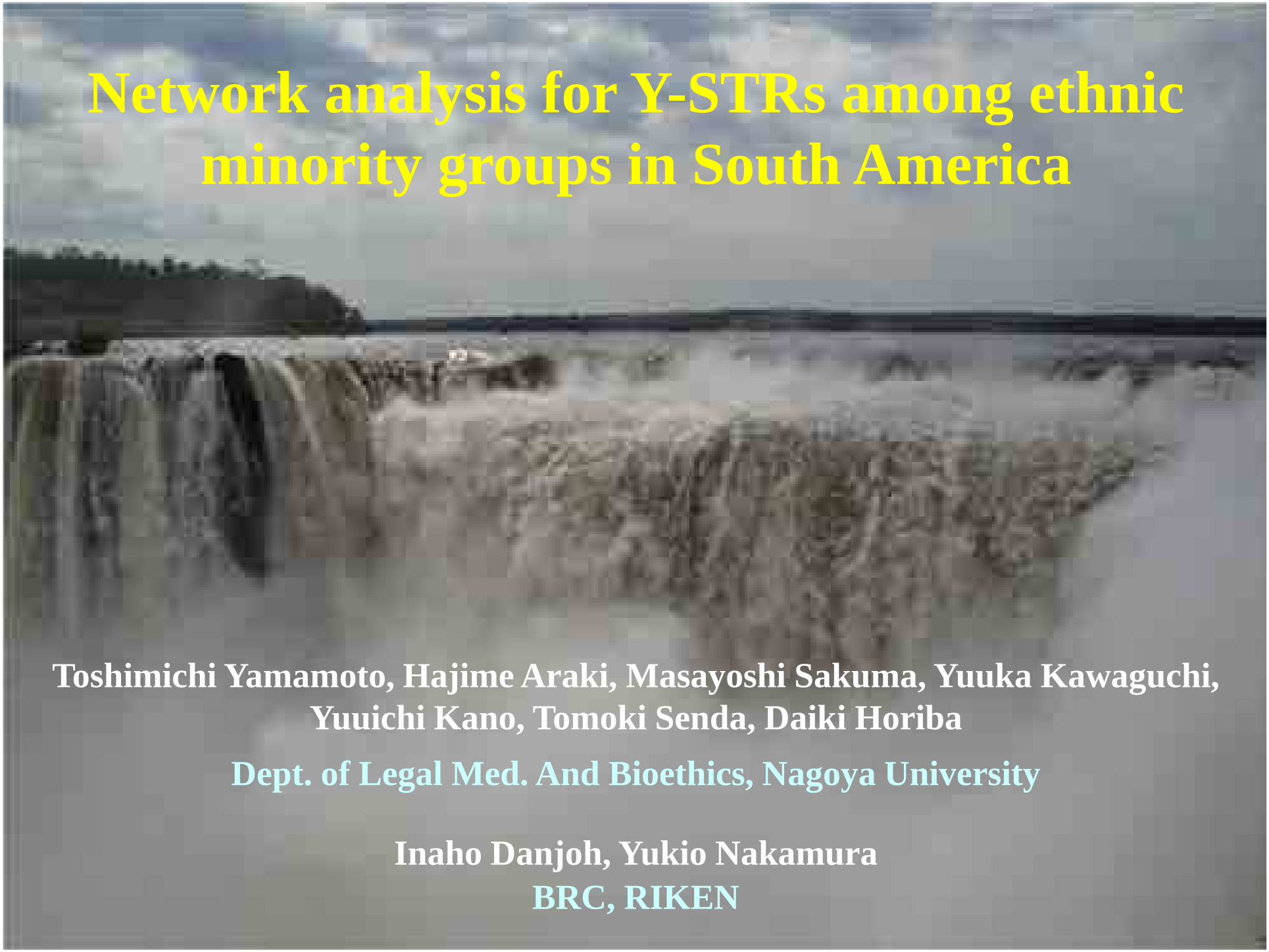
2. TCDL staff (Dr. K. Terali & D. K. Demirdov)

3. T. Zorlu & O. Bulbul @ Istanbul Univ. Institute of Forensic Science

4. Carolyn R. Hill & Dr. Peter Vallone @ NIST for variant sequencing

5. YHRD for Y-STR Haplotyping Quality Assurance Exercise 2013

6. CMP Turkish Cypriot Member Office (Mrs. G. Plumer-Kucuk)



Network analysis for Y-STRs among ethnic minority groups in South America

Toshimichi Yamamoto, Hajime Araki, Masayoshi Sakuma, Yuuka Kawaguchi, Yuuichi Kano, Tomoki Senda, Daiki Horiba

Dept. of Legal Med. And Bioethics, Nagoya University

Inaho Danjoh, Yukio Nakamura

BRC, RIKEN

“Sonoda-Tajima” collection

http://www.brc.riken.jp/lab/cell/english/index_hsc.shtml

“Sonoda-Tajima” collection

Totally more than 3,500 of blood samples from a variety of ethnic minorities across the world, especially about 2,500 samples from South America,

To study on phylogenetical classification of HTLV-I, and -II

They concluded that ancestors of the Amerind populations carried HTLV-I and -II into the South American continent from the Eurasian continent over 10,000 years ago, and that the indigenous South American populations could be divided into two major ethnic groups.

“Sonoda-Tajima” collection

Donation of the all of these blood samples to RBC in RIKEN, Japan

Unliked Anonymization of all of these blood samples

Except the information about

Age, Sex, Ethnicity, Collection site

Establishment of the cell lines transformed by B-lymphoid cells using the Epstein–Barr (EB) virus (B-LCLs)

About 500 B-LCLs originated from South America

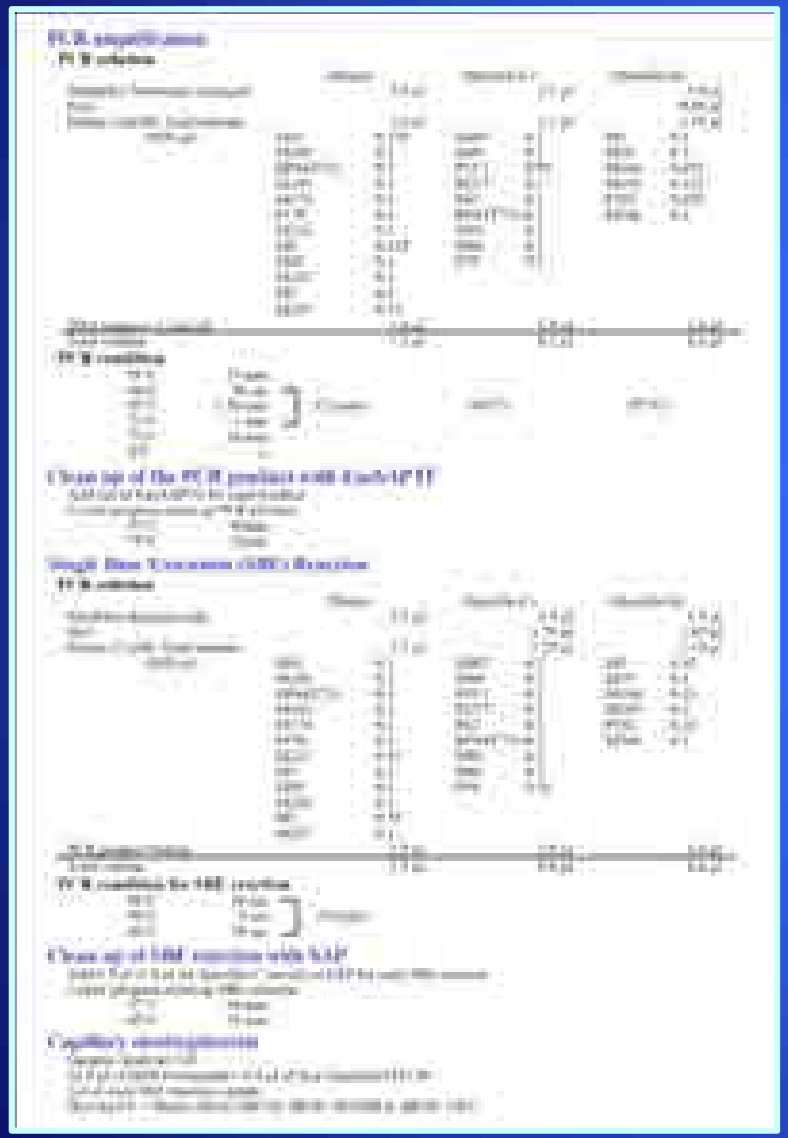
DNA extracted from first cultivated B-LCLs

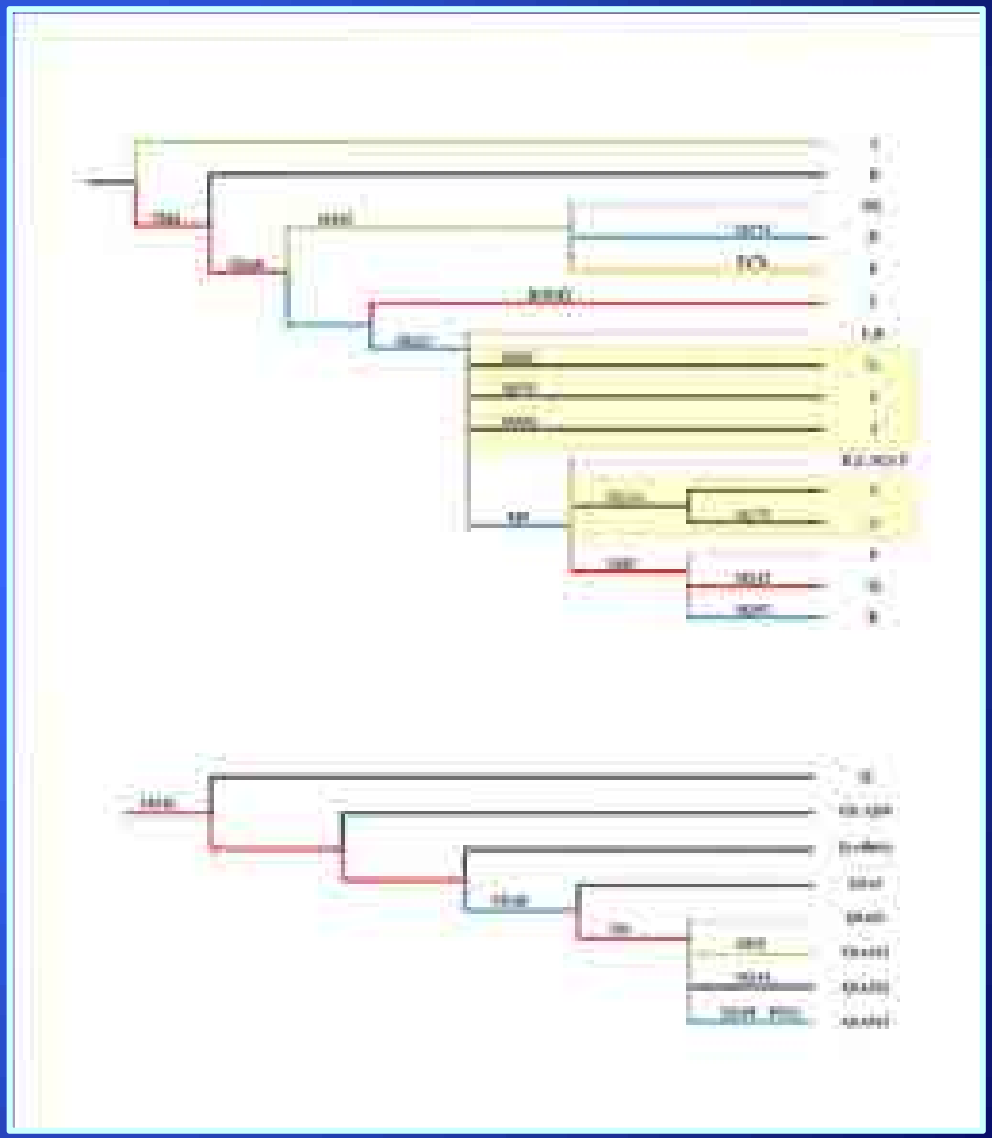
DNA samples

Tribe No.	Nation	Tribe Name	Sample (n)
1	Columbia	Wayuu	3
2		Guahibo	9
3		Ticuna	6
4		Matapi	1
5		Cumbal	4
6		Inga	4
7		Kamsa	3
8	Venezuela	Sanuma	14
9		Ye'Kuana	6
10		Piaroa	12
11	Ecuador	Canar	7
12		Saraguro	7
13	Peru	Aymara	8
14	Bolivia	Aymara	15
15		Chipaya	5
16		Quechua	15
17	Argentina	Puna	7
18	Paraguay	Chaco-Lengua	19
19		Chaco-Nivacle	8
20		Chaco-Sanapana	1
21	Chile	Huilliche	5
22		Mapuche	26
23		Atacama	14
	Missing etc.		5
	Total		204

Y-haplogrouping

- Experimental procedure & Flow chart of haplogrouping) -





Gepper M, et al. FSI genet. 5 (2011) 100-4

Y-haplogrouping

- Primers for multiplex PCR (Major & Specific-Q & -C) -

SNP	Branch	Forward Primer (5´- 3´)	bp	Reverse Primer (5´- 3´)	bp	Template (bp)	SNP	conc (µM)
Major								
M42	B-T	GAGGGAGATAACTTGTGTCAG	21	GCAAGTTAAGTCACCAGCTC	20	92	A/T	0.35
M168	C-T	TGTTTTGCAGAGAGCTTGGA	20	TGACTGTTCAGTTTTATTCCACAAA	25	150	C/T	0.28
M213	F-T	GGCCATATAAAAACGCAGCA	20	TGAATGGCAAATTGATTCCA	20	208	T/C	0.28
RPS4Y711	C	GATTTTGTTGGGTGGTGGTC	20	TGGCCAGCCTCTTATCTCTC	20	217	C/T	0.28
M145	DE*	GCATACTTGCCTCCACGACT	20	CCAGGAGCTCACAGTCACAA	20	173	G/A	0.28
M174	D	TTCTCCGTCACAGCAAAAATG	21	ATGCAAAAGGAGAAGGACAAGA	22	186	T/C	0.28
P170	E	CCTCCTGTGCCTCTTTCAGA	20	ACAGCAGCAAGCAGGTCTTT	20	243	G/A	0.28
M9	K-T	GCAGCATATAAAACTTTCAGG	21	AAAACCTAACTTTGCTCAAGC	21	340	C/G	0.35
M45	P*	GAGAGAGGATATCAAAAATTGG	22	TAGCTTACAACACAAGGATTC	21	229	G/A	0.28
M242	Q	TACGGCATAGAAAGTTTGTG	20	GAACAACTCTGAAGCGGTGG	20	133	C/T	0.28
M3	Q	AGGGCATCTTTCATTTTAGG	20	GTGGATTTGCTTTGTAGTAGG	21	156	G/A	0.28
M207	R	CTATGGGGCAAATGTAAGTC	20	TGAAGGAAAAGTGGAGTCTG	20	129	A/G	0.42
Specific-C								
M407	C3d	TACTGAAAGTTGGGGACAGTC	21	GTGATAATCGCTTGTCTCTTG	21	113	A/G	0.28
M48	C3c	TCCCTTCCACTCTTAGCTTGA	21	CAATGTAAATGTTAGTATAAGGATG	25	123	A/G	0.56
P53.1	C3e	AGATGTCACCTTCCGTCTATG	21	TTACACTATGAACCAATCCCAC	22	155	T/C	0.14
M217	C3*	CTCCAAAATCCTCTCGTACAG	21	TGCTGTGGCTTTCATCAAAATA	22	168	A/C	0.28
P62	C3f	TTGCCCTTCTTCAGAACTCC	20	TAACAGTCCCCACCAGGAAG	20	194	-(T)/G	0.28
RPS4Y711	C	GATTTTGTTGGGTGGTGGTC	20	TGGCCAGCCTCTTATCTCTC	20	217	C/T	0.56
M93	C3a	GAGGCAGGAGAATCACTTCAA	21	CTGGCTGAAAAAGATAATGGTG	22	244	C/T	0.28
M86	C3c	ATTTGCTACATACATCTAAGGTC	23	GCCACATTCCACGGGGTTC	19	250	T/G	0.28
P39	C3b	AAACCCTGTCTCTATGAAATAC	22	TTCTGTGAATTACAACCAGGCT	22	263	G/A	0.28
Specific-Q								
M3	Q	AGGGCATCTTTCATTTTAGG	20	GTGGATTTGCTTTGTAGTAGG	21	156	G/A	0.29
M19	Q1a3a1	GTTGCTGGTTGTTACGGGGT	20	CCACAAACTGATGTAGAGAC	20	115	T/A	0.29
P292	Q1a3a3	ATTCCAGTCATGATGAGGTGG	21	ATGCCGTTTGCCTAGGTTAG	20	134	-(T)/G	0.22
M194	Q1a3a2	GCCTGGATGAGGAAGTGAG	19	ATACAGTCGTTGCCTTCTCG	20	128	T/C	0.22
M199	Q1a3a3	CCTGGTTGGATTCTGGTCTT	20	TGATTTCAAGGATTTGTTAGTCTT	24	196	-(C)/G	0.37
M346	Q1a	CAGCCAAGAGGACAGTAAG	19	TATGTAGGAGGATATTCTTCCA	22	89	G/C	0.29

Gepper M, et al. FSI genet. 5 (2011) 100-4

Y-haplogrouping

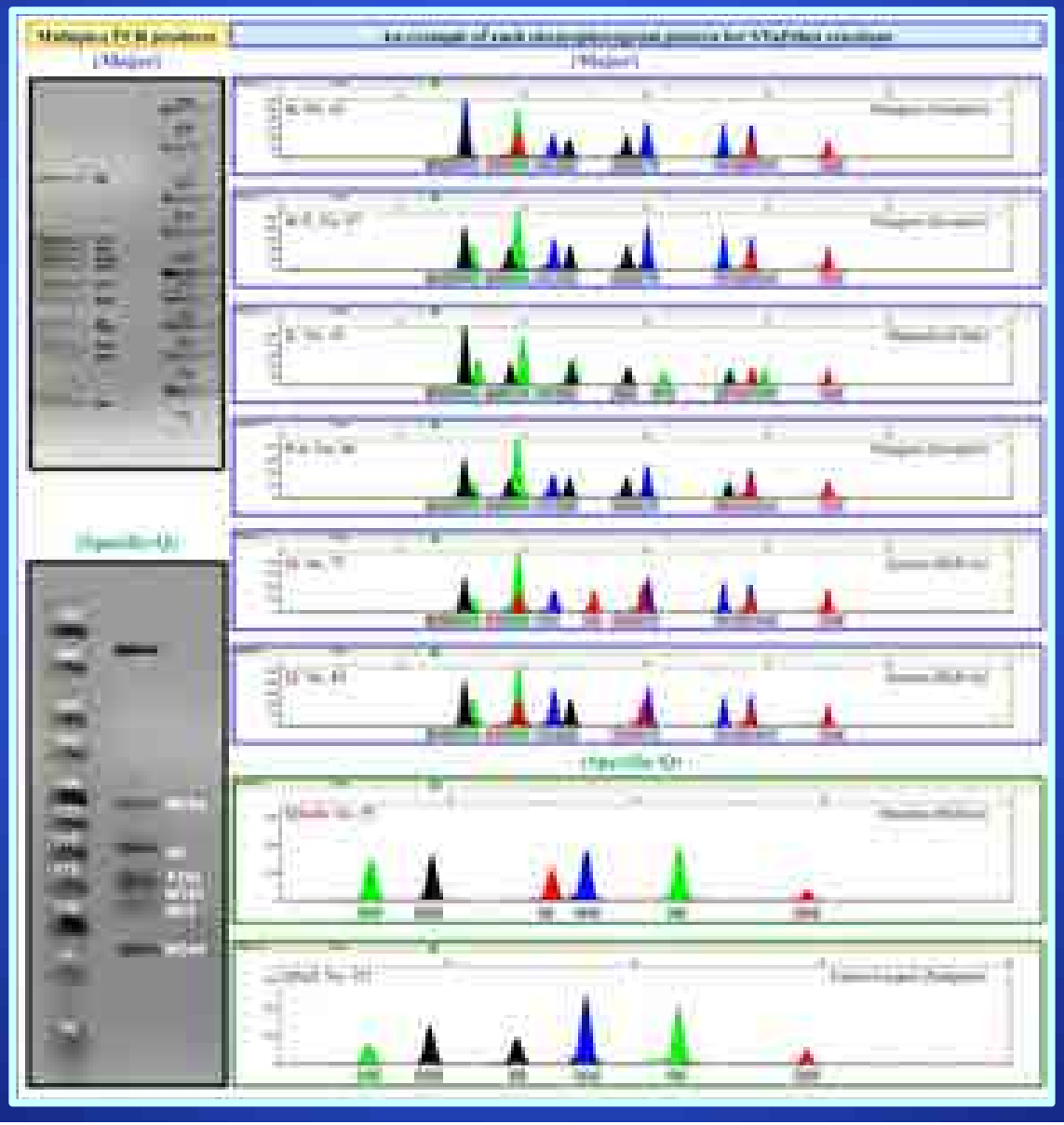
- Primers for multiplex SNaP shot reaction (Major & Specific-Q & -C) -

SNP	Primer sequence (5' - 3') target specific sequence black, neutral sequence blue	Size (bp)	Orientation Typing	conc (µM)	In analysis: ancestral/derived
Major					
RPS4Y711	GGCAATAAACCTTGGATTTC	20	Forward	0.28	black/red
M207	ATGTAAGTCAAGCAAGAAATTTA	23	Forward	0.09	blue/green
M45	TCTGACAACAGAAGGAGCTTTTTGC	25	Reverse	0.28	black/red
M174	CTGACAAGCACCCCTCACTTCTGCACT	27	Reverse	0.09	green/blue
M3	GTGAAAGTCTGACAACACCTCTGGGACTGA	30	Reverse	0.33	black/red
M213	AAGTCTGACAAGAACTTAAAACATCTCGTTAC	32	Reverse	0.14	green/blue
M242	ACGTCGTGAAAGTCTGACAAGGTGACCAAGGTGCT	35	Forward	0.09	black/red
P170	CGTCGTGAAAGTCTGACAATTTCCTTTGGCAAACTGA	37	Forward	0.28	blue/green
M9	GGTGCCACGTCGTGAAAGTCTGACAAGCCTAAGATGGTTGAAT	43	Forward	0.28	black/blue
M42	TGCCACGTCGTGAAAGTCTGACAATCAGATTTAGGACACAAAAGC	45	Forward	0.09	green/red
M145	CTAGGTGCCACGTCGTGAAAGTCTGACAAGACACCAGAAAGAAAGGC	47	Forward	0.09	blue/green
M168	AACTAGGTGCCACGTCGTGAAAGTCTGACAAGTATGTGTTGGAGGTGAGT	50	Forward	0.09	black/red
Specific-C					
RPS4Y711	GGCAATAAACCTTGGATTTC	20	Forward	0.27	black/red
M48	CAATTAGGATTAAGAATATGAT	22	Forward	0.27	green/blue
P53.1	ACAAGCAGAATCTGAACATATCTCC	25	Forward	0.09	red/black
M93	CTGACAAAAAAGCTTGGTGTGACTTGG	27	Reverse	0.18	blue/green
M217	AGTCTGACAAGTATTTTTCCTTCTGAAGAGTT	32	Reverse	0.18	red/blue
M407	CGTGAAAGTCTGACAACTAAATCAACTTCTCCTTTGG	37	Reverse	0.18	red/black
P39	GCCACGTCGTGAAAGTCTGACAACGGGAGGTGGAGGTTAT	40	Forward	0.14	blue/green
M86	GTGCCACGTCGTGAAAGTCTGACAACAAGTGGTTAACACACAAGC	45	Forward	0.18	red/blue
P62	TAGGTGCCACGTCGTGAAAGTCTGACAATACAGCACTTCTCATGGAAAGT	50	Forward	0.18	red/blue
Specific-Q					
P292	ATGAGAAATTGCTGTACTTA	20	Reverse	0.14	green/black
M199	GAAATGTTAAAATGGCTTACACTTG	25	Forward	0.18	black/blue
M3	GTGAAAGTCTGACAACACCTCTGGGACTGA	30	Reverse	0.32	black/red
M346	CGTGAAAGTCTGACAAAGCCAAGAGGACAGTAAGA	35	Forward		blue/black
M19	GCCACGTCGTGAAAGTCTGACAAAGACATCTGAAACCCAC	40	Reverse	0.09	green/red
M194	GCCACGTCGTGAAAGTCTGACAAAAACATACAGGGAGTGTTTTTT	45	Forward	0.14	red/black

Gepper M, et al. FSI genet. 5 (2011) 100-4

Y-haplogrouping

- Electropherogram patterns observed in this study (Major and Specific Q) -



Y-haplogrouping

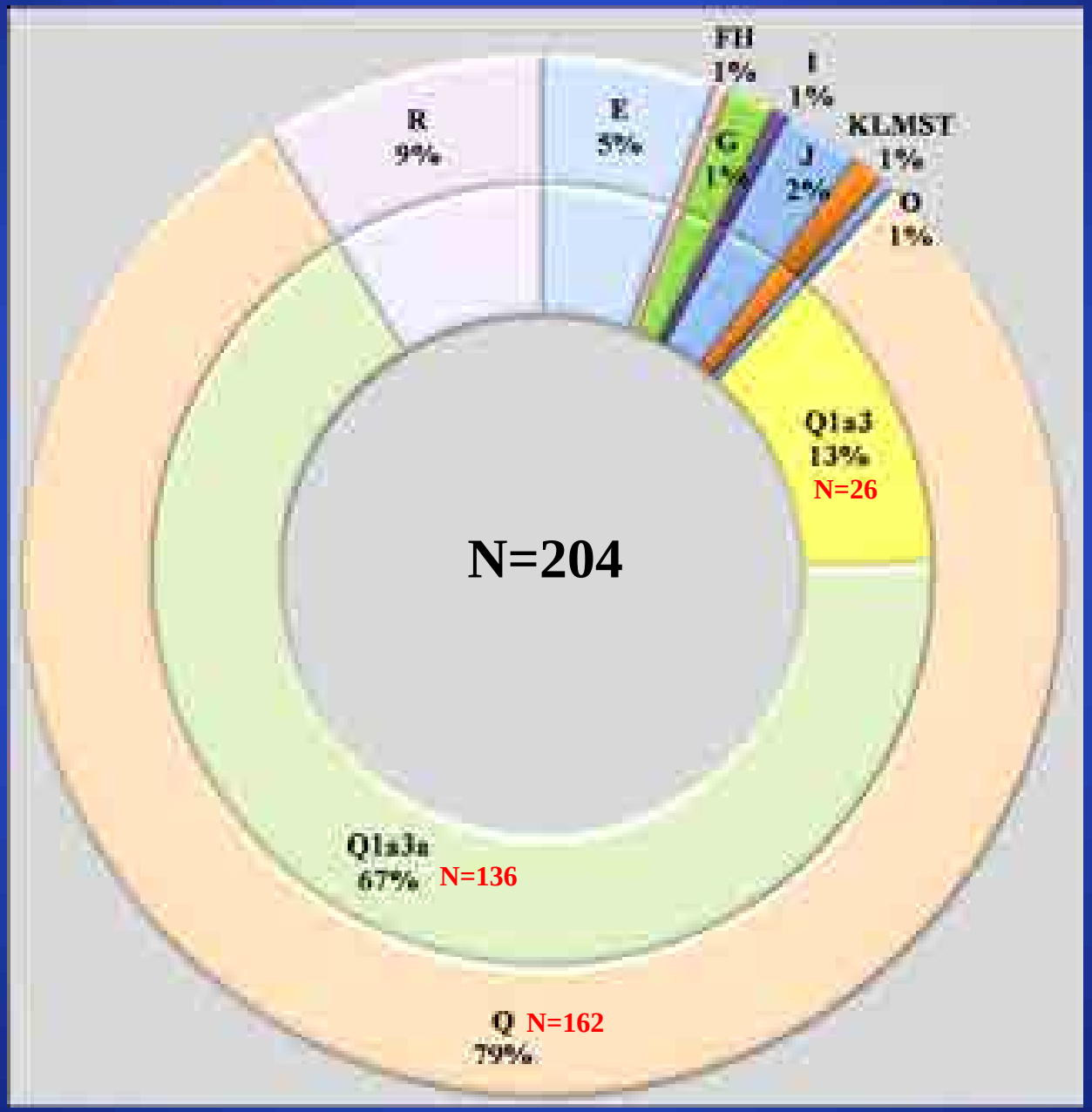
- Allele-specific PCR -

Y-HG	marker	SNP ID	variation	PCR amplification			
				amplicon length (bps)	Anneling temp.	Primer	
						allele-specific	common
C	M130	rs35284970	C→T	108	63	GGCAATAAACCTTGGATTTC GGCAATAAACCTTGGATTTCT	TGCAATTTAGCCACTGCTC
D	M174	rs2032602	T→C	124	66	AATACCTTCTGGAGTGCCCC AATACCTTCTGGAGTGCCCT	GAAGGTCCTGGAGATGCAAA
F to J G	M201	rs2032636	G→T	154	50	AATCCAGTATCAACTGAGGG AATCCAGTATCAACTGAGGT	CACTAAACATCATGGTGTGA
I	M170	rs2032597	A→C	74	52	CAACCCACACTGAAAAAAAG CAACCCACACTGAAAAAAAT	CATATTCTGTGCATTATACAAAT
J	M304	rs13447352	A→C	102	50	ATTTGAAAGTAACTTGTGAA ATTTGAAAGTAACTTGTGAC	TTTAAAATTACATCAGCTT
K to T O	M175	rs2032678	5-bp in/del	127 122	60	CACATGCCTTCTCACTTCTC CACATGCCTTCTCACTTCTC	TGTCCAATGCTGAAAGTAAG
N	M214	rs2032674	T→C	130	61	GACACTGTCTGAAAACAACG GACACTGTCTGAAAACAACA	AATATATGCCTGTAAAGCATC
Q	M242	rs8179021	C→T	216	66	ACGTTAAGACCAATGCCAAG ACGTTAAGACCAATGCCAAA	GTAATTGGCATCCCTTTAACT
R	M207	rs2032658	A→G	159	56	AAGTCAAGCAAGAAATTTAA AAGTCAAGCAAGAAATTTAG	TTTTTATTTCTAGGCTGTTC
C1	M8	rs3899	G→T	129	60	TGAAAAAGTTGGGTGACACA TGAAAAAGTTGGGTGACACC	CCATTGGAATTTAAGTGGCTT
C3	M217	rs2032668	A→C	114	60	ATAACTTGGACTGGGTTCAG ATAACTTGGACTGGGTTCAT	TGAATGTACCTTAGACCATCC

PCR condition

95 °C	11 min	
95 °C	45 sec	35 cycles
xx °C	45 sec	
72 °C	60 sec	
72 °C	10 min	

Frequency of Y-haplogroups in all DNA samples



Frequency of Y-haplogroups in each ethnic groups

Wayuu
3

Guahibo
E
9

Ticuna
6
J

Cumbal
4

Inga
4

Matapi

Kamsa
3

Canar
R
7
J

Saraguro
E
R
7
G

Aymara (Peru)
8

5

15
Q1a3a

Chipaya
R
E
15
J
Q1a3a

Aymara (Bol)

Puna
G
R
7
I

Quechua
Atacama
R
14
Q1a3a

5
Huilliche

Piaroa
12

Ye'Kuana
6

Sanuma
14
Q1a3

Chaco-Lengua
Q1a3
19
Q1a3a

Chaco-Nivacle
8

Chaco-Sanapana

Mapuche
R
E
26
Q1a3a

Missing
5

Loci included in anAmpFlSTR Yfiler kit

Locus Designation	Chromosome Location	Common Sequence Motif	Allele Range	Dye Label
DYS456	Yp11.2	$(AGAT)_n$	13-18	6-FAM
DYS389 I	Yq11.1	$(TCTG)_3(TCTA)_n$	10-15	6-FAM
DYS390	Yq11.221	$(TCTG)_8(TCTA)_n TCTG(TCTA)_4$	18-27	6-FAM
DYS389 II	Yq11.1	$(TCTG)_{4-5}(TCTA)_m \ldots (TCTG)_3(TCTA)_n$	24-34	6-FAM
DYS458	Yp11.2	$(GAAA)_n$	14-20	VIC
DYS19	Yp11.2	$(TAGA)_3 TAGG(TAGA)_n$	10-19	VIC
DYS385	Yq11.222	$(GAAA)_n$	7-25	VIC
DYS393	Yp11.31	$(AGAT)_n$	8-16	NED
DYS391	Yq11.1	$(TCTA)_n$	7-13	NED
DYS439	Yq11.1	$(GATA)_n$ or $(AGAT)_n$	8-15	NED
DYS635	Yq11.1	$(TCTA)_4(TGTA)_2(TCTA)_2(TGTA)_2 (TCTA)_2(TGTA)_2(TCTA)_n$	20-26	NED
DYS392	Yq11.222	$(TAT)_n$	7-18	NED
H4	Yq11.221	$(TAGA)_n$	8-13	PET
DYS437	Yq11.1	$(TCTA)_n(TCTG)_2(TCTA)_4$	13-17	PET
DYS438	Yq11.21	$(TTTTC)_n$	8-13	PET
DYS448	Yq11.223	$(AGAGAT)_n ATAGAGATAG(AGAGAT)_3 AGATAGATAGAGAA(AGAGAT)_{8-9}$	17-24	PET

YHRD. ORG.3.0

http://www.yhrd.org/

YHRD. ORG.3.0

http://www.yhrd.org/

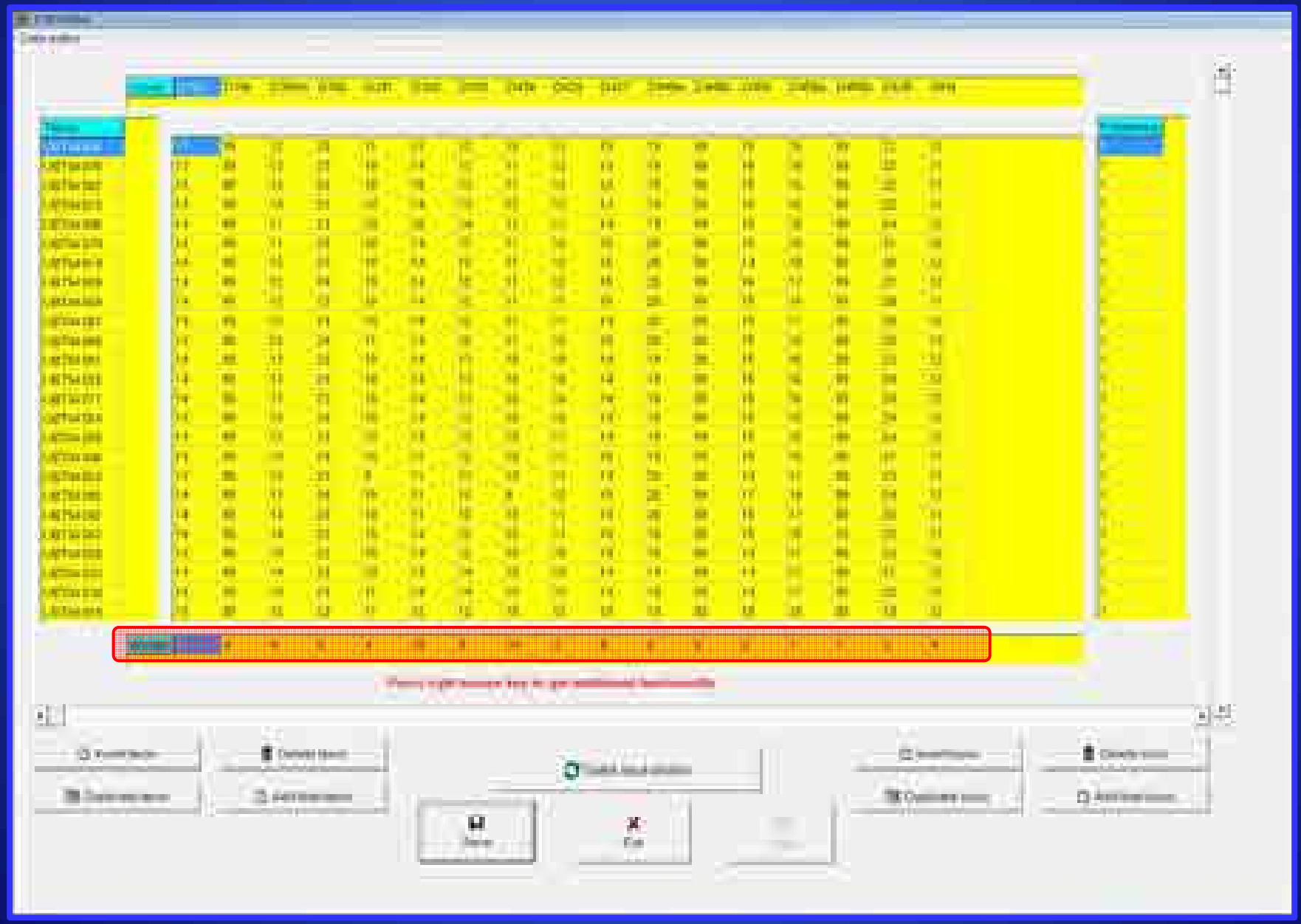
Mutation rates for DYS19 from 25 references.

Mutation Rates

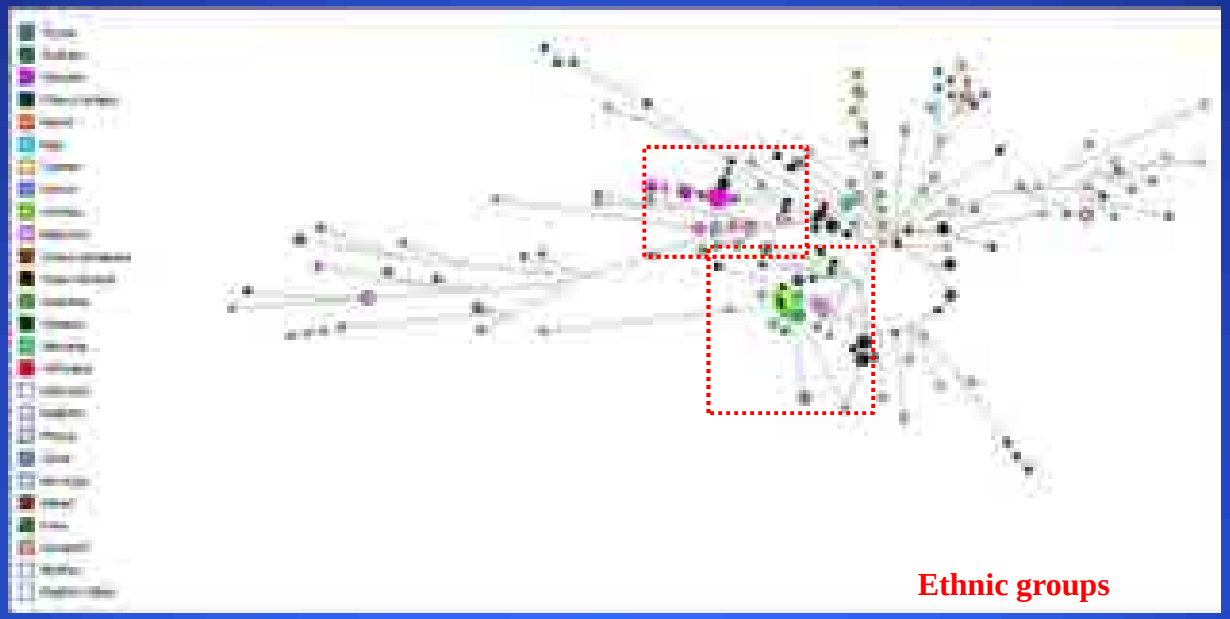
The weights for Median-joining Network analysis

From YHRD					Weight in median-joining network analysis
Locus	Mutations	Meioses	Mutation Rate		
DYS19	36	15539	2.3	$\times 10^{-3}$	4
DYS389I	37	13788	2.7	$\times 10^{-3}$	4
DYS389II	52	13759	3.8	$\times 10^{-3}$	3
DYS390	31	15061	2.1	$\times 10^{-3}$	5
DYS391	38	14935	2.5	$\times 10^{-3}$	4
DYS392	6	14867	0.4	$\times 10^{-3}$	25
DYS393	15	13713	1.1	$\times 10^{-3}$	9
DYS385	59	25620	2.3	$\times 10^{-3}$	4
DYS438	3	10122	0.3	$\times 10^{-3}$	34
DYS439	54	10096	5.3	$\times 10^{-3}$	2
DYS437	12	10101	1.2	$\times 10^{-3}$	8
DYS448	11	6678	1.6	$\times 10^{-3}$	6
DYS456	28	6678	4.2	$\times 10^{-3}$	2
DYS458	45	6677	6.7	$\times 10^{-3}$	1
DYS635	28	7525	3.7	$\times 10^{-3}$	3
Y-GATA-H4	19	7709	2.5	$\times 10^{-3}$	4
		Ave(14 loci)=	2.62	$\times 10^{-3}$	0.00262

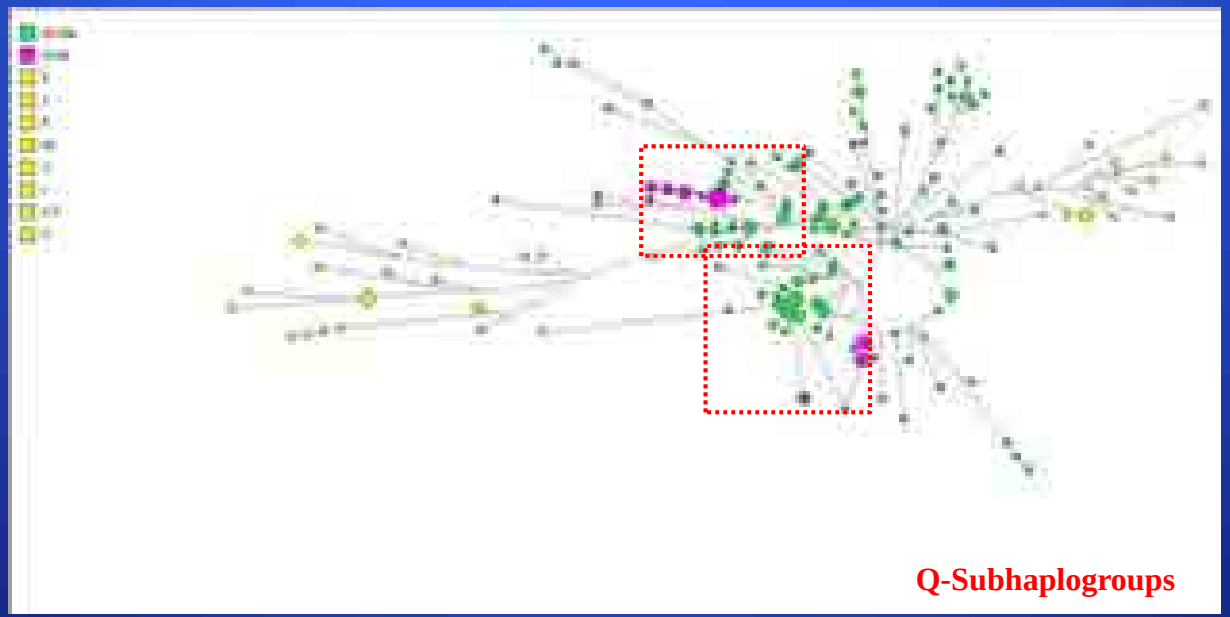
The weights for Median-joining Network analysis



Network analysis for 14 Y-STRs

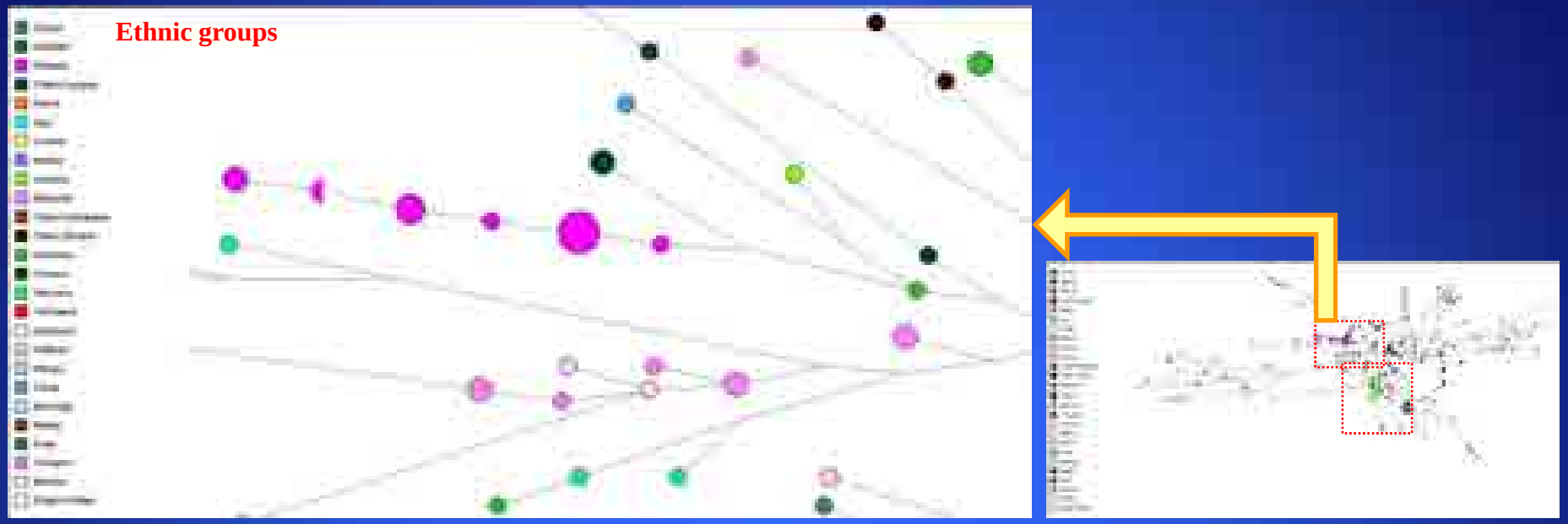


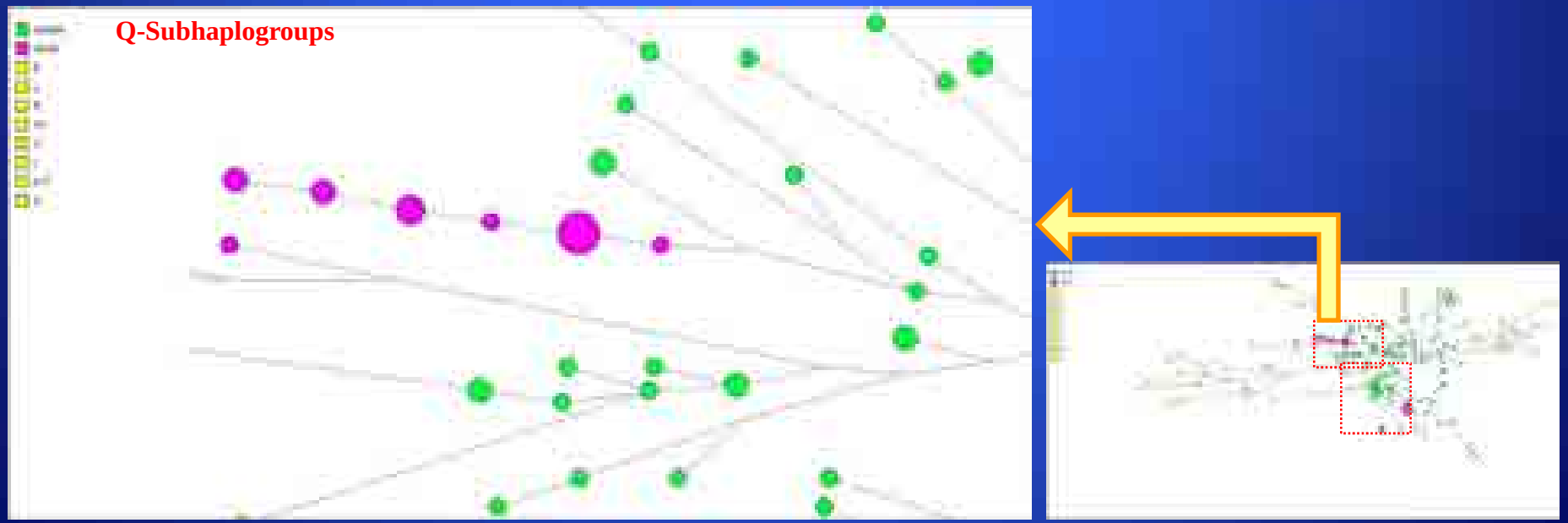
Ethnic groups



Q-Subhaplogroups

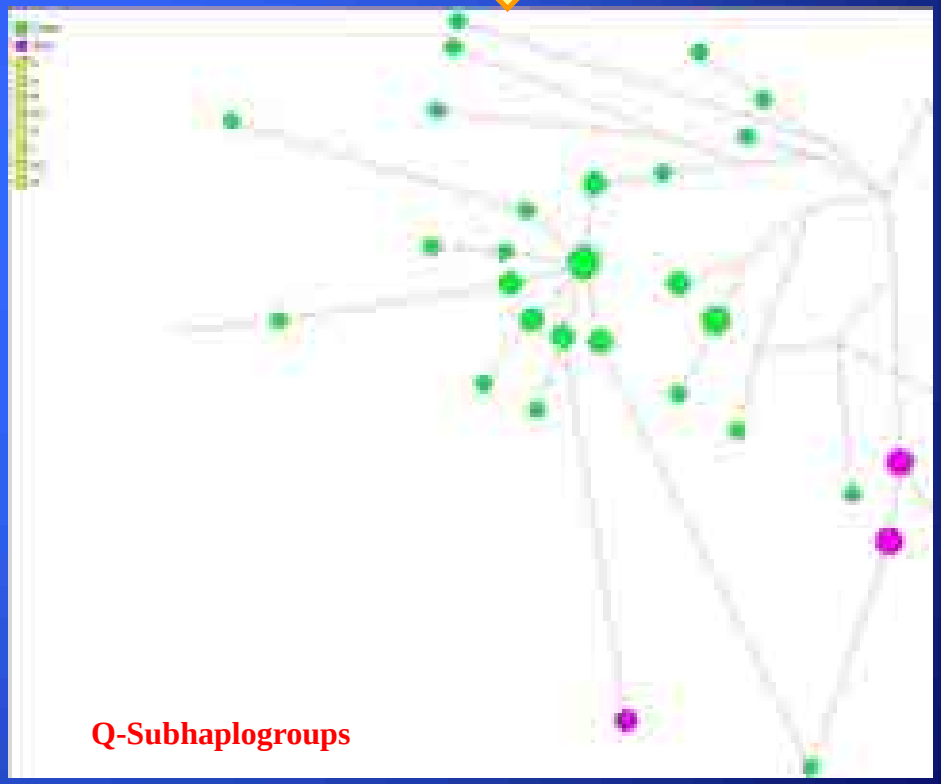
Network analysis for 14 Y-STRs





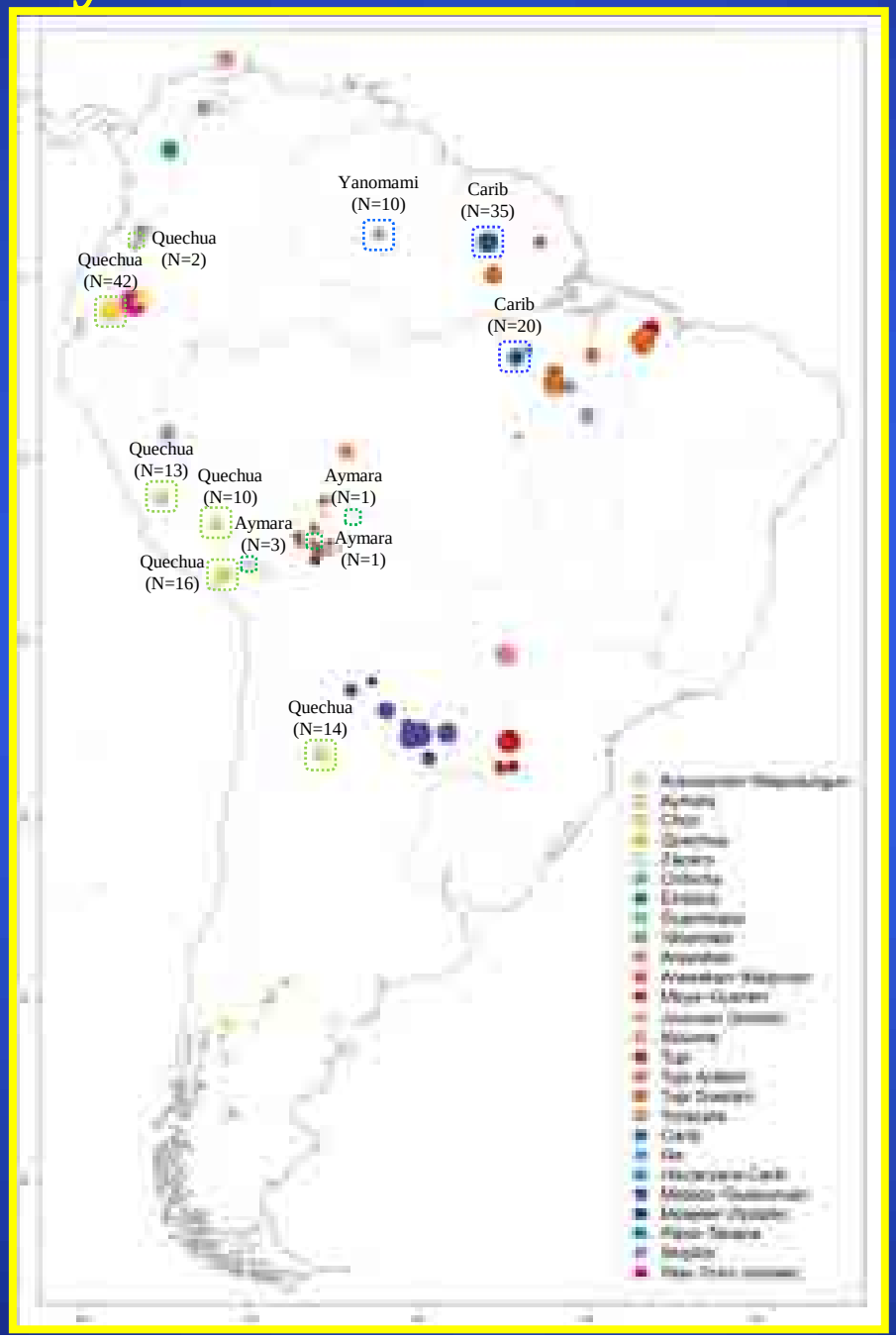
Network analysis for 14 Y-STRs

Ethnic groups



Q-Subhaplogroups

Network analysis for 14 Y-STRs with data published



Gyographically close to Sanuma

Yanomami : N=10

Carib : N=55

Same language cluster as Andean

Quechua : N=97

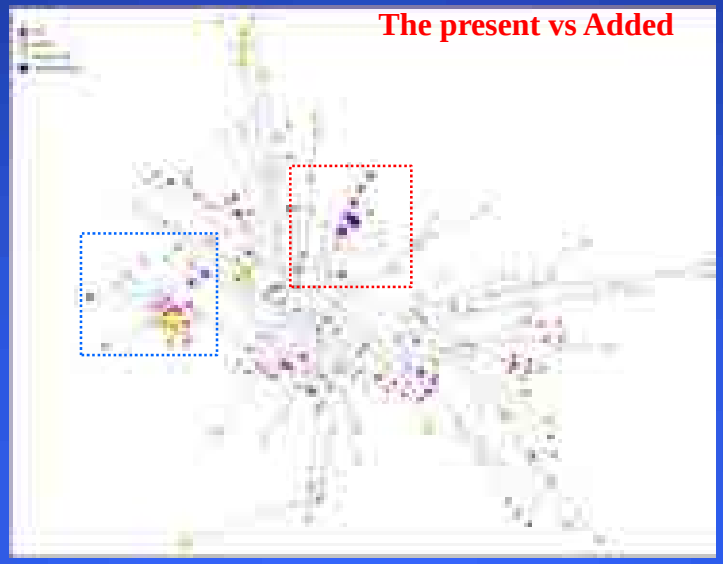
Aymara : N=5

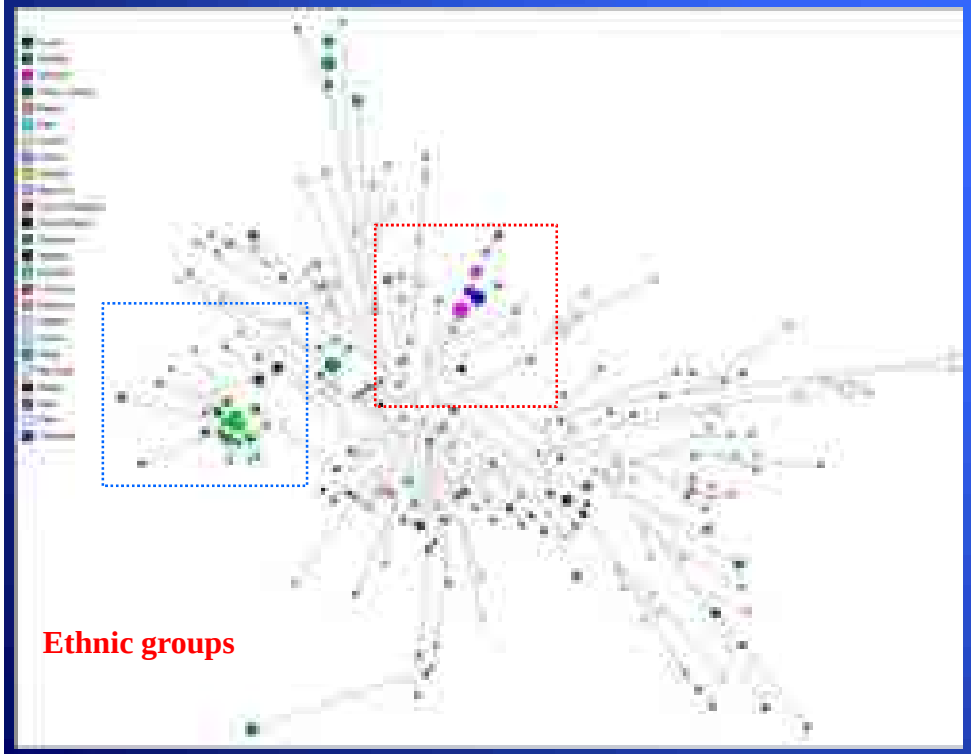
Total: 167
This study: 162
(Only Y-hg-Q)

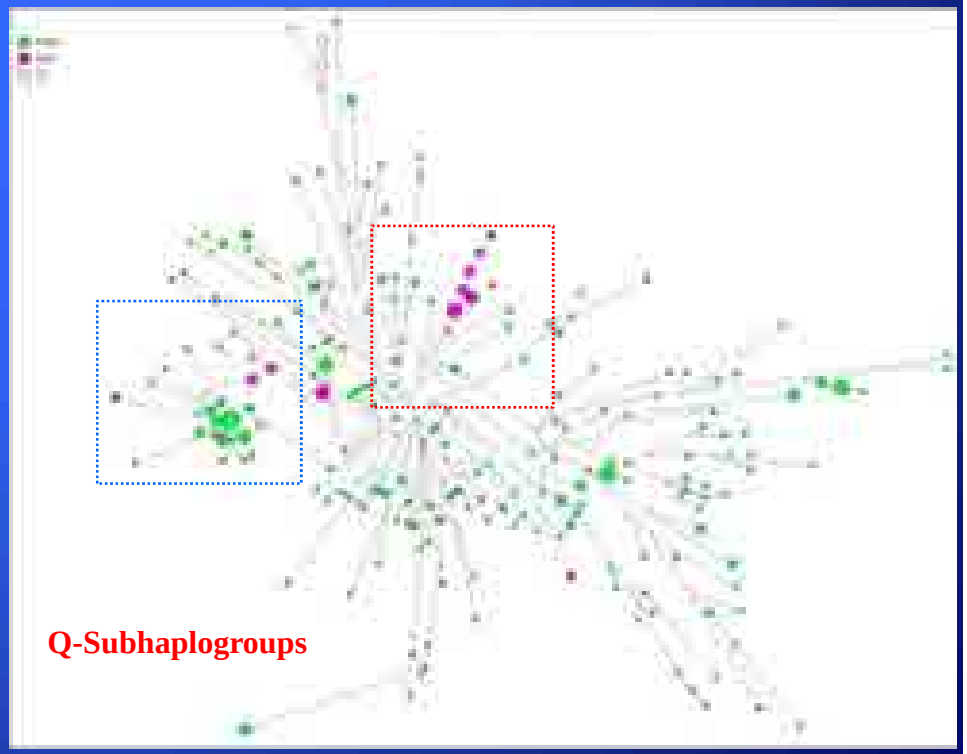
N=329

Roewer et al.
PLOS Genetics 9 (2013) e1003460

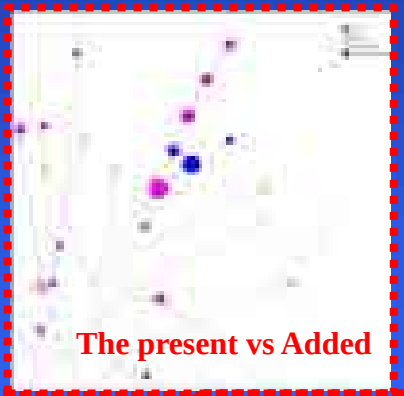
Network analysis for 14 Y-STRs with data published



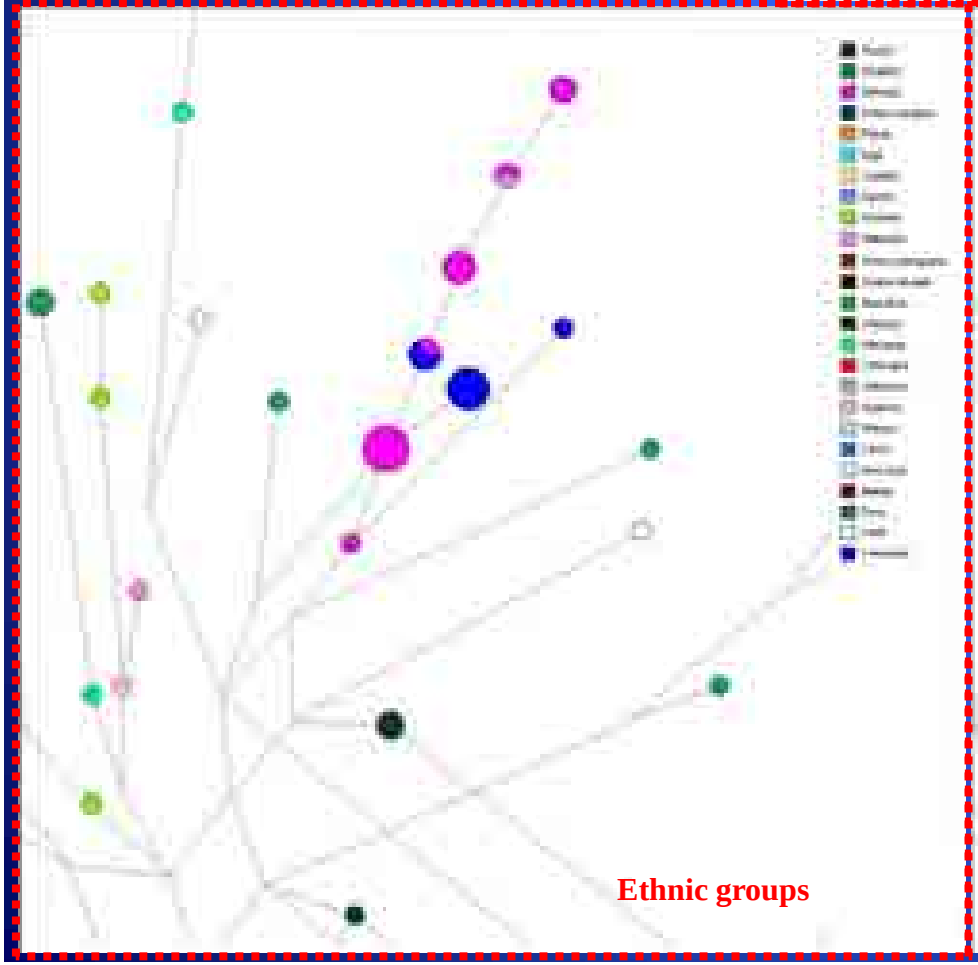




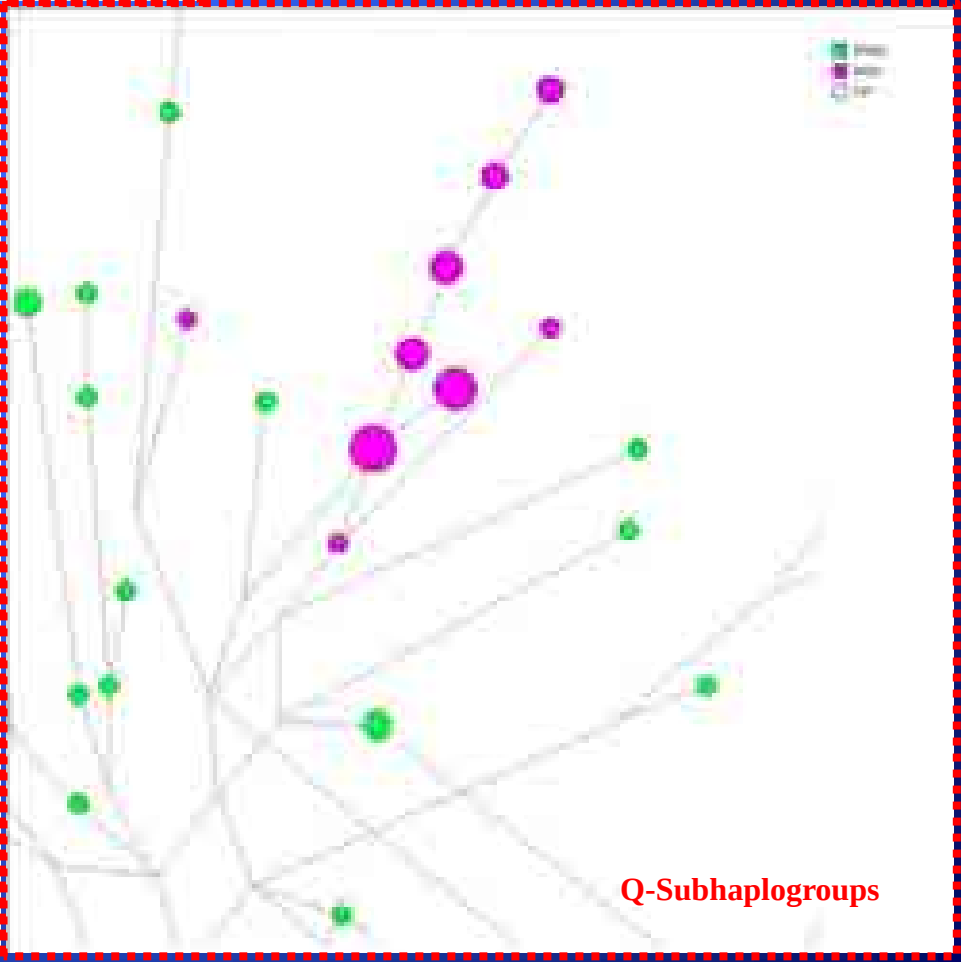
Network analysis for 14 Y-STRs with data published



The present vs Added

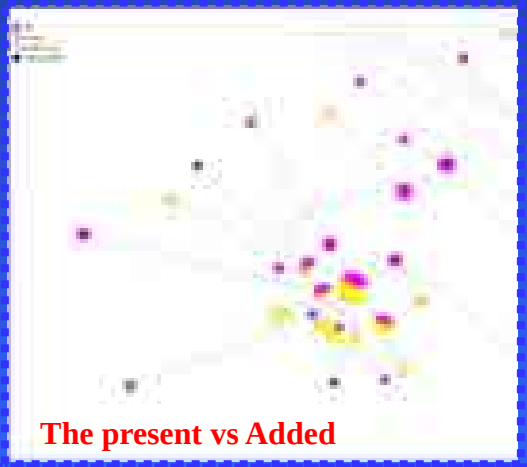


Ethnic groups

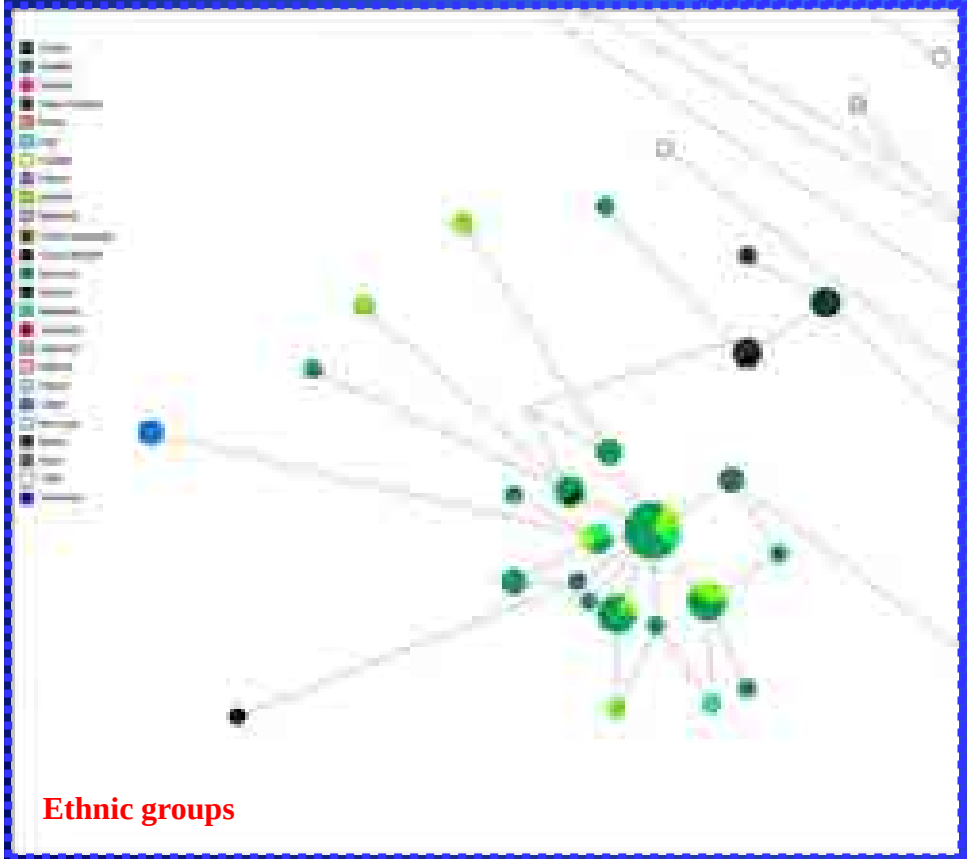


Q-Subhaplogroups

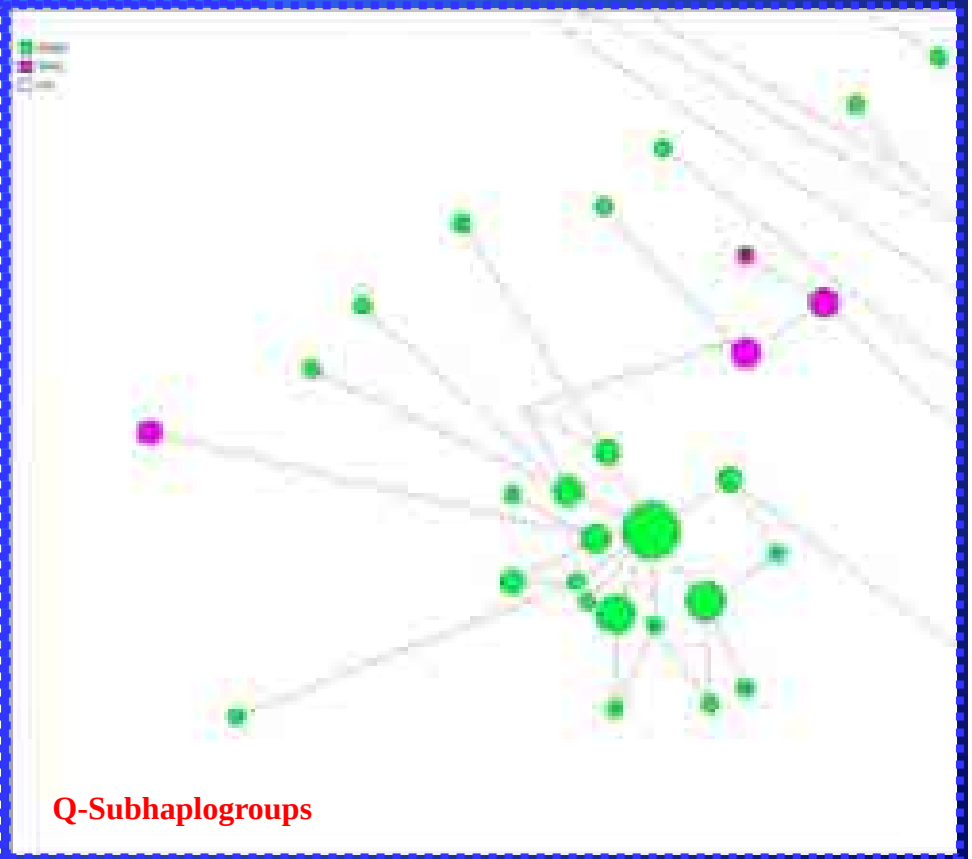
Network analysis for 14 Y-STRs with data published



The present vs Added



Ethnic groups

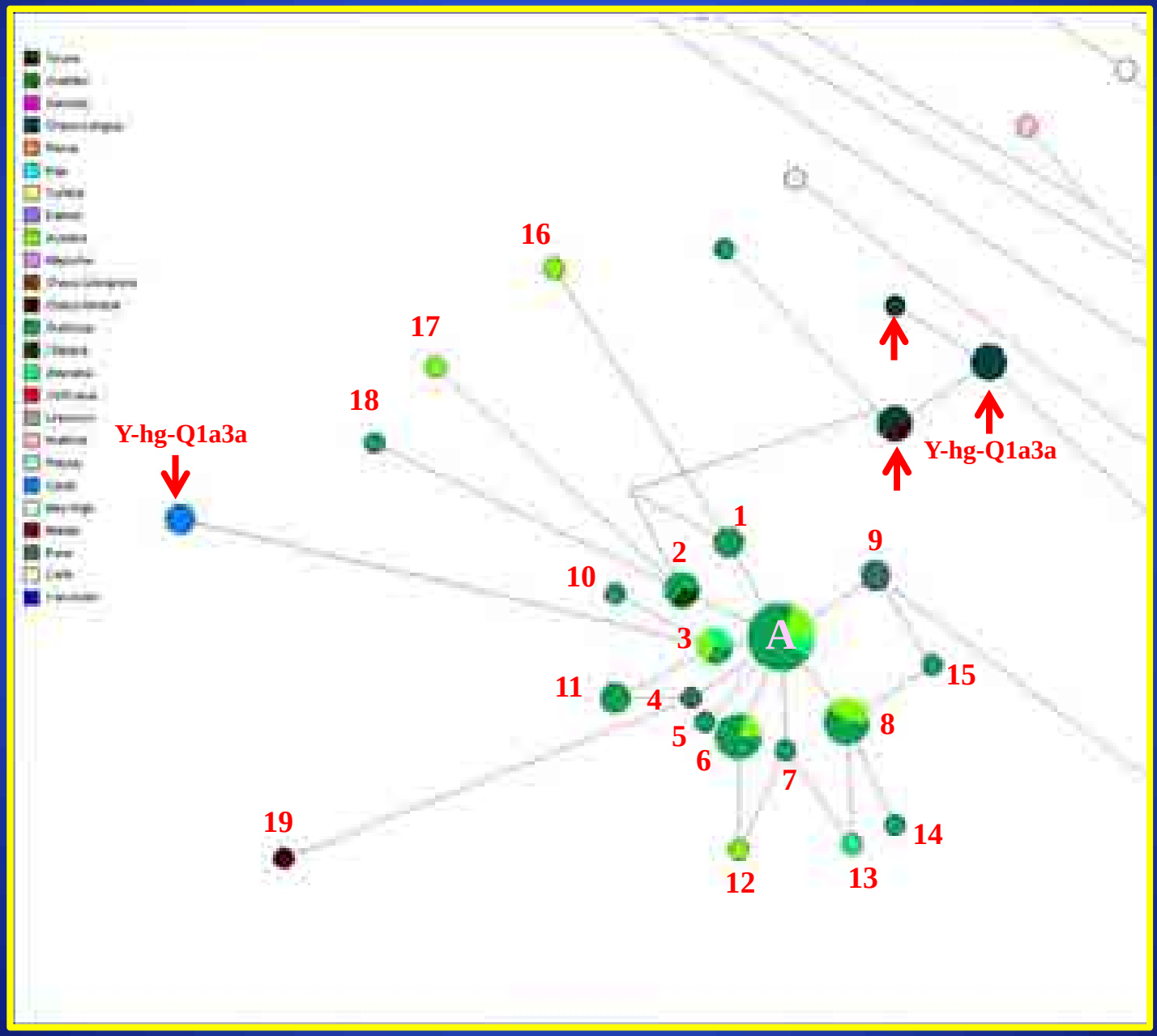


Q-Subhaplogroups

Calculation for average mutation for years

From YHRD					Weight in median-joining network analysis
Locus	Mutations	Meioses	Mutation Rate		
DYS19	36	15539	2.3	$\times 10^{-3}$	4
DYS389I	37	13788	2.7	$\times 10^{-3}$	4
DYS389II	52	13759	3.8	$\times 10^{-3}$	3
DYS390	31	15061	2.1	$\times 10^{-3}$	5
DYS391	38	14935	2.5	$\times 10^{-3}$	4
DYS392	6	14867	0.4	$\times 10^{-3}$	25
DYS393	15	13713	1.1	$\times 10^{-3}$	9
DYS385	59	25620	2.3	$\times 10^{-3}$	4
DYS438	3	10122	0.3	$\times 10^{-3}$	34
DYS439	54	10096	5.3	$\times 10^{-3}$	2
DYS437	12	10101	1.2	$\times 10^{-3}$	8
DYS448	11	6678	1.6	$\times 10^{-3}$	6
DYS456	28	6678	4.2	$\times 10^{-3}$	2
DYS458	45	6677	6.7	$\times 10^{-3}$	1
DYS635	28	7525	3.7	$\times 10^{-3}$	3
Y-GATA-H4	19	7709	2.5	$\times 10^{-3}$	4
		Ave(14 loci)=	2.62	$\times 10^{-3}$	0.00262
		A mutaiton for years	**20**	yrs/generat	**7630**
			25		**9537**

Time estimation



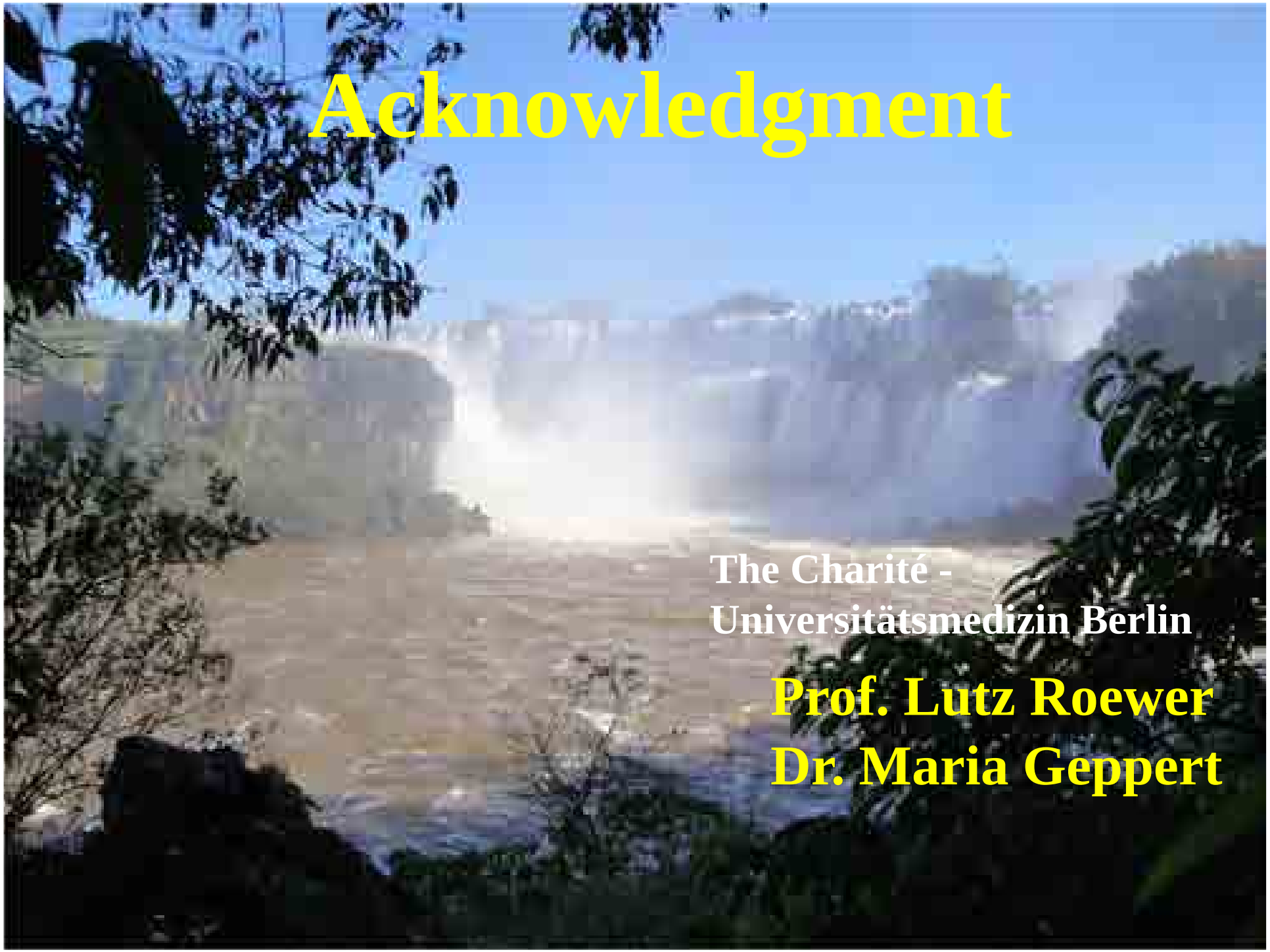
Time estimation

The time when a male-linage strongly influenced these tribes was estimated as 735 and 918 years ago, at 20 and 25 years per a generation, respectively.

Considering with the history of South America, this period (about 800 years ago) is almost corresponded to the period establishing the Cuzco Kingdom.

Summary

1. It is suggested that Sanuma tribe is very isolated, and very closed to Yanomami in male-linage on a Y-STR network analysis.
2. A “ Star-like ” cluster which consists of mainly Andean speaking such as Quechua and Aymara tribes was observed in this study.
3. As a result from the time estimation when a male-linage strongly influenced these tribes, it was suggested that rapid population growth started or a strong male (relatives) influenced these tribes about 800 years ago when are almost concordant with the period of the establishment of the Cuzco Kingdom before Inca Empire.



Acknowledgment

The Charité -
Universitätsmedizin Berlin

Prof. Lutz Roewer
Dr. Maria Geppert


Thank you very much for your attention!!