# Language and Genes of the Greater Himalayan Region

*Himalayan Languages Project*
George L. van Driem
Jean Robert M. L. Opgenort
Karma Tshering of Gaselô

*Tribhuvan University, Nepal*
Nirmal M. Tulandar

*Leiden University Medical Centre*
Thirsa Kraaijenbrink
Peter de Knijff

*The Wellcome Trust Sanger Institute*
Denise R. Carvalho-Silva
Chris Tyler-Smith

*University of Leicester*
Emma J. Parkin
Mark A. Jobling

*Università di Ferrara*
Guido Barbujani

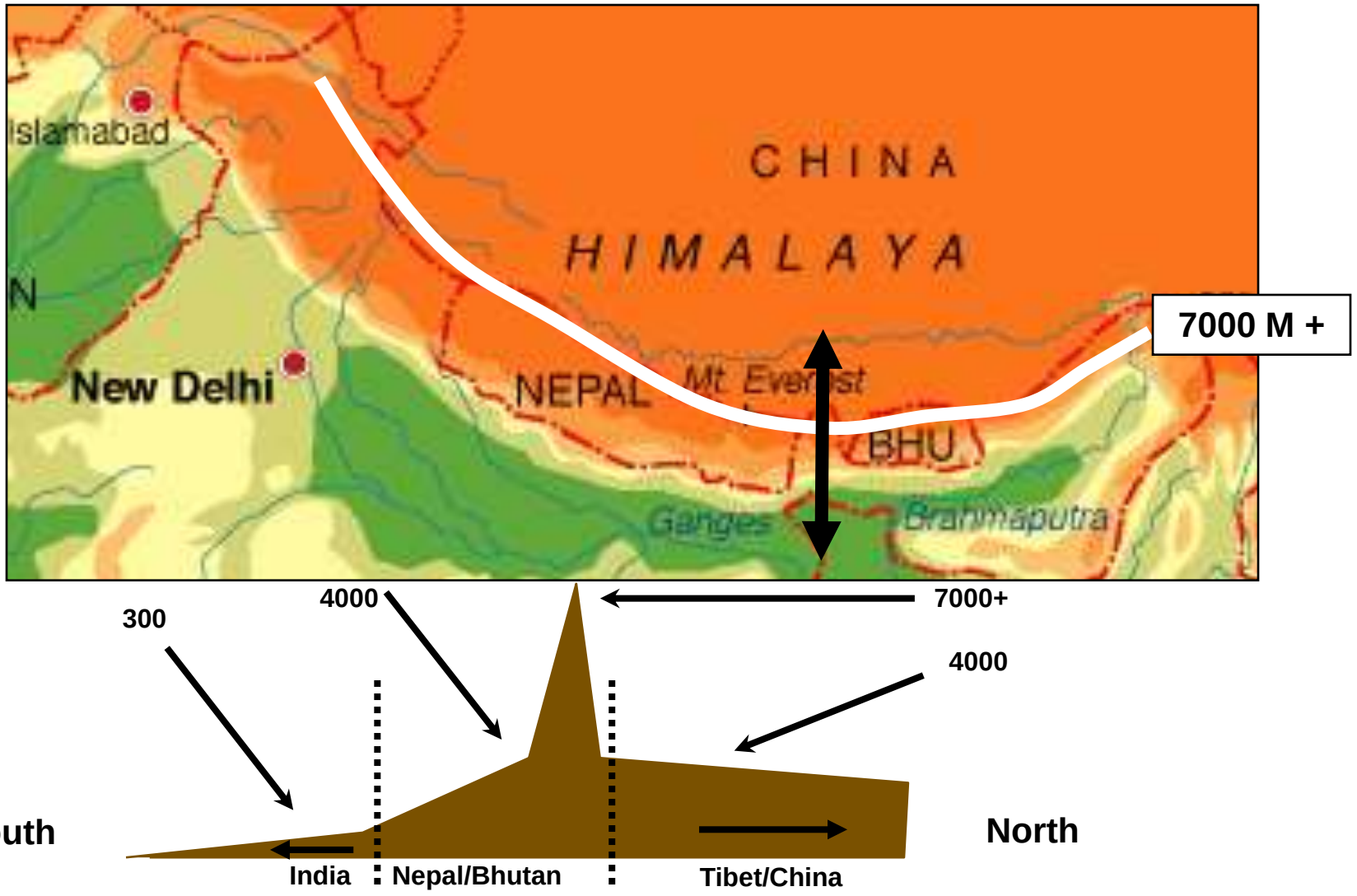

Arts & Humanities Research Council

# Outline

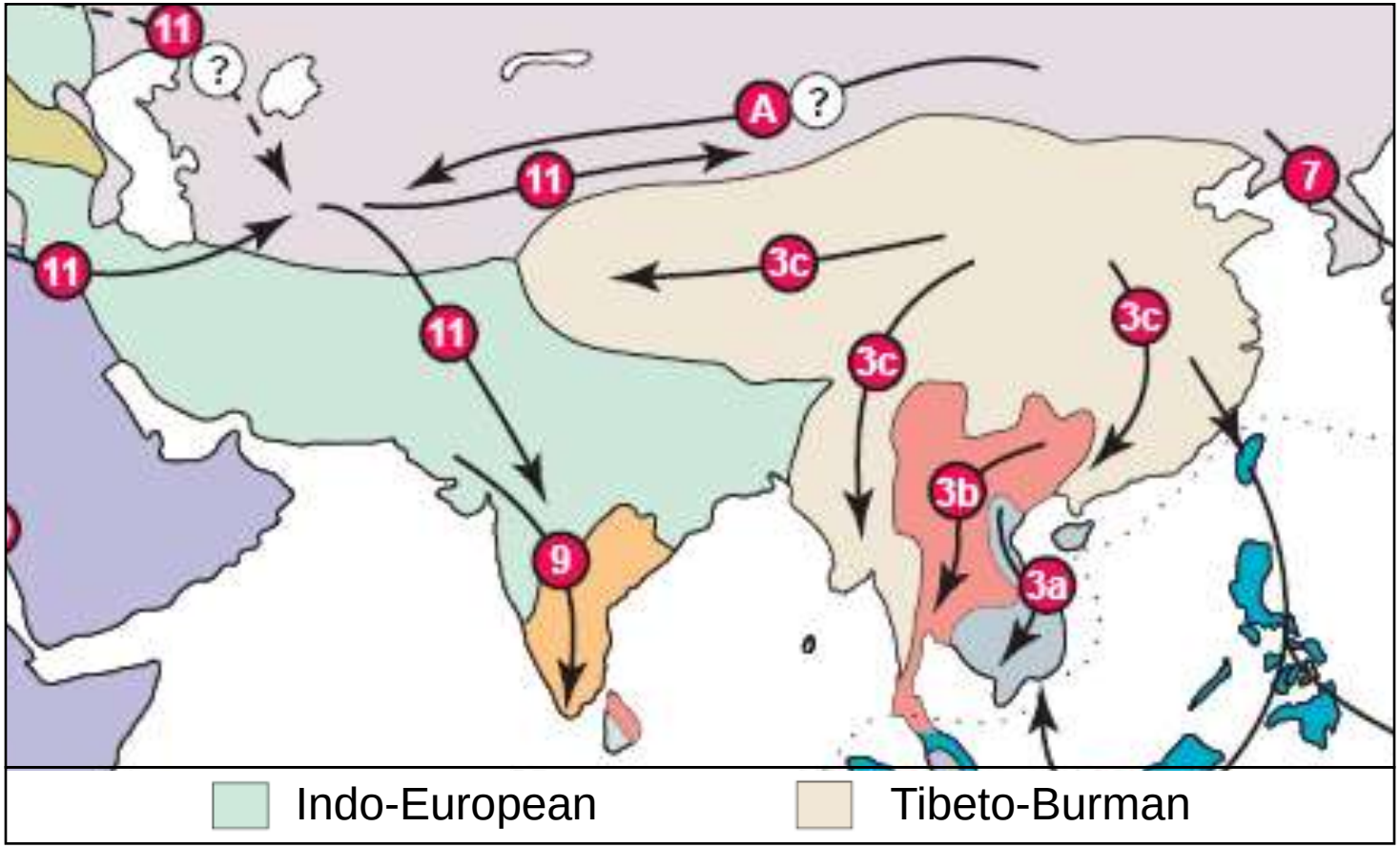
- Why the Himalayan region?
- The project
- Location of genetic boundaries
- The peopling of the area
- Patterning of diversity within Bhutan

# The geographical barrier is obvious

# The Himalaya

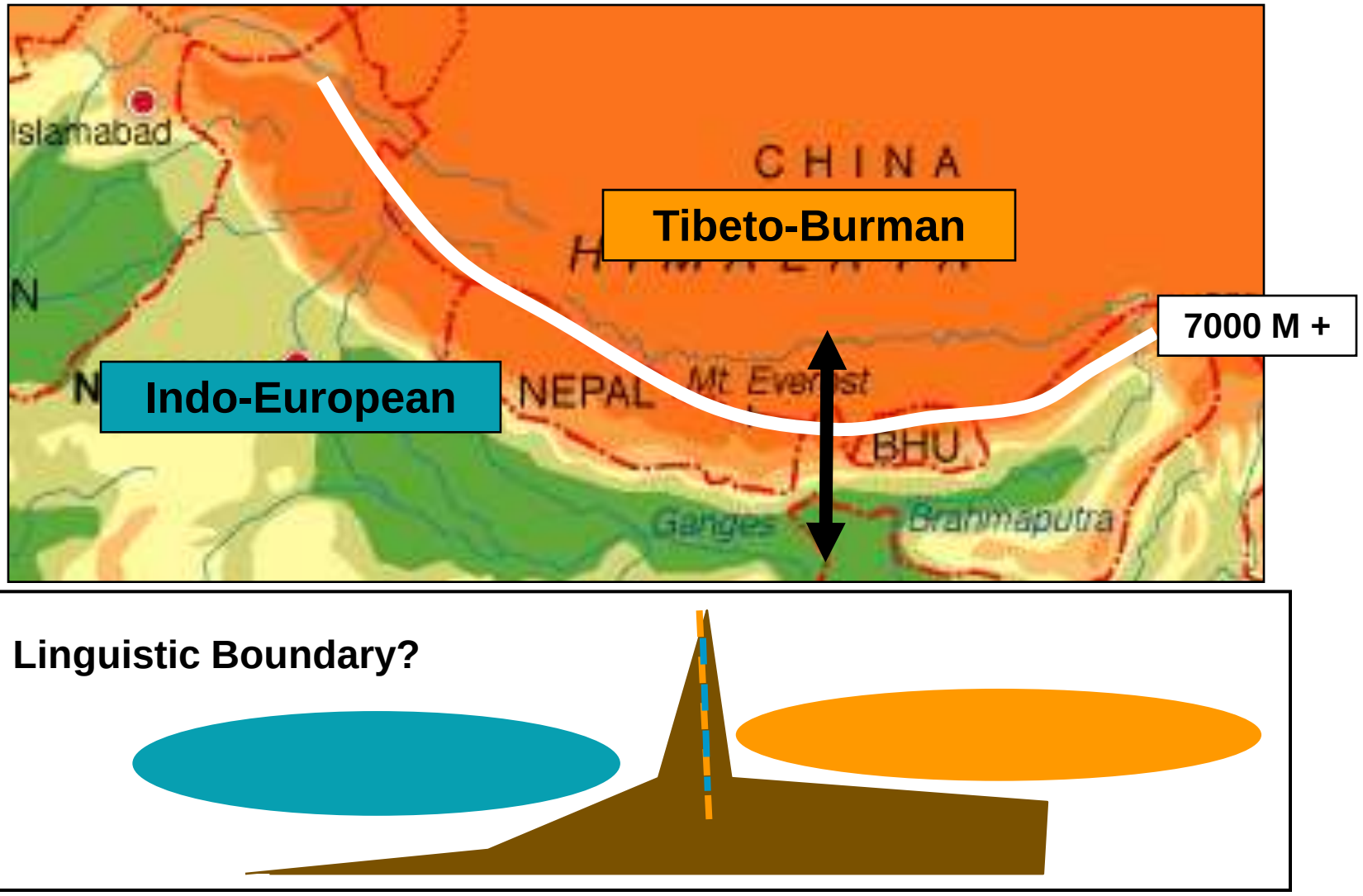


# There is also a linguistic boundary …

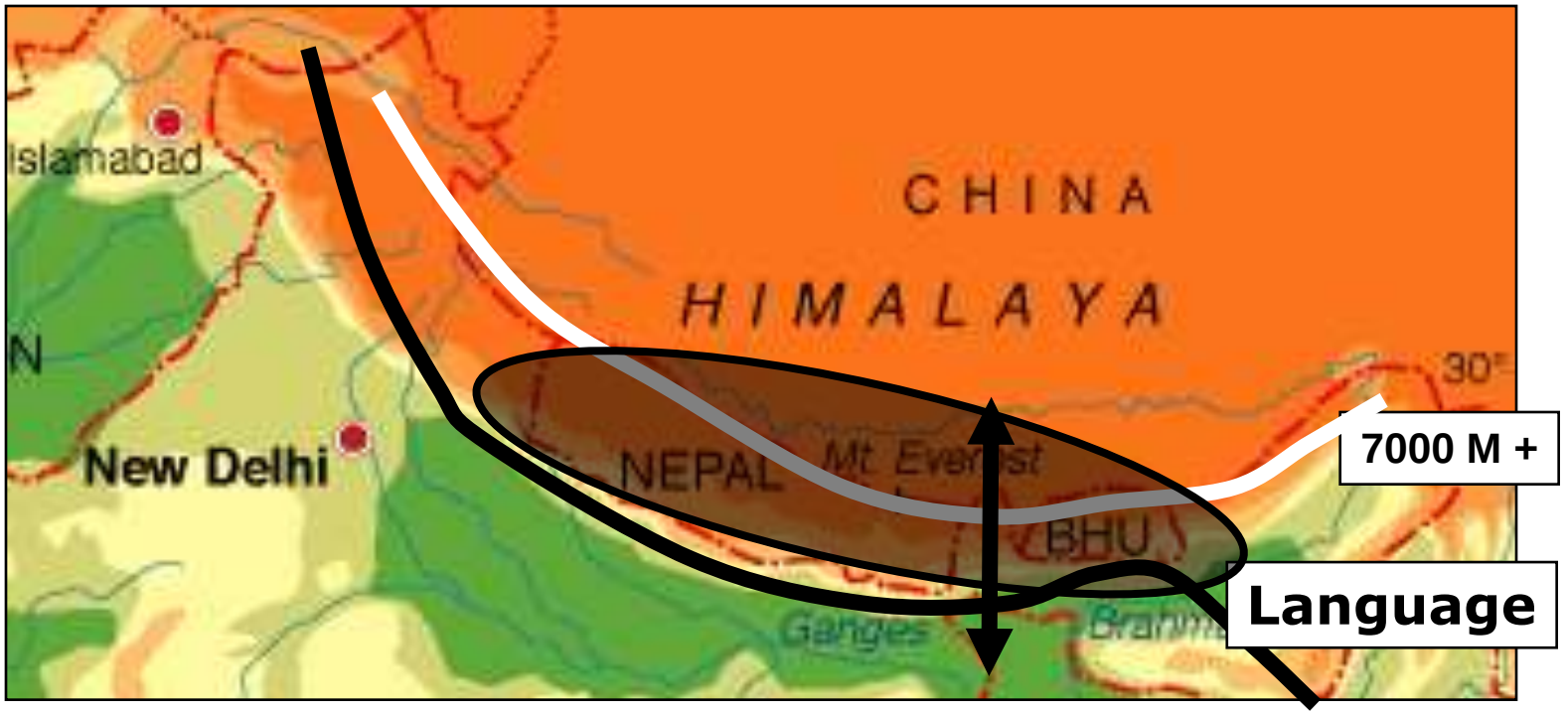


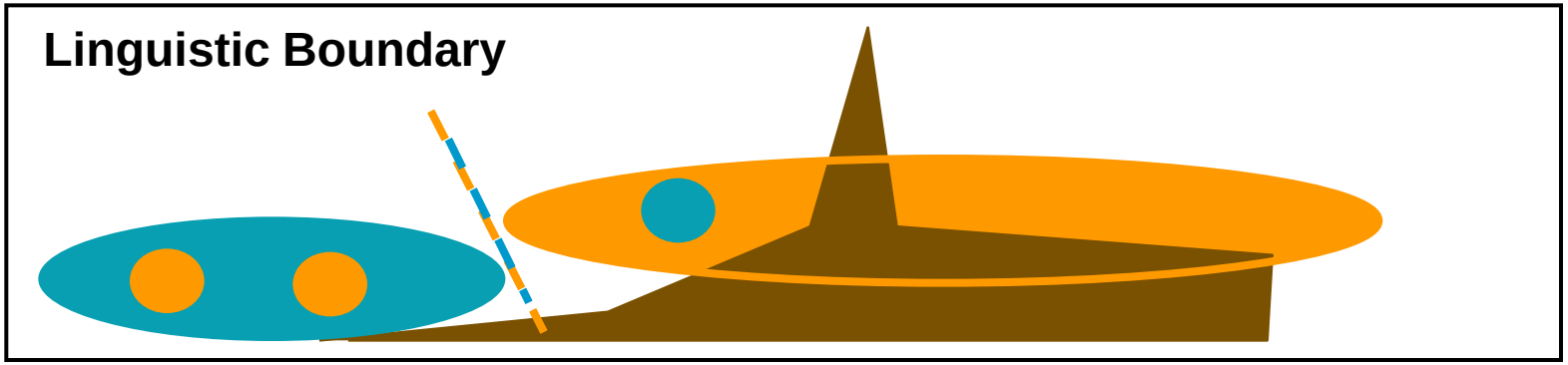

Diamond and Bellwood (2003) *Science* **300**, 597-603

# Where exactly is it?



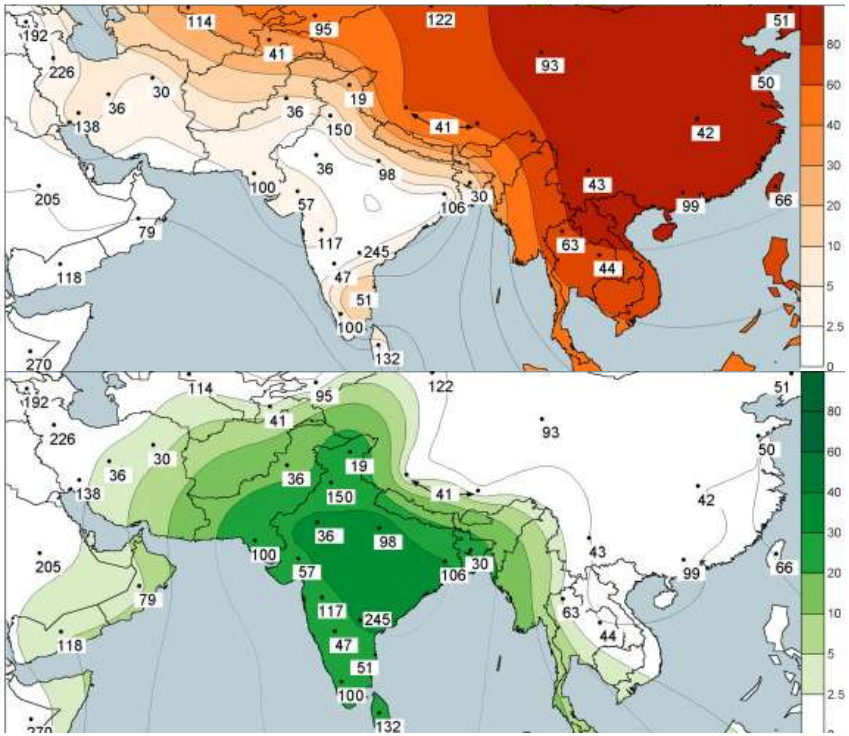
# Actually like this…



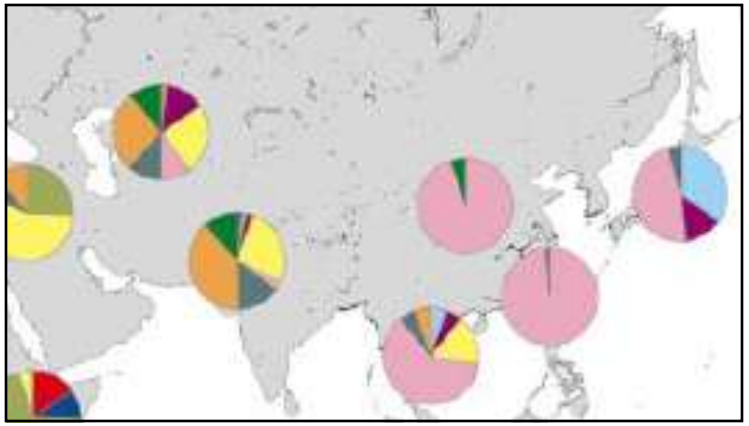
# Previous evidence for a genetic boundary

mtDNA



Metspalu *et al*. (2004) *BMC Genetics* **5**, 26

Y chromosome



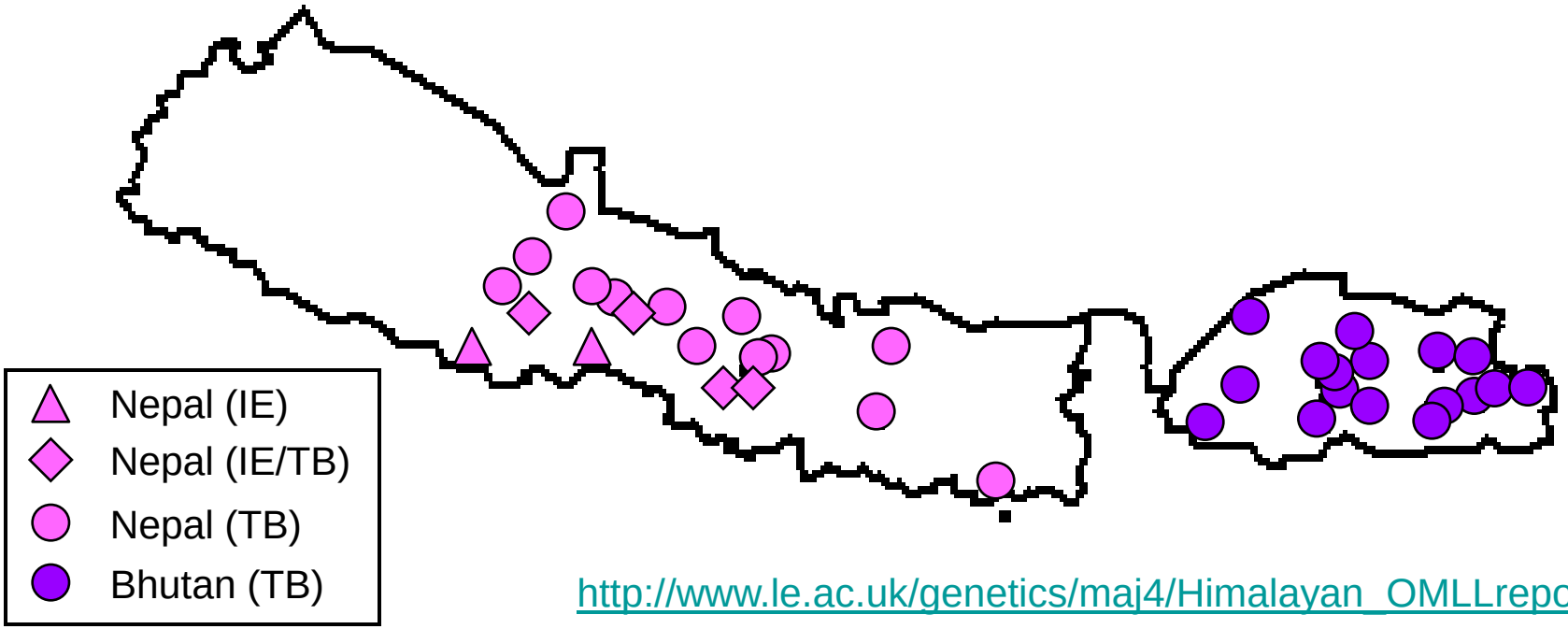
Underhill *et al*. (2001) *Ann. Hum. Genet*. **65**, 43-62

# Questions

- How precisely do the geographical, linguistic and genetic boundaries correspond?
- What can we say about the origins of the people now living in the Himalayan region?
- How has the extreme landscape patterned diversity at the local level? Is language important?

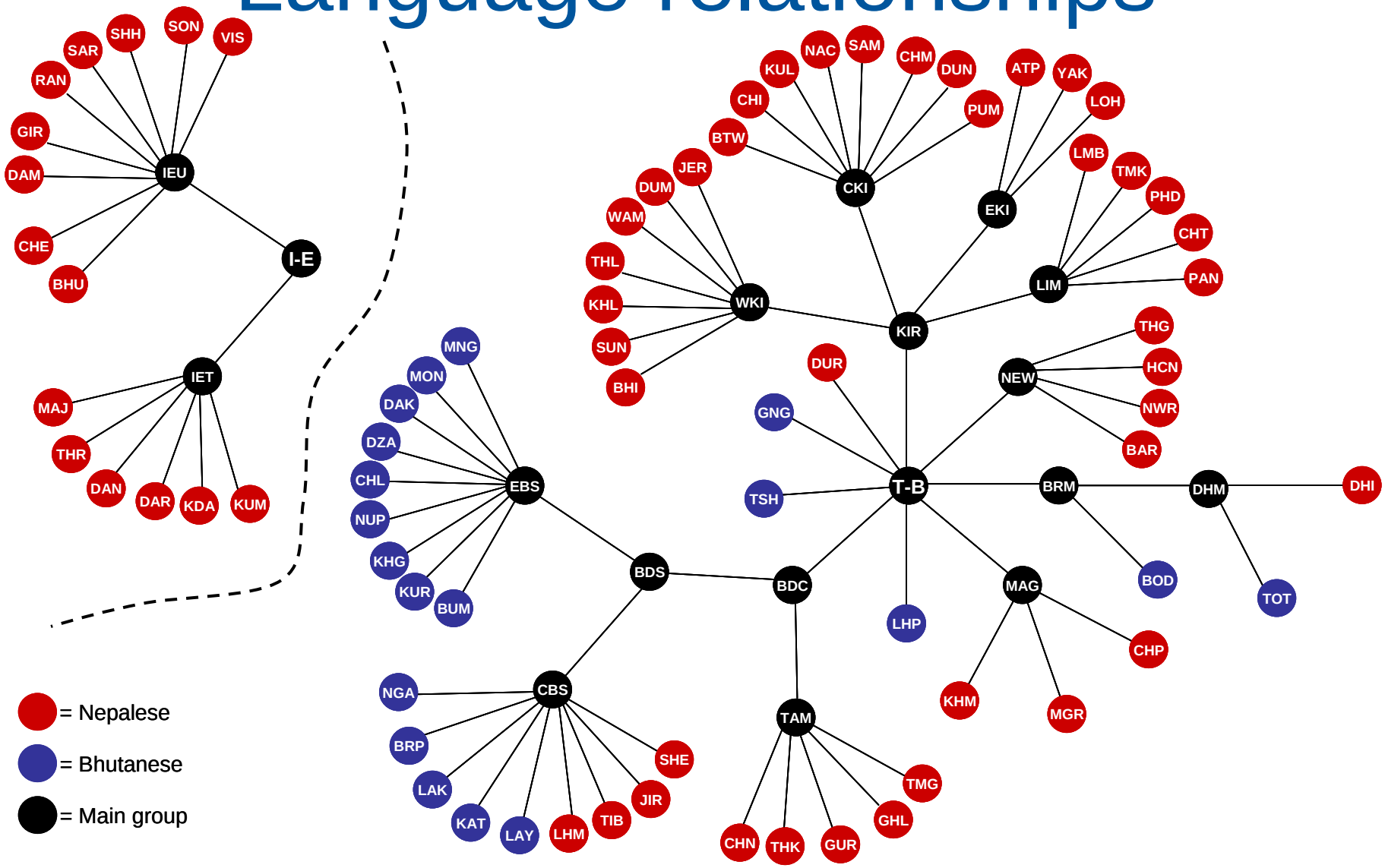
# Samples collected (2002-4)

- Nepal: 955 (764 males), 19 ethnolinguistic groups
- Bhutan: 941 (839 males) 17 ethnolinguistic groups
- Plus data from the literature on 5000-10,000 individuals from neighbouring populations

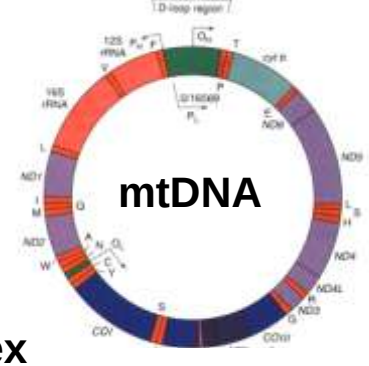


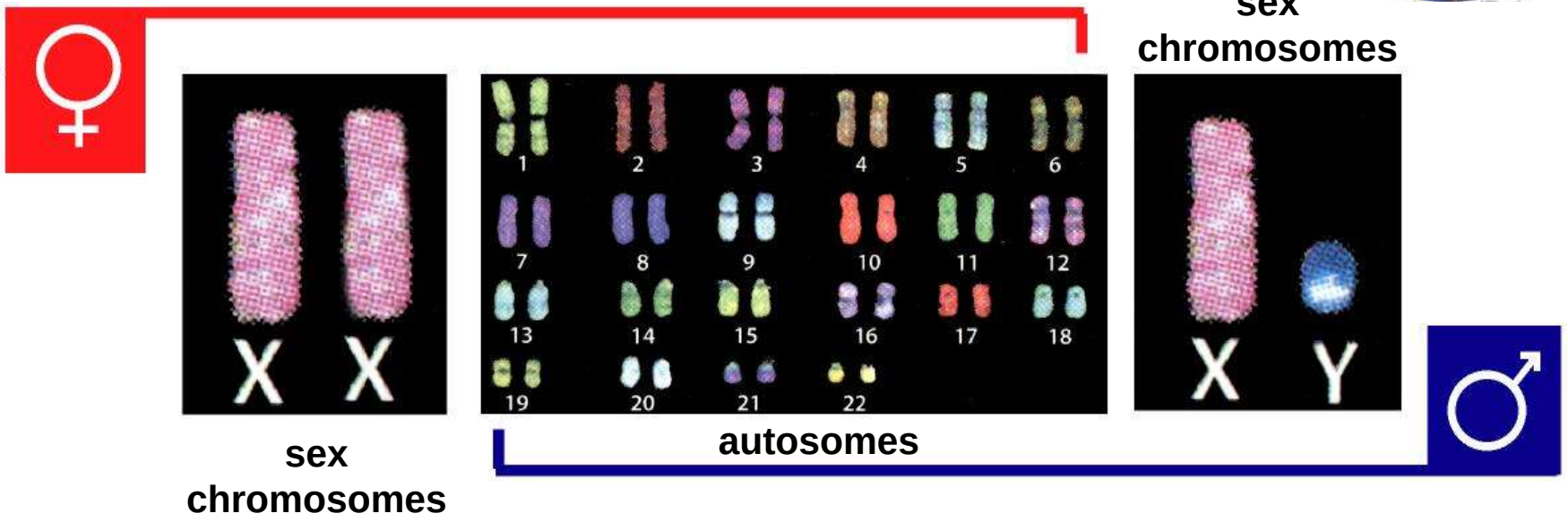

http://www.le.ac.uk/genetics/maj4/Himalayan_OMLLreport.pdf

# Language relationships



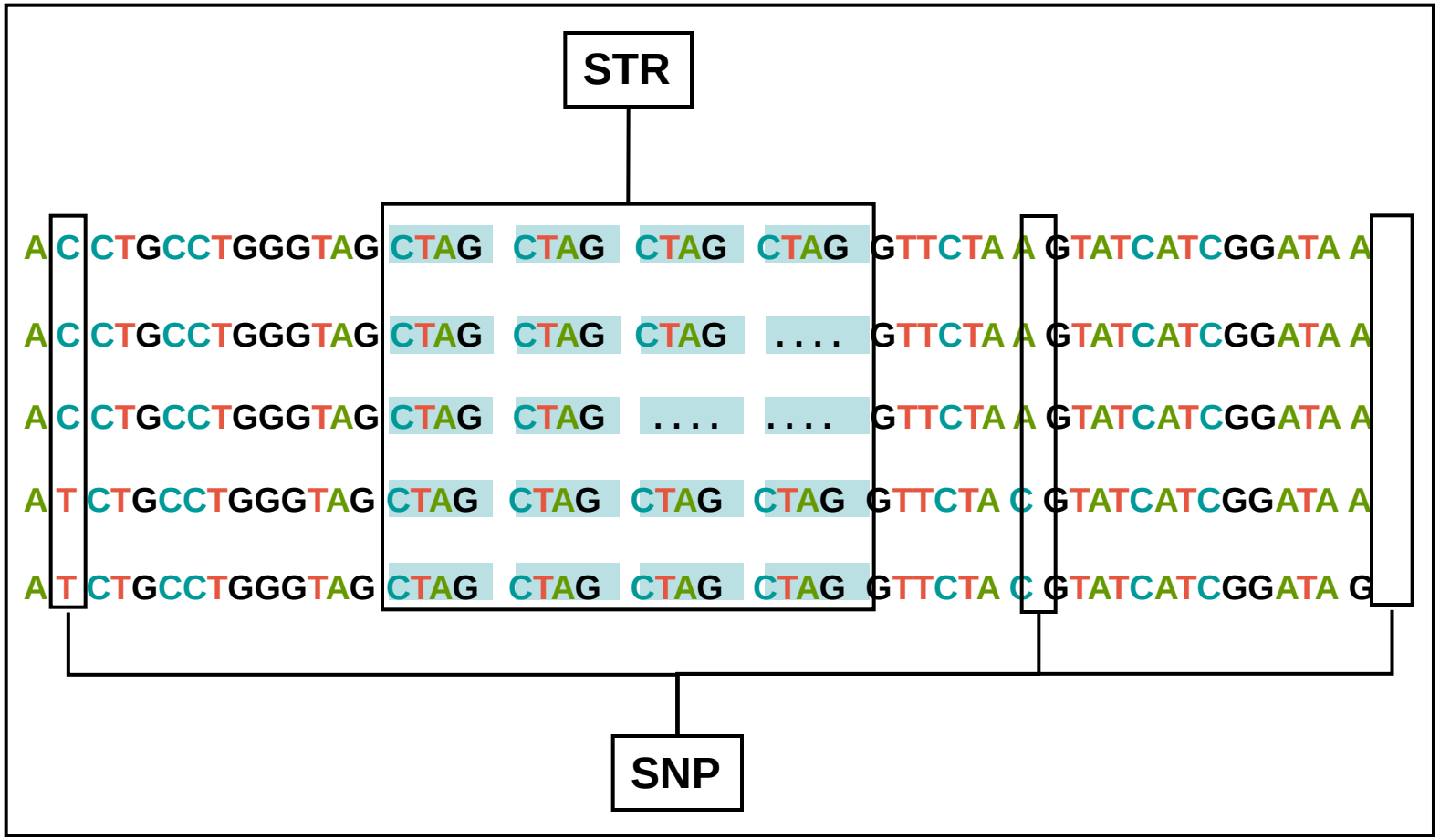
# Inheritance




- Y chromosome sex-determining - father to son
- Both sexes have mtDNA, but inherited only from mothers to children
- Autosomes inherited from both parents

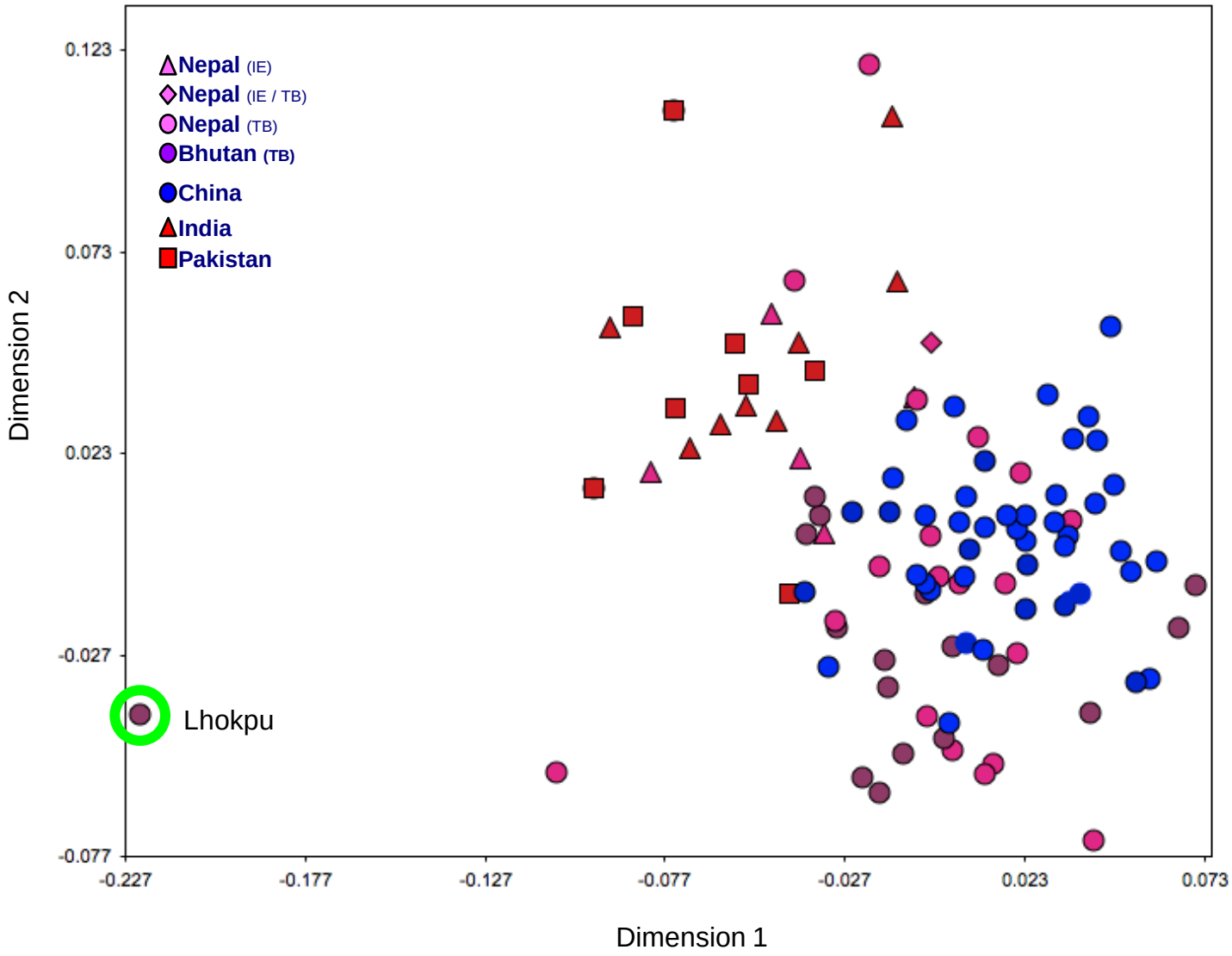
# STRs and SNPs



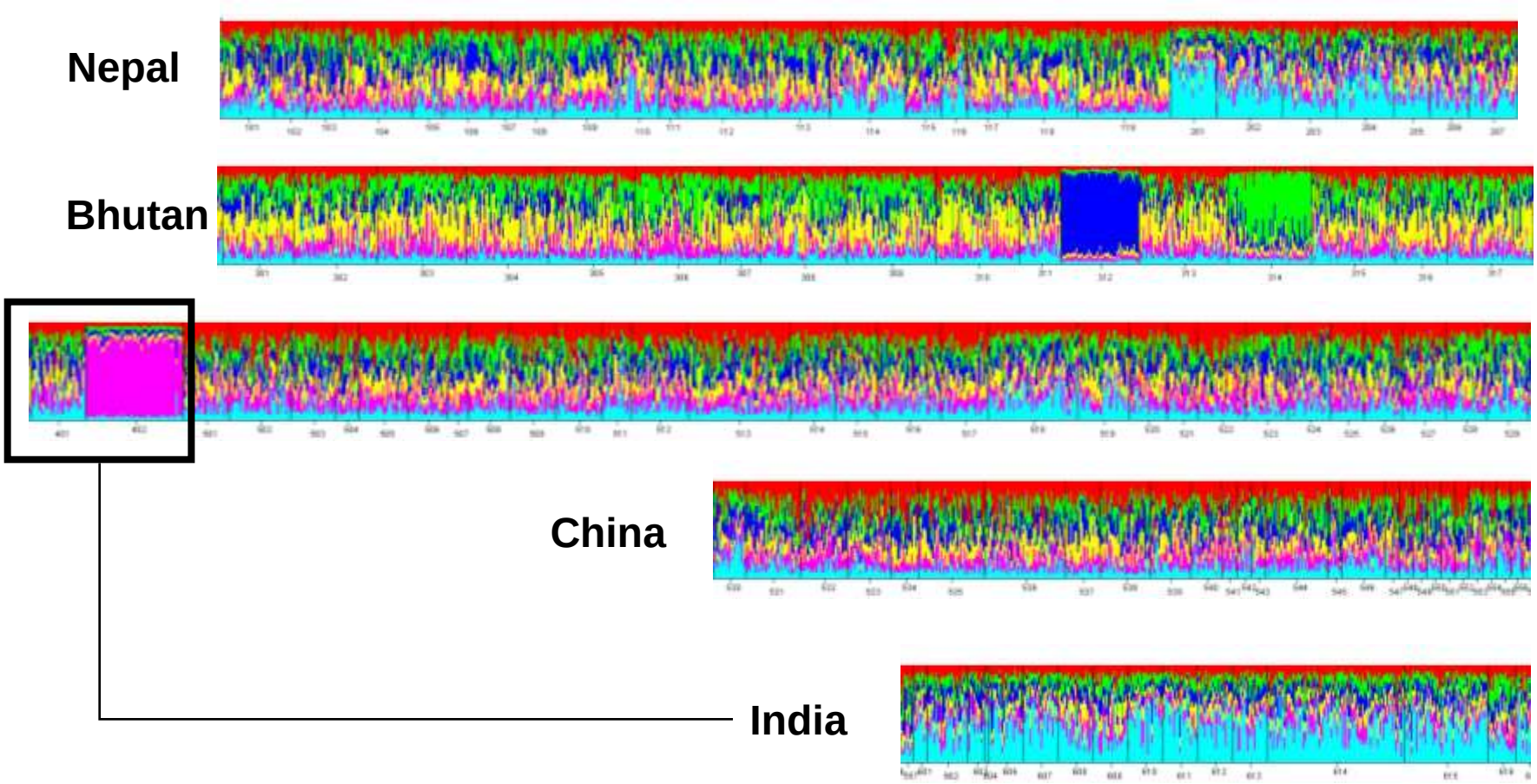
# Markers typed

- Autosomal
  - 21 forensic STRs
  - 13 high-differentiation SNPs
- Y-chromosomal
  - 26 Y-STRs
  - 60 Y-SNPs in hierarchical mode
- mtDNA
  - HVSI
  - 20 coding-region SNPs

# 15 autosomal STRs - MDS

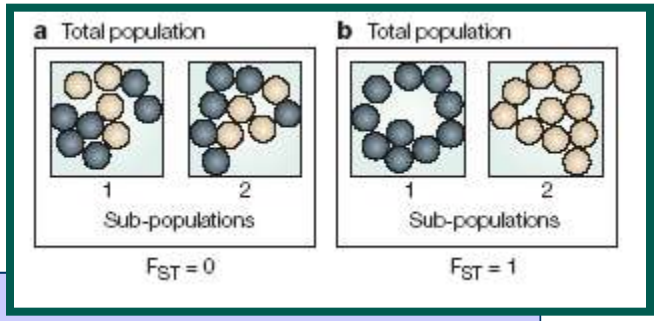


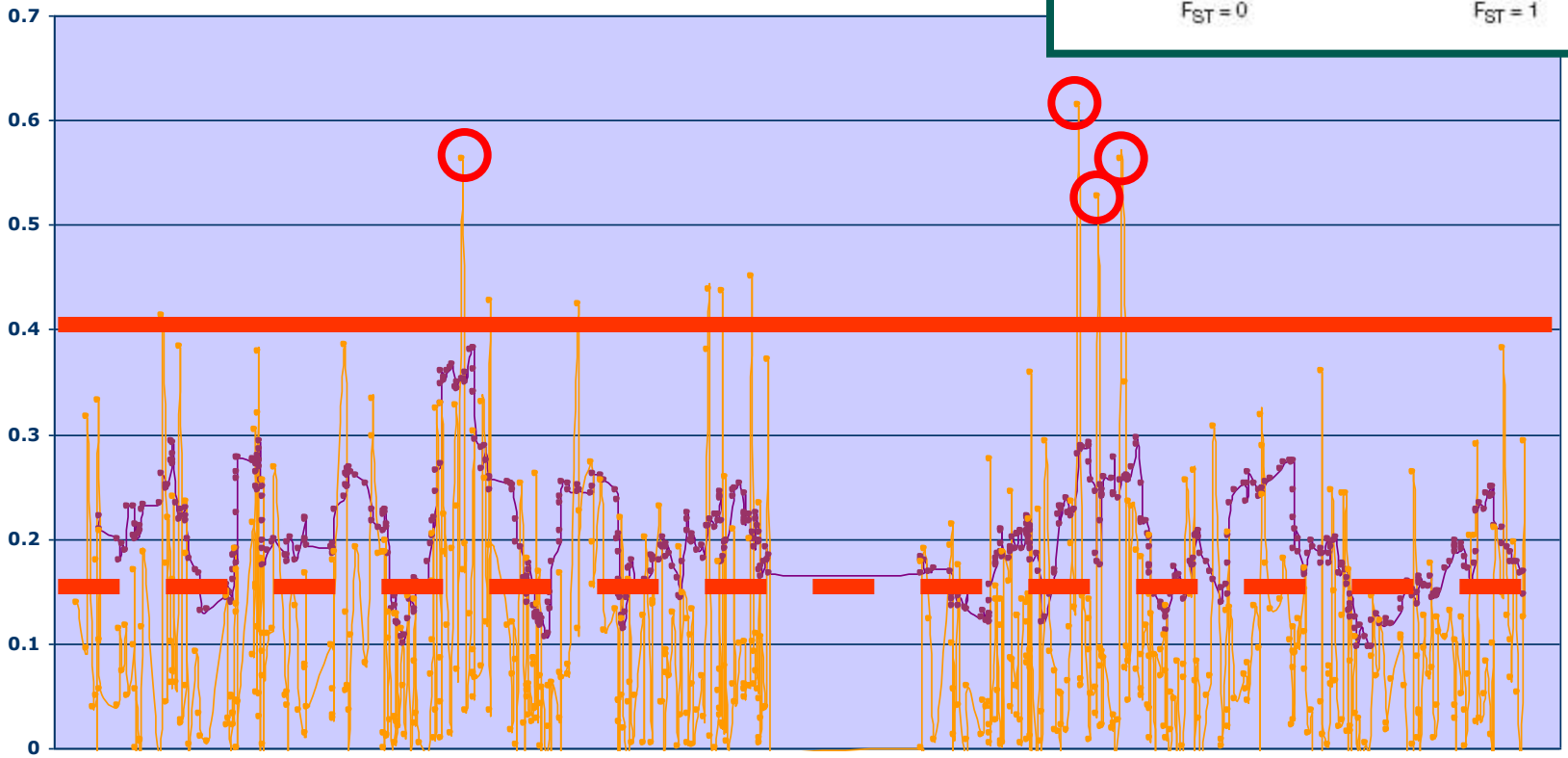
# STRUCTURE: 15 autosomal STRs



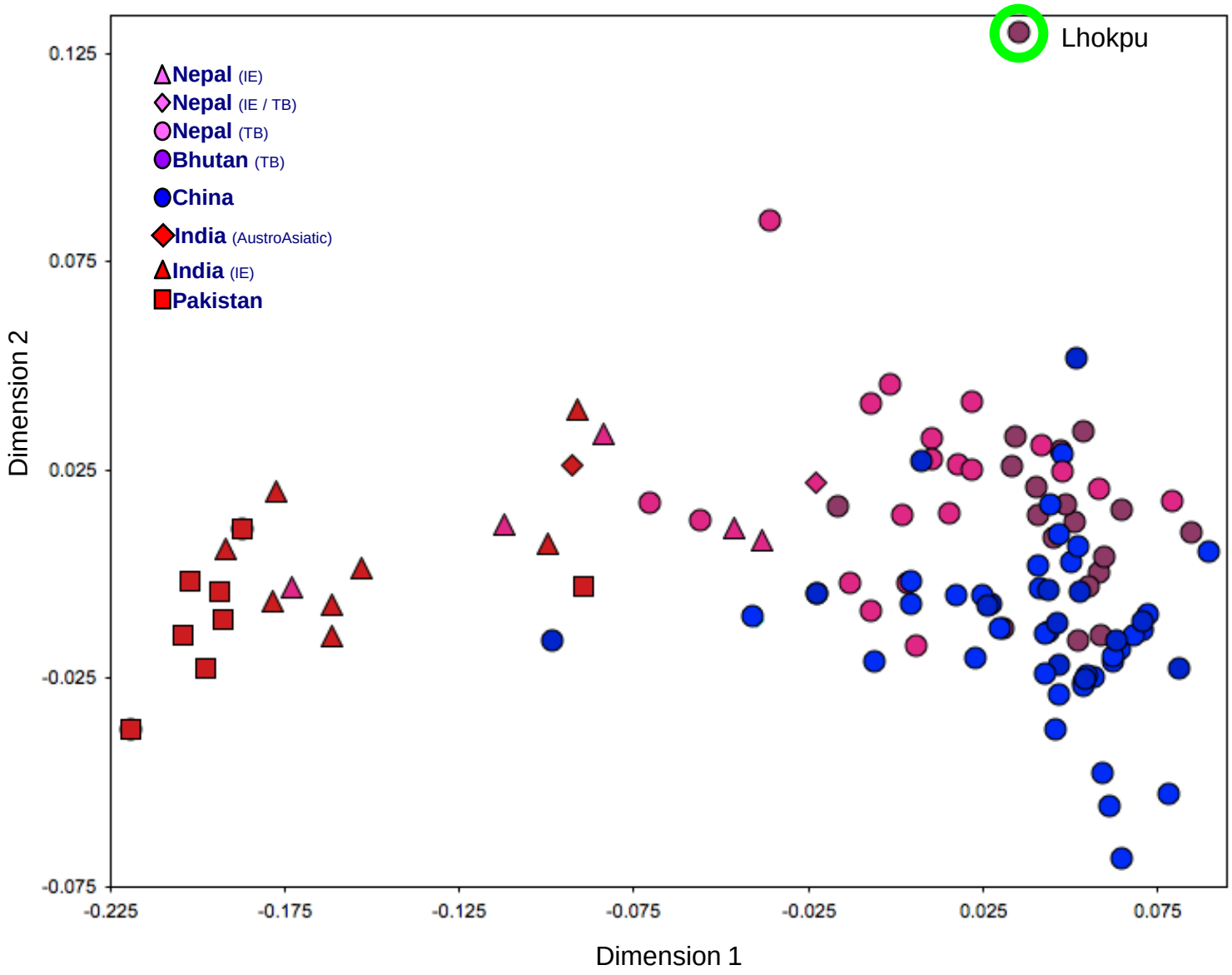
# Identifying informative autosomal SNPs

- $F_{ST}$ between Chinese and Pakistani samples in CEPH-HGDP panel

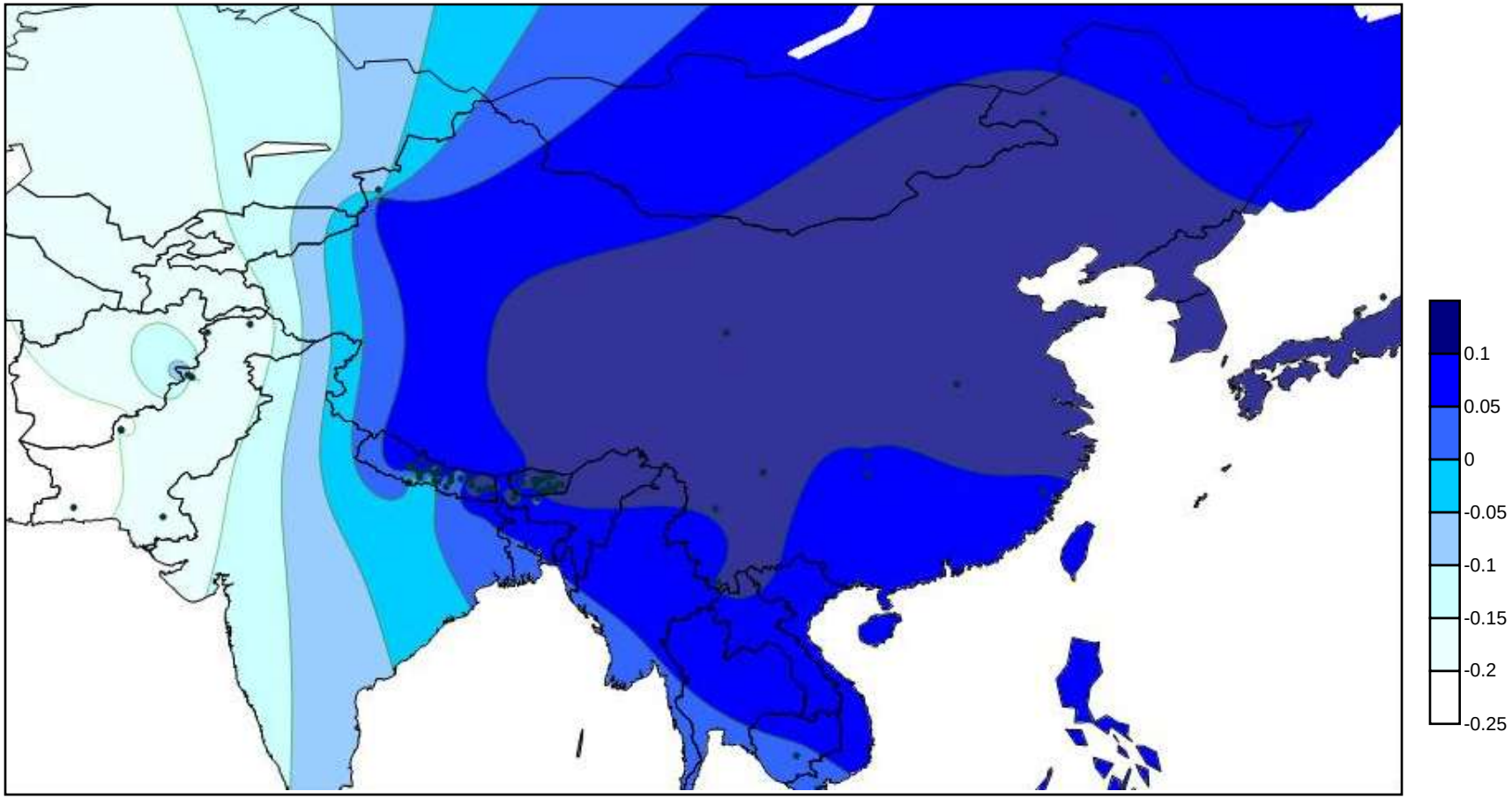


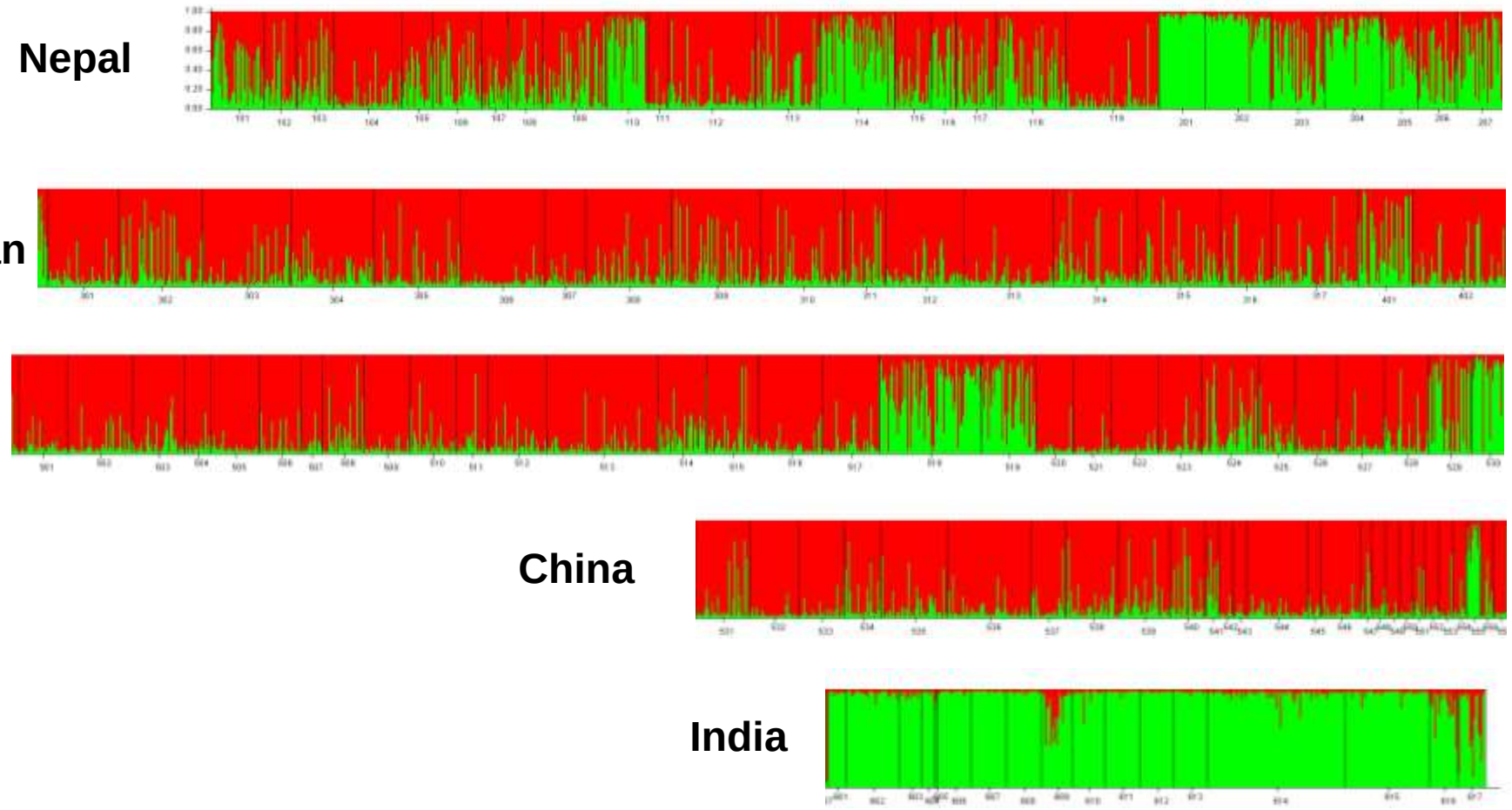
# 13 autosomal SNPs - MDS



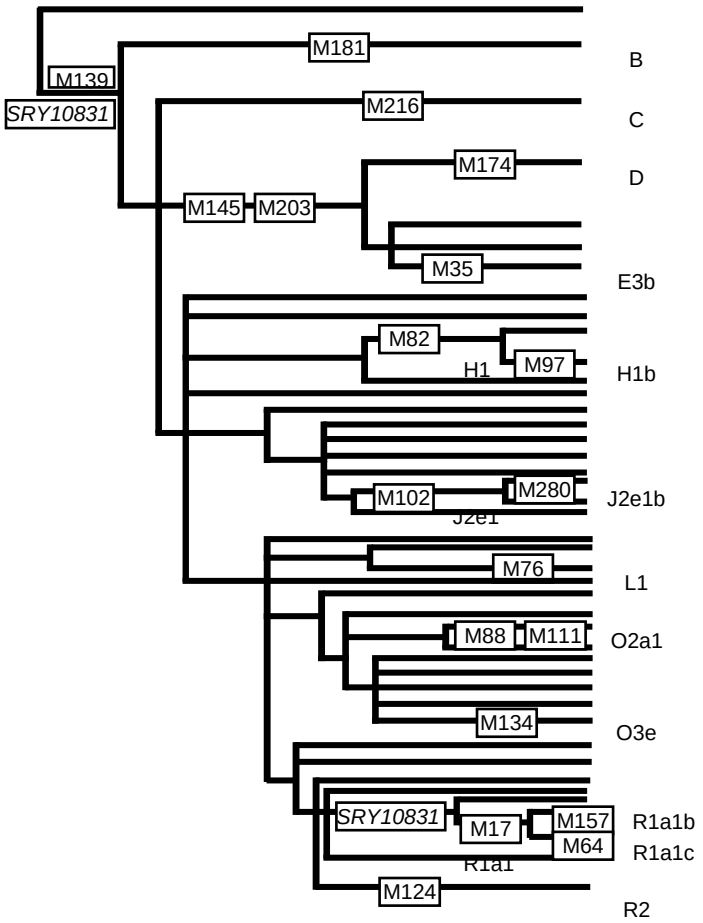
# Contour map of SNP MDS dimension 1

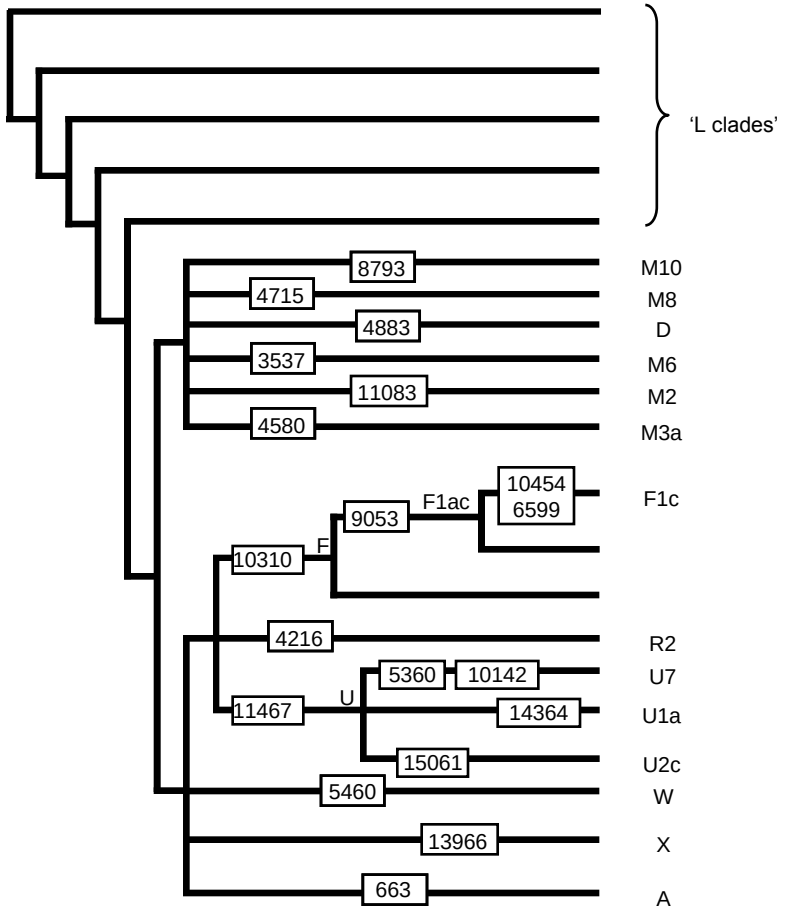


# STRUCTURE: 13 autosomal SNPs

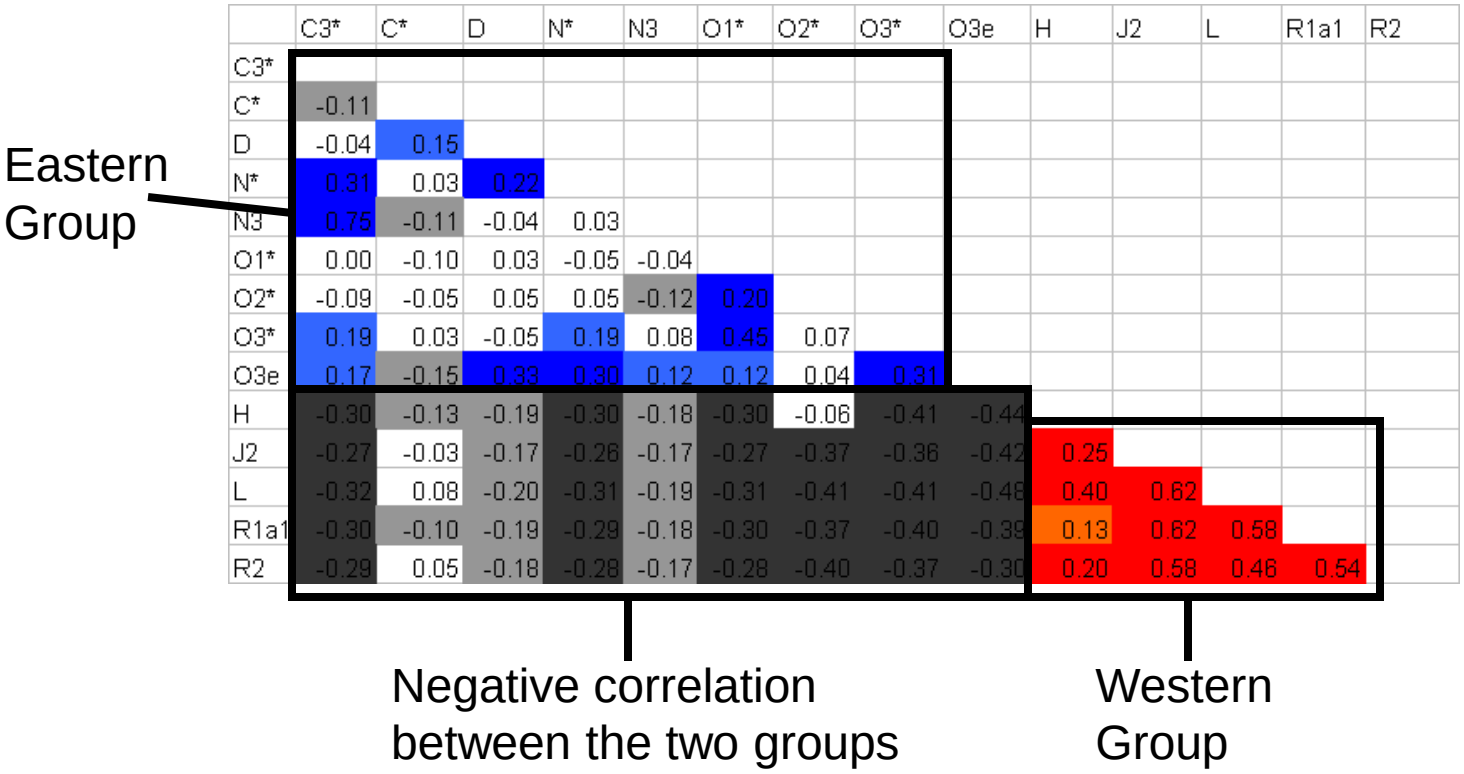


# Y and mtDNA


M181
M139
SRY10831
M216
M174
M145
M203
M35
M82
H1
M97
M102
M280
J2e1
M76
M88
M111
M134
SRY10831
M17
M157
M64
R1a1
M124
B
C
D
E3b
H1b
J2e1b
L1
O2a1
O3e
R1a1b
R1a1c
R2

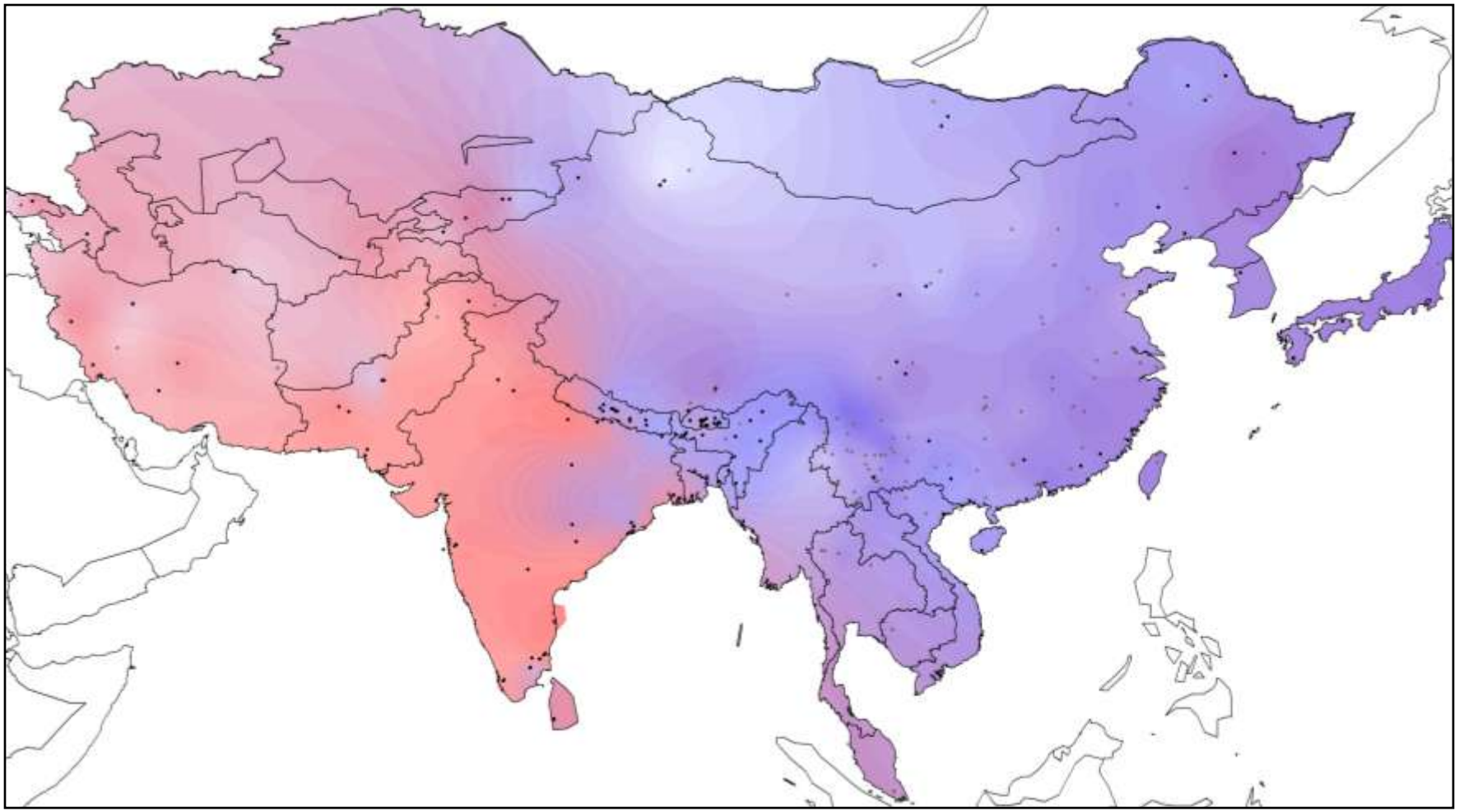

'L clades'
8793
4715
4883
3537
11083
4580
9053
F1ac
10454
6599
F
10310
4216
5360
10142
U
11467
14364
15061
5460
13966
663
M10
M8
D
M6
M2
M3a
F1c
R2
U7
U1a
U2c
W
X
A

# Identification of geographically-correlated Y-haplogroups

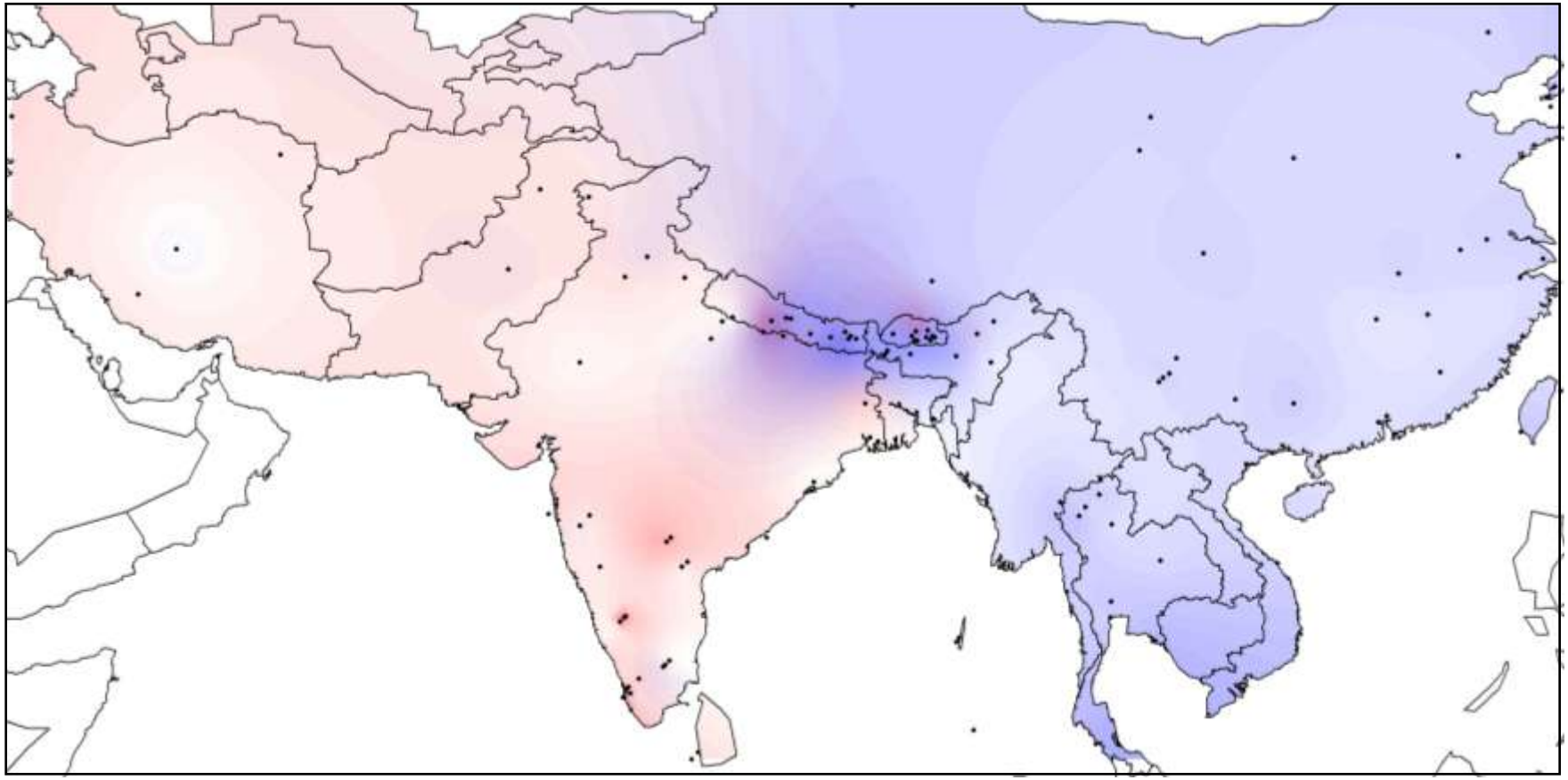


| | C3* | C* | D | N* | N3 | O1* | O2* | O3* | O3e | H | J2 | L | R1a1 | R2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C3* | | | | | | | | | | | | | | |
| C* | -0.11 | | | | | | | | | | | | | |
| D | -0.04 | 0.15 | | | | | | | | | | | | |
| N* | 0.31 | 0.03 | 0.22 | | | | | | | | | | | |
| N3 | 0.75 | -0.11 | -0.04 | 0.03 | | | | | | | | | | |
| O1* | 0.00 | -0.10 | 0.03 | -0.05 | -0.04 | | | | | | | | | |
| O2* | -0.09 | -0.05 | 0.05 | 0.05 | -0.12 | 0.20 | | | | | | | | |
| O3* | 0.19 | 0.03 | -0.05 | 0.19 | 0.08 | 0.45 | 0.07 | | | | | | | |
| O3e | 0.17 | -0.15 | 0.33 | 0.30 | 0.12 | 0.12 | 0.04 | 0.31 | | | | | | |
| H | -0.30 | -0.13 | -0.19 | -0.30 | -0.18 | -0.30 | -0.06 | -0.41 | -0.44 | | | | | |
| J2 | -0.27 | -0.03 | -0.17 | -0.26 | -0.17 | -0.27 | -0.37 | -0.36 | -0.42 | 0.25 | | | | |
| L | -0.32 | 0.08 | -0.20 | -0.31 | -0.19 | -0.31 | -0.41 | -0.41 | -0.48 | 0.40 | 0.62 | | | |
| R1a1 | -0.30 | -0.10 | -0.19 | -0.29 | -0.18 | -0.30 | -0.37 | -0.40 | -0.39 | 0.13 | 0.62 | 0.58 | | |
| R2 | -0.29 | 0.05 | -0.18 | -0.28 | -0.17 | -0.28 | -0.40 | -0.37 | -0.30 | 0.20 | 0.58 | 0.46 | 0.54 | |

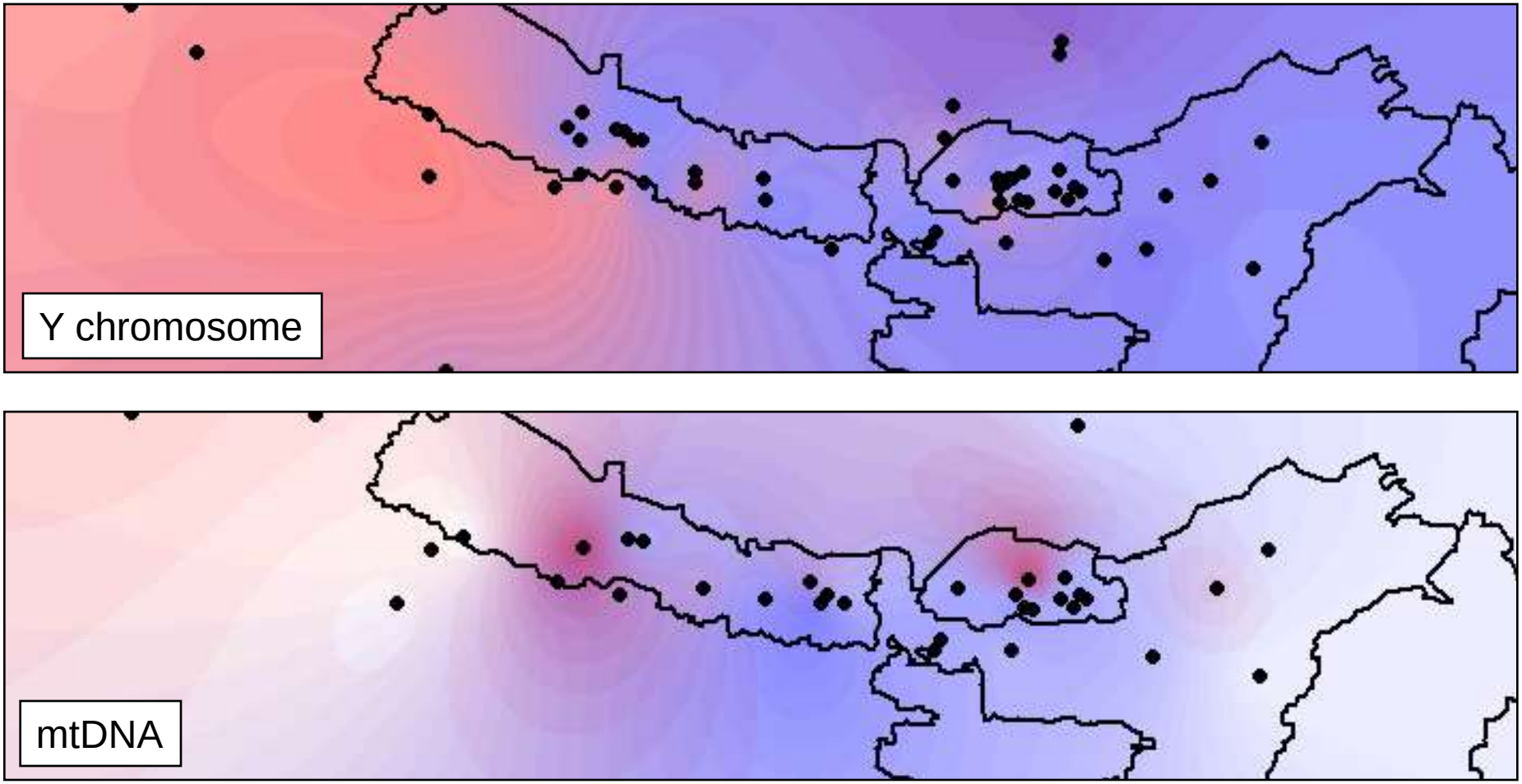
# The Y-chromosomal landscape



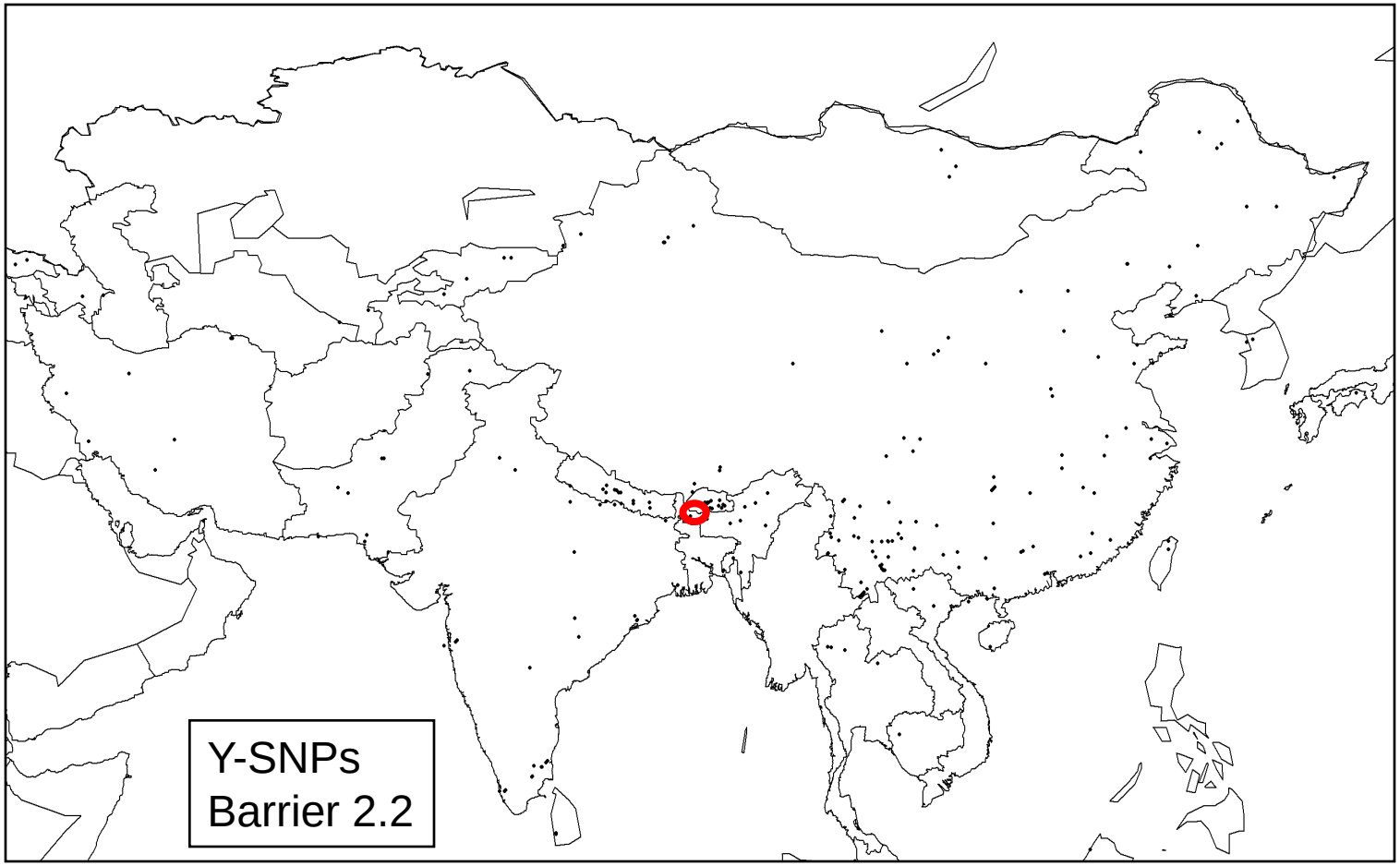
# The mtDNA landscape



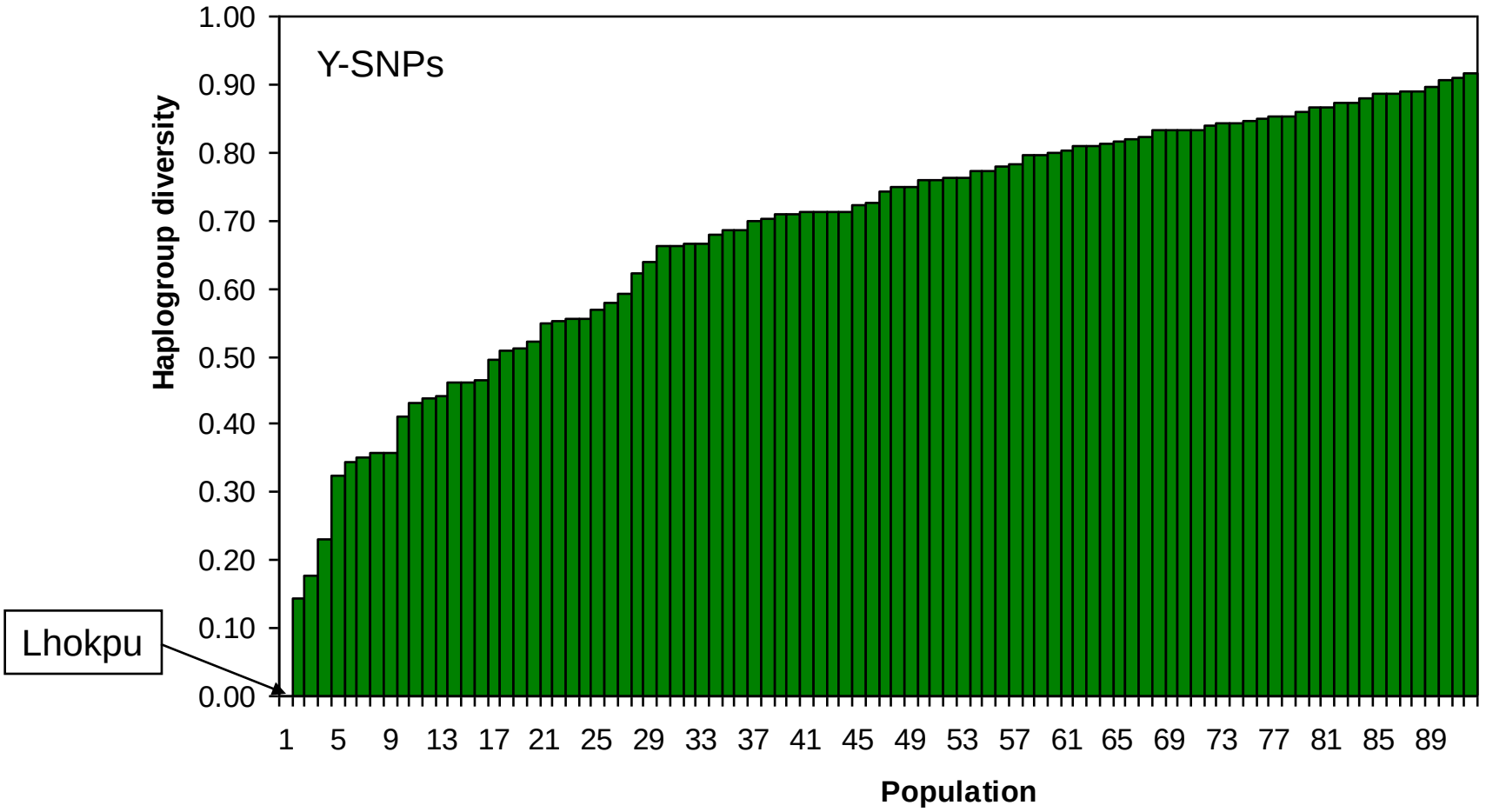
# The Himalayan region



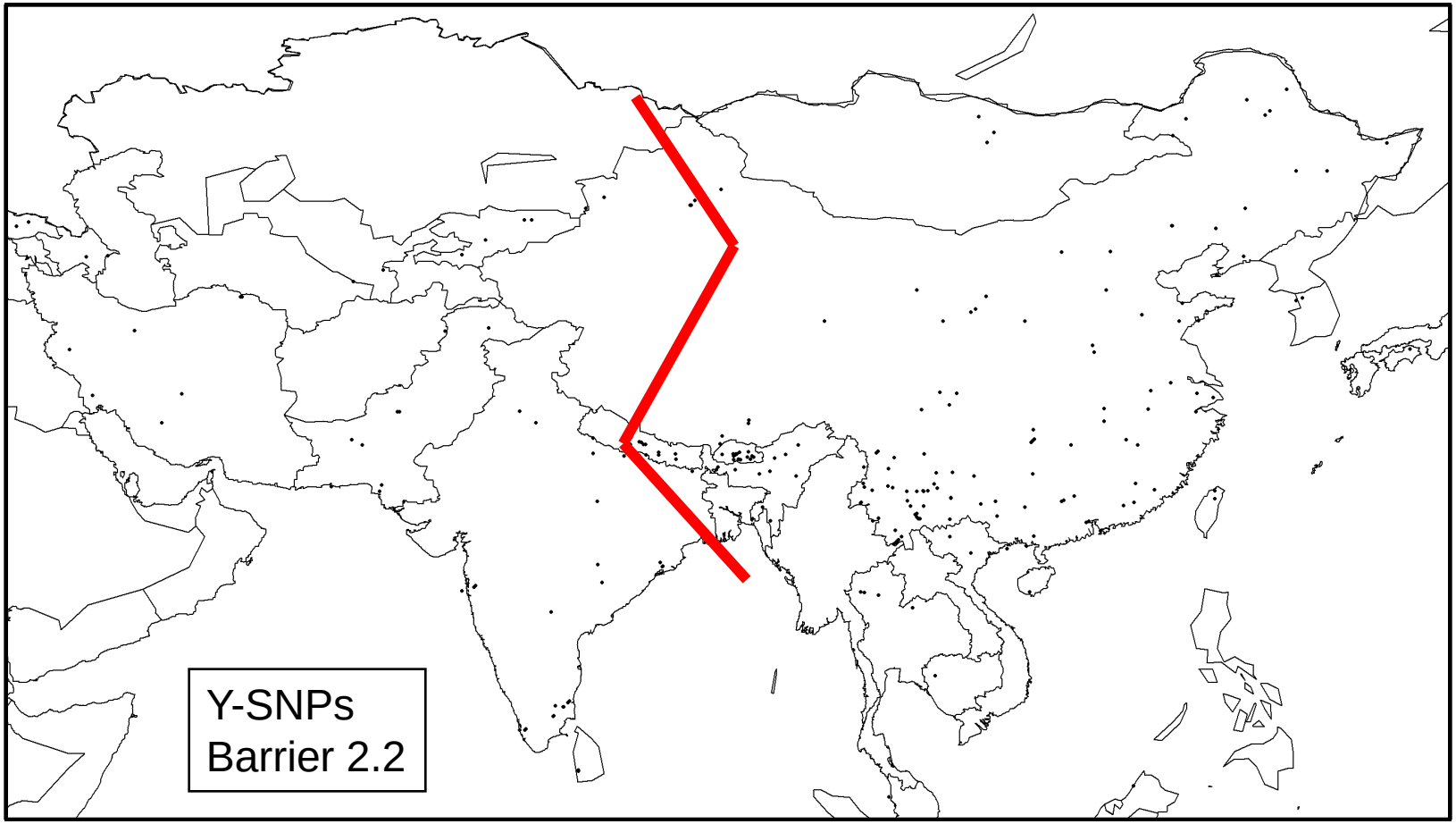
# Searching for boundaries with Barrier



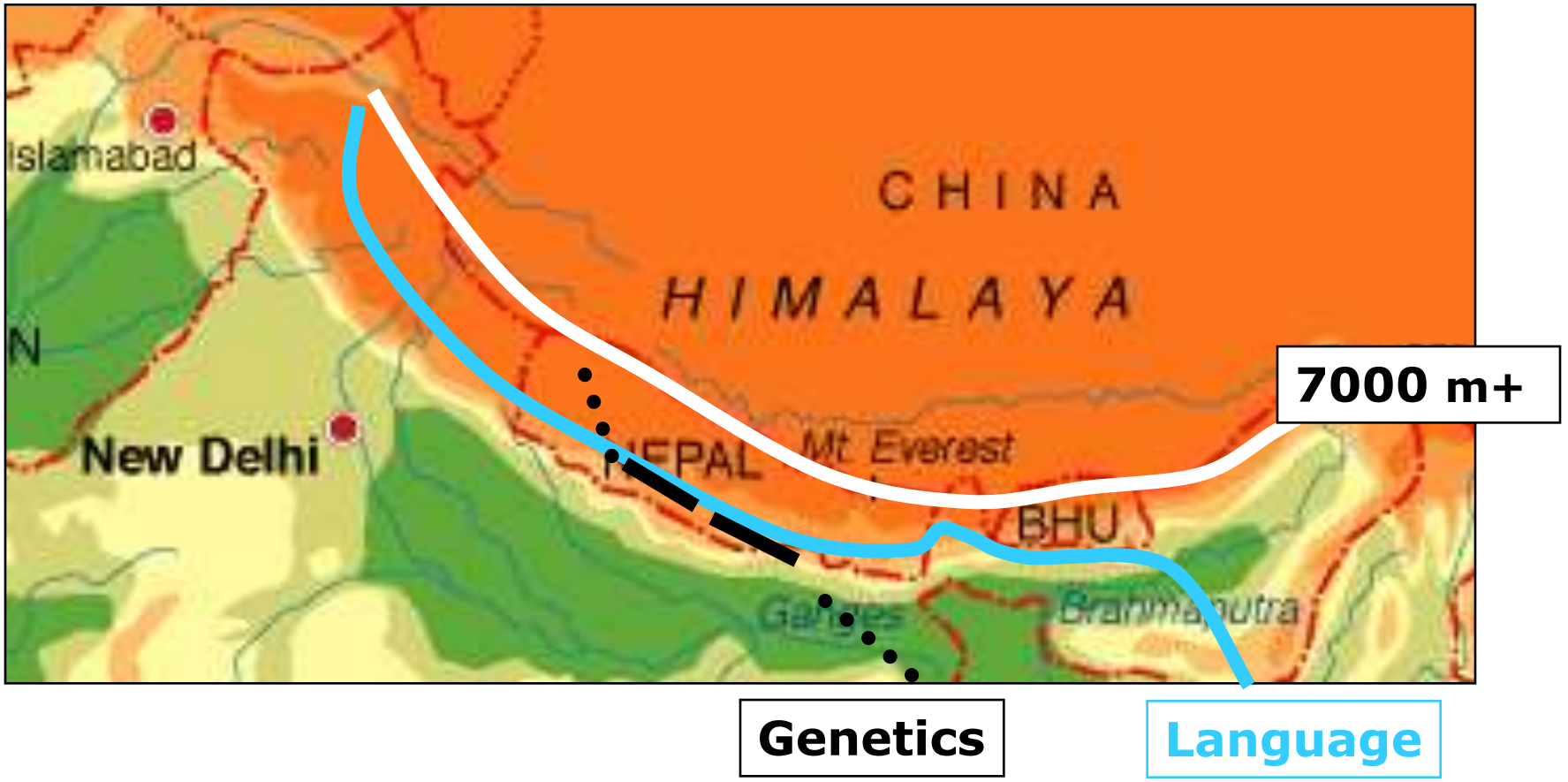
# Diversity varies between populations



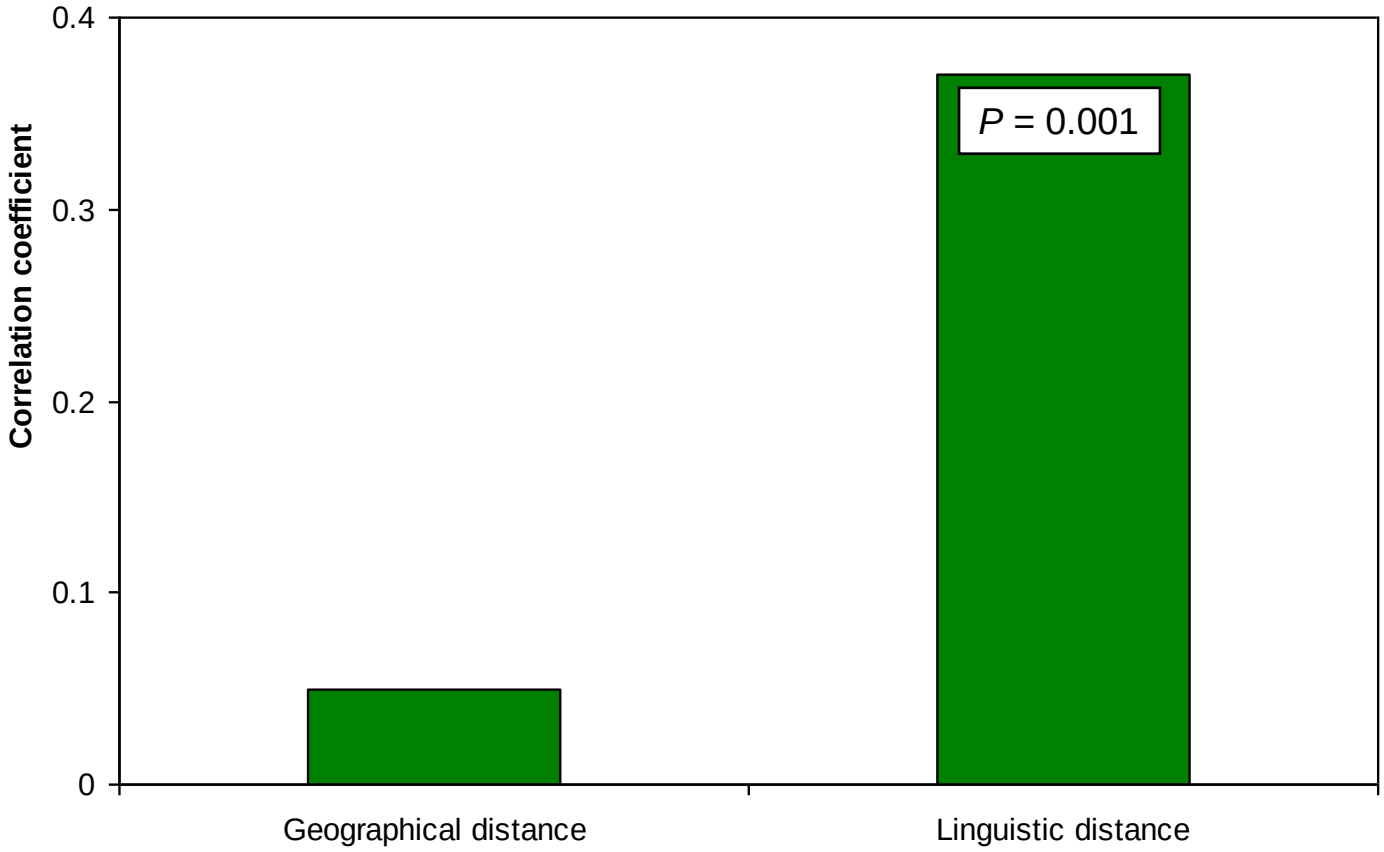
# Boundary after removing low-diversity populations



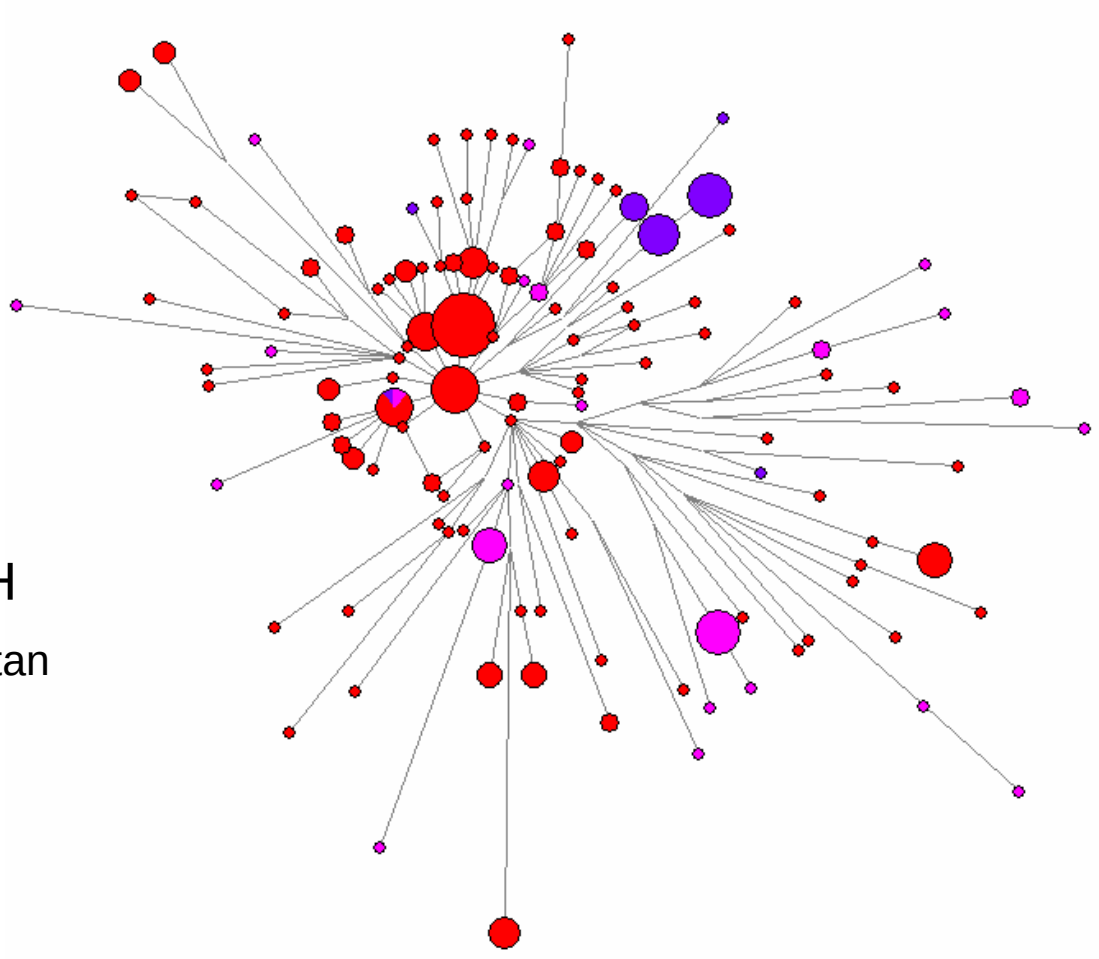
# Comparison of boundaries



# Mantel tests of genetic distance correlation (Y data)



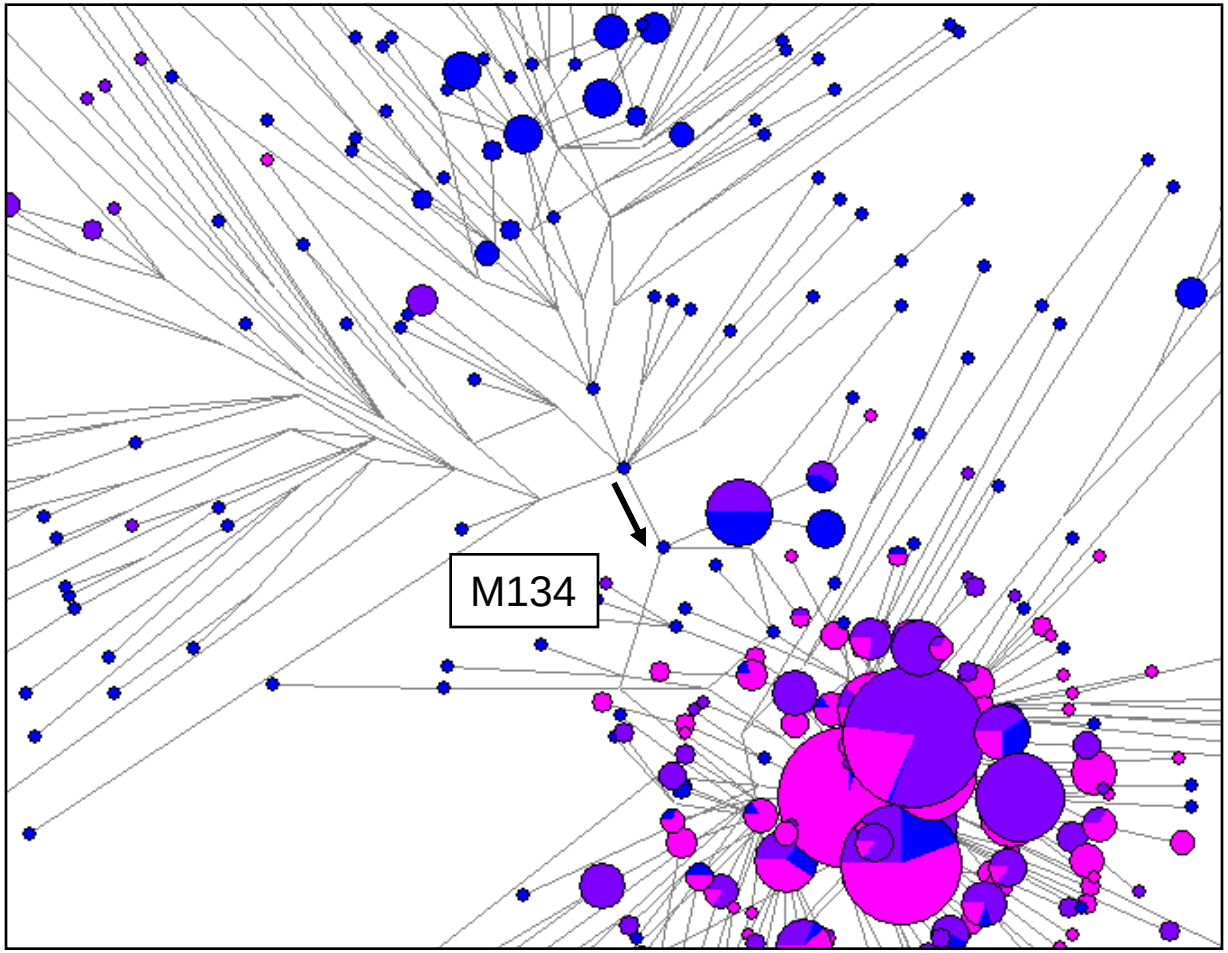
# Direction of gene flow?



Haplogroup H

- India, Pakistan
- Nepal
- Bhutan

# Direction of gene flow?




Haplogroup O3

- East Asia
- Nepal
- Bhutan

# Peopling of the region

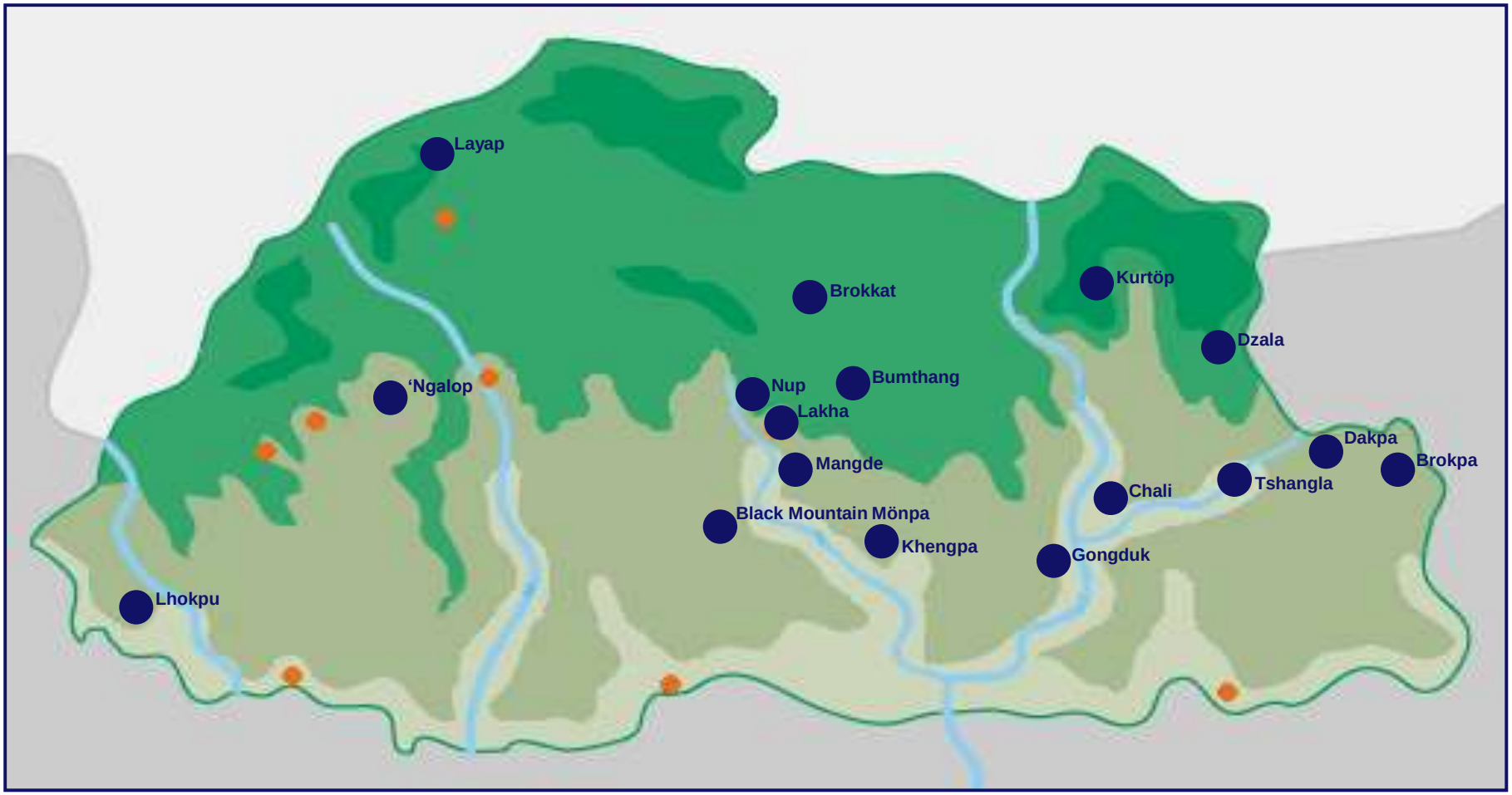

- When? Archaeology; palaeoclimatology; linguistics; genetics…

# Conclusions

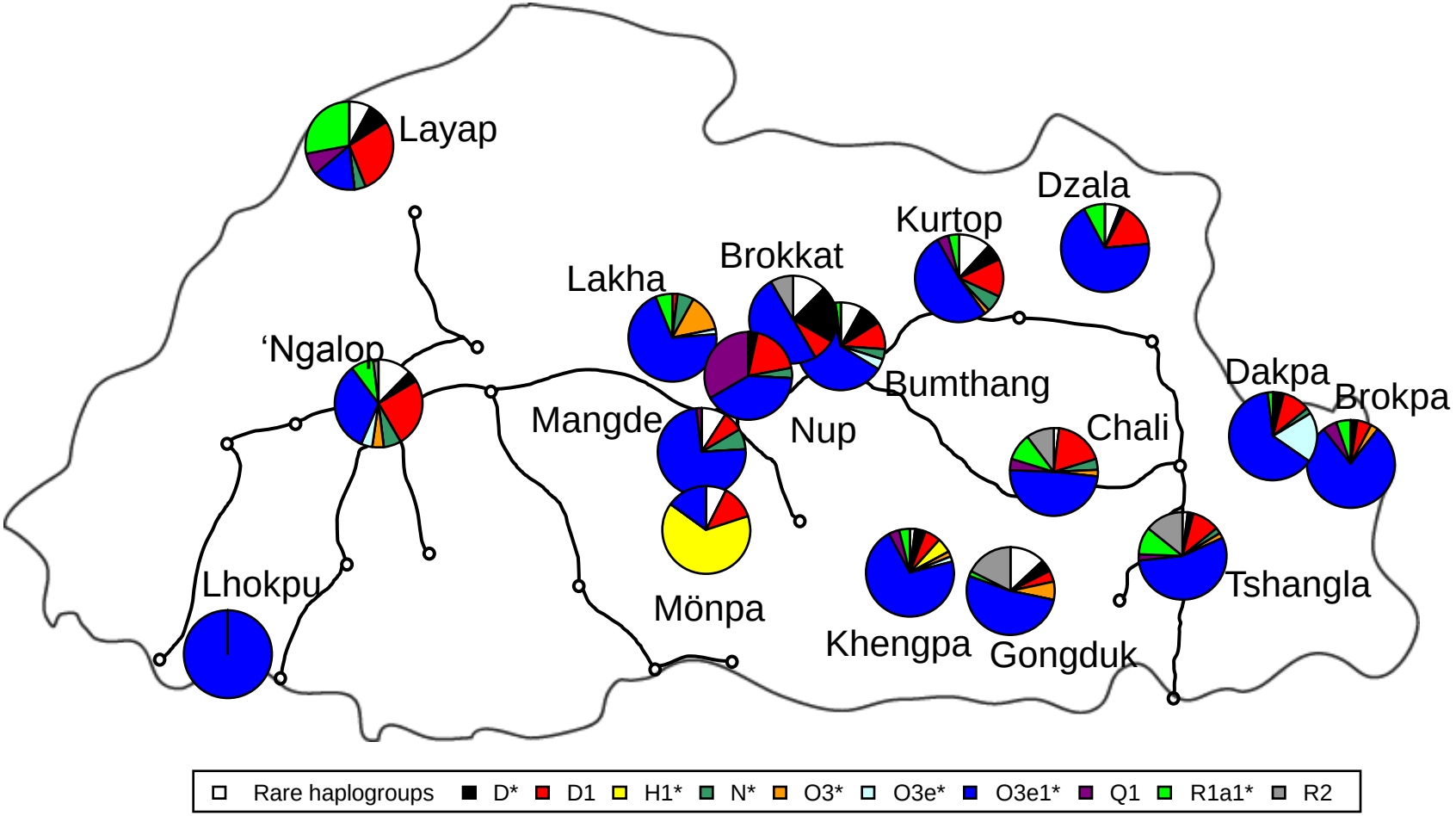
- Genetic variation correlates better with language than with geography
- Almost perfect fit between linguistic and genetic boundaries
- Consistent with TB-speaker origins from the east and IE-speaker origins from the west within the last few thousand years
- High diversity within Nepal and Bhutan

# Extreme geography in Bhutan

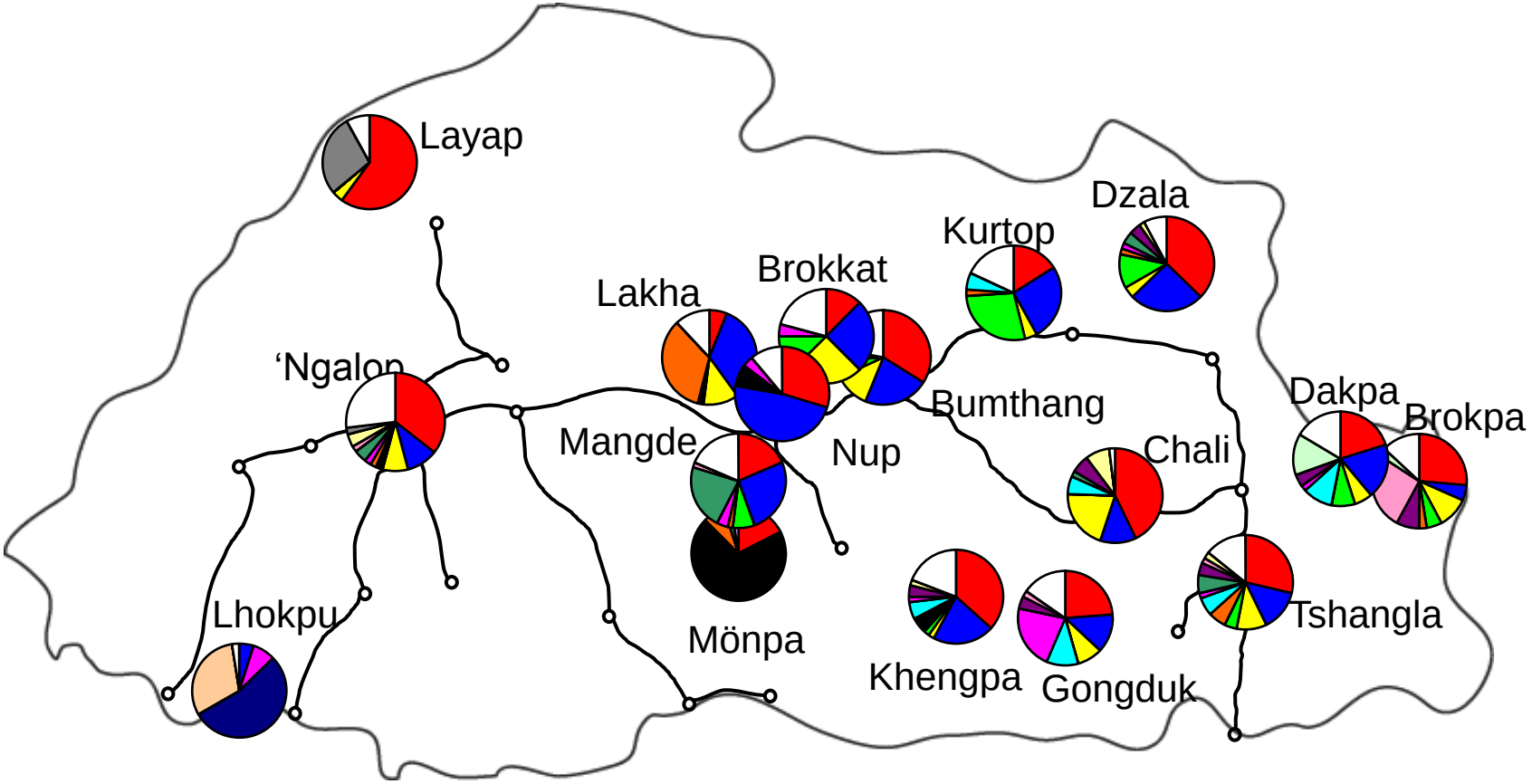
# Ethnolinguistic centres in Bhutan



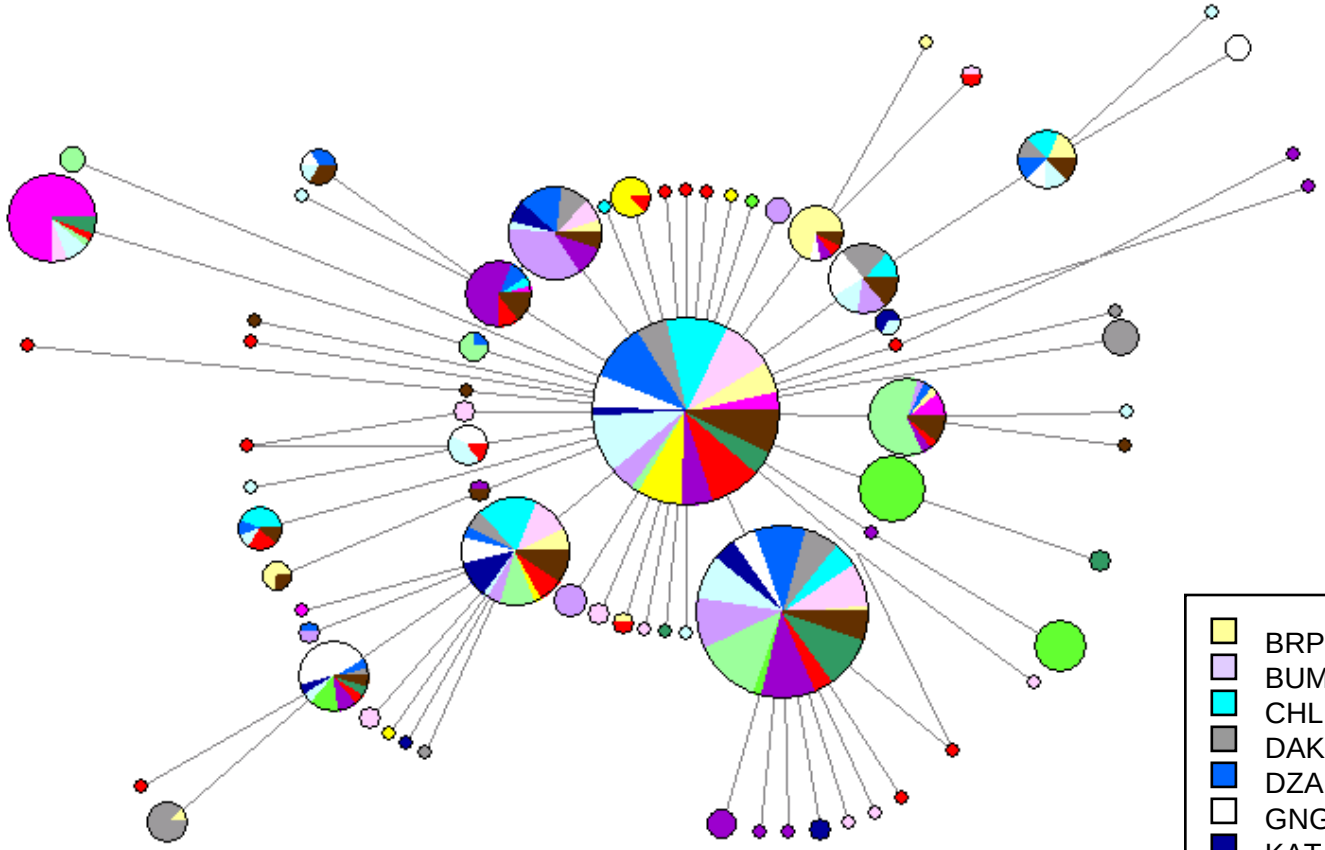
# Y haplogroups - high diversity, much population structure



# mtDNA haplotypes - high diversity, much population structure



# mtDNA haplotypes



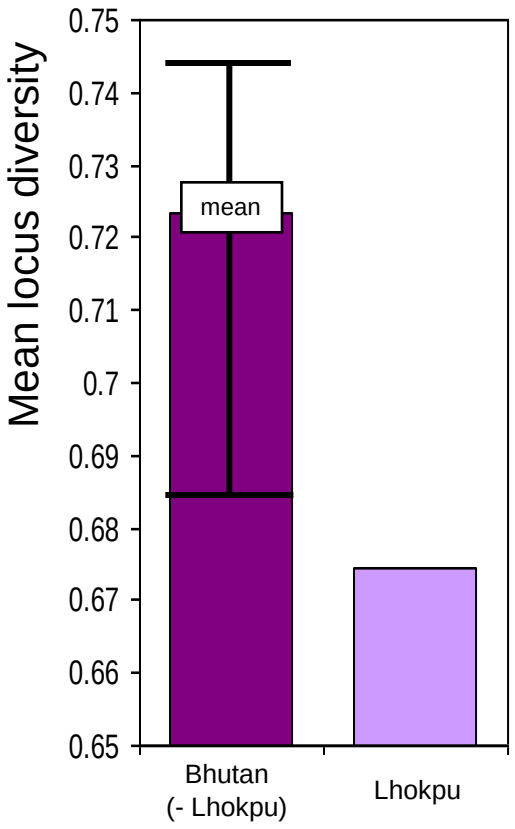

Median-Joining Network of HVSI sequences

All nucleotide sites with a recurrence greater than 2 removed

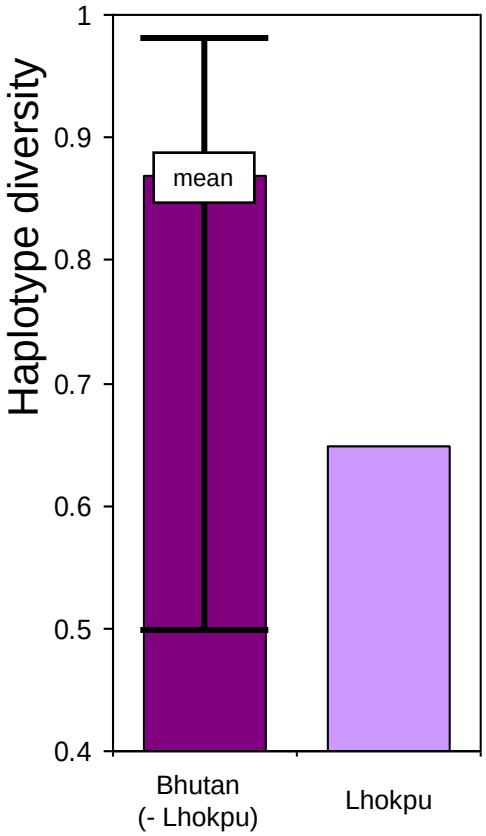
Network weighted with ts:tv of 1:15

# The Lhokpu: low diversity

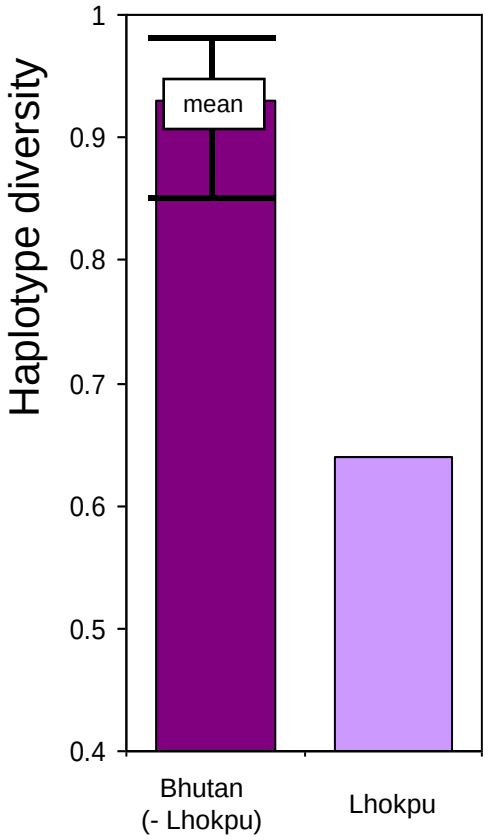
**Autosomal STRs**




**mtDNA HVSI**

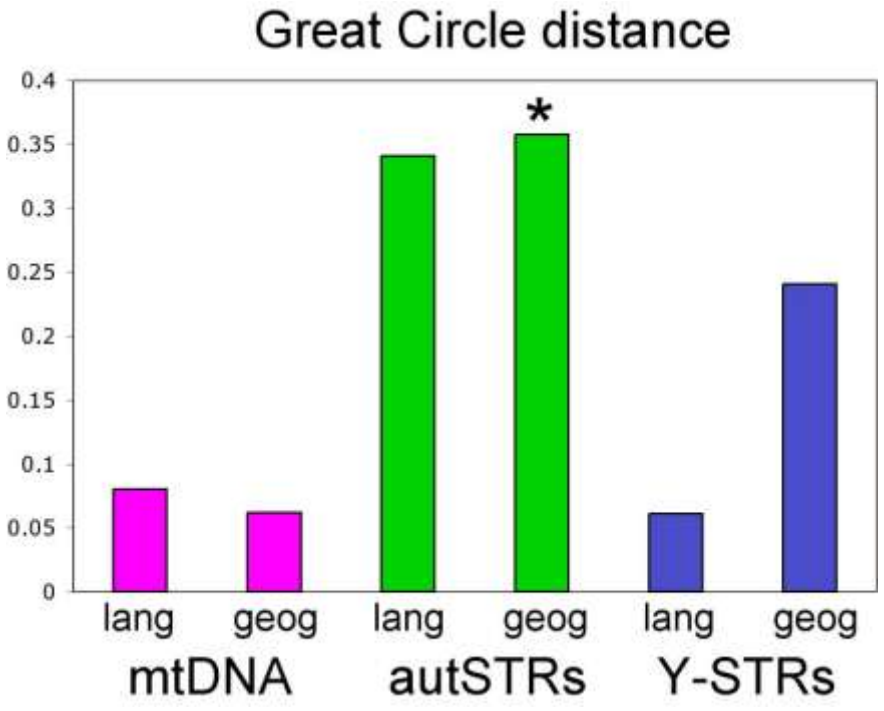


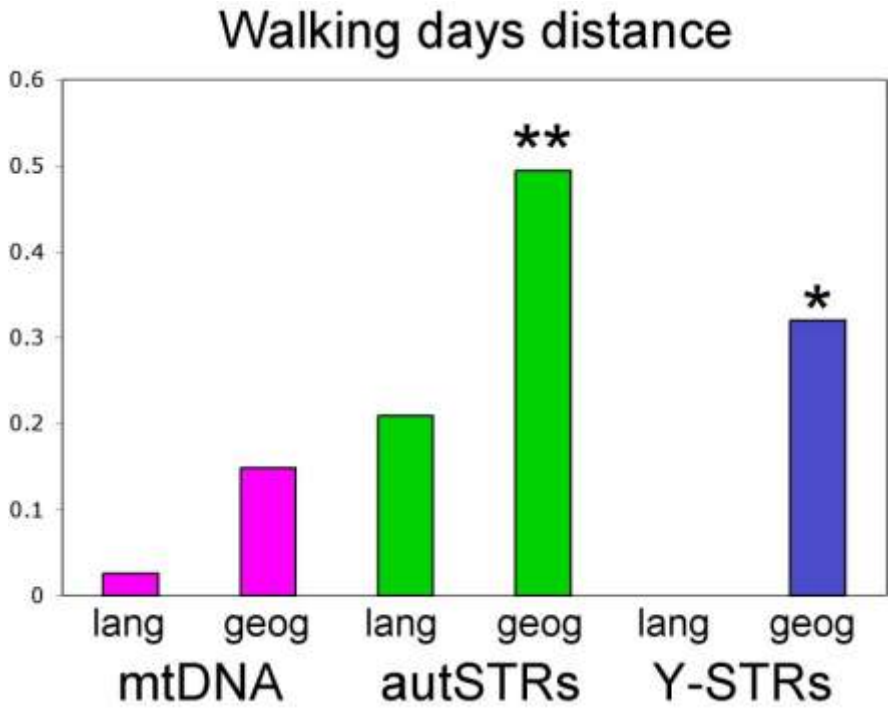

**Y-STRs**



# Diversity patterned more by geography than language

- Mantel testing, using distances:
  - Genetic: $F_{ST}$, $R_{ST}$
  - Geographic: Great Circle, or walking days
  - Linguistic: phylogenetic units within language tree



# Acknowledgements

The research in Nepal was conducted in association with the Centre for Nepal and Asian Studies (CNAS) at Kirtipur under the Bilateral Agreement for Academic Cooperation between Tribhuvan University (TU) and Leiden University. The research team received the organisational support and enthusiastic assistance of many grass-roots community service organisations and from the many informed volunteers of indigenous language communities who took an active interest in the research and discussed with interest the ramifications of genetic investigations for an enhanced understanding of our shared prehistorical past. Special gratitude is due the Arjun Limbu, Yograj Limbu and the Kirat Yakthung Chumlung headquartered at Mahalakshmisthan in Lalitpur; the Praja Capacity Development Programme (Praja Samudayik Vikash Karyakram) at Shaktikhor in Chitwan district; to Ajay Praja, Santa Bahadur Praja and Dambar Bahadur Chepang and the Nepal Chepang (Praja) Sangh at Kathmandu; to our many Tharu friends at Sauraha; to Bal Gopal Shrestha of Sankhu and the Newar community organisation Friends of Sankhu; to Bharat Rai of the National Youth Service Trust, our dear friend Ashok Rai and the Danuwar Rai community; to Dilendra Subba and the Limbu Literary Development Association; to the Kirat Yakkha Chuma headquartered at Mahalakshmisthan in Lalitpur; to Shree Mani Chand Chantyal and the Nepal Chantyal Samaj headquartered in Samakhushi in Kathmandu; to Buddhiman Dura, Lt. Col. John P. Cross, Ritu Kumar Dura and the Dura Seva Samaj Sampark Karyalaya at Ram Bajar in Pokhara; to Kishor Dura, Singh Raj Dura and the Dura Seva Samaj Samiti at Vasundhara in Kathmandu; to the Kshetriya Karyalaya of the Thakali Seva Samiti at Nadipur in Pokhara; to Dashrath Rai and the Kendriya Karyalaya of the Kirat Rai Yayokkha at Kathmandu; to Lila Thangmi, Kavi Raj Thangmi of Lapilang and the Nepal Thami Samaj; to Avinath Rai and Ganesh Rai of the Wambule Rai Samaj Nepal; to our many kind Baram friends of Gorkha district, for whom we are still completing a grammatical description of the language; to our wonderful friend Temba Bhote and the Buddhist half of the Lhomi (Shingsaba, Bhote) community; to Lt. Col. Michael Roe, Capt. Simon Garside and those troops of the British Gurkhas at Pokhara who came forward to volunteer blood; to our old friend Vishva 'Bishow' Bhatt of Gorkha; to Tek Bahadur Kumal and the Kumal Service Organisation at Chevetar in Gorkha district; to the kind director and staff of the Youth Awareness Environmental Forum and the Environmental Library at Badegaon at Godawari. Gratitude is also due to the Nepal Janjati Mahasangh at Anamnagar for their scholarly interest and kind advice. Furthermore, we owe gratitude to Cas F. de Stoppelaar, Honorary Consul-General of Nepal in the Netherlands and to Kari Cuelenare of the Dutch Consulate in Nepal. Last but not least, we thank our old and dear friends Narayan Prasad 'Yangsarumba' Pangyangu Subba and Gram Bahadur 'Sarumba' Pangyangu Subba of Tamphula in Tehrathum district and our dear friend Surendra Raj Dhakal of Gorkha

The research in the Kingdom of Bhutan was conducted as part of the long-standing bilateral cooperation between the Royal Government of Bhutan and Leiden University. The field campaign was carried out in accordance with the Memorandum of Understanding concluded between the Dzongkha Development Authority (DDA) and Leiden University. Much gratitude is due to the Chairmen of the Dzongkha Development Authority, Their Excellencies 'Lönpo Sangay Ngedup and 'Lönpo Thinley Gyamtsho, as well as to the Home Minister, His Excellency 'Lönpo Jigmi Yoezer Thinley, for encouraging and facilitating the research in every way. Logistics and preparations at district and village level were coordinated in a timely and thorough fashion by Dr'âsho Sangye Dorji, the Honourable Secretary of Dzongkha Development Authority, and by Tshewang Dorji, the Chief Research Officer. Much gratitude is due to all the village headmen and district officers who assisted the research team on site. Particular gratitude is due to our friends 'Adap Dôji and Seta of the Lhokpu language community at Loto Kucu, Karma Chen and Dhani Ram Toto at Phüntsho'ling, Tandri and friends at Riti in the Black Mountains, 'Ap Drakpa of Phajong Pam and friends in the Gongduk language community. Furthermore, we thank Cecila Keijzer, Peter Newsum and Yanchen Doma of SNV (Netherlands Development Organisation), Bhutan. Special thanks are due to our old friend Karma Tshering of Gaselô, who really made the work happen. Karma involved himself and made himself indispensable at every level, anticipated and prevented all possible difficulties, and facilitated every aspect of the research programme with unparalleled perspicacity.

Finally we would like to thank mr. H.N.W. van Gent of the Netherlands Ministry of Foreign Affairs in The Hague for his assistance in sending the materials to Nepal and Bhutan.

# Acknowledgements