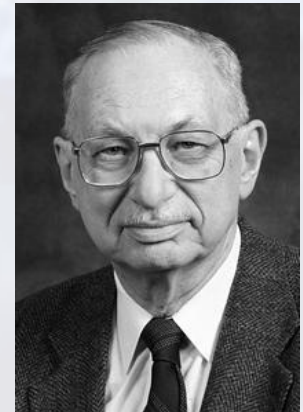
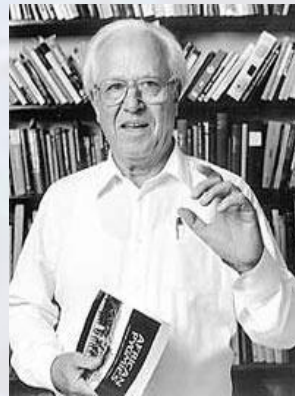
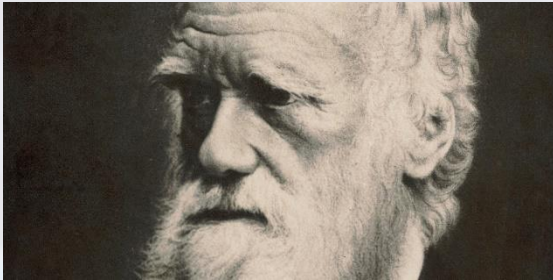


# Parametric history and population diversity



Andrea Ceolin, Dimitris Michelioudakis, Cristina Guardiano,  
Monica-Alexandrina Irimia, Giuseppe Longobardi, Nina Radkevich  
Mark Bartlett, Dimitar Kazakov, Andrea Brunelli  
Silvia Ghirotto, Alessio Boattini, Stefania Sarno  
Davide Pettener, Guido Barbujani, Donata Luiselli

Research in human populations has addressed the parallels between cultural and biological diversification since Darwin (1859), Cavalli Sforza et al. (1988), Sokal (1988)



A positive answer would allow researchers to support historical hypotheses using evidence coming from **two different domains**

# Creanza et al. (2015)



## A comparison of worldwide phonemic and genetic variation in human populations

Nicole Creanza<sup>a</sup>, Merritt Ruhlen<sup>b</sup>, Trevor J. Pemberton<sup>c</sup>, Noah A. Rosenberg<sup>a</sup>, Marcus W. Feldman<sup>a,1</sup>, and Sohini Ramachandran<sup>d,e,1</sup>

<sup>a</sup>Department of Biology and <sup>b</sup>Department of Anthropology, Stanford University, Stanford, CA 94305; <sup>c</sup>Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada R3E 0J9; and <sup>d</sup>Department of Ecology and Evolutionary Biology and <sup>e</sup>Center for Computational Molecular Biology, Brown University, Providence, RI 02912

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2013.

Contributed by Marcus W. Feldman, December 17, 2014 (sent for review July 16, 2014; reviewed by Quentin D. Atkinson and Keith Hunley)

**Worldwide patterns of genetic variation are driven by human demographic history. Here, we test whether this demographic history has left similar signatures on phonemes—sound units that distinguish meaning between words in languages—to those it has left on genes. We analyze, jointly and in parallel, phoneme inventories from 2,082 worldwide languages and microsatellite polymorphisms from 246 worldwide populations. On a global scale, both**

**compares the signatures of human demographic history in microsatellite polymorphisms from 246 worldwide populations (20) and complete sets of phonemes (phoneme inventories) for 2,082 languages; these are the largest available datasets of both genotyped populations and phonemes, the smallest units of sound that can distinguish meaning between words. Languages do not hold information about deep ancestry as genes do, and phoneme evolution is complex: phonemes can be transmitted**

# A comparison of worldwide phonemic and genetic variation in human populations

Nicole Creanza<sup>a</sup>, Merritt Ruhlen<sup>b</sup>, Trevor J. Pemberton<sup>c</sup>, Noah A. Rosenberg<sup>d</sup>, Marcus W. Feldman<sup>a,1</sup>, and Sohini Ramachandran<sup>a,c,1</sup>

<sup>a</sup>Department of Biology and <sup>b</sup>Department of Anthropology, Stanford University, Stanford, CA 94305; <sup>c</sup>Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada R3E 0R6; and <sup>d</sup>Department of Ecology and Evolutionary Biology and <sup>e</sup>Center for Computational Molecular Biology, Brown University, Providence, RI 02912

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2013.

Contributed by Marcus W. Feldman, December 17, 2014 (sent for review July 6, 2014; reviewed by Quentin D. Atkinson and Keith Hartley)

Worldwide patterns of genetic variation are driven by human demographic history. Here, we test whether this demographic history has left similar signatures on phonemes—sound units that distinguish meaning between words in languages—do those it has left on genes. We analyze, jointly and in parallel, phoneme inventories from 2,082 worldwide languages and microsatellite polymorphisms from 246 worldwide populations. On a global scale, both compares the signatures of human demographic history in microsatellite polymorphisms from 246 worldwide populations (20) and complete sets of phonemes (phoneme inventories) for 2,082 languages; these are the largest available datasets of both genotyped populations and phonemes, the smallest units of sound that can distinguish meaning between words. Languages do not hold information about deep ancestry as genes do, and

Languages are represented through binary characters, which code the absence/presence of phonemes:

	English	French	Japanese
/x/	1	0	0
/h/	1	0	1
/p/	0	1	1



## A comparison of worldwide phonemic and genetic variation in human populations

Nicole Creanza<sup>a</sup>, Merritt Ruhlen<sup>b</sup>, Trevor J. Pemberton<sup>c</sup>, Noah A. Rosenberg<sup>d</sup>, Marcus W. Feldman<sup>a,1</sup>, and Sohini Ramachandran<sup>a,1</sup>

<sup>a</sup>Department of Biology and <sup>b</sup>Department of Anthropology, Stanford University, Stanford, CA 94305; <sup>c</sup>Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada R3E 0V6; and <sup>d</sup>Department of Ecology and Evolutionary Biology and <sup>e</sup>Center for Computational Molecular Biology, Brown University, Providence, RI 02912

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2013.

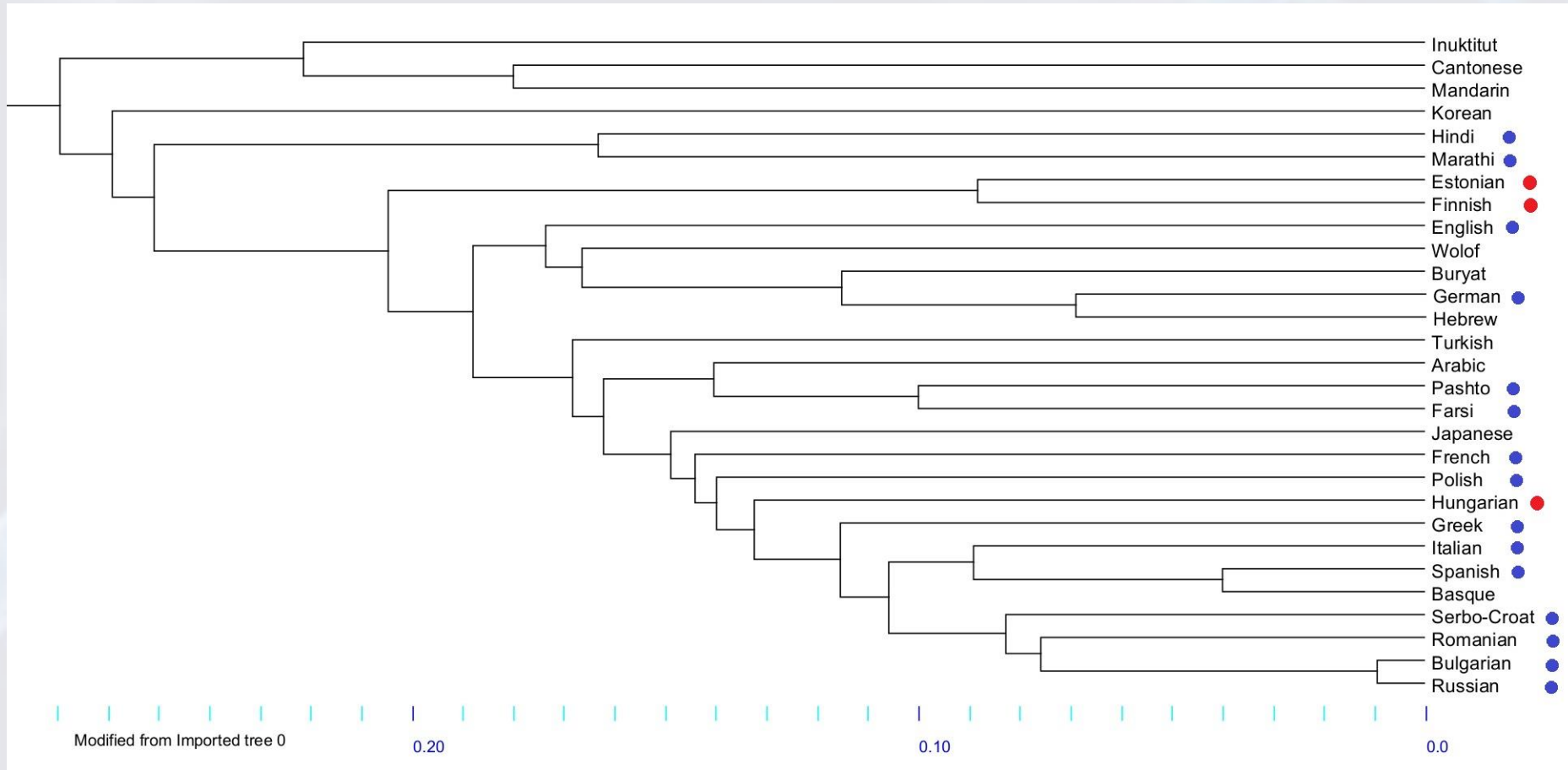
Contributed by Marcus W. Feldman, December 17, 2014 (sent for review July 6, 2014; received by Quentin D. Atkinson and Keith Hartley)

Worldwide patterns of genetic variation are driven by human demographic history. Here, we test whether this demographic history has left similar signatures on phonemes—sound units that distinguish meaning between words in languages—so those it has left on genes. We analyze, jointly and in parallel, phoneme inventories from 2,082 worldwide languages and microsatellite polymorphisms from 246 worldwide populations. On a global scale, both compares the signatures of human demographic history in microsatellite polymorphisms from 246 worldwide populations (20) and complete sets of phonemes (phoneme inventories) for 2,082 languages; these are the largest available datasets of both genotyped populations and phonemes, the smallest units of sound that can distinguish meaning between words. Languages do not hold information about deep ancestry as genes do, and

The major conclusions of the paper are:

- 1) Correlation between genes and languages is strong worldwide, but it is entirely predictable from geography (Partial Mantel Test:  $R = 0.05$ ,  $p = 0.16\text{--}0.17$ )
- 2) Geographical isolation leads to an increase in phonemic inventory sizes (vs. genetic drift)
- 3) High density areas display similarity in phonemic inventories between languages, even if they do not belong to the same family

Distance-based trees  
KITSCH (Phylip package)  
Felsenstein (2004)



Tree calculated from a sample of the  
Ruhlen phonemic database

## A comparison of worldwide phonemic and genetic variation in human populations

Nicole Creanza<sup>a</sup>, Merritt Ruhlen<sup>b</sup>, Trevor J. Pemberton<sup>c</sup>, Noah A. Rosenberg<sup>d</sup>, Marcus W. Feldman<sup>a,1</sup>, and Sohini Ramachandran<sup>a,b,1</sup>

<sup>a</sup>Department of Biology and <sup>b</sup>Department of Anthropology, Stanford University, Stanford, CA 94305; <sup>c</sup>Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada R3E 0Z6; and <sup>d</sup>Department of Ecology and Evolutionary Biology and <sup>e</sup>Center for Computational Molecular Biology, Brown University, Providence, RI 02912

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2013.

Contributed by Marcus W. Feldman, December 17, 2014 (sent for review July 6, 2014; reviewed by Quentin D. Atkinson and Keith Hartley)

Worldwide patterns of genetic variation are driven by human demographic history. Here, we test whether this demographic history has left similar signatures on phonemes—sound units that distinguish meaning between words in languages—so those it has left on genes. We analyze, jointly and in parallel, phoneme inventories from 2,082 worldwide languages and microsatellite polymorphisms from 246 worldwide populations. On a global scale, both compares the signatures of human demographic history in microsatellite polymorphisms from 246 worldwide populations (20) and complete sets of phonemes (phoneme inventories) for 2,082 languages; these are the largest available datasets of both genotyped populations and phonemes, the smallest units of sound that can distinguish meaning between words. Languages do not hold information about deep ancestry as genes do, and

“This suggests that phoneme inventories are affected by recent population processes and thus carry little information about the distant past”

(Creanza et al. 2015:1269)

Is there any other linguistic domain that can be used to study historical relationships at a global scale?



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)



Lingua

Lingua 119 (2009) 1679–1706

[www.elsevier.com/locate/lingua](http://www.elsevier.com/locate/lingua)

## Evidence for syntax as a signal of historical relatedness

Giuseppe Longobardi <sup>a,\*</sup>, Cristina Guardiano <sup>b</sup>

<sup>a</sup> *Laboratorio di Linguistica e antropologia cognitiva, DSA, Università di Trieste, Italy*

<sup>b</sup> *Dipartimento di Scienze del Linguaggio e della Cultura, Università di Modena e Reggio Emilia, Italy*

Received 15 January 2007; received in revised form 9 September 2008; accepted 9 September 2008

Available online 7 January 2009

Guardiano and Longobardi (2005) and Longobardi and Guardiano (2009) propose to look at parametric **Syntax** (Parametric Comparison Method, PCM)



# The syntax of the Nominal Domain (DP): 75 binary parameters (Guardiano and Longobardi 2016)

***Crosslinguistic morphosyntactic difference > parameter***

**if and only if** it entails

- (A) the presence of **obligatory formal expression** for a semantic or morphological distinction (*grammaticalisation*, i.e. the obligatory presence of a feature in the computation to obtain the relevant interpretation and its coupling with an uninterpretable counterpart)
- (B) the **variable form of a category** depending on the syntactic context (selection and feature agreement)
- (C) the **position of a category** (movement,  $\pm$ overt attraction triggered by grammaticalised features)
- (D) The **availability** in the lexicon of certain functional categories (e.g. functional genitive projections)

# The syntax of the Nominal Domain (DP): 75 binary parameters (Guardiano and Longobardi 2016)

## DP-subdomains:

- (1) the status of **features associated with D**, e.g. *person, number, gender, and definiteness*
- (2) the syntactic properties of **noun modifiers** (adjectives and relative clauses), **adnominal/genitival** arguments and **possessives, demonstratives**
- (3) the type and scope of '**N-movement**' along the functional spine of the nominal.

# Crossparametric Implications

**Languages are encoded as lists of binary parameters (+,-)**

Grammaticalized Person (FGP) and Strong Person (NSD)

	English	French	Chinese
FGP: gramm. person	+	+	-
NSD: strong person	-	+	?

# Crossparametric Implications

**Languages are encoded as lists of binary parameters (+,-)**

Grammaticalized Person (FGP) and Strong Person (NSD)

	<i>Conditions</i>	English	French	Chinese
<b>FGP</b> : gramm. person		+	+	-
NSD: strong person	<b>(+FGP)</b>	-	+	<b>0</b>

# Crossparametric Implications

Avoiding redundancy: A parameter is not assigned a value if the manifestations for that value are independently derivable.

Non-iterable pre-A genitive/GenS (GFS) and head marking with phi-agreement (GFN)

	<i>Conditions</i>	Italian	English	Hungarian
<b>GFN:</b> Gen-N agreement		-	-	+
<b>GFS:</b> GenS	<b>(-GFN)</b>	-	+	<b>0</b>



# TableA

It	-+---00+++++---0---+0--0000+++0+0+---+000-+-0--0+---00-+-
Sp	++---00+++++---0---+0+++++0+0+---+000-+-0--0+---00+++
Fr	++---00+-+---+0--0---+00-+-0+0+++0+0+---+000-+-0--0+---00+++
Ptg	++---00+++++---0---+0+-0000+++0+0+---+000-+-0--0+---00?+?
Rm	++---00+++++---0---+00+-+00+++0+0+---+000-+-0--0--0--0--0--
Grk	++---00+++++---0---+0-+-+---+0+0+---+0+---0--0-+-+0--0--0+
E	++---00+++++---0---+00-0000+-0+0+---+0+---0--0--0--0--0--
D	++---00+++++---0---+00-0000+-+0+0+---+0+---0--0--0--0--0--
Da	++---00+++++---+---+00-0000+++++0+---+0+---0--0--0--0--0--
Ice	++---00+++++---+---+00-0000+++++0+---+0+---0--0--0--0--0--
Nor	++---00+++++---+---+00-0000+++++0+---+0+---0--0--0--0--0--
Blg	++---00+++++---+0---+00-0000+++0+0+---+0+---0--0--0--0--0--
SC	++---00+-+---00-0000-+-0000+++0+0+0+---+0+---0--0--0--0--0--
Slo	++---00+-+---00-0000-+-0000+++0+0+0+---+0+---0--0--0--0--0--
Po	++---00+-+---00-0000-+-0000+++0+0+0+---+0+---0--0--0--0--0--
Rus	++---00+-+---00-0000-+-0000+++0+0+0+---+0+---0--0--0--0--0--
Ir	++---00+++++---0---+00+-0+---+0+---+0000-0--0--0--0--0--0--
Wel	++---00+++++---0---+00+-0+---+0+---+0000-0--0--0--0--0--0--
Ma	++---00+-+---00-0000-+-00+00+++0+---+0+---0--0--0--0--0--0--
Hi	++---00+-+---00-0000-+-00+00+++0+---+0+---0--0--0--0--0--0--
Pas	++---00+-+---00-0000--0000+++0+---0+---+0+---0--0--0--0--0--
Man	-0---++0000000000000000+0-0+++0-+000-0000-+-00+00-0-0-0-
Can	-0---++0000000000000000+0-0+++0-+000-0000-+-00+00-0-0-0-
Ar	++---00+++++---0+++0-0-+-+---+0+0+++000000+0--0+-0+++0--00--0-
Heb	++---00+++++---0---+0-+-0+++++0+0-++000000+0+-0+-0+++0--00--0-
Hu	++---00---+---+---0---+0---+00+-+0+0+---+0+---000000-0--0+0000
Est	++---00---+---+---000000---0000+++0+0+0+---+0+---0000-0--0+0-0-
Fin	++---00---+---+---000000---0000+++0+0+0+---+0+---000000-0--0+0000
Tur	++---00---+---+---000000---0-00+00+---0+---+00+00000000-0+0000
Bur	++---00---+---+---000000---0-00+00+---0+---+00+0--0+-0--0+-0+
cB	++---00-+00-0+0-000000-00+-0000+0+0+---+000000+---0--0+-00--000-0-
wB	++---00+00-0+0+000000-00+-0-0+0+0+---+000000+---0--0+-00--000-0-
Wo	++---00+00+-0+-0+00+0+-0000-0+0+000-000000-0--0-----+0+0+0?-0+

# Distances

How to choose a distance measure?

Since we have a lot of '0' values, we cannot simply “count” the number of differences (= Hamming distance)

We can use a **Jaccard-Tanimoto distance** between “comparable” values:

$$d(A,B) = \text{dif}(A,B) / [\text{dif}(A,B) + \text{id}(A,B)]$$

= differences / identities + differences

E.g.: Italian-English: (35 id., 6 diff.)  $d(\text{It}, \text{Eng}) = 6 / 41 = 0.146$

# Macro- and micro-classification

## ***Indo-European:***

Longobardi, G., Guardiano, C., Silvestri, G., Boattini, A., & Ceolin, A. (2013). Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics*, 3(1), 122-152.

## ***Greek and Romance micro-variation***

Guardiano, C., D. Michelioudakis, A. Ceolin, M. Irimia, G. Longobardi, N. Radkevic, G. Silvestri, A. Sitaridou (2016) South by SouthEast. A syntactic approach to Greek and Romance micro- variation. *L'Italia Dialettale*.

The classifications obtained largely match the results of well-established and sophisticated methods relying on phonology and vocabulary.

- extremely high correlation with distributions of distances in such classifications

Going beyond well-established families and beyond the historical depth of PIE, no other linguistic tools, e.g. etymology, can be used as benchmarks/standards of comparison.

- Do cross-family syntactic distances correlate with genetic distances? Is the correlation comparable to that of within-family distances?

# Genes and Languages in Europe (15 populations)

Research Article

**Across language families: Genome diversity mirrors linguistic variation within Europe**

Giuseppe Longobardi<sup>1,2</sup>, Silvia Ghirotto<sup>3</sup>, Cristina Guardiano<sup>4</sup>, Francesca Tassi<sup>3</sup>, Andrea Benazzo<sup>3</sup>, Andrea Ceolin<sup>1</sup> and Guido Barbujani<sup>3,\*</sup>

Article first published online: 8 JUN 2015  
DOI: 10.1002/ajpa.22758  
© 2015 Wiley Periodicals, Inc.

Issue

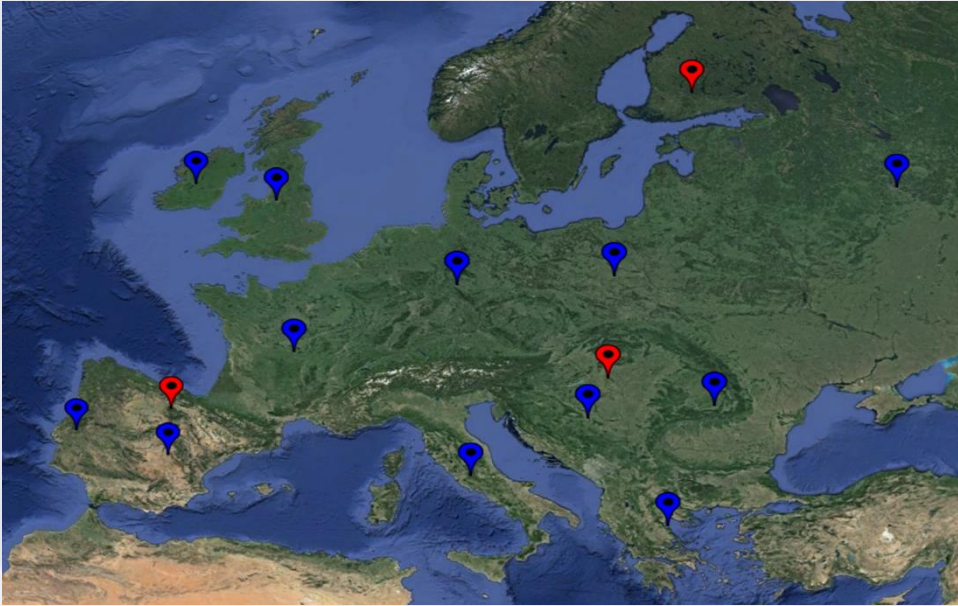


American Journal of Physical Anthropology  
Volume 157, Issue 4, pages 630–640, August 2015

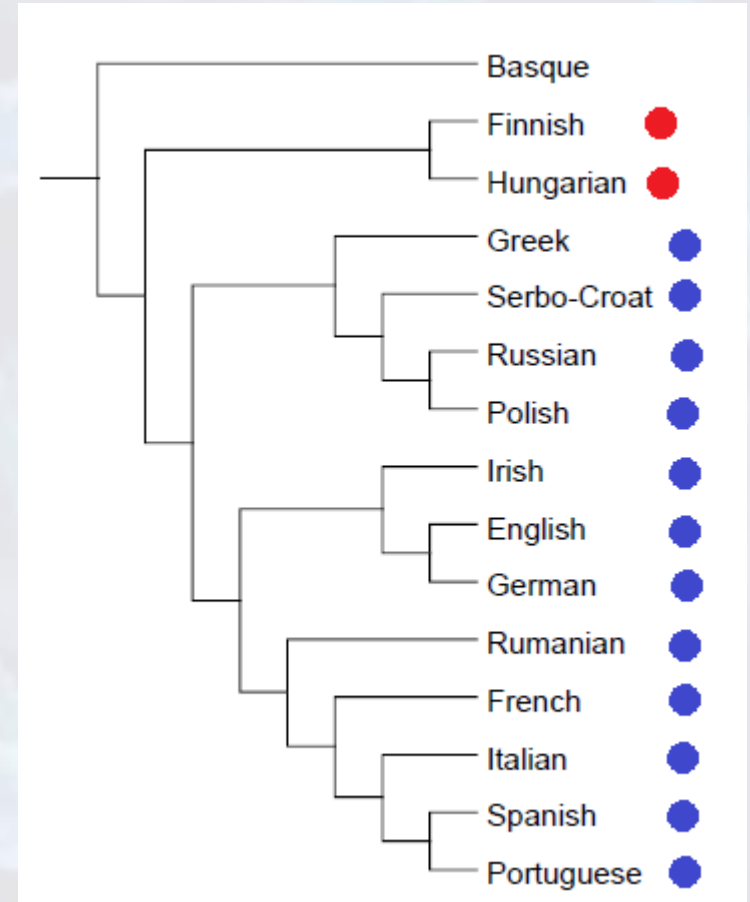
OPEN

12 IE populations and 3 non-IE populations (Basque, Hungarians, and Finns) are analyzed from the viewpoint of their syntactic, genetic and geographic distances

# Genes and Languages in Europe (15 populations)



Syntactic distances are distributed following historical patterns (contrary to the phonological data in Creanza et al. 2015)



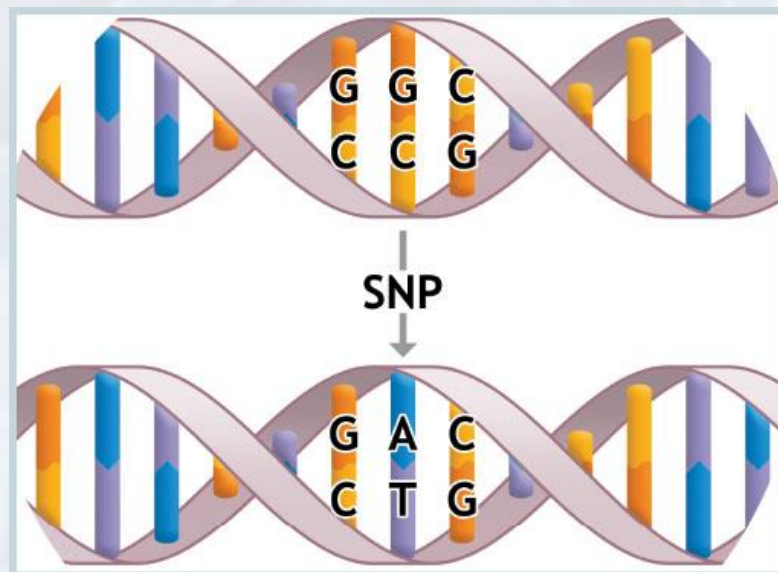
Tree from Longobardi et al. 2015



### The Population Reference Sample, POPRES: A Resource for Population, Disease, and Pharmacological Genetics Research

Matthew R. Nelson,<sup>1,\*</sup> Katarzyna Bryc,<sup>2</sup> Karen S. King,<sup>1</sup> Amit Indap,<sup>2</sup> Adam R. Boyko,<sup>2</sup>  
John Novembre,<sup>3,4</sup> Linda P. Briley,<sup>1</sup> Yuka Maruyama,<sup>1</sup> Dawn M. Waterworth,<sup>5</sup> Gérard Waeber,<sup>6</sup>  
Peter Vollenweider,<sup>6</sup> Jorge R. Oksenberg,<sup>7</sup> Stephen L. Hauser,<sup>7</sup> Heide A. Stirnadel,<sup>8</sup> Jaspal S. Kooner,<sup>9</sup>  
John C. Chambers,<sup>10</sup> Brendan Jones,<sup>1</sup> Vincent Mooser,<sup>5</sup> Carlos D. Bustamante,<sup>2</sup> Allen D. Roses,<sup>1</sup>  
Daniel K. Burns,<sup>1</sup> Margaret G. Ehm,<sup>1</sup> and Eric H. Lai<sup>1</sup>

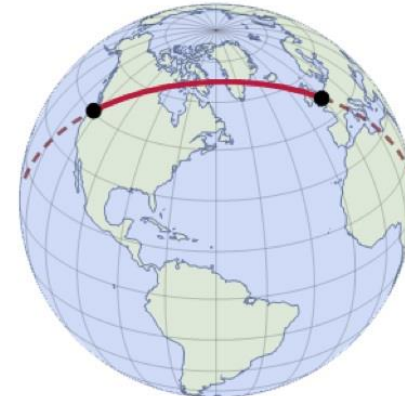
5,886 subjects genotyped at 500,568 loci using the Affymetrix 500K single nucleotide polymorphism



## Geographic distances

Great Circle Distances (the shortest distance between two points on the surface of a sphere)

	Basque	England	Finland	France	Germany	Greece	Hungary	Ireland	Italy	Poland	Portugal	Romania	Russia	Ser_Cro	Spain
Basque	0														
England	1187.32	0													
Finland	3598.33	3116.53	0												
France	703.27	930.48	2900.12	0											
Germany	1720.55	1441.89	1879.06	1021.06	0										
Greece	2842.08	3148.58	2070.09	2370.7	1808.53	0									
Hungary	2504.1	2515.3	1472.55	1894.75	1086.01	833.19	0								
Ireland	1301.76	652.09	3760.08	1386.19	2081.45	3726.67	3140.04	0							
Italy	1692.39	2083.82	2359.24	1234.44	1019.94	1155.1	948.34	2614.46	0						
Poland	2659.05	2423.91	1035.1	1986.03	1007.66	1309.18	514.5	3072.68	1325.44	0					
Portugal	693.88	1665.57	4292.19	1395.39	2414.33	3420.15	3164.26	1498.94	2299.75	3346.22	0				
Romania	3099.24	3151.21	1442.74	2518.15	1715.71	652.92	637.63	3777.63	1450.7	892.46	3738.87	0			
Russia	4669.32	4440.39	1463.4	4020.28	3053.41	2236.19	2181.1	5092.48	3103.42	2046.74	5340.66	1678.13	0		
Ser_Cro	2594.25	2759.21	1739.13	2048.29	1372.57	466.32	370.15	3361.71	924.19	868.04	3220.66	531.06	2204.72	0	
Spain	298.17	1473.97	3830.7	965.2	1962.6	2924.66	2666.35	1517.52	1798.83	2864.87	501.42	3237.63	4845.04	2719.27	0



## 12 IE populations + Basque, Hungarians, and Finns

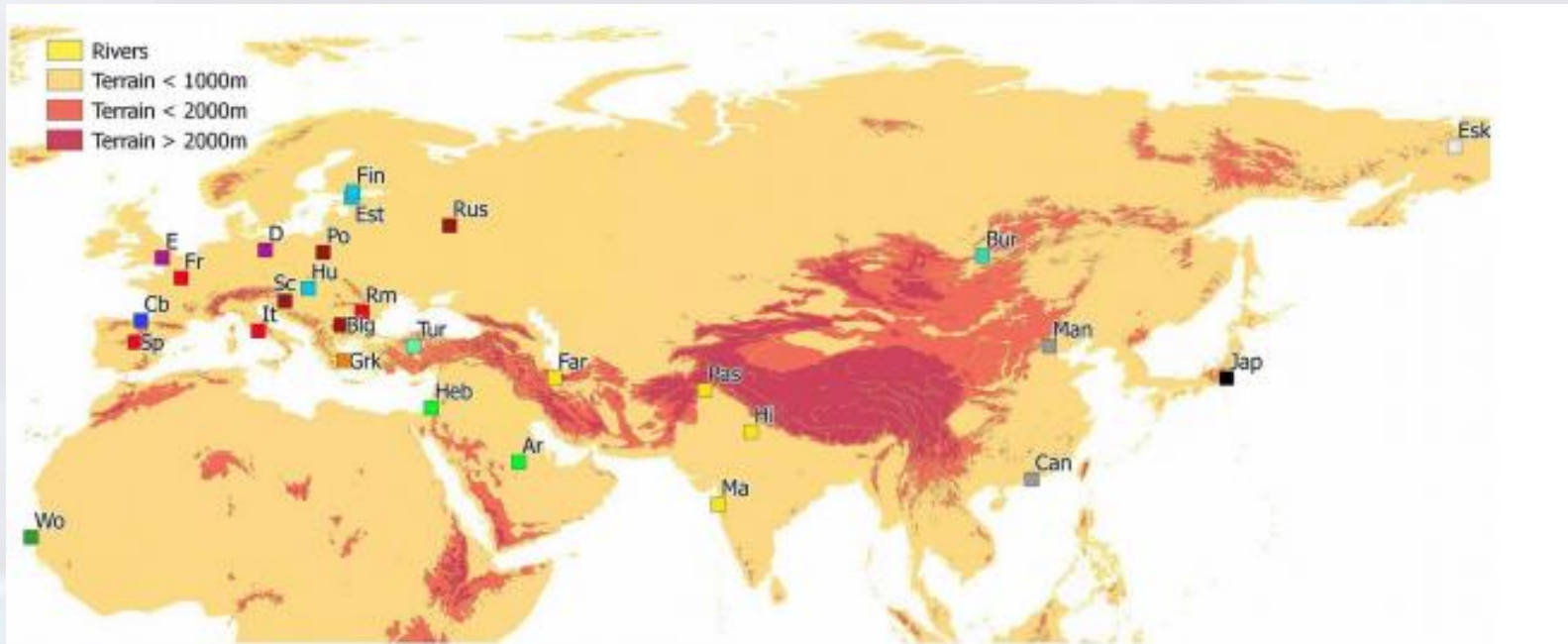
Distance matrices	<i>r</i>	<i>P</i>
$d_{\text{GEN}}$ $d_{\text{GEO}}$ Genetic - Geographic	0.299	0.030
$d_{\text{SYN}}$ $d_{\text{GEO}}$ Syntactic - Geographic	0.240	0.039
$d_{\text{SYN}}$ $d_{\text{GEN}}$ Syntactic - Genetic	0.599	0.001
$d_{\text{SYN}}$ $d_{\text{GEN}}$ ( $d_{\text{GEO}}$ ) Syntactic - Genetic (Geography held constant)	0.570	0.002

## An exception: Hungarian

“Careful analyses of 10th century ancient DNA in Hungary showed a predominance of European mitochondrial haplotypes in burials attributed to the lower classes, and a high incidence of Asian haplotypes in high-status individuals of that period (Tömöry et al. 2007), which points to the Asian immigrants as representing a social élite [...]

[...] when a Finno-Ugric language was introduced in Hungary, the genetic buildup of the population changed only in part, thus retaining similarities with its geographic neighbors, an example of the process called **élite dominance** by Renfrew (1992).”

## Next step: Eurasia (28 languages)



**Indo-European (15)**

**Finno-Ugric (3)**

**Altaic (2)**

**Semitic (2)**

**Sinitic (2)**

**Niger-Congo (1)**

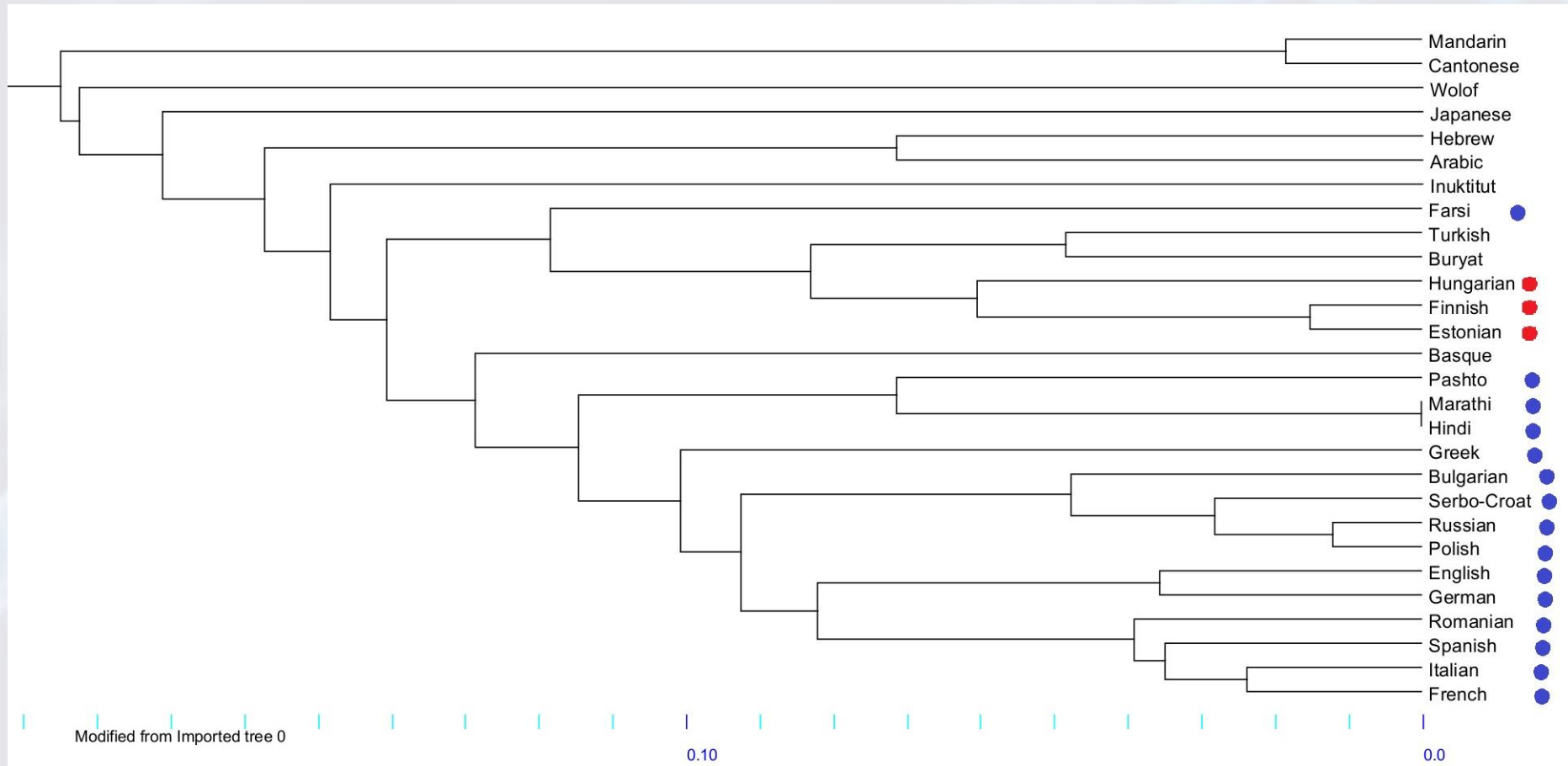
**Basque (1)**

**Japanese (1)**

**Inuit (1)**



Distance-based trees  
KITSCH (Phylip package)  
Felsenstein (2004)



## Correlations in Eurasia: 28 populations

Distance matrices	<i>r</i>	<i>P</i>
$d_{\text{GEN}}$ $d_{\text{GEO}}$ Genetic - Geographic	0.8319	0.0001
$d_{\text{SYN}}$ $d_{\text{GEO}}$ Syntactic - Geographic	0.4669	0.0001
$d_{\text{SYN}}$ $d_{\text{GEN}}$ Syntactic - Genetic	0.5286	0.0001
$d_{\text{SYN}}$ $d_{\text{GEN}}$ ( $d_{\text{GEO}}$ ) Syntactic - Genetic (Geography held constant)	0.2857	0.0036

## Syntactic and phonemic evidence in correlation with genes

	This study	Creanza et al. (PNAS 2015)	
	<i>Syntactic</i>	<i>Phonemic</i> (Ruhlen)	<i>Phonemic</i> (PHOIBLE)
$d_{\text{Gen}} - d_{\text{Lin}}$	<b>0.529</b> p=0.0001	<b>0.157</b> p=0.002	<b>0.240</b> p=0.0002
$d_{\text{Gen}} - d_{\text{Lin(Geo)}}$	<b>0.2857</b> p=0.0036	<b>0.05</b> p=0.16	<b>0.05</b> p=0.17

# Syntactic and phonemic evidence in correlation with genes (Eurasia)

## This study

### *Syntactic*

$d_{\text{Gen}} - d_{\text{Lin}}$

**0.529**

$p=0.0001$

$d_{\text{Gen}} - d_{\text{Lin(Geo)}}$

**0.2857**

$p=0.0036$

## Creanza et al. (PNAS 2015)

### *Phonemic (Ruhlen)*

**0.4232**

$p=0.005$

**0.0359**

$p=0.3344$

# Modeling geography

Great Circle Distances (GCD) are the standard measures in correlation studies

Can we test models closer to reality? We have four different models:

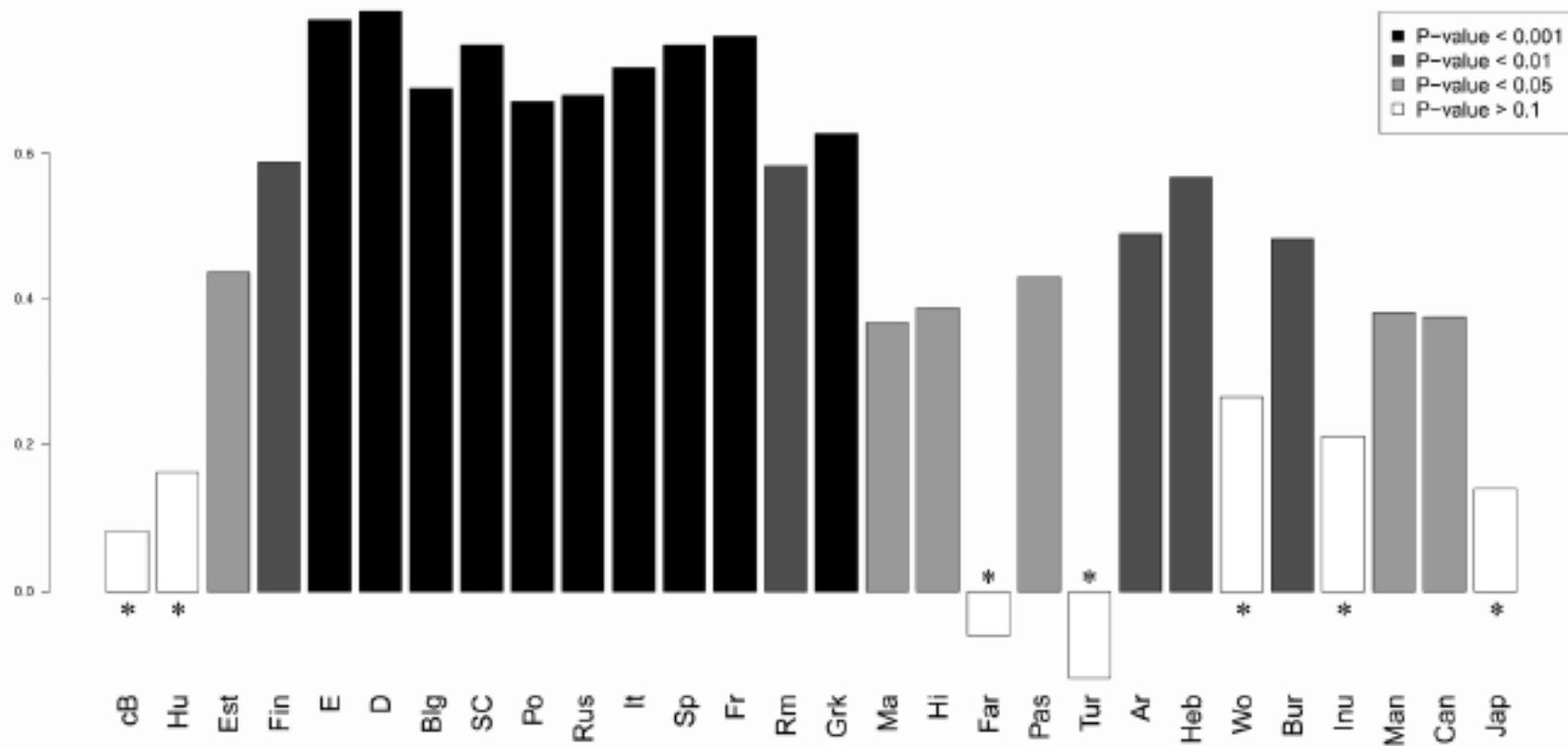
- 1- GCD with WayPoints
- 2- Road Maps
- 3- Least Cost Path
- 4- Resistance



## Correlations in Eurasia: 28 populations

Distance matrices	<i>r</i>	<i>P</i>
$d_{\text{SYN}}$ $d_{\text{GEN}}$ ( $d_{\text{GEO}}$ ) (GCD WayPoints)	0.2770	0.0063
$d_{\text{SYN}}$ $d_{\text{GEN}}$ ( $d_{\text{GEO}}$ ) (RoadMaps)	0.2641	0.0082
$d_{\text{SYN}}$ $d_{\text{GEN}}$ ( $d_{\text{GEO}}$ ) (Least Cost Path)	0.3049	0.0030
$d_{\text{SYN}}$ $d_{\text{GEN}}$ ( $d_{\text{GEO}}$ ) (Resistance)	0.3508	0.0011

# Partial correlations



# Exceptions

Wolof: it is more salient as an outlier genetically rather than linguistically. This can derive from grammatical variation being more constrained (by UG?)

Inuktitut: likely to be an insufficient sampling approximation: the language is spoken in Eastern Canada, while the nearest genetic proxy available was in North-Eastern Asia

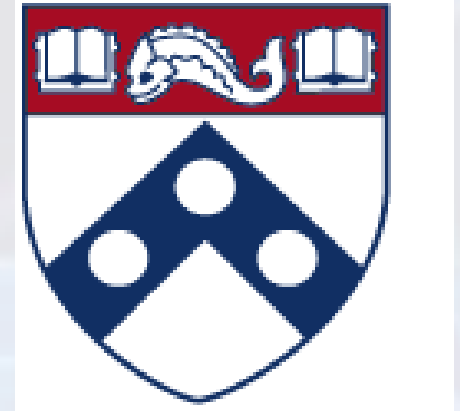
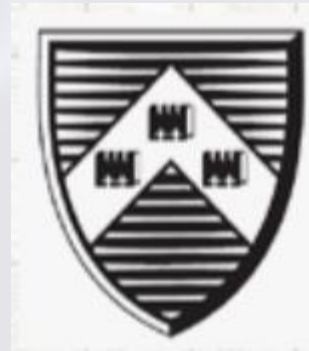
# Exceptions

Hungarian is still an exception, as it was in Longobardi et al. (2015)

Turkish, Farsi, Basque, Japanese can all be explained in terms of **élite dominance** (like Hungarian) and related demographic processes

# Conclusions

- Generative Syntax is a powerful tool to classify languages and investigate their history. It is only marginally affected by horizontal transmission and it allows to reach comparison at a global scale
- Languages and genes seem to follow the same axes of variation **independently** of geography (vs. Creanza et al. 2015)
- A single process (**élite dominance**, Renfrew 1992) can explain several cases of mismatch between linguistic and genetic variation
- These findings might provide new insights to study historical migrations in Eurasia and, potentially, in other continents



THANKS!

Langelin – LANguage GENes LINeages  
ERC Advanced Grant n. 295733

