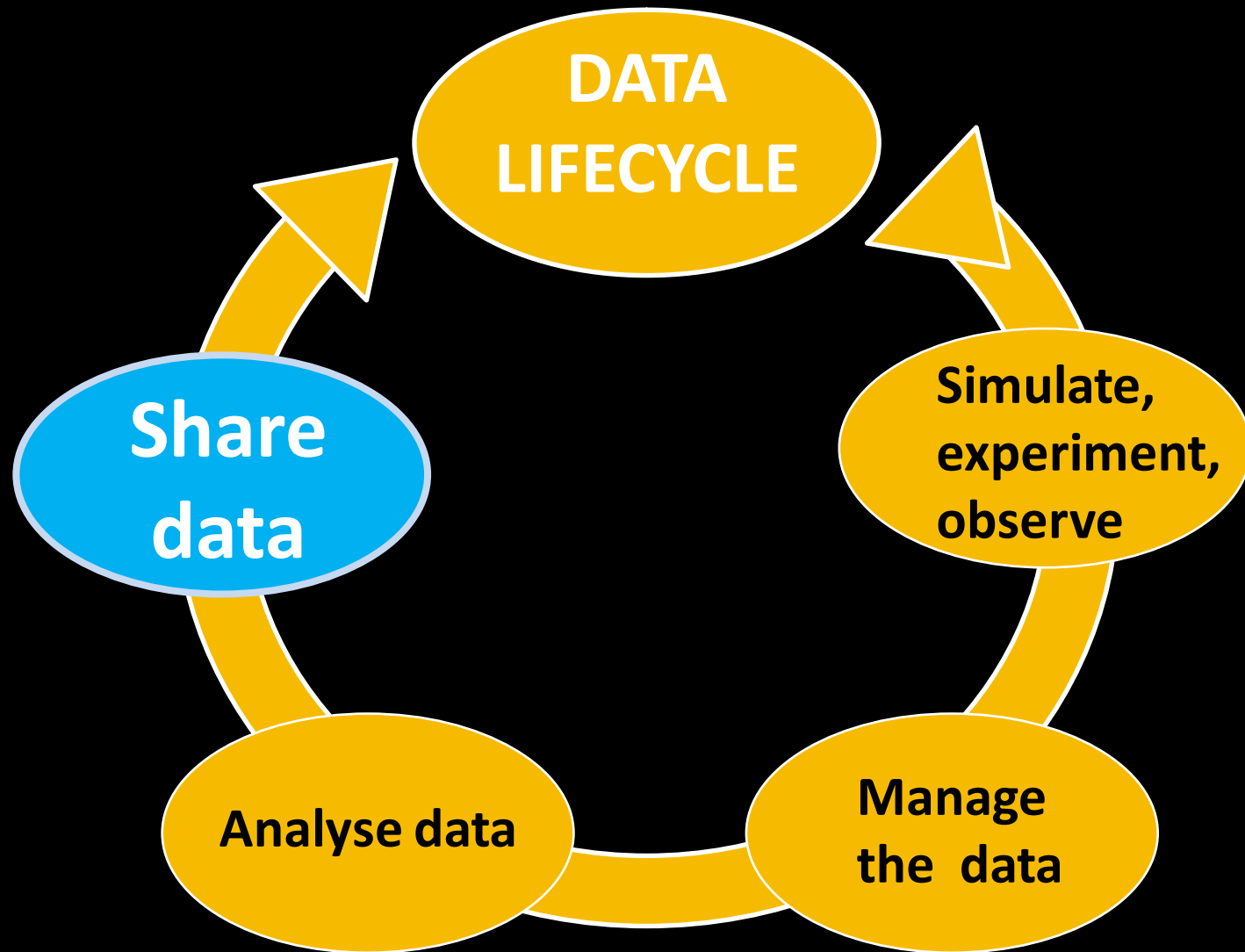


The culture of data sharing

Giovanni Destro Bisol

Sapienza Università di Roma

Istituto Italiano di Antropologia

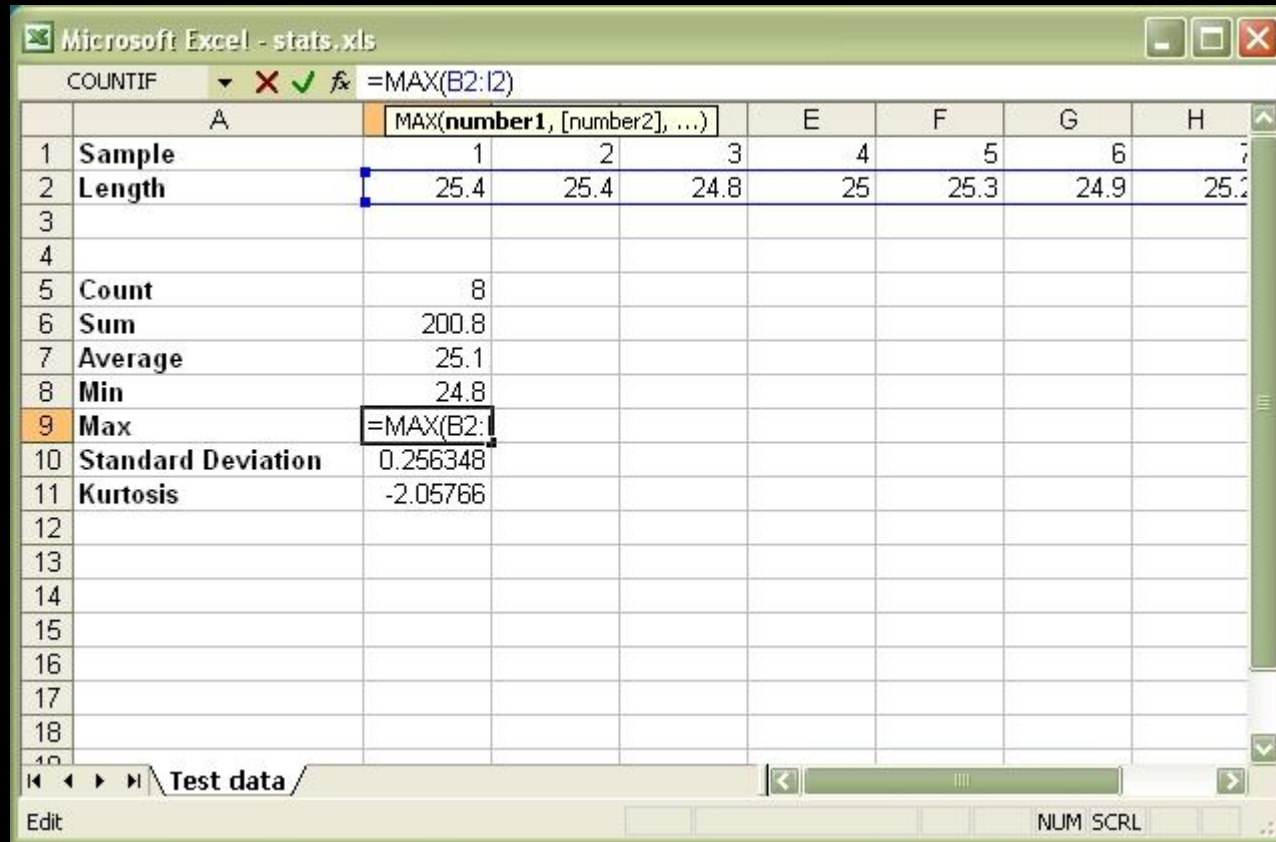




First...

get some clear ideas about...

...this isn't "data sharing"



Microsoft Excel - stats.xls

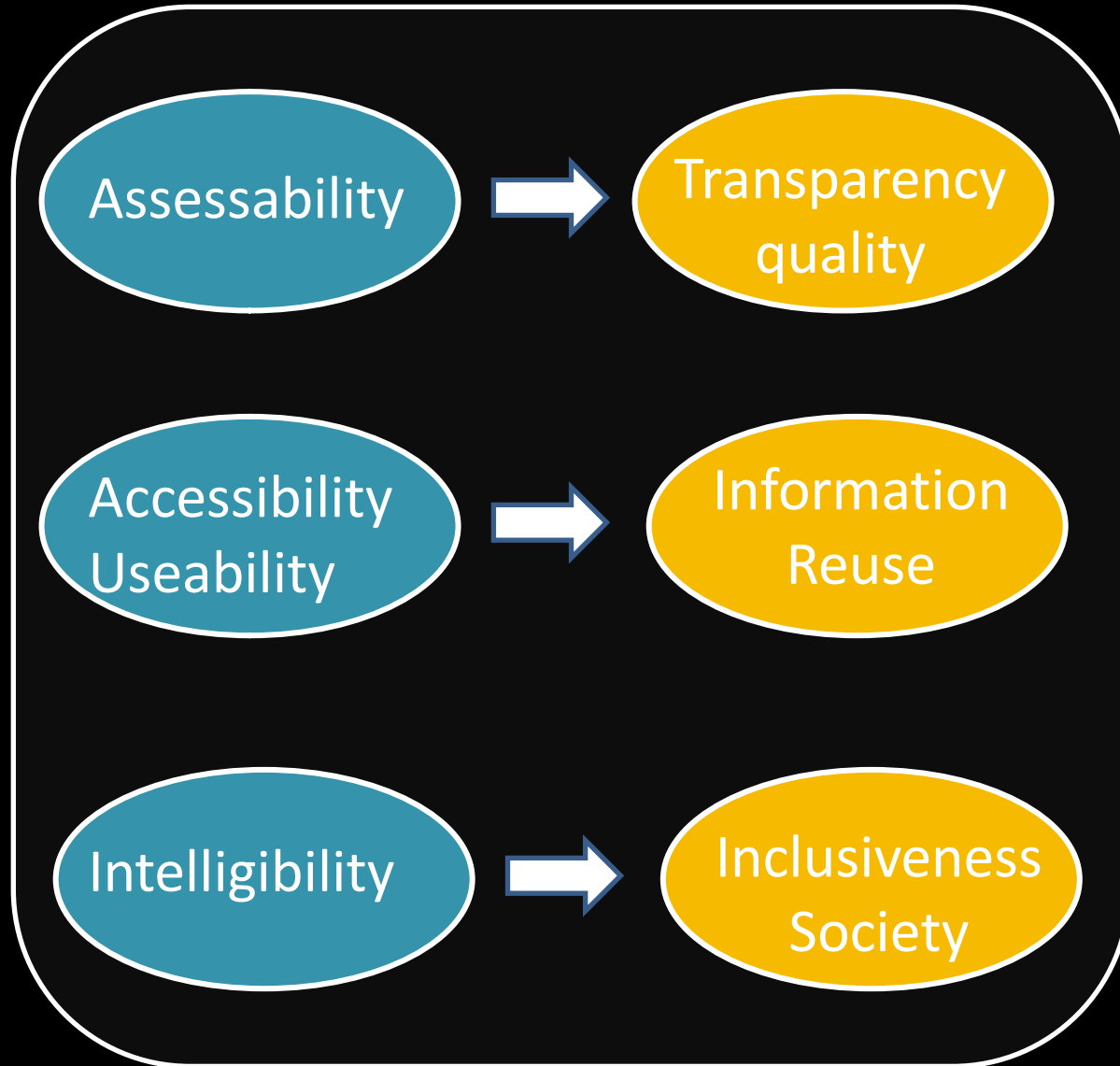
COUNTIF ✖ ✔ fx =MAX(B2:I2)

	A	B	C	D	E	F	G	H
1	Sample	1	2	3	4	5	6	7
2	Length	25.4	25.4	24.8	25	25.3	24.9	25.2
3								
4								
5	Count	8						
6	Sum	200.8						
7	Average	25.1						
8	Min	24.8						
9	Max	=MAX(B2:I2)						
10	Standard Deviation	0.256348						
11	Kurtosis	-2.05766						
12								
13								
14								
15								
16								
17								
18								
19								

Test data

Edit NUM SCRL

.. this is “data sharing”

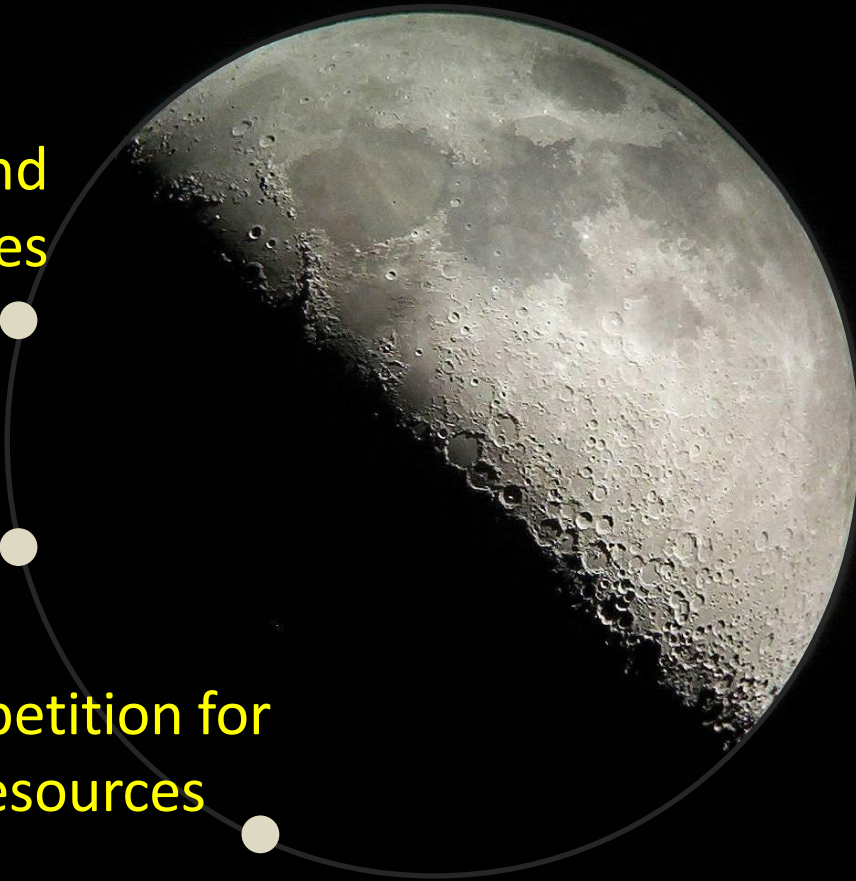


“data sharing” has cons...

Ethical and
legal issues

Misuse

Competition for
resources



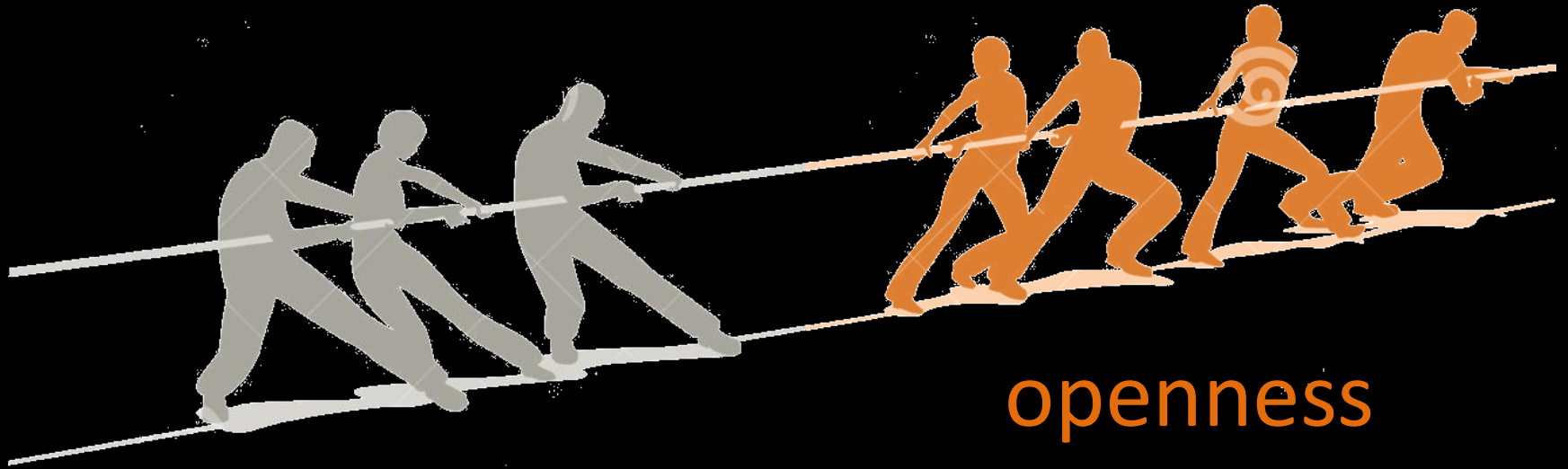
but pros outweigh...



.....the third view

tension

secrecy



openness



The round table model.

Soranno et al., *Bioscience*, 2015.

doi: 10.1093/biosci/biu169

Second

... Look at “classic” examples

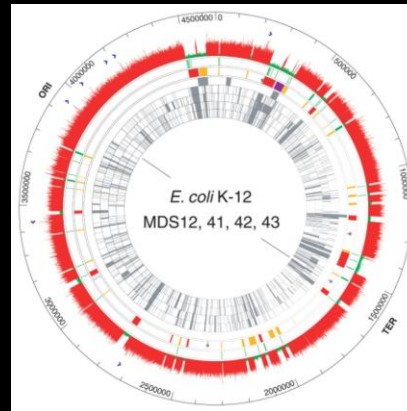
sharing saves lives!



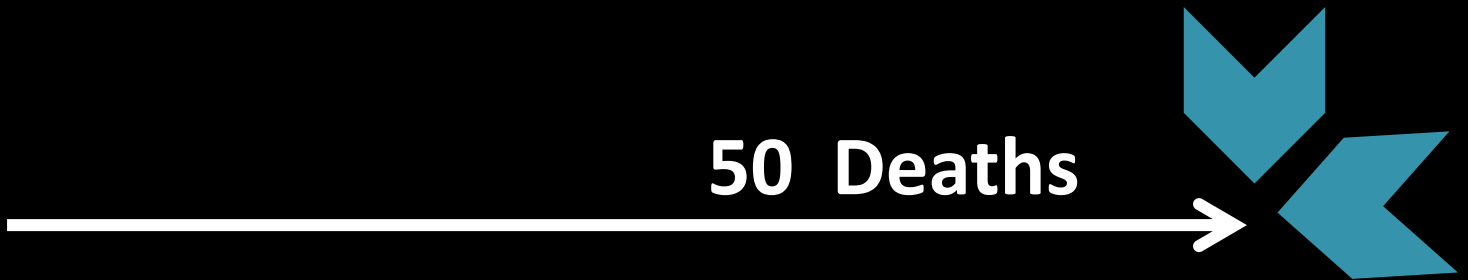
3 days ➤

4 days ➤

7 days



50 Deaths



sharing helps detect errors!



A screenshot of the Cornell University Library arXiv.org website. At the top, the Cornell University Library logo is on the left, and the text "Cornell University Library" is on the right. Below this is a red banner with "arXiv.org" in white. Under the banner, it says "Open access to 547,440 e-prints in Physics, Mathematics, and Engineering". Below that is a search bar with "Physics" selected. A date update "8 Apr 2009: Added public author identifiers, Facebook..." is visible. A warning "Robots Beware: indiscriminate automated downloads" is present. A "Physics" section header is followed by a list of sub-fields: Astrophysics (astro-ph), Condensed Matter (cond-mat), General Relativity and Quantum Cosmology (gr-qc), High Energy Physics - Experiment (hep-ex), High Energy Physics - Lattice (hep-lat), High Energy Physics - Phenomenology (hep-ph), and High Energy Physics - Theory (hep-th).

Third

... and your own field?

Le Scienze

Settembre 2014

€ 4,50

www.lescienze.it

edizione italiana di Scientific American

Geni italiani

La nostra diversità genetica è la più varia d'Europa: un viaggio scientifico tra DNA e cultura

Astrofisica

Ipotesi sull'origine delle misteriose «bolle di Fermi» nella Via Lattea

Neuroscienze

L'inaspettato legame tra neurogenesi e disturbi d'ansia

doi 10.4436/JASS.92001

JASS Reports

Journal of Anthropological Sciences

Vol. 92 (2014), pp. 201-231

Linguistic, geographic and genetic isolation: a collaborative study of Italian populations

Marco Capocasa^{2,3}, Paolo Anagnostou^{1,2}, Valeria Bachis⁴, Cinzia Battaggia¹, Stefania Bertoncini⁵, Gianfranco Biondi⁶, Alessio Boattini⁷, Iliaria Boschi⁸, Francesca Brisighelli^{1,9}, Carla Maria Calò⁴, Marilisa Carta⁷, Valentina Coia^{10,11}, Laura Corrias⁴, Federica Crivellaro¹², Sara De Fanti⁷, Valentina Dominici^{1,2}, Gianmarco Ferri¹³, Paolo Francalacci¹⁴, Zeldia Alice Franceschi¹⁵, Donata Luiselli⁷, Laura Morelli¹⁴, Giorgio Paoli⁵, Olga Rickards¹⁶, Renato Robledo¹⁷, Daria Sanna¹⁴, Emanuele Sanna⁴, Stefania Sarno⁷, Luca Sineo¹⁸, Luca Taglioli⁵, Giuseppe Tagarelli¹⁹, Sergio Tofanelli⁵, Giuseppe Vona⁴, Davide Pettener⁷ & Giovanni Destro Bisol^{1,2}

Summary - The animal and plant biodiversity of the Italian territory is known to be one of the richest in the Mediterranean basin and Europe as a whole, but does the genetic diversity of extant human populations show a comparable pattern? According to a number of studies, the genetic structure of Italian populations retains the signatures of complex peopling processes which took place from the Paleolithic to modern era. Although the observed patterns highlight a remarkable degree of genetic heterogeneity, they do not, however, take into account an important source of variation. In fact, Italy is home to numerous ethno-linguistic minorities which have yet to be studied systematically. Due to their difference in geographical origin and demographic history, such groups not only signal the cultural and social diversity of our country, but they are also potential contributors to its bio-anthropological heterogeneity. To fill this gap, research groups from four Italian Universities (Bologna, Cagliari, Pisa and Roma Sapienza) started a collaborative study in 2007, which was funded by the Italian Ministry of Education, University and Research and received partial support by the Istituto Italiano di Antropologia. In this paper, we present an account of the results obtained in the course of this initiative. Four case-studies relative to linguistic minorities from the Eastern Alps, Sardinia, Apennines and Southern Italy are first described and discussed, focusing on their micro-evolutionary and anthropological implications. Thereafter, we present the results of a systematic analysis of the relations between linguistic, geographic and genetic isolation. Integrating the data obtained in the course of the long-term study with literature and unpublished results on Italian populations, we show that a combination of linguistic and geographic factors is probably responsible for the presence of the most robust signatures of genetic isolation. Finally, we evaluate the magnitude of the diversity of Italian populations in the European context. The human genetic diversity of our country was found to be greater than observed throughout the continent at short (0-200 km) and intermediate (700-800km) distances, and accounted for most of the highest values of genetic distances observed at all geographic ranges. Interestingly, an important contribution to this pattern comes from the "linguistic islands" (e.g. German speaking groups of Sappada and Luserna from the Eastern Italian Alps), further proof of the importance of considering social and cultural factors when studying human genetic variation.

Keywords - Genetic structure, Linguistic diversity, Minority languages, Linguistic islands.

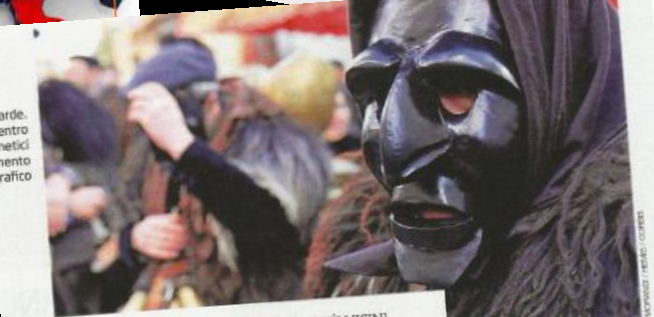
Gli italiani sono il popolo con la varietà genetica più ricca d'Europa

Estensione latitudinale dell'Italia, flussi migratori e biodiversità sono le principali ragioni di questa scoperta. Lo studio è stato condotto dai ricercatori dell'Università La Sapienza di Roma e degli atenei di Pisa, Cagliari e Bologna

di AGNESE FIORETTI



DAL NORD AL SUD DELL'ITALIA - ma in realtà anche solo della Sardegna - le nostre differenze genetiche sono dalle sette alle 30 volte maggiori rispetto a quelle registrate tra i portoghesi e ungheresi. Insomma, gli italiani sono il popolo con la più alta diversità genetica in Europa.



SPAGNOLI E RUMENI PER DNA RISULTANO PIÙ VICINI A CERTE NOSTRE COMUNITÀ NELLA STESSA REGIONE

DALLE ALPI ALLA SARDEGNA NOI ITALIANI SIAMO UN POPOLO DI BIODIVERSI

MACRO
 www.macro.it

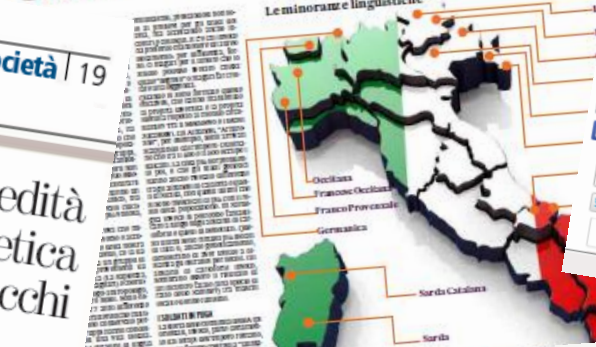
il libro **Donna Tartù e il successo** di "The Goldfinch" (intervista a pag. 34)

Televisione **Canale 5, da stasera il peccato e la vergogna** 2 (intervista a pag. 35)

Moda **La fantasia e l'originalità** di Pirelli Uomo e super Milano (intervista a pag. 33)

Letteratura **Gusto** **Ambiente** **Società** **Cinema** **Viaggi** **Architettura** **Tes** **Arte** **Moda** **Tecnologia** **Musica** **Scienza** **Archeologia** **Televisione** **Sa**

Italia, il Paese dei geni



Un'eredità genetica che ci fa ricchi

Marco Malvaldi
 SEGUE DALLA PRIMA PAGINA

Talvolta accoglievamo, talvolta schiavizzavamo, altre volte ci invadono italiani tutti, eredi di questo patrimonio genetico, non facciamo l'errore di fare sfoggio in modo molto italico di tale diadema: facendo vedere al resto d'Europa la nostra incredibile varietà genetica, che fu, mentre intanto il presente va a rovescio come si diceva di un glorioso tempo. Avere un patrimonio intrinseco così vasto è un'ottima base di partenza; però, a garantire ottime prospettive, vorrei rammentare un piccolo esempio.



Nel 2008, alcuni ricercatori annunciarono di aver scoperto il segreto della velocità degli sprinter giamaicani. Il nocciolo della scoperta risiedeva nel fatto che particolari variazioni in un gene chiamato Actn3 favorivano la funzionalità delle fibre muscolari «veloci», e che il 98% circa

STEFANO RIZZATO
 ROMA

A tenerli insieme ci sono il passaporto e una bandiera. A dividerli quasi tutto il resto. Lirali e colore degli occhi. E, più in profondità, anche il patrimonio genetico: tutto il codice nascosto tra le eliche del Dna è destinato a passare di padre in figlio. Ecco qua, i 167 diversi tipi di italiani. Diversi tra loro più di quanto lo siano uno spagnolo e un ungherese. Siamo il Paese con la biodiversità umana più estesa d'Europa e a ricordarlo ora c'è uno studio specifico, che ha coinvolto quattro atenei: la Sapienza di Roma insieme alle università di Bologna, Cagliari e Pisa.

Il lavoro è partito nel 2007 e ha unito genetica e antropologia. Da una parte, la raccol-



progetti PRIN 2007-2009

ITALIAN ISOLATES

lingua, geografia e geni

[Home](#)

[Data](#)

[Risultati](#)

[Photogallery](#)

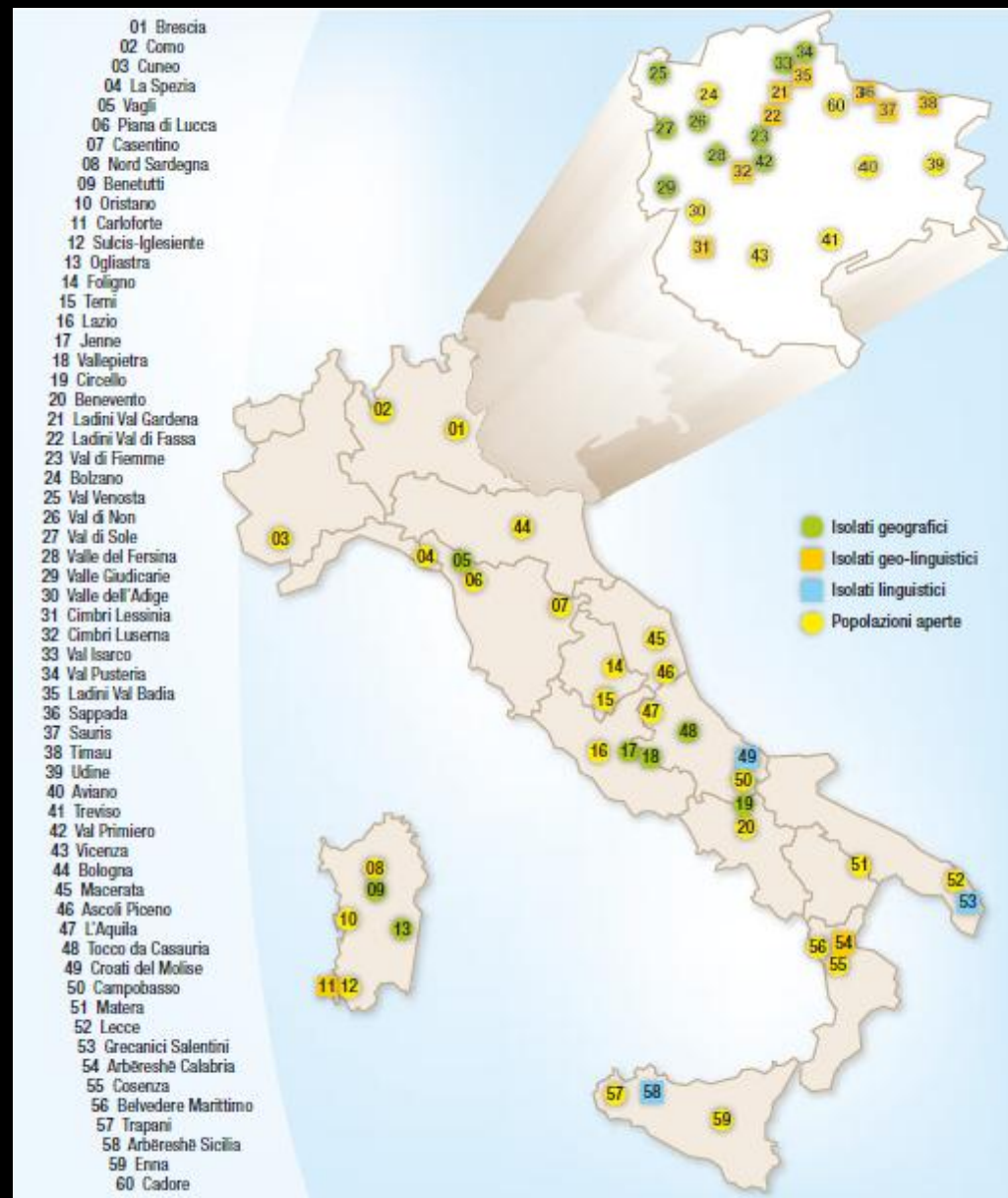
[Notizie](#)

[Persone](#)

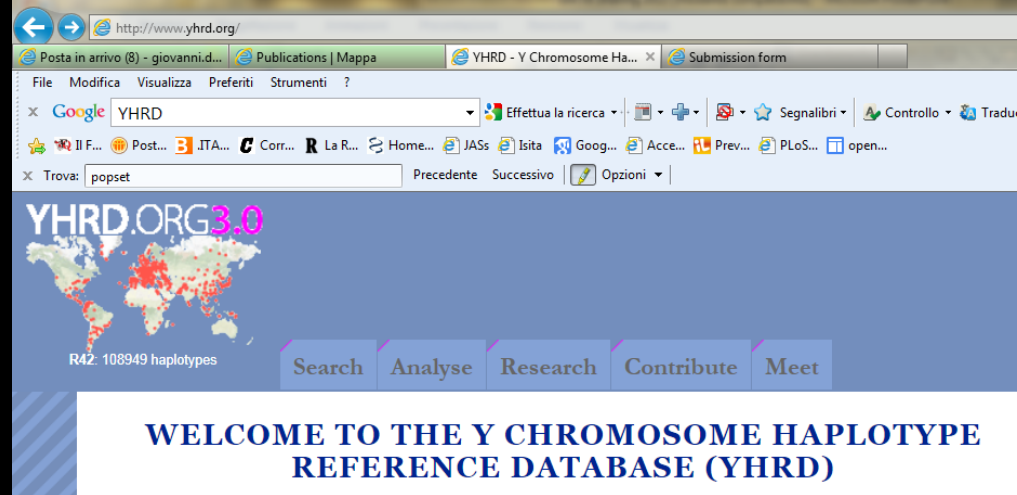
- Project duration: 2007-2014
- 4 research units (Roma, Bologna, Pisa Cagliari)
- 20 papers in peer reviewed journals
- 33 contributions to Congress (10 international)
- average grant by MIUR RU/year: 8,000 euros

... data sharing helped our
project in 3 ways

1. building the dataset



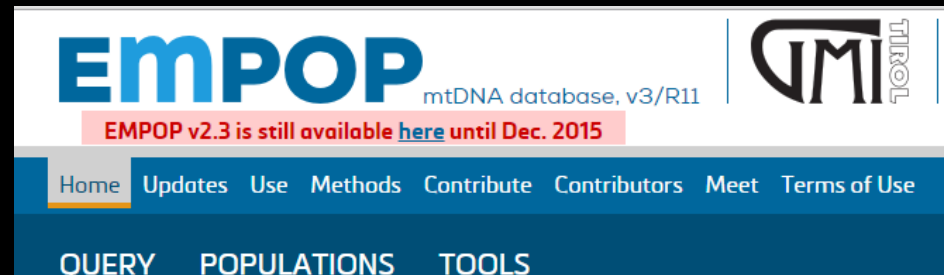
2. Data quality



The screenshot shows the YHRD.ORG 3.0 website. The header includes a navigation bar with links like 'Posta in arrivo', 'Publications', 'Mappa', 'YHRD - Y Chromosome Ha...', and 'Submission form'. Below the navigation bar is a search bar with 'YHRD' entered. The main content area features a world map with red dots representing haplotypes, labeled 'R42: 108949 haplotypes'. To the right of the map are buttons for 'Search', 'Analyse', 'Research', 'Contribute', and 'Meet'. Below the map, a large banner reads 'WELCOME TO THE Y CHROMOSOME HAPLOTYPE REFERENCE DATABASE (YHRD)'.



The screenshot shows the NCBI GenBank Overview page. The header includes the NCBI logo and the text 'GenBank Overview'. Below the header is a navigation bar with links for 'PubMed', 'Entrez', 'BLAST', 'OMIM', and 'Bo'. A search bar is present with 'Entrez' selected and 'eat-4 elegans' entered. On the left side, there is a sidebar with links for 'NCBI', 'SITE MAP', and 'Submit to GenBank'. On the right side, there is a section titled 'What is GenBank?' with a description: 'GenBank® is the NIH genetic sequence collection of all publicly available'.



The screenshot shows the EMPPOP website. The header includes the EMPPOP logo and the text 'mtDNA database, v3/R11'. Below the header is a navigation bar with links for 'Home', 'Updates', 'Use', 'Methods', 'Contribute', 'Contributors', 'Meet', and 'Terms of Use'. A red banner across the middle of the page reads 'EMPOP v2.3 is still available [here](#) until Dec. 2015'. At the bottom, there is a section with links for 'QUERY', 'POPULATIONS', and 'TOOLS'.

3.

Sharing data (and ideas) with study participants



Acknowledgements

Paolo Anagnostou (Sapienza Università di Roma)

Marco Capocasa (Istituto Italiano di Antropologia)

Nicola Milia (Università di Cagliari)

TO CONCLUDE...

“Sharing isn’t easy...and learning to share is a long process.... It can be hard...to understand what’s mine, yours, and ours. Sharing grows little by little, as [we] develop the ability to see things from another person’s point of view and to trust that what they share will be given back.”

Fred Rogers (2004)*

*

