ORIGINE E DISTRIBUZIONE DEI COGNOMI

P. Rossi – Dipartimento di Fisica "E. Fermi" - Ottobre 2008

Introduzione

Lo studio dell'origine e della distribuzione dei cognomi è per sua stessa natura intrinsecamente interdisciplinare. Un'analisi non settoriale della questione coinvolge infatti non soltanto le discipline storiche e demografiche, ma anche la genetica delle popolazioni e, in alcuni degli sviluppi più recenti, certi aspetti della teoria della complessità, studiata dalla fisica statistica.
In questa sede vogliamo ripercorrere con qualche dettaglio le linee di sviluppo e i principali risultati che sono emersi in quest'ambito dagli studi svolti in prevalenza dai genetisti, e negli ultimi anni anche da ricercatori che usano i concetti e applicano i metodi della fisica statistica.

L'avvio dell'uso dei cognomi in biologia umana

Il principale motivo di interesse per lo studio dei cognomi nell'ambito della biologia umana nasce dal fatto che i cognomi, pur essendo di origine culturale e non biologica, si trasmettono di generazione in generazione secondo regole ben precise e legate al comportamento riproduttivo, esattamente come avviene per i geni, e più specificamente per i cosiddetti "geni neutri", che pur facendo parte integrante del corredo genetico non hanno influenza sul fenotipo e pertanto non sono soggetti a nessun tipo di pressione selettiva. La loro distribuzione è quindi governata soltanto dalle leggi della statistica, che devono però tener conto dell'esistenza di fenomeni anche importanti di mutazione e di migrazione (e ciò vale tanto per i cognomi quanto per i geni).
In particolare, nelle culture in cui il cognome dei figli è mutuato da quello paterno, le sue leggi di trasmissione sono esattamente le stesse che governano il cromosoma Y, che essendo posseduto soltanto dal sesso maschile viene ereditato soltanto dal padre e resta invariato nella linea di discendenza maschile diretta con la sola eccezione delle mutazioni (e delle false paternità).
La prima idea di utilizzo dei cognomi nella biologia delle popolazioni risale, certo non incongruamente, a George Darwin[1] (figlio del ben più famoso Charles) che nel 1875 propose di analizzare i casi di matrimoni tra persone portatrici dello stesso cognome per stimare la proporzione dei matrimoni tra cugini primi e valutare gli effetti biologici della consanguineità. Egli suppose che il numero dei matrimoni isonimi che non fossero tra cugini primi dovesse essere proporzionale alla frequenza del cognome nella popolazione, e quindi frequente soltanto per i cognomi comuni. Sulla base della frequenza dei 50 cognomi più comuni in Inghilterra nel 1853 egli stimò che il numero atteso per i matrimoni isonimici tra persone non imparentate fosse prossimo all'uno per mille, e che la deviazione osservata da tale frequenza fosse da attribuirsi ai matrimoni tra cugini primi, che egli stimò al 4,5% per l'aristocrazia, al 3,5% per la borghesia, al 2,5% per la popolazione rurale e al 2% per la popolazione urbana.
L'idea fu abbandonata per circa trent'anni, ripresa da Arner[2] nel 1908, di nuovo abbandonata, e riproposta infine nel 1960 da Shaw[3] che sottolineò l'opportunità offerta dal sistema spagnolo che attribuisce a ciascun individuo due cognomi, corrispondenti nell'ordine al primo cognome del nonno paterno e di quello materno, per cui la frequenza dell'identità tra i due cognomi permetterebbe una più rapida misura (statistica) della consanguineità.

---

[1] G.H. Darwin, *Marriages between first cousins in England and their effects,* Journal of the Statistical Society 38 (1875) 153-184
[2] G.B.L. Arner, *Consanguineous marriages in the American population*, Columbia U. Studies in History, Economics and Public Law (XXXI) 3, New York 1908
[3] R.F. Shaw, *An index of consanguinity based in the use of the surname in Spanish speaking countries,* Journal of Heredity 51 (1960) 221-230

Isonimia e consanguineità

La vera svolta nell'utilizzo dei cognomi nello studio della genetica delle popolazioni umane si ebbe però soltanto nel 1965, quando Crow e Mange[4] proposero un modello matematico per la stima della consanguineità a partire dai cognomi. Essi definirono la consanguineità totale (Ft) di una popolazione e le sue componenti casuale (Fr) e non casuale (Fn), delle quali soltanto la prima può essere spiegata dall'accoppiamento casuale (*random mating)* degli individui presenti in una popolazione. Fecero quindi le ipotesi restrittive (e non realistiche sul piano storico) che tutti i cognomi siano monofiletici (ovvero che condividere il cognome significhi condividere l'antenato da cui esso deriva) e che i due sessi siano ugualmente rappresentati nei migranti, Sotto queste ipotesi essi osservarono che, per un grande numero di tipologie di parentela, l'isonimia nei matrimoni I (*marital isonymy*) è indicativa del grado di consanguineità nella popolazione indipendentemente dal grado di consanguineità nei singoli matrimoni, in quanto la probabilità che due discendenti da un antenato comune abbiano lo stesso cognome varia, nella maggior parte dei casi in misura proporzionale al grado di consanguineità. Per l'esattezza, la probabilità P che due individui abbiano lo stesso cognome è ovviamente legata al grado di parentela (e vale 1 per i fratelli, 1/2 per la parentela zio-nipote, 1/4 per i cugini, etc), ma anche il grado di consanguineità F tra i loro figli dipende dal grado di parentela (e vale 1/4 per i figli di fratelli, 1/16 per i figli di cugini etc), ed è abbastanza facile convincersi che tipicamente vale la relazione $F = P/4$, con rare eccezioni legate a relazioni di parentela complesse e incrociate la cui rilevanza statistica è comunque assai scarsa. Di conseguenza il numero effettivo di matrimoni tra consanguinei è tanto maggiore di quello indicato dall'isonimia quanto più remota è la parentela, e il rapporto numerico cresce in ragione inversa di P, ma in compenso la consanguineità tra i discendenti di matrimoni tra parenti è ridotta in proporzione a F, e i due effetti si compensano in virtù della relazione $F/P = 1/4$. Pertanto la percentuale dei matrimoni isonimi è una misura diretta del grado di consanguineità Ft della popolazione nel suo complesso, pari a 1/4 dell'isonimia matrimoniale: $Ft = 1/4\ I$. Per separare la componente casuale della consanguineità da quella non casuale, Crow e Mange notarono che, dette p(i) e q(i) le frazioni della popolazione maschile e femminile individuate dal cognome i-esimo, la frequenza dei matrimoni casuali tra coppie con lo stesso cognome è p(i) q(i) e il coefficiente di consanguineità casuale Fr vale quindi, per quanto in precedenza dimostrato, $Fr = 1/4 \sum p(i)\, q(i)$, mentre il coefficiente di consanguineità non casuale Fn si ricava facilmente dalla relazione $Ft = Fn + (1 - Fn)\, Fr$, dove a sua volta Ft è misurato dall'isonimia nei matrimoni. A livello teorico Holgate[5] nel 1971 mostrò che il coefficiente di isonimia tende a crescere nel tempo in modelli di popolazione monogama con accoppiamenti casuali. Nel 1977 Lasker[6], osservando che il coefficiente di parentela per una coppia di genitori è esattamente il doppio del coefficiente di consanguineità dei loro figli, propose un'estensione della formula di Crow e Mange che permettesse di usare l'isonimia come misura del grado di parentela tra due popolazioni qualunque. La parentela misurata dall'isonimia vale $RI = 1/2 \sum p1(i)\, p2(i)$, dove p1(i) e p2(i) sono le frequenze con cui il cognome i-esimo compare nelle due popolazioni. I risultati di Crow e Mange e di Lasker sono stati presto applicati a differenti popolazioni, anche con correzioni atte a tener conto degli effetti di polifiletismo e migrazioni[7].

---

[4] J.F. Crow and A.P. Mange, *Measurement of inbreeding from the frequency of marriages between persons of the same surname,* Eugenics Quarterly 12 (1965) 199-203;
J.F. Crow, *The estimation of inbreeding from isonymy,* Human Biology 52 (1980) 1-14

[5] P. Holgate, *Drifts in the Random Component of Isonymy,* Biometrics 27 (1971) 448-451

[6] G.W. Lasker, *A coefficient of relationship by isonymy: A Method for Estimating the genetic Relationship between Populations,* Human Biology 49 (1977) 489-493

[7] W.S. Ellis and W.T. Starmer, *Inbreeding as measured by Isonymy, Pedigrees, and Population Size in Törbel, Switzerland,* Am. J. Hum. Genet. 30 (1978) 366-376

Per una ricognizione dei dati raccolti fino al 1980 e per un inquadramento generale si vedano l'articolo di rivista di Lasker[8] e il volume pubblicato dallo stesso nel 1985.
Tra gli sviluppi successivi merita di segnalare soprattutto lo studio dell'isonimia matrimoniale nel contesto iberico e iberoamericano[9], particolarmente adatto per i motivi citati in precedenza.

La distribuzione in frequenza e l'estinzione dei cognomi

Una linea di ricerca indipendente dalla precedente fu avviata anch'essa nel XIX secolo dallo studio di Galton e Watson[10] che nel 1874, stimolati dall'osservazione che numerose famiglie dell'aristocrazia britannica si erano estinte o rischiavano l'estinzione, si posero il problema di calcolare la probabilità di tale estinzione, giungendo inizialmente all'errata conclusione che la probabilità di estinzione di qualunque cognome dopo un tempo sufficientemente lungo fosse sempre 1. In realtà tale risultato vale soltanto nel caso in cui il numero medio di discendenti maschi sia minore o uguale a uno, mentre in caso contrario la probabilità è certamente maggiore di 0 ma inferiore a 1. Lo studio di Galton e Watson diede comunque il via allo studio matematico dei fenomeni di ramificazione (*branching processes*), che nel contesto dell'analisi evolutiva delle popolazioni fu ripreso da Lotka[11] a partire dal 1931.
Tuttavia soltanto l'apparizione nel 1967 dell'articolo di Karlin e McGregor[12] che formulava la teoria del comportamento delle mutazioni neutre in popolazioni finite di dimensione costante, unita alla constatazione che i cognomi possono essere considerati come alleli trasmessi lungo la linea maschile, attrasse l'attenzione degli studiosi di genetica verso lo studio della distribuzione e dell'estinzione dei cognomi, a partire dal lavoro di Yasuda *et al.*[13] pubblicato nel 1974, nel quale la teoria di Karlin e McGregor fu applicata ai dati sui cognomi raccolti tra la popolazione della valle del Parma da Cavalli-Sforza e collaboratori fin dal 1954. La teoria permette di predire il valore atteso della distribuzione di $N^*(k)$, il numero dei cognomi che sono rappresentati da k individui (maschi) in una popolazione di dimensione N, assumendo che in ogni transizione (passaggio da una generazione alla successiva) sia assegnata (e uguale per tutti gli individui) la probabilità ν di un cambio di cognome. La somma dei valori $N^*(k)$ è il numero atteso S di cognomi nella popolazione, e si può calcolare il valore atteso dell'isonimia casuale, ossia la probabilità che due individui presi a caso abbiano lo stesso cognome, ottenendo $I = 1/(N\nu + 1 - \nu)$. Si può dimostrare che in questo modello $S > 1/I$. Il confronto tra la distribuzione teorica e quella empirica permette di verificare la bontà delle assunzioni e di fissare i valori dei parametri N e ν. Poiché le vere mutazioni di cognome sono relativamente rare, un valore elevato di ν (come quello trovato nel caso della valle del Parma, superiore a 0,2) sembra indicare l'importanza dei fenomeni di immigrazione nella distribuzione dei cognomi. Yasuda *et al.*, nello stesso articolo, affrontarono poi il tema della distribuzione statistica della discendenza, in

---

[8] G.W. Lasker, *Surnames in the Study of Human Biology,* American Anthropologist 82 (1980) 525-538; G.W. Lasker, *Surnames and Genetic Structure,* Cambridge University Press 1985
[9] J. Pinto-Cisternas, L. Pineda and I. Barrai, *Estimation of Inbreeding by Isonymy in Iberoamerican Populations,* Am. J. Hum. Genet. 37 (1985) 373-385;
M. Esparza, C. Garcia-Moro and M. Hernandez, *Inbreeding From Isonymy and repeated Pairs of Surnames in the Ebro Delta Region,* Am. J. Hum. Biology 18 (2006) 849-852
[10] F. Galton and H.W. Watson, *On the Probability of the Extinction of Families,* Journal of the Anthropological Institute of Great Britain and Ireland 4 (1874) 138-144
[11] A.J. Lotka, *The extinction of families I-II*, J. Wash. Acad. Sci. 21 (1931) 377-380, 453-459
[12] S. Karlin and J. McGregor, *The number of mutant forms maintained in a population*, Proc. 5th Berkeley Symp. Math. Stat. Prob. 4 (1967) 415-438
[13] N. Yasuda, L.L. Cavalli-Sforza, M. Skolnick and A. Moroni, *The Evolution of Surnames: An Analysis of Their Distribution and Extinction*, Theor. Pop. Biol. 5 (1974) 123-142

particolare per l'aspetto relativo alla probabilità di estinzione dei cognomi dopo un numero sufficientemente alto di generazioni. Le probabilità di estinzione da essi calcolate (sulla base di un modello di decrescita geometrica per le probabilità di nascita di n figli maschi) risultarono in ragionevole accordo con i dati empirici raccolti.
Un risultato fondamentale nello studio della distribuzione delle frequenze dei cognomi fu pubblicato da Fox e Lasker[14] nel 1983 (ma già annunciato nell'articolo di Lasker del 1980). Fox e Lasker mostrarono che i dati empirici relativi alla distribuzione in frequenza dei cognomi di 4794 individui viventi nell'area di Reading (Inghilterra) potevano essere descritti con buona precisione utilizzando una distribuzione di Pareto discreta (ossia una legge di potenza).
L'ubiquità delle leggi di potenza nella descrizione di fenomeni, sia naturali che sociali e culturali, ogni qual volta il fenomeno non sia caratterizzato da una scala intrinseca, è spiegata nel contesto della moderna teoria della complessità sulla base della nozione di "criticità auto-organizzata" ed è il motivo che ha indotto in tempi recenti gli studiosi di fisica statistica a investigare numerosi soggetti (inclusa appunto la distribuzione dei cognomi) al di fuori dell'ambito tradizionale di indagine della loro disciplina.[15]
Un altro importante contributo all'indagine sulla distribuzione in frequenza dei cognomi fu portato da una serie di lavori pubblicato a partire dal 1983, Zei *et al.*[16] e specificamente dedicati allo studio della distribuzione dei cognomi in Sardegna. Fu mostrato che la distribuzione empirica mostrava un buon accordo con la distribuzione degli alleli neutri di Karlin-McGregor, e anche con la distribuzione logaritmica di Fisher[17], peraltro derivabile dalla precedente nel limite in cui N diventa infinito. Vale in questo caso $N^*(k) = \alpha/k \; x^k$, dove si sono introdotte le nuove variabili $x = 1 - \nu$ e $\alpha = N \nu/(1 - \nu)$, ed è possibile valutare il numero totale S dei cognomi mediante la semplice relazione $S = - \alpha \ln \nu$. Infine fu considerata la dipendenza spaziale della distribuzione dei cognomi, notando come la variazione spaziale fosse molto più marcata della variazione temporale, e largamente indipendente dalla frequenza dei cognomi (salvo il caso di bassissima frequenza). Questi risultati sono tutti consistenti con l'ipotesi che i cognomi si comportino come alleli neutri.
Ulteriori proposte di parametrizzazione della distribuzione in frequenza dei cognomi, basate su modelli statistici dei comportamenti riproduttivi e della conseguente evoluzione dei cognomi, furono avanzati da Panaretos[18] che nel 1989 propose l'impiego della distribuzione di Yule, e da Consul[19], che nel 1991 ricavò, sia da un modello di nascita e morte che da un modello di ramificazione, una descrizione basata sulla distribuzione di Geeta, in buon accordo con i dati.

---

[14] W.R. Fox and G.W. Lasker, *The Distribution of Surname Frequencies,* Int. Stat. Review 51 (1983) 81-87
[15] M.E.J. Newman, *Power laws, Pareto distributions and Zipf's law*, Cont. Phys. 46 (2005) 323
[16] G. Zei, C.R. Guglielmino, E. Siri, A. Moroni and L.L. Cavalli-Sforza, *Surnames as Neutral Alleles: Observations in Sardinia,* Human Biology 55 (1983) 357-365; G. Zei, R. Guglielmino Matessi, E. Siri, A. Moroni and L. Cavalli-Sforza, *Surnames in Sardinia I. Fit of frequency distributions for neutral alleles and genetic population structure,* Ann. Hum. Genet. 47 (1983) 329-352; G. Zei, A. Piazza, A. Moroni and L.L. Cavalli-Sforza, *Surnames in Sardinia III. The spatial distribution of surnames for testing neutrality of genes,* Ann. Hum. Genet. 50 (1986) 169
[17] R.A. Fisher, A.S. Corbet and C.B. Williams, *The Relation Between the Number of Species and the Number of Individuals in a random sample of an Animal Population,* J. Animal Ecology 12 (1943) 42-58
[18] J. Panaretos, *On the Evolution of Surnames*, Int. Stat. Review 57 (1989) 161-167; J. Panaretos, *A Probability Model Involving the Use ofthe Zero-Truncated Yule Distribution for Analysing Surname Data,* IMA Journ. Math. Appl, Med. Biol. 6 (1989) 133-136
[19] P.C. Consul, *Evolution of Surnames*, Int. Stat. Review 59 (1991) 271-278; M.N. Islam, *A Stochastic Model for Surname Evolution,* Biom. J. 37 (1995) L119-126

Distribuzione dei cognomi e dinamiche migratorie

Nel lavoro di Wijsman *et al.*[20] del 1984 furono calcolate le matrici di migrazione tra nove differenti aree della Sardegna, sulla base di una matrice di parentela definita in modo analogo a quello introdotto da Lasker e valutata tra popolazioni considerate a tempi diversi (1850-1970).
Fu così evidenziata una dipendenza forte e non lineare dei coefficienti di immigrazione dalla distanza tra i territori esaminati.
Il tema dell'utilizzo dei cognomi per lo studio delle dinamiche migratorie fu ulteriormente sviluppato in alcuni lavori successivi. In particolare Piazza *et al.*[21] nel 1987 confrontarono le stime dei coefficienti di migrazione in Italia da loro ottenute analizzando le distribuzioni di cognomi estratte dagli elenchi telefonici con le corrispondenti stime ricavate dalle fonti demografiche ufficiali, mostrando che le stime basate sulle distribuzioni dei cognomi mediante il calcolo del coefficiente ν potevano fornire una stima attendibile dei coefficienti di migrazione.
Lo studio dei fenomeni migratori mediante l'analisi della distribuzione dei cognomi fu condotto in seguito anche da Darlu e Ruffié[22] per le migrazioni interne francesi, e da Degioanni *et al.*[23] per le migrazioni dall'Italia verso la Francia.
A loro volta Zei *et al.*[24] ricostruirono le dinamiche migratorie italiane a partire dall'analisi della distribuzione geografica dei cognomi (e in particolare dallo studio dei confini tra regioni in cui gli stessi cognomi sono presenti con frequenze molto differenti), confrontando i risultati con le barriere geografiche e linguistiche e con gli esiti delle analisi genetiche delle popolazioni. I confini genetici e quelli tra i cognomi mostrarono locazioni molto simili, e correlate soprattutto alle barriere fisiche, ma anche a quelle linguistiche.
La relazione tra la trasmissione genetica (puramente verticale) e quella culturale (che è spesso anche orizzontale) è stata investigata confrontando la distribuzione dei cognomi con quella di altri marcatori culturali in una serie di lavori relativi alla Sicilia, a partire da Guglielmino *et al.*[25]

---

[20] E. Wijsman, G. Zei, A. Moroni and L.L. Cavalli-Sforza, *Surnames in Sardinia II. Computation of migration matrices from surname distributions in different periods,* Ann. Hum. Genet. 48 (1984) 65-78;
[21] A. Piazza, S. Rendine, G. Zei, A. Moroni and L.L. Cavalli-Sforza, *Migration rates of human populations from surname distributions,* Nature 329 (1987) 714; A. Piazza, N. Cappiello, E. Olivetti and S. Rendine, *A genetic history of Italy*, Ann. Hum. Genet. 52 (1988) 203-213
[22] P. Darlu and J. Ruffié, *L'immigration dans les départements français étudiée par la méthode des patronymes,* Population 47 n.3 (1992) 719; P. Darlu and J. Ruffié, *Relationships between consanguinity and migration rate from surname distributions and isonymy in France*, Ann. Hum. Biol. 19 (1992) 133; P. Darlu, A. Degioanni, J. Ruffié, *Quelques statistiques sur la distribution des patronymes en France*, Population 52 n.3 (1997) 607
[23] A. Degioanni, A. Lisa, G. Zei, P. Darlu, *Patronymes italiens et migration italienne en France*, Population 51 n.6 (1996) 1153
[24] G. Zei, G. Barbujani, A. Lisa, O. Fiorani, P. Menozzi, e. Siri and L.L. Cavalli-Sforza, *Barriers to gene flow estimated by surname distribution in Italy,* Ann. Hum. Genet. 57 (1993) 123
[25] C.R. Guglielmino, G. Zei and L.L. Cavalli-Sforza, *Genetic and Cultural transmission in Sicily as Revealed by Names and Surnames,* Human Biology 63 (1991) 607-627; A. Rodriguez-Larralde, A. Pavesi, C. Scapoli, F. Conterio, G. Siri and I. Barrai, *Isonymy and the genetic structure of Sicily,* J. Biosoc. Sci. 26 (1994) 9-24; A. Piazza *et al.*, *Towards a genetic history of Sicily,* J. Cult. Heritage 1 (2000) Sup.39-42; A. Pavesi, P. Pizzetti, E. Siri, E. Lucchetti and F. Conterio, *Coexistence of Two Distinct Patterns in the Surname Structure of Sicily,* Am. J. Phys. Anthrop. 120 (2003) 195-199; A. De Silvestri and C.R. Guglielmino, *Sicilian Provinces: Population Subdivisions Revealed by Surname Frequencies,* Human Biology 76 (2004) 901-920

Cognomi e cromosoma Y

Tra i recenti sviluppi dell'indagine sull'origine e l'evoluzione biologica dei cognomi uno dei più interessanti è quello avviato dall'articolo di Sykes e Irven[26] del 2000, basato sull'osservazione che nella maggior parte delle culture, e in particolare in quelle europee, la trasmissione ereditaria del cognome avviene in forma patrilineare, in perfetta analogia con quanto avviene per il cromosoma Y, che essendo posseduto soltanto dai maschi viene ereditato esclusivamente dal padre. Fu quindi analizzato il patrimonio genetico di un significativo campione casuale di soggetti maschi portatori del cognome Sykes. Dal fatto che quasi il 50% del campione condivideva un particolare aplotipo non osservato nei gruppi di controllo, Sykes e Irven inferirono che, ammettendo un numero molto limitato di casi di nonpaternità nell'arco dei 700 anni in cui il cognome è documentato (in media 1,3% per generazione), il cognome Sykes dovesse ritenersi monofiletico (mentre fonti scritte parevano indicare un'origine multipla).
Una prima descrizione dello stato delle ricerche sul cromosoma Y e sul loro intreccio con lo studio dell'origine e della distribuzione dei cognomi si trova nell'articolo di Jobling[27] del 2001 e nella relativa bibliografia. Tra i risultati più importanti vanno ricordati quello sulla relazione tra gli aplotipi di un gruppo di 221 maschi irlandesi e l'origine (gaelica o anglosassone) dei loro cognomi, e quello relativo al patrimonio genetico dei religiosi ebraici il cui cognome (Cohen e sue varianti) indica, per la tradizione biblica, una discendenza patrilineare diretta da Aronne, fratello di Mosé (XIII sec. a.C.). In entrambi i casi la forte correlazione sembra indicare uno stretto legame tra il dato genetico e quello culturale.
Il metodo di analisi della distribuzione spaziale dei cognomi è stato raffinato nell'articolo di Manni *et al.*[28], con l'uso del metodo di raggruppamento SOM (*self-organizing maps*), che gli autori hanno applicato al caso dei cognomi dei Paesi Bassi, suggerendo che il campionamento genetico di un'area soggetta a fenomeni di immigrazione debba essere effettuato con selezione preliminare basata sul cognome, se si vogliono mettere in evidenza i tratti genetici caratteristici della popolazione da maggior tempo residente nell'area in esame.

Lo stato dell'arte

Le più recenti messe a punto dello stato delle ricerche in forma di articolo di rivista sono quella di Colantonio *et al.*[29] del 2003, con un'amplissima bibliografia ragionata, ripartita per aree geografiche, e quella di P. Darlu[30], che aveva l'obiettivo esplicito di trasmettere le acquisizioni della ricerca effettuata dai genetisti alla comunità dei demografi storici, spesso interessati a problemi molto simili (dinamica delle popolazioni, fenomeni migratori) ma finora non particolarmente interessati allo studio della distribuzione dei cognomi.

---

[26] B. Sykes and C. Irven, *Surnames and the Y Chromosome,* Am. J. Hum. Genet. 66 (2000) 1417-1419
[27] M. A. Jobling, *In the name of the father: surnames and genetics,* Trends in Genetics 17 (2001) 353-357
[28] F. Manni, B. Toupance, A. Sabbagh and E. Heyer, *New Method fir Surname Studies of Ancient Patrilineal Population Structures, and possible Application to Improvement of Y-Chromosome Sampling,* Am. J. Phys. Anthrop. 126 (2005) 214-228
[29] S.E. Colantonio, G.W. Lasker, B.A. Kaplan and V. Fuster, *Use of Surname Models in Human Population Biology: A Review of Recent Developments,* Human Biology 75 (2003) 785-807
[30] P. Darlu, *Patronymes et démographie historique,* Ann. Dem. Hist. 2 (2004) 53-65

I cognomi in Europa Occidentale

Un'altra importante messa a punto è quella contenuta nell'articolo del 2007 di Scapoli *et al.*[31], in cui vengono riassunti e nuovamente analizzati soprattutto i dati raccolti in una ventina d'anni di attività da un ampio gruppo di ricercatori il cui principale esponente è I. Barrai.
I primi lavori del gruppo (fino al 1994) si concentrarono su alcune realtà locali italiane (Ferrara, Perugia, la già citata Sicilia)[32], indagando sia i fenomeni di consanguineità che quelli di migrazione e di isolamento dovuto alla distanza.
In seguito, quando per molti Paesi divennero largamente disponibili le versioni digitalizzate degli elenchi telefonici, il gruppo ampliò la propria indagine a tali Paesi, e a un numero elevatissimo di soggetti, fino a coprire mediamente il 9% della popolazione dei principali Stati dell'Europa occidentale (Svizzera, Germania. Italia, Austria, Paesi Bassi, Belgio Spagna, Francia)[33]

---

[31] C. Scapoli, E. Mamolini, A. Carrieri, A. Rodriguez-Larralde, I. Barrai, *Surnames in Western Europe: A comparison of the subcontinental populations through isonymy,* Theor. Pop. Biol. 71 (2007) 37-48

[32] I. Barrai, G. Barbujani, M. Beretta, I. Maestri and A. Russo, *Surnames in Ferrara: distribution, isonymy and levels of inbreeding,* Ann. Hum. Biol. 14 (1987) 415; I. Barrai, G. Formica, R. Barale, C. Scapoli, R. Camella and M. Beretta, *Isonymy in immigrants from Ferrara in 1981-1988,* Ann. Hum. Biol. 17 (1990) 7; I. Barrai, G. Formica, R. Barale and M. Beretta, *Isonymy and migration distance,* Ann. Hum. Genet. 53 (1989) 249; I. Barrai, C. Scapoli, R. Carella, G. Formica, R. Barale and M. Beretta, *Isonymy in records of births and deaths in Ferrara,* Ann. Hum. Biol. 18 (1991) 395; I. Barrai, G. Formica, C. Scapoli, M. Beretta. E. Mamolini, S. Volinia, R. Barale, P. Ambrosino and F. Fontana, *Microevolution in Ferrara: Isonymy 1890-1990,* Ann. Hum. Biol. 19 (1992) 371; A. Rodriguez-Larralde, G. Formica, C. Scapoli, M. Beretta. E. Mamolini and I. Barrai, *Microevolution in Perugia: isonymy 1890-1990,* Ann. Hum. Biol. 20 (1993) 261; M. Beretta, E. Mamolini, A. Ravani, C. Vullo, C. Scapoli, R. Barale, A. Rodriguez-Larralde and I. Barrai, *Comparison of Structures from Frequencies of Genes and Surnames in the Population of Ferrara,* Human Biology 65 (1993) 225

[33] I. Barrai, C. Scapoli, M. Beretta, C. Nesti, E. Mamolini and A. Rodriguez-Larralde, *Isonimy and the genetic structure of Switzerland I. The distribution of surnames,* Ann. Hum. Biol. 23 (1996) 431; C. Scapoli, A. Rodriguez-Larralde, M. Beretta, C. nesti, A. Lucchetti, I. Barrai, *Correlations between Isonimy Parameters,* Int. J. of Anthropology 12 (1997) 17; A. Rodriguez-Larralde, C. Scapoli, M. Beretta, C. Nesti, E. Mamolini and I. Barrai, *Isonimy and the genetic structure of Switzerland II. Isolation by distance,* Ann. Hum. Biol. 25 (1998) 533; A. Rodriguez-Larralde, I. Barrai, C. Nesti, E. Mamolini and C. Scapoli, *Isonimy and Isolation by Distance in Germany,* Human Biology 70 (1998) 1041; I. Barrai, A. Rodriguez-Larralde, E. Mamolini and C. Scapoli, *Isonimy and isolation by distance in Italy,* Human Biology 71 (1999) 947; I. Barrai, A. Rodriguez-Larralde, E. Mamolini, F. Manni and C. Scapoli, *Elements of the surname structure of Austria,* Ann. Hum. Biol. 27 (2000) 607; I. Barrai, A. Rodriguez-Larralde, F. Manni and C. Scapoli, *Isonimy and Isolation by Distance in the Netherlands,* Human Biology 74 (2002) 263; I. Barrai, A. Rodriguez-Larralde, F. Manni, V. Ruggiero, D. Tartari and C. Scapoli, *Isolation by Language and Distance in Belgium,* Ann. Hum. Genetics 68 (2003) 1; A. Rodriguez-Larralde, A. Gonzales-Martin, C. Scapoli, I. Barrai, *The names of Spain: A study of the isonymy structure of Spain,* Am. J. Phys. Anthropol. 121 (2003) 280; C. Scapoli, H. Goebl, S. Sobota, E. Mamolini, A. Rodriguez-Larralde, I. Barrai, *Surnames and dialects in France: Population structure and cultural evolution,* J. Theor. Biol. 237 (2005) 75

Alcune indagini del gruppo hanno riguardato anche Venezuela, Argentina, Stati Uniti.[34]
Nel principale articolo metodologico, pubblicato da Scapoli *et al.*[35] nel 1997, sono indicati i principali parametri che costituiscono l'oggetto di studio in tutti i lavori del gruppo: l'isonimia I intesa come misura della componente casuale della consanguineità, il coefficiente α di Fisher che misura l'abbondanza dei cognomi in un gruppo, l'indice ν di Karlin e McGregor, che risulta proporzionale al coefficiente di immigrazione, e anche l'entropia della distribuzione dei cognomi espressa dalla relazione $H = \Sigma\, p(i) \ln p(i)$. Si noti che nel limite di N grande, in cui vale la teoria di Fisher, vale anche la relazione $I = 1/\alpha + 1/N$. Viene calcolato anche l'isolamento dovuto alla distanza, misurando la correlazione tra il grado di isonimia e la distanza geografica di coppie di popolazioni.
Un elemento comune a tutti i lavori, ispirato al risultato di Fox e Lasker, è la rappresentazione della distribuzione dei cognomi mediante grafici log-log, che trasformano le leggi di potenza in relazioni lineari, e la stima del corrispondente coefficiente (l'esponente nella legge di potenza) per i diversi Paesi. Nel lavoro riassuntivo sull'Europa occidentale i dati relativi agli otto Paesi esaminati sono stati aggregati, mostrando che anche nel caso generale la rappresentazione della distribuzione come legge di potenza fornisce ancora una descrizione abbastanza soddisfacente dell'andamento generale della distribuzione stessa.
In relazione a questi risultati un'analisi dei meccanismi che possono portare all'instaurarsi di una legge di potenza è stata effettuata anche da Bartley e collaboratori, che hanno preso in esame un modello includente processi di nascita, morte e creazione di nuovi tipi.[36]

---

[34] A. Rodriguez-Larralde, J. Morales and I. Barrai, *Surname Frequency and the Isonimy Structure of Venezuela,* Am. J. Human Biol. 12 (2000) 352; I. Barrai, A. Rodriguez-Larralde, E. Mamolini, F. Manni and C. Scapoli, *Isonymy structure of USA population,* Am. J. Phys. Anthropol. 114 (2001) 109; J.E. Dipierri, E.L. Alfaro, C. Scapoli, E. Mamolini, A. Rodriguez-Larralde and I. Barrai, *Surnames in Argentina: A population study through isonimy* Am. J. Phys. Anthropol. 128 (2005) 199; A. Rodriguez-Larralde, C. Scapoli, E. Mamolini, I. Barrai, *Surnames in Texas: a population study through isonymy,* Human Biology 79 (2007) 215
[35] C. Scapoli, A Rodriguez-Larralde, M. Beretta, C. Nesti, A. Lucchetti, I. Barrai, *Correlations between Isonymy Parameters*, Int. J. of Anthropology 12 (1997) 17
[36] D.L. Bartley, T. Ogden, R. Song, *Frequency distributions from birth, death and creation processes,* BioSystems 66 (2002) 179-191

Lo studio dei cognomi in meccanica statistica

L'interesse dei fisici teorici per la creazione e lo studio di modelli stocastici capaci di cogliere alcuni aspetti della dinamica dell'evoluzione biologica risale alla fine degli anni Ottanta del XX secolo, in corrispondenza dell'affermarsi della teoria neutralista dell'evoluzione (per cui la gran parte della variabilità individuale non ha effetti importanti sulla *fitness*), che si presta molto bene a rappresentazioni che sono tipiche dei sistemi studiati in meccanica statistica, e grazie ai notevoli sviluppi della teoria dei sistemi disordinati, nel cui ambito entrano abbastanza naturalmente i modelli di evoluzione neutra. Ricordiamo qui, tra i lavori che in qualche modo anticipano le problematiche di interesse in questo contesto, l'articolo di Derrida e Peliti[37] del 1991 sull'evoluzione di un modello per popolazioni di individui che si riproducono in modo asessuato, per il quale risulta possibile calcolare non solo la statistica delle genealogie e la variabilità genetica (che sono i corrispettivi del grado di consanguineità e della distribuzione in frequenza) ma anche le fluttuazioni di queste quantità, tipiche dei sistemi disordinati. Nell'articolo di Serva e Peliti[38] dello stesso anno viene invece esaminato un modello di riproduzione sessuata, trovando che in questo caso le fluttuazioni nella distanza genetica tra gli individui tendono a svanire nel limite di popolazione infinita.
Le proprietà statistiche degli alberi genealogici furono studiate per la prima volta nel 1999 in modo quantitativo, mediante simulazioni numeriche, nel lavoro di Derrida, Manrubia e Zanette[39], in cui fu mostrato che un semplice modello neutro di riproduzione sessuata era in grado di riprodurre i valori empirici della ripetizione di antenati nell'albero genealogico di Edoardo III (1312-1377) re d'Inghilterra. La distribuzione delle ripetizioni di antenati, dopo un numero sufficiente di generazioni, acquista una forma universale, sulla base della quale è anche possibile calcolare la frazione di individui della più antica generazione considerata che risulta mediamente assente in un dato albero genealogico (pari al 20,3% della popolazione), e l'esponente della legge di potenza che governa l'andamento della distribuzione per bassi valori del numero delle ripetizioni. Dopo la pubblicazione dei risultati di Miyazima *et al.*[40] a proposito dei cognomi nelle città giapponesi, Zanette e Manrubia[41] si indirizzarono al problema della distribuzione in frequenza dei cognomi, dimostrando che in un modello stocastico (proposto da Simon[42] nel 1955) con popolazione (esponenzialmente) crescente e con probabilità non nulla α per la comparsa di nuovi cognomi a ogni generazione, la distribuzione tende asintoticamente nel tempo a una legge di potenza, con esponente dipendente da α, e tendente a 2 quando α tende a 0.

---

[37] B. Derrida and L. Peliti, *Evolution in a flat fitness landscape,* Bull. Math. Biol. 53 (1991) 355-382

[38] M. Serva and L. Peliti, *A statistical model of an evolving population with sexual reproduction,* J. Phys. A: Math. Gen. 24 (1991) L705-L709

[39] B. Derrida, S.C. Manrubia and D.H. Zanette, *Statistical Properties of Genealogical Trees*, Phys. Rev. Lett. 82 (1999) 1987-1990; B. Derrida, S.C. Manrubia and D.H. Zanette, *Distribution of repetitions of ancestors in genealogical trees,* Physica A 281 (2000) 1-16; B. Derrida, S.C. Manrubia and D.H. Zanette, *On the genealogy of a population of biparental individuals,* J. Theor. Biol. 203 (2000) 303

[40] S. Miyazima, Y. Lee, T. Nagamine and H. Miyajima, *Power Law Distribution of Family Names in Japanese Societies,* Physica A 278 (2000) 282-288

[41] D. H. Zanette and S. C. Manrubia, *Vertical transmission of culture and distribution of family names*, Physica A 295 (2001) 1-8; S. C. Manrubia and D. H. Zanette, *At the Boundary between Biological and Cultural Evolution: the Origin of Surname Distributions*, J. Theor. Biol. 216 (2002) 461-477; S.C. Manrubia, B. Derrida and D.H. Zanette, *Genealogy in the Era of Genomics,* Am. Scient. 91 (2003) 158-165

[42] H.A. Simon, *On a Class of Skew Distribution Functions,* Biometrika 42 (1955) 425-440

Un modello alternativo di spiegazione fu proposto nel 2002 da Reed e Hughes[43], che mostrarono che, se un processo stocastico caratterizzato da una crescita esponenziale è interrotto (oppure osservato) in modo casuale, la distribuzione dello stato osservato segue (almeno asintoticamente) una legge di potenza. Considerando quindi l'evoluzione della distribuzione dei cognomi come un processo di diramazione di tipo Galton-Watson con l'aggiunta di una probabilità finita per la comparsa di nuovi cognomi (per mutazione o per immigrazione), si dimostra che la distribuzione dei cognomi segue una legge di potenza, con un esponente dipendente dalla probabilità di comparsa di nuovi cognomi e dal coefficiente di crescita della popolazione.
Più di recente alcuni studi originati dall'analisi sia diacronica che sincronica dei cognomi coreani[44] hanno condotto alla formulazione di un modello di dinamica delle popolazioni formulato in termini di una *master equation*, nella quale appaiono come variabili le probabilità P(j,k;s,t) che una famiglia abbia k membri al tempo t se aveva j membri al tempo s, mentre compaiono come parametri le probabilità di nascita, di morte e di mutazione di cognome di un membro della famiglia al tempo t (che sono tutte proporzionali a k e hanno, per ipotesi, la stessa dipendenza dal tempo). Quest'equazione può essere formalmente risolta per valori dati di j e s, e in particolare nel caso j=1 (ossia assumendo la comparsa della famiglia al tempo s), e pertanto la distribuzione complessiva dei cognomi nella popolazione si può ricavare dalla conoscenza del numero di cognomi introdotti a ciascun tempo s. Il modello così formulato è molto generale, e comprende per esempio il modello di Zanette e Manrubia come caso particolare. È possibile rappresentare il caso di una popolazione in cui siano presenti sia mutazione che immigrazione assumendo che il numero di cognomi introdotti al tempo s abbia due componenti, di cui una costante che rende conto dell'immigrazione, e l'altra proporzionale alla popolazione totale N, per tener conto delle mutazioni. In una popolazione che cresce esponenzialmente il numero di cognomi cresce nel tempo come il logaritmo di N, mentre la probabilità che una famiglia abbia k membri è (a grandi tempi) proporzionale a 1/k : questi risultati sembrano descrivere bene il caso della Corea (e quello della Cina), coerentemente con i dati storici noti. Se invece si ammette la possibilità di mutazioni il numero dei cognomi cresce proporzionalmente a N, mentre l'esponente della distribuzione è prossimo a -2, ma comunque dipende dal coefficiente di crescita della popolazione e dal coefficiente di mutazione.
Questi risultati sono stati di recente riprodotti da De Luca e Rossi mediante l'applicazione di tecniche di gruppo di rinormalizzazione mutuate dalla teoria dei campi statistica.[45]

---

[43] W.J. Reed and B.D. Hughes, *From gene families to incomes and internet file sizes: Why power laws are so common in nature*, Phys. Rev. E 66 (2002) 067103; W.J. Reed and B.D. Hughes, *On the distribution of family names*, Physica A 319 (2003) 579-590

[44] B.J. Kim and S. M. Park, *Distribution of Korean Family Names,* Physica A 347 (2005) 683-694; H.A.T. Kiet, S.K. Baek, B.J. Kim, H. Jeung, *Korean Family Name Distribution in the Past,* J. Korean Phys. Soc. 51 (2007) 1812-1816; S.K. Baek, H.A.T. Kiet and B.J. Kim, *Family name distributions: Master equation approach,* Phys. Rev. E 76 (2007) 046113

[45] A. De Luca and P. Rossi, *Renormalization group evaluation of exponents in family name distributions* (2008, to appear)