

La genealogia come disciplina ausiliaria della genetica

P. Rossi e R. Bizzocchi – Università di Pisa

Per una storia delle interazioni tra genetica e genealogia

La genealogia è normalmente classificata tra le discipline ausiliarie della storia, alla stregua della sfragistica o della numismatica, ma talvolta con minore dignità scientifica, per motivi diversi che concorrono a ridurre la credibilità di molta parte del materiale reperibile nella letteratura specifica. Da un lato è certamente presente, più che in altre discipline, un'ampia fascia di "dilettanti" i cui contributi spesso mancano del necessario rigore filologico, nell'ansia di stabilire collegamenti e di completare ricostruzioni, anche in mancanza di un'adeguata documentazione. D'altra parte è da sempre presente anche una tendenza che possiamo definire "nobiliare", anche se va molto al di là della ristretta cerchia della nobiltà di sangue, e il cui obiettivo è quello di attribuire illustri natali anche ben oltre il confine della verità e della credibilità storica.¹ Questa tendenza si prolunga, e anche si rinnova e modifica, nel pieno della contemporaneità anche per motivi di natura economico-sociale, quale l'esigenza dei (supposti) discendenti di emigrati di vedersi riconoscere diritti ereditari o meglio ancora un'ascendenza tale da consentire per *ius sanguinis* l'attribuzione di una cittadinanza altrimenti difficilmente conseguibile.

È spiacevole constatare che proprio in Italia, per motivi che certamente affondano le proprie radici nella storia passata e recente della nazione, le tendenze sopra indicate sono particolarmente presenti, ed è quindi altrettanto particolarmente presente una forte diffidenza verso questo genere di studi. Certamente diverso è lo *status* delle ricerche genealogiche in Paesi in cui la disciplina ha ben più solide tradizioni e radici, a partire dall'Inghilterra, ma anche in gran parte dell'Europa continentale.

Anche in virtù di questa maggiore credibilità, non v'è dubbio che, fuori dall'Italia, il valore conoscitivo della genealogia ha trovato un diverso riconoscimento, oltre che nelle tradizionali ricerche storiche (come qui riferisce Marie Lezowski), anche in un contesto, largamente interdisciplinare, che fa riferimento alla genetica (e più in generale alla dinamica) delle popolazioni. Ci riferiamo agli studi volti a identificare le relazioni, anche remote, di consanguineità che intercorrono tra individui e tra popoli: ricerche che dovrebbero condurre, nelle ambizioni degli studiosi della disciplina, alla ricostruzione di una tassonomia genetica dell'umanità, una sorta di "albero genealogico virtuale" (ovviamente non direttamente associabile a singoli individui storicamente identificati) dell'intera specie *Homo sapiens sapiens*.

Per comprendere il ruolo che la genealogia storica può, e anzi dovrebbe, giocare in questo genere di studi occorre tuttavia ripercorrere brevemente un itinerario di ricerca che, nell'ambito della genetica, è iniziato nella seconda metà del XIX secolo.²

La prima idea di utilizzo di dati genealogici nella biologia delle popolazioni risale, certo non incongruamente, a George Darwin (figlio del ben più famoso Charles e di sua cugina Emma Wedgwood) 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.³ 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

¹ R. Bizzocchi, *Genealogie incredibili. Scritti di storia nell'Europa moderna*, Il Mulino, Bologna 2009

² Per una rassegna cfr P. Rossi, *Surname distribution in population genetics and in statistical physics*, *Physics of Life Reviews* 10 (2013) pp. 395-415

³ G.H. Darwin, *Marriages between first cousins in England and their effects*, *Journal of the Statistical Society* 38 (1875) pp. 153-184

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. A causa della sua apparente ingenuità quest'idea fu abbandonata per circa trent'anni, ripresa da Arner⁴ nel 1908 e di nuovo abbandonata.

Una linea di ricerca indipendente dalla precedente fu avviata nello stesso periodo dallo studio di F. Galton (1822-1911), secondo cugino di C. Darwin, e del matematico Rev. H.W. Watson (1827-1903), i quali 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, si elevasse fino alla quasi certezza.⁵ In realtà tale risultato vale soltanto nel caso in cui il numero medio di discendenti maschi sia minore o uguale a uno, come accade in una popolazione numericamente stabile o decrescente mentre in caso contrario esiste sempre un'elevata probabilità (superiore al 50%) che un cognome prima o poi si estingua, ma non si arriva mai alla certezza. Questo scenario corrisponde perfettamente a ciò che osserviamo nelle nostre società passate e presenti, sia nel caso di popolazione crescente sia in quello opposto. 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 Lotka⁶ a partire dal 1931.

Non v'è dubbio tuttavia che l'interesse per gli effetti genetici della consanguineità non era destinato a diminuire, specialmente dopo la riscoperta delle leggi di Mendel nel 1900 ad opera dei biologi H. de Vries, C. Correns e E. Tschermack von Seysenegg (ognuno dei quali arrivò indipendentemente alle stesse conclusioni cui era arrivato nel 1866 il dimenticato monaco agostiniano); in particolare la diffusione dell'emofilia nelle famiglie reali europee tra la fine dell'Ottocento e l'inizio del Novecento obbligò a dettagliate analisi della discendenza della regina Vittoria per poter finalmente comprendere la complessa dinamica che sta alla base della trasmissione ereditaria della patologia.

Una svolta nell'utilizzo delle relazioni genealogiche nello studio della genetica delle popolazioni umane si ebbe però soltanto nel 1965, quando Crow e Mange⁷ proposero un modello matematico per la stima della consanguineità a partire dai cognomi. Essi definirono la consanguineità totale di una popolazione e le sue componenti casuale e non casuale, delle quali soltanto la prima può essere spiegata dall'accoppiamento casuale (*random mating*) degli individui presenti in una popolazione. Fecero quindi le ipotesi restrittive 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 tra i nuovi arrivati nella comunità per effetto di fenomeni di immigrazione. Sotto queste ipotesi essi osservarono che, per la quasi totalità delle tipologie di parentela l'isonimia nei matrimoni (*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. Notiamo infatti che, secondo la nozione *standard* di consanguineità (che si ricava dal grado di parentela), il grado di consanguineità tra

⁴ G.B.L. Arner, *Consanguineous marriages in the American population*, Columbia U. Studies in History, Economics and Public Law (XXXI) 3, New York 1908

⁵ 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) pp. 138-144

⁶ A.J. Lotka, *The extinction of families I-II*, J. Wash. Acad. Sci. 21 (1931) pp. 377-380, 453-459

⁷ 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) pp. 199-203;

J.F. Crow, *The estimation of inbreeding from isonymy*, Human Biology 52 (1980) pp. 1-14

fratelli è $\frac{1}{4}$, quelli tra cugini è $\frac{1}{16}$, quello tra secondi cugini è $\frac{1}{64}$; ma notiamo che i fratelli (germani) hanno tutti lo stesso patronimico, mentre in media solo un cugino primo su quattro e solo un secondo cugino su sedici ha lo stesso cognome. Come si vede, in tutti i casi considerati (e in quasi tutti gli altri casi possibili) il grado di isonimia è esattamente quattro volte maggiore del grado di consanguineità. Di conseguenza il numero effettivo di matrimoni tra consanguinei è tanto maggiore di quello indicato dall'isonimia quanto più remota è la parentela, ma in compenso la consanguineità tra i discendenti di matrimoni tra parenti è ridotta in proporzione, e i due effetti si compensano. Pertanto la percentuale dei matrimoni isonimi (nelle ipotesi indicate) sarebbe una misura diretta del grado di consanguineità della popolazione nel suo complesso.

Nel 1977 l'antropologo Lasker⁸, 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.

Nel frattempo un crescente interesse dei genetisti per la distribuzione dei cognomi nasceva dal progressivo affermarsi della teoria neutrale dell'evoluzione, che poneva l'enfasi sullo studio del comportamento dei cosiddetti "geni neutri", ovvero quei geni che, non essendo portatori di informazioni determinanti per lo sviluppo del fenotipo, non sono soggetti alla pressione evolutiva causata dalla selezione naturale e quindi si trasmettono da una generazione all'altra con meccanismi governati puramente dalle leggi della statistica. Non sfuggì agli studiosi il fatto che le modalità di trasmissione del cognome nelle società occidentali sono (o almeno sono state fino a tempi recentissimi) perfettamente analoghe a quelle con cui passa da una generazione all'altra il cromosoma Y, che si trasmette pressoché inalterato (a meno di mutazioni casuali) dal padre a ciascun figlio maschio. Pertanto la distribuzione dei cognomi dovrebbe essere assimilabile a quella dei geni neutri, ma certamente assai più facile da studiare empiricamente anche su larghissima scala (fino a milioni di individui).⁹

L'apparizione nel 1967 dell'articolo di Karlin e McGregor¹⁰ che formulava la teoria del comportamento delle mutazioni neutre in popolazioni finite di dimensione costante attirò quindi l'attenzione degli studiosi di genetica verso lo studio della distribuzione e dell'estinzione dei cognomi, a partire dal lavoro di Yasuda *et al.*¹¹ pubblicato nel 1974, nel quale la teoria di Karlin e McGregor fu applicata ai dati raccolti tra la popolazione della valle del Parma da Cavalli-Sforza e collaboratori fin dal 1954 con un'ampia ricerca basata sullo studio delle dispense per matrimoni tra consanguinei e dei registri parrocchiali.¹²

Tuttavia proprio il confronto tra i modelli teorici e i risultati degli studi empirici mise presto in evidenza alcuni gravi limiti concettuali di questa metodologia di ricerca. In primo luogo è assai evidente la debolezza dell'ipotesi di monofiletismo, in quanto la documentazione disponibile permette di constatare che numerosi cognomi, e in particolare proprio i più diffusi, traggono origine da caratteristiche fisiche, professioni o provenienza geografica che si ripropongono in modo del tutto indipendente in luoghi e tempi differenti dando luogo a un polifiletismo diffuso di cui risulta però molto difficile valutare l'impatto quantitativo sull'insieme della popolazione, se non tramite uno studio analitico dei singoli cognomi, della loro origine e della loro distribuzione geografica. Ma

⁸ G.W. Lasker, *A coefficient of relationship by isonymy: A Method for Estimating the genetic Relationship between Populations*, Human Biology 49 (1977) pp. 489-493

⁹ Per una rassegna e una discussione *cfr* G. Redmonds, T. King and D. Hey, *Surnames, DNA and family History*, Oxford University Press, Oxford 2011

¹⁰ S. Karlin and J. McGregor, *The number of mutant forms maintained in a population*, Proc. 5th Berkeley Symp. Math. Stat. Prob. 4 (1967) pp. 415-438

¹¹ 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) pp. 123-142

¹² Per una rivista *cfr* L.L. Cavalli-Sforza, A. Moroni and G. Zei, *Consanguinity, Inbreeding and Genetic drift in Italy*, Princeton University press, Princeton 2004

anche l'ipotesi di una equa ripartizione tra i generi delle dinamiche migratorie sembra priva di qualsiasi fondamento storiografico, e anche in questo caso soltanto studi di demografia storica potrebbero permettere una valutazione quantitativa del fenomeno. Si noti che in entrambi i casi si tratta di ricerche che esorbitano completamente dall'ambito delle competenze professionali dei genetisti i cui studi sono quasi sempre avvenuti in contesti del tutto estranei a quello degli studi genealogici, nella pretesa (o nell'illusione) che una risposta di tipo statistico potesse in qualche modo aggirare l'esigenza di un solido fondamento storiografico, che invece proprio le criticità sopra menzionate avrebbero reso non solo auspicabile, ma probabilmente addirittura indispensabile.

L'effetto delle nascite illegittime, con conseguente abbandono del neonato e attribuzione allo stesso di un cognome non correlato a quello dell'effettiva paternità potrebbe essere a sua volta oggetto di studi statistici basati sulla documentazione disponibile e quindi in qualche modo quantificato e introdotto nel modello per correggere le stime per eccesso del numero di mutazioni. Tuttavia un ulteriore problema, che purtroppo nemmeno la ricerca storica potrebbe affrontare in modo conclusivo, riguarda l'effettiva entità del fenomeno delle "false paternità" originate da vicende adulterine. Su questo tema non si può nemmeno procedere per inferenza da dati statistici (eventualmente) ricavabili dai campionamenti effettuati al tempo presente, trattandosi di eventi fortemente condizionati dall'epoca e dalla condizione sociale, per cui non si può applicare lo stesso schema interpretativo da un lato a società caratterizzate da una forte segregazione di genere e dall'altro a situazioni di grande promiscuità in cui è ridotta la certezza della paternità biologica.

Di fatto alcune di queste difficoltà sono oggi talvolta considerate come questioni di cui non più preoccuparsi, in quanto l'evoluzione delle tecniche di analisi e di decrittazione del DNA permetterebbero di ricostruire, sempre su base statistica, le relazioni e le distanze genetiche esistenti tra gli individui e tra i gruppi, in particolare mediante la costruzione di un "orologio genetico" che individua la data di separazione tra due linee di discendenza sulla base del numero di differenze presenti nel rispettivo patrimonio genetico e del tempo medio intercorrente tra due mutazioni. Questi presupposti hanno dato origine a una vasta letteratura, che ha portato all'individuazione e alla datazione della cosiddetta "Eva africana" (che sarebbe l'antenata comune, per linea puramente femminile, di tutte le donne attualmente viventi, e questo grazie allo studio della parte condivisa del DNA mitocondriale) e di "Adamo Y-cromosomico" (antenato comune di tutti i maschi viventi individuato dalla comunanza di importanti segmenti del cromosoma Y), salva la recentissima scoperta di un gruppo sufficientemente diverso dagli altri da costringere ad antedatate di più di centomila anni il supposto antenato maschio comune all'intera umanità¹³. Ma altri risultati sono stati proposti di recente, che vanno dall'attribuzione di una comune origine per tutti i Cohen (Aronne?) alla fine del secondo millennio a.C.¹⁴ alla stima della presenza percentuale dei discendenti di Gengis Khan nelle popolazioni asiatiche¹⁵, o alla valutazione del numero di discendenti del mitico sovrano gaelico Niall, antenato degli O'Neill ma anche, secondo alcuni (ma non altri) genetisti, di almeno due o tre milioni di cittadini irlandesi.¹⁶

Un limite evidente dei risultati appena menzionati consiste nel riferimento a una singola linea di ascendenza/discendenza, quella agnaticia, nel caso del cromosoma Y, o quella matrilineare, nel caso del DNA mitocondriale. I problemi posti da questa limitazione dovrebbero essere evidenti dal punto di vista dell'antropologia fisica, poiché la "diluizione" del patrimonio genetico, per la parte più rilevante ai fini della definizione del fenotipo, è esponenziale, in quanto il contributo del corredo genico di un singolo antenato si dimezza ad ogni generazione, ma sono gravi anche dal

¹³ F. Mendez *et al.*, "An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree", *Am. Journ. of Human Genetics* 92 (2013) pp. 454-459

¹⁴ M.G. Thomas *et al.*, *Origins of Old testament priests*", *Nature* 394 (1998) pp. 138-140

¹⁵ T. Zerjal *et al.*, *The Genetic Legacy of the Mongols*, *American Journal of Human Genetics* 72 (2003) pp. 717-721

¹⁶ L.T. Moore *et al.*, *A Y-Chromosome Signature of Hegemony in Gaelic Ireland*, *American Journal of Human Genetics* 78 (2006) pp. 334-338

punto di vista dell'antropologia culturale, in quanto anche in una società patriarcale è inimmaginabile che i modelli comportamentali e i rapporti sociali fossero determinati esclusivamente dall'appartenenza al *clan* paterno, senza alcun riferimento ai legami cognatizi che invece sono sempre stati fondamentali nella formazione delle "alleanze" (vocabolo il cui corrispettivo francese non a caso designa anche i matrimoni).

Verifiche genealogiche dei modelli genetici: esperienze e proposte

Lungi da noi ogni facile ironia sullo "stile" di queste analisi scientifiche, che hanno comunque alle spalle, al di là del *format* della comunicazione, un'ampia messe di interessanti dati sperimentali. Ci pare comunque abbastanza evidente che un rinnovato ed esteso allargamento delle concrete conoscenze genealogiche potrebbe invece rappresentare un fondamentale correttivo all'ingenuità (e di fatto alla limitata scientificità) di un approccio puramente statistico, per di più irrigidito dalle limitazioni "tecniche" di cui abbiamo appena parlato. Si tratterebbe di applicare un protocollo che dovrebbe essere proprio di tutte le ricerche che, in una scienza "naturale" come la genetica, abbiano la pretesa del rigore scientifico, ovvero il controllo empirico delle ipotesi formulate mediante procedure volte ad attestare la coerenza delle osservazioni con le predizioni del modello adottato.

I modelli matematici che stanno alla base di tutte le teorie relative alla consanguineità sono fondati su ipotesi semplificatrici di cui non è possibile effettuare verifiche su base sincronica, mentre soltanto un approccio diacronico consentirebbe il loro consolidamento o la loro eventuale falsificazione. Ma nel caso in esame un approccio diacronico significa esattamente un approccio genealogico, perché soltanto una concreta ricostruzione delle effettive parentele consentirebbe di controllare se le stime teoriche di distanza genetica abbiano un effettivo riscontro fattuale.

Date queste premesse dovrebbe essere abbastanza chiaro che il tipo di ricerche funzionali alle verifiche di interesse per la genetica delle popolazioni non può essere soltanto quello tradizionalmente legato a singoli personaggi o specifici *clan* familiari, ma dovrebbe estendersi per quanto possibile a gruppi abbastanza ampi da poterne trarre conclusioni statisticamente attendibili.

In effetti alcune iniziative sono già in atto da qualche anno nel mondo anglosassone, grazie anche a una consolidata tradizione di ricerca genealogica per così dire "di massa", affidata non soltanto a ricercatori professionisti ma anche a volontari, usualmente motivati dal desiderio di ricostruire reti ampie di relazioni familiari, spesso ricercando antenati condivisi di soggetti accomunati da uno stesso cognome. Senza alcuna pretesa di effettuare una ricognizione sistematica degli studi in cui i risultati delle ricerche genealogiche sono stati utilizzati per trarne conclusioni rilevanti per l'antropologia fisica, vogliamo qui ricordare soltanto, a mero titolo di esempio, il recentissimo utilizzo critico di un immenso *database* (Geni), disponibile *online* e comprendente 86 milioni di profili, al fine di estrarne importanti informazioni sulla possibile relazione tra patrimonio genetico e longevità e sulla velocità media della dispersione geografica delle famiglie.¹⁷ Alcuni risultati in tale direzione erano stati peraltro anticipati già nel 2015 da M. Fire e Y. Elovici.¹⁸

Anche senza raggiungere le dimensioni di studi di ampiezza quantitativa tale da poter supplire con la mera forza della statistica ai limiti di scarso rigore dei singoli dati disponibili, sarebbero certamente rilevanti analisi che si concentrassero, con la necessaria maggiore acribia, su specifici gruppi sociali o su popolazioni relativamente stabili e riferibili ad ambiti geografici ben definiti, come ad esempio studi che riguardino *élites* politico-sociali, per le quali è più facile raccogliere una documentazione estesa su lunghi archi di tempo, oppure si concentrino su comunità locali con scarsa dinamica migratoria su cui si possano effettuare ricerche specifiche e sistematiche. Diremo fra poco qualcosa di queste due tipologie, specie la prima. Non sono pochi gli esempi di studi effettuabili, o anche in parte già effettuati, nello spirito qui invocato di un confronto tra modelli

¹⁷ J. Kaplanis *et al.*, *Quantitative analysis of population-scale family trees using millions of relatives*, <http://biorxiv.org/content/early/2017/02/07/106427>

¹⁸ M. Fire and Y. Elovici, *Data mining of Online Genealogy Datasets for Revealing Lifespan Patterns in Human Population*, ACM TIST 6(2): 28:1-28:22 (2015)

astratti di dinamica delle popolazioni e la concreta storicità delle vicende genealogiche, individuali e di gruppo.

Richiamiamo a questo proposito la problematica aperta dall'articolo seminale di B. Derrida, S.C. Manrubia e D.H. Zanette, che nel 1999 pubblicarono nelle *Physical Review Letters* (la più autorevole rivista di settore) la loro proposta di un modello matematico volto a predire la distribuzione asintotica (tendenziale) delle ripetizioni degli antenati per il caso di una popolazione costante (o crescente con un tasso definito) nell'ambito della quale avvengano in ciascuna generazione accoppiamenti prevalentemente casuali con distribuzioni altrettanto casuali del numero di discendenti per coppia.¹⁹

Il tema affrontato si lega strettamente al tentativo di superare alcune delle aporie concettuali che abbiamo sopra evidenziato, tenendo conto da un lato dell'importanza dei legami "orizzontali" tra le famiglie dovuti alle relazioni cognatizie, dall'altro della conseguente possibilità, o addirittura della necessità, in una popolazione numericamente chiusa, di ritrovare più volte alcuni individui nelle *Ahnentafeln* dei loro discendenti, mentre per altri individui si assiste, dopo alcune generazioni, alla scomparsa non solo del nome, ma dell'intera discendenza. Non occorre sottolineare la rilevanza genetica e sociale di queste dinamiche, e l'importanza di una loro piena comprensione, se possibile anche quantitativa.

Tuttavia non si può mancare di rilevare, in un lavoro di tale valore concettuale, la mancanza di un serio tentativo di suffragare i risultati teorici con una significativa evidenza sperimentale. Gli autori si limitarono a "scaricare" dalla rete la tavola degli antenati di un sovrano inglese di epoca medievale, Edoardo III (1312-1377) e confrontarla (peraltro con un esito assai poco convincente) con qualche elementare predizione del loro modello. È facile convincersi dei limiti di qualità, di quantità e di estensione temporale dei dati disponibili: nella tavola degli antenati di Edoardo III i primi seri dubbi compaiono già alla sesta generazione (inizio del XIII secolo) e alla decima generazione manca già più del 10% degli antenati, mentre dai nostri studi sappiamo che occorrerebbero diverse altre generazioni "quasi" complete per entrare nel regime in cui il modello diventa realmente predittivo. Ma soprattutto esiste un'evidente contraddizione tra asserzioni di natura statistica e una "verifica" basata su un singolo caso.

Ciò detto, ci sembra che sarebbe invece del tutto praticabile un'investigazione che, usando come punto di partenza un congruo numero di soggetti vissuti in epoche assai più a noi prossime, ne ricostruisca *Ahnentafeln* più lunghe e più complete, e le analizzi con adeguati strumenti statistici, al fine di verificare il modello di Derrida *et al.* e, in caso di esito positivo, ne fissi i non banali parametri (dimensioni tipiche delle popolazioni di riferimento e relativi tassi di crescita).

Esistono, come già detto, almeno due differenti tipologie di popolazioni per le quali si può immaginare, ed è stato in parte già avviato, un lavoro di ricerca che muova nella direzione testé indicata.

Da un lato candidati naturali per questo tipo di investigazioni sono gli appartenenti ai ceti nobiliari, sia per la maggior abbondanza della documentazione genealogica, sia per la relativa "chiusura" dei gruppi sociali di riferimento, che offre la possibilità di una stima quantitativa *a priori* delle dimensioni della popolazione in esame. Nel caso (da noi studiato) di alcuni recenti esponenti della casa di Bourbon, si conoscono almeno quarantamila antenati distinti, ripartiti su una trentina di generazioni "storiche" (ovvero poco contaminate da genealogie fittizie o improbabili); questo dato ne fa un campione straordinariamente più significativo di quello esaminato da Derrida *et al.*, cosa che ha reso possibile una prima (anche se non definitiva) verifica di alcune importanti

¹⁹ B. Derrida, S.C. Manrubia and D.H. Zanette, *Statistical Properties of Genealogical Trees*, *Physical Review Letters* 82 (1999) pp. 1987-1990; *Distribution of repetitions of ancestors in genealogical trees*, *Physica A* 281 (2000) pp. 1-16; *On the genealogy of a population of biparental individuals*, *Journal of Theoretical Biology* 203 (2000) pp. 303-315

predizioni del modello.²⁰ È bene a tale proposito prevenire qui una facile obiezione legata a una presunta (o presumibile) specificità dei comportamenti matrimoniali e riproduttivi dei ceti nobiliari rispetto alla maggioranza della popolazione coeva. Notiamo innanzitutto che le proibizioni canoniche relative ai matrimoni tra consanguinei venivano applicate, e violate quando lo si ritenesse opportuno, in tutte le classi sociali, come dimostrano ad esempio i già citati studi di Cavalli-Sforza e collaboratori sulla popolazione agricola della valle del Parma nel corso dell'età moderna. Aggiungiamo che le dimensioni tipiche del gruppo relativamente chiuso costituito dall'alta nobiltà europea nell'età moderna (mai molto più di un migliaio di individui "fertili" per generazione) lo rendono perfettamente paragonabile, beninteso soltanto dal punto di vista statistico, a una qualunque piccola comunità locale riferibile allo stesso periodo, che costituirebbe l'altro contesto, individuato qui sopra e su cui torneremo in chiusura, di estremo interesse per questo tipo di indagini.

Un'altra problematica nella quale può emergere una stretta connessione tra considerazioni di natura statistica basate su modelli genetici e risultati della ricerca storiografica in ambito genealogico riguarda la ricerca del cosiddetto "antenato comune più recente" (comunemente indicato con l'acronimo MRCA, *Most Recent Common Ancestor*).

Precisiamo che questo tema è particolarmente rilevante nel quadro della "teoria della coalescenza", un modello matematico della genetica delle popolazioni proposto da J. Kingman nel 1982²¹ nel quadro della teoria neutrale dell'evoluzione. Uno dei principali obiettivi del modello di Kingman e dei suoi successivi raffinamenti è quello di valutare il tempo che può essere intercorso tra il momento in cui si prende in esame una determinata popolazione e il momento in cui è vissuto il più recente antenato comune all'intera popolazione (in sigla tMRCA).

In genetica usualmente per "popolazione" si intende la distribuzione delle differenti varianti (alleli) di un determinato gene presenti nel gruppo in esame, e il tMRCA dovrebbe essere determinato, da un punto di vista statistico, dal numero di differenti mutazioni e dalla probabilità di mutazione per unità di tempo. Ma così come la distribuzione dei cognomi rappresenta una buona "simulazione" della distribuzione dei geni neutri, la ricerca del più recente antenato comune (in realtà tipicamente una coppia) per un'assegnata popolazione umana può rappresentare un modello empirico verificabile per le predizioni della teoria della coalescenza. Chiaramente gli esiti della ricerca del MRCA in una popolazione umana hanno implicazioni di antropologia culturale che vanno molto al di là dell'interesse puramente genetico, con significative ricadute anche sull'immaginario collettivo. Sulle implicazioni culturali di queste ricerche, lette in un'ottica antropologica, anche in relazione con la sempre maggior diffusione degli strumenti di comunicazione e di ricerca via rete, merita particolare attenzione il saggio del 2015 *Ancestry*, di P. G. Solinas²².

Fu seminale in quest'ambito l'articolo di Chang del 1999 in cui, sulla base di un modello non molto diverso da quello di Derrida *et al.* fu stimato il numero di generazioni che tipicamente dovrebbero intercorrere tra il MRCA e il tempo presente per una popolazione di dimensioni assegnate e sufficientemente interconnessa (escludendo quindi casi di permanente isolamento).²³ Il risultato, a prima vista abbastanza stupefacente, fu la predizione che il MRCA dell'intera umanità (escludendo eventualmente piccole comunità ancora del tutto isolate) potrebbe essere vissuto in tempi storici. Un corollario ancor meno intuitivo del modello di Chang è l'esistenza, su una scala temporale poco meno che doppia della precedente, di una "popolazione ancestrale universale" (IAP: *Identical Ancestors Point*) i cui componenti sono antenati di tutti i membri della popolazione attuale,

²⁰ P.Rossi, *Self-Similarity in Population Dynamics: Surname Distributions and Genealogical Trees*, *Entropy* 17 (2015) pp. 1-13

²¹ J.F.C. Kingman, *On the Genealogy of Large Populations*, *Journal of Applied Probability* 19 A (1982) pp. 27-43

²² P.G. Solinas, *Ancestry. Parentele elettroniche e lignaggi genetici*, Editpress, Firenze 2015

²³ J.T. Chang, *Recent common ancestors of all present-day individuals*, *Advances in Applied Probability* 31 (1999) pp. 1002-1026

o di nessuno.²⁴ Non sfuggano tuttavia due importanti considerazioni: in primo luogo il “peso” (inteso come numero delle ripetizioni nell’*Ahnentafel*) di uno specifico soggetto appartenente all’IAP è certamente molto diverso per ciascun individuo presentemente vivente, e in secondo luogo la presenza nell’albero genealogico non ha alcuna implicazione diretta sulla presenza nel corredo genetico individuale, in quanto la “diluizione” dei geni per effetto della riproduzione sessuata è esponenzialmente rapida, e quindi il “peso” del patrimonio genetico di un determinato antenato su quello di un soggetto attuale dipende fortemente dalle dimensioni della popolazione al tempo dell’IAP e dal numero di ripetizioni dell’antenato stesso.

Chiaramente la genealogia storica può fornire un contributo sostanziale alla validazione del modello di Chang e/o all’individuazione dei suoi eventuali limiti concettuali. In questa direzione si è mosso di recente il lavoro di tesi di G. Spugnesi.²⁵

Partendo da una popolazione di 48 individui vissuti intorno all’inizio del secolo scorso e adeguatamente rappresentativi della intera “popolazione” dell’alta nobiltà tedesca (*Hochadel*) e delle case sovrane europee dell’epoca, Spugnesi ne ha ricostruito le *Ahnentafeln* fino alla sedicesima generazione, (XIV-XV secolo) creando quindi più di tre milioni di schede “virtuali”, che però in virtù delle ripetizioni corrispondevano a circa 190 mila individui, dei quali circa 165 mila sconosciuti, cosa non preoccupante in quanto si trattava prevalentemente di coppie della 15esima o 16esima generazione, raramente presenti in più di una *Ahnentafel* e quindi non rilevanti per la ricerca del MRCA; in realtà grazie al meccanismo delle ripetizioni soltanto circa il 20% degli antenati non risultava noto.

Il lavoro di Spugnesi ha permesso di evidenziare che in un gruppo sociale abbastanza compatto e coeso, le cui dimensioni non hanno mai superato il migliaio di individui per generazione, è possibile, almeno per le generazioni nate nel XVIII e nel XIX secolo, individuare un antenato comune a tutto il gruppo, risalente approssimativamente a 250 anni prima della data di riferimento per il gruppo, come predetto dal modello. A mero titolo esemplificativo segnaliamo che il più recente antenato comune di tutta l’alta nobiltà europea vivente nell’anno 1900 risulta essere Joachim Ernst I, Graf zu Oettingen-Oettingen (1612-1659), che ebbe due mogli, entrambe fertili, e quindi non MRCA. Tenuto conto della durata media di una generazione (circa 33 anni nella base di dati utilizzata) e dell’interdizione al matrimonio tra fratelli, che sposta di una generazione la valutazione del tMRCA rispetto al modello statistico di Chang (che assume una totale casualità degli accoppiamenti), i risultati di Spugnesi offrono una sostanziale conferma empirica del modello, aprendo nel contempo la strada per ulteriori approfondimenti su tutte le tematiche sin qui individuate (distribuzione delle ripetizioni di antenati, distribuzione dei cognomi, consanguineità, mobilità geografica, probabilità di estinzione delle famiglie), con ricadute non soltanto sugli studi di dinamica delle popolazioni²⁶ ma anche sull’“antropologia” fisica e culturale (fertilità, longevità, politiche matrimoniali) delle “popolazioni” esaminate.

Nello spirito degli studi fin qui citati, ma spostando l’attenzione a un ambito diverso, ci sentiamo di concludere segnalando un’ipotesi di lavoro da noi avanzata qualche anno fa, e consistente nell’utilizzo del materiale raccolto in molti anni di accurate ricerche dal collega Sergio Nelli, che ha analizzato parecchie migliaia di documenti e ricostruito un gran numero di alberi genealogici

²⁴ Per una rassegna cfr S.C. Manrubia, B. Derrida and D.H. Zanette, *Genealogy in the Era of Genomics*, *American Scientist* 91 (2003) pp. 158-165

²⁵ G. Spugnesi, *MRCA e MCRA: proprietà statistiche delle genealogie nobiliari europee nell’età moderna* (tesi), Università di Pisa, A.A. 2013/14; l’acronimo MCRA (*Most Common Recent Ancestor*), qui introdotto, si riferisce agli antenati che in tempi recenti, grazie alle ripetizioni, hanno potuto dare il maggior contributo al patrimonio genetico della popolazione studiata.

²⁶ P.Rossi, *Self-Similarity in Population Dynamics: Surname Distributions and Genealogical Trees*, *Entropy* 17 (2015) pp. 1-13

relativi alla popolazione di Montecarlo di Lucca nel corso dell'età moderna.²⁷ Si tratta anche in questo caso, come in quello della nobiltà, di una comunità relativamente chiusa, per la quale quindi alcuni parametri sarebbero in qualche misura prefissati, ma con il vantaggio di esplorare un contesto sociale radicalmente differente da quelli che sono stati usualmente oggetto degli studi genealogici. Si tratta di un'esplorazione ancora molto preliminare, che richiede una sostanziale riorganizzazione dell'ampio materiale disponibile e probabilmente anche qualche ulteriore approfondimento, ma che certamente, se il progetto giungerà a compimento, offrirà un importante complemento ai risultati già acquisiti in altri contesti. Avremmo così un esempio significativo- diverso ma convergente con quelli d'ambito nobiliare- dell'utilità della genealogia storica come verifica del lavoro più teorico degli scienziati.

²⁷ S. Nelli, P. Rossi, R. Bizzocchi, *Un progetto di analisi statistica dei dati genealogici relative a Montecarlo di Lucca in età moderna*, in "L'Italia dei cognomi" (cura A. Addobbati, R. Bizzocchi, G. Salinero), pp. 209-212, Pisa University Press, Pisa 2012