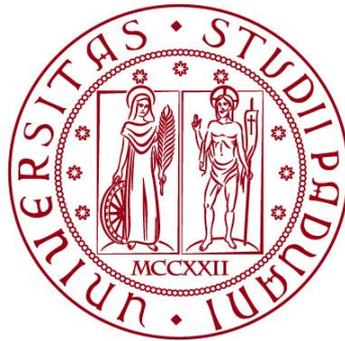


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea magistrale in Biologia Evoluzionistica



TESI DI LAUREA

**Tracce dei crolli demografici del XIV secolo
nel genoma degli europei contemporanei**

Relatore: Prof. Luca Pagani
Dipartimento di Biologia

Correlatore: Prof. Cristian Taccioli
Dipartimento di Medicina Animale, Produzioni e Salute

Laureando: Federico De Pizzol

ANNO ACCADEMICO 2021/2022

Indice

0. Abstract.....	1
1. Introduzione	3
1.1. La Peste Nera	3
1.2. Dimensione effettiva della popolazione.....	5
1.3. Le variazioni del passato nel genoma dei contemporanei.....	6
1.4. Scopo della tesi	7
2. Materiali e Metodi	8
2.1. Dati genetici utilizzati	8
2.2. Dati storici utilizzati.....	11
2.3. PCA.....	12
2.4. IBDseq	12
2.5. IBDNe.....	13
2.5.1. Consistenza tra cromosomi	13
2.5.2. Pacchetto di R per identificare picchi e valli	13
2.6. Simulazioni	14
3. Risultati.....	15
3.1. Rimozione di outliers e identificazione di sottogruppi	15
3.2. Confronto WGS e SNP array	16
3.3. Validazione sui vari cromosomi	18
3.4. Picchi e Valli.....	19
3.5. Confronto fra profilo simulato e profilo inferito da IBDNe	20
3.6. Europa.....	22
3.7. Asia Orientale	26
3.8. Subcontinente indiano.....	28
3.9. Via della Seta	30
3.10. Etiopia ed Egitto.....	34
3.11. Africa e America.....	35
4. Discussione.....	38
4.1. Europa.....	38
4.1.1. Spagna.....	38
4.1.2. Italia	39

4.1.3.	Gran Bretagna	40
4.1.4.	Finlandia e CEU	40
4.1.5.	Estonia.....	41
4.2.	Asia orientale	42
4.3.	Subcontinente indiano	42
4.4.	Via della Seta	43
4.5.	Etiopia ed Egitto.....	43
4.6.	Africa e America	44
4.7.	Interpretazione generale dei risultati	44
5.	Conclusioni.....	46
6.	Bibliografia	47

0. Abstract

Il XIV secolo è ricordato da tutti come il periodo più buio della storia Europea. Dall'oriente, una nave genovese in fuga dall'assedio di Caffa giunge sulle coste messinesi nel 1347, dando inizio a un'epidemia destinata a decimare la popolazione europea: la tristemente nota Peste Nera. Nonostante la grande fama di questo fenomeno, i dati storici che riguardano la reale dimensione della popolazione dei vari paesi europei (e dunque i relativi crolli demografici) sono pochi e spesso discordanti tra loro. Inoltre, non è chiaro quanto questo calo demografico abbia avuto un impatto sulla genetica di popolazione Europea. A sostegno delle testimonianze storiche, ad oggi è possibile impiegare strumenti genetici per stimare la dimensione effettiva (N_e) di una popolazione del passato, utilizzando i genomi dei discendenti odierni di coloro che vissero la Peste Nera in prima persona. Partendo dai dati genomici di 5789 individui del *1000 Genomes Project* e di altri studi precedenti, lo studio si pone l'obiettivo di ricostruire le modifiche nella dimensione effettiva di popolazione di vari paesi eurasiatici e non, negli ultimi 2000 anni, tramite segmenti di DNA "identici per discendenza" (IBD), per confrontarli con i dati storici disponibili. Per quanto le varie popolazioni mostrino andamenti differenti nella variazione della loro dimensione, nei casi Europei notoriamente colpiti dalla peste del 1300 (e dei secoli successivi), si nota un calo drastico della popolazione, che vede il proprio recupero solo a partire dal 1600. Le stime a partire dai dati genetici, per quanto approssimative, sono in grado di dare un'idea piuttosto corretta dell'andamento di una popolazione nel corso del tempo, fornendo così la possibilità di esplorare la storia di quei popoli per i quali il record storiografico è ad oggi carente.

14th century is known to all as the darkest period in European history. From the east, a Genoese ship fleeing from the siege of Kaffa docked on the shores of Messina in 1347, giving rise to an epidemic destined to decimate European population: the infamous Black Death. Despite the fame of this phenomenon, there is little and conflicting historical data about the real population size of the European countries and the related demographic collapses. Moreover, it is not clear how much this demographic collapse affected the European population genetics. In support of historical testimonies, it is currently possible to employ genetic tools to estimate the effective size (N_e) of a population in the past, using the genomes of modern descendants of those who experienced first-hand the Black Death. Starting from the genomes of 5789 individuals of the *1000 Genomes Project* and other previous studies, this study aims to reconstruct the variations in the effective population of several Eurasian and non-Eurasian countries, in the last 2000 years, by the use of segments of DNA “identical by descent” (IBD), to compare them with historical records. Even if the trends of the various populations are different from one another, we are able to notice a drastic decline in European populations notoriously hit by the plague in 1300 (and the further centuries), seeing a recovery only in 1600. Genetic estimates, despite being imprecise, are able to give a correct idea of the trend of a population through time, giving the possibility to explore the history of those people lacking an historical record.

1. Introduzione

1.1. La Peste Nera

Quello che si conosce dell'arrivo della peste in Europa si colloca a metà tra storia e leggenda e vede l'assedio di Caffa tra il 1344 e il 1346 come il primo evento noto che coinvolga il morbo. La narrazione che si fa dell'assedio di Caffa, in Crimea, vedrebbe le truppe di Yanibeg, Khan dell'Orda d'Oro, lanciare corpi infetti con le catapulte oltre le mura della città per sterminarne la popolazione, mentre le navi genovesi assediata fuggono via mare portando con se il bacillo di *Yersinia pestis* (Naphy & Spicer, 2004). Quanto ci sia di veritiero sul precoce utilizzo di "armi batteriologiche" nella storia non è noto, ma l'arrivo della peste in Europa è ben documentato da centinaia di fonti dell'epoca tra cronisti, religiosi, artisti e poeti. Dopo la partenza da Caffa, nel 1347 la peste giunse a Costantinopoli, Alessandria e Cipro e in breve tempo attraverso le rotte commerciali del mediterraneo raggiunse i porti occidentali di Messina, Genova, Firenze e Venezia. Nel 1348 arrivò a Marsiglia e iniziò a diffondersi nell'entroterra Francese, Italiano ed Egiziano. Negli anni successivi si diffuse in tutta Europa, fino in Scandinavia (C. Wickham, 2016). Nel 1350 il morbo aveva raggiunto La Svezia, la Norvegia, la Scozia e perfino la Groenlandia portando all'abbandono degli ultimi insediamenti europei in quest'ultima (Naphy & Spicer, 2004). Non tutte le città europee vennero colpite dalla prima ondata dell'epidemia, la cosiddetta "Peste Nera" (1347-1353), ma nessun luogo sfuggì alle successive (Barbieri et al., 2021). L'arrivo della peste in Europa, come detto, è ben documentato, ma la sua origine resta ancora oggi un campo dibattuto. L'ipotesi più plausibile per quanto riguarda la culla del morbo è che questo si sia sviluppato agli inizi del XIV secolo in Asia centrale e che sia stato diffuso dall'espansione dell'impero mongolo. La prima testimonianza certa che si ha della peste sono due cimiteri nella valle di Chüy in Kirghizistan, che riportano sepolture di massa nel 1338 e 1339 recando sulle pietre tombali riferimenti a una non precisata epidemia. Di recente le analisi dei resti di alcuni corpi in questi cimiteri hanno rivelato che non solo erano affetti da *Y. pestis*, ma che si tratta di un ceppo che pare aver dato origine sia ai ceppi che invasero l'Europa del tempo, che a quelli moderni (Spyrou et al., 2022).

La peste è una zoonosi, ovvero una patologia che può essere trasmessa dagli animali all'uomo, la cui persistenza nell'ambiente è dovuta a una grande varietà di roditori e (soprattutto) alle pulci che parassitano questi ultimi. La pulce è il vero serbatoio del batterio Gram negativo *Yersinia pestis* poiché non sviluppa alcuna patologia in seguito all'infezione. Esistono cinque forme principali di peste: la peste bubbonica, la peste polmonare, la peste faringea, la peste meningea e la peste setticemica (Perry & Fetherston, 1997) caratterizzate da diverse modalità di infezione e sintomi ma causate dallo stesso agente eziologico. La peste bubbonica è la più nota, la più frequente ed è l'unica malattia conosciuta a portare

rigonfiamenti dei linfonodi chiamati “bubboni” e causare epidemie mortali (Barbieri et al., 2021). Tra i due e i sei giorni circa dall’infezione, generalmente causata dal morso di una pulce, l’individuo infetto sviluppa febbre, mal di testa, brividi, dolori muscolari e affaticamento; i linfonodi nella regione dell’ingresso del patogeno diventano gonfi ed estremamente dolorosi, dopodiché anche i linfonodi ascellari e del collo vanno incontro alla medesima sorte e se non trattata adeguatamente, la malattia procede evolvendosi in peste setticemica con l’ingresso del patogeno nel sangue (Adamovicz & Worsham, 2012; Perry & Fetherston, 1997), causando il decesso nel 50% dei casi non trattati. La peste setticemica porta alla morte di cellule e tessuti e gangrena delle estremità. Se il batterio raggiunge i polmoni la peste setticemica può portare a peste polmonare, che può essere contratta anche per infezione tramite via aerea da un essere umano all’altro. Si tratta della forma più rara, poiché generalmente il malato decede (o guarisce) prima di giungere a questo stadio e perché la trasmissione della peste per via aerea è inefficiente se confrontata con altre patologie respiratorie. La peste polmonare ha un alto tasso di mortalità anche con trattamento poiché porta a insufficienza respiratoria e shock. La peste meningea generalmente è una conseguenza della peste bubbonica nei bambini, che si sviluppa appunto in meningite. La peste faringea è invece solitamente conseguenza dell’ingestione di carne cruda o (più raramente) dell’inalazione di aerosol infetto e se non trattata può evolvere in peste polmonare (Adamovicz & Worsham, 2012; Perry & Fetherston, 1997). Ad oggi la peste si presenta in modo sporadico o in epidemie molto limitate. La mortalità dei malati di peste è piuttosto bassa se adeguatamente trattati con terapie antibiotiche in tempi rapidi, senza dare modo al batterio di diffondersi eccessivamente nell’organismo. Naturalmente nel XIV secolo l’assenza di terapie e di conoscenza riguardo la malattia e la microbiologia, unitamente alle scarse condizioni igieniche, portava alla morte con una frequenza molto maggiore rispetto al giorno d’oggi.

Le fonti riguardanti la quantità di morti in Europa nel medioevo sono molteplici, spesso discordanti e riportate goffamente nel corso degli anni compiendo numerosi errori (Barbiera et al., 2016; Roosen & Curtis, 2018). Generalmente si parla della morte di un terzo della popolazione europea (Barbieri et al., 2021), ma non è raro trovare fonti che parlino della morte di metà della popolazione totale (Rosenwein, 2014). Per quanto riguarda le fonti dell’epoca, Papa Clemente VI valutò che la prima ondata avesse ridotto la popolazione europea del 31% ma alcune grandi città come Parigi pare persero metà della popolazione (Rosenwein, 2014) e Venezia oltre il 60% (Naphy & Spicer, 2004). Osservando la totalità delle fonti si rischia di essere sovrachiati dalla quantità di dati sulla mortalità della peste riguardanti specifiche città, regioni o addirittura precise stagioni, ma sempre senza ottenere una risposta chiara sul fenomeno complessivo. Oltre a variare molto i dati disponibili all’interno dello stesso paese, pare che l’effetto della peste sia stato molto vario tra gli stati europei non avendo quindi un impatto ugualmente dannoso ovunque (Barbiera et al., 2016; Izdebski et al., 2022). Non è raro inoltre

trovare fonti che spieghino come l'impatto della peste sia stato sovrastimato (Izdebski et al., 2022), così come viceversa, altri scritti sostengano che abbia avuto un effetto più disastroso di ciò che si pensa (Green, 2020). In sintesi, la grande varietà di informazioni sulla Peste Nera, di dati storici, di stime e di opinioni è andata a formare una matassa, arricchita anche da suggestioni letterarie e pittoriche, che la ricerca continua a tentare di sbrogliare un passo alla volta.

1.2. Dimensione effettiva della popolazione

Con dimensione effettiva della popolazione (N_e) si intende il numero di individui di una popolazione, nelle condizioni del modello scelto (come ad esempio Wright-Fisher), che produrrebbe lo stesso effetto di deriva genetica che si osserva nella popolazione reale in esame. Si può basare la definizione di N_e sulla varianza campionaria delle frequenze alleliche, sul grado di inbreeding della popolazione, oppure si può definire come il numero di individui che contribuiscono alla generazione successiva, ma generalmente queste definizioni sono sovrapponibili l'una all'altra (Jobling et al., 2014). Il concetto di N_e è importante poiché le popolazioni reali tendono a evolvere in modi complessi, difficili da adattare alla perfezione a un modello (Nielsen & Slatkin, 2013). Sono tre le assunzioni fondamentali per cui una popolazione reale è inadatta per essere considerata all'interno di un modello: la dimensione reale di una popolazione non è quasi mai costante nel tempo, le generazioni si sovrappongono le une alle altre e infine non si verifica l'accoppiamento casuale tra gli individui (Jobling et al., 2014). È importante tenere presente che per come è strutturato il genoma umano, vi è una differenza tra la trasmissione degli autosomi, dei cromosomi X e Y e del DNA mitocondriale (Jobling et al., 2014). Nel presente studio verranno presi in considerazione unicamente gli autosomi per facilità di analisi ma la storia dei cromosomi sessuali e del mtDNA sono di grande importanza nello studio dell'evoluzione della nostra specie e delle popolazioni umane e sono state in grado di rivelare importanti informazioni ad esempio sulla storia di Neanderthal (Meyer et al., 2014), delle popolazioni africane (Arredi et al., 2004) o sull'espansione austronesiana (Skoglund et al., 2016).

Non è facile collegare la dimensione effettiva di popolazione (N_e) alla popolazione censita (N_c) ovvero il numero di individui realmente esistenti. Molti parametri influenzano questa relazione tra cui le fluttuazioni in N_c , il successo riproduttivo degli individui, strutture interne alla popolazione e l'accoppiamento assortativo per ciascuna delle generazioni considerate. Generalmente N_e è inferiore rispetto a N_c ma bisogna tenere conto che ci sono delle differenze tra la popolazione effettiva recente e quella sul lungo periodo e che esistono metodi differenti per calcolarle e diverse stime per risalire a N_c . Questo studio si occupa di popolazione effettiva recente e per legarla alla popolazione censita ci si è affidati ai calcoli svolti in uno studio precedente (Pankratov et al., 2020) che,

tenendo conto di varie assunzioni calibrate sulla biologia delle popolazioni umane, propone la seguente relazione per gli autosomi:

$$N_c = 4.89 \times N_e$$

Uno studio della N_e delle generazioni passate a partire da dati genomici moderni (ovvero dei sopravvissuti), può portare evidenze complementari a quelle storiche che, tenendo conto delle limitazioni intrinseche a questo approccio, possono fornire un nuovo punto di vista nel dibattito sull'impatto demografico reale della prima grande epidemia di peste e di quelle dei secoli successivi.

1.3. Le variazioni del passato nel genoma dei contemporanei

Per considerare le popolazioni moderne come discendenti dei sopravvissuti del passato, è necessario capire dove guardare nel genoma dei moderni. Un collegamento naturale tra ieri e oggi sono i segmenti IBD (identical by descent), segmenti di DNA di due individui che possono essere fatti risalire a un antenato comune nel passato. Per tentare di spiegare come IBD e variazioni nella dimensione di popolazione siano legati tra loro è bene chiarire che ciascun individuo umano è immaginabile come costituito da 2 individui aploidi ovvero metà del proprio padre e metà della propria madre. Se consideriamo ciascun individuo diploide come la somma di due individui aploidi, possiamo parlare di popolazione di individui aploidi anche trattando di esseri umani.

La teoria del coalescente (Kingman, 1982) indica che a uno specifico locus del loro genoma, due gameti aploidi da una popolazione Wright-Fisher di dimensione costante (N_e) hanno una probabilità di $1/N_e$ di trovare un antenato comune alla generazione precedente. Il tempo (in generazioni dal presente "gbp") per questi due gameti per risalire all'antenato comune più recente (Time from the Most Recent Common Ancestor, TMRCA) ha quindi un valore atteso in unità di N_e . Più in generale, se una popolazione è composta da N individui aploidi alla generazione g , allora la probabilità di trovare un antenato comune in quella generazione è $N(g)^{-1}$. La relazione tra la probabilità di trovare un antenato comune e la dimensione di popolazione è la chiave per tentare di ricostruire la dinamica demografica della popolazione. Si può infatti studiare la distribuzione delle medie dei TMRCA lungo il genoma per varie coppie di individui e ottenere quindi informazioni sulla dimensione di popolazione a diverse scale temporali. L'estensione di un aplotipo IBD determina probabilisticamente il TMRCA per i due individui al locus considerato. Sfortunatamente i segmenti individuali portano poche informazioni riguardo a specifici siti se il MRCA non è molto recente. Tuttavia essendo interessati in analisi genome-wide e di popolazione, informazioni significative possono essere raccolte dal grande numero di segmenti coereditati da diverse coppie di individui della stessa popolazione. Il numero considerato cresce quadraticamente con la dimensione del campione e il numero

di segmenti IBD attesi aumenta man mano che le lunghezze dei segmenti diminuiscono (Palamara et al., 2012). In sintesi, più è lungo un segmento IBD, più recente sarà l'antenato comune tra i due individui che condividono il segmento e come visto in precedenza la generazione a cui appartiene l'antenato comune è collegata a N_e . Più antenati comuni avrò a una generazione e meno persone ci si aspetta fossero vive o attive in quella generazione.

1.4. Scopo della tesi

Dopo aver realizzato le curve N_e /Anno per tutti i dati di *1000 Genomes Project* e altri dati aggiuntivi, lo studio si propone di confrontare questi dati di origine genetica con i record storici riguardanti la dimensione delle varie popolazioni (ove disponibili), con l'obiettivo di identificare, se presenti nel dato genetico, tracce del crollo demografico dovuto alla Peste Nera del 1347-53. Inoltre tenta di dare una spiegazione plausibile e collegabile a eventi storici particolarmente rilevanti, degli andamenti ottenuti anche quando non correlabili alla peste, come ad esempio nel caso dell'America. La grande applicabilità di questo metodo di ricostruzione della dimensione di popolazione consiste nel poter utilizzare dati di DNA moderni per ottenere informazioni aggiuntive sulla storia umana o ricostruire l'andamento demografico di paesi per cui i record storici sono carenti o assenti.

2. Materiali e Metodi

2.1. Dati genetici utilizzati

Il dataset principale utilizzato è quello di 1000 Genomes Project (Auton et al., 2015), con tutte le popolazioni disponibili. Per la presente tesi si è scelto di utilizzare i genomi sequenziati a media copertura (8x), allineati alla build 37 del genoma umano e successivamente riannotati sulla build 38 scaricati dal seguente link in data 24/11/2021:

http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/phase3_liftover_nygc_dir/.

I dati utilizzati sono WGS (whole genome sequencing, solo autosomi) e si tratta di 81647806 snps biallelici per ciascuna delle popolazioni di 1KG (1000 Genomes Project). Sono in seguito riportate nella *Tabella 2.1.1* le varie popolazioni con i relativi ID e numero di individui.

In aggiunta ai dati del 1000 Genomes Project, si è deciso di integrare anche dati di individui provenienti da vari paesi lungo la “via della seta” e genotipizzati su SNP array (Omni Express 700k Illumina Chip), definiti in seguito come SILK ROAD (Mezzavilla et al., 2014) ottenuti dal laboratorio del Prof. Paolo Gasparini dell’Università di Trieste. Dal dataset iniziale sono state selezionate solo le popolazioni che disponessero di almeno 40 individui (Tab 2.1.1 giallo). Questi dati non sono WGS, e si tratta di un set di soli 579384 SNPs. Per verificare che l’utilizzo di un numero inferiore di SNP non inficiasse il risultato di IBDNe è stato creato un subset degli SNP di TSI (TSI-snpchip) e confrontato con TSI-wgs (Vedi paragrafo 3.2).

Considerando la vicinanza con il continente eurasiatico sono state aggiunte anche due popolazioni del nord-est africano: EGITTO ed ETIOPIA (Pagani et al., 2015). Per quanto riguarda le popolazioni interne al set ETIOPIA sono state selezionate solo quelle omogenee tra loro con admixture Est-Africa/Ovest-Asia circa del 50%: AMHARA, OROMO, WOLAYTA, SOMALI (Pagani et al., 2015), riassunte per semplicità in “ETIOPIA”. Nonostante il dato fosse di tipo WGS, per compatibilità con le altre popolazioni non Europee si è deciso di sottocampionarlo selezionando lo stesso set di SNPs di SILK ROAD, che svolge quindi una funzione di fattore limitante per il panel di SNPs utilizzato.

Avendo a disposizione una grande quantità di dati estoni, è stata aggiunta al gruppo di popolazioni europee anche ESTONIA (Pankratov et al., 2020). Anche in questo caso utilizzando i 579384 SNPs limitati da SILK ROAD. Oltre al totale di 2420 individui che rappresentano la popolazione estone nella sua totalità, sono disponibili due sottogruppi “Harju” (302 individui) e “Parnu” (117 individui).

I formati utilizzati per i dati sono: VCF (variant call format) (Danecek et al., 2011) e PLINK file formats (Chang et al., 2015).

Popolazione	ID	N° individui
African Caribbean in Barbados	ACB	98
African Ancestry in SW USA	ASW	61
Bengali in Bangladesh	BEB	88
Dai Chinese in Xishuangbanna, China	CDX	93
Northern Europeans from Utah	CEU	108
Han Chinese in Beijing, China	CHB	104
Han Chinese South	CHS	105
Colombian in Medellín, Colombia	CLM	94
Esan in Nigeria	ESN	99
Finnish in Finland	FIN	100
British From England and Scotland	GBR	91
Gujarati Indians in Houston, Texas, USA	GIH	99
Gambian in Western Division – Mandinka	GWD	114
Iberian Populations in Spain	IBS	110
Indian Telugu in the U.K.	ITU	99
Japanese in Tokyo, Japan	JPT	104
Kinh in Ho Chi Minh City, Vietnam	KHV	99
Luhya in Webuye, Kenya	LWK	99
Mende in Sierra Leone	MSL	87
Mexican Ancestry in Los Angeles CA USA	MXL	64
Peruvian in Lima Peru	PEL	87
Punjabi in Lahore, Pakistan	PJL	96
Puerto Rican in Puerto Rico	PUR	104
Sri Lankan Tamil in the UK	STU	102
Toscani in Italia	TSI	107
Yoruba in Ibadan, Nigeria	YRI	115
ARMENIA	ARMENIA	105
AZERBAIJAN	AZERBAIJAN	67
CRIMEA	CRIMEA	87
GEORGIA	GEORGIA	128
KAZAKISTAN	KAZAKISTAN	58
KIRGHIZISTAN	KIRGHIZISTAN	184
MONGOLIA	MONGOLIA	40
TAJIKISTAN	TAJIKISTAN	102
UZBEKISTAN	UZBEKISTAN	104
EGITTO	EGITTO	108
ETIOPIA	ETIOPIA	104
ESTONIA	ESTONIA	2420

Tab 2.1.1 Panel di popolazioni utilizzate con rispettivi ID e numero di individui. Le popolazioni che presentano un ID differente dal nome della popolazione corrispondono al panel di 1000 Genomes Project. I colori sono stati scelti per indicare le aree geografiche che coinvolgono le popolazioni. In verde le popolazioni africane, in rosso le popolazioni del subcontinente indiano, in blu le popolazioni europee, in rosa le popolazioni est-asiatiche, in grigio le popolazioni americane e in giallo le popolazioni di SILK ROAD. Lo stesso schema di colori sarà mantenuto più avanti.

Per rendere più chiara la distribuzione delle popolazioni utilizzate, la Fig 2.1.2 mostra una mappa dell'Eurasia, riportando le popolazioni eurasiatiche (quindi più plausibilmente coinvolte nel fenomeno della Peste Nera).



Fig 2.1.2 Mappa dell'Eurasia con le popolazioni più rilevanti. In blu "1000 Genomes Project". In arancione SILK ROAD. In verde Egitto ed Etiopia. In rosso Estonia. La linea viola indica il percorso della via della seta nel 1347.

Tutti i grafici presentati sono stati utilizzati mediante l'ausilio di ggplot2 (H. Wickham, 2016) per R (RCore, 2021).

Tutte le tabelle sono state realizzate con Microsoft Excel (Microsoft, 2010).

La mappa in Fig 2.1.2 è stata realizzata utilizzando MyMaps (Maps, 2021).

Per la manipolazione dei dati genetici sono stati utilizzati i seguenti software:

Vcftools (Danecek et al., 2011) scaricato da:

https://vcftools.github.io/man_latest.html

Bcftools (Li, 2011) scaricato da: <https://github.com/samtools/bcftools>

Plink (Chang et al., 2015) scaricato da: <https://www.cog-genomics.org/plink/1.9/>

2.2. Dati storici utilizzati

Ai dati genetici ottenuti dalle varie popolazioni sono stati accostati dati storici legati ai vari paesi considerati nello studio. Esiste una vasta quantità di informazioni storiche riguardanti la Peste Nera ma negli ultimi quaranta anni si sono verificati numerosi casi di uso acritico e di riscrittura imprecisa di queste informazioni che ha reso molto complesso conoscere l'affidabilità dei dati riguardanti la Peste Nera e mostrato una grave sovrastima dell'effetto della peste nelle grandi città e una sottostima degli effetti nelle aree rurali e nei centri abitati più piccoli (Roosen & Curtis, 2018). Pertanto non è stato tenuto conto delle stime totali per quanto riguarda l'Europa, dell'emergere dei vari focolai o del presunto numero di defunti in uno specifico periodo legati alla Peste Nera. Per limitare potenziali distorsioni interpretative, sono state utilizzate unicamente stime per paese del numero di abitanti (dunque dati non legati direttamente al fenomeno della peste). Anche per questo motivo ci si è affidati ai dati storici non utilizzandoli come punto di partenza, ma accostandoli ai dati genetici in un secondo momento. Un altro problema riguardante i record storici sta nel fatto che si tratta generalmente di dati piuttosto vecchi che non hanno subito revisioni nel corso degli ultimi anni. In alcuni casi inoltre questi dati risultano discordanti da una fonte all'altra, pertanto si è deciso di non creare un dataset "medio" ma di accostare i vari dati contemporaneamente. I dati maggiormente disponibili riguardano per ragioni geopolitiche i paesi europei, mentre in altri casi sono del tutto assenti, difficili da reperire o poco affidabili (come ad esempio per i paesi dell'Asia centrale).

Il dataset principale e più affidabile utilizzato verrà indicato come McEvedy per praticità (McEvedy & Jones, 1979). Anche i dati proposti da Angus Maddison e scaricati da:

<https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-database-2010> , sono stati utilizzati in tutti i casi in cui fosse possibile nonostante dispongano di poche stime per il periodo preso in esame.

Sono stati utilizzati inoltre i seguenti dati:

Per vari paesi d'Europa: Russell (Russell, 1972) e Uralanis (Uralanis, 1941)

Per l'Italia: Lo Cascio e Malanima (Lo Cascio & Malanima, 2005), Beloch e Bellettini (Bellettini, 1987), Breschi e Malanima (Malanima & Breschi, 2002).

Per la Gran Bretagna: dati di Our World in Data per l'Inghilterra (BankOfEngland, 2021) scaricati da:

<https://ourworldindata.org/grapher/population-of-england-millennium>

Per l'Estonia: Pankratov (Pankratov et al., 2020).

Per la Cina: Durand (Durand, 1960), Banister (Banister, 1992)

2.3. PCA

L'Analisi delle Componenti Principali (Principal Component Analysis, PCA) è uno dei metodi più vecchi e maggiormente utilizzati per ridurre drasticamente la dimensionalità di un dataset e per rendere le informazioni più semplici da interpretare, senza perdere informazione e mantenendo più diversità possibile. Questo si traduce nel creare nuove variabili che sono funzioni lineari di quelle nel dataset originale, in grado di massimizzarne la varianza senza essere correlate tra loro (Jolliffe & Cadima, 2016). Nel caso di questo studio, l'analisi delle componenti principali ci permette di osservare l'omogeneità interna a una singola popolazione per identificare la presenza di individui particolarmente distanti geneticamente dal gruppo di appartenenza (outliers) o di strutture interne alle singole popolazioni (una popolazione che in realtà è costituita da due sottopopolazioni simili), anziché realizzare una panoramica generale di tutte le popolazioni insieme (Novembre et al., 2008). Per svolgere l'analisi delle componenti principali è stata utilizzata la funzione "--pca" di PLINK 1.9 (Chang et al., 2015).

2.4. IBDseq

Ibdseq (B. L. Browning & Browning, 2013) è un software che serve per identificare segmenti di DNA identici per discendenza (IBD) in dati di DNA non necessariamente in fase. IBDseq confronta a coppie gli individui e stima la probabilità dei genotipi osservati in due possibili situazioni: un modello IBD e un modello non-IBD (quindi in ciascun caso chiedendosi se ogni SNP identico tra i due individui discenda da un antenato comune recente oppure no). Il rapporto tra le due stime offre il LOD-score per ciascuno SNP: se positivo ci troviamo in una situazione di identità per discendenza, se negativo no. Sequenze di SNPs con punteggio positivo vanno a formare segmenti IBD. Osservando due alleli (in due individui) alla ricerca di una possibile discendenza da un antenato comune è importante sapere anche se questa discendenza si estende da entrambe le parti del locus in esame. Più il segmento IBD risulta lungo e più è recente l'antenato che ha originato quel segmento

(B. L. Browning & Browning, 2013). Il software riceve in input un file in formato VCF (Danecek et al., 2011) e una lista di individui che rappresenti la popolazione per cui si vuole estrarre i segmenti IBD.

2.5. IBDNe

IBDNe (S. R. Browning et al., 2018) è un software che partendo da segmenti lunghi di IBD è in grado di stimare l'Ne della popolazione in esame per periodi recenti: da 4 a oltre 200 generazioni per dati di sequenziamento e da 4 a 50 generazioni per SNP array.

Fornendo al software una lista di segmenti IBD (ottenuti con IBDseq), questi vengono filtrati in vari modi (per lunghezza, per vicinanza sul cromosoma, per posizione nel cromosoma) e viene restituita una stima di Ne e un intervallo di confidenza del 95% per ciascuna generazione da 0 al limite inserito dall'utente (S. R. Browning & Browning, 2015; S. R. Browning et al., 2018).

I parametri utilizzati per l'analisi sono i seguenti:

mincm=2 (la soglia minima di dimensione dei segmenti IBD in cM)

gmin= 2 (la generazione più recente che viene stimata efficacemente dal software)

gmax=300 (la generazione più antica stimata dal software 600/mincm)

nboots=80 (numero di repliche di bootstrap)

2.5.1. Consistenza tra cromosomi

Sapendo che ciascun cromosoma ha una storia a sé a causa della casualità degli eventi di ricombinazione, può avere senso controllare che ci sia consistenza nelle stime di IBDNe per i singoli cromosomi e che la stima sul totale (WG) sia riconducibile a un valore intermedio tra tutti gli autosomi. A questo scopo è stata scelta una popolazione qualsiasi (CHB in questo caso) ed è stata eseguita l'analisi di IBDNe sui singoli cromosomi, con la consapevolezza che non viene svolto il bootstrap avendo un singolo cromosoma per volta.

2.5.2. Pacchetto di R per identificare picchi e valli

Per identificare rapidamente e in maniera discreta la presenza di picchi e valli nelle curve stimate da IBDNe sono state usate le funzioni "Findpeaks" e "Findvalleys" del pacchetto "quantmod" di R <https://CRAN.R-project.org/package=quantmod>.

In caso di picchi o valli che corrispondono a più generazioni con lo stesso livello di Ne, è stata scelta come coordinata temporale di tale evento la media fra le generazioni o la più recente delle due generazioni medie possibili. È stata inoltre calcolata la riduzione di popolazione dovuta a un evento demografico, come il rapporto fra Ne del picco e Ne della valle successiva.

2.6. Simulazioni

Essendo noto dai dati storici a disposizione che il crollo demografico europeo sia stato repentino durante il periodo della Peste Nera, può essere utile comprendere quanto siano identificabili dai software a nostra disposizione, rapide variazioni nella dimensione di popolazione. A questo scopo sono state realizzate due simulazioni di andamento per 3 popolazioni (A,B,C) utilizzando il software msprime 1.0.2 (Kelleher et al., 2016). Le simulazioni sono state realizzate variando la dimensione di popolazione di partenza, l'entità del crollo demografico e la dimensione di popolazione dopo la ripresa. Sono state realizzate due simulazioni differenti per ciascuna popolazione, in un primo caso considerando il calo demografico come un evento istantaneo con ripresa immediata. Nel secondo caso la simulazione prevede un periodo di 10 generazioni tra la fine del crollo e l'inizio della ripresa.

3. Risultati

3.1. Rimozione di outliers e identificazione di sottogruppi

Per ciascuna popolazione è stata svolta un'analisi delle componenti principali per identificare eventuali strutture interne e la presenza di outliers. Come illustrato in *Tab 3.1-a* la maggior parte delle popolazioni presentava outliers che sono stati rimossi per lo svolgimento delle analisi successive. Inoltre, alcune popolazioni presentavano una struttura interna (*Tab 3.1-b*); in particolare sono presenti dei sottogruppi (almeno 30 individui) in GIH ed Etiopia e una divisione in due gruppi distinti in KHV.

pop	n° outliers	individui trattenuti	pop	n° outliers	individui trattenuti	pop	n° outliers	individui trattenuti
ACB	4	94	IBS	2	108	ARMENIA	13	92
ASW	5	56	ITU	27	72	AZERBAIJAN	0	67
BEB	4	84	JPT	0	104	CRIMEA	11	76
CDX	2	91	KHV	0	99	GEORGIA	4	124
CEU	13	95	LWK	5	94	KAZAKISTAN	3	55
CHB	1	103	MSL	6	81	KIRGHIZISTAN	16	168
CHS	0	105	MXL	0	64	MONGOLIA	0	40
CLM	0	94	PEL	6	81	TAJKISTAN	20	82
ESN	4	95	PJL	4	92	UZBEKISTAN	37	67
FIN	1	99	PUR	3	101	ESTONIA	15	2405
GBR	4	87	STU	7	95	EGITTO	5	103
GIH	4	95	TSI	2	105	ETIOPIA	2	102
GWD	3	111	YRI	12	103	TOT	245	5789

Tab 3.1-a: Elenco delle popolazioni con rispettivi outliers rimossi e individui trattenuti per l'analisi.

pop	n individui
GIH subset	66
KHV_1	69
KHV_2	30
Etiopia subset	79

Tab 3.1-b: Sottogruppi creati a partire dall'analisi delle componenti principali.

3.2. Confronto WGS e SNP array

Il dataset principale utilizzato nello studio è costituito dai genomi completi delle popolazioni di *1000 Genomes Project* (81647806 snps), tuttavia le altre popolazioni prese in esame sono disponibili solo in forma di SNP array, non di genoma completo. 579384 è il numero di SNPs disponibili per le popolazioni di Silk Road che sono state scelte come fattore limitante e a questi sono stati adattati gli SNP array di Etiopia, Egitto ed Estonia. Per verificare che l'utilizzo di un numero inferiore di SNP non inficiasse il risultato di IBDNe è stato creato un subset degli SNP di TSI.

In *Fig 3.2* è sono mostrati i risultati di IBDNe sul genoma di TSI e sul sottocampione di 579384 SNPs. Confrontando i profili ottenuti con TSI-wgs e TSI-snpchip possiamo notare che sono molto simili, specialmente nel periodo preso in esame (1700-1000). I dati SNP array mostrano una stima di N_e più bassa man mano che si procede a ritroso nel tempo se confrontata con il genoma completo, tuttavia per lo scopo di questo studio, la somiglianza dei due profili è sufficiente per rendere possibile l'utilizzo di SNP array con 579384 SNPs e dunque i dataset non disponibili come genomi completi. Nonostante i risultati non siano identici, può essere utile accettare di avere un dato leggermente più impreciso rispetto a quelli basati su WGS per poter includere più popolazioni nello studio. Questo confronto è confermato inoltre dalla documentazione legata al software utilizzato, secondo cui la determinazione di N_e funziona in modo efficiente fino a 50 generazioni sui dati SNP array e oltre 200 per i dati da sequenziamento (S. R. Browning & Browning, 2015). Il periodo preso in considerazione va da 8 a 32 generazioni fa circa, considerando una durata di 30 anni per ciascuna generazione e fissando la data di campionamento al 1950. Si apre quindi la possibilità di utilizzare senza particolari accorgimenti anche i dati SNP array.

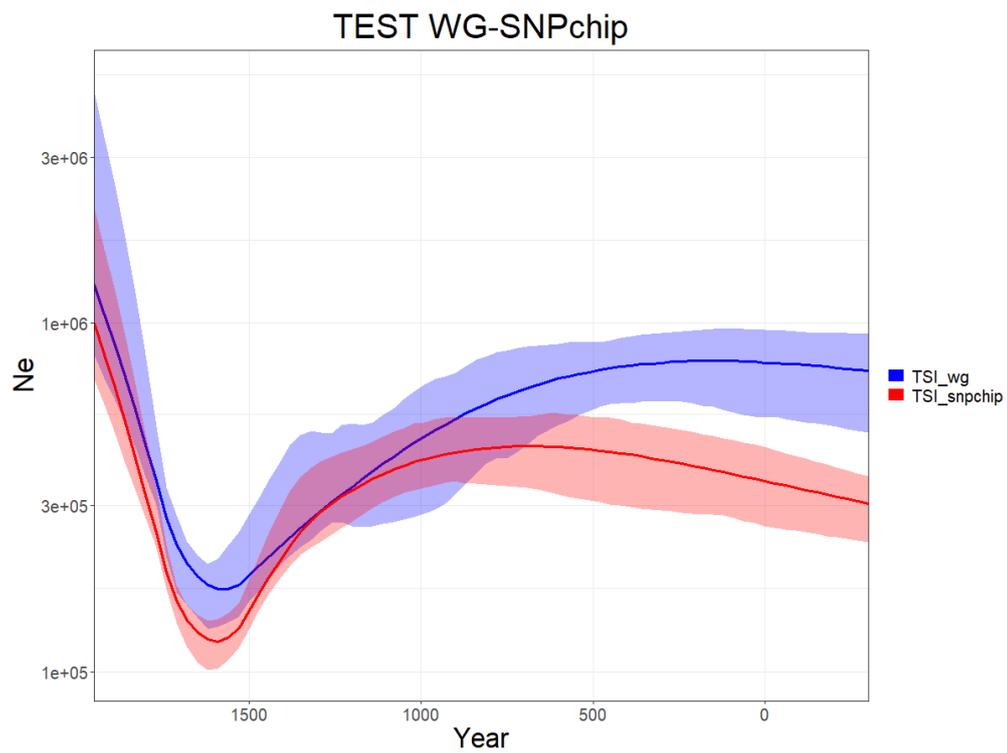


Fig 3.2: Profili di Ne di TSI con genoma completo (TSI_wg) e sotto-campione corrispondente ai 579384 SNPs di Silk Road ($TSI_snpchip$).

3.3. Validazione sui vari cromosomi

Il risultato dell'analisi dei singoli cromosomi mostra, come atteso, che la curva stimata per il genoma completo si pone come intermedia tra tutti gli autosomi. I vari cromosomi mostrano andamenti simili ma con magnitudini molto variabili al momento della generazione 0.

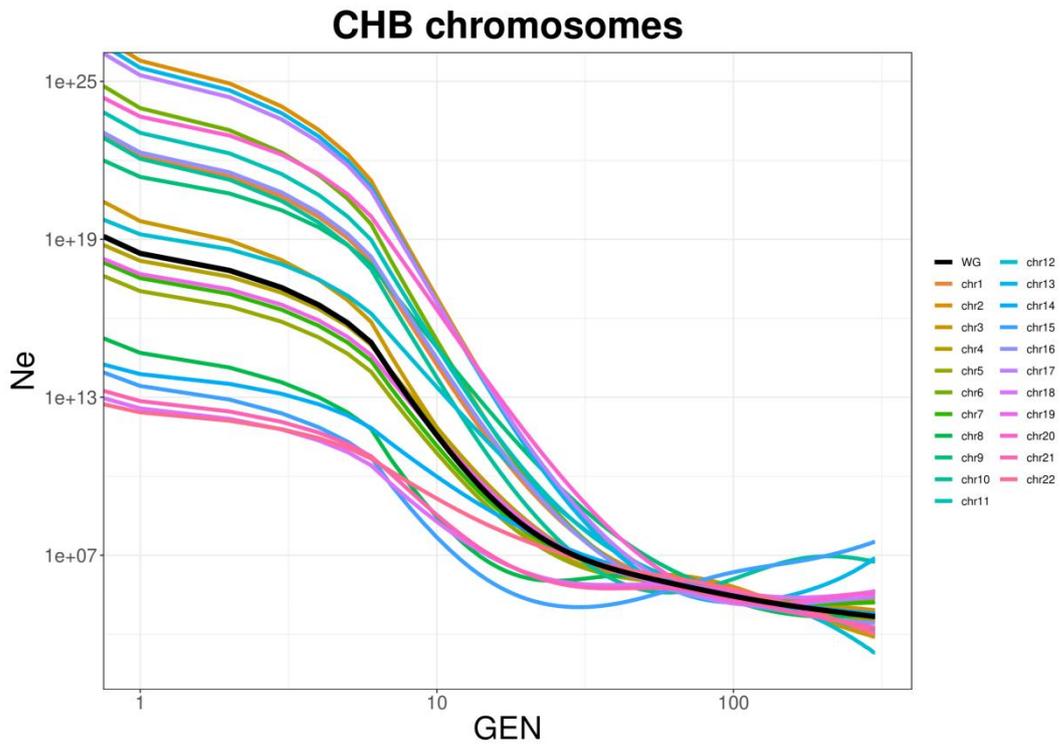


Fig 3.3 Rappresentazione delle dimensioni di popolazione stimate con IBDNe per CHB sul totale del Genoma (WG, in nero) e sui singoli autosomi.

3.4. Picchi e Valli

Raccogliendo in modo schematico i picchi e le valli (Fig 3.4) degli andamenti delle varie popolazioni (dove disponibili, nelle ultime 70 generazioni), si nota che tutte le popolazioni che presentano un calo seguito da un recupero, mostrano la ripresa in un momento antecedente alla rivoluzione industriale. Non essendoci altri fenomeni ben chiari a una prima occhiata, vale la pena osservare le popolazioni una per una e provare a dare una spiegazione dell'andamento che vediamo.



Fig 3.4 Rappresentazione schematica della presenza di picchi (1 Blu) e valli (-1 Rosso) in corrispondenza di anno e generazione. La banda grigia si riferisce al periodo intorno alla Peste Nera. La banda rossa si riferisce alla rivoluzione industriale.

3.5. Confronto fra profilo simulato e profilo inferito da IBDNe

Dai risultati ottenuti dalle simulazioni (Fig 3.5) si può notare che la qualità della stima del picco di crescita prima del crollo demografico è fortemente legata alle dinamiche demografiche: sono infatti determinanti N_e prima e dopo il crollo, l'intensità del crollo stesso, l'intensità del recupero e la durata del periodo di tempo tra la fine del crollo demografico e l'inizio della ripresa. Riguardo questo ultimo aspetto in particolare, la Simulazione 2 (in cui è stato scelto un periodo di ripresa di 10 generazioni) mostra una stima molto più accurata se confrontata con la Simulazione 1 a ripresa istantanea. Per quanto le valli siano generalmente ben stimate, non si può dire lo stesso per quanto riguarda i picchi: è importante notare che la presenza di un crollo istantaneo (come quello rappresentato dalla Peste Nera) risulta difficile da stimare e le curve di IBDNe presentano dunque un picco arretrato rispetto al picco reale. Osservando i risultati ottenuti si potrebbe speculare che la valle reale si trovi in corrispondenza di quella inferita da IBDNe oppure leggermente arretrata e che il picco reale invece possa essere più recente di quanto stimato.

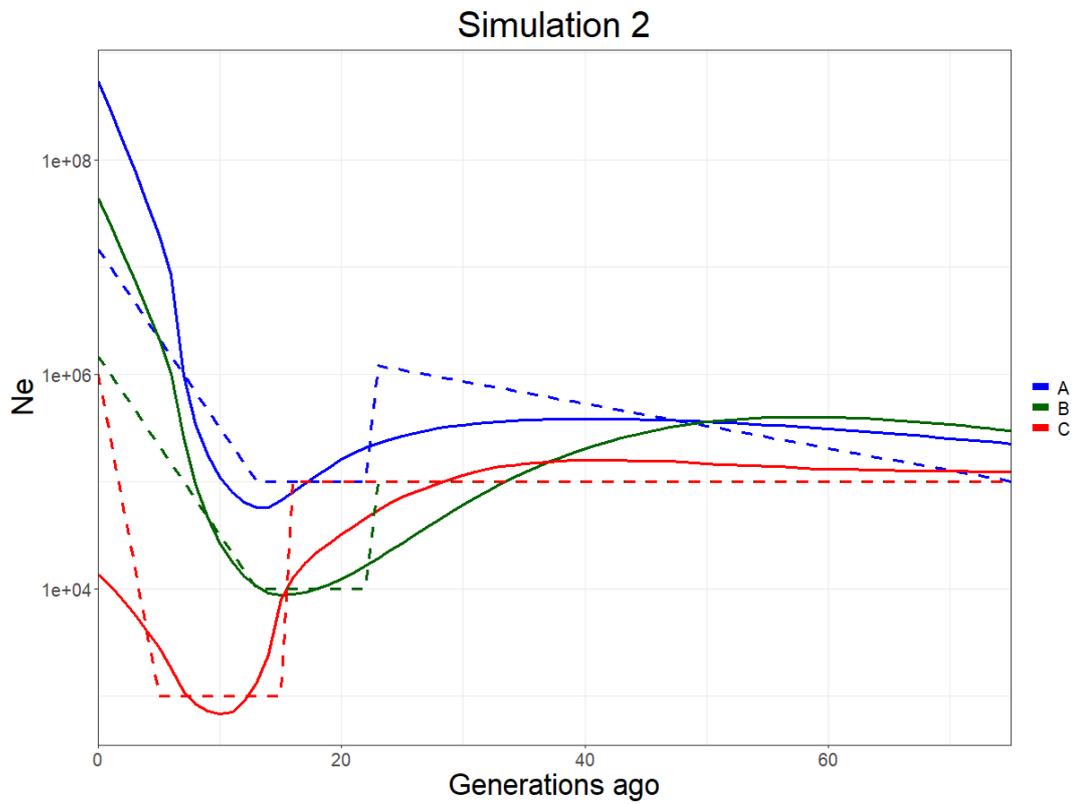
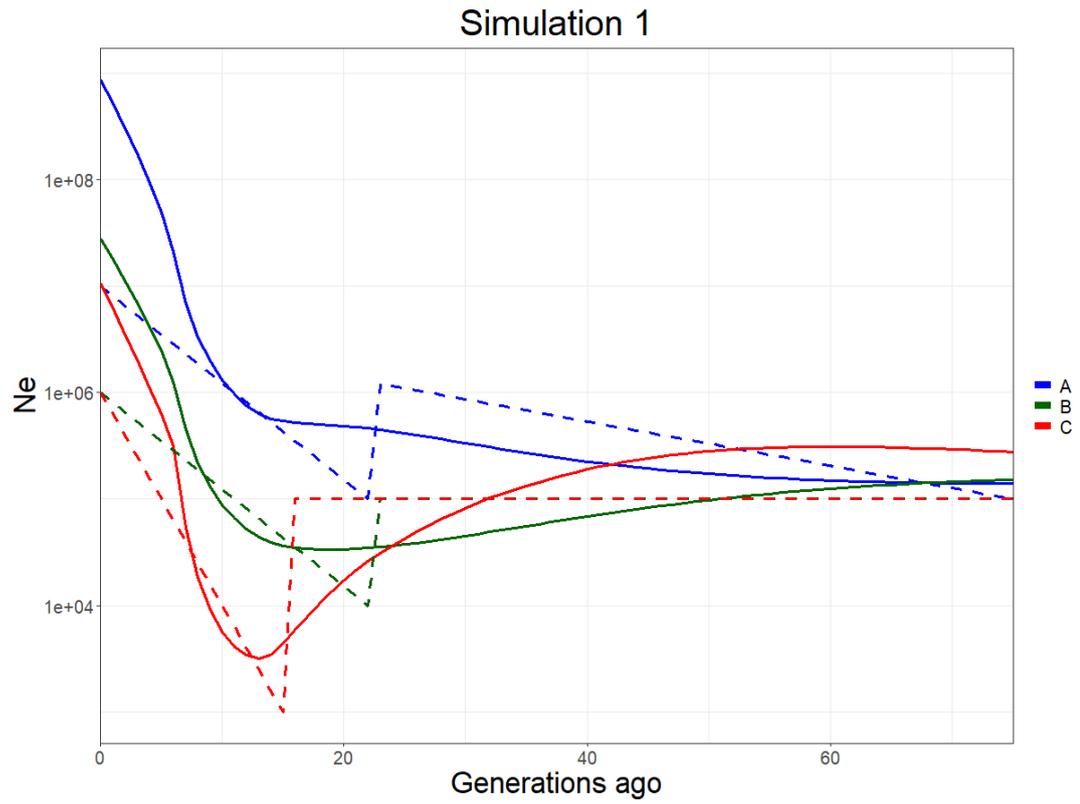
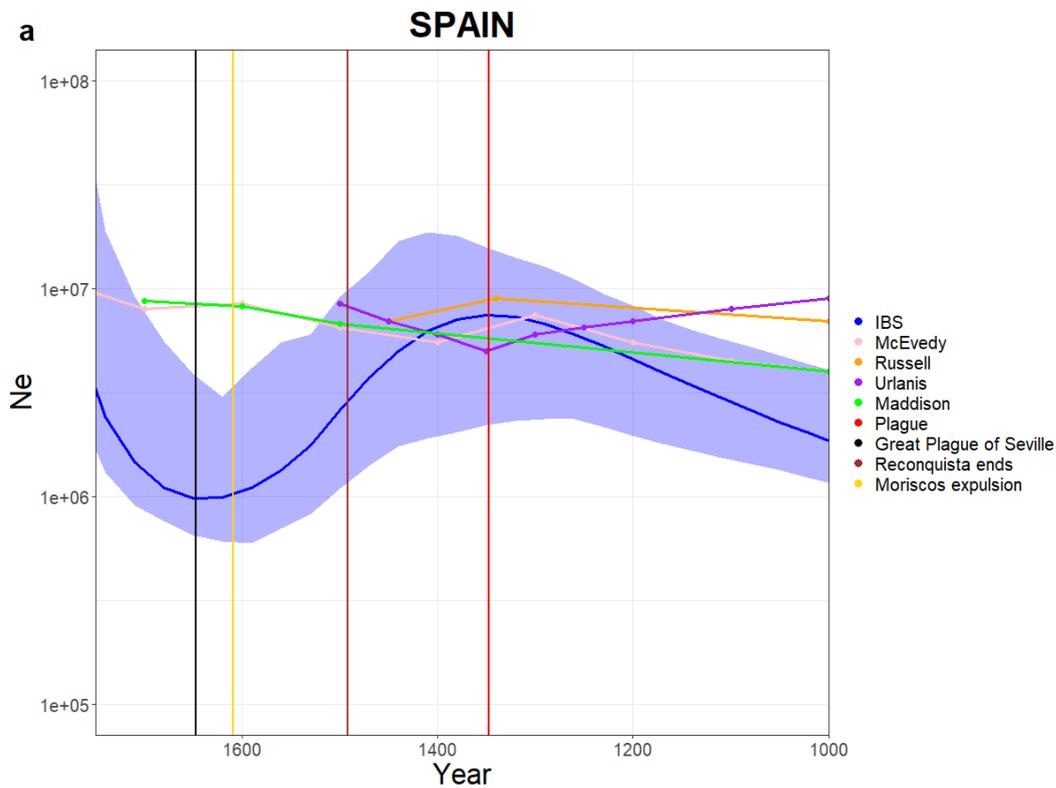
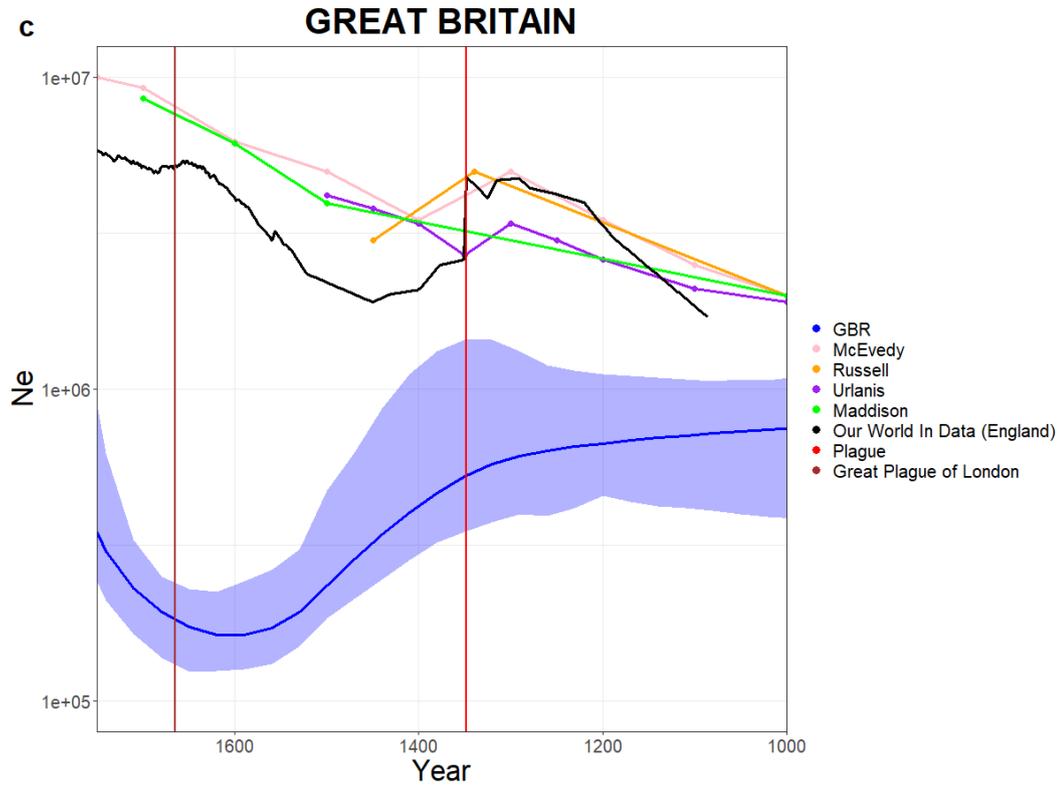
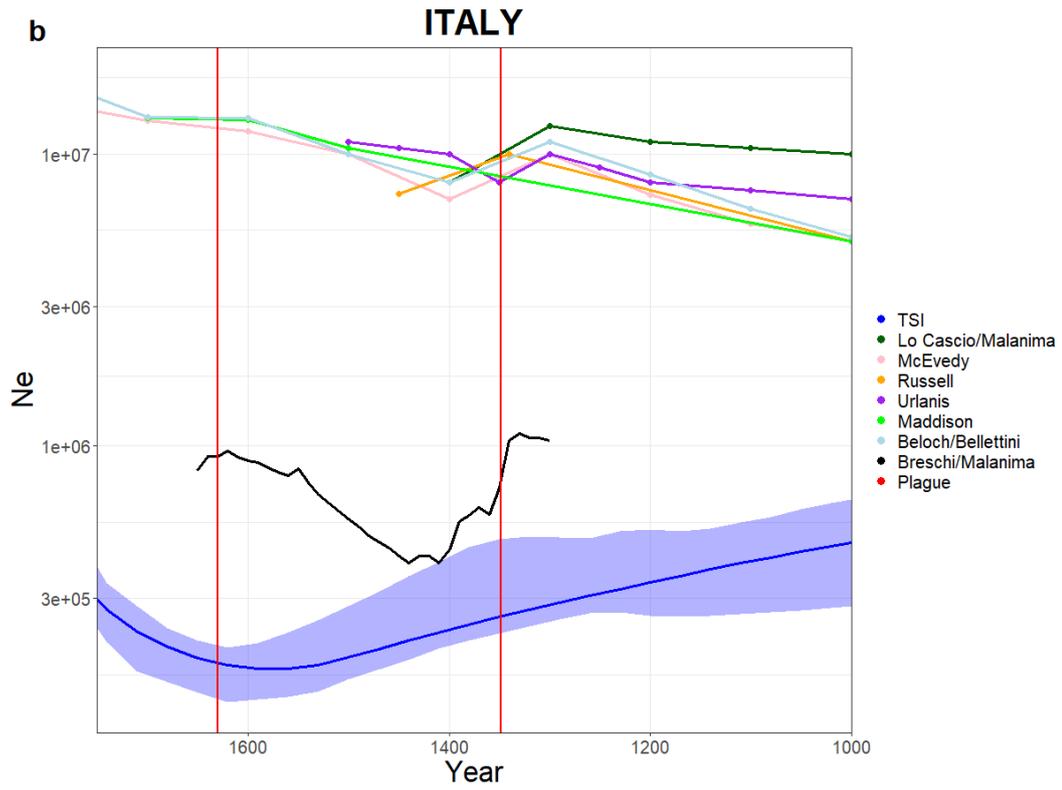


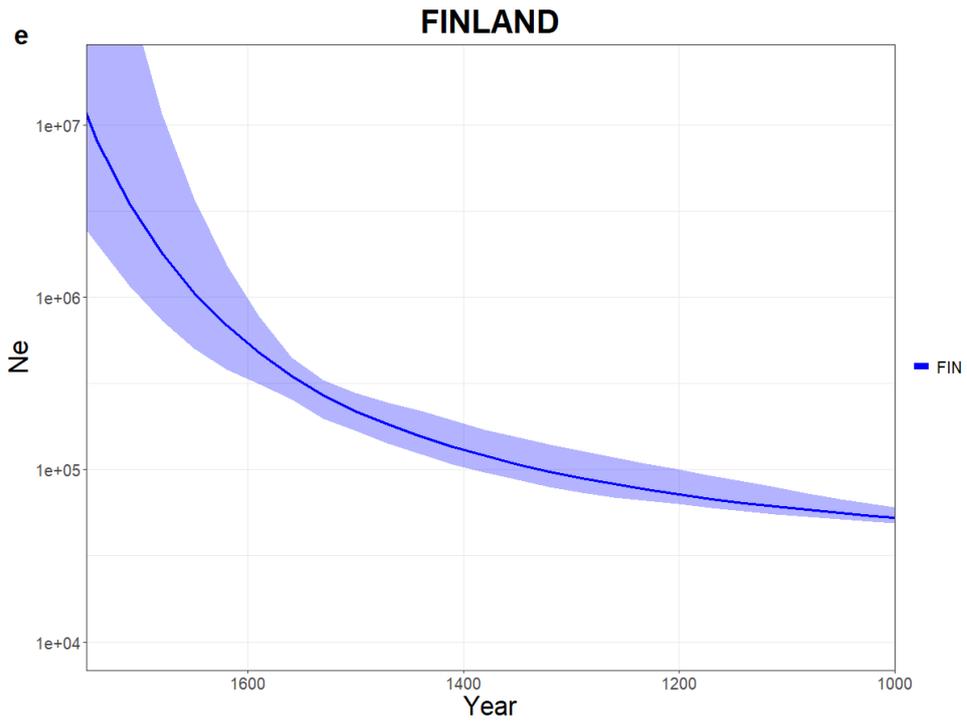
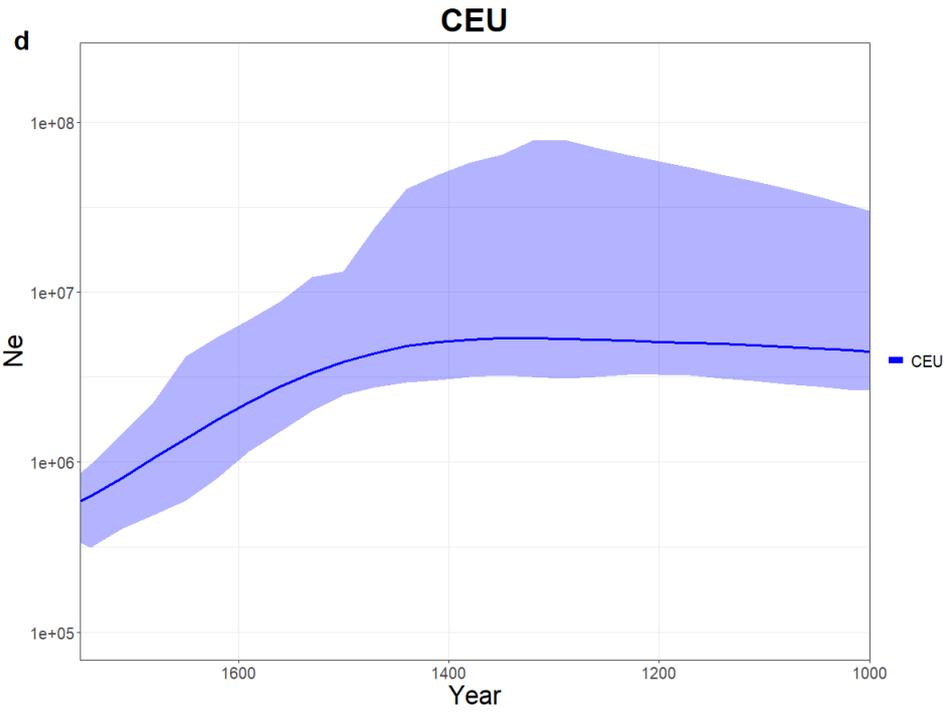
Fig 3.5: Le simulazioni demografiche prodotte per 3 popolazioni A,B e C sono riportate con le linee tratteggiate e i rispettivi N_e calcolati tramite l'uso di IBDNe sono riportati come linee continue

3.6. Europa

Osservando le curve ottenute è possibile apprezzare una grande varietà di andamenti tra le N_e stimate da IBDNe (Fig 3.6). Italia, Spagna e Gran Bretagna (Fig 3.6-a-b-c) mostrano tutte una valle intorno al 1600 ma andamenti di declino diversi tra loro. I CEU (Fig 3.6-d), non rispecchiando un paese preciso, sono stati riportati in quanto popolazione Europea e mostrano un declino costante dal 1350 in poi. La Finlandia (Fig 3.6-e) mostra una crescita continua che inizia a impennarsi intorno al 1600. In tutti i casi in cui sono disponibili dati storici per il periodo preso in esame è chiaro che la stima storica e quella genetica non coincidano in magnitudine (ad eccezione dell'Estonia (Fig 3.6-f) per cui si dispone di un pannello di individui molto ampio e rappresentativo dell'intero paese). Ciascun caso necessiterà di essere analizzato singolarmente per provare a spiegare i risultati ottenuti.







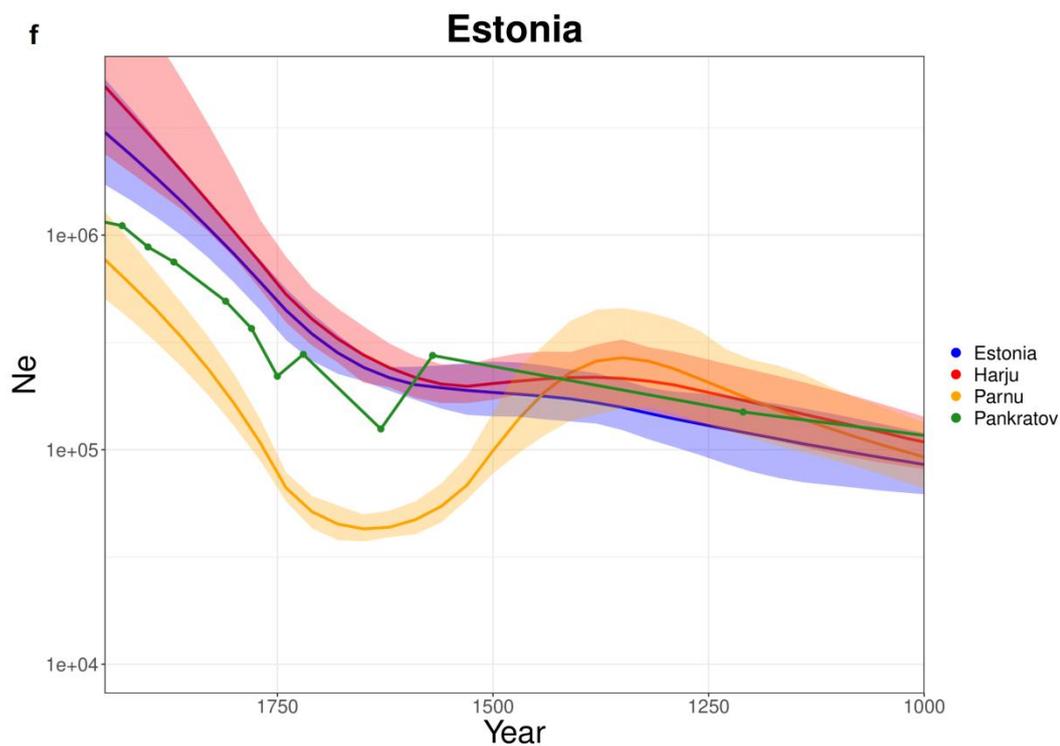
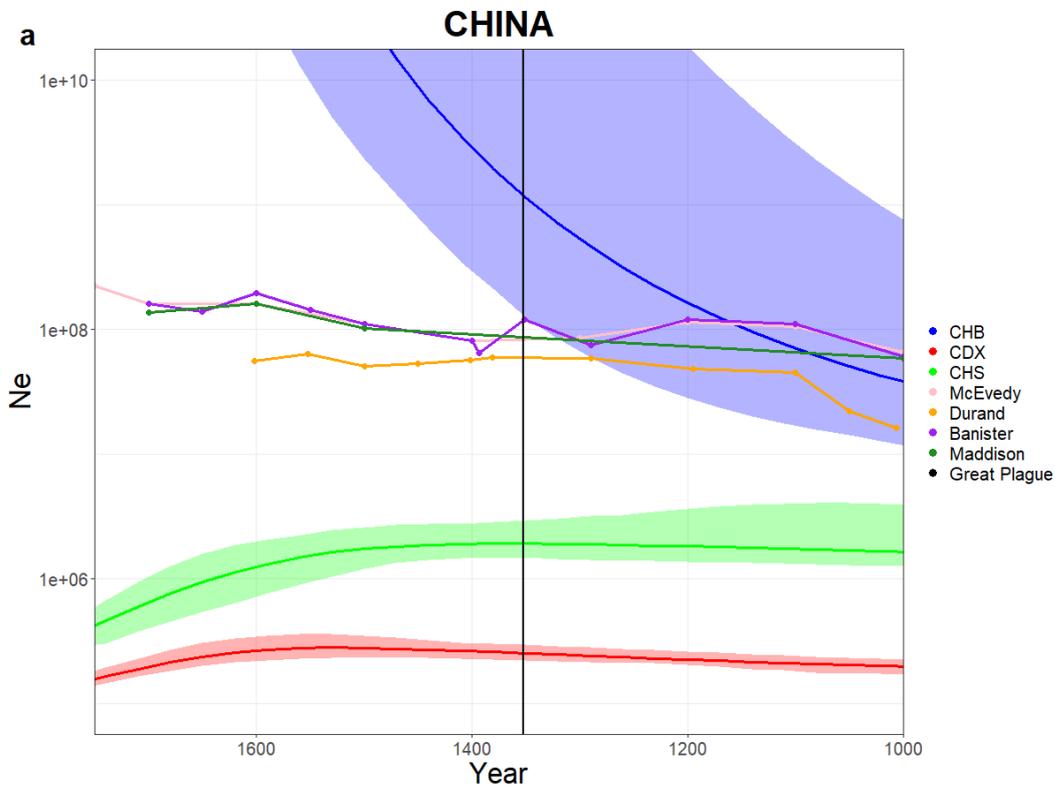


Fig 3.6: Curve stimate con IBDNe per le varie popolazioni europee. Dove possibile sono state riportate alcune stime storiche ed eventi noti legati al paese di riferimento. In ciascun caso N_e è stato moltiplicato per 4.89 per avvicinarsi a N_c . Le curve stimate sono accostate a una banda di errore. (a) Spagna, (b) Italia, (c) Gran Bretagna, (d) CEU. (e) Finlandia, (f) Estonia avendo due sottogruppi aggiuntivi (Harju e Parnu), presenta tre curve stimate da IBDNe.

3.7. Asia Orientale

Osservando l'andamento delle popolazioni cinesi (Fig 3.7-a) non stupisce il comportamento di CHB, che mostra una crescita fuori scala se confrontata con tutte le altre popolazioni del mondo. Le popolazioni CHS e CDX invece sembrano mostrare un calo intorno al 1600. Il Giappone mostra un andamento particolare e una dimensione di popolazione di N_e che supera in alcuni momenti di 10 volte la stima storica. Per quanto riguarda il Vietnam, la stima di IBDNe sul totale della popolazione KHV è leggermente inferiore (come ci si potrebbe aspettare) alla stima storica per la totalità dell'Indocina. I due subset di KHV mostrano curve differenti dal 1500 in poi, che vengono riassunte da una curva totale (Blu) intermedia tra le due.



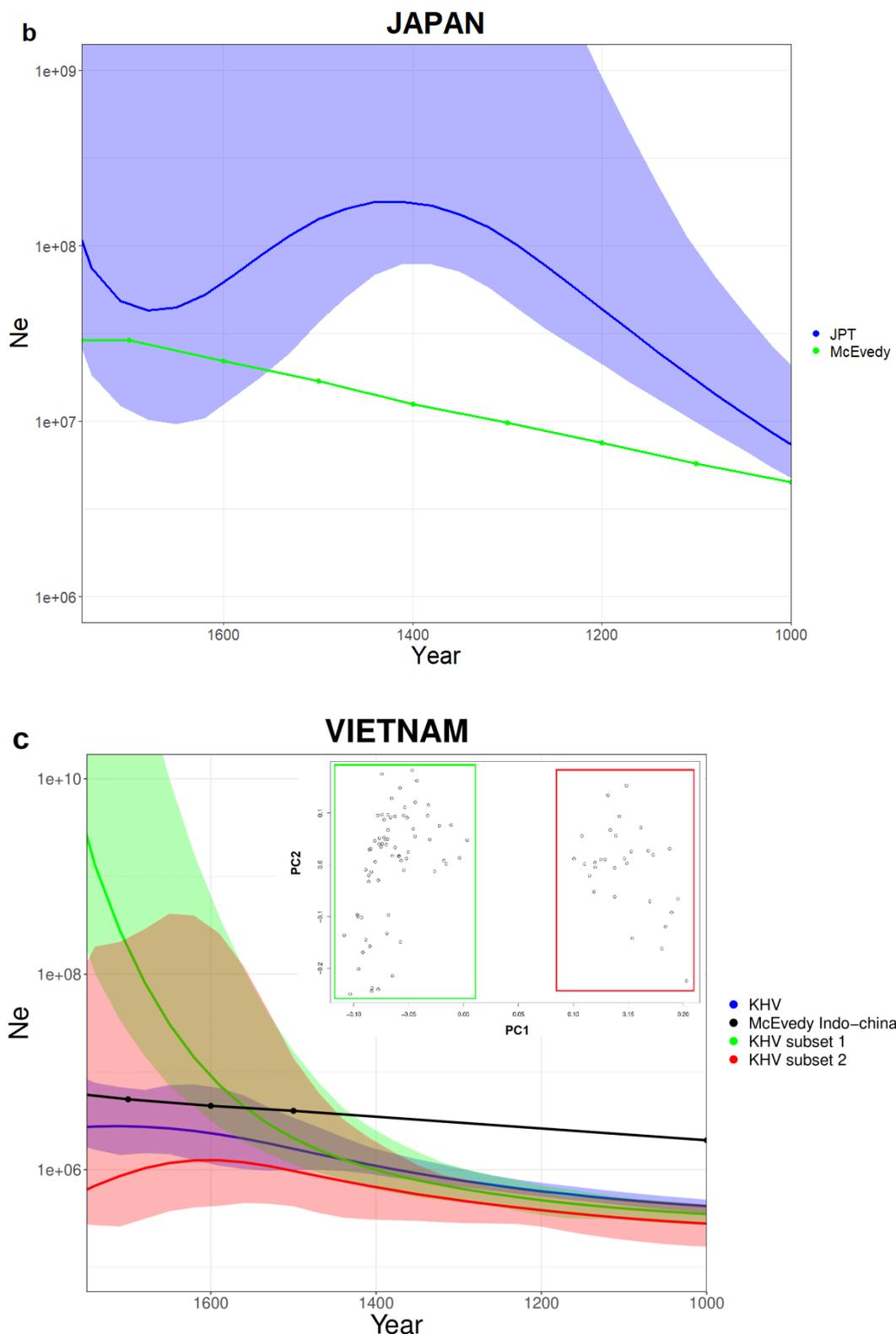


Fig 3.7: Curve di Ne ($\times 4.89$) stimate da IBDNe associate ai dati storici disponibili per i relativi paesi. (a) Cina, rappresentata in “1000 Genomes Project” da tre popolazioni diverse CHB, CHS, CDX con stime storiche e Grande Peste del 1352. (b) Giappone. (c) Vietnam con stima storica dell’Indocina, PCA e subset creati sulla base di quest’ultima.

3.8. Subcontinente indiano

Il subcontinente indiano è abitato da molte popolazioni diverse culturalmente e geneticamente, confermate dalla grande varietà di andamenti che riusciamo a osservare tramite IBDNe (Fig 3.8-a). Tutte le curve dei singoli gruppi si trovano al di sotto della stima storica della popolazione del subcontinente nel suo complesso.

Considerando la peculiare distribuzione degli individui di GIH nella PCA (Fig 3.8-b) poteva essere rilevante controllare che i due gruppi osservati non fossero due popolazioni separate. La presenza di un solo gruppo sufficientemente grande da poter essere analizzato però consente di realizzare un solo subset. Le curve del totale di GIH e del sottogruppo sono quasi identiche, quindi non riscontriamo la presenza di reali sottopopolazioni.

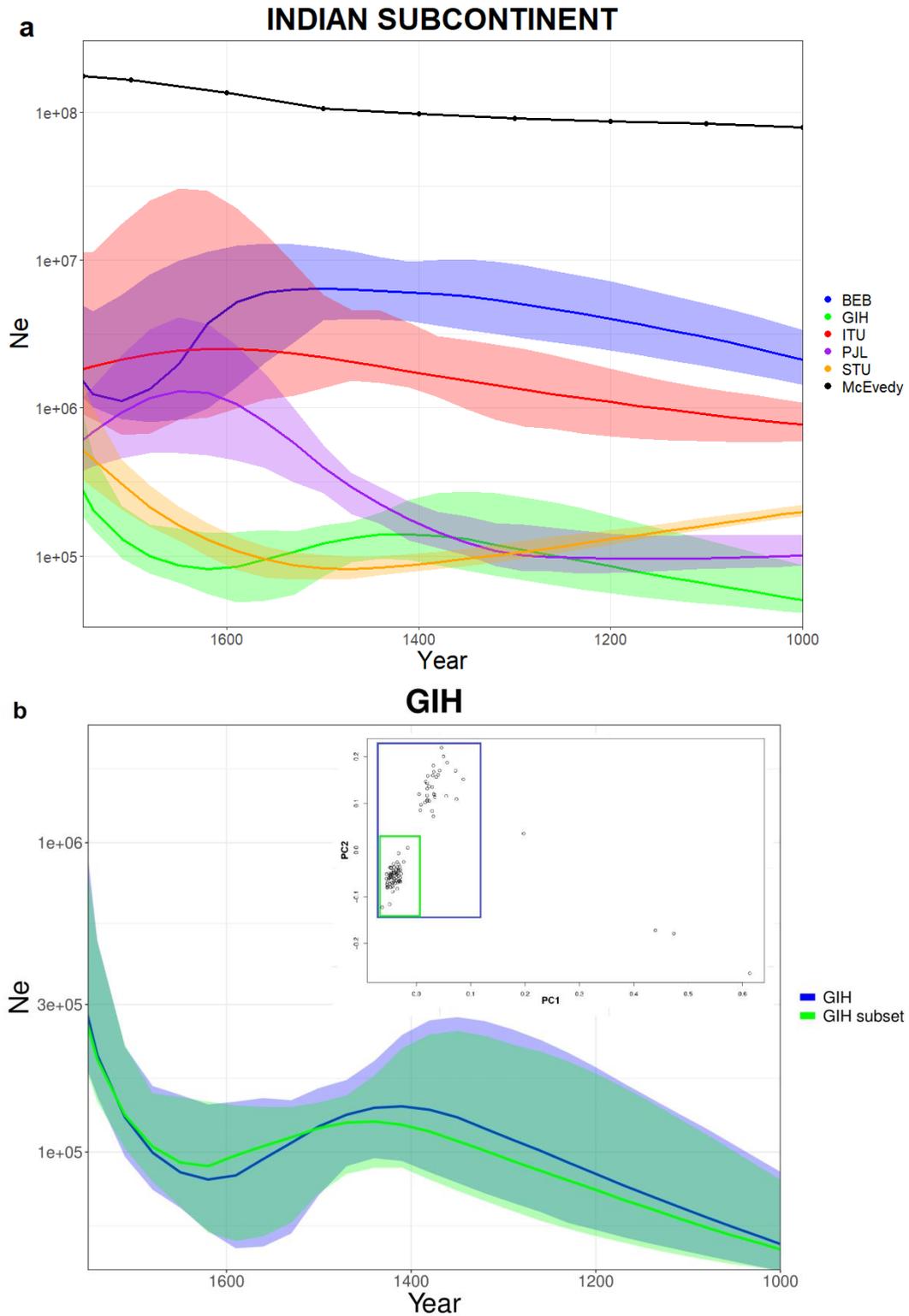
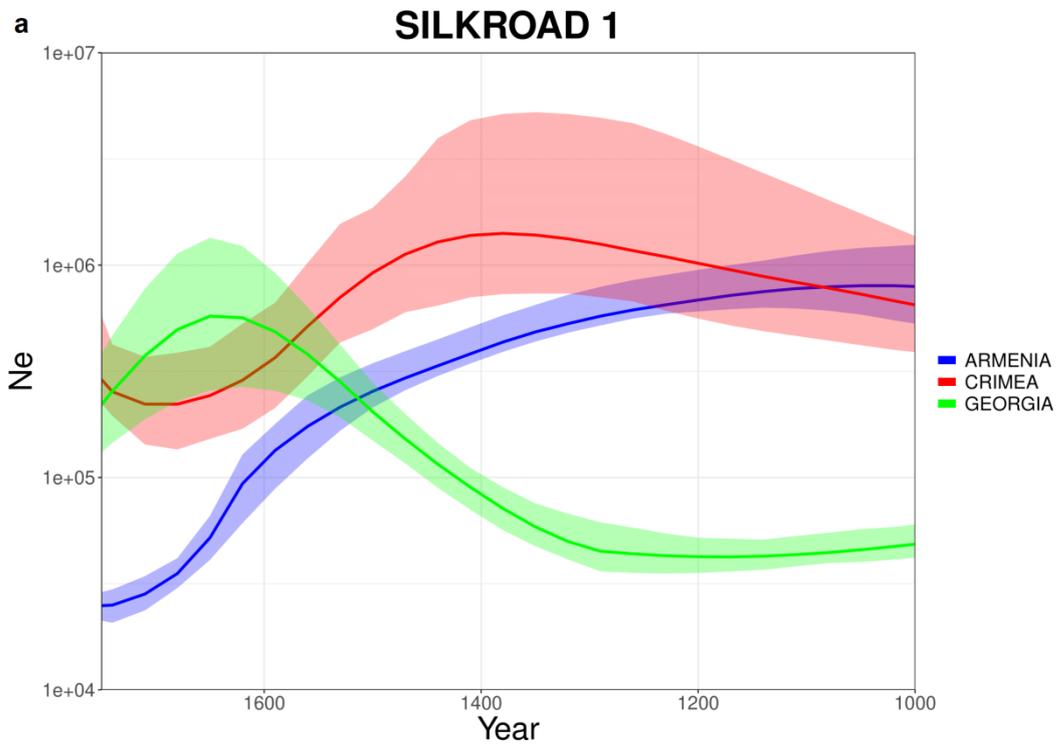


Fig 3.8 (a) Curve stimate da IBDNe ($\times 4.89$) per il subcontinente indiano con stima storica totale. (b) GIH e subset di GIH costruito sulla base di PCA.

3.9. Via della Seta

Osservando le curve prodotte per SILK ROAD (Fig 3.9.1) nonostante gli andamenti siano vari, è possibile notare come alcune popolazioni inizino a subire un calo demografico intorno al 1500. Tre popolazioni (Fig 3.9.1-c) presentano un andamento particolare con un declino continuo partendo da dimensioni di popolazione estremamente alte se confrontate con le altre popolazioni. Questo andamento così particolare e poco realistico potrebbe essere collegato alla distribuzione anomala degli individui delle tre popolazioni, osservabili nelle PCA (Fig 3.9.2).



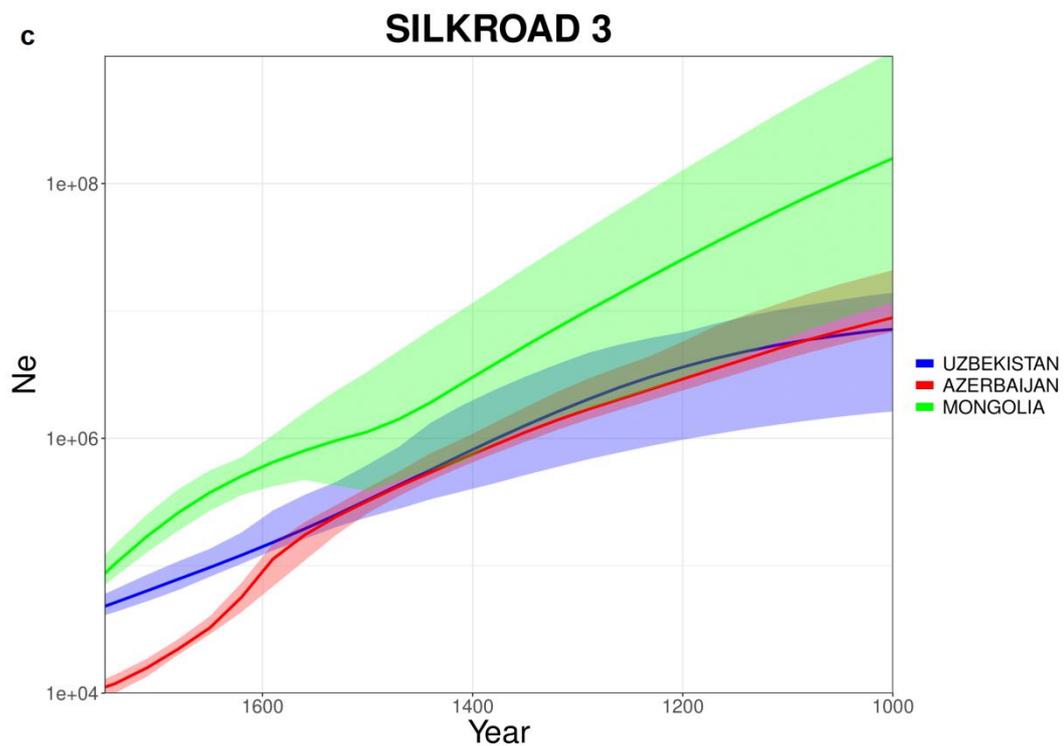
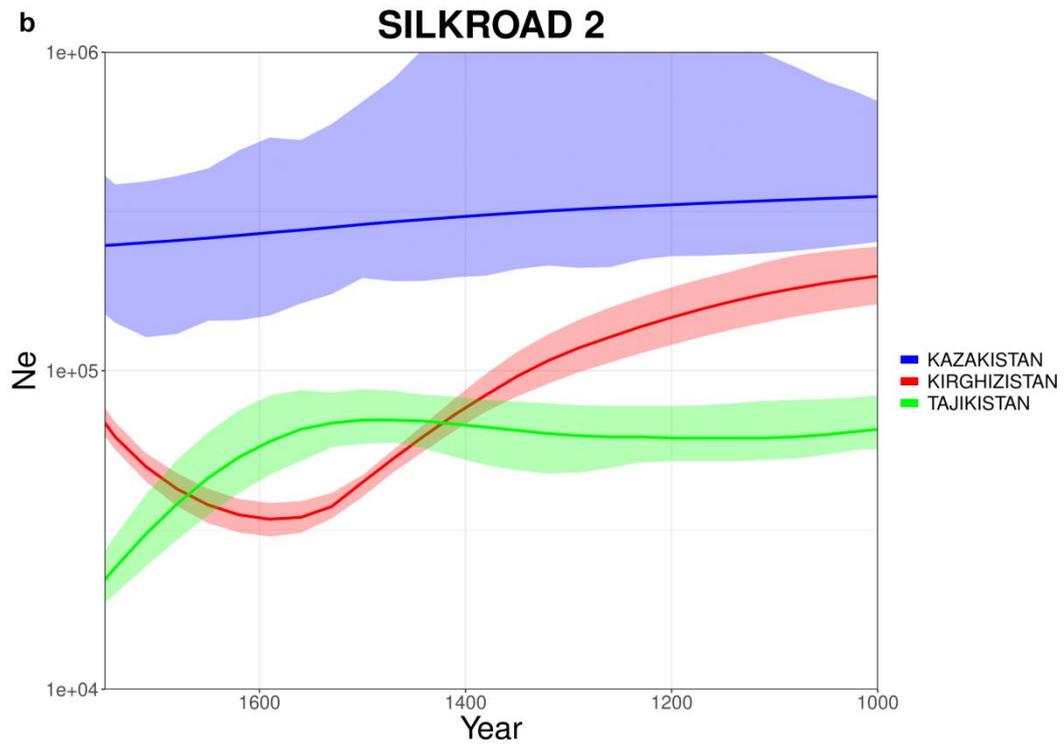
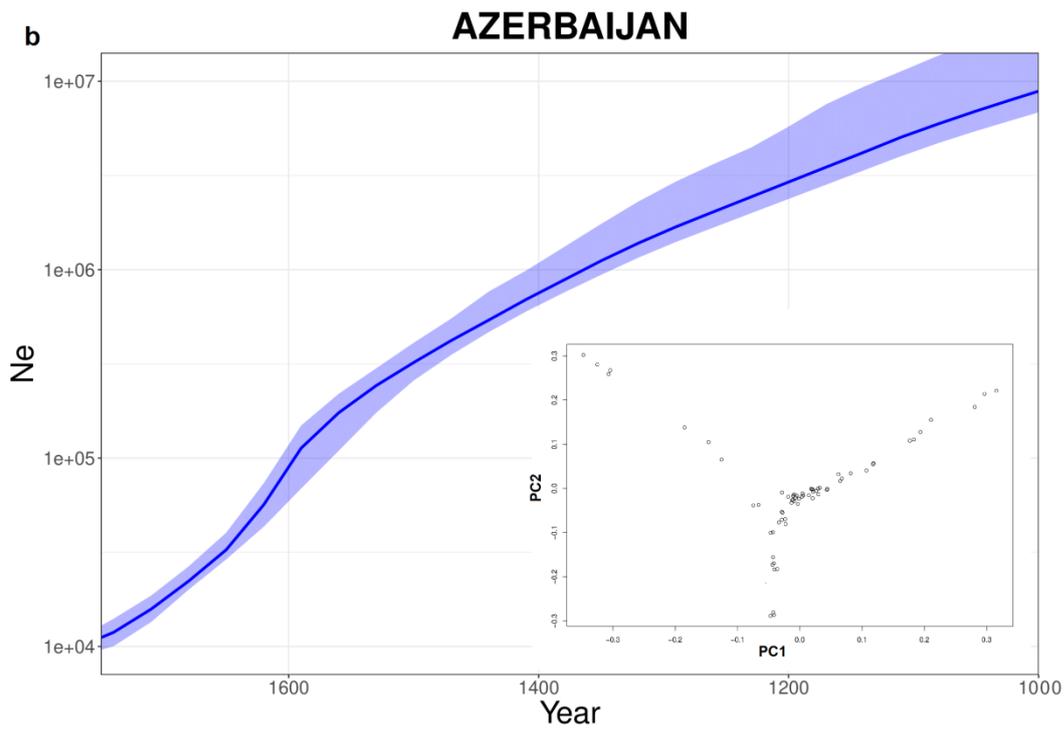
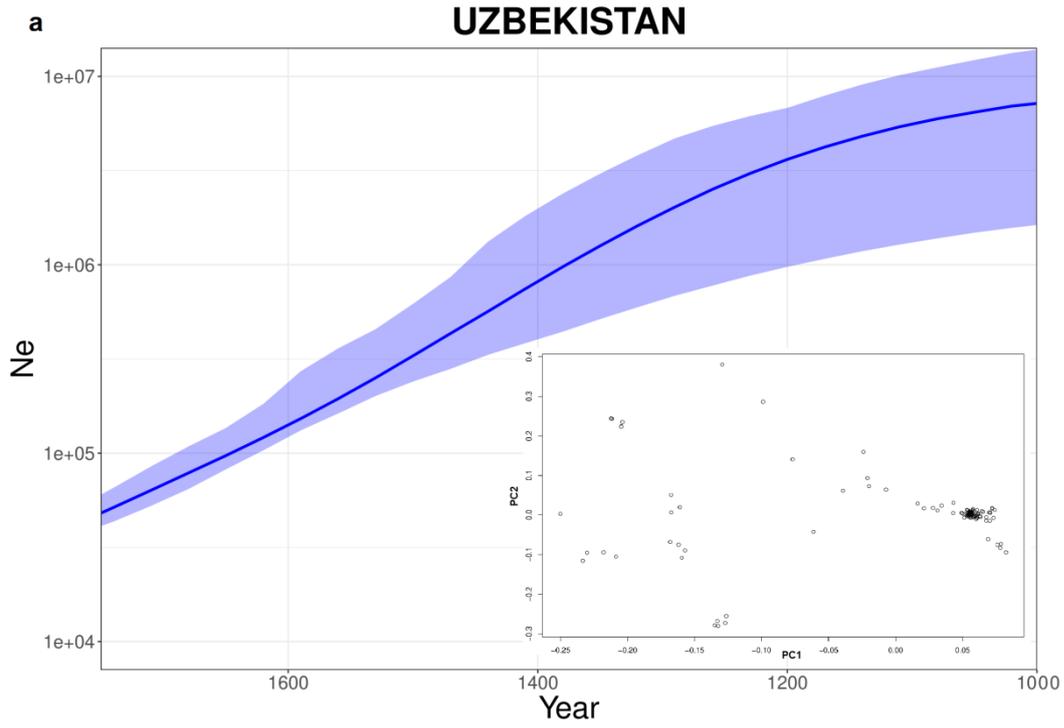


Fig 3.9.1: Curve stimate da IBDNe ($\times 4.89$) per il dataset SILK ROAD. Le curve sono state separate in tre gruppi (a) Armenia, Crimea, Georgia (b) Kazakistan, Kirghizistan, Tajikistan (c) Uzbekistan, Azerbaijan, Mongolia .



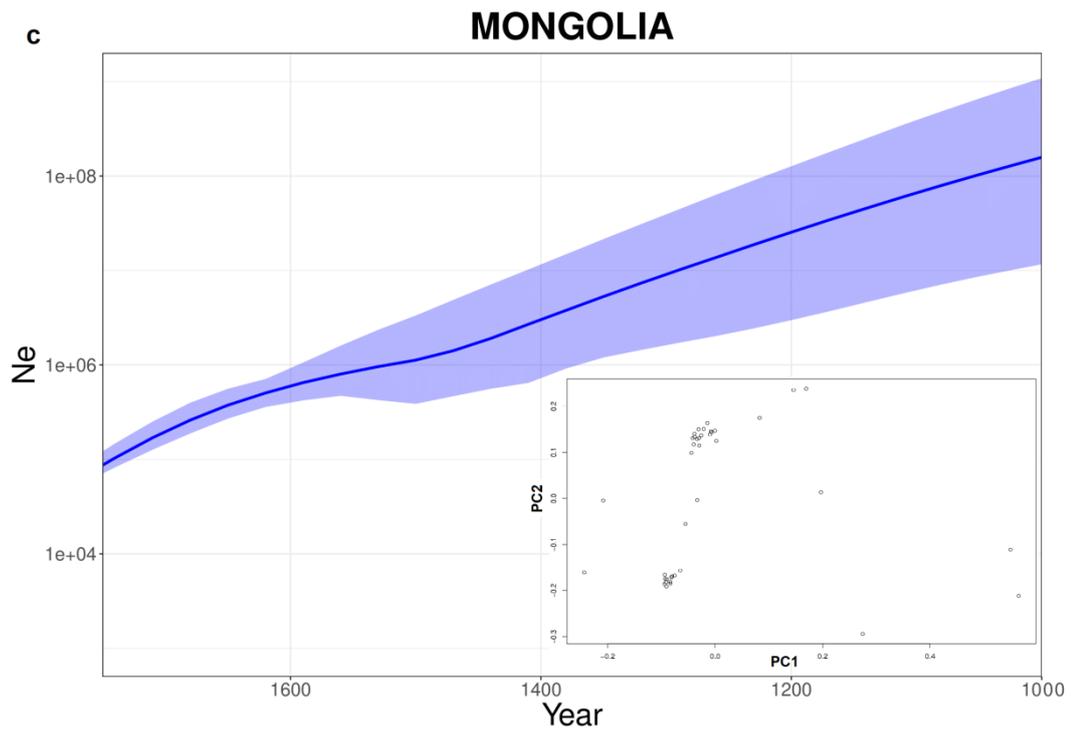
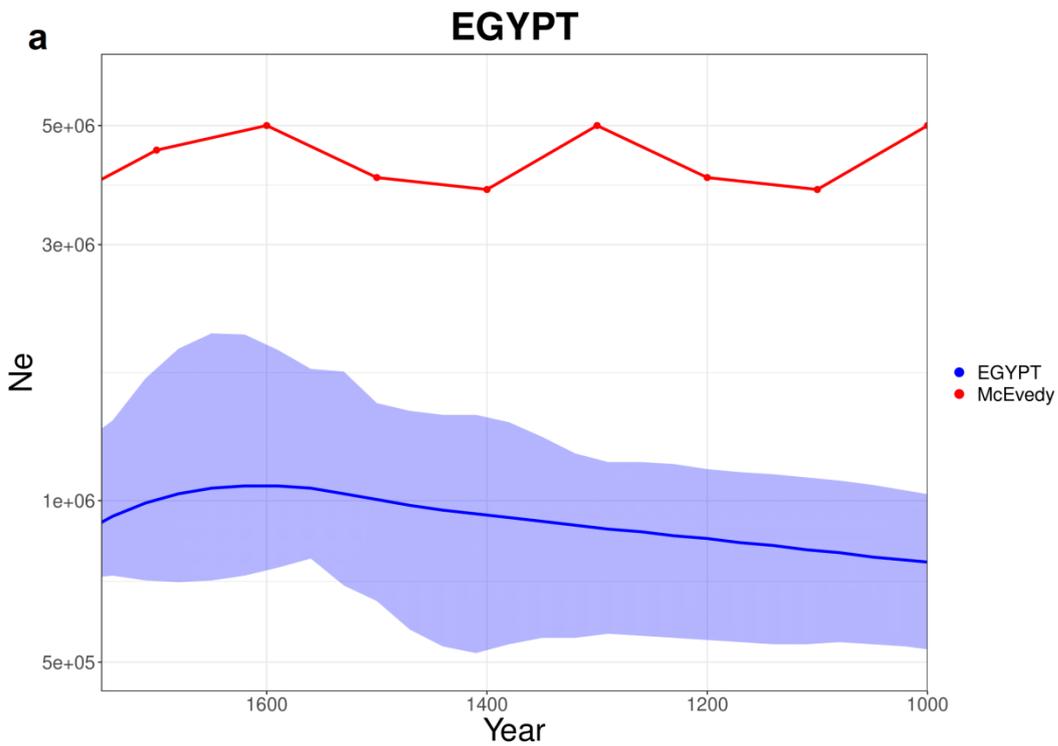


Fig 3.9.2: Curve stimate da IBDNe ($\times 4.89$) e PCA per (a) Uzbekistan, (b) Azerbaijan e (c) Mongolia.

3.10. Etiopia ed Egitto

Per quanto riguarda l'Egitto la curva stimata da IBDNe (Fig 3.10-a) mostra un andamento molto lineare con un calo intorno al 1600 e una stima di N_e inferiore alle stime storiche della popolazione. Le curve di Etiopia (Fig 3.10-b) mostrano una crescita seguita da un calo verso la metà del 1400 e da una successiva ripresa a metà del 1600. Il subset di Etiopia non mostra differenze sostanziali nell'andamento ma una stima di N_e leggermente più elevata.



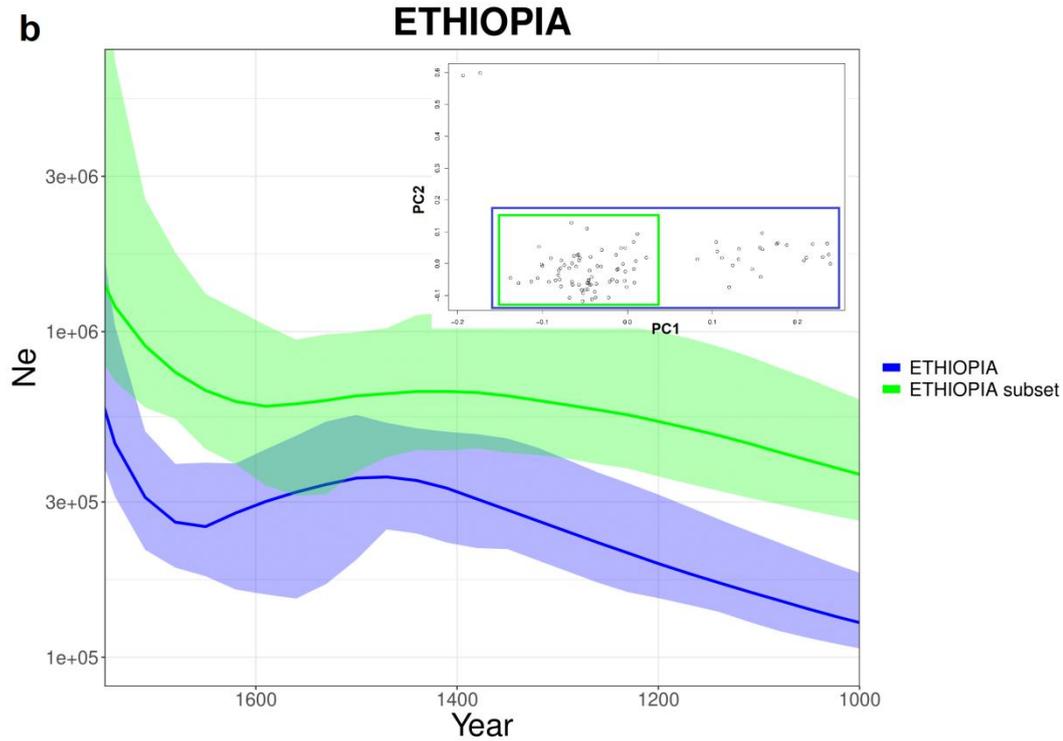


Fig 3.10: Curve stimate da IBDNe ($\times 4.89$) per (a) Egitto, (b) Etiopia con PCA e subset costruito basandosi su quest'ultima.

3.11. Africa e America

Avendo compiuto le analisi sulla totalità dei dati di “1000 Genomes Project”, può essere interessante osservare anche i risultati di Africa subsahariana (Fig 3.11.1) e America (Fig 3.11.2) nonostante non siano coinvolti direttamente nello studio e non si disponga di dati storici affidabili sulla dimensione di popolazione.

È possibile notare come (escludendo YRI), tutte le popolazioni africane presentino un calo demografico successivo al 1500 e solo due mostrino una ripresa (ESN e ACB). Per quanto riguarda l'America possiamo osservare un calo drastico della popolazione di CLM e PUR, un calo significativo di PEL e una certa costanza di MXL.

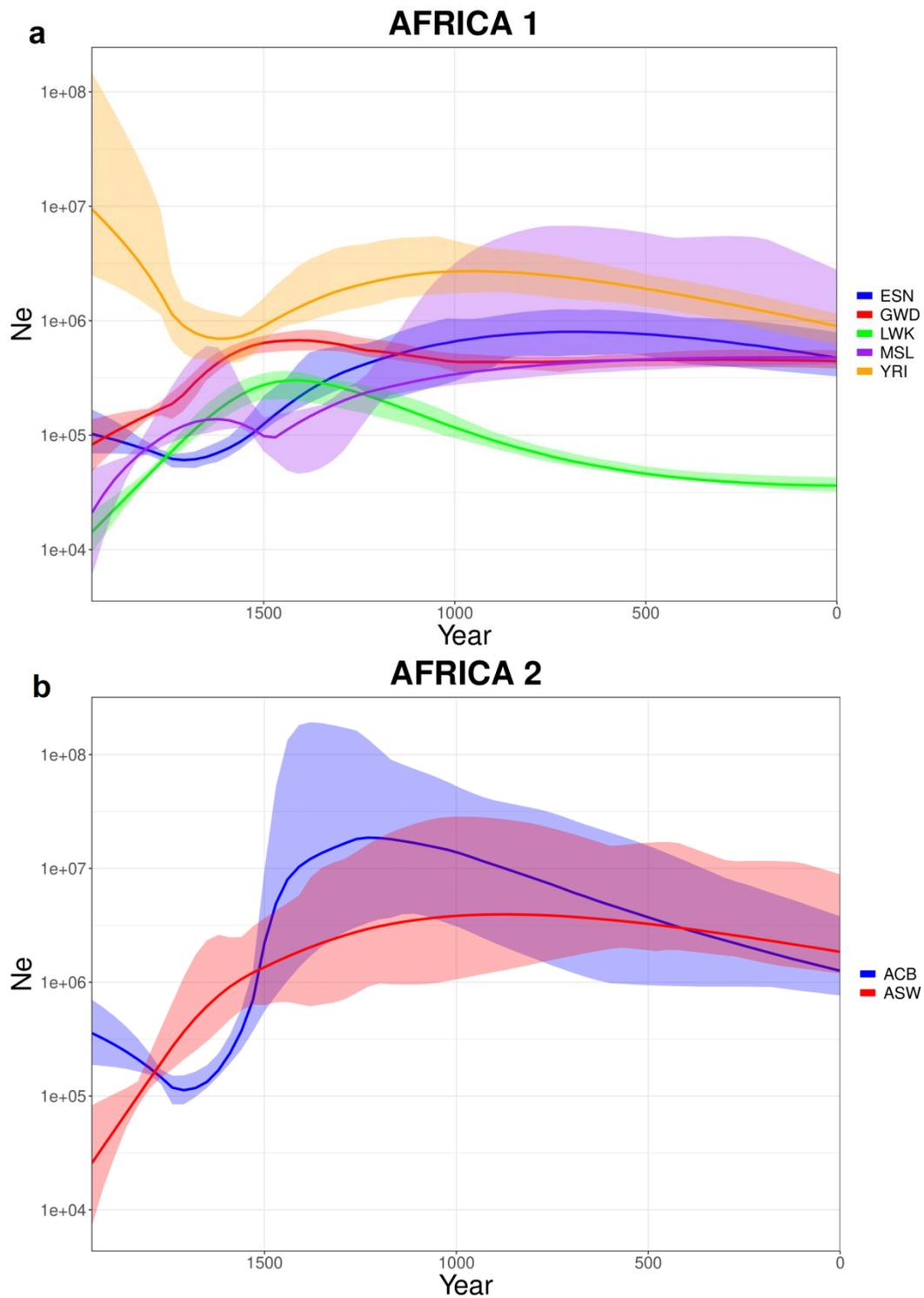


Fig 3.11.1: Curve stimate da IBDNe ($\times 4.89$) per varie popolazioni africane (a) ESN, GWD, LWK, MSL, YRI. (b) ACB e ASW essendo popolazioni africane residenti negli Stati Uniti d'America, sono state separate dalle precedenti per rendere i grafici più chiari

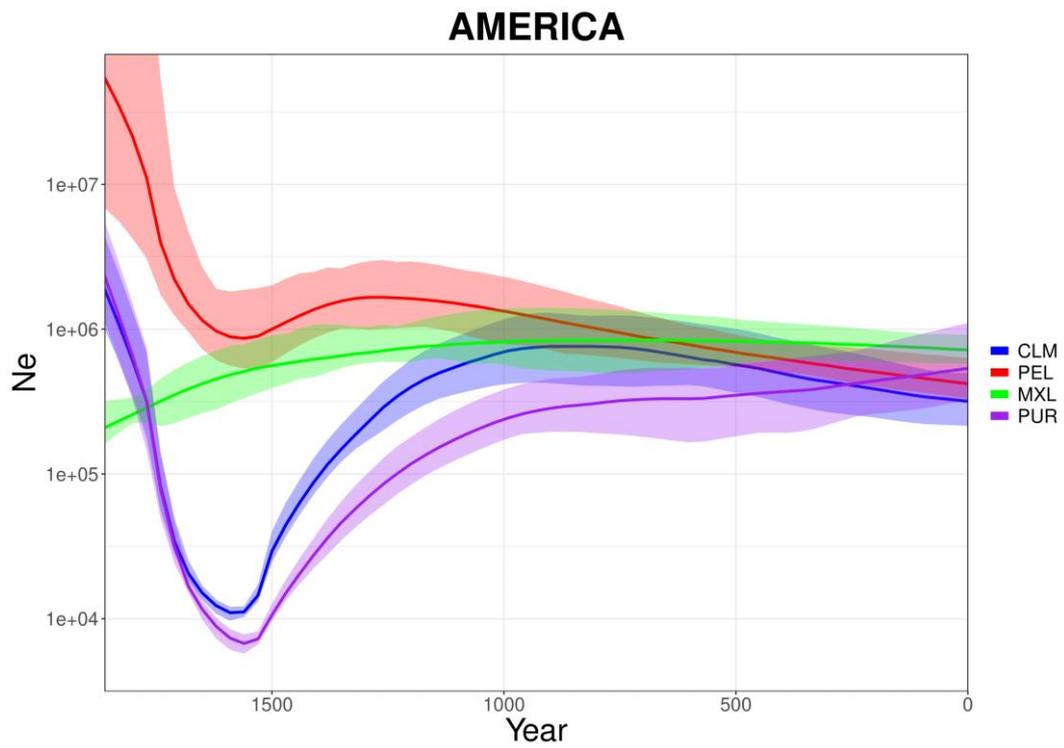


Fig 3.11.2: Curve stimate da IBDNe ($\times 4.89$) per varie popolazioni americane.

4. Discussione

4.1. Europa

4.1.1. Spagna

La curva prodotta da IBDNe (Fig 3.6-a) mostra una crescita costante e consistente della popolazione spagnola e il successivo andamento coincide con l'aspettativa di un calo della popolazione intorno al 1348: si vede infatti un picco nel 1350 che dà inizio a un calo continuo della popolazione del 87% fino al 1650. I dati storici mostrano una ripresa della crescita già nel 1400 mentre la curva di IBDNe mostra un crollo drastico della popolazione. Il periodo tra il 1348 e il 1650 in Spagna è segnato non solo dall'arrivo della Peste Nera e l'inizio di varie epidemie che dureranno tre secoli, ma anche dalla Reconquista spagnola e la conseguente cacciata dei Moriscos dalla Spagna, entrambi eventi dal forte impatto demografico (Vanoli, 2009; Ortiz & Bernard, 1993). A differenza della morte indiscriminata di individui dovuta alla peste, l'espulsione dei musulmani dalla Spagna potrebbe aver inciso sulla genetica spagnola più di quanto abbia inciso sulla sua dimensione di popolazione censita, essendo una rimozione forzata di una precisa componente popolazionistica (per lo più di origine mediorientale). La ripresa della popolazione dalla seconda metà del 1600 potrebbe essere legata al termine di questo periodo turbolento per i geni spagnoli poiché tra il 1647 e il 1652 si vive l'ultima grande epidemia (la "Grande Peste di Siviglia") e da circa 40 anni è terminata l'espulsione dei Moriscos. È plausibile dunque che il calo della popolazione osservato tramite IBDNe sia dovuto a una commistione di eventi e continuativo nel tempo, anziché un crollo istantaneo. Per quanto riguarda la magnitudine, fino alla prima metà del 1400 la stima di IBDNe coincide con i dati storici presenti, poi il crollo demografico si mostra più consistente di quello riportato dai record storici e non mostra una ripresa rapida.

4.1.2. Italia

Per l'Italia è stata utilizzata come popolazione "TSI" (Toscani in Italia). Tuttavia non si tratta di individui né da tutta Italia né da tutta la Toscana e dunque potrebbe essere una popolazione solo parzialmente rappresentativa. Osservando la curva prodotta da IBDNe (Fig 3.6-b) notiamo un calo costante della popolazione dal 150 CE fino al 1600 CE con una riduzione della popolazione dal picco alla valle del 77.8%. Non è chiaramente visibile un crollo della popolazione tra 1300 e 1400 che possa corrispondere all'effetto della peste in Europa, ma è possibile provare a spiegare ciò che si vede anche accostandosi alle simulazioni svolte (Paragrafo 3.5). I vari dati storici indicano che la popolazione Italiana pare essere rimasta piuttosto stabile (o leggermente oscillante intorno a una media) tra la caduta dell'impero romano (476 CE) e il 1348 (l'anno della Peste Nera). Questi dati somigliano all'andamento che vediamo nella popolazione simulata B (Fig 3.5), con il mantenimento di una certa costanza seguito da un crollo istantaneo. Vediamo che nello stimare la curva di IBDNe per la popolazione B, il software produce una curva molto simile a quella per TSI, spostando indietro il picco e mostrando un calo lieve e duraturo piuttosto che istantaneo. È importante però tenere conto del fatto che, anche se la curva di TSI somiglia alla stima di IBDNe per la simulazione B non significa necessariamente che la reale curva di TSI corrisponda alla reale curva di simulazione B poiché a due curve uguali stimate da IBDNe potrebbero corrispondere andamenti demografici differenti. IBDNe tenta di dare la propria spiegazione ai dati che riceve compiendo approssimazioni e non essendo in grado di rilevare piccole fluttuazioni nella popolazione. Ciò non toglie che esista una buona somiglianza tra la curva reale di simulazione B e i dati storici per l'Italia e una tra la curva inferita da IBDNe per TSI e quella per Simulazione B. Al termine del calo demografico si osserva una valle con un recupero intorno al 1600 e potrebbe essere rilevante notare che questo periodo coincide con gli ultimi focolai di peste importanti in Europa e precede la crescita demografica dovuta alla rivoluzione industriale un secolo dopo. Un aspetto importante che vale la pena menzionare al di là dell'andamento della curva è la magnitudine: osserviamo una stima di popolazione molto inferiore a quella data dai record storici (sia per l'Italia che solo per la Toscana). È plausibile che la spiegazione di questo fenomeno sia la ridotta rappresentanza che l'Italia ha all'interno della popolazione TSI. Probabilmente per altri tipi di studi (Novembre et al., 2008), TSI rappresentano un buon proxy per gli italiani ma in questo caso, nel tentativo di stimare la dimensione di popolazione portano a una sottostima. Per quanto riguarda la Toscana vale il medesimo discorso, si tratta di un campione di circa 100 individui che potrebbe non essere totalmente rappresentativo, tenendo anche conto del fatto che ci interfacciamo con i geni dei discendenti dei sopravvissuti ma non necessariamente tutti coloro che sopravvissero alla peste hanno avuto dei discendenti, come spiegato dalla teoria del coalescente (Kingman, 1982).

4.1.3. Gran Bretagna

I dati storici per la Gran Bretagna sono generalmente discordanti tra loro, mostrando tutti un declino nel 1300, ma di entità e durata variabile. La curva generata da IBDNe rispecchia bene l'andamento generale dei dati storici, mostrando un declino accentuato nel 1300 (nonostante il reale picco della curva sia nel 480, ma con una decrescita poco rilevante fino al 1350) e un recupero nel 1600. Osservando la curva, si potrebbe pensare che la Gran Bretagna abbia avuto una crescita rapida subito prima del crollo nel 1350 e che quindi non viene identificata da IBDNe (se non nella banda di errore) che mostra un declino costante dal 480 al 1620 con una riduzione della popolazione del 81%. Anche in questo caso abbiamo una curva che sembra sottostimare la popolazione ma corrispondere nell'andamento, potrebbe nuovamente trattarsi di una questione di rappresentatività del campione.

4.1.4. Finlandia e CEU

Per Finlandia e CEU non disponiamo di dati storici affidabili. Nel caso della Finlandia manca un record storico e dobbiamo quindi limitarci ad apprezzare la curva generata da IBDNe, che mostra una crescita costante. La Finlandia non è stata coinvolta dalle epidemie di peste grazie al suo isolamento dagli altri paesi europei e alla bassa densità di popolazione (Armstrong, 2021) e questo pare essere confermato dalla curva che riusciamo a ottenere, priva di alcun calo demografico. Con CEU ci si trova ad affrontare una popolazione che non rappresenta un paese odierno vero e proprio. CEU sono abitanti degli Stati Uniti d'America emigrati dall'Europa nei secoli passati (presumibilmente dal centro-Europa tra Francia e Paesi Bassi) e appartenenti a comunità piuttosto isolate e chiuse (con quindi poche possibilità di rimescolare il proprio DNA con "estranei"). Osservando la curva di IBDNe possiamo notare un andamento simile a IBS (Fig 4.1.4) ma dopo il crollo intorno al 1350, la popolazione non vede un recupero. Questo potrebbe essere la diretta conseguenza dell'emigrazione di questa popolazione che quindi vedrebbe un secondo declino (l'emigrazione non è altro che un collo di bottiglia), quando invece le altre popolazioni iniziano a mostrare una ripresa; dunque siamo in grado di osservare unicamente una decrescita costante.

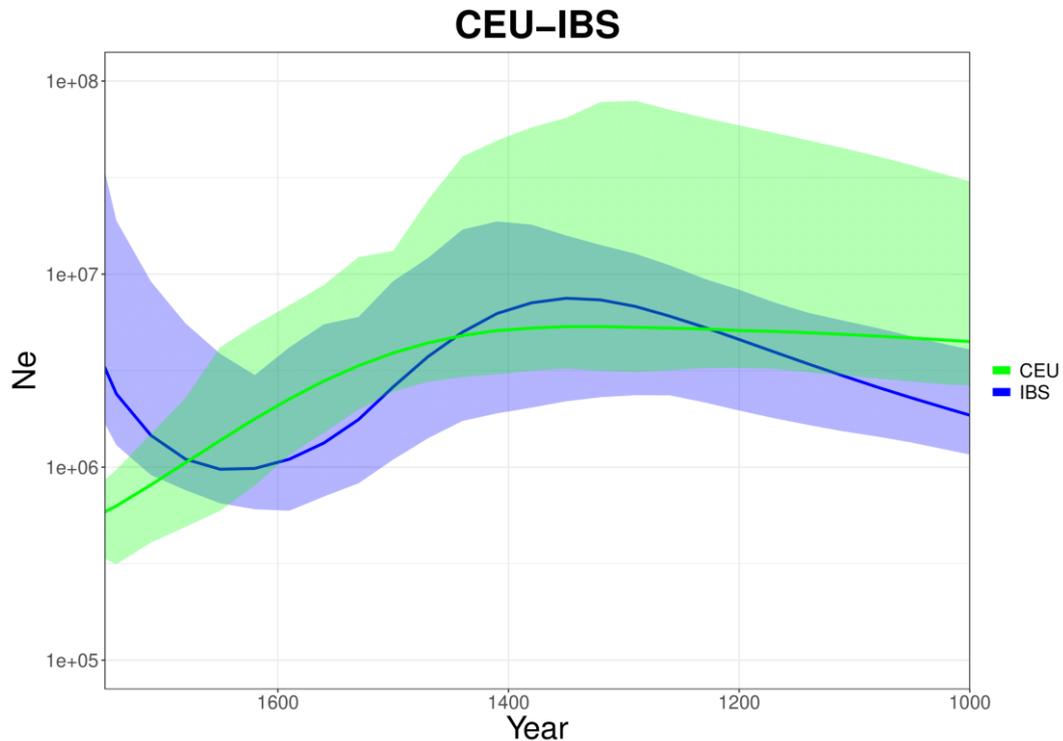


Fig 4.1.4: Confronto tra le curve di CEU e IBS

4.1.5. Estonia

L'Estonia è stata aggiunta ai dati Europei per la presenza di un grande numero di individui disponibili e come controllo per il lavoro svolto, poiché la popolazione estone è stata già precedentemente analizzata con gli stessi strumenti di questo studio (Pankratov et al., 2020). I risultati ottenuti da questa analisi coincidono con quelli precedentemente ottenuti da Pankratov. Osserviamo nel caso dell'Estonia una corrispondenza perfetta tra i dati storici e la curva generata da IBDNe, se confrontata con gli altri paesi osservati fino ad ora. Questo non invalida le analisi svolte sulle altre popolazioni, pone invece l'attenzione sull'importanza di avere un panel di individui il più completo e rappresentativo possibile. Nel caso dell'Estonia sono stati utilizzati i dati SNP array di 2420 individui provenienti da tutto il paese, paragonati ai ~100 individui di tutte le popolazioni di 1KG project, prelevati spesso da aree ristrette. Ad eccezione della sottopopolazione "Parnu", non osserviamo un calo demografico nel periodo corrispondente alle epidemie in Europa, è plausibile dunque che anche l'Estonia, come la Finlandia, sia riuscita a sfuggire alla peste, probabilmente per ragioni simili.

4.2. Asia orientale

Sappiamo per certo che la peste ha colpito la Cina, alcune fonti parlano di una riduzione del 25% della popolazione (Naphy & Spicer, 2004), ma in una popolazione così numerosa, vale la pena chiedersi se una riduzione da 120 a 90 milioni di individui possa aver avuto un effetto sulla genetica tale da permetterci di vedere questo declino tramite le tecnologie a nostra disposizione. Osservando le curve generate da IBDNe per la Cina, in particolare per CHB, non notiamo nulla di simile, anzi la crescita esplosiva della popolazione cinese pare mettere in difficoltà le nostre stime, proponendo una popolazione decisamente non compatibile con i numeri reali. Osservando la curva di IBDNe la popolazione nipponica, invece, pare subire un drastico calo dall'inizio del 1400 alla fine del 1600, sovrapponibile ai due secoli ritenuti i più bui della storia giapponese. Quest'epoca vede infatti sanguinosi scontri tra i vari signori della guerra, la caduta del potere dello Shōgun con il conseguente scoppio della guerra civile e l'inizio del periodo Sengoku "L'era del paese in guerra" (Ferejohn & Rosenbluth, 2010). Nonostante questo, le stime storiche della popolazione non sembrano mostrare alcuna decrescita in quel periodo.

Per quanto riguarda il sud est asiatico, la curva di IBDNe per KHV combacia bene con i dati storici, mostrando una lenta crescita, leggermente inferiore in magnitudine alle stime complessive dell'Indocina. Osservando la suddivisione in due sottogruppi di KHV, la stima complessiva (blu) potrebbe essere il risultato di un gruppo in crescita (verde) e un gruppo in calo (rosso).

4.3. Subcontinente indiano

Esistono numerose fonti occidentali medievali che narrano di come la peste abbia devastato l'India, ma spesso non si tiene in considerazione di come la parola India non avesse un significato preciso all'epoca e poteva riferirsi sia all'India vera e propria, che all'Etiopia, al Vietnam o all'Asia in generale. Stando alle fonti medievali indiane, anch'esse numerose, non si trovano menzioni di epidemie paragonabili alla peste; si parla di epidemie, ma con descrizioni ben diverse da quelle che riportano le fonti occidentali, alcuni cronisti sostengono anche di aver contratto queste malattie sopravvivendone e non riportando sintomi ascrivibili alla Peste Nera che colpì l'occidente. Questi stessi cronisti documentano però la peste in Siria descrivendola esattamente come l'epidemia europea (Anandavalli, 2007). Per quanto riguarda le curve di IBDNe, possiamo apprezzare una grande varietà di andamenti e magnitudini che ci confermano come il subcontinente indiano sia composto da un'incredibile ricchezza di popoli ben distinti oltre che per cultura anche per storia demografica.

4.4. Via della Seta

Non ci sono record storici affidabili per quanto riguarda la via della seta e i paesi del centro-Asia, poiché nel corso dei secoli questi paesi sono stati divisi in molteplici stati e imperi cambiando la propria bandiera numerose volte. È dunque complesso risalire a dati come i censimenti per aree che formalmente non erano considerate uno stato vero e proprio, né rappresentate da zone geografiche ben delimitate come penisole o isole. Possiamo però pensare al fatto che la via della seta abbia goduto di grande prestigio prima del suo declino poiché era l'unica via di comunicazione sicura che collegasse la Cina con l'Europa. Sappiamo che la via della seta divenne sempre meno sicura e popolata dopo che l'impero ottomano si impose come barriera tra oriente e occidente nel 1453 con la conquista di Costantinopoli (Rosenwein, 2014). Dalla metà del 1400 l'Asia centrale smise di essere una via percorribile e nei due secoli successivi vennero scoperte nuove rotte commerciali via mare che permisero all'Europa di commerciare con l'estremo oriente e l'India senza dover passare attraverso i paesi occupati dall'impero ottomano. Osservando le curve prodotte con IBDNe possiamo apprezzare come varie popolazioni subiscano un calo demografico consistente proprio intorno alla metà del 1400 o poco successivamente (Crimea, Armenia, Kirghizistan, Tajikistan). Osservando i paesi dell'Asia centrale, ci troviamo dinnanzi a una grande varietà di andamenti demografici difficili da spiegare senza una conoscenza capillare della storia di quest'area geografica. Possiamo osservare però tre popolazioni con un andamento decisamente peculiare, la cui spiegazione risiede probabilmente in un problema tecnico e non in un andamento demografico poco plausibile: le curve di Azerbaijan, Mongolia e Uzbekistan mostrano un costante declino. La spiegazione più verosimile per questo fenomeno sta nella poca omogeneità interna delle tre popolazioni, come mostrato nelle PCA svolte.

4.5. Etiopia ed Egitto

La diffusione della peste bubbonica in Etiopia nel XIV secolo è ad oggi dibattuta ma esistono alcune testimonianze che sembrerebbero confermare la presenza di epidemie ascrivibili ad essa, forse provenienti dal medio oriente, proprio nel periodo in cui l'Europa soffriva la Peste Nera (Chouin, 2018). La curva prodotta da IBDNe mostra un discreto calo demografico che però ha inizio un secolo dopo e che ha termine a metà del 1600.

È ben noto che la Peste Nera sia giunta in Egitto e abbia causato enormi danni alla società egiziana dell'epoca (Chouin, 2018) ma dai dati storici sulla dimensione di popolazione si riscontra solo un lieve calo, non primo e non ultimo in quella che sembra essere una storia demografica soggetta a molteplici fluttuazioni. La curva prodotta da IBDNe mostra una certa stabilità demografica con una leggera crescita fino al 1620 dove si inizia a vedere un declino. Uno sguardo alla storia dell'Egitto

ci rivela che il XVII e il XVIII secolo sono stati un periodo di grande crisi dovuto al susseguirsi di sei carestie (Quataert, 2005).

4.6. Africa e America

La grande complessità della storia dell’Africa subsahariana, unita alla scarsità di fonti facilmente accessibili e affidabili, rendono l’indagine della demografia africana particolarmente insidiosa. Ad oggi la diffusione della Peste Nera nell’Africa subsahariana resta una affascinante ipotesi (Chouin, 2018); l’analisi delle varie popolazioni africane e i relativi andamenti demografici richiederebbe uno studio a sé. Potrebbe però essere rilevante notare la grande somiglianza nell’andamento generale delle curve generate da IBDNe per le varie popolazioni africane.

Per quanto riguarda la storia americana invece, la “scoperta delle Americhe” e il conseguente genocidio delle popolazioni indigene da parte degli europei (Diamond, 2014), vengono chiaramente confermate anche dai dati genetici che mostrano un crollo demografico indiscutibilmente legato all’arrivo degli europei con una riduzione della popolazione (per CLM e PUR) di oltre il 99%. Per quanto riguarda PEL e MXL la spiegazione più semplice per gli andamenti che osserviamo, piuttosto diversi da CLM e PUR, potrebbe risiedere nell’alto grado di admixture di queste popolazioni con i conquistatori europei.

4.7. Interpretazione generale dei risultati

Nell’osservazione delle curve generate da IBDNe e soprattutto nel tentativo di ricondurle a dati e avvenimenti storici, non va dimenticato che esistono molteplici assunzioni da tenere a mente alla base della genetica di popolazione, della bioinformatica e della statistica. Di conseguenza, il software utilizzato si basa su modelli teorici che tentano di descrivere la realtà ma che non ne sono la perfetta rappresentazione. Esistono differenze radicali tra quella che è la dimensione effettiva di popolazione (N_e) e la dimensione censita (N_c) e per quanto siano state sviluppate metodiche per avvicinare il più possibile una misura all’altra (Palamara et al., 2012; Pankratov et al., 2020), restano due dati diversi condizionati da bias differenti nelle loro stime.

Ne corrisponde a N_c solamente se specifiche condizioni sono soddisfatte: alla luce dei risultati e delle simulazioni svolte sembrano avere grande importanza la dimensione di popolazione prima e dopo un crollo demografico, l’intensità del crollo stesso e la durata del periodo di tempo tra la fine del crollo e l’inizio della ripresa. La magnitudine di un crollo demografico (o di una crescita) è determinante perché questo sia rilevabile tramite IBDNe e piccole oscillazioni non sono rilevabili; ciò che è invece osservabile è il risultato generale dell’andamento di un periodo. È molto importante che il campione utilizzato per l’analisi genetica sia sufficientemente grande e rappresentativo della popolazione cui fanno

riferimento i dati storici. È fondamentale inoltre tenere a mente che l'analisi svolta riguarda solo gli antenati di individui del presente, che non rappresentano necessariamente tutta la popolazione che ha vissuto in uno specifico momento e in uno specifico luogo, come spiegato dalla teoria del coalescente (Kingman, 1982). Osservando i dati storici è importante ricordare che spesso si dispone solo di punti scollegati e non di una curva continua. È possibile avere un dato per il 1300, uno per il 1400 e non notare nulla nel mezzo semplicemente perché non esiste un dato intermedio che potrebbe mostrare ad esempio un picco tra i due punti o una valle. Nel definire date e intervalli temporali è importante considerare che l'unità di tempo di cui si dispone per l'analisi genetica è quella della generazione (circa 30 anni) e che si inizia a contare le generazioni da una data arbitraria (il 1950), basata sulla nascita degli individui campionati, che comunque risultano essere svariate migliaia di persone non necessariamente tutte coetanee. Non è possibile avere una risoluzione che permetta di osservare fenomeni repentini come l'arrivo della peste dall'Asia che ha impiegato pochi anni a diffondersi in Europa, se la scala di misura di cui si dispone va di trenta anni in trenta anni.

5. Conclusioni

Generalmente i risultati ottenuti con IBDNe sono convincenti (alla luce del Cap 4.7) e confrontabili con eventi storici e date. Nell'Europa occidentale assistiamo a crolli demografici più o meno in corrispondenza della Peste Nera, ma sicuramente influenzati dalla stessa, generalmente con una riduzione di N_e maggiore del 75%. In ciascun caso è visibile un recupero della popolazione nel XVII secolo, quando le epidemie che hanno segnato l'Europa per trecento anni, cessarono. Il periodo della Peste Nera e delle successive epidemie sembra aver lasciato una chiara traccia genetica nelle popolazioni dell'Europa occidentale. Per quanto riguarda le aree dell'Europa nord-orientale invece, non si riscontra alcun crollo demografico, coerentemente con i dati storici a disposizione. Non siamo in grado di osservare una chiara traccia della peste nei paesi dell'Est Asiatico, probabilmente a causa della popolazione già numerosa all'epoca, che pare non aver risentito geneticamente della peste. Per quanto riguarda i paesi dell'Asia centrale, probabilmente la bassa densità di popolazione ha reso difficile il propagarsi dell'epidemia localmente, rendendo le aree della via della seta protagoniste solo in quanto via di passaggio per l'epidemia verso occidente, ma senza lasciare una vera impronta nella popolazione locale. Sembra invece più evidente in queste popolazioni un declino successivo alla chiusura della via della seta. Il subcontinente indiano non sembra aver risentito della Peste Nera come i cronisti europei sembravano narrare, a differenza dei cronisti indiani che non menzionano mai epidemie di peste in India. È bene ricordare però che dove vediamo un crollo intorno al XIV secolo, non è necessariamente detto che si tratti della Peste Nera poiché come è noto "la correlazione non implica causalità". Probabilmente alla peste vanno accostate numerose altre cause del declino della popolazione europea, che già nei 20 anni precedenti all'epidemia, ad esempio, stava attraversando un periodo di crisi dovuto alle carestie (Naphy & Spicer, 2004). Viceversa dove non vediamo un crollo di N_e , non significa che la peste non abbia avuto un impatto su N_e e sulla società dell'epoca. Nonostante la correlazione non implichi causalità, la situazione dei popoli nativi del centro e sud America, che mostrano un crollo del 99% in seguito all'arrivo dei conquistatori europei, lascia poco spazio ai dubbi e a eventuali concause per spiegare una tale riduzione di N_e .

In conclusione è di fondamentale importanza tenere a mente che il presente studio non tenta di sostituire i dati storici con dati genetici, ma di fornire una prospettiva differente con cui guardare agli stessi eventi e pertanto non si pone come un'alternativa ma come un complemento ai dati già noti e diffusi.

6. Bibliografia

- Adamovicz, J. J., & Worsham, P. L. (2012). Plague. In *Biodefense Research Methodology and Animal Models, Second Edition*.
<https://doi.org/10.1201/b11523>
- Anandavalli, L. (2007). The Black Death In Medieval India. *Tangents*, 6, 20–25.
- Armstrong, D. (2021). *How Iceland, Finland, and Poland Were Spared from the Wrath of the Black Death*. Wondrium Daily.
<https://www.wondriumdaily.com/how-iceland-finland-and-poland-were-spared-from-the-wrath-of-the-black-death/>
- Arredi, B., Poloni, E. S., Paracchini, S., Zerjal, T., Fathallah, D. M., Makrelouf, M., Pascali, V. L., Novelletto, A., & Tyler-Smith, C. (2004). A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *American Journal of Human Genetics*, 75(2), 338–345.
<https://doi.org/10.1086/423147>
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurler, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., ... Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74.
<https://doi.org/10.1038/nature15393>
- Banister, J. (1992). A Brief History of China's Population. In D. (eds) Poston, D.L., Yaukey (Ed.), *The Population of Modern China* (pp. 51–57). Springer.
- BankOfEngland. (2021). *Population (England)*.
<https://www.bankofengland.co.uk/statistics/research-datasets>
- Barbiera, I., Castiglioni, M., & Dalla-Zuanna, G. (2016). LA MORTALITÀ AI TEMPI DELLA PESTE NERA: INDAGINE PALEODEMOGRAFICA DELL'ITALIA DI XIV SECOLO. *Archeologia Medievale*, XLIII.
- Barbieri, R., Signoli, M., Chev , D., Costedoat, C., Tzortzis, S., Aboudharam, G., Raoult, D., & Drancourt, M. (2021). Yersinia pestis: The natural history of Plague. *Clinical Microbiology Reviews*, 34(1), 1–44.
<https://doi.org/10.1128/CMR.00044-19>
- Bellettini, A. (1987). *La popolazione italiana*. Einaudi.
- Browning, B. L., & Browning, S. R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. *American Journal of Human Genetics*, 93(5), 840–851. <https://doi.org/10.1016/j.ajhg.2013.09.014>
- Browning, S. R., & Browning, B. L. (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *American Journal of Human Genetics*, 97(3), 404–418.
<https://doi.org/10.1016/j.ajhg.2015.07.012>
- Browning, S. R., Browning, B. L., Daviglus, M. L., Durazo-Arvizu, R. A.,

- Schneiderman, N., Kaplan, R. C., & Laurie, C. C. (2018). Ancestry-specific recent effective population size in the Americas. *PLoS Genetics*, *14*(5), 1–22. <https://doi.org/10.1371/journal.pgen.1007385>
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 1–16. <https://doi.org/10.1186/s13742-015-0047-8>
- Chouin, G. (2018). Reflections on plague in African history (14th–19th c.). *Afriques*, *09*. <https://doi.org/10.4000/afriques.2228>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Diamond, J. (2014). *Armi, Acciaio e Malattie. Breve storia del mondo negli ultimi tredicimila anni*. Einaudi.
- Durand, J. D. (1960). The Population Statistics of China, A.D. 2-1953. *Population Studies*, *13*(3), 209–256.
- Ferejohn, J. A., & Rosenbluth, F. M. (2010). *War and State Building in Medieval Japan* (J. A. Ferejohn & F. M. Rosenbluth (eds.)). Stanford University Press.
- Green, M. H. (2020). The four black deaths. *American Historical Review*, *125*(5), 1601–1631. <https://doi.org/10.1093/ahr/rhaa511>
- Izdebski, A., Guzowski, P., Poniak, R., Masci, L., Palli, J., Vignola, C., Bauch, M., Coccozza, C., Fernandes, R., Ljungqvist, F. C., Newfield, T., Seim, A., Abel-Schaad, D., Alba-Sánchez, F., Björkman, L., Brauer, A., Brown, A., Czerwiński, S., Ejarque, A., ... Masi, A. (2022). Palaeoecological data indicates land-use changes across Europe linked to spatial heterogeneity in mortality during the Black Death pandemic. *Nature Ecology and Evolution*, *6*(3), 297–306. <https://doi.org/10.1038/s41559-021-01652-4>
- Jobling, Hollox, Hurles, Kivisild, & Tyler-Smith. (2014). *Human Evolutionary Genetics* (Second Edi). CRC Press.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). <https://doi.org/10.1098/rsta.2015.0202>
- Kelleher, J., Etheridge, A. M., & McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Computational Biology*, *12*(5), 1–22. <https://doi.org/10.1371/journal.pcbi.1004842>
- Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and Their Applications*, *13*(3), 235–248. [https://doi.org/10.1016/0304-4149\(82\)90011-4](https://doi.org/10.1016/0304-4149(82)90011-4)

- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Lo Cascio, E., & Malanima, P. (2005). Cycles and Stability. Italian Population before the Demographic Transition (225 B.C.- A.D. 1900). *Cycles and Stability. Italian Population before the Demographic Transition (225 B.C.- A.D. 1900)*, 21(3), 197–232. <https://doi.org/10.1410/20933>
- Malanima, P., & Breschi, M. (2002). Demografia ed economia in Toscana: il lungo periodo (secoli XIV-XIX). *Prezzi, Redditi, Popolazioni in Italia: 600 Anni (Dal Secolo XIV Al Secolo XX)*, July, 109–142.
- Maps, G. (2021). MyMaps. In *Google maps* (p. 1). Available on <https://www.google.com/maps/d/u/0/>.
- McEvedy, C., & Jones, R. (1979). *Atlas of world population history*. Penguin Books Ltd. and Allen Lane.
- Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J. L., Martínez, I., Gracia, A., De Castro, J. M. B., Carbonell, E., & Pääbo, S. (2014). A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature*, 505(7483), 403–406. <https://doi.org/10.1038/nature12788>
- Mezzavilla, M., Vozzi, D., Pirastu, N., Giroto, G., d’Adamo, P., Gasparini, P., & Colonna, V. (2014). Genetic landscape of populations along the Silk Road: Admixture and migration patterns. *BMC Genetics*, 15(1), 1–10. <https://doi.org/10.1186/s12863-014-0131-6>
- Microsoft, C. (2010). *Microsoft Excel*. Retrieved from <https://office.microsoft.com/excel>.
- Naphy, W., & Spicer, A. (2004). *Plague. Black Death and Pestilence in Europe*. Tempus Publishing Limited.
- Nielsen, R., & Slatkin, M. (2013). *An Introduction to Population Genetics, Theory and application* (A. Sinauer (ed.)). Sinauer Associates Inc.
- Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., Indap, A., King, K. S., Bergmann, S., Nelson, M. R., Stephens, M., & Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, 456(7219), 274–274. <https://doi.org/10.1038/nature07566>
- Ortiz, A. D., & Bernard, V. (1993). *Historia de los moriscos. Vida y tragedia de una minoría*. Alianza Editorial. <http://www.muniloslagos.cl/index.php/patentes-y-rentas/item/84-historia-de-los-lagos#>
- Pagani, L., Schiffels, S., Gurdasani, D., Danecek, P., Scally, A., Chen, Y., Xue, Y., Haber, M., Ekong, R., Oljira, T., Mekonnen, E., Luiselli, D., Bradman, N., Bekele, E., Zalloua, P., Durbin, R., Kivisild, T., & Tyler-Smith, C. (2015). Tracing the Route of Modern Humans out of Africa by Using 225

- Human Genome Sequences from Ethiopians and Egyptians. *American Journal of Human Genetics*, 96(6), 986–991.
<https://doi.org/10.1016/j.ajhg.2015.04.019>
- Palamara, P. F., Lencz, T., Darvasi, A., & Pe'er, I. (2012). Length distributions of identity by descent reveal fine-scale demographic history. *American Journal of Human Genetics*, 91(5), 809–822.
<https://doi.org/10.1016/j.ajhg.2012.08.030>
- Pankratov, V., Montinaro, F., Kushniarevich, A., Hudjashov, G., Jay, F., Saag, L., Flores, R., Marnetto, D., Seppel, M., Kals, M., Võsa, U., Taccioli, C., Möls, M., Milani, L., Aasa, A., Lawson, D. J., Esko, T., Mägi, R., Pagani, L., ... Metspalu, M. (2020). Differences in local population history at the finest level: the case of the Estonian population. *European Journal of Human Genetics*, 28(11), 1580–1591. <https://doi.org/10.1038/s41431-020-0699-4>
- Perry, R. D., & Fetherston, J. D. (1997). *Yersinia pestis* - Etiologic agent of plague. *Clinical Microbiology Reviews*, 10(1), 35–66.
<https://doi.org/10.1128/cmr.10.1.35>
- Quataert, D. (2005). *The Ottoman Empire, 1700-1922*. Cambridge University Press.
- RCore, T. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Roosen, J., & Curtis, D. R. (2018). Dangers of noncritical use of historical plague data. *Emerging Infectious Diseases*, 24(1), 103–110.
<https://doi.org/10.3201/eid2401.170477>
- Rosenwein, B. H. (2014). *A short history of the middle Ages* (Fourth Edi). University of Toronto press.
- Russell, J. C. (1972). *Population in Europe, 500-1500* (C. M. Cipolla (ed.); The Fontan).
- Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G. R., Reepmeyer, C., Petchey, F., Fernandes, D., Fu, Q., Harney, E., Lipson, M., Mallick, S., Novak, M., Rohland, N., Stewardson, K., Abdullah, S., Cox, M. P., ... Reich, D. (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature*, 538(7626), 510–513.
<https://doi.org/10.1038/nature19844>
- Spyrou, M. A., Musralina, L., Ruscone, G. A. G., Kocher, A., Borbone, P., Khartanovich, V. I., Buzhilova, A., Djansugurova, L., Bos, K. I., Kühnert, D., Haak, W., Slavin, P., & Krause, J. (2022). The source of the Black Death in fourteenth-century central Eurasia. *Nature*, July 2021.
<https://doi.org/10.1038/s41586-022-04800-3>
- Urlanis, B. T. (1941). *Rost naseleniia v Evrope : opyt ischisleniia*. OGIZ-Gospolitizdat.
- Vanoli, A. (2009). *La Reconquista*. Il Mulino.

Wickham, C. (2016). *Medieval Europe*. Yale Univeristy Press.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

