



UNIVERSITÀ DI PARMA

# UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN  
“*Biologia Evoluzionistica ed Ecologia*”  
CICLO XXXV

**Analisi paleogenetiche su resti umani antichi.**

**Gli Etruschi di *Felsina* e dintorni: evidenze genomiche dall'Etruria padana**

Coordinatore:  
Prof. Pierluigi Viaroli

Tutore:  
Prof.ssa Martina Lari

Co-tutore:  
Prof. Cosimo Posth

Dottorando: Valentina Zaro

Anni Accademici 2019/2020 – 2021/2022

Il dottorato in Biologia Evoluzionistica ed Ecologia – XXXIV ciclo è in convenzione tra le Università di Ferrara, Firenze e Parma (sede amministrativa).



# Sommario

1.	Introduzione: parte I – La paleogenetica .....	1
1.1	Dai primi passi al Nobel.....	1
1.2	Gli aspetti critici del DNA antico.....	5
1.3	I diversi campi di applicazione e il loro potenziale informativo.....	8
2.	Introduzione: parte II – Storia genetica dell’Europa dal Neolitico all’età del Ferro .....	10
3.	Introduzione: parte III – Gli Etruschi .....	17
3.1	Cenni storici generali.....	17
3.2	Bologna- <i>Felsina</i> : fulcro dell’Etruria padana .....	20
3.3	Origine ed eredità del profilo genetico etrusco .....	24
4.	Obiettivo dello studio.....	30
5.	Materiali e metodi.....	32
5.1	Contesto archeologico e datazione dei campioni .....	32
5.1.1	Necropoli di Piazza VIII Agosto .....	33
5.1.2	Necropoli di Piazza Azzarita .....	35
5.1.3	Necropoli dell’area ex Manifattura Tabacchi.....	37
5.1.4	Necropoli di Via Saffi .....	38
5.1.5	Necropoli di Giardini Margherita .....	39
5.1.6	Necropoli di Casteldebole .....	40
5.1.7	Necropoli di Casalecchio di Reno .....	41
5.1.8	Recenti aggiunte al dataset etrusco.....	43
5.2	Metodi di laboratorio: dalla fase di campionamento al sequenziamento.....	45
5.3	Analisi bioinformatica .....	49
5.3.1	Processamento dei dati grezzi.....	50
5.3.2	Determinazione del sesso e stima della contaminazione .....	51
5.3.3	Markers uniparentali: assegnazione degli aplogruppi .....	52
5.3.4	Genotyping e costruzione del dataset di analisi .....	53
5.3.5	Determinazione delle parentele e calcolo delle runs of homozygosity.....	53
5.3.6	Analisi delle componenti principali e ADMIXTURE .....	55
5.3.7	Raggruppamento in clusters .....	56
5.3.8	F-statistics.....	57
5.3.9	qpAdm modelling.....	59
6.	Risultati e discussione .....	62
6.1	Analisi dei dati di sequenziamento <i>shotgun</i> .....	62

6.2	Statistiche relative al processamento dei dati di cattura nucleare.....	70
6.3	Attribuzione del sesso biologico e valutazione della contaminazione.....	77
6.4	Markers uniparentali e legami biologici.....	84
6.5	PCA e ADMIXTURE.....	92
6.6	Definizione dei clusters genetici.....	99
6.7	Evidenze popolazionistiche emerse dalle statistiche F .....	105
6.8	Modelli di <i>admixture</i> ottenuti con qpAdm.....	112
7.	Conclusioni .....	125
	Riferimenti.....	129



## Abstract

During the Early Iron Age (~9<sup>th</sup> century BCE), the Etruscan civilization flourished in a large area of central Italy with local expansions into neighboring regions, such as Campania and the Po Valley. Contrary to previous suggestions supporting an Anatolian origin of the Etruscan civilization, recent genomic studies on Etruscan-associated human remains from central Italy show that the Etruscans carried a local genetic profile, shared with both contemporary Italic groups (such as the Latins) and preceding Bronze Age populations. Despite the Etruscan gene pool remaining highly homogeneous for almost 800 years (8<sup>th</sup>-1<sup>st</sup> century BCE), individuals with a nonlocal genetic ancestry of likely central European origin are attested in Etruria from the 7<sup>th</sup> century BCE onwards, suggesting direct contacts between the Etruscan population and Celtic-related groups as early as the Hallstatt period. To date, the available genomic data on the Etruscan population are only representative of classical Etruria which includes the territories of present-day Tuscany, northern Lazio and Umbria. Given the extent of the Etruscan expansion and the relevance of some of the cities founded in these regions, the present study focused on the genomic analyses of over 50 ancient individuals associated with the Etruscan culture and coming from several necropolises excavated in Bologna (and surroundings: Casteldebole and Casalecchio di Reno) linked to the ancient city of *Felsina*, the main Etruscan center of the Po Valley. Most of the samples are dated between the VIII and the V century BCE and come from mixed-rite necropolises where, during this time, inhumation starts to become the prevalent funerary practice. Radiocarbon dating performed on some of the samples revealed the presence of a few individuals belonging to the Roman Imperial period (all coming from Casalecchio).

Besides contributing to a better definition of the Etruscan-related gene pool by adding new genomic evidence from a region not yet investigated, the newly generated genomes provide information on the genetic affinity shared between the Etruscan communities of the Tyrrhenian and the Po Valley areas. In addition, the time transect covered by the Etruscan dataset corresponds to a moment of great growth and radical transformation during which the town of *Felsina* quickly develops into the major political and economic center of the region. Archaeological findings show clear evidence of extensive contacts between the Etruscans of *Felsina* and Celtic-related groups from both north Italy and transalpine territories before the well-known Celtic/Gallic invasions of the IV-III century BCE (La Tène phase). Palaeogenomic interest in the newly produced dataset also lies on the possibility to better investigate the dynamics underlying the relationship between these two culturally distinct groups and to evaluate the resulting biological impact on the genetic makeup of *Felsina*.

Moreover, we have recently added to the described dataset additional samples from the Po Valley dated to the migration period (V-II century BCE) and coming from sites (Monterenzio Vecchio and Monte Bibeale) with strong signal of Celtic cultural influence. Only preliminary results are shown in the present study for this second set of samples.

Starting from mainly petrous bones, we generated genome-wide data building double stranded and double indexed DNA libraries for Illumina platform and following an established in-solution capture protocol to enrich for ~1.24 million informative nuclear SNPs. After checking for preservation (% of endogenous DNA) and damage patterns, we inferred biological sex and estimated both nuclear and mitochondrial contamination. Moreover, we assigned mitochondrial (mtDNA) and Y chromosome haplogroups and investigated biological relatedness through kinship identification and the analyses of runs of homozygosity (ROH). Individuals showing a good coverage on the analysis panels and low contamination rates were considered suitable for downstream population genetics analyses which mainly involved principal component analysis (PCA), ADMIXTURE models, F-statistics ( $F_3$  and  $F_4$ ) and qpAdm admixture modelling.

The vast majority of the Etruscans samples from the Po Valley likely carries a local genetic profile which seems to resemble the one previously found in Etruria and, overall, in the rest of the Italian Peninsula during the Iron Age (and the Bronze Age), albeit additional radiocarbon dates are needed to confirm the time coordinates of the samples. Moreover, it is possible to observe the presence of a relevant fraction of individuals (~17%) who carries a nonlocal component associated to central or northern European populations (as demonstrated by F-statistics and admixture tests) and shows an extra affinity to steppe-related ancestry with respect to the local Etruscan cluster. Although it has not been possible to pinpoint the specific source of this nonlocal component so far, the observed signal is consistent with an incoming northern ancestry, in most of the cases probably admixed with a local source, which is found in all the examined necropolises of Bologna and in individuals dated from the VIII to the V-III century BCE. This suggests continuity in the source of this central-northern European ancestry which is consistent with deriving from contacts with Celtic-related groups.

Unexpectedly, almost all analysed individuals from the more recent necropolises of Monterenzio and Monte Bibeale, dated to the migration period, seem to carry a local genetic profile.

Individuals dated to the Roman Imperial period are genetically distinct from the previous Iron Age samples and show an increased affinity to Iranian-related ancestry accompanied by a close genetic link to ancient Caucasian populations (mainly from Armenia and Iran). Unlike to previously reported findings from central Italy, we do not find evidence for admixture with eastern Mediterranean and Levantine populations in the Po Valley samples.

Among the Imperial samples we found one individual carrying an ancestry of African origin, possibly sub-Saharan, supported by the assignment of the mitochondrial haplogroup L3e1.



The results described here, combined with further analyses that will be performed on the newly generated dataset, can provide new insights into the genetic history of Bologna and the Po Valley during the Etruscan period and add relevant information to address historical and archaeological questions on specific contexts.

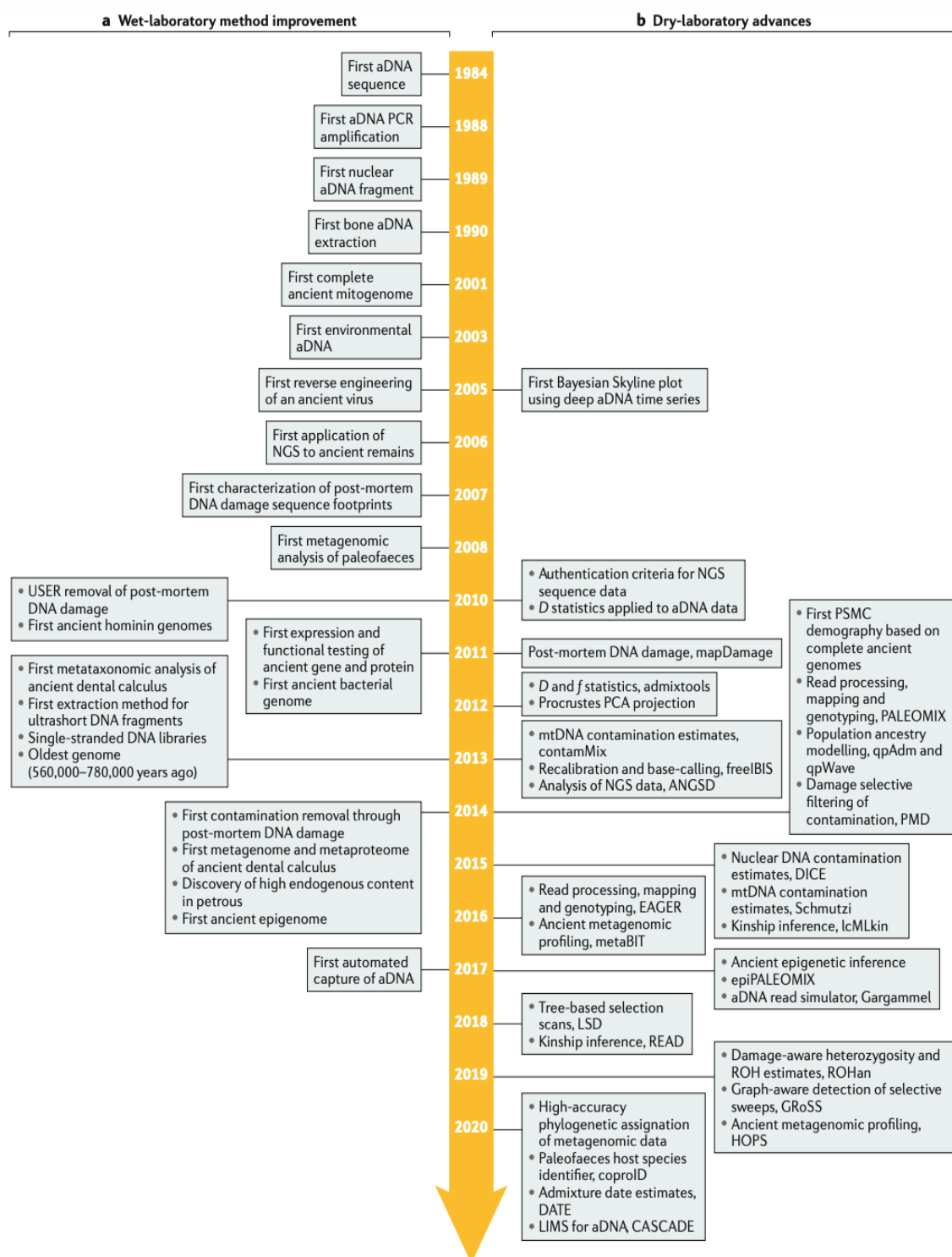
# 1. Introduzione: parte I – La paleogenetica

## 1.1 Dai primi passi al Nobel

Nel 1984, il recupero di corte sequenze di genoma mitocondriale (mtDNA) da un tessuto muscolare essiccato di un campione museale di quagga (*Equus quagga*), un equide affine alla zebra estintosi negli ultimi decenni del XIX secolo, ha dato avvio ad un primo sviluppo della ricerca nell'ambito del DNA antico (aDNA) [1]. Nonostante il dato ottenuto fosse limitato a circa 230 paia di basi (bp), il lavoro di Higuchi e collaboratori testò con successo la possibilità di acquisire informazioni genetiche dirette dai resti di un organismo vissuto più di 100 anni prima, mettendo per la prima volta in risalto le effettive potenzialità dal dato molecolare nel contribuire allo studio e alla comprensione di eventi passati. Pochi decenni dopo, l'età dei reperti biologici dai quale è stato possibile recuperare aDNA si è estesa notevolmente: i più antichi resti umani caratterizzati geneticamente, provenienti dal sito spagnolo di Sima de los Huesos e identificati verosimilmente come pre-Neandertal, sono datati a circa 430.000 anni fa (Pleistocene Medio) [2], mentre sono riconducibili ad oltre 1 milione di anni fa (Pleistocene Inferiore) i molar di due esemplari di mammut provenienti dalla Siberia nord-orientale che attualmente costituiscono i campioni più antichi dai quali è stato estratto materiale genetico endogeno [3]. Durante i primi due decenni del 2000 sono stati ricostruiti, inoltre, i genomi completi di numerose specie animali da tempo scomparse, quali il mammut lanoso (*Mammuthus primigenius*) [4] e l'orso delle caverne (*Ursus spelaeus*) [5], nonché di specie umane estinte come l'uomo di Neandertal [6-9] e di Denisova [10, 11].

Grazie alle informazioni estratte dal DNA antico, la nostra conoscenza e percezione della variabilità genetica non è rimasta limitata alla sola diversità osservabile nelle popolazioni moderne ma risulta arricchita da un dato temporale che conserva traccia degli eventi e dei cambiamenti genetici che nel corso della storia hanno interessato diversi gruppi umani, animali, vegetali e persino microbici. Il rapido sviluppo della paleogenetica e la significativa crescita del suo potenziale applicativo sono da associarsi, oltre che ad una più approfondita conoscenza dei meccanismi di degrado chimico-fisico che alterano la struttura molecolare dell'aDNA [12], ad innovazioni scientifiche rivoluzionarie. Accanto all'invenzione della PCR (Polymerase Chain Reaction, avvenuta nel 1986 ad opera di Kary Mullis) [13], che ha offerto l'opportunità di amplificare esigui quantitativi di molecole di DNA *target* provenienti da resti antichi fino ad un livello compatibile con le successive analisi [14], l'avvento delle tecnologie di sequenziamento di nuova generazione (New Generation Sequencing, NGS) agli inizi del XXI secolo ha rappresentato il momento di massima trasformazione nel settore dell'aDNA,

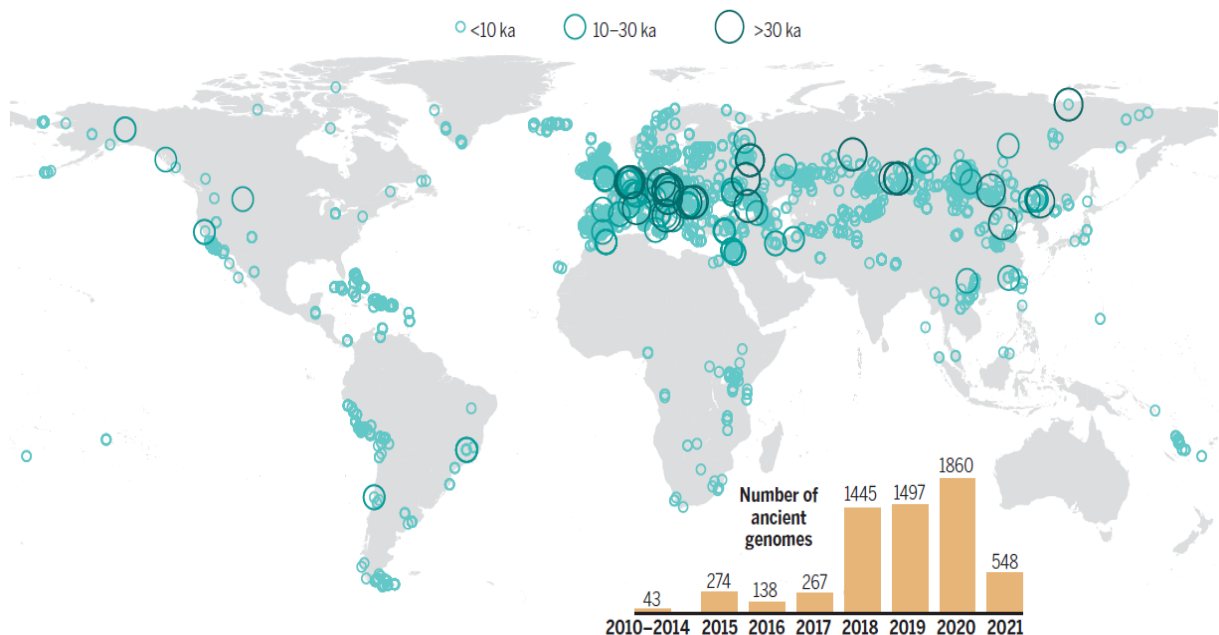
determinando cambiamenti radicali sia dal punto di vista delle metodologie di laboratorio (*wet lab*) che delle strategie di analisi del dato (*dry lab*) [15] (Figura 1.1).



**Figura 1.1** Principali innovazioni nell'ambito delle metodologie di laboratorio (parte a) e dell'analisi del dato bioinformatico (parte b); (immagine estratta da [15]).

Il successo dell'approccio NGS rispetto all'utilizzo di metodi basati sulla PCR deriva dalla possibilità di sequenziare contemporaneamente milioni di loci a partire da quantitativi ridotti di frammenti di DNA di corte dimensioni (anche inferiori a 50 bp), nonché di recuperare ed analizzare tali molecole nella loro interezza, consentendo di discriminare l'informazione endogena da quella contaminante sulla base dell'identificazione di specifici *patterns* di degrado che contraddistinguono il DNA antico (come l'incremento del tasso di deaminazioni all'estremità dei filamenti ed una ridotta lunghezza media delle sequenze ottenute; vedi *paragrafo 1.2*). Questi aspetti hanno permesso di affrontare con successo alcune delle principali problematiche connesse alla natura del DNA antico (quali scarsa concentrazione, danni strutturali, contaminazione ambientale ed elevato rischio di contaminazione con DNA moderno) che nell'era pre-NGS avevano rappresentato dei fattori limitanti all'ottenimento di dati autentici e riproducibili e, dunque, al progredire della disciplina stessa.

Lo sviluppo delle tecniche NGS ha facilitato, inoltre, il moltiplicarsi di studi su scala genomica (*Figura 1.2*), rendendo accessibile la produzione e l'analisi di nuovi dati in grado di arricchire e completare l'informazione molecolare restituita da ristrette porzioni di DNA nucleare o dalla caratterizzazione dei soli marcatori uniparentali (in particolare l'mtDNA), dati sui quali si basava la maggior parte dei precedenti studi.



**Figura 1.2** Mappa dei genomi antichi pubblicati, aggiornata a giugno 2021 (sono compresi genomi completi, esomi e dati genomici derivanti dalla cattura di SNPs nucleari). I cerchi grandi di colore blu scuro indicano gli individui datati a più di 30.000 anni fa; i cerchi blu di medie dimensioni corrispondono agli individui più recenti di 30.000 anni fa; i cerchi piccoli blu chiaro indicano, invece, individui più recenti di 10.000 anni fa (immagine estratta da [39]).

Il delinarsi di questa tendenza trova origine anche nel frequente impiego di soluzioni di *target enrichment*, ovvero di specifici protocolli per l'arricchimento di determinate regioni genomiche che consentono di generare selettivamente dati di sequenza a livello genico [16], cromosomico [17], mitocondriale [18] o genomico [19]. Tale approccio, oltre a fornire un'alternativa economicamente vantaggiosa al sequenziamento *shotgun* di genomi completi (in particolare nel caso di campioni scarsamente preservati che richiederebbero un'elevata profondità di sequenziamento), consente di catturare ed arricchire una precisa frazione endogena di interesse, minimizzando in questo modo l'incidenza nell'*output* di sequenziamento di materiale contaminante (solitamente in percentuale preponderante) o non rilevante agli scopi dello studio. Alcuni progetti di genetica di popolazione, ad esempio, sfruttano apposite sequenze oligonucleotidiche (sonde) in soluzione per effettuare la cattura di specifiche posizioni a singolo polimorfismo (SNPs) presenti nel genoma nucleare umano e identificate come informative per l'*ancestry* [20].

Oltre ai progressi appena riportati, hanno rivestito un ruolo fondamentale anche l'implementazione dei metodi di estrazione [21-24] e di preparazione delle librerie genomiche [10, 25, 26], unitamente all'individuazione di specifici distretti scheletrici che, per le loro caratteristiche intrinseche e la loro posizione anatomica, favoriscono una migliore preservazione del materiale molecolare e, quindi, una maggiore resa in termini quantitativi e qualitativi [27-30].

Anche il *workflow* di analisi del dato antico è stato oggetto di forti trasformazioni in concomitanza alla diffusione delle tecniche NGS, tra le quali l'introduzione di strumenti statistici in grado di gestire i problemi legati alla contaminazione [31-33] ed alle alterazioni post-mortem tipiche dell'aDNA [34], ma anche di rilevare sottili cambiamenti nella struttura di una popolazione imputabili al verificarsi in passato di possibili eventi di deriva genetica, *admixture* o selezione naturale [35].

I promettenti risultati ottenuti dalla paleogenetica negli ultimi decenni hanno messo in risalto le potenzialità del dato molecolare antico nel rispondere a molteplici quesiti in ambito evolutivo, storico, antropologico ed archeologico. Il processo di crescita ed innovazione è tuttora in corso e in costante accelerazione, favorito da un più esteso impiego nel *wet lab* di procedure di automazione (soprattutto per la costruzione delle librerie ed il *target enrichment*) [26, 36, 37], dall'individuazione di nuovi materiali come valide fonti di informazioni molecolari (coproliti, *dental calculus*, sedimenti, archivi bioculturali come la pergamena e i materiali tessili) [15] e dalla sperimentazione in ambito computazionale di strumenti sempre più potenti (come l'applicazione di metodi di imputazione ad integrazione del frammentario dato genetico antico [38]).

La ricerca di nuovi materiali di partenza per lo studio dell'aDNA risiede anche nel carattere invasivo delle analisi paleogenetiche, fattore limitante in presenza di resti di particolare unicità o rilevanza storico-antropologica (che rientrano, come tutti i reperti scheletrici, nella sfera dei beni culturali) e

nel caso in cui l'integrità dell'elemento da campionare sia di fondamentale interesse anche per altre aree di studio. In alcuni casi può rendersi necessario, quindi, l'impiego di metodi di campionamento o estrazione minimamente invasivi [39, 40] e lo sfruttamento di un quantitativo minimo di materiale di partenza per la conduzione di più analisi in parallelo [41].

Sfide e prospettive future riguardano aspetti di varia natura, tra cui l'acquisizione di dati antichi da regioni geografiche ancora poco esplorate dal punto di vista genetico, come l'Africa e l'Oceania, (anche a causa delle condizioni ambientali poco favorevoli alla conservazione del DNA) [42] o da transetti temporali ad oggi sottorappresentati [43]; prevedono, inoltre, il passaggio da studi in prevalenza su larga scala a progetti focalizzati su contesti locali o sito-specifici [15, 43], con una maggiore attenzione verso le questioni etiche ed una più sistematica integrazione dell'informazione genetica con dati isotopici, proteomici, antropologici ed archeologici, seguendo un approccio di tipo multidisciplinare volto ad una descrizione più esauriente delle popolazioni in esame e delle loro condizioni di vita [15, 42, 43].

A coronamento dei primi 40 di ricerca sul DNA antico, è doveroso citare il recente conferimento del premio Nobel per la Medicina 2022 al biologo e genetista Svante Pääbo per le sue scoperte sulla storia evolutiva dell'uomo e per il suo contributo al progresso degli studi paleogenetici [44]. Un prestigioso riconoscimento alla sua carriera personale ma anche agli ambiziosi successi raggiunti in pochi decenni dalla paleogenetica.

## 1.2 Gli aspetti critici del DNA antico

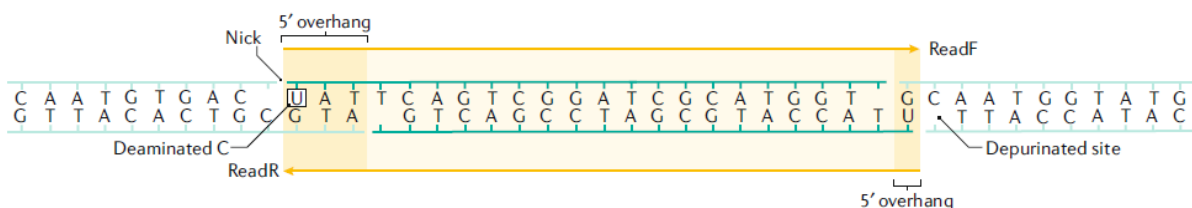
Come dimostrato dagli esempi riportati nel precedente paragrafo, in presenza di condizioni favorevoli, il DNA può conservarsi per migliaia di anni all'interno dei resti di un organismo dopo la sua morte. Durante questo lasso temporale il materiale genetico è però soggetto a fenomeni post-mortem di natura autolitica (dovuti all'attività enzimatica delle nucleasi cellulari) e diagenetica che determinano la comparsa di caratteristici *patterns* di degrado [12], la cui incidenza è indicativa della storia tafonomica dei resti in esame e, in alcuni casi, della loro antichità [45].

A differenza di quanto accade *in vivo*, i danni subiti dall'adna permangono nel tempo a causa della mancanza di meccanismi di riparazione enzimatica. I processi degenerativi, principalmente causati da reazioni di idrolisi ed ossidazione, sono influenzati dai parametri ambientali propri del contesto di deposizione dei resti, come pH, temperatura, umidità, presenza di ossigeno ed esposizione alle radiazioni UV. Inoltre, ulteriori fattori quali l'attività microbica e lo stress meccanico indotto dalle stratificazioni sedimentarie influiscono sulla possibilità di preservazione del materiale genetico.

I fenomeni di degrado del DNA antico, che si manifestano a livello di alterazioni chimico-fisiche della sua struttura molecolare, sono alla base della ridotta concentrazione e della limitata lunghezza delle molecole di DNA, che a loro volta si traducono in una perdita di materiale informativo.

Il processo di frammentazione dell'aDNA è dovuto principalmente alla comparsa di siti abasici lungo le sequenze nucleotidiche, evento causato dall'idrolisi dei legami N-glicosidici sussistenti tra le basi azotate e lo zucchero. Tale processo degenerativo viene definito depurinazione dal momento che coinvolge con una maggior frequenza le basi puriniche (in particolare la guanidina). L'indebolimento del doppio filamento in corrispondenza di un sito abasico può risolversi nella rottura dello stesso, con conseguente frammentazione della molecola e formazione di estremità sporgenti (Figura 1.3). In corrispondenza di quest'ultimi, si verifica con maggiore incidenza il processo di deaminazione, la forma più comune di degrado chimico per idrolisi a cui è soggetto l'aDNA. Tale reazione interessa nello specifico i gruppi amminici secondari che caratterizzano la struttura molecolare della citosina, causando la sua conversione in uracile [46] (Figura 1.3). In presenza di questa tipologia di lesioni, durante le fasi di amplificazione del DNA, si verificano delle transizioni  $C \rightarrow T$  e  $G \rightarrow A$  (quest'ultime lungo il filamento complementare) dovute all'errata lettura della base azotata danneggiata [47]; in particolare, i residui di deossi-uridina causano l'incorporazione di una deossi-adenosina invece di una deossi-guanosina nel corso del processo di ricostruzione del filamento complementare da parte della DNA polimerasi. Il tasso di sostituzioni  $C \rightarrow T$  e  $G \rightarrow A$  presenta un andamento specifico: tende ad aumentare rispettivamente verso le estremità 5' e 3' dei filamenti [48] come conseguenza del fatto che le estremità, spesso ridotte a singolo filamento, risultano maggiormente esposte (rispetto alle porzioni centrali) all'attività diagenetica.

Come già accennato, l'aDNA può riportare anche lesioni di natura ossidativa, in seguito alle quali si riscontrano ulteriori modificazioni nucleotidiche che, come nel caso precedente, possono condurre ad un'errata ricostruzione della sequenza (o ad un arresto dell'estensione della molecola durante i processi replicativi) [12].



**Figura 1.3** Schematizzazione di una molecola di DNA antico e delle tipiche alterazioni che la caratterizzano. Il fenomeno di depurinazione determina la frammentazione della molecola con conseguente formazione di estremità sporgenti, in corrispondenza delle quali si verificano con maggiore probabilità reazioni di deaminazione (immagine estratta da [15]).

L'ottenimento di sequenze che non rispecchiano l'informazione originale può condurre a stime alterate di parametri quali il grado di eterozigosi, la variabilità nucleotidica, il contenuto in GC ed il tempo di divergenza.

Le caratteristiche tipiche dell'aDNA implicano l'adozione di specifiche accortezze durante le fasi di analisi sperimentale, le quali sono improntate alla massimizzazione del recupero di molecole di DNA endogeno. Oltre alla necessità di isolare l'esigua frazione endogena dalla maggioritaria componente esogena, principalmente di natura microbica, risulta di fondamentale importanza minimizzare il rischio di contaminazione con DNA moderno. A tale scopo si rendono necessari alcuni accorgimenti, espressi nei *Golden Criteria* [49], tra i quali le norme di vestiario e le procedure comportamentali a cui gli operatori devono attenersi, nonché una precisa organizzazione del flusso di lavoro ed una suddivisione fisica del laboratorio in spazi appositamente adibiti ai differenti *step* di analisi (aree a contaminazione controllata pre-amplificazione ed aree post-amplificazione). In questo contesto risulta essenziale anche la possibilità di tracciare eventuali episodi di contaminazione avuti luogo non solo negli spazi di laboratorio ma anche nei luoghi di provenienza o conservazione dei resti (per esempio siti archeologici o musei) [50].

Per monitorare la presenza di potenziali contaminanti negli spazi di laboratorio o in relazione all'uso di determinati reagenti, nel corso delle fasi sperimentali vengono inseriti e processati in parallelo ai campioni dei controlli negativi. Gli ambienti e gli strumenti di lavoro sono abitualmente interessati da specifici trattamenti di pulizia e sterilizzazione (con apposite sostanze chimiche e irraggiamento UV) volti all'eliminazione o all'inattivazione di biomolecole potenzialmente contaminanti. Durante gli stadi iniziali del *workflow* di analisi, i campioni stessi vengono sottoposti ad opportuni processi di pulizia meccanica e decontaminazione mirati a ridurre l'impatto della componente esogena e a minimizzare l'influenza di sostanze inibitrici (come acidi umidi e tannini del suolo) che potrebbero ostacolare l'efficienza dei successivi *step* di reazione.

Nonostante le peculiarità del DNA antico rappresentino un'inevitabile fonte di criticità per le analisi molecolari, la loro identificazione fornisce, tuttavia, un criterio di autenticazione del dato finale, consentendo di discriminare il DNA endogeno da eventuali contaminanti moderni.



### 1.3 I diversi campi di applicazione e il loro potenziale informativo

L'ambito di analisi che maggiormente ha tratto beneficio dall'introduzione delle nuove tecnologie biomolecolari, oltre all'archeologia microbica [51, 52], è la genetica di popolazione umana.

La ricchezza di genomi umani antichi disponibili in letteratura ha avuto un ruolo fondamentale nel trasformare la conoscenza del nostro passato: l'iniziale ipotesi di prolungata continuità genetica ed isolamento tra popolazioni ha lasciato spazio ad una visione decisamente più dinamica della storia umana, in cui mobilità ed eventi di flusso genico hanno avuto una parte predominante [53].

Nonostante il dato genetico relativo alle popolazioni pre-Olocene sia ancora molto frammentario, a causa della scarsità di materiale fossile e della limitata preservazione del DNA antico, sostanziali progressi sono stati compiuti in seguito all'ottenimento di dati genomici nucleari da gruppi umani pre-Neandertal [2], individui afferenti alle linee di Neandertal e Denisova [54, 55], nonché esemplari di *Homo sapiens* datati al Paleolitico Superiore [17, 20, 56-61].

L'Europa, e più in generale l'Eurasia occidentale, sono state finora il focus principale della maggior parte dei lavori di genetica di popolazione [43]; di recente, però, numerosi studi si sono orientati verso la scoperta della storia genomica di altre aree geografiche quali l'Africa subsahariana [62, 63], l'America [64], l'Oceania [65], l'Asia orientale e sud-orientale [66-68], dimostrando come l'adDNA possa conservarsi anche in zone tropicali o caratterizzate da condizioni ambientali poco favorevoli. Nei contesti archeologici, il dato molecolare antico ha inoltre il potenziale di fornire informazioni su questioni puntuali come l'identificazione delle relazioni biologiche tra individui di uno stesso sito o necropoli, consentendo di rintracciare in questo modo le pratiche culturali e le caratteristiche della struttura sociale delle popolazioni in questione [69-71].

Oltre alla ricostruzione della storia dell'uomo, la ricerca si è focalizzata anche sulla comprensione della biologia umana: un recente lavoro mostra come le principali varianti genetiche associate ad un elevato rischio di complicanze gravi nell'uomo in seguito alla contrazione della malattia COVID-19 potrebbero essere state ereditate da Neandertal [72]. Con gli avanzamenti nel campo delle tecniche di *gene-editing*, prossimi studi potranno anche direzionarsi verso la caratterizzazione di varianti adattative individuate in gruppi antichi al fine di comprendere come l'evoluzione abbia dato forma all'odierna struttura genomica umana [15].

Un diverso campo applicativo riguarda l'analisi di microorganismi patogeni e del microbioma antico. Insieme ai più tradizionali approcci paleopatologici e paleodemografici, il DNA antico derivante da microorganismi (sia patogeni che commensali) può fornire rilevanti informazioni circa lo stato di salute di popolazioni antiche, la diffusione di epidemie nel passato, l'interazione tra agenti patogeni ed organismo ospite, la composizione della dieta ed eventuali cambiamenti legati a quest'ultima.

Se i metodi basati sulla PCR avevano inizialmente permesso di identificare in modo mirato patogeni che causano lesioni visibili sui resti scheletrici (come *Mycobacterium tuberculosis* e *Mycobacterium leprae*) [73] o altri specifici microorganismi di interesse, le tecniche NGS hanno offerto l'opportunità di individuare simultaneamente molteplici segnali microbici/virali (anche di agenti patogeni che non lasciano evidenze sul tessuto osseo come *Yersinia pestis* e *Salmonella enterica*) [51], di ricostruire il profilo composizionale e funzionale del microbioma antico a partire da campioni quali sedimenti di latrine, coproliti o *dental calculus* [74-76], e di valutare l'autenticità del dato ottenuto.

Anche il tema della domesticazione di piante e animali nel passato è stato in più occasioni affrontato sotto l'aspetto genetico; tale approccio ha consentito di rivelare una storia di complesse dinamiche popolazionistiche che hanno rivelato eventi di completo rimpiazzamento, introgressione adattativa tra forme addomesticate e selvatiche, inaspettati *patterns* geografici e temporali circa l'origine e la selezione di alcune specie ed un ampio ruolo dei processi di dispersione naturale [77, 78]. Oltre a ciò, l'analisi dell'aDNA si è occupata anche di indagare le relazioni filogenetiche tra specie viventi e specie estinte, nonché di identificare le potenziali cause della scomparsa di quest'ultime andando a valutare possibili attività/pressioni antropiche o cambiamenti climatici riconducibili al medesimo periodo temporale [79]. Ad esempio, la frequentazione delle grotte da parte dell'uomo nel corso dell'Aurignaziano ha inciso negativamente sulle dinamiche di vita dell'orso delle caverne, le quali si basavano sullo sfruttamento dello stesso habitat per il letargo invernale [80, 81]. Analisi incentrate sulla teoria della coalescenza hanno permesso di scoprire, inoltre, come alcuni gruppi umani antichi siano riusciti a sopravvivere per lungo tempo all'estinzione malgrado la ridotta dimensione effettiva della loro popolazione [7, 8, 61].

Infine, un'area applicativa particolarmente innovativa è quella rappresentata dallo studio del DNA ambientale o sedimentario (dall'inglese eDNA o sedDNA). Basandosi su strategie metagenomiche e di arricchimento mirato, l'aDNA estratto da sedimenti di diversa natura (anche lacustri, marini o di ghiaccio) favorisce non solo l'accesso ad un'ampia variabilità molecolare, dalle specie microbiche a quelle vertebrate, anche in assenza di macro-fossili [82], ma permette inoltre di caratterizzare i differenti livelli di un campione stratigrafico; ciò fornisce rilevanti informazioni circa la distribuzione spaziale e temporale delle specie riscontrate e l'impatto ambientale di cambiamenti climatici su scala globale [37, 83]. In aggiunta, l'utilizzo di campioni sedimentari ha meno implicazioni sotto l'aspetto della distruttività delle analisi ed offre un'informazione genetica comprensiva dell'intera diversità tassonomica dell'ecosistema antico in esame. Possibili limiti riguardano la specificità della determinazione tassonomica, che risulta influenzata dall'esteso grado di frammentazione del DNA ambientale e dal livello di accuratezza dei database comparativi a disposizione [15].

## 2. Introduzione: parte II – Storia genetica dell'Europa dal Neolitico all'età del Ferro

L'assetto genetico delle attuali popolazioni europee è stato plasmato nel corso del tempo da diversi fenomeni migratori ed eventi di flusso genico che hanno avuto un forte impatto popolazionistico, culturale e linguistico sui gruppi umani antichi. Questi eventi comprendono il primo popolamento dell'Europa, la transizione neolitica e le successive migrazioni dell'età del Bronzo [84].

A questi specifici avvenimenti sono correlate le tre principali componenti ancestrali che descrivono la variabilità genetica delle moderne popolazioni europee [85]; tali componenti corrispondono al segnale genetico caratteristico dei primi cacciatori-raccoglitori dell'Europa occidentale (WHG da *western hunter-gatherers*) [86], a quello introdotto in Europa da popolazioni strettamente correlate agli agricoltori-allevatori neolitici dell'Anatolia (attuale Turchia) [87] e, infine, all'*ancestry* associata alla migrazione di popoli di pastori afferenti alla cultura Yamnaya, originari delle steppe pontico-caspiche (~5000 a.C.) [88].

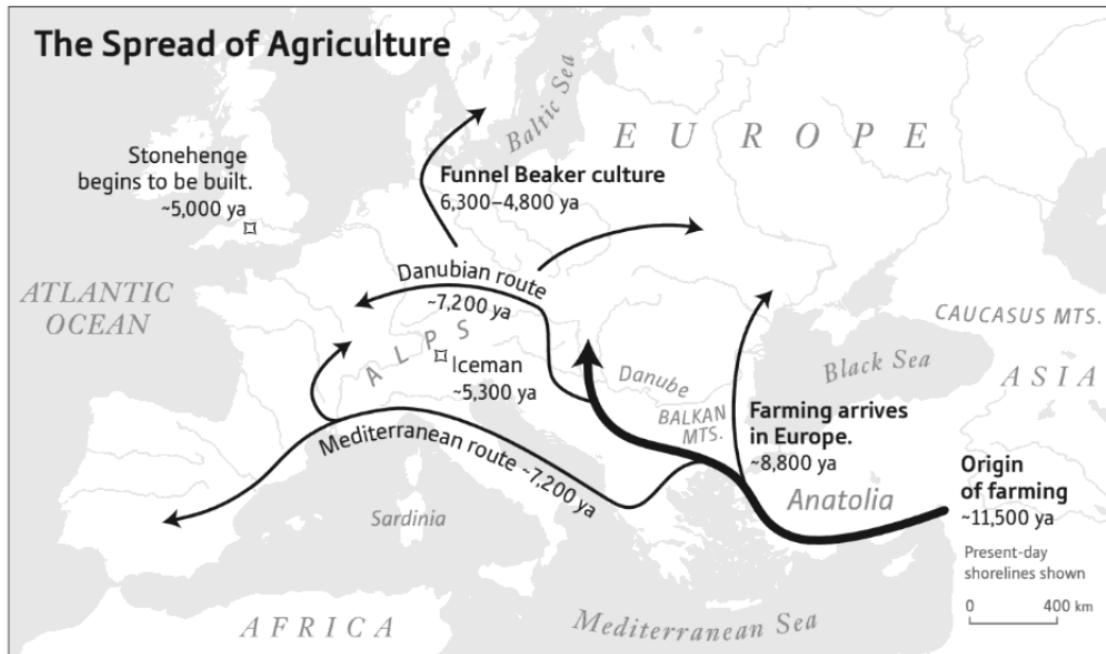
La rivoluzione neolitica ha determinato la transizione da uno stile di vita nomade, caratteristico dei gruppi di cacciatori-raccoglitori, ad una condizione di sedentarietà strettamente legata allo sviluppo e alla diffusione delle pratiche agricole e della domesticazione degli animali.

L'origine dello stile di vita neolitico è riconducibile al Vicino Oriente, in particolare alla regione della Mezzaluna Fertile (antica Mesopotamia), e risale a circa 12.000-11.500 anni fa [84, 89]. Un recente studio condotto da Lazaridis *et al.*, identifica almeno due distinti eventi di mobilità dalla zona della Mesopotamia e del Levante verso l'Anatolia, rispettivamente durante la fase neolitica pre-ceramica (~10.000-6500 a.C.) e ceramica (~7000 a.C.), i quali avrebbero contribuito ad accelerare lo sviluppo, con tutta probabilità già in corso, delle innovazioni neolitiche nella regione [90].

Nell'intervallo compreso tra ~8800-6000 anni fa, le tecniche neolitiche vengono diffuse in tutto il territorio europeo da popolazioni di agricoltori e allevatori connessi al contesto anatolico, i quali costituiscono la fonte primaria della componente genetica *Anatolia Neolithic*, la quale caratterizza il *pool* genico degli europei [84]. Tale componente genetica presenta, infatti, un'elevata affinità con le attuali popolazioni dell'Europa sud-occidentale, mentre si differenzia in maniera significativa da quelle del Vicino Oriente; inoltre, risulta massimizzata nei popoli sardi e baschi, probabilmente a seguito di condizioni di isolamento che hanno reso meno incisivo l'impatto dei successivi eventi demografici in queste regioni [84].

La rivoluzione neolitica è giunta in Europa attraverso due rotte principali, ben documentate a livello archeologico (*Figura 2.1*): una lungo le coste del Mediterraneo verso la penisola iberica (delineata

dalle culture della ceramica impressa e della ceramica cardiale), l'altra in direzione nord-ovest verso l'Ungheria e l'Europa centrale, seguendo le valli del Danubio e del Reno (rappresentata dalla cultura della ceramica lineare, LBK da *Linearbandkeramik culture*) [91].



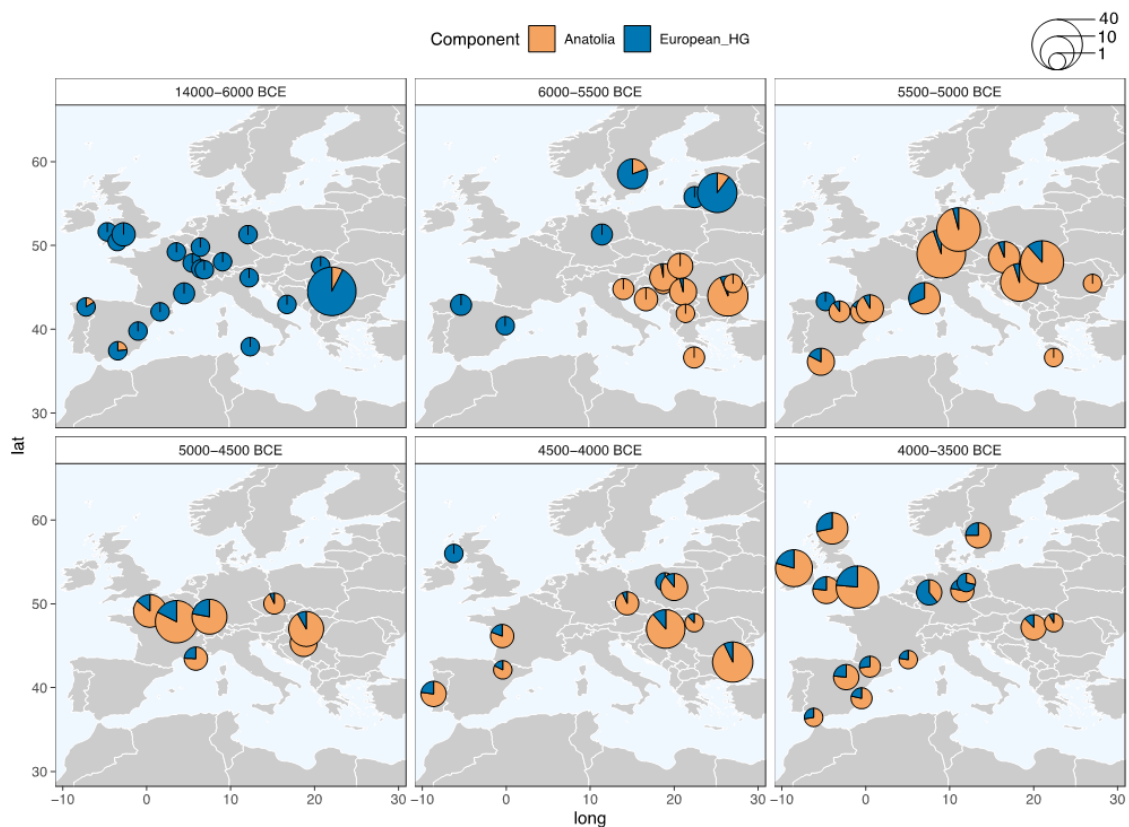
**Figura 2.1** Schematizzazione delle due rotte migratorie principali (lungo le coste del Mediterraneo e lungo la valle del Danubio) che a partire da circa 8800 anni fa hanno determinato la diffusione dello stile di vita neolitico nel continente Europeo (immagine estratta da [89]).

Se studi su larga scala, basati sull'analisi di genomi antichi, hanno confermato l'ipotesi del modello demico (inizialmente in contrapposizione a quello di diffusione culturale) [28, 87, 88, 92-96], focus locali hanno rivelato complessi processi di *admixture*, variabili a livello regionale, tra le popolazioni agricole in arrivo e i gruppi indigeni di cacciatori-raccoglitori [87, 93, 97]. Le evidenze genetiche raccolte lungo la rotta continentale (dal sud-est fino al centro Europa) supportano l'ipotesi di una rapida espansione, inizialmente caratterizzata da ridotte interazioni biologiche tra gruppi migranti e locali, soprattutto nel periodo della cultura LBK (~5500-4500 a.C.), e da una successiva fase di coesistenza e scambi culturali durata più di un millennio. Anche lungo la rotta mediterranea si osservano dei *patterns* simili, con la comparsa delle prime evidenze di mescolanza genica solamente dopo almeno un millennio dallo stanziamento dei primi agricoltori-allevatori [98].

Lo scenario generale è caratterizzato, dunque, da un limitato flusso genico tra i gruppi indigeni e migranti durante la prima fase del Neolitico e da un aumento del segnale genetico degli HGs europei in popolazioni associate ad uno stile di vita agricolo soltanto in uno stadio più avanzato (fenomeno delle "rinascita mesolitica", in parte dovuto ad un probabile indebolimento del processo migratorio;

Figura 2.2). Un'eccezione all'andamento generale è riportata nello studio di Rivollat *et al.* [98], nel quale viene osservata un'elevata percentuale di componente HG nei resti di agricoltori neolitici provenienti da siti francesi, in particolare dall'area meridionale, e risalenti ad una fase iniziale del Neolitico (31% in media, contro il 3% osservato in Europa centrale ed il 13% nella penisola iberica [99]; Figura 2.2). È importante sottolineare che l'attuale Francia costituisce il punto di incontro delle due rotte di espansione neolitica.

Tale studio ha permesso anche di evidenziare alcune dinamiche locali nelle interazioni biologiche e culturali tra le popolazioni mesolitiche e neolitiche, che risultano ben distinte lungo le due rotte: in Europa centrale gli agricoltori neolitici presentano un'esigua frazione di componente HG, giunta probabilmente in forma già mescolata in seguito alle prime fasi di espansione nel sud-est europeo; ad ovest del Reno, invece, la componente HG deriva direttamente dai gruppi mesolitici locali. Questi elementi confermano la rapida diffusione dei gruppi neolitici lungo la rotta continentale, avvenuta senza significative interazioni con i cacciatori-raccoglitori del luogo, mentre suggeriscono un tardivo processo di *admixture* locale nei territori della Francia, Spagna e Gran Bretagna [98].



**Figura 2.2** Serie di mappe raffiguranti la variazione nel corso del tempo delle proporzioni di componente *Anatolia Neolithic* (arancione) e HG europea (blu) negli individui esaminati in [99]. La variabilità osservata regionalmente e temporalmente delinea la complessità delle interazioni biologiche in corso durante l'espansione neolitica in Europa (immagine estratta da [99]).

In Gran Bretagna la cultura neolitica compare intorno al 4000 a.C., un millennio dopo rispetto alle vicine regioni dell'Europa del nord. Anche in questo caso l'evento coincide con un'ampia diffusione della componente anatolica che, da recenti studi genomici, sembra essere stata introdotta dalla migrazione di popolazioni neolitiche correlate alla rotta mediterranea [98, 100]. Ciò è supportato anche dalle relativamente alte percentuali della componente HG europea riscontrate negli individui neolitici della regione (*Figura 2.2*) [98].

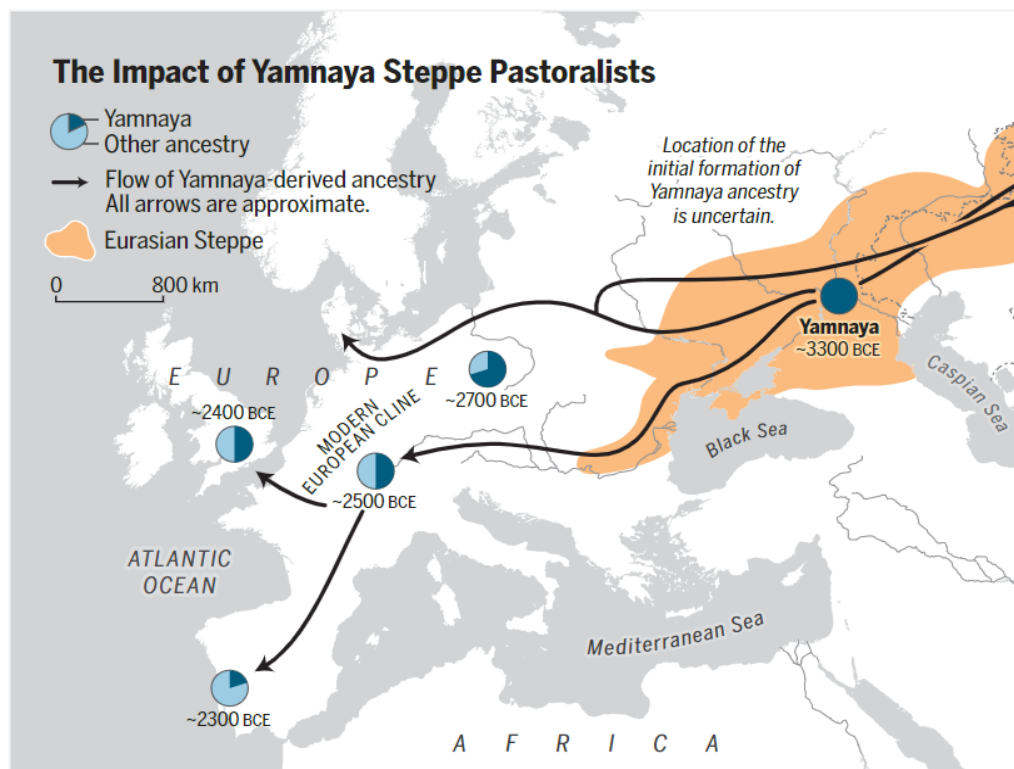
Come conseguenza delle dinamiche appena descritte, l'Europa raggiunge un nuovo equilibrio: il profilo genetico associato ai cacciatori-raccoglitori europei sopravvive solo in mescolanza a quello degli agricoltori-allevatori neolitici, ad eccezione di aree geografiche molto limitate come le isole al largo della Svezia meridionale [99].

Il cambiamento a livello popolazionistico che si osserva nel continente europeo con la rivoluzione neolitica non è l'ultimo ad aver profondamente influenzato l'assetto genetico dell'Europa antica. Oltre alla ricolonizzazione del continente da parte degli HGs europei dopo il *Last Glacial Maximum* (LGM, ~19.000 anni fa) e alla transizione neolitica, le migrazioni del tardo Neolitico/prima età del Bronzo dall'Eurasia orientale costituiscono il terzo principale evento che ha contribuito a definire il profilo genetico dei moderni europei e i diversi gradienti di variabilità che lo caratterizzano [84].

A partire da 5000 anni fa, le culture neolitiche dell'area temperata dell'Europa orientale vengono rapidamente sostituite dalla diffusione della cultura Yamnaya, originaria delle steppe situate a nord del Mar Nero e del Mar Caspio; questa cultura deve almeno in parte il suo successo all'impiego del carro con ruote (forse adottato dall'adiacente cultura Maikop) e del cavallo, innovazioni che hanno favorito uno sfruttamento più efficiente delle risorse della steppa. Oltre alla migrazione verso ovest, evidenza di un'ulteriore espansione degli Yamnaya anche nella direzione opposta è rappresentata dal profilo genetico di individui della regione dei monti Altai-Saiani associati alla cultura Afanasievo (della prima età del Bronzo), il quale risulta indistinguibile da quello Yamnaya [101].

La migrazione verso occidente è ben documentata dal punto di vista archeologico ed è connessa alla diffusione in Europa di una componente genetica di derivazione caucasica, la cosiddetta *steppe-related ancestry* (*Figura 2.3*) [88, 101]. La struttura genetica della popolazione Yamnaya può essere descritta, infatti, come una combinazione di due componenti principali in proporzioni simili tra loro: una derivante da un gruppo geneticamente simile ai cacciatori-raccoglitori del Caucaso (*Caucasus hunter-gatherers*, CHG) e una associata agli HGs dell'Europa orientale (*Eastern hunter-gatherers*, EHG) [102, 103]. L'eredità EHG che caratterizza la popolazione Yamnaya sembra essere alla base dell'affinità genetica esistente tra i moderni europei e l'*Ancient North European ancestry* (ANE), la cui miglior *proxy* è rappresentata da un campione siberiano di 24.000 ritrovato nei pressi del sito di Mal'ta [84, 88, 104]. Questa componente genetica, in precedenza assente nei territori dell'Europa

centrale ed occidentale, inizia a comparire in concomitanza alle migrazioni dall'area delle steppe. Alcuni studi suggeriscono il ruolo determinante di questo evento migratorio nello sviluppo della cultura della ceramica cordata (*Corded Ware*) nelle zone dell'Europa centrale e nella propagazione delle lingue indoeuropee [88, 101, 105]. Haak e colleghi stimano che l'espansione della popolazione Yamnaya abbia comportato una sostituzione di ~75% dell'ancestry che in precedenza caratterizzava l'Europa centrale [88]; si registra, invece, una graduale diminuzione di tale impatto nei territori situati più ad occidente (*Figura 2.3*).



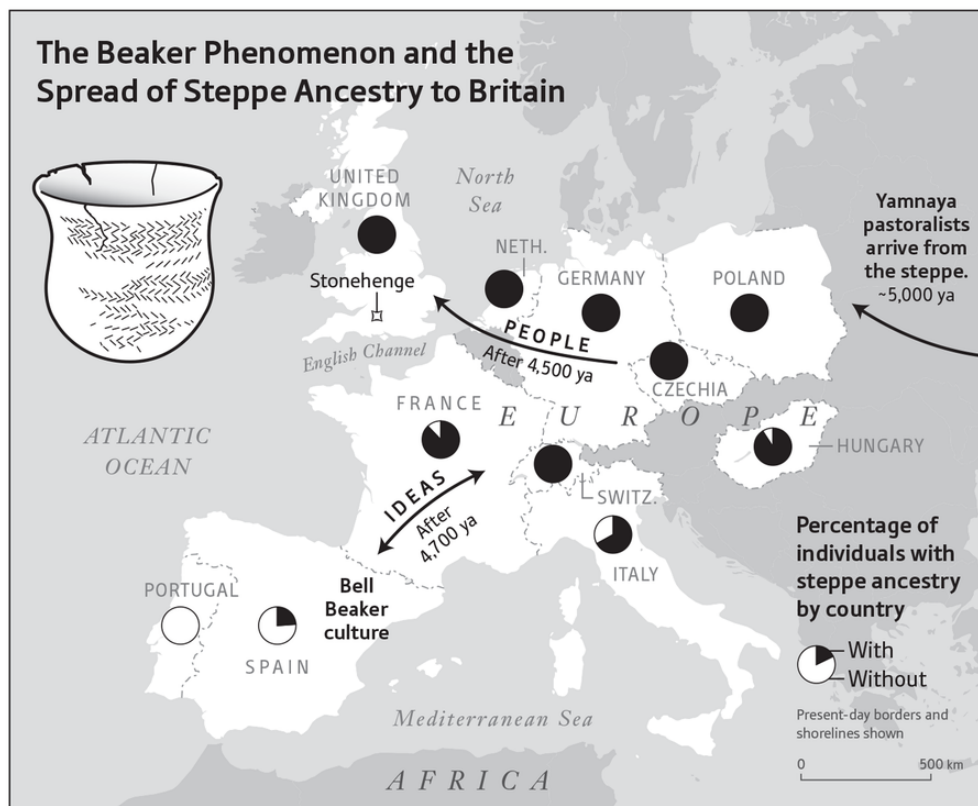
**Figura 2.3** Schematizzazione della migrazione verso occidente del popolo Yamnaya durante la prima età del Bronzo. I grafici a torta rappresentano una stima della proporzione di componente Yamnaya nelle varie regioni, mentre le date si riferiscono al genoma più antico in cui è stata riscontrata tale *ancestry* (immagine modificata da [105]).

Approssimativamente 4700 anni fa, un altro importante evento interessa il continente europeo: la diffusione della cultura del vaso campaniforme (*Bell Beaker*), probabilmente a partire dalla penisola iberica. Se l'arrivo della cultura *Bell Beaker* in Europa centrale sembra essere stato determinato da un movimento di idee piuttosto che di persone, confermato dalla diversità del profilo genetico di individui associati al tale cultura e provenienti rispettivamente dalla penisola iberica e dall'Europa centrale, la sua diffusione in Gran Bretagna appare legata a dinamiche del tutto differenti [106].

La componente genetica della steppa, assente nell'isola prima della diffusione della cultura della ceramica campaniforme, inizia a comparire in larga percentuale nel periodo successivo a 4500 anni fa. Questa evidenza, accompagnata dall'assenza di una particolare affinità genetica con la penisola iberica, ha permesso di ipotizzare un'importante migrazione di gruppi con elevata *steppe ancestry* e di cultura *Bell Beaker* dall'Europa centrale, la quale avrebbe determinato la sostituzione di circa il 90% del precedente *pool* genico (Figura 2.4) [106].

Nonostante le culture presenti in Europa centrale durante il tardo Neolitico e l'età del Bronzo siano geneticamente molto simili tra loro, è possibile comunque notare un gradiente di affinità genetica con gli Yamnaya: i valori più alti si riscontrano in campioni *Corded Ware* provenienti dalla Germania, i più bassi in individui ungheresi di varie culture (per esempio Vatyá e Maros), mentre la posizione intermedia è occupata da campioni *Bell Beaker* sempre provenienti dalla Germania [91, 101].

Durante l'età del Bronzo si osserva un declino generale della componente Yamnaya a favore di un nuovo incremento delle componenti WHG e *Anatolia Neolithic*; si è probabilmente trattato, quindi, di un evento migratorio confinato nel tempo piuttosto che di un costante flusso genico dalle steppe [88, 91]. Oggi la componente della steppa caratterizza in percentuale più elevata le popolazioni dell'Europa settentrionale rispetto a quelle dell'area meridionale [88].



**Figura 2.4** Schematizzazione della diffusione della cultura della ceramica campaniforme (*Bell Beaker*) in Europa centrale e in Gran Bretagna (immagine estratta da [89]).



Diversi eventi migratori hanno avuto luogo anche in periodi successivi a quello della prima età del Bronzo [107, 108]. I gruppi umani coinvolti, però, non risultano altamente differenziati tra loro dal punto di vista genetico così come lo erano, ad esempio, le popolazioni di cacciatori-raccoglitori mesolitici ed i primi agricoltori-allevatori neolitici [84]. A partire dall'età del Ferro la composizione genetica delle popolazioni delle diverse regioni d'Europa inizia a riflettere quella degli attuali gruppi della medesima area geografica [84, 109, 110]. Ciò che si osserva in Eurasia occidentale durante il periodo storico è un'ampia continuità genetica a livello regionale, accompagnata però anche da significative evidenze di mobilità, connessioni inter-regionali ed elevati livelli di eterogeneità locale [109]. Nonostante quest'ultimi, il mantenimento di una struttura genetica in linea generale stabile nel tempo e nello spazio potrebbe essere stato favorito da diversi fattori quali: fenomeni di mobilità transitoria, accentuati anche da una maggior velocità degli spostamenti sia via terra che via mare (motivo per cui numerosi eventi di mobilità potrebbero aver comportato una maggiore variabilità genetica ma non necessariamente consistente flusso genico), e un minor impatto demografico delle migrazioni dovuto all'incremento della dimensione della popolazione europea (che si traduce in una ridotta frazione relativa di migranti) [84, 109].

Recentemente, numerosi studi si sono focalizzati sul dato genomico antico per andare a ricostruire le dinamiche popolazionistiche che hanno interessato alcune regioni dell'Europa durante l'età del Ferro come, ad esempio, la Gran Bretagna [107, 111, 112], la Francia [113, 114], la penisola iberica [97], l'Italia [115-118], l'area del Mediterraneo [110]) o specifici contesti di particolare interesse storico-archeologico [119].

### 3. Introduzione: parte III – Gli Etruschi

#### 3.1 Cenni storici generali

Nel passaggio tra l'età del Bronzo e l'età del Ferro (intorno ai decenni finali del X secolo a.C.) nei territori che saranno sede della fioritura della civiltà etrusca si sviluppa la cultura villanoviana, così definita dal sito presso Bologna in cui è stata per la prima volta identificata (Villanova di Castenaso). Si tratta, al di là delle articolazioni locali, di una *facies* archeologica fortemente omogenea che si manifesta essenzialmente nell'adozione del rituale funerario dell'incinerazione [120]. Con l'inizio dell'età del Ferro si avvia un imponente fenomeno di concentrazione insediativa che, superando il modello di occupazione tipico del Bronzo finale, porta alla formazione di abitati accentrati, estesi e disposti su siti naturalmente difesi, esternamente ai quali sono dislocate le necropoli [120].

La cultura villanoviana si estende oltre i confini dell'Etruria propria (Toscana, Lazio settentrionale e Umbria), interessando in particolare l'area della pianura Padana e della Campania, due regioni che anche in età storica (fissata convenzionalmente per l'Italia peninsulare alla fine dell'VIII secolo a.C. [121]) continuano ad essere caratterizzate da una preponderante presenza etrusca [120, 122].

Infatti, oltre alla dodecapoli (lega delle 12 città-stato) localizzata nell'Etruria tirrenica, il controllo politico etrusco si estende tra l'VIII e il VI secolo a.C. alla costa campana e all'area padana, dove le città, rispettivamente, di Campeva e Felsina (attuale Santa Maria Capua Vetere e Bologna) rivestono un ruolo di particolare rilievo (*Figura 3.1*). È da sottolineare anche la presenza di una colonia etrusca in Corsica (Aleria), ottenuta in seguito alla vittoria della flotta alleata etrusca e cartaginese contro i Focesi (abitanti della colonia ellenica di Focea, in Asia Minore) intorno al 540 a.C., grazie alla quale gli Etruschi consolidano il proprio controllo sul Mar Tirreno e respingono l'avanzata greca [120].

La questione dell'origine del popolo etrusco e delle sue relazioni con le comunità villanoviane del periodo precedente è stata oggetto di interesse e dibattito fin dall'antichità. Secondo gli storici greci Erodoto ed Ellanico di Lesbo, gli Etruschi sarebbero originari dell'area anatolica o, più in generale, del Mar Egeo. Il primo li identifica, infatti, in una popolazione di migranti fuggiti dalla Lidia (regione dell'Anatolia corrispondente all'attuale Turchia sud-occidentale) in seguito ad una carestia, mentre il secondo nel leggendario popolo pre-ellenico dei Pelasgi che, secondo la tradizione antica, aveva colonizzato alcune isole dell'Egeo quali Lemno ed Imbro [122]. Tali ipotesi hanno trovato supporto nella presenza di influenze riconducibili all'area Mediterraneo orientale nella cultura etrusca del cosiddetto periodo orientalizzante (tra l'VIII ed il VII secolo a.C.) e nella similarità riscontrata tra i caratteri dell'alfabeto etrusco e quelli incisi nella stele di Caminia, ritrovata a Lemno.



**Figura 3.1** Cartina in cui sono indicati (In rosso chiaro) i principali centri dell'Etruria classica e le aree di espansione etrusca nella pianura Padana e in Campania, in cui dominano rispettivamente le città di Bologna (*Felsina*) e Capua (immagine modificata da [123]).

Inoltre, il capovolgimento del rito funebre dalla pratica dell'incinerazione a quella dell'inumazione, riscontrata più frequentemente nel periodo etrusco, è stato interpretato come ulteriore evidenza di discontinuità tra la fase villanoviana e quella successiva. Differente, però, è la visione dello storico greco di età augustea Dioniso d'Alicarnasso, il quale sostiene la tesi di uno sviluppo autoctono della cultura etrusca a partire dalle precedenti comunità villanoviane, ammettendo comunque impulsi culturali di provenienza orientale [122]; quest'ultimo scenario è quello che attualmente trova il maggior consenso tra gli archeologi. Accanto a queste due teorie principali, è stata anche avanzata l'ipotesi di una possibile provenienza settentrionale, che si ricollega in campo linguistico all'affinità tra la lingua etrusca ed il retico, documentato nella valle dell'Adige e nelle regioni circostanti [121, 122]. Nonostante la vicinanza tra queste due lingue non-indoeuropee, la diversità riscontrata nelle

relative manifestazioni archeologiche non consentirebbe, in assenza di documentazione epigrafica, di ipotizzare un legame tra la civiltà etrusca e quella retica [121].

Le elevate abilità nel campo della metallurgia e della navigazione [123], la sofisticata cultura e la peculiare conservazione della loro tradizione linguistica in un contesto caratterizzato in prevalenza da gruppi italici di lingua indoeuropea (come i Latini), hanno contraddistinto gli Etruschi dalle coeve popolazioni dell'Italia pre-romana.

Informazioni storiche sulle vicende che hanno interessato il popolo etrusco giungono per la maggior parte da fonti di ambito greco e romano, che rivelano però solo pochi eventi significativi.

Nel V-IV secolo a.C. alcune aree del mondo etrusco, in modo particolare quelle più periferiche, sono interessate da avvenimenti che ne determinano un profondo decadimento. Infatti, verso la fine del V secolo a.C., la Campania settentrionale etrusca affronta un radicale mutamento di identità con la perdita dei tradizionali tratti culturali e linguistici etruschi, sostituiti dall'adozione della lingua osca e, in generale, da un'entità etnica di tipo campano. La pianura Padana, invece, viene investita da intensi flussi migratori da parte di tribù celtiche di origine transalpina, tra cui popolazioni galliche che, oltrepassando gli Appennini e seguendo la valle del Tevere, assediano e devastano Roma nel 390 a.C. L'archeologia riscontra segnali evidenti di un collasso dei delicati equilibri economici dei centri etrusco-padani, una rapida riduzione della densità di popolamento dell'area e la presenza di elementi culturali di origine gallica, spesso in combinazione a caratteri tipici del substrato locale [121].

L'espansione politica romana in Etruria avviene come risultato di diverse azioni separate e in gran parte indipendenti, che prendono avvio dalla conquista definitiva di Veio nel 396 a.C. a conclusione di una guerra decennale. Nonostante la sequenza delle guerre tra Etruschi e Romani non sia sempre ricostruibile con precisione, è molto probabile che intorno al 270 a.C. tutte le città etrusche fossero incluse nel sistema di alleanze con Roma [121]; la distruzione della città di *Volsinii* (Orvieto) nel 264 a.C. viene convenzionalmente considerata l'atto finale dell'istituzione della supremazia romana sui territori dell'Etruria [120]. Nonostante la civiltà etrusca non costituisca più un'entità politica autonoma, evidenze delle sue tradizioni culturali e religiose persistono anche nei secoli successivi. Con le leggi del 90-89 a.C. viene offerta a tutte le comunità alleate di Roma, situate a sud del Po, la possibilità di richiedere la cittadinanza romana; durante il periodo della Roma imperiale, dopo il 27 a.C., l'Etruria viene incorporata all'interno della VII *regio* della divisione augustea.

Il processo di conquista e assimilazione lascia traccia di elementi culturali mutuati dall'Etruria nel mondo romano, soprattutto nella sfera della religione, della letteratura religiosa, dell'arte e delle prassi divinatorie, che si manifestano anche durante l'età imperiale (dal 27 a.C.) [122].

### 3.2 Bologna-*Felsina*: fulcro dell'Etruria padana

La geografia storica del territorio dell'Etruria padana (*Figura 3.2*) è caratterizzata da due elementi naturali: il fiume Po a nord e la catena degli Appennini a sud, attraversata dai suoi affluenti; anziché rappresentare una barriera, la posizione e conformazione naturale del territorio hanno consentito agli Etruschi dell'area padana di sviluppare contatti diretti con i centri urbani dell'Etruria tirrenica e con i popoli confinanti, tra cui i Veneti, i gruppi di cultura golasecchiana ed i Liguri verso nord, i Piceni e gli Umbri verso sud. L'Etruria padana è stata per secoli, quindi, una terra di incontri e di scambi [124]. Un passo dello storico romano Tito Livio riferisce la presenza nell'area padana di una dodecapoli fondata dalle corrispettive dodici città tirreniche; nonostante non vi siano evidenze concrete di una vera e propria dodecapoli padana, la presenza stabile degli Etruschi nella valle del Po e la loro efficiente struttura organizzativa a livello politico e commerciale rappresentano un dato certo, di cui anche gli autori antichi avevano consapevolezza [125]. Il fulcro di questo sistema è rappresentato dalla città di Bologna (*Felsina*, poi *Bononia* in età romana), definita *princeps Etruriae* da Plinio nella sua *Naturalis Historia* ad esaltare non tanto la sua indubbia rilevanza, quanto il suo precoce sviluppo (insieme a Verucchio, situata nell'area adriatica) rispetto agli altri centri padani e, dunque, il suo ruolo chiave nella genesi e formazione dello stesso *ethnos* etrusco [126].

La conquista dei territori della valle del Po è attribuita, secondo una duplice tradizione mitologica, ai personaggi leggendari di Tarconte, eroe eponimo di Tarquinia, e Ocno, re di Perugia e fondatore di Bologna e Mantova. La compresenza di queste due credenze riflette probabilmente due distinte fasi di occupazione della valle del Po: la più antica, di impronta villanoviana, finalizzata soprattutto al reperimento di nuove terre per lo sfruttamento agricolo (IX secolo a.C.), e quella più recente che, attraverso un notevole processo di riorganizzazione territoriale, ha determinato lo sviluppo di nuovi centri urbani e la creazione di un potente sistema commerciale. [125, 127].

Analogamente a quanto accade per la maggior parte dei centri etruschi dell'area tirrenica, anche attorno a Bologna si verifica, a partire dagli inizi dell'età del Ferro (IX secolo a.C., Villanoviano I), un addensamento demografico che comporta la formazione di una serie di abitati topograficamente distinti che, nel loro insieme, occupano approssimativamente il settore sud-occidentale della futura città di epoca storica [125-127]. Nella prima metà dell'VIII secolo a.C. (avanzato Villanoviano II), il progressivo declino e abbandono di questi abitati e la concomitante formazione di un insediamento stabile nell'area del successivo centro urbano costituiscono il segnale di un movimento sinecistico derivante da una precisa scelta insediativa che presuppone un considerevole stadio di evoluzione dell'assetto e delle relazioni socio-politiche dei gruppi coinvolti (riflesso nella comparsa all'interno dei corredi delle prime forme di distinzione sociale) [126]. Bologna appare, quindi, come un centro

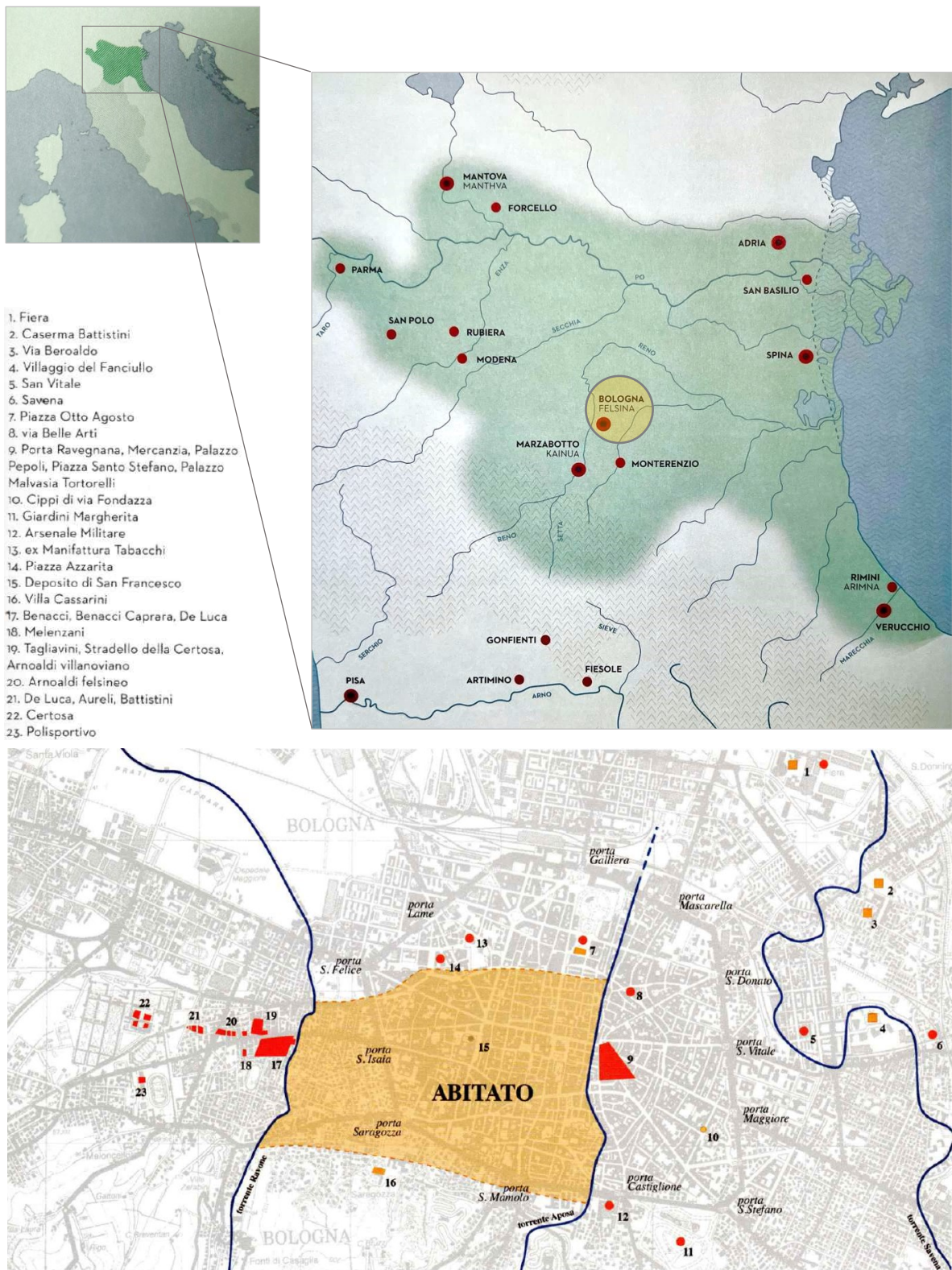
dai chiari connotati proto-urbani, delimitato ad est e ad ovest rispettivamente dai fiumi Aposa e Ravone; le necropoli si dispongono, invece, lungo quasi tutto il perimetro dell'abitato, in prossimità delle principali direttrici di traffico (*Figura 3.2*) [120, 126, 127]. È molto probabile che la formazione del centro proto-urbano di Bologna tra il IX e l'VIII secolo a.C. sia riferibile all'azione di comunità locali sotto però il possibile impulso di gruppi provenienti dall'Etruria tirrenica [124].

Tra la fine dell'VIII ed il VII secolo a.C., Bologna ha ormai raggiunto una complessa articolazione politica e sociale, fondata su clan aristocratici, e un consolidato sviluppo produttivo e commerciale che la pone al centro di un vasto circuito di scambi [124]; chiare testimonianze della sua floridezza e preminenza sono il grande deposito di bronzo noto come "ripostiglio di San Francesco", indicativo della rilevanza dell'attività metallurgica bolognese e degli scambi a lungo raggio ad essa connessi, e la ramificata distribuzione di piccoli centri agricoli e itinerari nella pianura, manifestazione di una progressiva espansione sui territori circostanti mirata allo sfruttamento agricolo e al controllo delle direttrici commerciali verso i paesi d'oltralpe e l'Etruria tirrenica [126]. Il processo di colonizzazione del territorio da parte di Bologna sembra aver avuto risvolti militari a difesa dei confini minacciati da popolazioni celtiche, come conferma Tito Livio nella narrazione della battaglia del Ticino (intorno al 600 a.C.) combattuta tra Celti ed Etruschi [122, 125]. Tra questa iniziale fase di assestamento territoriale e le invasioni "storiche" degli inizi del IV secolo a.C., si colloca un periodo di convivenza relativamente pacifica, nonché di relazioni culturali e commerciali molto strette tra gli Etruschi e i gruppi limitrofi dell'Italia settentrionale, in particolare le popolazioni celtiche, ormai solidamente attestate a nord del Po, i Liguri ed i Veneti. Questi precoci contatti dovuti ad un'intensa circolazione di merci, artigiani e tecnologie, oltre che di mercenari, devono essere considerati ed inseriti in un contesto generale di ampia mobilità e permeabilità all'interno dei diversi gruppi, che sembrano accogliere con facilità individui da altre zone, i quali si integrano e si fondono con le usanze locali, mantenendo comunque traccia della loro provenienza originaria. A *Felsina*, così come in altre aree dell'Etruria padana, è attestata, per esempio, la presenza di elementi materiali di impronta celtica golasecchiana e transalpina (come fibule tardo-halstattiane occidentali) [128, 129].

Il periodo tra il VII ed il VI secolo a.C. viene definito Orientalizzante; durante quest'ultimo, i contatti culturali e commerciali di ampio respiro che avevano contraddistinto *Felsina* fin dalle sue prime fasi, sono veicolo di ideologie, simboli e iconografie vicino-orientali, in linea con quanto accade nel resto del Mediterraneo [126]. Un avvenimento significativo in termini culturali, riconducibile alle prime fasi del VII secolo a.C., è la precoce acquisizione della scrittura, attestata da iscrizioni ritrovate nel contesto bolognese [125]. L'assetto politico ed economico dell'Etruria padana, fondato in questo stadio più antico sui centri di Bologna e Verucchio, subisce una radicale trasformazione a partire dalla metà del VI secolo a.C., periodo in cui nell'area del Mediterraneo occidentale la conflittualità

tra Etruschi, Greci e Cartaginesi, non più trattenuta nei limiti della concorrenza commerciale, sfocia in una progressiva perdita del dominio etrusco sull'alto Tirreno, con conseguente contrazione degli scambi con le popolazioni transalpine (che fino ad allora avevano seguito la rotta in direzione di Marsiglia e della valle del Rodano). È proprio in questo contesto che l'Etruria padana assume un ruolo decisivo grazie al facile accesso al Mar Adriatico e alla maggiore vicinanza ai valichi alpini. La valle del Po è interessata, quindi, da un fenomeno di generale riorganizzazione e dalla fondazione *ex novo* dei centri di Marzabotto (nell'appennino bolognese), di Spina (sull'Adriatico) e di Mantova (subito a nord del Po), nonché dalla rifondazione della stessa Bologna, capitale della nuova struttura economica basata su una rete di intense connessioni con l'Etruria tirrenica, la Grecia, e i territori d'oltralpe [125, 127]. Questo processo di riorganizzazione territoriale e strutturale, spesso associato ad una "seconda colonizzazione" dell'area padana ad opera dell'Etruria tirrenica, sembra invece essere stato promosso in larga parte da popolazioni locali (come dimostra la netta prevalenza di nomi gentilizi padani nelle città di nuova fondazione), senza escludere una possibile spinta esterna sul piano organizzativo ed economico. In questo quadro storico si colloca la definitiva evoluzione in senso urbano di Bologna-*Felsina*, che può essere considerata a tutti gli effetti una città, anche sotto l'aspetto urbanistico ed architettonico [127]. L'urbanistica regolare, l'evidenza di strade ed itinerari organizzati, le tecniche di gestione delle acque sulla base delle risorse del territorio, un'efficiente sistema commerciale fondato su tutti i poli urbani della regione, ciascuno indispensabile per gli altri, sono in estrema sintesi alcuni dei caratteri del notevole fenomeno di ristrutturazione del territorio avuto luogo nel corso del VI secolo [124].

Con la discesa dei Galli, nella prima metà del IV secolo a.C., il modello urbano creato dagli Etruschi nella pianura padana inizia a disgregarsi, con la conseguente frantumazione dell'organizzazione del territorio, all'interno del quale mutano sia le direttive viarie e commerciali che la distribuzione degli insediamenti; ciò si verifica a seguito dell'assunzione da parte dei nuovi gruppi celtici del ruolo di intermediari tra il Mar Mediterraneo e l'Europa continentale. Nonostante, dopo una prima fase di assestamento, si osservi una sorta di integrazione tra popolazioni etrusche e galliche, la struttura e la funzione economica del sistema padano risultano ormai profondamente mutati rispetto alla fase precedente [127]. Anche il processo di crescita di *Felsina* viene interrotto dall'arrivo dei Galli Boi, che provoca una contrazione dell'abitato ed un forte declino delle attività produttive e commerciali. Malgrado ciò, la città, rinominata dai Galli *Bona*, continua a mantenere un ruolo preminente nella rete urbana dell'Italia settentrionale. Nel 196 a.C., la città viene sottratta ai Galli Boi e rifondata dai romani nel 189 sotto il nome di *Bononia*, forse dal toponimo celtico *Bona* che richiama il significato augurale dell'analogo vocabolo latino (allusivo a beni, ricchezza) [130].



**Figura 3.2** In alto: cartina che mostra l'estensione dell'Etruria padana e la localizzazione dei suoi principali centri urbani. In basso: cartina di Bologna con i principali rinvenimenti di abitato (arancione chiaro) e necropoli (arancione scuro) di epoca etrusca (IX-IV secolo a.C.); i nomi dei siti sono visibili nella legenda riportata sopra la cartina, a sinistra (immagini modificate da [123, 125]).



### 3.3 Origine ed eredità del profilo genetico etrusco

Come accennato in precedenza, il tema dell'origine della popolazione etrusca ha suscitato grande interesse fin dall'antichità. In tempi più recenti, l'analisi genetica ha cercato di fornire informazioni utili ad una miglior comprensione delle dinamiche popolazionistiche legate allo sviluppo della civiltà etrusca sia attraverso l'analisi di sequenze o posizioni informative del genoma dei moderni abitanti della Toscana (assumendo una condizione di continuità genetica tra quest'ultimi e gli Etruschi), sia mediante il recupero di DNA antico da resti scheletrici di individui associati alla cultura etrusca.

Uno studio effettuato alla fine degli anni Novanta e incentrato sull'analisi delle porzioni ipervariabili (HVR1 e HVR2, localizzate nella regione di controllo D-loop) dell'mtDNA ha evidenziato per la prima volta la posizione intermedia delle sequenze toscane tra la variabilità genetica europea e quella caratteristica del Vicino Oriente [131]. Nel 2004, Vernesi e collaboratori riscontrano una particolare affinità tra il genoma mitocondriale etrusco ed il *pool* genico delle moderne popolazioni turche, suggerendo la possibilità di un effettivo flusso genico avuto luogo nel passato a partire dalle zone del Mediterraneo orientale [132]. Sebbene anche altri studi (sempre basati sull'mtDNA) abbiano successivamente evidenziato la presenza di una componente di derivazione anatolica in campioni etruschi, l'aspetto discordante tra i diversi lavori riguarda la definizione dell'arco temporale in cui l'evento di flusso genico avrebbe avuto luogo. In accordo con modelli inferenziali testati mediante metodi statistici bayesiani, alcuni lavori ipotizzano che la connessione genetica tra gli attuali toscani e l'Anatolia risalirebbe ad uno stadio remoto della Preistoria, potenzialmente riconducibile alla fase di diffusione delle comunità di agricoltori durante il Neolitico [133, 134]; ciò avvalorerebbe la teoria di uno sviluppo locale della civiltà etrusca a sfavore di un'ipotetica recente migrazione dall'Egeo. Altre ricerche [135-137] suggeriscono lo scenario opposto che, in linea con la teoria proposta dallo storico Erodoto, prevedrebbe un legame genetico più recente imputabile all'arrivo in Italia centrale di popolazioni provenienti dal Vicino Oriente intorno alla tarda Età del Bronzo.

Due ulteriori lavori, condotti su dati nucleari e genomi mitocondriali completi generati da moderne popolazioni toscane, rilevano evidenze di *admixture* con il Vicino Oriente. In particolare, l'analisi del genoma mitocondriale identifica nell'attuale Iran o Caucaso meridionale l'origine della componente vicino-orientale in esame [138]; tale segnale risulta in linea generale confermato dall'informazione nucleare che la colloca anche in questo caso nella regione del Caucaso meridionale o dell'Anatolia orientale [139]. Per cercare di interpretare i risultati divergenti ottenuti dai diversi studi sull'origine genetica degli Etruschi, il modello demografico proposto in [139] prevede una migrazione avvenuta secondo *step* multipli: gli antenati degli Etruschi, originari della zona dell'attuale Iran, si sarebbero dapprima espansi verso il Caucaso meridionale ed in seguito verso le coste occidentali della Turchia;

da qui, attraversando il Mediterraneo, avrebbero poi raggiunto l'Italia Centrale intorno a 2600-3100 anni fa, entrando in contatto con le popolazioni locali.

Un ulteriore quesito che è stato indagato da più studi riguarda l'eredità genetica della popolazione etrusca: analisi di sequenze mitocondriali hanno dimostrato l'improbabilità di una continuità diretta tra gli Etruschi e i moderni abitanti della Toscana [140, 141], ad eccezione di rari casi evidenziati da ricerche condotte su scala geografica ristretta [133].

Il primo studio *genome-wide* svolto direttamente a partire dai resti di individui afferenti alla cultura etrusca (e non su popolazioni moderne della medesima area geografica), viene pubblicato nel 2019 da Antonio e collaboratori [115]. Si tratta di un rilevante progetto paleogenomico incentrato sulla città di Roma (e dintorni) nel contesto di un transetto temporale che ricopre gli ultimi 12.000 anni, all'interno del quale il periodo dell'età del Ferro (e della Roma repubblicana) è rappresentato da 11 individui datati tra il 900 ed il 200 a.C. (di cui 2 identificati culturalmente come Villanoviani, 3 come Latini e 3 come Etruschi). Nonostante la limitatezza del dataset dell'età del Ferro e, in particolare di quello etrusco, le analisi hanno consentito di sottolineare interessanti aspetti genetici, fino ad allora poco esplorati, che caratterizzano gli abitanti della città di Roma durante tale periodo. Essi possono essere modellati con le stesse componenti genetiche che risultano predominanti nella maggior parte dei gruppi europei a seguito delle migrazioni del tardo Neolitico/prima età del Bronzo (trattate in precedenza). L'introduzione della *steppe ancestry*, probabilmente dovuta a contatti con popolazioni intermedie, determina anche in Italia un chiaro *shift* genetico tra le popolazioni dell'età del Rame e quelle dei periodi successivi (*Figura 3.3*), come confermato dallo studio di Sauppe *et al.* [142] in cui per la prima volta vengono caratterizzati a livello genomico campioni italiani dell'età del Bronzo e viene stimato l'arrivo della *steppe ancestry* in Italia centrale almeno a partire dal 1650 a.C.

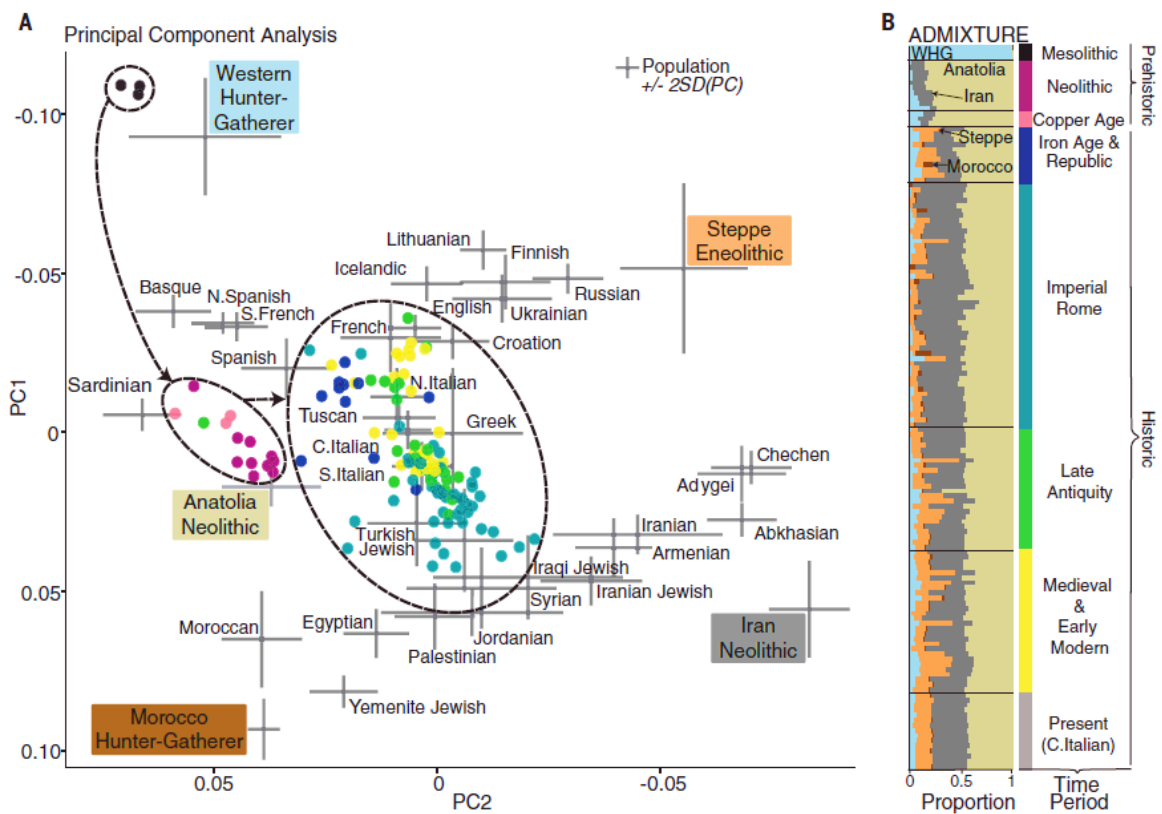
Inoltre, contrariamente agli individui dell'epoca preistorica, i campioni dell'età del Ferro iniziano ad apparire geneticamente simili alle moderne popolazioni europee e del bacino del Mediterraneo.

In aggiunta, mostrano un elevato livello di eterogeneità favorito dall'incremento delle connessioni con gruppi geograficamente distanti attraverso rapporti commerciali, campagne di colonizzazione e conflitti (agevolati dagli avanzamenti tecnologici nel campo degli spostamenti sia via terra che via mare); due individui provenienti da siti latini mostrano, infatti, una componente di influenza vicino-orientale, mentre un individuo etrusco presenta un'*ancestry* di derivazione africana.

Nel contesto della questione circa l'origine della popolazione etrusca, è interessante notare come in Antonio *et al.* non vengano riscontrate differenze tra individui etruschi e latini nella condivisione di alleli con gruppi di periodi precedenti o successivi, suggerendo una possibile sovrapposibilità dei loro profili genetici. Ad esclusione dei tre *outliers* citati sopra, tutti i restanti campioni dell'età del Ferro possono essere modellati, infatti, come il risultato della mescolanza di due componenti: una

locale (rappresentata da campioni italiani dell'età del Rame, utilizzati come *proxy*) ed una legata alla *steppe ancestry*.

Una considerazione aggiuntiva di particolare interesse riguarda il periodo seguente all'età del Ferro: a Roma, durante l'età imperiale (a partire dal 27 a.C.), si osserva un ulteriore cambiamento genetico su larga scala caratterizzato non solo da un incremento della variabilità genetica della popolazione ma, soprattutto, da uno spostamento verso componenti tipiche del Mediterraneo orientale/Vicino Oriente (visibile nell'analisi delle componenti principali, PCA di *Figura 3.3* in cui i campioni imperiali, indicati in verde smeraldo, risultano spostati in direzione del cline del Vicino Oriente); ciò si traduce nella presenza nel *pool* genetico degli individui imperiali di una consistente frazione di *Iran Neolithic ancestry*, componente genetica massimizzata nei primi gruppi di pastori neolitici dell'area iraniana, già attestata in Sicilia a partire dall'età del Bronzo (II millennio a.C.) [116].



**Figura 3.3** Parte **A**: PCA dell'Eurasia occidentale in cui sono evidenziati i principali *shifts* genetici che hanno interessato Roma (e dintorni) dal Mesolitico ad oggi. Ad eccezione di alcuni *outliers*, gli individui dell'età del Ferro (rappresentati in blu) formano un cluster nelle vicinanze delle attuali popolazioni spagnole; inoltre, si distinguono geneticamente dai precedenti campioni Neolitici e dell'età del Rame (cluster rosa e violetto) e dai successivi del periodo imperiale (verde smeraldo). Parte **B**: grafico di ADMIXTURE in cui è visibile la comparsa della componente della steppa (in arancione) a partire dell'Età del Ferro/Roma repubblicana; durante l'età imperiale si evidenzia, invece, una presenza significativa della componente *Iran\_Neolithic* (in grigio). (Immagine estratta da [115]).

Per quanto fondamentali dal punto di vista storico e popolazionistico, i diversi cambiamenti rilevati da Antonio *et al.* riguardano specificamente Roma e la sua area metropolitana e, in assenza (in quel momento) di dati genetici da altre zone d'Italia relativi all'età del Ferro e al periodo imperiale, non potevano essere considerati rappresentativi di tutta la penisola; in aggiunta, come già accennato, il numero di campioni etruschi risulta troppo esiguo e quindi statisticamente poco significativo per poter avanzare con un certo grado di affidabilità teorie sulla loro origine ed eredità genetica.

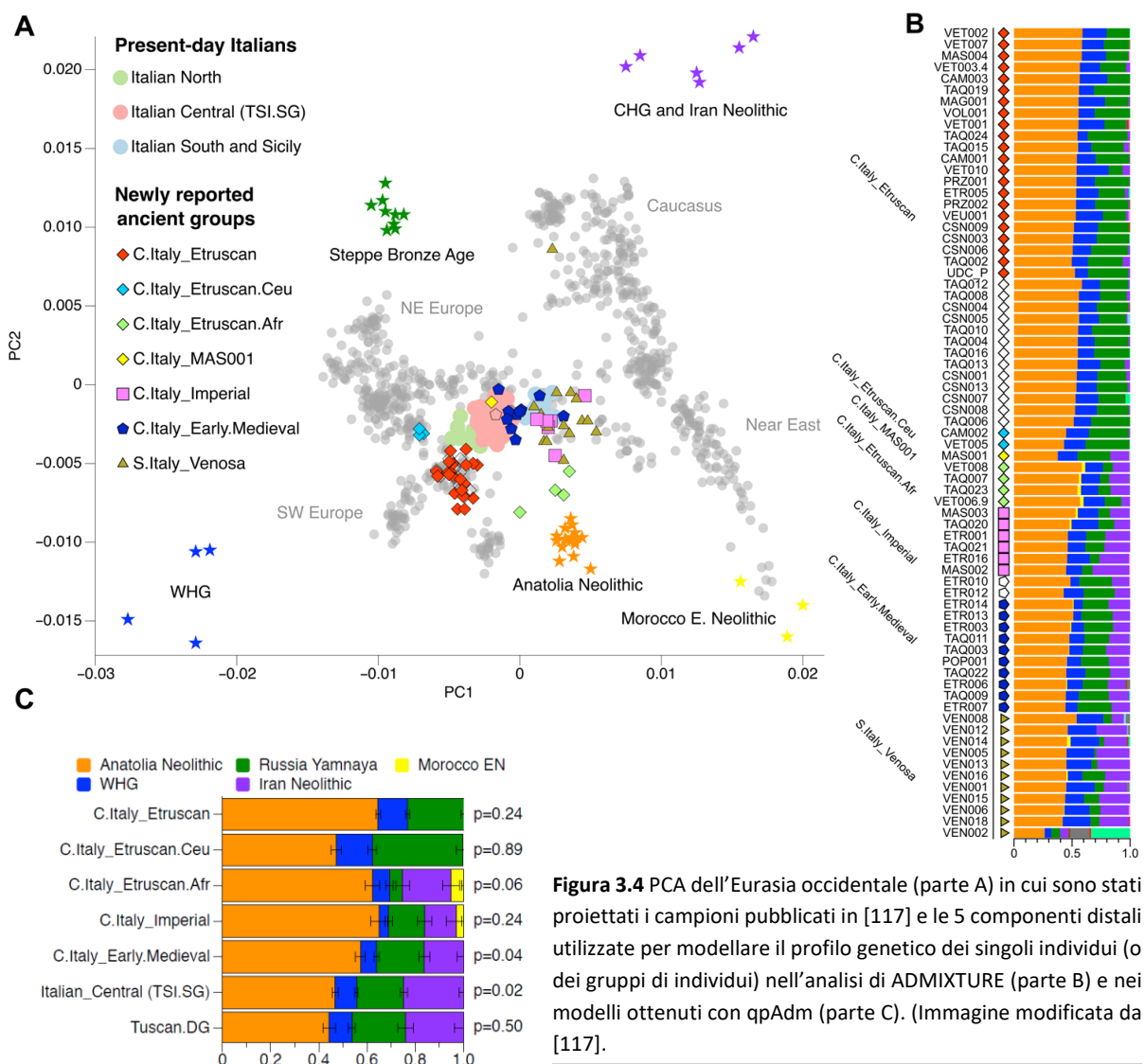
Nel 2021 viene pubblicato uno studio condotto dal Max Planck Institute for the Science of Human History di Jena, in collaborazione con l'Università degli Studi di Firenze ed altri istituti di ricerca (che mi ha vista direttamente coinvolta): si tratta del primo dataset comprendente un elevato numero di individui associati alla cultura etrusca provenienti dai territori dell'Etruria classica (N=48; Toscana e Lazio settentrionale), che ricoprono un intervallo temporale che si estende dall'800 all'1 a.C. (età del Ferro e Roma repubblicana) [117]. In aggiunta ai campioni etruschi, il dataset si compone anche di una frazione di individui datati al periodo imperiale e medievale.

Osservando il grafico della PCA dell'Eurasia occidentale di *Figura 3.4*, si nota come la maggior parte dei campioni etruschi (indicati in rosso) formi un cluster in corrispondenza delle attuali popolazioni spagnole, punto in cui ricade anche la variabilità genetica degli individui riferibili all'età del Ferro ed alla Roma repubblicana pubblicati da Antonio e colleghi (ad eccezione degli *outliers*). In assenza di evidenze di recente *admixture* con l'Anatolia, questo segnale (sostenuto anche dai modelli testati con qpAdm) è in accordo con lo scenario di un'origine locale della popolazione etrusca, il cui profilo genetico risulta condiviso con i precedenti gruppi dell'età del Bronzo (verosimile continuità genetica con le comunità Villanoviane) e con le coeve popolazioni italiche, quali i Latini, sebbene le differenze culturali e linguistiche che li caratterizzano. L'inclusione nel cluster locale di campioni che ricoprono l'intero intervallo 800-1 a.C., suggerisce il mantenimento per lungo tempo di un *pool* genetico ampiamente omogeneo, malgrado la presenza di segnali di natura alloctona in una minoranza della popolazione etrusca, verosimilmente derivanti dal Vicino Oriente (individuo MAS001), dal Nord Africa (cluster C.Italy\_Etruscan.Afr) e dal centro Europa (cluster C.Italy\_Etruscan.Ceu).

Nel *plot* di ADMIXTURE (*Figura 3.4*) e nei modelli testati con qpAdm (anche nella forma distale di *Figura 3.4*) si riscontra la presenza nella popolazione etrusca, caratterizzata da una lingua di ceppo non indoeuropeo, di un'alta proporzione di componente della steppa, comunemente associata alla diffusione delle lingue indoeuropee. La persistenza di una tradizione linguistica non indoeuropea, combinata ad un evidente cambiamento genetico (acquisizione della *steppe ancestry*), è indicativa di dinamiche complesse che oltrepassano l'assunzione semplicistica per cui modificazioni genetiche e linguistiche seguono la stessa direzione. L'assenza della componente *Iran Neolithic* nel modello di qpAdm (*Figura 3.4*) potrebbe inoltre suggerire che la stretta connessione linguistica osservata con

l'isola di Lemno non derivi da espansioni dal Mediterraneo orientale ma, piuttosto, dalla penisola italiana.

I grafici di ADMIXTURE e di qpAdm mostrano, inoltre, come gli individui con profilo genetico non locale si distinguono dal cluster principale per la presenza di componenti aggiuntive o in proporzioni differenti. È visibile, quindi, una maggior percentuale di componente della steppa negli Etruschi con profilo genetico associato a popolazioni dell'Europa centrale (C.Italy\_Etruscan.Ceu), la presenza nel cluster C.Italy\_Etruscan.Afr di un'ancestry (derivante probabilmente da contatti con popolazioni Cartaginesi) attestata in Marocco nel Neolitico iniziale, ed una frazione di *Iran Neolithic ancestry* nel campione MAS001 paragonabile a quella che caratterizza gli individui del periodo imperiale.



**Figura 3.4** PCA dell'Eurasia occidentale (parte A) in cui sono stati proiettati i campioni pubblicati in [117] e le 5 componenti distali utilizzate per modellare il profilo genetico dei singoli individui (o dei gruppi di individui) nell'analisi di ADMIXTURE (parte B) e nei modelli ottenuti con qpAdm (parte C). (Immagine modificata da [117]).

Appartiene al cluster Etruscan.Ceu il campione alloctono più antico, riconducibile al VII secolo a.C. Il fatto che il medesimo segnale genetico venga riscontrato in un altro individuo, proveniente da un sito differente e datato al III secolo a.C., consente di ipotizzare una certa continuità nell'origine di questa componente caratteristica dell'Europa centrale che, seppur identificata sporadicamente in Etruria, riflette contatti ripetuti e costanti nel tempo con gruppi di cultura celtica dalla fase Hallstatt (prima età del Ferro) alla successiva fase La Tène (dal IV secolo a.C. circa).

Nel corso del periodo imperiale si osserva un rimpiazzamento fino al 50% del *pool* genetico etrusco secondo una tendenza, già evidenziata in Antonio *et al.* e visibile, quindi, anche nei territori dell'ex Etruria, che prevede un consistente *shift* verso componenti del Mediterraneo orientale e del Vicino Oriente. Ciò rispecchia l'impatto dell'Impero romano nel determinare movimenti di risorse umane (schiavi, mercanti, soldati, mercenari) su larga scala, comportando sostanziali cambiamenti a livello di composizione e variabilità nell'assetto genetico delle popolazioni imperiali.

Un recente studio (in preprint) pubblicato su bioRxiv [110] discute i risultati dell'analisi genomica di campioni dell'età del Ferro provenienti dal bacino del Mediterraneo, in particolare dalla città di Kerkouane in Tunisia (primi genomi di questo periodo provenienti dal Nord Africa) e dalla necropoli etrusca di Tarquinia nel Lazio. Ciò che si osserva per l'Italia centrale è, ancora una volta, un segnale di continuità locale con i gruppi dell'età del Bronzo, combinato ad un incremento della variabilità genetica rispetto a quest'ultimo periodo. In quasi la metà degli individui provenienti da Tarquinia (intorno al 40%) si rileva la presenza di componenti genetiche non locali, associate al Nord Africa e al Levante, nonché un'affinità con gruppi celtici e popoli dell'Europa centrale. In generale, durante l'età del Ferro il bacino del Mediterraneo è interessato da un'intensa rete di interazioni tra popoli differenti, da consistente mobilità e flusso genico, che si riflettono nel ritrovamento di sepolture con caratteri locali in associazione ad individui con profilo genetico non autoctono [110].

Ricade proprio all'interno della variabilità pan-mediterranea la caratterizzazione a livello genomico dei Dauni, popolazione dell'età del Ferro che occupava i territori della Puglia settentrionale [118]. I dati prodotti, gli unici disponibili per l'Italia meridionale peninsulare nel contesto dell'Età del Ferro, mostrano, oltre che un'elevata eterogeneità, una significativa affinità genetica con le popolazioni coeve dell'Italia centrale e della Croazia; questo suggerisce possibili input genetici dalla penisola balcanica, nello specifico dall'Illiria, con la quale sono documentate strette affinità culturali oltre che relazioni politiche e commerciali legate al controllo dell'area che si estende tra la Dalmazia ed il Gargano.

## 4. Obiettivo dello studio

L'origine della civiltà etrusca è stata per lungo tempo oggetto di interesse e dibattito, alimentato in modo particolare dalle peculiarità linguistiche e culturali che hanno contraddistinto fin dalle prime fasi il popolo etrusco dagli adiacenti gruppi italici (come, ad esempio, i Latini) e che hanno condotto all'ipotesi di una sua provenienza non autoctona.

Recentemente, studi di carattere archeologico e genetico (quest'ultimi basati sull'analisi del DNA degradato estratto da resti umani antichi) hanno permesso di evidenziare segnali di un'origine con ogni probabilità locale della civiltà etrusca. Gli Etruschi mostrano, infatti, un profilo genetico affine a quello osservato in altre popolazioni pre-romane del contesto italiano e in continuità genetica con quello dei precedenti gruppi dell'età del Bronzo [115, 117].

Nonostante l'attuale assenza di un confronto diretto a livello genetico tra Villanoviani ed Etruschi, reso difficile dalla scarsa preservazione del materiale genetico in resti scheletrici inceneriti (pratica funeraria prevalente nella cultura villanoviana), i dati fino ad ora raccolti puntano in modo concorde verso uno sviluppo locale della civiltà etrusca a partire dalla precedente villanoviana.

I dati genetici al momento disponibili sul popolo etrusco risultano essere rappresentativi però delle sole aree dell'Etruria classica (Toscana e Lazio); data l'estensione del dominio etrusco nei territori adiacenti e l'importanza di alcuni centri urbani sorti in tali contesti, il presente studio ha previsto l'analisi genomica di oltre 50 individui associati alla cultura etrusca, provenienti dall'Etruria padana e in particolare da necropoli situate nella città di Bologna e dintorni (Casteldebole e Casalecchio di Reno), correlate all'antico abitato di *Felsina*. La maggior parte dei materiali scheletrici su cui è stato condotto lo studio risultano datati (alcuni anche al radiocarbonio) al periodo compreso tra l'VIII e il VI-V secolo a.C. e provengono da necropoli a rito misto, indicative di un processo di transizione che si manifesta in un più frequente utilizzo della pratica inumatoria.

La caratterizzazione di tali resti ha permesso non solo di contribuire ad una più completa definizione del *pool* genetico etrusco mediante l'aggiunta di informazioni derivanti da una regione non ancora indagata, ma anche di valutare il grado di affinità genetica esistente tra la comunità padana e quella dell'Etruria classica.

Inoltre, l'intervallo temporale coperto dai campioni inclusi nello studio corrisponde ad un periodo di radicali trasformazioni che conducono ad un rapido sviluppo in senso urbano di *Felsina* e alla sua assunzione di un ruolo politico e commerciale predominante all'interno dell'intera Etruria padana. Testimonianze archeologiche nell'area di Bologna evidenziano contatti tra gli Etruschi di *Felsina* ed

il mondo celtico transalpino e transpadano già in una fase precedente all'invasione storica del 388-387 a.C. Questi contatti si collocano in un quadro generale di scambi commerciali e di apertura della comunità felsinea ad elementi stranieri, ben documentati dai chiari rimandi di alcuni corredi funebri alla cultura celtica transalpina, a quelle della zona golasecchiana e ad altri gruppi culturali non locali, come i Veneti.

L'inclusione e l'integrazione di individui non autoctoni all'interno della comunità etrusca risultano evidenti anche dalle analisi genomiche riportate nel precedente studio sull'Etruria classica, dove già dal VII secolo a.C. si rileva la presenza di individui geneticamente affini a popolazioni del dell'Europa centrale, di probabile origine celtica, nelle stesse aree sepolcrali dedicate ai cittadini etruschi [117]. Le nuove analisi condotte hanno consentito di verificare l'effettiva presenza anche nell'area padana di individui di provenienza non locale e di valutare il possibile impatto di contatti, in particolare con popolazioni celtiche, sul profilo genetico degli abitanti di *Felsina* al fine di comprendere più a fondo le dinamiche relazionali esistenti tra questi due gruppi culturalmente distinti.

Recentemente, al dataset di analisi sono stati inclusi nuovi campioni provenienti sempre dall'area padana ma in prevalenza datati, su base archeologica, al periodo delle migrazioni storiche. L'analisi di questi individui (provenienti dai siti di Monterenzio Vecchio e Monte Tamburino), al momento ancora in corso e di cui verranno riportati soltanto i risultati preliminari, permetterà di indagare le implicazioni dal punto di vista genetico dell'invasione ed occupazione dei centri urbani situati tra il Po e l'Appennino da parte di gruppi celtici rappresentati, secondo le fonti storiche, da tribù di Galli Boi. Interessante sarà valutare se alle diverse dinamiche esistenti tra Etruschi e popolazioni celtiche nelle fasi Hallstatt e La Tène corrispondano anche specifici *patterns* osservabili nell'assetto genetico dei gruppi umani in esame.

Infine, completano il dataset un individuo perinatale proveniente dal tempio di Uni a Marzabotto, interessante per particolare contesto di ritrovamento, e di tre campioni provenienti dalla necropoli di San Silvestro a Faenza, identificati in accordo alle evidenze archeologiche come villanoviani.



## 5. Materiali e metodi

### 5.1 Contesto archeologico e datazione dei campioni

I campioni coinvolti nel presente studio (N=53) provengono da scavi archeologici condotti in diverse necropoli etrusche situate nell'area urbana e metropolitana della città di Bologna; nella prima zona rientrano i nuclei sepolcrali ritrovati presso i Giardini Margherita, Piazza Azzarita, Piazza VIII Agosto, Via Saffi e l'ex Manifattura Tabacchi, mentre nella seconda quelli di Casteldebole e Casalecchio di Reno (Figura 5.1). Le necropoli, caratterizzate da evidenze di rito misto (incineratorio e inumatorio), sono riferibili, in accordo con le evidenze archeologiche, al periodo compreso tra l'VIII ed il V secolo a.C. (datazioni più puntuali sono riportate nei paragrafi dedicati a ciascun sito).

I resti scheletrici da sottoporre alla caratterizzazione genetica sono stati selezionati dal Laboratorio di Antropologia Fisica dell'Università di Bologna (Dipartimento di Scienze Biologiche, Geologiche e Ambientali), al quale erano stati in precedenza affidati dalla Soprintendenza Archeologia, Belle Arti e Paesaggio per la città metropolitana di Bologna e le province di Modena, Reggio Emilia e Ferrara per l'analisi del profilo biologico.



**Figura 5.1** Mappa di Bologna e dintorni in cui sono indicati i siti da cui provengono i resti scheletrici coinvolti nello studio e le rispettive indicazioni temporali (datazione su base archeologica/genetica o al radiocarbonio).

I campioni in esame rientrano nel quadro di un progetto di ricerca più ampio focalizzato sullo studio paleogenomico dei resti di individui antichi associati alla cultura etrusca, condotto in collaborazione dall'Università degli Studi di Firenze e l'Eberhard Karls University of Tübingen (Germania); pertanto, per quanto riguarda la caratterizzazione genomica, metà dei campioni sono stati processati presso il Laboratorio di Antropologia Molecolare e Paleogenetica dell'Università di Firenze (Dipartimento di Biologia), mentre i restanti presso la sede dell'Archeo- and Paleogenetics Group dell'Università di Tubinga (Prehistory and Archaeological Sciences), nell'ambito di progetti di mobilità Erasmus.

Sulla base dei segnali genetici, i resti di alcuni individui (N=15) sono stati sottoposti ad una datazione diretta al radiocarbonio presso il CEZ Archaeometry gGmbH di Mannheim (Germania).

La selezione dei campioni da datare si è basata sui risultati dell'analisi delle componenti principali (PCA, *Paragrafo 6.5*) e ha coinvolto principalmente individui che risultavano degli *outliers* rispetto al cluster genetico di riferimento per il periodo in esame. I risultati ottenuti dalla datazione hanno evidenziato la presenza di una minoranza di individui appartenenti al periodo della Roma imperiale (N=4), tutti provenienti dal sito di Casalecchio; i restanti campioni hanno restituito, invece, datazioni in linea con il periodo stimato su base archeologica e sono in larga parte attribuibili ad una fase iniziale dell'età del Ferro (*Figura 5.1*).

Alcune delle informazioni antropologiche ed archeologiche riportate nei paragrafi seguenti, che non risultano ancora disponibili in letteratura, sono state fornite dalla Prof.ssa Maria Giovanna Belcastro (Laboratorio di Antropologia Fisica dell'Università di Bologna) e dal Funzionario Archeologo Dott.ssa Sara Campagnari (Soprintendenza Archeologia, Belle Arti e Paesaggio per la città metropolitana di Bologna e le province di Modena, Reggio Emilia e Ferrara), entrambe coinvolte in questo progetto.

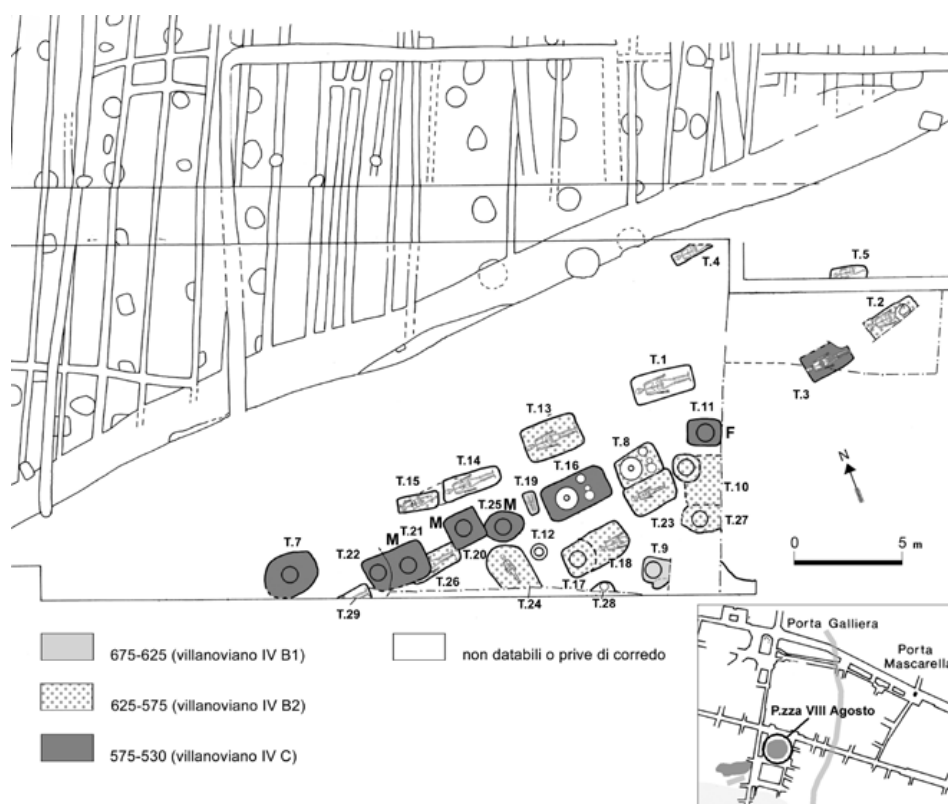
#### 5.1.1 Necropoli di Piazza VIII Agosto

Grazie agli scavi archeologici condotti nella necropoli di Piazza VIII Agosto (VIII-VI secolo a.C., *Figura 5.2*), situata in prossimità del margine nord-orientale dell'abitato di *Felsina*, sono state rinvenute 28 sepolture, 14 delle quali ad inumazione e attribuibili al VI secolo a.C. [129, 143].

Fin dalle sue prime fasi, la necropoli appare caratterizzata da un tenore generalmente modesto dei corredi, peculiarità che ricorre non solo nelle sepolture ad inumazione (in certi casi addirittura prive di corredo), ma anche in quelle a dolio, contraddistinte in altre necropoli di Bologna da apparati di accompagnamento piuttosto rilevanti. Questi elementi suggeriscono che la necropoli di Piazza VIII Agosto possa essersi sviluppata in rapporto ad uno specifico contesto sociale [129]; è stata avanzata anche l'ipotesi che questo sepolcreto possa riferirsi ad un unico gruppo familiare o, comunque, ad una comunità piuttosto ristretta [143]. Si tratta, in aggiunta, di una necropoli caratterizzata da chiari

riferimenti all'ambito adriatico veneto e veneto-orientale, con rimandi all'area renano-ticinese ed isontina (in particolare al sito di Santa Lucia di Tolmino). Questi caratteri suggeriscono la verosimile sepoltura di individui con origini veneto-orientali e/o isontine, nonché legate all'ambito alpino sud-orientale della cerchia hallstattiana [129]. Quello di Piazza VIII Agosto non sembra costituire un caso isolato: simili evidenze di carattere non locale sono attestate anche nel vicino sepolcreto della zona dell'ex Manifattura Tabacchi, anche esso collocato nei pressi delle direttrici che si allontanavano dalla città in direzione nord [129].

I resti scheletrici di 10 inumati (rocche petrose) derivanti dalle sepolture indicate in *compreso* tra il VII-V secolo a.C. sono stati sottoposti all'analisi paleogenetica. Le sepolture 1, 13 e 26 erano in cassa lignea, mentre le restanti in semplice fossa terragna; solo la T.4 e la T.29 erano prive di corredo. In particolare, la T.4 risulta databile alla seconda metà del VI secolo a.C. in quanto impostata sulla spoliazione della struttura lignea attestata in Piazza VIII Agosto dalla metà dell'VIII secolo, probabilmente riferibile ad un edificio a destinazione pubblica [126]. Dalla T.18 proviene, invece, un corredo femminile con confronti con l'ambito ligure e versiliese (si ricorda in questo contesto il tema della celtizzazione dei Liguri [144]). La sovrapposizione parziale osservata per alcune sepolture di questa necropoli indica una possibile deposizione dei defunti secondo eventi temporalmente distinti [145].



**Figura 5.2** Planimetria della necropoli di Piazza VIII Agosto con indicazione delle fasi cronologiche a cui le diverse sepolture sono state attribuite (immagine estratta da [143]).

Sulla base delle evidenze genetiche, gli individui provenienti dalla T.4 e T.18 (AGOS003 e AGOS001, rispettivamente) sono stati datati al radiocarbonio ( $^{14}\text{C}$ ). L'ampio intervallo di datazione, visibile in *compreso* tra il VII-V secolo a.C., dipende dall'andamento della curva di calibrazione associata al periodo dell'età del Ferro compreso tra il VII-V secolo a.C.

N° sepoltura	Sesso	Età individuo	ID laboratorio Tübingen	Elemento scheletrico	Datazione $^{14}\text{C}$ (CI 95%)
T.1	F	AG	AGOS004	Rocca petrosa	-
T.4	M	AG	AGOS003	Rocca petrosa	cal BC 744-406
T.13	F	AG	AGOS007	Rocca petrosa	-
T.14	NI	AG (18-24)	AGOS008	Rocca petrosa	-
T.15	NI	10-11	AGOS006	Rocca petrosa	-
T.18	F	AM	AGOS001	Rocca petrosa	cal BC 769-516
T.23	NI	4-5	AGOS005	Rocca petrosa	-
T.24	NI	6-7	AGOS010	Rocca petrosa	-
T.26	M	AG (o AM)	AGOS002	Rocca petrosa	-
T.29	F	AG	AGOS009	Rocca petrosa	-

**Tabella 5.1** Principali informazioni relative agli individui provenienti dalla necropoli di Piazza VIII Agosto coinvolti nello studio. Viene indicata la sepoltura di provenienza, il sesso e l'età (in anni o fascia di appartenenza: AM=adulto maturo, AG=adulto giovane) dell'individuo, identificati tramite analisi morfometrica presso il Laboratorio di Antropologia Fisica dell'Università di Bologna. Viene riportato, inoltre, l'ID attribuito all'individuo per la sua identificazione durante le analisi paleogenetiche, il distretto scheletrico pervenuto per quest'ultime e l'eventuale datazione al  $^{14}\text{C}$ .

### 5.1.2 Necropoli di Piazza Azzarita

Nella necropoli di Piazza Azzarita sono state rinvenute 60 sepolture, di cui 30 ad inumazione che si attestano in un arco temporale compreso tra la seconda metà del VI e gli inizi del V secolo a.C.

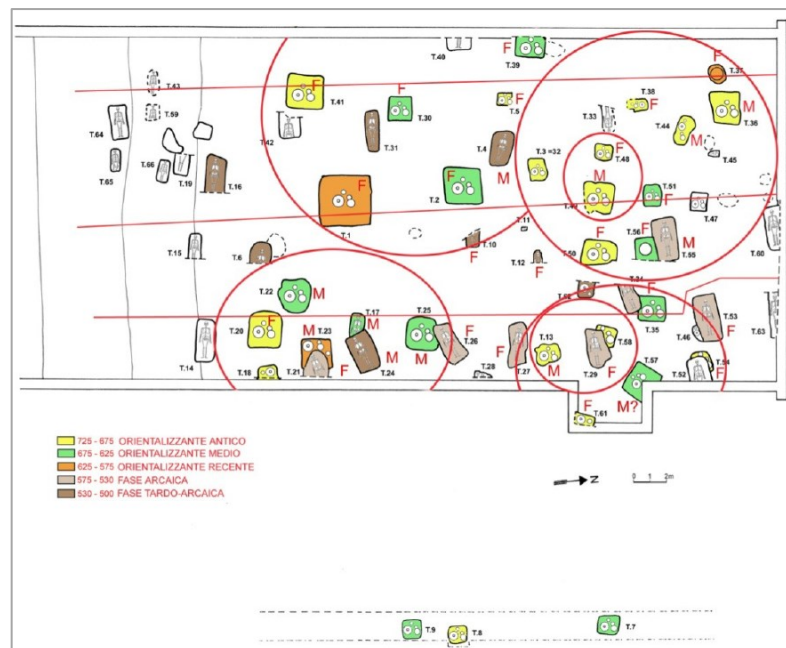
La necropoli di Piazza Azzarita, così come altri scavi che hanno interessato negli ultimi anni il settore settentrionale della città (da ovest ad est: ex Manifattura Tabacchi, Piazza VIII Agosto, via Belle Arti), suggeriscono tra la fine del VII e l'inizio del VI secolo a.C., l'avvio di nuove dinamiche di occupazione degli spazi sepolcrali. Nelle medesime aree, durante i decenni centrali del VI secolo a.C., accanto a rari complessi che manifestano ricchezza aristocratica attraverso materiali pregiati e importati, si registra il passaggio preferenziale verso il rito dell'inumazione, con fosse uniformemente orientate in senso est-ovest. Con la loro generale modestia e uniformità dei corredi, tali sepolture indicano la presenza di gruppi sociali emergenti con nuove necessità di autorappresentazione, probabile segno di trasformazioni sociali in atto nella comunità felsinea [126]. Affianco alla standardizzazione e alla sobrietà dei corredi, durante la fase arcaica emerge il fenomeno di distinzione di individui (o gruppi)

di origine non locale, evidente sia negli aspetti di organizzazione topografica che nel rituale e nella composizione dei corredi, caratterizzati da elementi fortemente connotativi di una diversa identità culturale accanto ad oggetti caratteristici del *milieu* locale [146].

I 14 individui coinvolti nello studio genomico provengono dalle inumazioni indicate in *Tabella 5.2*. Gli individui erano quasi tutti in posizione supina con il cranio orientato ad ovest ad esclusione di quelli provenienti dalla T.27 e T.33, in cui il capo era rivolto verso est. Gli unici individui deposti in posizione differente risultano quelli provenienti dalle tombe 27 (sul fianco sinistro con arti inferiori flessi), 43 (rannicchiato sul fianco destro) e, infine, 66 (rannicchiato sul fianco sinistro). L'ultima è l'unica sepoltura in cui, oltre ai resti di un giovane uomo, sono stati rinvenuti anche pochi elementi scheletrici appartenenti ad un bambino di cui non è stato possibile però identificare l'età.

Molti resti risultavano incompleti o danneggiati a causa di arature di età romana, interventi edilizi moderni o costruzione di canali fognari; lo stato di preservazione dei singoli elementi scheletrici era anche esso in molti casi precario [143].

Sulla base del profilo genetico ottenuto, l'inumato proveniente dalla T.55 (AZZA006) è stato datato al radiocarbonio (*Tabella 5.2*). Per questo individuo sono disponibili anche informazioni relative al corredo che lo collocano intorno al 550 a.C.; il corredo caratterizzato da rimandi all'Etruria tirrenica nord-occidentale, include una fibula serpeggiante (di norma attribuita al genere maschile), la cui diffusione interessa l'Etruria settentrionale costiera, l'area dell'Etruria padana, la zona atestina e golasecchiana e il mondo celtico transalpino.



**Figura 5.3** Planimetria della necropoli di Piazza Azzarita con indicazione cronologica delle diverse sepolture ritrovate (immagine estrapolata da [146]).

N° sepoltura	Sesso	Età individuo	ID laboratorio Tübingen	Elemento scheletrico	Datazione <sup>14</sup> C (CI 95%)
T.6	F	AM	AZZA001	Rocca petrosa	-
T.12	F	AG	AZZA003	Rocca petrosa	-
T.15	M	AM	AZZA002	Rocca petrosa	-
T.26	F	AG	AZZA011	Rocca petrosa	-
T.27	F	AG	AZZA013	Rocca petrosa	-
T.28	M	AG (19-25)	AZZA014	Rocca petrosa	-
T.33	M	AG	AZZA010	Rocca petrosa	-
T.43	NI	AG	AZZA012	Rocca petrosa	-
T.53	F	AM	AZZA004	Rocca petrosa	-
T.55	F	AG	AZZA006	Rocca petrosa	cal BC 792-550
T.60	F	12-16	AZZA005	Rocca petrosa	-
T.64	F	AM	AZZA007	Rocca petrosa	-
T.65	NI	6-7	AZZA008	Rocca petrosa	-
T.66a	M	AG (19-20)	AZZA009	Rocca petrosa	-

**Tabella 5.2** Principali informazioni riguardanti gli individui ritrovati nella necropoli di Piazza Azzarita coinvolti in questo studio (vedi didascalia di *Tabella 5.1* per la descrizione delle diverse colonne). Nella stima del sesso, effettuata su base morfometrica, NI=non identificabile.

### 5.1.3 Necropoli dell'area ex Manifattura Tabacchi

Il nucleo dell'ex Manifattura Tabacchi ha restituito 19 sepolture per la maggior parte ascrivibili al VI secolo a.C., di cui 9 ad inumazione. Le inumazioni, nello specifico, sono associate ad un'occupazione antica della necropoli, databile all'ultima fase del Villanoviano IV, tra la fine del VII e l'inizio del VI secolo a.C. Le inumazioni erano tutte in cassa lignea e orientate in direzione est-ovest (cranio ad ovest), ad esclusione delle sepolture 16 (unica orientata nord-sud, con cranio a nord, e unica priva di corredo) e 19 (non facente parte nello studio), che erano in fossa terragna.

Alcune tombe risultavano spoliate in antico, tra cui le T.10 e 12, violate in epoca etrusca, e la T.11, fortemente intaccata da interventi post-medievali. Al momento dello scavo archeologico, tutte le inumazioni apparivano in cattivo stato di conservazione e con gli scheletri non in connessione, forse a causa di allagamenti che avrebbero interessato le tombe in seguito a possibili innalzamenti della falda acquifera [147]. È stato ipotizzato che anche questa necropoli possa essere riferibile ad un unico gruppo familiare o ad una comunità ristretta che avrebbe utilizzato tale spazio sepolcrale per circa tre generazioni, tra il 590 ed il 510 a.C. [143].

La necropoli presenta alcuni corredi che rimandano all'area golasecchiana per tipologia di oggetti e per la rielaborazione locale delle produzioni venete a fasce rosse e nere.

Tra gli individui inclusi nello studio paleogenomico (N=5, *Tabella 5.3*), quelli provenienti dalle tombe numero 10 (TAB010) e 16 (TAB016) sono stati datati al radiocarbonio. Il primo tentativo di datazione

del campione TAB016 non ha restituito risultati a causa della scarsa preservazione del collagene. Data la peculiarità della sepoltura e del profilo genetico dell'inumato, una seconda prova di datazione è attualmente in corso. Per quanto concerne il corredo, in associazione alla T.10, è stata ritrovata un'olla di grandi dimensioni con rimandi all'Etruria tirrenica (in particolare a Lucca); non sono presenti però elementi utili a circoscrivere la cronologia della sepoltura che, in linea generale, potrebbe essere attribuibile alla seconda metà del VI secolo.

N° sepoltura	Sesso	Età individuo	ID laboratorio Firenze	Elemento scheletrico	Datazione <sup>14</sup> C (CI 95%)
T.10	M	AM	TAB010	Rocca petrosa	cal BC 790-550
T.11	M	AG	TAB011	Rocca petrosa	-
T.12	F	AG	TAB012	Rocca petrosa	-
T.14	F	AG	TAB014	Rocca petrosa	-
T.16	F	AG (o AM)	TAB016	Rocca petrosa	collagene insufficiente

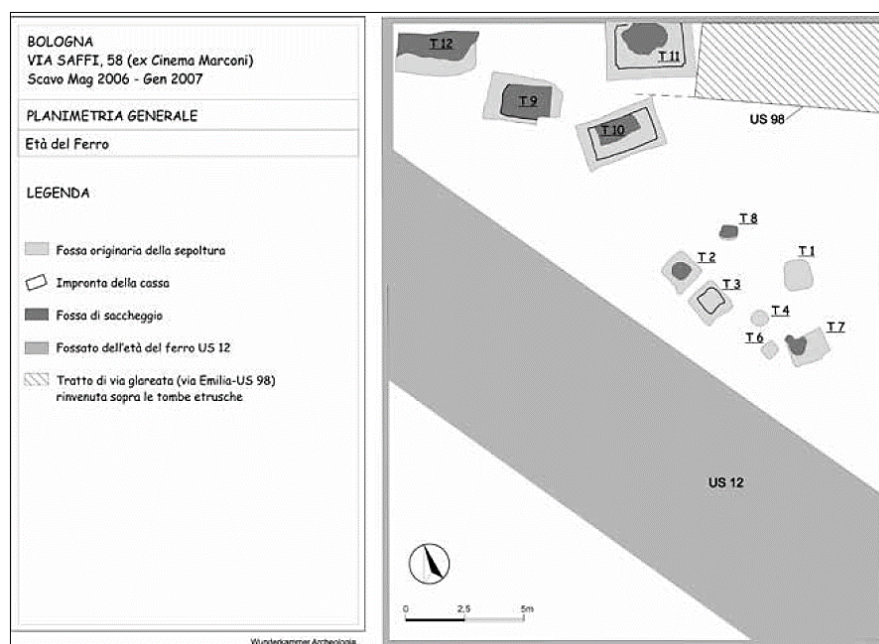
**Tabella 5.3** Principali informazioni sugli individui inclusi nello studio, provenienti dalla necropoli scavata nell'area dell'ex Manifattura Tabacchi (vedi didascalia di *Tabella 5.1* per la descrizione delle diverse colonne).

#### 5.1.4 Necropoli di Via Saffi

La piccola necropoli di Via Saffi (riferibile a fine VII-inizi VI e V secolo a.C., *Figura 5.4*) comprende 11 sepolture di cui 5 inumazioni in cassa lignea rettangolare che si attestano nell'ambito del V secolo [148]. Tutte le sepolture ritrovate si dispongono lungo la sponda settentrionale di un ampio canale di epoca etrusca che sembra segnalare un vero e proprio confine, in quanto nessun elemento di rilevanza archeologica è stato individuato sulla sua sponda meridionale (*Figura 5.4*) [149].

Tutte le inumazioni risultano saccheggiate in antico (in un momento anteriore al disfacimento della cassa) e conservano solo scarsi frammenti ossei, rinvenuti non in connessione anatomica.

La sepoltura numero 10, da cui proviene l'unico individuo della necropoli incluso nel presente studio (per il quale è disponibile la datazione al radiocarbonio, riportata in *Tabella 5.4*), fa parte del gruppo settentrionale del sepolcreto (T.9-12) che, con ogni evenienza, proseguiva oltre i limiti dello scavo. Le tombe 10-12 hanno conservato alla sommità segnacoli tombali che, nonostante risultino spostati dalla posizione originaria in seguito ad eventi di violazione passati, sono comunque mantenuti entro i limiti dei rispettivi spazi sepolcrali [148]. La sepoltura 11 (non compresa nello studio genomico) ha restituito una stele figurata di eccezionale qualità e dai significati e rinvii ideologici estremamente complessi [149].



**Figura 5.4** Planimetria generale della necropoli scavata presso Via Saffi (immagine estratta da [148]).

N° sepoltura	Sesso	Età individuo	ID laboratorio Tübingen	Elemento scheletrico	Datazione <sup>14</sup> C (CI 95%)
T.10	M	AG	VISA001	Rocca petrosa	cal BC 756-420

**Tabella 5.4** Principali informazioni relative all'unico individuo proveniente dalla necropoli di Via Saffi esaminato nel presente progetto (vedi didascalia di *Tabella 5.1* per la descrizione delle diverse colonne).

### 5.1.5 Necropoli di Giardini Margherita

A differenza dei siti appena trattati, gli scavi delle necropoli di Giardini Margherita, Casalecchio di Reno e Casteldebole (condotti dalla Soprintendenza Archeologica dell'Emilia-Romagna) e le analisi morfometriche su alcuni dei resti umani ad esse associati risalgono a tempi meno recenti. È inoltre possibile che parte del materiale scheletrico non sia ancora stato studiato o, perlomeno, pubblicato. Per questi motivi, la documentazione relativa ai contesti e ai singoli inumati provenienti da questi siti risulta, al momento, piuttosto frammentaria.

Le inumazioni di Giardini Margherita (sito riferibile a metà VI-inizio IV secolo a.C.) appartengono ad un ampio sepolcreto (236 sepolture), più volte esplorato tra il 1876 e il 1986, anno di rinvenimento del materiale scheletrico esaminato e riportato in [150]. I corredi riscontrati nelle sepolture di tale necropoli risultano piuttosto ricchi; ne è esempio quello della cosiddetta Tomba Grande (T.6, metà V secolo a.C.), appartenente con ogni probabilità ad un individuo femminile di alto rango, che viene considerato il più ricco di Giardini Margherita e forse di tutta *Felsina*. Tale corredo comprende un



raffinato servizio di vasellame da banchetto con ceramiche d'importazione greca e oggetti di bronzo di produzione etrusca, un grande candelabro di bronzo, un portafiaccole ed oggetti di ornamento personale della defunta, tra cui un anello d'oro, una fibula d'argento e una di bronzo rivestita da una sottile lamina d'oro [151].

In *Tabella 5.5* vengono mostrate le informazioni principali (anche se parziali) relative agli 8 individui analizzati a livello genomico in questo studio. Per tre di essi è stata effettuata una datazione diretta al  $^{14}\text{C}$ ; quest'ultima ha evidenziato l'appartenenza dell'individuo MAR012 (associato alla T.12) ad un periodo dell'età del Ferro successivo rispetto a quello in cui ricadono i campioni finora descritti.

Durante lo step di pulizia superficiale del cranio intero fornitoci per il campione MAR021 è stato ritrovato un ossicino dell'udito (incudine) all'interno del meato acustico della rocca petrosa destra, il quale è stato utilizzato come materiale di partenza per le analisi molecolari.

N° sepoltura	Sesso	Età individuo	ID laboratorio Firenze	Elemento scheletrico	Datazione $^{14}\text{C}$ (CI 95%)
T. 1	M	AG (17-20)	MAR001	Rocca petrosa	-
T.3	-	-	MAR003	Rocca petrosa	-
T. 6	F	AG (20-29)	MAR006	Rocca petrosa	cal BC 772-541
T.12	-	-	MAR012	Rocca petrosa	cal BC 400-212
T.17	-	-	MAR017	Rocca petrosa	-
T. 18	-	-	MAR018	Rocca petrosa	-
T. 20	-	-	MAR020	Rocca petrosa	-
T. 21	-	-	MAR021	Cranio	cal BC 734-405

**Tabella 5.5** Principali informazioni relative agli individui provenienti dagli scavi di Giardini Margherita, inclusi nel presente studio (vedi didascalia di *Tabella 5.1* per la descrizione delle diverse colonne). Le informazioni relative alla stima del sesso e dell'età sono state in parte estratte da [150] e in parte fornite dal Laboratorio di Antropologia Fisica dell'Università di Bologna.

### 5.1.6 Necropoli di Casteldebole

Una piccola zona sepolcrale riferibile al V secolo a.C. e costituita da 4 sepolture è stata rinvenuta a Casteldebole in una cava di ghiaia, al di sotto di una villa romana. I resti umani ritrovati, in discreto stato di conservazione, sono riferibili a 4 soggetti adulti, due dei quali inclusi nel presente progetto (*Tabella 5.6*). Tra questi, l'individuo proveniente dalla sepoltura T.19 presenta un'età alla morte superiore a 60 anni; l'analisi morfologica dei resti scheletrici ha evidenziato, inoltre, la presenza di alcuni aspetti patologici, quali ascessi mandibolari e mascellari a livello dei denti molari [150].

Per le sue caratteristiche genetiche, il campione CAST002 (proveniente dalla T.4) è stato sottoposto a datazione al radiocarbonio, il cui risultato lo colloca in un periodo potenzialmente precedente al

V secolo a.C. Avendo a disposizione dei crani interi, le analisi molecolari di CAST001 e CAST002 sono state condotte sfruttando rispettivamente un dente molare (RM<sup>1</sup>) e un ossicino dell'udito (incudine) ritrovato in fase di pulizia superficiale del campione dai residui di suolo.

N° sepoltura	Sesso	Età individuo	ID laboratorio Tübingen/Firenze	Elemento scheletrico	Datazione <sup>14</sup> C (CI 95%)
T.4	M	AM (30-39)	CAST002	Cranio	cal BC 796-546
T.19	F	AM (> 60)	CAST001	Cranio	-

**Tabella 5.6** Principali informazioni riguardanti i resti scheletrici coinvolti in questo studio, provenienti dalla piccola area sepolcrale di Casteldebole (vedi didascalia di *Tabella 5.1* per la descrizione delle diverse colonne). Le informazioni relative alla stima del sesso e dell'età sono state estratte da [150].

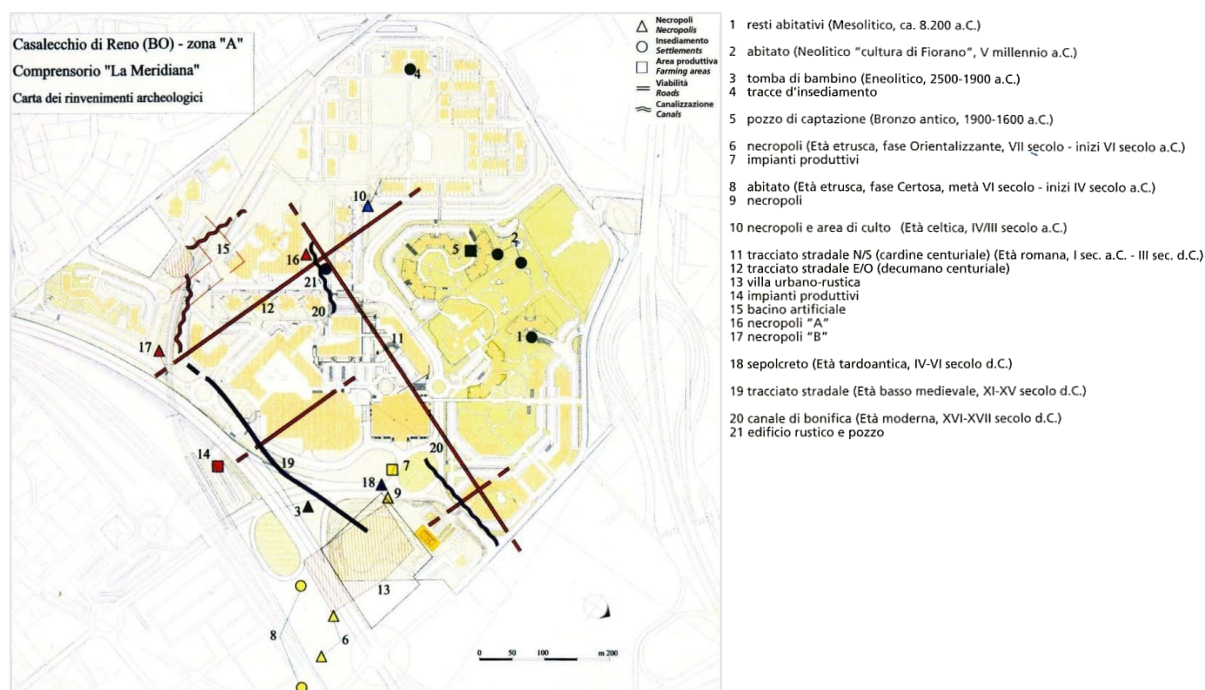
### 5.1.7 Necropoli di Casalecchio di Reno

Il materiale scheletrico proveniente dalla necropoli di Casalecchio di Reno comprende 13 inumati di cui 4 indicati come subadulti (*Tabella 5.7*).

Non sono disponibili al momento informazioni puntuali relative ai resti umani coinvolti nello studio, se non quelle deducibili dalle note che accompagnano la numerazione delle sepolture (*Tabella 5.7*), come nel caso dei due individui entrambi associati alla T.5, uno descritto come villanoviano e l'altro come etrusco; queste indicazioni suggeriscono la presenza, al momento dello scavo, archeologico di evidenze stratigrafiche o materiali che potessero indicare l'appartenenza di CAS005-VILL ad un periodo precedente rispetto a CAS005-ETR.

Nonostante tutte le inumazioni risultassero associate a individui di cultura etrusca, riferibili alla fase compresa tra il VII ed il VI secolo a.C., le datazioni al radiocarbonio hanno evidenziato la presenza di 2 individui (CAS054 e CAS067, provenienti rispettivamente dalle T.54 e T.67) attribuibili al periodo imperiale romano. Sulla base delle evidenze genetiche, anche gli inumati associati alle T.90 e T.95 sembrano poter essere ascrivibili allo stesso periodo. Tale risultato, nel contesto di Casalecchio, è in linea con le evidenze archeologiche (quali tracce di strutture abitative e necropoli), che attestano un'occupazione continuativa del sito tra il Mesolitico ed il Basso Medioevo (*Figura 5.5*). Sono state rinvenute in questa necropoli anche sepolture che rientrano nell'ambito celtico [152-154]. Infatti, i dati archeologici rivelano che, con l'arrivo dei Galli, la zona "A" della necropoli viene abbandonata dagli Etruschi durante la prima fase del IV secolo, per poi essere nuovamente occupata dai gruppi celtici (al di sopra delle rovine etrusche) qualche decennio dopo [155].

Per quanto riguarda le datazioni al <sup>14</sup>C, nella tabella riportata di seguito sono indicate anche quelle relative ai campioni CAS005-VILL, CAS022 e CAS041-1.



**Figura 5.5** Carta dei ritrovamenti archeologici di Casalecchio di Reno. I resti di strutture abitative e di necropoli testimoniano una continua occupazione del sito dal Mesolitico al Basso Medioevo (immagine estratta da <https://tinyurl.com/mu6vzsvr>).

N° sepoltura	Sesso	Età individuo	ID laboratorio Firenze	Elemento scheletrico	Datazione <sup>14</sup> C (CI 95%)
T. 1	-	subadulto	CAS001	Rocca petrosa	-
T. 5 (etrusco)	-	-	CAS005-ETR	Rocca petrosa	-
T. 5 (villanoviano)	-	-	CAS005-VILL	Rocca petrosa	cal BC 794-556
T. 20	-	-	CAS020	Rocca petrosa	-
T. 22	-	-	CAS022	Rocca petrosa	cal BC 753-417
T. 28	-	-	CAS028	Rocca petrosa	-
T. 36	-	-	CAS036	Rocca petrosa	-
T. 41(1)	-	-	CAS041-1	Rocca petrosa	cal BC 786-547
T. 41(2)	-	-	CAS041-2	Rocca petrosa	-
T. 54	-	subadulto	CAS054	Rocca petrosa	cal AD 129-313
T. 67	-	-	CAS067	Rocca petrosa	cal AD 231-345
T. 90	-	subadulto	CAS090	Rocca petrosa	-
T. 95	-	subadulto	CAS095	Rocca petrosa	-

**Tabella 5.7** Principali informazioni relative ai resti umani analizzati nello studio e provenienti dalla necropoli di Casalecchio di Reno (vedi didascalia di *Tabella 5.1* per la descrizione delle diverse colonne). Le informazioni riguardanti stima del sesso e dell'età alla morte per l'inumato della sepoltura 1 sono state estratte da [150].

### 5.1.8 Recenti aggiunte al dataset etrusco

Recentemente sono stati aggiunti al dataset descritto finora ulteriori 30 campioni sempre rientranti nell'ambito dell'Etruria padana (*Tabella 5.8*). La maggior parte di questi deriva dalle necropoli di Monte Tamburino (a Monte Bibebe, N=15) e Monterenzio Vecchio (N=11), entrambe in provincia di Bologna (circa 30 km da essa in direzione sud-est) e datate, in accordo alle evidenze archeologiche, rispettivamente al V-II e al IV-III secolo a.C. Si tratta di campioni che ricoprono una fase dell'età del Ferro successiva rispetto a quasi tutti i contesti trattati in precedenza e che coincide, nei territori dell'Italia settentrionale, al periodo delle ripetute ondate migratorie di gruppi celtici provenienti dal centro Europa.

La necropoli di Monte Tamburino è considerata la più grande di ambito celtico in Italia e comprende almeno 171 sepolture. Si caratterizza per la presenza simultanea di materiali riferibili alla tradizione sia etrusca che celtica, in aggiunta ad oggetti di corredo potenzialmente derivanti da un processo di mescolanza in atto tra le due culture [156]. Un'analisi degli isotopi dello stronzio effettuata su 21 individui ritrovati nella necropoli ha evidenziato che il 19% di questi non presenta un'origine locale; non è stato possibile, però, determinare l'area geografica di provenienza a causa della similarità del rapporto  $^{87}\text{Sr}/^{86}\text{Sr}$  con quello rilevato per la zona di Monte Bibebe [157].

Anche nella vicina necropoli di Monterenzio Vecchio sono state individuate chiare evidenze di sincretismo culturale che si manifestano nella coesistenza di pratiche funerarie basate su costumi nordici (ideologia funeraria dominata dalle armi) e riti locali di impronta mediterranea (ideologia incentrata sul banchetto), a conferma delle intense interazioni esistenti tra la civiltà etrusca e celtica. Il processo di scambio e mescolanza culturale sfocia agli inizi del IV secolo a.C. nello sviluppo di una cultura che viene definita etrusco-celtica [155]. Nel lavoro di Sorrentino e collaboratori [155], l'effettiva presenza di individui di origine non locale, è confermata dall'analisi isotopica dello stronzio e dallo studio di tratti non metrici dentali; in modo particolare, quest'ultimi evidenziano un'elevata affinità biologica tra gli individui non autoctoni della necropoli e popolazioni celtiche dell'Europa non continentale, provenienti dallo Yorkshire (Regno Unito).

Risalgono, invece, ad una fase più antica i tre campioni provenienti dalla necropoli di San Silvestro (frazione di Faenza, provincia di Ravenna) classificati come villanoviani e riconducibili, sulla base del contesto stratigrafico, all'intervallo compreso tra l'VIII ed il VI secolo a.C.

Infine, è stato incluso nelle analisi paleogenomiche un individuo perinatale (MARZ001), ritrovato nel Tempio di Uni a Marzabotto (provincia di Bologna), per il quale erano già disponibili i dati relativi al genoma mitocondriale (aplogruppo U5a1g), alla determinazione del sesso (maschile, effettuata anche su base genetica) e alla datazione al radiocarbonio (*Tabella 5.8*) [158].

Per i campioni provenienti dai contesti brevemente descritti in questo paragrafo, verranno esposti solo i primi risultati preliminari dell'analisi genomica, che è tuttora in corso.

Sito	ID archeologico	Sesso	Età individuo	ID laboratorio Tübingen/Firenze	Elemento scheletrico	Datazione <sup>14</sup> C (CI 95%)
Monterenzio Vecchio	MV T 2	F	35-50	MOVE001	Rocca petrosa	-
	MVT 3	M	20-25	MOVE002	Rocca petrosa	-
	MVT 14	F	> 50	MOVE003	Rocca petrosa	-
	MV T 15	M	35-50	MOVE004	Rocca petrosa	-
	MVT 24	-	3-12	MOVE005	Dente (Rdm <sup>2</sup> )	-
	MV T35	-	3-12	MOVE006	Dente (Ldm <sup>1</sup> )	-
	MV T39	M	20-35	MOVE007	Rocca petrosa	-
	MV T A	F	20-35	MOVE008	Rocca petrosa	-
	MV T B	M	20-35	MOVE009	Dente (LM <sup>1</sup> )	-
	MV T C	-	-	MOVE010	Dente (RM <sub>1</sub> )	-
	MV T F	-	-	MOVE011	Dente (LM <sup>2</sup> )	-
Monte Tamburino/ Monte Bibebe	MT 13	F	Adulto	MOTE001	Rocca petrosa	-
	MT 19	F	20-30	MOTE002	Rocca petrosa	-
	MT 20	F	21-40	MOTE003	Dente (LM <sub>1</sub> )	-
	MT 41	F	30-35	MOTE004	Rocca petrosa	-
	MT 42	M	40-45	MOTE005	Dente (LM <sup>1</sup> )	-
	MT 43	F	20-25	MOTE006	Dente (RM <sup>2</sup> )	-
	MT 62	F	~20	MOTE007	Rocca petrosa	-
	MT 67	M	Adulto	MOTE008	Rocca petrosa	-
	MT 78	F	30-40	MOTE009	Dente (LM <sub>1</sub> )	-
	MT 83	F	50-60	MOTE010	Dente (RM <sub>1</sub> )	-
	MT 90	F	~20	MOTE011	Rocca petrosa	-
	MT 91	M	50-60	MOTE012	Dente (LM <sub>1</sub> )	-
	MT 107	M	~20	MOTE013	Dente (RM <sub>1</sub> )	-
	MT 119	M	55-60	MOTE014	Rocca petrosa	-
	MT 150	F	Adulto	MOTE015	Dente (LM <sup>3</sup> )	-
San Silvestro (Faenza)	FZN 6 US 322	-	-	SILV001	Rocca petrosa	-
	FZN 7 US S336	-	-	SILV002	Dente (RI <sub>2</sub> )	-
	FZN 8 US S357	-	-	SILV003	Dente (LM <sub>2</sub> )	-
Marzabotto	-	-	-	MARZ001	Rocca petrosa	cal BC 760-390

**Tabella 5.8** Principali informazioni relative ai nuovi campioni aggiunti recentemente al dataset etrusco. I dati riguardanti la determinazione morfometrica del sesso e dell'età alla morte sono stati estratti da [155] per il sito di Monterenzio Vecchio e da [156,157] per la necropoli di Monte Tamburino.

## 5.2 Metodi di laboratorio: dalla fase di campionamento al sequenziamento

Come accennato in precedenza, le indagini molecolari sui resti umani inclusi nel presente progetto sono state svolte presso il Laboratorio di Antropologia e Paleogenetica dell'Università di Firenze ed il laboratorio dell'Archeo- and Paleogenetics Group dell'Università di Tubinga (Germania), entrambi dotati di apposite strutture (cleanroom) dedicate all'analisi dell'aDNA. Le diverse fasi sperimentali sono state condotte adottando tutte le precauzioni e seguendo tutte le procedure necessarie per minimizzare, e monitorare, il rischio di contaminazione con DNA moderno.

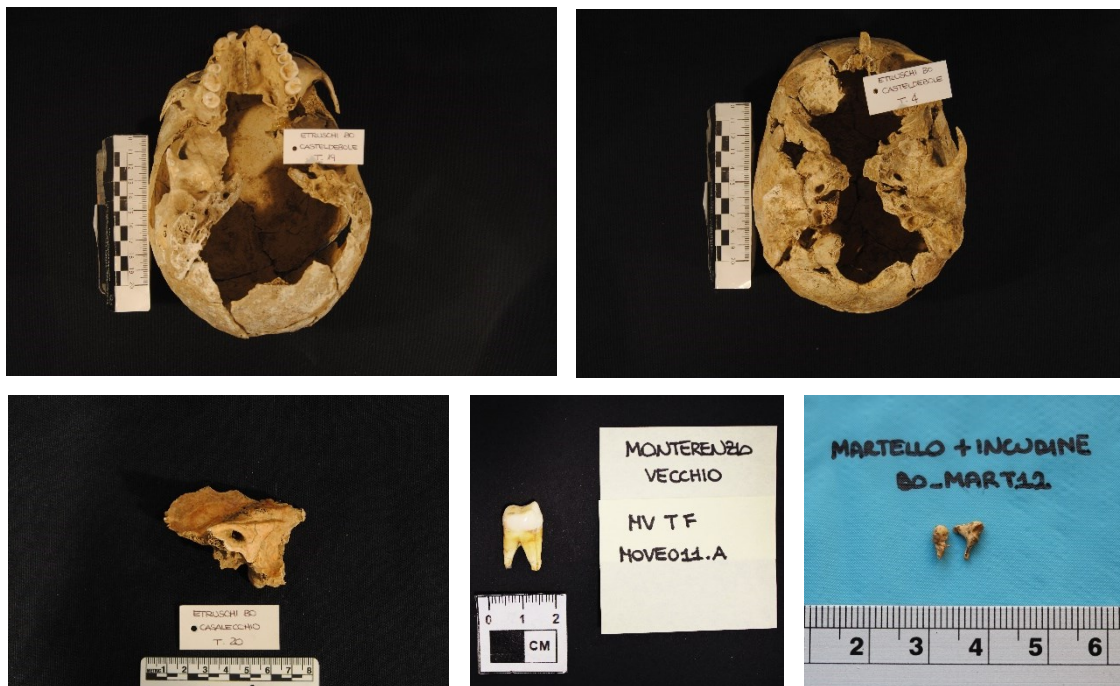
Per tutti i resti scheletrici è stata ottenuta l'autorizzazione a procedere con il campionamento di un quantitativo minimo di materiale di partenza necessario per l'estrazione del DNA antico; prima di procedere con il campionamento, tutti i materiali scheletrici pervenuti per l'analisi genomica sono stati documentati fotograficamente. I resti erano costituiti principalmente da rocche petrose isolate e, nel caso dell'ultimo set di campioni (Monterenzio Vecchio, Monte Tamburino/Monte Bibebe, San Silvestro), anche da denti isolati. Solo per i siti di Casteldebole e Giardini Margherita sono pervenuti in totale tre crani parzialmente restaurati (*Figura 5.6*); per l'individuo MAR021 è stato isolato l'osso temporale destro inclusivo dell'unica rocca petrosa preservata, per CAST001 è stato prelevato un dente (RM<sup>1</sup>, vista la conservazione della porzione mascellare) e, infine, per l'individuo CAST002 è stato isolato l'osso temporale sinistro comprensivo anche in questo caso della rocca petrosa.

Prima dell'effettivo campionamento i resti sono stati sottoposti ad un'iniziale fase di pulizia a secco volta alla rimozione dei residui di suolo più consistenti e ad un successivo *step* di decontaminazione mediante esposizione a radiazione ultravioletta (UV, alla lunghezza d'onda di 254 nm). Per quanto riguarda le rocche petrose, la pulizia del meato acustico esterno ha permesso di recuperare in alcuni casi l'incudine e/o il martello (ossicini dell'orecchio medio, insieme alla staffa) che, quando integri, sono stati utilizzati per la successiva fase di estrazione del DNA, dopo essere stati sottoposti a pulizia superficiale mediante acqua distillata ed esposizione a radiazione UV (*Figura 5.6*) [30]. Quando non presenti gli ossicini dell'orecchio, le rocche petrose sono state tagliate a metà con l'impiego di lame diamantate circolari impostate su un trapano dentistico al fine di esporre la porzione più densa del tessuto osseo, dalla quale sono stati ricavati circa 50 mg di polvere d'osso destinata all'estrazione del DNA (come descritto in [159]). La polverizzazione è stata effettuata mediante microfresce rotanti impostate su un trapano dentistico utilizzando una bassa velocità di rotazione per non surriscaldare l'area interessata dal campionamento.

Le rocche petrose processate a Tubinga (dai siti di Piazza VIII Agosto, Piazza Azzarita e Via Saffi) sono state campionate secondo un approccio metodologico differente: con un seghetto manuale sono stati effettuati due tagli longitudinali sul lato della rocca petrosa che in posizione anatomica rivolto

verso l'interno del cranio, consentendo in questo modo la rimozione di un piccolo cuneo di tessuto osseo e la conseguente esposizione di una porzione della regione densa da polverizzare. L'incisione è stata effettuata all'incirca all'altezza del meato acustico (collocato sul lato opposto della rocca petrosa) e al di sopra della fossa giugulare. Sono state campionate seguendo la stessa strategia anche le rocche petrose incluse del secondo set di campioni.

I denti sono stati processati, invece, andando a separare la corona dalla radice con l'impiego di un seghetto o di una lama diamantata circolare e polverizzando direttamente l'interno della camera pulpare per ottenere, anche in questo caso, circa 50 mg di polvere.



**Figura 5.6** Alcuni dei resti scheletrici pervenuti per l'analisi paleogenomica. In alto: due crani parzialmente restaurati provenienti dal sito di Casteldebole; in basso (da sinistra a destra): esempio di rocca petrosa isolata, di dente molare, e di ossicini dell'orecchio medio recuperati in alcuni casi in seguito alla pulizia del meato acustico esterno.

La polvere prodotta è stata utilizzata per l'estrazione del materiale genetico, condotta seguendo un protocollo appositamente sviluppato per il recupero di frammenti di DNA di ridotte dimensioni [22]; inoltre, durante ogni estrazione è stato inserito un controllo negativo per tracciare eventuali episodi di contaminazione o cross-contaminazione. Ciascun campione di polvere è stato incubato *overnight* in stufa a 37 °C in 1 mL di *extraction* buffer a base di EDTA (0.45M, pH 8.0) e contenente proteinasi K (0.25 mg/mL). Nei casi in cui erano stati ritrovati gli ossicini dell'orecchio, l'incudine o il martello sono stati direttamente immersi nell'*extraction* buffer saltando lo *step* di polverizzazione. Il giorno

seguinte, dopo aver centrifugato la sospensione, il surnatante è stato dapprima trasferito in un *binding* buffer preparato con guanidina cloridrato e successivamente in falcon dotate di colonnine con membrana in silice (High Pure Viral Nucleic Acid Large Volume Kit, Roche), come descritto in [160]. Dopo due cicli di lavaggio effettuati con un *wash* buffer a base di etanolo (High Pure Viral Nucleic Acid Large Volume Kit, Roche), il DNA è stato eluito in 100 µL totali di *elution* buffer (EB, 10 mM Tris-Cl, pH 8.5) attraverso due ripetizioni successive da 50 µL ciascuna.

La concentrazione degli estratti è stata misurata mediante Qubit 3.0 Fluorometer utilizzando il kit Qubit 1X dsDNA HS (High Sensitivity, Thermo Fischer Scientific).

In seguito, 30 µL di estratto sono stati utilizzati per la costruzione di librerie Illumina *double stranded* e *dual indexed* seguendo un protocollo adattato in modo specifico per il materiale genetico antico [161]. La preparazione delle librerie ha previsto un'iniziale fase di trattamento con l'enzima uracil-DNA glicosilasi (UDG-half) che ha permesso di rimuovere parzialmente i danni tipici del DNA antico, conservandoli solamente nei due nucleotidi terminali alle estremità 3' e 5' dei filamenti di DNA al fine di poter consentire, in un secondo momento, l'autenticazione del dato [162]. Per il campione MARZ001 era già disponibile una libreria preparata in precedenza senza trattamento UDG, per la quale erano già stati ottenuti i risultati di un primo sequenziamento *shotgun*.

Anche durante la preparazione delle librerie sono stati inseriti dei controlli negativi che, come quelli aggiunti nel corso dell'estrazione, sono stati portati avanti nelle successive fasi di analisi.

La doppia indicizzazione delle librerie è stata effettuata mediante l'aggiunta di combinazioni uniche di sequenze oligonucleotidiche ciascuna di lunghezza pari a 8 bp; l'*indexing* PCR è stata effettuata preparando una mix di reazione con l'enzima Pfu Turbo DNA Polymerase (Agilent Technologies) e impostando 15 cicli di amplificazione. I prodotti ottenuti da questa reazione sono stati sottoposti ad un passaggio di purificazione effettuato con l'impiego di colonnine MinElute (Qiagen), al termine del quale il DNA amplificato ed indicizzato è stato eluito in 20 µL di buffer EB.

Gli eluiti così ottenuti sono stati quantificati mediante Agilent 4150 TapeStation System (kit D1000) e ne sono stati valutati anche i profili a livello qualitativo. Nei rari casi in cui il profilo evidenziava la presenza di molecole eteroduplici, le librerie sono state sottoposte a 2 cicli aggiuntivi di PCR volti a favorire il distacco dei frammenti di materiale genetico unitisi in modo casuale in corrispondenza di porzioni complementari. La mix di reazione è stata preparata utilizzando l'enzima Herculase II Fusion DNA Polymerase (Agilent Technologies) e il set di primers IS5/IS6. Al termine della reazione è seguita un'ulteriore fase di purificazione con colonnine MinElute; il DNA è stato eluito in 20 µL di buffer EB e lo strumento Agilent 4150 TapeStation System (kit D1000) è stato nuovamente utilizzato per valutare la concentrazione del DNA e per verificare l'effettiva rimozione degli eteroduplici.



Terminati i passaggi appena descritti, le librerie sono state combinate in maniera equimolare in un unico *pool* di sequenziamento con concentrazione 2 nM per una prima fase di *screening* mirata alla valutazione del grado di preservazione del materiale genetico. La concentrazione del *pool* è stata valutata mediante Agilent 4150 TapeStation System (kit HS). Successivamente alle fasi preliminari di denaturazione e diluizione delle librerie, previste dal protocollo Illumina, il *pool* finale è stato sottoposto a sequenziamento *shotgun* su piattaforma Illumina NovaSeq 6000 in modalità *paired-end*, con un *output* pari a 5 milioni di *reads* per campione (2x50+8+8 cicli). Il sequenziamento è stato effettuato presso il Laboratorio di Genomica Avanzata del Dipartimento di Biologia dell'Università di Firenze.

I dati di sequenziamento ottenuti sono stati analizzati mediante il software EAGER (v.1.92.59) al fine di valutare la quantità e qualità del DNA endogeno preservato [163]. I campioni che mostravano un quantitativo di DNA endogeno superiore allo 0,1% e l'effettivo segnale di sostituzioni C-T e G-A corrispondentemente alle estremità 3' e 5' dei filamenti di DNA (a conferma dell'autenticità del dato genetico insieme alla ridotta dimensione dei frammenti) sono stati selezionati per il successivo *step* di analisi molecolare che ha previsto l'arricchimento delle librerie per ~1.24 milioni di SNPs distribuiti lungo il genoma nucleare umano, informativi per gli studi di genetica di popolazione.

Per tutti i campioni selezionati, la fase di arricchimento è stata effettuata presso il *modern lab* del gruppo di Tubinga, seguendo un consolidato protocollo di cattura in soluzione (1240K capture) [20]. Quando necessario, prima di essere sottoposte alla cattura degli SNPs nucleari, le librerie sono state amplificate (utilizzando l'enzima Herculase II Fusion DNA Polymerase e i primers IS5/IS6) allo scopo di raggiungere una concentrazione compresa tra 200 e 400 ng/μL. La concentrazione finale delle librerie è stata misurata mediante lo spettrofotometro NanoDrop (Thermo Fisher Scientific).

Per i campioni caratterizzati da una scarsa conservazione del materiale genetico endogeno (< 1%) sono stati eseguiti 2 *rounds* di cattura al fine di massimizzare l'efficienza della reazione.

I prodotti di cattura sono stati uniti in un unico *pool* 10 nM e sequenziati su piattaforma Illumina HiSeq 4000 in modalità *single-end* (1x76+8+8 cicli), con un *output* di 20 milioni di *reads* per ciascun campione, presso il Max Planck Institute for the Science of Human History di Jena (Germania).

I catturati dell'ultimo set di campioni sono stati sequenziati invece su piattaforma Illumina NovaSeq 6000 in modalità *single-end* (1x122+8+8) presso il Quantitative Biology Center (QBiC) dell'Università di Tubinga. Per questioni di tempistiche, le catture nucleari ed il loro successivo sequenziamento sono stati condotti per questi campioni in parallelo al sequenziamento *shotgun* (effettuato presso il Laboratorio di Genomica Avanzata di Firenze), senza attendere i risultati di quest'ultimo che sono però risultati fondamentali per valutare l'efficienza della reazione di arricchimento. In assenza delle

informazioni relative alla preservazione del DNA, tutti i campioni sono stati sottoposti a 2 *rounds* di cattura.

In tutti i casi in cui non è stato possibile ottenere una buona copertura sul genoma mitocondriale a partire dai dati di cattura nucleare, le librerie genomiche sono state arricchite in modo specifico per l' mtDNA seguendo il metodo descritto in [18]. Dopo essere stati amplificati, purificati e quantificati, i prodotti della cattura mitocondriale sono stati sequenziati su piattaforma Illumina NovaSeq 6000 in modalità *paired-end* e definendo un *output* di 1 milione di *reads* per ciascun campione (2x50+8+8 cicli, presso il Laboratorio di Genomica Avanzata).

La cattura dell' mtDNA non è ancora stata effettuata per il secondo set di campioni, per i quali sono stati recentemente generati i dati di cattura nucleare. Solo per il campione MARZ001 erano già stati in precedenza ottenuti, sequenziati ed analizzati i prodotti di cattura del genoma mitocondriale.

### 5.3 Analisi bioinformatica

L'analisi bioinformatica ha previsto un' iniziale fase di processamento dei dati grezzi generati dal sequenziamento *shotgun* al fine di effettuare uno *screening* preliminare dei campioni in esame.

Il processamento delle *reads* grezze è stato effettuato secondo le stesse modalità anche per i dati di cattura nucleare e mitocondriale, i quali sono stati successivamente utilizzati per la conduzione di specifiche analisi di carattere sia individuale che popolazionistico. Quest' ultime hanno previsto la determinazione del sesso biologico degli individui, la stima della contaminazione a livello nucleare e mitocondriale, lo studio dei marcatori uniparentali (DNA mitocondriale e cromosoma Y [chrY]), l'identificazione di eventuali legami di parentela, la valutazione delle runs of homozygosity (ROH) e, in particolare, la caratterizzazione genomica dei campioni tramite analisi di genetica di popolazione basate sugli SNPs nucleari selezionati in fase di cattura.

Come precisato nel precedente paragrafo, per l'ultimo set di campioni la produzione dei catturati ed il sequenziamento *shotgun* sono stati condotti in parallelo al fine di velocizzare la generazione dei dati di cattura nucleare. I dati *shotgun* sono stati comunque prodotti allo scopo di calcolare e valutare l'efficienza della reazione di arricchimento.

### 5.3.1 Processamento dei dati grezzi

Una volta effettuato il *demultiplexing*, i dati grezzi generati dal sequenziamento (in formato FASTQ compresso) sono stati analizzati mediante una *pipeline* specifica per il DNA antico integrata nel *tool* EAGER (v.1.92.59) [163]. La qualità di sequenziamento è stata valutata per le *reads* prodotte per ciascun campione mediante il modulo FastQC. Attraverso AdapterRemoval (v2.2.0) [164] sono state rimosse le sequenze degli adattatori e, in caso di sequenziamento in *paired-end*, è stato effettuato il *merging* delle *reads forward* e *reverse* (R1 ed R2) impostando una sovrapposizione minima di 10 bp, in modo tale da scartare tutte le sequenze più lunghe, potenzialmente contaminanti. Inoltre, sono state eliminate le *reads* di lunghezza inferiore alle 30 bp dal momento che avrebbero potuto mappare in modo aspecifico sul genoma di riferimento.

Dopo questa preliminare fase di filtraggio, le sequenze derivanti dal sequenziamento *shotgun* e dal sequenziamento delle catture nucleari sono state mappate sul genoma di riferimento umano (hg19) mediante BWA (v.0.7.12) [165], disabilitando l'opzione di *seeding* e impostando i parametri *-n* 0,01 (indicativo del numero di *mismatches* permessi in fase di mappaggio tra le *reads* e la sequenza di riferimento) e *-q* 30 (corrispondente alla soglia minima definita per la qualità di mappaggio).

Al fine di compiere successive analisi sull' mtDNA, le *reads* generate dalle catture nucleari sono state mappate anche sulla sequenza di riferimento mitocondriale umana rCRS (NC\_012920.1) mediante CircularMapper, un *tool* specificamente sviluppato per il mappaggio su genomi circolari [163].

Anche i dati prodotti a partire dalle catture mitocondriali sono stati mappati allo stesso modo sulla sequenza rCRS, dopo essere stati precedentemente processati con AdapterRemoval secondo le specifiche riportate sopra.

Una volta effettuato il mappaggio, i duplicati di PCR sono stati rimossi mediante il software DeDup (v0.12.2)[163], il quale identifica le molecole clonali tenendo in considerazione le coordinate sia di inizio che di fine delle sequenze.

Le *reads* mappanti sulle sequenze di riferimento sono state analizzate tramite mapDamage2.0 [34] al fine di valutare l'effettiva presenza dei tipici *patterns* di degrado che caratterizzano il DNA antico: ridotta dimensione dei frammenti e aumento del tasso di misincorporazioni nelle porzioni terminali delle molecole di DNA.

L'analisi dei dati *shotgun* ha previsto il lancio della sola *pipeline* di EAGER, sufficiente per valutare il grado di preservazione del DNA endogeno e la sua autenticità attraverso le statistiche restituite da MapDamage; i dati genomici derivanti dalla cattura degli SNPs nucleari (e dell' mtDNA) sono stati ulteriormente analizzati attraverso le modalità descritte nei prossimi paragrafi.

Prima di procedere con le successive fasi di analisi, i files .bam prodotti dal mappaggio sull'intero genoma umano sono stati processati andando a mascherare i 2 nucleotidi terminali di entrambe le estremità di ciascuna *reads* mediante TrimBam, incluso nel pacchetto BamUtil [166]. Tale passaggio (*trimming*) risulta fondamentale al fine di ridurre l'impatto delle misincorporazioni conservate alle estremità delle sequenze (dovute al parziale trattamento UDG) nelle analisi successive. Nel caso del campione MARZ001, data l'elevata incidenza di deaminazioni anche nelle porzioni più interne delle molecole di DNA (derivanti almeno in parte dall'assenza di trattamento UDG), è stato effettuato un *trimming* di 10 bp.

### 5.3.2 Determinazione del sesso e stima della contaminazione

La determinazione del sesso degli individui, a livello genetico, è stata effettuata a partire dai files .bam precedentemente sottoposti allo *step* di *trimming*. Il sesso è stato determinato sulla base dei valori normalizzati ottenuti dal calcolo della X-ratio (rapporto tra il numero di SNPs *target* coperti sul cromosoma X e il numero di SNPs *target* coperti sugli autosomi) e Y-ratio (rapporto tra il numero di SNPs *target* coperti sul cromosoma Y e il numero di SNPs *target* coperti sugli autosomi). I risultati attesi per campioni non contaminati corrispondono a valori di X-ratio  $\sim 1$  ed Y-ratio  $\sim 0$  di per gli individui di sesso femminile e valori di X-ratio ed Y-ratio  $\sim 0,5$  per gli individui di sesso maschile. Ai campioni che cadono in un intervallo intermedio viene assegnato sesso indeterminato; solitamente tale risultato è indicativo della presenza di DNA contaminante.

Per i campioni di sesso maschile è stato possibile effettuare una stima della contaminazione a livello nucleare andando a misurare il grado di eterozigosi osservato in siti polimorfici del cromosoma X, dopo aver valutato il tasso di incidenza di eventuali errori di sequenziamento nelle regioni adiacenti alle posizioni considerate, che dovrebbero invece essere monomorfiche. Data la presenza di un solo cromosoma X negli individui di sesso maschile, l'evidenza di un significativo tasso di eterozigosi è di norma segnale di contaminazione. Per questa analisi è stato utilizzato il programma ANGSD (0.935) [167] ed è stata presa in considerazione la stima restituita dal "method1 new version", effettuata sulla base del "Methods of Moments" (MoM).

Tutti i campioni maschili con contaminazione inferiore al 6% e con almeno 100 SNPs coperti 2 volte sul cromosoma X sono stati considerati idonei per essere sottoposti alle seguenti fasi di analisi.

Sia gli individui maschili che femminili sono stati sottoposti alla stima della contaminazione a livello mitocondriale mediante il *tool* schmutzi [168], il quale ricostruisce attraverso un processo iterativo la sequenza consenso endogena dell'mtDNA (tenendo in considerazione i *patterns* di deaminazione e la distribuzione della dimensione dei frammenti) e stima, al contempo, il livello di contaminazione

umana mediante il confronto con un database di riferimento di frequenze alleliche rappresentative della moderna variabilità mitocondriale euroasiatica (possibili fonti di contaminazione). Il comando `ContDeam.pl` è stato lanciato con le opzioni `--library double` e `--lengthDeam 2` (default), in accordo con la tipologia di librerie genomiche preparate; per il comando `schmutzi.pl` sono stati utilizzati, invece, i parametri `--notusepredC` e `--uselength`.

Nei casi in cui i dati di cattura nucleare sono risultati insufficienti per ottenere una stima affidabile della contaminazione (a causa della bassa copertura sull' mtDNA), quest'ultima è stata determinata a partire dai dati ottenuti dalla cattura specifica del genoma mitocondriale.

Per gli individui con un basso rapporto tra le sequenze mappanti sul genoma mitocondriale e quelle mappanti sul genoma nucleare ( $mt/nuclear\ ratio < 200$ , statistica ottenuta da EAGER), la stima della contaminazione mitocondriale è stata considerata indicatore affidabile del grado di contaminazione nucleare [169]. La correlazione osservata tra  $mt/nuclear\ ratio$  e contaminazione è particolarmente stabile nel caso in cui il DNA provenga da rocche petrose, motivo per cui il rapporto  $mt/nuclear$  è stato valutato solo per gli individui per i quali era stato campionato tale distretto scheletrico [169]. Campioni con contaminazione mitocondriale maggiore del 5% sono stati esclusi dalle successive indagini, ad eccezione degli individui maschili per i quali non è stata riscontrata contaminazione a livello nucleare.

### 5.3.3 Markers uniparentali: assegnazione degli aplogruppi

Come già accennato, sia per la stima della contaminazione mitocondriale che per la ricostruzione della sequenza consenso sono stati utilizzati i files in formato `.bam` derivanti dal mappaggio dei dati di cattura (nucleare e, quando necessario, mitocondriale) sulla sequenza di riferimento rCRS.

La ricostruzione della sequenza consenso mitocondriale è stata effettuata utilizzando il comando `log2fasta` del pacchetto `schmutzi` [168], applicando filtri di qualità crescente ( $q10$ ,  $q20$ ,  $q30$ ). I files `.fasta` generati, contenenti la sequenza consenso determinata da `schmutzi` secondo i diversi criteri di qualità imposti, sono stati processati mediante la versione online del *tool* Haplogrep2 (v.2.4.0, basato su Phylotree 17) [170, 171] per ottenere l'assegnazione dell'aplogruppo mitocondriale.

L'aplogruppo del cromosoma Y è stato definito, invece, utilizzando il software Yleaf (v2.2) [172] che si basa su una lista di posizioni marker localizzate sul chrY riportate nel database ISOGG (v.15.73) [173]. Per l'analisi sono stati considerati soltanto i siti marker coperti almeno 2 volte ( $-r\ 2$ ) e le *reads* con qualità di mappaggio pari o superiore a 30 ( $-q\ 30$ ). La chiamata per un determinato sito è stata effettuata solo quando supportata da almeno il 90% delle sequenze ( $-b\ 90$ ). Per gli individui con

scarsa copertura sul chrY, è stato possibile ottenere l'assegnazione dell'aplogruppo impostando il parametro `-r 1` e adottando, quindi, un approccio meno stringente.

L'attribuzione automatica di Yleaf è stata confermata attraverso un controllo manuale e verificata tramite l'uso di un ulteriore software dedicato allo scopo, pathPhynder [174].

#### 5.3.4 Genotyping e costruzione del dataset di analisi

Ciascun individuo è stato genotipizzato mediante pileupCaller [175] con una chiamata random di tipo pseudo-aploide effettuata sul pannello 1240K (comprensivo degli SNPs nucleari selezionati in fase di cattura) e sul pannello Human Origins (HO, contenente la metà delle posizioni presenti nel 1240K) a partire dai files .bam trimmati. La chiamata è stata condotta restringendo ad una qualità di mappaggio pari o superiore a 30 (`-Q 30`); inoltre, sono state prese in considerazione solo le basi con qualità (*Phred quality score*) anche in questo caso pari o superiore a 30 (`-q 30`). Il parametro `-B` è stato inserito allo scopo di disabilitare il meccanismo di ricalibrazione della qualità delle basi, non consigliato in presenza di dati a basso coverage (come quelli prodotti da DNA antico).

Una volta effettuato il *genotyping*, per ciascun campione è stato calcolato il numero di SNPs coperti su entrambi i pannelli di riferimento.

I dati genotipizzati, in formato eigenstrat (composto dai tre files .geno, .snp, .ind, rispettivamente contenenti le informazioni relative ai genotipi, agli SNPs del pannello in uso e agli individui tipizzati), sono stati uniti ai datasets pubblici HO e 1240K (entrambi v.44.3, composti da 597.573 e 1.233.013 SNPs rispettivamente) [176] tramite il comando `mergeit` del pacchetto EIGENSOFT (v.7.0.2) [177, 178].

Ai datasets sono stati aggiunti anche i dati di *genotyping* relativi ad individui antichi di interesse non ancora inclusi nella versione in uso, estrapolati direttamente dalle rispettive pubblicazioni [63, 107, 110, 114, 117, 118, 142, 179, 180].

#### 5.3.5 Determinazione delle parentele e calcolo delle runs of homozygosity

I dati di *genotyping*, trasformati in formato PED, sono stati utilizzati per individuare possibili legami biologici tra gli individui, in particolare tra quelli provenienti dalla medesima necropoli. L'analisi è stata effettuata tramite il software READ [181], dopo aver selezionato con PLINK (v.1.9) [182] i soli SNPs localizzati sugli autosomi. READ effettua l'identificazione delle parentele dividendo il genoma in finestre non sovrapposte di 1 Mbps e calcolando, all'interno di ciascuna di esse, la proporzione

di alleli non condivisi tra coppie di individui (P0, il cui valore normalizzato consente di classificare secondo i diversi gradi le parentele riscontrate).

Per confermare i risultati ottenuti, l'identificazione dei legami di parentela è stata effettuata anche calcolando (indipendentemente da READ) il tasso di *mismatches* (PMR, pairwise mismatch rate) tra coppie di individui, prendendo in considerazione tutti gli SNPs autosomici coperti almeno una volta in entrambi i campioni analizzati [183]. I valori di PMR ottenuti sono stati poi utilizzati per calcolare il coefficiente di relazione ( $\alpha$ ) tra le coppie di individui:  $\alpha = [(PMR \text{ non relazionati} - PMR \text{ calcolato}) / PMR \text{ identici}]$  che corrispondono, in ordine, al PMR calcolato per individui non relazionati tra loro, a quello relativo alla coppia di individui in esame e, infine, a quello associato a due campioni identici (due gemelli o due campioni riferibili allo stesso individuo). Il "PMR non relazionati" corrisponde al valore medio di tutti i PMR calcolati dal software nell'intera popolazione considerata, mentre "PMR identici" corrisponde alla metà di quest'ultimo. Per coppie con valori di  $\alpha \sim 0,5$  o  $0,25$ , gli individui in questione sono stati classificati, rispettivamente, come relazionati di primo e secondo grado.

I risultati ottenuti relativamente alle parentele sono stati ulteriormente convalidati effettuando un controllo degli aplogruppi dell'mtDNA e del chrY assegnati in precedenza.

Per evitare di introdurre *bias* nelle analisi di genetica di popolazione, nei casi di rapporti di parentela di primo o secondo grado, solo l'individuo con il più elevato numero di SNPs coperti è stato incluso nelle fasi di analisi successive.

È stata effettuata, inoltre, un'analisi delle *runs of homozygosity* (ROH), porzioni genomiche contigue in omozigosi derivanti dall'aver ereditato dai propri genitori aplotipi identici. Più recente è il legame biologico esistente tra i genitori di un dato individuo, maggiori saranno le dimensioni e la frequenza delle ROH osservate. In particolare, ROH di lunghe dimensioni (>20 Cm) sono indicative del diretto legame di parentela tra i genitori, mentre ROH di dimensioni minori (4-8 Cm) sono correlate alla cosiddetta *background relatedness* e forniscono informazioni circa la dimensione della popolazione a cui appartiene l'individuo in esame [184]. Tale analisi, effettuata a partire dai dati di *genotyping*, è stata condotta con il programma hapROH [184], appositamente sviluppato per un'identificazione affidabile delle ROH anche per campioni a basso *coverage* e da dati pseudo-aploidi; hapROH si basa su un pannello di riferimento di 5008 aplotipi rappresentativi dell'attuale variabilità genetica umana (1000 Genomes [185]). L'analisi è stata condotta per gli individui con almeno 400.000 SNPs coperti sul pannello 240K.

### 5.3.6 Analisi delle componenti principali e ADMIXTURE

il comando `smartpca`, incluso nel pacchetto EIGENSOFT (v. 7.0.2) [177, 178], è stato adottato per l'analisi delle componenti principali (PCA), costruita a partire dai dati genomici (genotipizzati sul pannello HO) di 62 popolazioni moderne informative dell'attuale variabilità genetica dell'Eurasia occidentale. Per l'analisi sono state utilizzate le opzioni “`lsqproject: YES`”, suggerita nel momento in cui vengono proiettati sulle prime componenti principali (PCs) campioni con elevata percentuale di dato mancante (come nel caso di individui antichi), e “`shrinkmode: YES`”, che consente di correggere l'effetto di stiramento degli assi prodotto dai campioni moderni utilizzati per il calcolo delle PCs (e la conseguente tendenza degli individui antichi a cadere nell'area centrale della PCA) [177].

Una volta costruito lo spazio di variabilità genetica, gli individui in esame con almeno 20.000 SNPs coperti sul pannello HO ed ulteriori campioni antichi di interesse disponibili in letteratura sono stati proiettati sugli assi -PC1. A causa dallo scarso numero di SNPs coperti il campione AZZA012 è stato escluso dalla PCA e, in generale, dallo studio popolazionistico; sono stati esclusi dalla PCA per la bassa copertura anche i campioni MOTE008, MOVE005, MOVE006, MOVE011 e tutti quelli del sito di San Silvestro. Il grafico della PCA è stato ottenuto in RStudio utilizzando la funzione `ggplot`.

Inoltre, al fine di stimare, secondo un approccio di *maximum likelihood*, i diversi contributi genetici (*ancestries*) caratterizzanti il profilo degli individui in esame, è stata lanciata un'analisi sul pannello HO attraverso l'algoritmo ADMIXTURE (v.1.3.0), in modalità *unsupervised* [186]. Prima dell'effettiva analisi, il dataset HO è stato ristretto alle sole popolazioni dell'Eurasia occidentale; in aggiunta, sono stati rimossi dal dataset tutti gli individui con una percentuale di dato mancante superiore al 99%, gli individui legati da rapporti di parentela e tutti i campioni sardi del periodo pre-nuragico (per evitare che venisse massimizzata una componente neolitica ad essi correlata a basse K) [116, 187]. L'analisi è stata quindi condotta su un totale di 6159 individui moderni ed antichi ed ha riguardato solo gli SNPs trattenuti in seguito ad un filtraggio, effettuato con PLINK (v.1.9) [182], basato sul calcolo del grado di correlazione esistente tra coppie di SNPs. Tale *step*, definito *pruning*, è stato effettuato per rimuovere gli SNPs in *linkage disequilibrium* impiegando i parametri `--indep-pairwise 200 25 0.2`, dove 200 bp indica l'ampiezza delle finestre considerate, con uno scorrimento di 25 bp, e 0,2 indica la soglia massima accettata per il parametro  $R^2$  affinché gli SNPs vengano trattenuti nell'analisi.

L'analisi è stata lanciata definendo un numero di K (componenti) compreso tra 3 e 16, per ciascuna delle quali è stato poi valutato il *cross validation error* (CV error) restituito dall'analisi. Il *plot* è stato creato con la funzione `barplot` di R in RStudio a partire dai dati ottenuti da K=13, quella che mostrava



il minor *CV error* e in cui *Iran\_Neolithic* e *steppe-related ancestry* venivano distinte in due diverse componenti.

### 5.3.7 Raggruppamento in clusters

Sulla base dei *patterns* osservati dall'analisi delle componenti principali, sono state condotte analisi aggiuntive volte alla definizione, quando possibile, dei clusters genetici. In primo luogo, sono state calcolate delle statistiche  $F_4$  sul pannello 1240K utilizzando la funzione qpDstat (modalità "f4 mode: YES") del pacchetto ADMIXTOOLS (v.7.0.2) [188, 189].

Le  $F_4$ -statistics sono state condotte nella forma  $f_4(\text{Onge.DG}, X; \text{Individuo}, \text{Cluster})$  dove Onge.DG (N=2) è l'outgroup, mentre  $X$  è rappresentato dai *cornerstones* della variabilità genetica dell'Eurasia occidentale: WHG [che include Spain\_HG\_published.SG, Luxembourg\_Loschbour\_published.DG e Hungary\_EN\_HG\_Koros\_published.SG (N=3)], EHG [che include Russia\_EHG, Russia\_HG\_Karelia, Russia\_HG\_Samara e Russia\_Popovo\_HG (N=6)], Russia\_Samara\_EBA\_Yamnaya (N=10), Turkey\_N (N=24), Iran\_GanjDareh\_N (N=8), Morocco\_EN.SG (N=3), Morocco\_Iberomaurusian (N=6); *Cluster* include gli individui appartenenti allo stesso intervallo temporale (Bologna\_Etruscan per l'età del Ferro e Bologna\_Imperial per il periodo imperiale) che in PCA risultano strettamente relazionati a livello genetico (formano un cluster), mentre *Individuo* rappresenta i campioni che, per la posizione che occupano, sono stati identificati come possibili *outliers*. Ciascun *Individuo* è stato considerato un *outlier* in presenza di almeno una paragone con  $Z\text{score} > |3|$  (è stata prestata attenzione anche ai segnali osservati per  $Z\text{score} > |2.5|$ ) e, quindi, trattato separatamente (come individuo singolo) nella successiva analisi.

Per confermare i risultati delle statistiche  $F_4$  e per meglio valutare se alcuni *outliers* potessero essere raggruppati in un unico cluster, è stata effettuata un'analisi con il programma qpWave, incluso nel pacchetto ADMIXTOOLS [188, 189]. Gli individui sono stati testati a coppie per valutare se fossero relazionati tra loro in modo simmetrico o differente rispetto ad un set di popolazioni di riferimento [Mbuti.DG, Russia\_Ust\_Ishim.DG, CHG (che include Georgia\_Kotias.SG e Georgia\_Satsurblia.SG), EHG, Spain\_ElMiron, Czech\_Vestonice16, Russia\_MA1\_HG.SG, Israel\_Natufian, Jordan\_PPNB, Turkey\_N, WHG, Iran\_GanjDareh\_N, Russia\_Samara\_EBA\_Yamnaya, Morocco\_LN.SG], ovvero se potessero costituire o meno un clade (vedi [109, 110, 116]). Gli individui sono stati raggruppati sulla base del parametro p-value restituito dall'analisi, definendo la soglia p-value  $> 0.01$ . L'analisi è stata condotta sul pannello 1240K utilizzando le opzioni "allsnps: YES", al fine di massimizzare il numero di SNPs coinvolti in ciascun confronto a coppie [190], e "inbreed: NO"; nei test effettuati sono stati inclusi, per confronto, anche i principali clusters genetici di interesse descritti nel precedente lavoro

sull'Etruria classica [117]. I risultati ottenuti da qpWave sono stati convertiti graficamente in una *heatmap* creata mediante la funzione `heatmap.2` di R.

L'analisi ha permesso di evidenziare l'affinità genetica tra alcuni individui riferibili all'età del Ferro che si discostano per il loro profilo dal cluster principale Bologna\_Etruscan. Tali individui sono stati riuniti in un cluster denominato Bologna\_Etruscan.Ceu (dove Ceu abbrevia Central Europe), rispetto al quale i campioni AGOS001 e AZZA006 (datati alla prima età del Ferro) risultano essere *outliers*. Allo scopo di confermare e ulteriormente indagare questo segnale è stata effettuata la statistica  $f_4(\text{Onge.DG}, X; \text{Outlier}, \text{Bologna\_Etruscan.Ceu})$  in cui  $X$  è rappresentato dalle diverse componenti distali riportate ad inizio paragrafo e *Outlier* alternativamente dagli individui AGOS001 e AZZA006; in aggiunta, sono stati ripetuti i test con il programma qpWave confrontando le coppie AGOS001-Bologna\_Etruscan.Ceu, AZZA006-Bologna\_Etruscan (considerando prima l'intero cluster e poi solo gli individui con datazione al radiocarbonio) e, infine, AZZA006-Bologna\_Etruscan.Ceu.

Lo stesso confronto con qpWave è stato effettuato anche per CAS090-Bologna\_Imperial e CAS095-Bologna\_Imperial per indagare alcuni *patterns* ottenuti dalla prima analisi con qpWave.

Sulla base dei modelli ottenuti con qpWave il campione TAB016 non rientra in nessuno dei gruppi genetici definiti in precedenza, mentre il campione CAS054 si conferma un *outlier* rispetto al cluster rappresentativo del periodo imperiale (Bologna\_Imperial).

I campioni che per le loro caratteristiche genetiche non rientrano in nessuno dei clusters identificati sono stati considerati come singoli individui nelle successive analisi.

### 5.3.8 F-statistics

La presenza di eventuali segnali di *admixture* tra il cluster Bologna\_Etruscan e 562 popolazioni antiche dell'Eurasia occidentale e dell'Africa settentrionale (datate tra il Mesolitico e il periodo della Roma imperiale e precedentemente selezionate dal dataset di analisi), è stata verificata tramite il calcolo di statistiche  $F_3$  sul pannello 1240K (con il comando `q3Pop` di ADMIXTOOLS [188, 189]).

Le statistiche sono state condotte nella forma  $f_3(\text{Bologna\_Etruscan}, \text{Popolazioni antiche}; \text{Test})$  dove *Test* corrisponde a Bologna\_Etruscan.Ceu o Bologna\_Imperial, testati separatamente per valutare se il profilo genetico ad essi associato potesse derivare da possibili eventi di flusso genico avuti luogo tra Bologna\_Etruscan ed una o più popolazioni antiche tra quelle selezionate. L'analisi è stata condotta impostando il parametro "inbreed: YES" e sono state considerate significative statistiche negative con  $Z\text{score} < -3$ . Nel caso del cluster Bologna\_Etruscan.Ceu, è stata effettuata anche la statistica nella forma  $f_3(\text{Italy\_Broion\_BA}, \text{Popolazioni antiche}; \text{Bologna\_Etruscan.Ceu})$  nell'ipotesi di un fenomeno di *admixture* meno recente che possa aver coinvolto popolazioni dell'età del Bronzo,

ovvero del periodo precedente a quello in esame (sono stati utilizzati come *proxy* i dati genomici di individui provenienti dal sito Grottina dei Covoloni del Broion, in Italia settentrionale, alcuni dei quali datati ad una fase iniziale all'età del Bronzo) [142]).

Sono state effettuate, inoltre, delle statistiche  $F_4$  sempre sul dataset 1240K utilizzando il comando qpDstat (modalità "f4 mode: YES") del pacchetto ADMIXTOOLS [188, 189]. Per misurare il grado di affinità genetica tra il cluster principale Bologna\_Etruscan, associato all'area dell'Etruria padana, ed il gruppo C.Italy\_Etruscan, comprendente gli individui di cultura etrusca con profilo genetico locale relativi alle zone dell'Etruria classica e riportati in Posth *et al.* [117]), è stata costruita la statistica  $f_4(\text{Onge.DG}, X; \text{C.Italy\_Etruscan}, \text{Bologna\_Etruscan})$  in cui  $X$  è rappresentato, di volta in volta, da una delle popolazioni antiche *cornerstones* della variabilità genetica dell'Eurasia occidentale elencate nel paragrafo precedente. La statistica è stata ripetuta una seconda volta trattenendo nei clusters C.Italy\_Etruscan e Bologna\_Etruscan esclusivamente gli individui con datazione al radiocarbonio (rispettivamente  $N=21$  e  $N=3$ ). Ulteriori statistiche con costruzione simile sono state effettuate per valutare l'affinità tra alcuni clusters o per evidenziare le componenti distali che meglio descrivono le differenze genetiche osservate tra quest'ultimi. Sono state quindi effettuate le seguenti analisi:  $f_4(\text{Onge.DG}, X; \text{C.Italy\_Etruscan.Ceu}, \text{Bologna\_Etruscan.Ceu})$ ,  $f_4(\text{Onge.DG}, X; \text{Bologna\_Etruscan.Ceu}, \text{Bologna\_Etruscan})$  e  $f_4(\text{Onge.DG}, X; \text{C.Italy\_Imperial}, \text{Bologna\_Imperial})$  dove C.Italy\_Etruscan.Ceu e C.Italy\_Imperial corrispondono ai clusters genetici descritti in Posth *et al.* [117].

È stato possibile confermare che Bologna\_Etruscan e *Test* non costituiscono un clade mediante la statistica  $f_4(\text{Onge.DG}, \text{Popolazioni antiche}; \text{Test}, \text{Bologna\_Etruscan})$  dove *Popolazioni antiche* indica i 562 gruppi selezionati in precedenza dal dataset 1240K per le statistiche  $f_3$ , e *Test* è rappresentato alternativamente dagli individui AGOS001, AZZA006, TAB016 e dai clusters Bologna\_Etruscan.Ceu e Bologna\_Imperial. In maniera simile sono state calcolate le statistiche  $f_4(\text{Onge.DG}, \text{Popolazioni antiche}; \text{AGOS001/AZZA006}, \text{Bologna\_Etruscan.Ceu})$  e  $f_4(\text{Onge.DG}, \text{Popolazioni antiche}; \text{CAS054}, \text{Bologna\_Imperial})$ . Per quest'ultimo confronto sono stati aggiunti alla lista di popolazioni antiche 34 gruppi/individui provenienti dall'Africa subsahariana, datati ad un periodo precedente rispetto a CAS054, e pubblicati in [62, 63, 191-193].

In aggiunta, la statistica  $f_4(\text{Onge.DG}, \text{Popolazioni antiche}; \text{AGOS001}, \text{Italy\_North\_BellBeaker\_2})$  è stata effettuata al fine di individuare il gruppo antico con maggior affinità genetica con AGOS001 (il campione più spostato in PCA verso il nord del cline europeo) rispetto a Italy\_North\_BellBeaker\_2, individuo di cultura *Bell Beaker*, proveniente dal nord Italia e datato al 2000 a.C. circa, per il quale non è stata riscontrata la presenza di *steppe-related ancestry*, a differenza di altri due campioni dallo stesso sito e periodo (Italy\_North\_BellBeaker\_1 e Italy\_North\_BellBeaker\_3) [106].

Questa analisi è stata condotta per meglio localizzare a livello geografico e temporale la possibile sorgente di *steppe-ancestry* che avrebbe contribuito al profilo genetico del campione AGOS001 (e verosimilmente anche del gruppo Bologna\_Etruscan.Ceu), utilizzando come termine di paragone un campione dell'età del Bronzo proveniente dall'Italia settentrionale, non ancora interessato da *admixture* con gruppi associati alla componente della steppa.

Inoltre, per confermare alcune evidenze già parzialmente riscontrate nella corrispondente statistica  $F_4$  distale, C.Italy\_Imperial e Bologna\_Imperial sono stati confrontati mediante il test  $f_4$ (Onge.DG, *Popolazioni antiche*; Bologna\_Imperial, C.Italy\_Imperial). In questo modo è stato possibile verificare se alcune delle popolazioni antiche incluse nell'analisi fossero geneticamente più affini ad uno dei due cluster del periodo imperiale; ciò ha permesso di individuare, in sostanza, i contributi genetici che contraddistinguono i due gruppi in questione.

È importante aggiungere, infine, che nella costruzione delle statistiche  $F_4$  è stata prestata attenzione a minimizzare possibili *bias* derivanti dalla differente tipologia di dati utilizzati in ciascun confronto (*shotgun* o derivanti dalla cattura di SNPs nucleari).

### 5.3.9 qpAdm modelling

Il software qpAdm (del pacchetto ADMIXTOOLS [188, 189]) è stato utilizzato sul pannello 1240K per verificare l'ipotesi che due gruppi fossero equamente relazionati rispetto ad un set di popolazioni di riferimento, definite *right populations* ( $p\text{-value} > 0.05$ ) o per testare modelli di *admixture* tra due popolazioni sorgente, chiamate *left populations* (*two-way models*, considerati validi con  $p\text{-value} > 0.05$ ). Per l'analisi sono state impostate le opzioni "allsnps: YES" e "inbreed: NO"; il set principale di *right populations* utilizzato nei test si compone dei seguenti gruppi/individui disponibili nel dataset 1240K (già selezionati in fase di *grouping* per i modelli qpWave): Mbuti.DG, Russia\_Ust\_Ishim.DG, CHG, EHG, Spain\_ElMiron, Czech\_Vestonice16, Russia\_MA1\_HG.SG, Israel\_Natufian, Jordan\_PPNB, Turkey\_N, WHG, Iran\_GanjDareh\_N, Russia\_Samara\_EBA\_Yamnaya, Morocco\_LN.SG (CHG, EHG e WHG includono gli stessi individui riportati nel *Paragrafo 5.3.7*).

Per ciascun individuo/cluster dell'età del Ferro precedentemente identificato (Bologna\_Etruscan, Bologna\_Etruscan.Ceu, AGOS001, AZZA006, TAB016,) sono stati verificati, in primo luogo, modelli *one-way* che prevedono una derivazione diretta dei campioni in esame da una singola popolazione contemporanea o antecedente, presupponendo quindi una condizione di continuità genetica con quest'ultima. I test sono stati effettuati considerando 365 popolazioni dell'Eurasia occidentale e dell'Africa settentrionale datate tra il Neolitico medio e l'età del Ferro.

Il confronto tra Bologna\_Etruscan e C.Italy\_Etruscan è stato ripetuto selezionando da entrambi i gruppi solamente gli individui per i quali era disponibile la datazione al radiocarbonio.

In seguito, gli individui/clusters dell'età del Ferro in esame sono stati modellati come derivanti da fenomeni di admixture tra due gruppi distinti (*2-way models*): il primo rappresentato da individui provenienti dall'Italia continentale o insulare precedentemente pubblicati e datati tra il Neolitico e l'età del Bronzo [101, 106, 115-117, 142, 187, 194], ed il secondo da popolazioni dell'età del Bronzo associate alla diffusione della componente della steppa (alternativamente Germany\_BellBeaker e Russia\_Samara\_EBA\_Yamnaya, quest'ultima rimossa dal set di popolazioni di riferimento quando adottata come popolazione sorgente).

In aggiunta, per il cluster Bologna\_Etruscan.Ceu e gli individui AGOS001 ed AZZA006 è stata testata l'ipotesi di admixture tra Italy\_Broion\_BA (popolazione rappresentativa del *pool* genetico osservato in Italia settentrionale durante l'età del Bronzo, comprendente anche individui datati alla sua fase iniziale) e 362 popolazioni antiche corrispondenti a quelle selezionate per i precedenti modelli *one-way* ad eccezione di Italy\_North\_Remedello\_EBA.SG e Italy\_North\_Remedello\_C.SG, aggiunti alle *right populations* sotto il *label* Italy\_North\_Remedello\_C-EBA.SG, e Italy\_Broion\_BA utilizzata come popolazione sorgente. Analogamente, anche per i campioni Bologna\_Imperial e CAS054 sono stati valutati diversi modelli costruiti per testare le ipotesi di *admixture* tra il cluster Bologna\_Etruscan (rappresentativo del periodo precedente) e le 362 popolazioni menzionate sopra; per CAS054 sono stati inclusi nei test anche i gruppi di origine subsahariana già precedentemente considerati per le statistiche  $F_4$  (Paragrafo 5.3.8).

L'individuo TAB016 è stato momentaneamente escluso dal test descritto sopra in attesa dei risultati di un secondo tentativo di datazione al radiocarbonio; nel caso di questo campione, infatti, il dato temporale risulta indispensabile per una corretta costruzione ed interpretazione dei modelli.

Sulla base della sua posizione in PCA e dei segnali evidenziati dalle analisi condotte in precedenza, per il campione AZZA006 è stato testato anche un modello di *admixture* tra i 2 clusters dell'età del Ferro (Bologna\_Etruscan e Bologna\_Etruscan.Ceu), a seguito del quale è stata valutata la possibilità di un modello 1-way valido in presenza del solo gruppo Bologna\_Etruscan.Ceu, anche dopo aver spostato Bologna\_Etruscan dalle *left* alle *right populations*.

In aggiunta, i cluster genetici Bologna\_Etruscan, Bologna\_Etruscan.Ceu e Bologna\_Imperial sono stati modellati in modo tale da essere caratterizzati da massimo 5 componenti distali costituite da Turkey\_N, WHG, Russia\_Samara\_EBA\_Yamnaya, Iran\_GanjDareh\_N e Morocco\_EN.SG, utilizzando lo stesso set di popolazioni di riferimento indicato ad inizio paragrafo (ad eccezione dei gruppi qui utilizzati come popolazione sorgente). Per ciascun gruppo è stato inizialmente testato il modello comprensivo di tutte le 5 componenti (*5-way admixture models*) e valutata la presenza di un *nested*

*model* funzionante che consentisse di minimizzare il numero di componenti richieste per descrivere il profilo genetico in questione. Nel caso di un *nested model* valido, il programma è stato rilanciato includendo soltanto le componenti previste da quest'ultimo; ciò è stato effettuato per i due gruppi dell'età del Ferro, per i quali è risultato funzionante un modello di *3-way admixture* tra Turkey\_N, WHG e Russia\_Samara\_EBA\_Yamnaya, anche spostando Iran\_GanjDareh\_N nel set di popolazioni di riferimento. Infine, seguendo lo stesso approccio, è stata valutata la possibilità di ottenere dei modelli distali validi anche per gli individui isolati AGOS001, AZZA006, TAB0016 e CAS054. Nel caso di AGOS001, è stato possibile ottenere un modello funzionante solo escludendo il gruppo EHG dalle *right populations*.

## 6. Risultati e discussione

### 6.1 Analisi dei dati di sequenziamento *shotgun*

I dati *shotgun* delle 53 librerie ottenute dai resti provenienti dall'area di Bologna hanno consentito di selezionare, sulla base della preservazione del DNA antico, i campioni idonei da poter sottoporre alla successiva cattura degli SNPs nucleari. Per il secondo set di campioni, i dati *shotgun* non sono stati utilizzati per uno screening iniziale dal momento che sono stati prodotti in concomitanza alle catture nucleari, effettuate su tutti gli individui; tali dati hanno permesso però di stimare l'efficienza della reazione di cattura e verificare l'iniziale grado di preservazione del materiale genetico.

Entrambi i sequenziamenti sono stati effettuati in modalità paired-end per un output di 5 milioni di reads per campione.

Per l'individuo MARZ001 erano già disponibili i dati di un precedente sequenziamento *shotgun*, i quali avevano evidenziato una scarsa percentuale di DNA endogeno, pari allo 0,671%, ma comunque sufficiente per procedere con la cattura degli SNPs nucleari (DNA endogeno > 0,1%). L'analisi aveva confermato, inoltre, la presenza dei tipici *patterns* di degrado del DNA antico: tassi di deaminazione particolarmente elevati alle estremità dei filamenti 3' e 5' (corrispondentemente 37,68% e 35,53%), dovuti all'assenza del trattamento UDG, e ridotte dimensioni dei frammenti (55,14 bp) [158].

Per gli altri individui, le principali statistiche risultanti dall'analisi di EAGER [163] a partire dalle *reads* grezze prodotte dai sequenziamenti *shotgun* sono mostrate in *Tabella 6.1* (a fine paragrafo), mentre in *Tabella 6.2* (sempre a fine paragrafo) sono riportati i risultati restituiti da MapDamage [34], utili per una prima autenticazione del dato.

Nella prima tabella sono riportati il totale di *reads* grezze prodotte (*Raw reads*), il numero di *reads* ottenute dalla fase di *merging* (*Reads post AdapterRemoval*), il numero di molecole clonali scartate (*Duplicati rimossi*) e quello di *reads* mappanti sulla sequenza di riferimento in seguito alla rimozione dei duplicati (*Reads mappanti post DeDup*); sono indicate, inoltre, la percentuale di DNA endogeno calcolata sul genoma umano completo, quella associata alle sole posizioni target selezionate per la cattura nucleare (necessaria per il calcolo, in un secondo momento, dell'efficienza di reazione) ed il valore relativo al cluster factor.

La percentuale di DNA endogeno, che corrisponde alla frazione di *reads* mappanti sul genoma di riferimento umano rispetto al numero totale di *reads* trattenute in seguito alla fase di *merging*, è il parametro principale che è stato valutato nella selezione degli individui idonei alla cattura nucleare.

Per quanto riguarda il primo set di campioni, sono stati ottenuti valori di DNA endogeno compresi tra un minimo di 0,08 (per il campione AGOS008) ed un massimo di 85,81 (per CAST002).

In generale, per il 64% dei campioni si osservano proporzioni di DNA endogeno superiori al 10%, mentre per il 18% tale frazione supera addirittura il 50%; si tratta di valori decisamente elevati per il DNA antico, indicativi di un grado di preservazione ottimale del materiale genetico in esame. Solo il campione AGOS008 è stato escluso dall'arricchimento degli SNPs nucleari poiché la percentuale di DNA endogeno risultava inferiore alla soglia dello 0,1%.

In tutti i campioni, il numero di duplicati rimossi risulta piuttosto esiguo rispetto al numero di *reads* uniche mappanti sulla sequenza di riferimento. Ciò traspare anche dai valori di cluster factor che, essendo intorno al valore ideale di 1, evidenziano un'elevata complessità delle librerie sequenziate. La percentuale di DNA endogeno on target, calcolata sulle specifiche posizioni nucleari da catturare, è per tutti i campioni molto bassa, in accordo alla tipologia di sequenziamento e all'output richiesto; è atteso un aumento di tale valore successivamente alla fase di arricchimento delle librerie.

Per quanto concerne il secondo set di campioni, le percentuali di DNA endogeno risultano, in linea generale, più basse di quelle registrate in precedenza e comprese tra un valore minimo di 0.03% (per SILV002 e SILV003) ed un massimo di 58,81 (per MOTE011). Questa volta, la maggior parte dei campioni presenta una frazione di endogeno inferiore al 10%, soglia superata solo nel 34% dei casi; in aggiunta, il 34% dei campioni non arriva a superare l'1% di DNA endogeno. Nonostante la scarsa preservazione del materiale genetico, soltanto gli individui SILV002 e SILV003 sarebbero stati esclusi dalla fase di cattura nucleare se fosse stato effettuato uno screening preliminare.

È da sottolineare il fatto che, rispetto al precedente set di campioni, risulta maggioritario il numero di individui per i quali sono stati campionati denti invece di rocche petrose. Considerando solo le rocche petrose, 11 su 14 hanno restituito valori di DNA endogeno superiori al 9%. Il fatto che i tre campioni associati alla cultura villanoviana provenienti dalla necropoli di San Silvestro abbiano tutti restituito quantitativi di DNA endogeno molto bassi, indipendentemente dal distretto scheletrico di partenza, ha fatto ipotizzare ad un possibile trattamento dei resti correlato a riti incineratori.

Anche in questo caso il numero di duplicati rimossi risulta decisamente più basso (di un ordine di grandezza) rispetto al numero di *reads* uniche mappanti dopo la rimozione delle molecole clonali; ciò è segnalato anche dal valore del cluster factor che in tutti i casi si aggira intorno all'1 ed indica, nel caso dei campioni preservati, una buona complessità delle librerie.

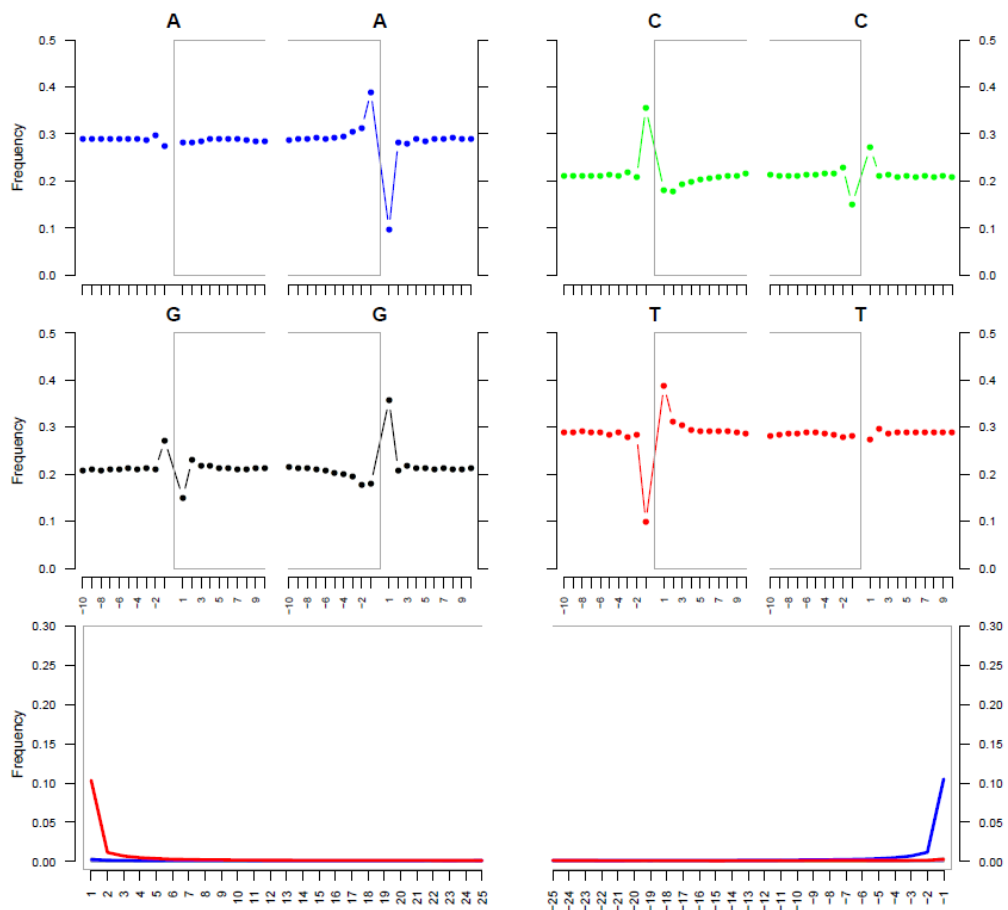
Oltre al grado di preservazione dei campioni, sono stati analizzati il tasso di misincorporazioni C-T e G-A, valutato al primo nucleotide rispettivamente delle estremità 5' e 3' dei filamenti di DNA, e le dimensioni medie dei frammenti stessi.



Considerando nell'insieme i due gruppi di campioni (ad esclusione di MARZ001, la cui libreria è stata costruita senza trattamento UDG), i valori ottenuti per le transizioni C-T risultano compresi tra 0,044 (SILV003) e 0,227 (AZZA001), mentre le modificazioni G-A ricadono tra 0,037 (AGOS006) e 0,223 (AZZA001). Ad entrambe le estremità è possibile quindi osservare un'incidenza di misincorporazioni compatibile con l'antichità dei reperti e in linea con quanto atteso sulla base del trattamento UDG-half effettuato in fase di preparazione delle librerie.

La lunghezza media dei frammenti di DNA, variabile tra 39 e 54 bp, è anche essa conforme a quanto di norma osservato per campioni di DNA antico.

Per una visualizzazione grafica delle statistiche prodotte da MapDamage viene mostrato, a scopo esemplificativo, il plot ottenuto per il campione CAST002 (Figura 6.1).



**Figura 6.1** Plot ottenuto dal software MapDamage rappresentante le sostituzioni rilevate alle estremità 5' e 3' delle molecole di DNA per il campione CAST002, indicate per ciascun nucleotide singolarmente nei quattro mini-plot in alto e come percentuale di transizioni C-T (in rosso) e G-A (in blu) nella pannello inferiore dell'immagine.

Nella parte superiore della figura sono riportati quattro *mini-plot* che illustrano la frequenza delle diverse basi azotate (A, C, G e T) in riferimento ai primi 10 nucleotidi situati alle estremità 5' e 3' dei filamenti di DNA; nello specifico, le parentesi grigie racchiudono i valori medi di frequenza calcolati sulle *reads* del campione dalla posizione 5' delle molecole degradate (parentesi di sinistra) fino alla posizione 3' (parentesi di destra), mentre esternamente sono rappresentati i valori riguardanti la sequenza di riferimento. Ciò che si osserva in corrispondenza delle parentesi è un incremento di T all'estremità 5' del filamento (accompagnato da una diminuzione nella frequenza delle purine) ed un aumento di A all'estremità opposta. Nel pannello sottostante è riportata in rosso la frequenza delle sostituzioni C-T e in blu quella relativa alle transizioni G-A per i primi 25 nucleotidi ad entrambe le estremità del filamento. Nuovamente, è visibile l'aumento delle transizioni C-T all'estremità 5' (a sinistra) e delle G-A all'estremità 3' (a destra), in accordo alle modificazioni a cui è comunemente soggetto il DNA antico. L'andamento osservato nel *plot* è quello atteso per librerie UDG-half, nelle quali i danni post-mortem sono pressoché totalmente riparati nella porzione interna dei frammenti mentre risultano ancora osservabili in corrispondenza dei primi due nucleotidi terminali.

ID campioni	Raw reads	Reads post AdapterRemoval	Duplicati rimossi	Reads mappanti post DeDup	% DNA endogeno	% DNA endogeno on target	Cluster factor
AGOS001	9635226	4504409	55586	577055	28.18	0.55	1.10
AGOS002	12594872	5944893	27502	250315	9.36	0.19	1.11
AGOS003	12308210	5440801	84100	757069	28.41	0.56	1.11
AGOS004	8222058	3253986	13901	157043	8.77	0.19	1.09
AGOS005	10913576	4325212	777	9073	0.40	0.01	1.09
AGOS006	11745540	4520401	715	7602	0.35	0.01	1.09
AGOS007	12219102	5146610	9707	113943	4.29	0.08	1.09
<b>AGOS008</b>	<b>10802896</b>	<b>4951477</b>	<b>161</b>	<b>1471</b>	<b>0.08</b>	<b>0.00</b>	<b>1.11</b>
AGOS009	7556716	3555304	938	13910	0.85	0.02	1.07
AGOS010	13250842	5955194	7715	86950	3.01	0.07	1.09
AZZA001	9206108	4821034	20864	224271	11.79	0.20	1.09
AZZA002	8333490	4005902	5299	66853	3.60	0.07	1.08
AZZA003	8888764	4009318	42592	435283	22.03	0.44	1.10
AZZA004	9421118	4656197	36272	419971	19.22	0.38	1.09
AZZA005	11056628	5529428	25520	309668	12.07	0.25	1.08
AZZA006	10490246	5298048	2734	35261	1.47	0.03	1.08
AZZA007	9863542	5146860	19245	243140	10.71	0.21	1.08
AZZA008	9767552	5082372	34621	389846	17.20	0.37	1.09
AZZA009	7890032	3618946	468	6413	0.38	0.01	1.07
AZZA010	10549724	4896746	21320	265318	11.38	0.22	1.08
AZZA011	10948462	5463703	5122	58643	2.64	0.05	1.09
AZZA012	11975638	5728866	211	2398	0.11	0.00	1.09
AZZA013	11063600	5422783	36640	416315	17.45	0.35	1.09
AZZA014	10529824	5311030	49255	519600	22.72	0.45	1.10
CAS001	23876454	11065888	32770	397706	6.98	0.14	1.08
CAS005-ETR	15674566	8263136	124626	1409334	37.19	0.78	1.09
CAS005-VILL	16260034	8196086	155199	1712826	44.60	0.81	1.09
CAS020	21797302	10754934	156120	1830334	33.74	0.65	1.09
CAS022	10647876	5311846	177774	1978658	73.80	1.36	1.09
CAS028	18465004	9378882	141973	1990030	43.67	0.86	1.07
CAS036	21310416	9059963	3664	40458	0.82	0.02	1.09
CAS041-1	9927310	4883505	163360	2031247	75.98	1.44	1.08
CAS041-2	13588272	6953859	201881	2352873	67.88	1.24	1.09
CAS054	19327684	10043650	130431	1479369	31.97	0.67	1.09
CAS067	21900848	10951823	381123	4565091	77.99	1.54	1.08
CAS090	14273242	7241724	120933	1293901	36.74	0.75	1.09
CAS095	11909604	6200775	83525	978303	33.89	0.71	1.09
CAST001	10188410	6266437	1880	18736	1.25	0.02	1.10
CAST002	9173234	4581056	180794	2202330	85.81	1.66	1.08
MAR001	11892758	6299221	107498	1054149	40.20	0.77	1.10
MAR003	12038418	6068901	96862	1022948	39.35	0.63	1.10
MAR006	14675906	7136656	315691	3287286	85.35	1.61	1.10
MAR012	12888528	6320396	193100	2340742	69.89	1.33	1.08
MAR017	10844372	5531480	173966	1922559	72.52	1.37	1.09
MAR018	15503820	7930738	231637	2635734	67.90	1.32	1.09
MAR020	12391884	6142089	150786	1716372	58.51	1.03	1.09
MAR021	12232398	5765527	39888	526000	18.12	0.33	1.08
TAB010	13213084	6581702	93686	1081230	35.98	0.64	1.09
TAB011	13054510	6034421	2567	29068	1.00	0.02	1.09
TAB012	18348402	7499887	4251	44983	1.11	0.02	1.10
TAB014	10162934	5274137	17736	218907	9.67	0.18	1.08
TAB016	15728234	7840722	88023	1048593	25.91	0.52	1.08
VISA001	10339486	5866563	27006	337909	16.63	0.36	1.08

ID campioni	Raw reads	Reads post AdapterRemoval	Duplicati rimossi	Reads mappanti post DeDup	% DNA endogeno	% DNA endogeno on target	Cluster factor
MOTE001	11770462	6065704	104480	1020780	31.85	0.65	1.10
MOTE002	10048218	5695949	109924	1133734	41.70	0.89	1.10
MOTE003	12886584	7719684	1348	10103	0.39	0.01	1.13
MOTE004	11974370	6835942	74226	641525	20.50	0.42	1.12
MOTE005	10191758	6594097	5302	46001	2.06	0.04	1.12
MOTE006	8729138	5367165	5331	53159	2.53	0.05	1.10
MOTE007	10706968	6143597	31391	248390	9.57	0.18	1.13
MOTE008	9738334	5818478	2786	26922	1.24	0.02	1.10
MOTE009	9386370	5758882	17084	172945	7.71	0.16	1.10
MOTE010	9571058	6005076	4197	46588	2.05	0.05	1.09
MOTE011	9991906	5403716	168428	1557677	58.81	1.20	1.11
MOTE012	9443984	6041840	8871	81733	3.83	0.08	1.11
MOTE013	9490604	5552960	7500	75491	3.23	0.06	1.10
MOTE014	11976738	6947271	107021	954842	32.95	0.66	1.11
MOTE015	9891192	6301314	2820	23076	1.04	0.02	1.12
MOVE001	24755278	11076940	107748	804541	12.21	0.26	1.13
MOVE002	11490386	6079063	125981	1235134	38.30	0.81	1.10
MOVE003	16911402	8247619	66269	571078	12.11	0.25	1.12
MOVE004	15152700	7126136	20040	190604	4.46	0.09	1.11
MOVE005	9917636	5910947	225	6028	0.25	0.00	1.04
MOVE006	9591312	4584081	225	5176	0.23	0.00	1.04
MOVE007	13799106	7047496	156181	1271300	34.08	0.74	1.12
MOVE008	12645604	6534145	180422	1553884	42.09	0.94	1.12
MOVE009	11862274	6507021	1565	14984	0.47	0.01	1.10
MOVE010	19353268	9636771	7971	68220	1.36	0.03	1.12
MOVE011	17427246	8348522	634	5519	0.13	0.00	1.12
SILV001	13551844	7105523	234	5057	0.14	0.00	1.05
<b>SILV002</b>	<b>11387342</b>	<b>6610741</b>	<b>24</b>	<b>623</b>	<b>0.03</b>	<b>0.00</b>	1.04
<b>SILV003</b>	<b>14636510</b>	<b>7592721</b>	<b>34</b>	<b>826</b>	<b>0.03</b>	<b>0.00</b>	1.04

**Tabella 6.1** Principali statistiche relative al sequenziamento *shotgun* dei 53 campioni etruschi dell'area bolognese e dei nuovi campioni recentemente aggiunti al dataset. È stato evidenziato in rosso l'unico campione del primo set (AGOS008) escluso dalla successiva fase di cattura degli SNPs nucleari a causa della scarsa preservazione del DNA endogeno (< 0.1%); allo stesso modo sono indicati gli unici due campioni del sito di San Silvestro che sarebbero stati esclusi dalla cattura se fosse stato effettuato lo screening preliminare.

ID campioni	% Deaminazioni 3'	% Deaminazioni 5'	Dimensione media frammenti
AGOS001	0.21	0.21	49
AGOS002	0.13	0.14	49
AGOS003	0.08	0.08	48
AGOS004	0.05	0.05	51
AGOS005	0.04	0.05	47
AGOS006	0.04	0.05	43
AGOS007	0.06	0.07	50
AGOS008	0.08	0.06	39
AGOS009	0.14	0.15	50
AGOS010	0.07	0.07	52
AZZA001	0.22	0.23	47
AZZA002	0.19	0.20	49
AZZA003	0.17	0.17	50
AZZA004	0.10	0.11	51
AZZA005	0.06	0.07	51
AZZA006	0.07	0.07	49
AZZA007	0.15	0.15	51
AZZA008	0.07	0.07	52
AZZA009	0.07	0.06	50
AZZA010	0.13	0.13	49
AZZA011	0.22	0.22	48
AZZA012	0.05	0.04	45
AZZA013	0.19	0.19	50
AZZA014	0.13	0.13	50
CAS001	0.08	0.08	46
CAS005-ETR	0.07	0.07	52
CAS005-VILL	0.12	0.12	47
CAS020	0.11	0.11	48
CAS022	0.12	0.12	48
CAS028	0.09	0.09	48
CAS036	0.10	0.09	49
CAS041-1	0.12	0.12	49
CAS041-2	0.13	0.14	48
CAS054	0.05	0.05	51
CAS067	0.08	0.08	50
CAS090	0.06	0.06	49
CAS095	0.06	0.06	50
CAST001	0.20	0.22	47
CAST002	0.11	0.10	49
MAR001	0.10	0.10	48
MAR003	0.17	0.18	44
MAR006	0.09	0.09	48
MAR012	0.12	0.12	49
MAR017	0.10	0.10	48
MAR018	0.08	0.08	49
MAR020	0.11	0.11	46
MAR021	0.09	0.09	47
TAB010	0.12	0.12	46
TAB011	0.06	0.05	47
TAB012	0.05	0.05	45
TAB014	0.10	0.10	47
TAB016	0.07	0.07	49
VISA001	0.14	0.14	54

ID campioni	% Deaminazioni 3'	% Deaminazioni 5'	Dimensione media frammenti
MOTE001	0.07	0.07	51
MOTE002	0.08	0.08	53
MOTE003	0.09	0.09	49
MOTE004	0.08	0.08	50
MOTE005	0.07	0.07	50
MOTE006	0.09	0.09	50
MOTE007	0.12	0.12	49
MOTE008	0.13	0.13	48
MOTE009	0.09	0.09	51
MOTE010	0.09	0.10	54
MOTE011	0.09	0.09	51
MOTE012	0.09	0.09	49
MOTE013	0.08	0.08	48
MOTE014	0.10	0.10	51
MOTE015	0.09	0.08	49
MOVE001	0.09	0.09	53
MOVE002	0.09	0.09	52
MOVE003	0.10	0.10	51
MOVE004	0.09	0.09	52
MOVE005	0.08	0.10	50
MOVE006	0.09	0.08	51
MOVE007	0.07	0.07	54
MOVE008	0.07	0.07	54
MOVE009	0.08	0.08	49
MOVE010	0.09	0.09	47
MOVE011	0.09	0.07	44
SILV001	0.09	0.07	49
SILV002	0.04	0.07	46
SILV003	0.04	0.04	45

**Tabella 6.2** Statistiche ottenute da MapDamage sui dati *shotgun*: percentuali di transizioni G-A e C-T rilevate ai primi nucleotidi rispettivamente delle estremità 3' e 5' dei filamenti e dimensione media delle molecole di DNA.

## 6.2 Statistiche relative al processamento dei dati di cattura nucleare

I dati di sequenza ottenuti dai prodotti di cattura nucleare sono stati inizialmente processati anche in questo caso tramite la pipeline di EAGER [163]. In *Tabella 6.3* (a fine paragrafo) vengono riportate le principali statistiche ottenute dal sequenziamento in modalità single-end (output di 20 milioni di reads per campione) delle 82 librerie precedentemente sottoposte al processo di arricchimento; sono mostrati gli stessi parametri discussi anche per i dati *shotgun* con l'aggiunta dei valori relativi all'efficienza di cattura (rappresentata rapporto tra la percentuale di DNA endogeno *on target* dopo e prima dell'arricchimento) e alla copertura media sul genoma umano.

Considerando la percentuale di DNA endogeno valutata sull'intero genoma umano, si nota come in tutti i casi i valori risultino superiori a quelli in precedenza ottenuti a partire dai dati *shotgun*. Infatti, la fase di cattura nucleare ha consentito di arricchire le librerie per posizioni specifiche del genoma nucleare umano, riducendo quindi l'incidenza nel dato finale di materiale genetico di diversa natura (presente in proporzione maggiore nel caso in cui si sequenzi in modo aspecifico l'intero contenuto genomico di un campione). La proporzione di materiale genetico endogeno osservata nei campioni ricopre un intervallo che va da un minimo di 0,04 (SILV002) ad un massimo di 88% (MARZ006).

Come già osservato nel precedente paragrafo, il primo set di campioni risulta caratterizzato da un miglior grado di preservazione rispetto al secondo, nel quale più del 23% degli individui mostra una percentuale di DNA endogeno inferiore all'1%.

Come atteso, è possibile notare anche un aumento della percentuale di DNA endogeno *on target* (0,01-28,69 rispettivamente in SILV003 e MOVE002), evidente anche dal calcolo dell'efficienza di cattura, che costituisce il fattore di incremento relativo alla copertura sugli SNPs nucleari in seguito alla reazione di arricchimento. I valori di efficienza più elevati che si osservano analizzando il primo set di campioni corrispondono a quelli associati ad un doppio *round* di cattura, effettuato nel caso di scarsa preservazione del DNA endogeno rilevata in fase di screening; sul secondo set di campioni, in assenza di tale informazione, sono stati condotti in tutti i casi 2 rounds di cattura.

In *Tabella 6.3* sono indicati in rosso i campioni con percentuale di DNA *on target* inferiore all'1%.

Il cluster factor, variabile tra 1,15 (CAS028) e 11,35 (AGOS005), indica che per alcuni campioni (con valore intorno all'1) potrebbe valere la pena investire in un sequenziamento più profondo qualora si ritenesse utile incrementare il dato; i restanti campioni (con valori lontani da 1) si caratterizzano, invece, per una bassa complessità della libreria, evidenziata anche dal limitato numero di sequenze uniche mappanti sulla *reference* rispetto al numero di duplicati di PCR rimossi.

La copertura media sul genoma umano, per quanto non particolarmente elevata trattandosi di dati di cattura, è in quasi la metà dei campioni superiore all'1% e nel 13% dei campioni maggiore del 2%.

Anche nel caso dei dati di cattura sono stati valutati i *patterns* di degrado restituiti da MapDamage, mostrati in *Tabella 6.4*. Le sostituzioni G-A, valutate al primo nucleotide in 3', compaiono con una frequenza compresa tra 0,034 (SILV003) e 0,208 (AZZA001, escludendo MARZ001 con 0,363 dovuto alla tipologia di libreria preparata); le modificazioni C-T, valutate al primo nucleotide in 5', rientrano nell'intervallo 0,017-0,230 (rispettivamente in SILV002 ed AZZA001, non considerando MARZ001 con una frequenza pari a 0,231). Nonostante i valori minimi risultino decisamente bassi anche per librerie di tipo UDG-half, bisogna considerare il fatto che sono associati ad individui caratterizzati da un numero molto limitato di *reads* mappanti sul genoma umano, come SILV002 e SILV003.

La dimensione media dei frammenti varia tra 45 (MOTE008) e 73 bp (MOVE006) e risulta coerente a quanto atteso per molecole di DNA antico. In accordo con quanto solitamente osservato in seguito al processo di cattura, le dimensioni dei frammenti risultano più elevate rispetto a quelle riscontrate nei rispettivi dati *shotgun*.

In *Tabella 6.4* sono riportati, inoltre, i valori relativi all'mt/nuclear ratio, i quali possono costituire un valido indicatore nella stima della contaminazione, soprattutto nel caso in cui il dato genetico è stato prodotto a partire dalla rocca petrosa (come descritto in [169] e riportato nei Paragrafi 5.3.2 e 6.3). Tutti i valori associati rocche petrose si mantengono sotto la soglia di 200, consentendo di estendere con un buon grado di affidabilità la stima della contaminazione mitocondriale anche al genoma nucleare (correlazione utile in presenza di individui di sesso femminile), mentre i valori più elevati corrispondono al campionamento di denti.

Nella *Tabella 6.5* è possibile osservare il numero di SNPs coperti sui pannelli Human Origins (HO, comprensivo di 597.573 SNPs) e 1240K (1.233.013 SNPs) in seguito alla fase di genotipizzazione. Sul pannello HO si osservano coperture comprese tra 273 (per SILV003) e 443.874 SNPs (per MARZ006) che corrispondono sul pannello 1240K a circa il doppio delle posizioni (544 per SILV003 e 820.554 per MARZ006). I valori riportati in *Tabella 6.5* riflettono il grado di preservazione del DNA endogeno già discusso in precedenza. Gli individui con meno di 20.000 SNPs coperti sul pannello HO sono stati successivamente esclusi dalle analisi popolazionistiche (AZZA012, MOTE008, MOVE005, MOVE006, MOVE011, SILV001, SILV002 e SILV003, marcati in rosso in *Tabella 6.5*).



ID campioni	Raw reads	Reads post AdapterRemoval	Duplicati rimossi	Reads mappanti post DeDup	% DNA endogeno	% DNA endogeno on target	Efficienza di cattura	Cluster factor	Copertura media
AGOS001	26886966	23881549	1566649	6162342	44.03	10.74	19.6	1.25	1.61
AGOS002	23203192	20654088	631459	2692497	21.68	6.14	33.0	1.24	0.85
AGOS003	27241203	23877717	1722010	6589584	46.36	11.87	21.3	1.26	1.76
AGOS004	19396740	16025867	701701	2824503	28.40	8.82	47.4	1.25	0.92
AGOS005	20249214	17017520	1740516	168171	13.41	7.79	973.8	11.35	0.05
AGOS006	21745033	17891913	906397	145229	8.32	4.02	669.7	7.24	0.04
AGOS007	21245223	17819073	424786	1363010	13.54	3.69	43.9	1.31	0.37
AGOS009	18600910	16906709	2393837	1350214	25.35	16.22	954.1	2.77	0.68
AGOS010	21028892	18082633	275647	1167261	10.05	3.02	46.4	1.24	0.36
AZZA001	20056546	18945527	384291	1985966	19.46	3.67	18.6	1.19	0.47
AZZA002	20055420	18047678	191278	1020819	9.56	2.36	34.1	1.19	0.30
AZZA003	24364861	21295566	1191835	5120752	39.02	9.87	22.4	1.23	1.36
AZZA004	20954415	19349473	787933	4149322	34.31	8.10	21.3	1.19	1.06
AZZA005	20263109	18749579	665208	3101067	25.98	6.81	27.3	1.22	0.86
AZZA006	15929715	14722640	76021	394963	4.39	1.18	40.7	1.19	0.12
AZZA007	21272758	20140182	463057	2578942	20.34	4.86	22.7	1.18	0.68
AZZA008	17259577	16328667	742771	3674852	34.28	9.11	24.9	1.20	1.01
AZZA009	4576808	4069248	341324	97703	12.30	7.82	976.9	4.49	0.04
AZZA010	26137581	23183641	771932	3527586	24.60	6.51	29.2	1.22	0.98
AZZA011	21623949	19614499	153971	704167	6.29	1.47	29.5	1.22	0.19
AZZA012	20552918	18372363	252282	80364	2.32	1.26	630.5	4.14	0.03
AZZA013	18846330	17102312	701678	3345540	31.51	8.06	23.2	1.21	0.93
AZZA014	22068492	20443772	984318	4589216	36.55	8.61	19.1	1.21	1.18
CAS001	20449461	18307846	310026	1825454	16.47	3.81	28.2	1.17	0.49
CAS005-ETR	21195955	20235802	1359808	6902267	52.05	12.16	15.5	1.20	1.68
CAS005-VILL	24647883	23332865	1510441	7243876	54.00	10.08	12.5	1.21	1.58
CAS020	19466591	18232222	1025858	5600599	49.68	11.17	17.2	1.18	1.39
CAS022	25211674	23718161	2437637	10500976	74.25	14.01	10.3	1.23	2.08
CAS028	18250830	17276975	960992	6230512	56.47	12.07	14.1	1.15	1.49
CAS036	25392393	22289535	2559766	1314498	20.24	12.40	729.5	2.95	0.61
CAS041-1	20040852	18947757	1641244	9512824	79.24	15.19	10.5	1.17	1.95
CAS041-2	20771123	19792249	1775851	8679787	72.04	13.40	10.8	1.21	1.75
CAS054	21019117	20044222	1204347	6245504	48.53	11.26	16.9	1.19	1.56
CAS067	21938493	20998292	1982281	10669667	80.98	15.00	9.8	1.19	2.18
CAS090	22687138	21462162	1596297	6788097	51.47	11.75	15.7	1.24	1.65
CAS095	19504233	18388351	1058138	5586978	47.34	10.47	14.8	1.19	1.31
<b>CAST001</b>	<b>23947314</b>	<b>23203308</b>	<b>63225</b>	<b>200689</b>	<b>1.59</b>	<b>0.33</b>	<b>14.4</b>	<b>1.32</b>	<b>0.04</b>
CAST002	18866081	18010624	1758097	9814617	85.32	15.81	9.5	1.18	1.91
MAR001	23857818	22505887	1743703	6860835	52.02	11.40	14.8	1.25	1.59
MAR003	19888704	18385645	921177	4833997	47.77	8.13	12.9	1.19	1.00
MAR006	20483173	19241162	2121406	10640848	88.01	17.85	11.1	1.20	2.30
MAR012	21446695	20137654	1805866	9626364	76.60	15.34	11.6	1.19	2.09
MAR017	22938764	21568875	2364891	9800781	76.13	14.83	10.8	1.24	1.98
MAR018	20015433	18980551	1618954	8878728	73.50	14.95	11.3	1.18	1.95
MAR020	21351605	19893749	1575938	7751644	66.80	13.05	12.7	1.20	1.70
MAR021	17242234	15609504	639176	3303724	35.31	8.44	25.7	1.19	0.88
TAB010	25649214	23600252	1378906	6790969	50.02	10.12	15.8	1.20	1.59
TAB011	24474140	21033201	132051	386108	3.36	0.92	51.2	1.34	0.11
TAB012	15920111	12641090	34743	173100	2.46	0.48	25.3	1.20	0.04
TAB014	16680267	15634011	329177	1579233	17.51	3.82	21.5	1.21	0.39
TAB016	21262614	20140150	1280240	5857929	46.12	12.77	24.7	1.22	1.72
VISA001	20148644	19339251	725561	3693598	29.09	8.18	22.5	1.20	1.10

ID campioni	Raw reads	Reads post AdapterRemoval	Duplicati rimossi	Reads mappanti post DeDup	% DNA endogeno	% DNA endogeno on target	Efficienza di cattura	Cluster factor	Copertura media
MOTE001	14314098	13430100	1096128	5531181	63.87	24.75	37.8	1.20	2.29
MOTE002	13945798	13462199	1399518	6024414	70.36	27.30	30.6	1.23	2.37
MOTE003	15154798	14102848	121240	128695	2.48	1.03	171.0	1.94	0.05
MOTE004	13782850	13273974	1009790	4553056	55.14	21.08	50.8	1.22	1.85
MOTE005	13423487	12863623	717021	781140	15.06	6.66	166.5	1.92	0.27
MOTE006	14131870	13335958	868627	817782	15.74	7.46	152.3	2.06	0.28
MOTE007	15695449	14705353	1266334	2612739	35.04	14.09	77.9	1.49	0.99
<b>MOTE008</b>	<b>13616203</b>	<b>12444590</b>	<b>16902</b>	<b>29468</b>	<b>0.57</b>	<b>0.15</b>	<b>6.5</b>	<b>1.57</b>	<b>0.01</b>
MOTE009	14993118	14177060	1200534	2225346	31.01	13.03	84.1	1.54	0.80
MOTE010	14922361	13992427	619161	801428	12.10	6.23	138.5	1.77	0.32
MOTE011	17012176	16066125	1893335	7925245	78.30	27.61	23.0	1.24	2.81
MOTE012	16000351	15200069	1907363	5531181	22.75	11.85	153.9	2.97	0.28
MOTE013	13311376	12369417	1056107	774157	19.04	8.32	136.3	2.36	0.21
MOTE014	14451059	13651086	1698356	5024989	64.60	24.01	36.2	1.34	1.79
MOTE015	18635538	17656866	387974	385222	5.48	2.50	131.5	2.01	0.13
MOVE001	13984791	12705075	1057877	4603495	56.03	26.23	101.7	1.23	2.26
MOVE002	11606324	11069462	1106216	5060891	69.95	28.69	35.6	1.22	2.11
MOVE003	13337034	12434214	1149645	4133207	53.18	25.06	99.9	1.28	1.96
MOVE004	15522134	14067301	741455	2882879	32.40	15.62	166.2	1.26	1.43
<b>MOVE005</b>	<b>14544030</b>	<b>13615894</b>	<b>31562</b>	<b>30866</b>	<b>0.68</b>	<b>0.24</b>	<b>60.8</b>	<b>2.02</b>	<b>0.01</b>
<b>MOVE006</b>	<b>12184574</b>	<b>10426492</b>	<b>63204</b>	<b>54832</b>	<b>1.85</b>	<b>0.70</b>	<b>174.8</b>	<b>2.15</b>	<b>0.02</b>
MOVE007	13768275	12895739	1209232	5496569	64.43	26.34	35.4	1.22	2.28
MOVE008	16672672	15910544	1576345	7416979	70.15	27.04	28.7	1.21	2.92
MOVE009	15463970	14363269	272754	230898	4.59	2.22	246.4	2.18	0.09
MOVE010	15403184	13696279	438305	910492	13.52	5.99	230.5	1.48	0.42
<b>MOVE011</b>	<b>15706028</b>	<b>13890519</b>	<b>47327</b>	<b>64937</b>	<b>1.16</b>	<b>0.47</b>	<b>235.5</b>	<b>1.73</b>	<b>0.03</b>
<b>SILV001</b>	<b>14235383</b>	<b>13064095</b>	<b>19961</b>	<b>20879</b>	<b>0.62</b>	<b>0.18</b>	<b>60.0</b>	<b>1.96</b>	<b>0.01</b>
<b>SILV002</b>	<b>15661181</b>	<b>14529991</b>	<b>815</b>	<b>1903</b>	<b>0.04</b>	<b>0.01</b>	-	<b>1.43</b>	<b>0.00</b>
<b>SILV003</b>	<b>18033971</b>	<b>16367829</b>	<b>784</b>	<b>1950</b>	<b>0.08</b>	<b>0.01</b>	-	<b>1.40</b>	<b>0.00</b>
MARZ001	12648738	12289795	212071	352948	5.92	2.93	-*	1.60	0.16

**Tabella 6.3** Statistiche relative al sequenziamento e successivo processamento tramite EAGER dei dati di cattura nucleare. In rosso sono indicati gli individui per i quali, nonostante la reazione di arricchimento, si registra una percentuale di DNA endogeno on target inferiore all'1%. \*Valore non calcolato per MARZ001 dal momento che non risulta disponibile il dato relativo alla percentuale di DNA endogeno on target precedente alla fase di cattura.

ID campioni	MT/NUC ratio	% Deaminazioni 3'	% Deaminazioni 5'	Dimensione media frammenti
AGOS001	70.72	0.17	0.20	56
AGOS002	64.75	0.11	0.13	58
AGOS003	80.25	0.08	0.08	56
AGOS004	49.72	0.05	0.06	61
AGOS005	139.80	0.05	0.05	55
AGOS006	20.91	0.05	0.04	50
AGOS007	25.17	0.06	0.07	58
AGOS009	13.39	0.13	0.16	60
AGOS010	36.73	0.06	0.07	62
AZZA001	57.50	0.21	0.23	52
AZZA002	68.35	0.16	0.20	57
AZZA003	47.09	0.13	0.16	58
AZZA004	46.67	0.10	0.11	58
AZZA005	47.24	0.06	0.06	60
AZZA006	91.89	0.06	0.07	59
AZZA007	42.67	0.12	0.15	59
AZZA008	72.50	0.06	0.07	62
AZZA009	12.49	0.05	0.06	62
AZZA010	41.74	0.11	0.13	58
AZZA011	76.33	0.19	0.22	56
AZZA012	24.71	0.07	0.08	57
AZZA013	47.71	0.15	0.19	58
AZZA014	88.67	0.11	0.13	58
CAS001	152.60	0.08	0.08	54
CAS005-ETR	74.26	0.06	0.07	59
CAS005-VILL	82.03	0.12	0.13	52
CAS020	69.48	0.10	0.11	55
CAS022	88.33	0.12	0.13	53
CAS028	77.51	0.08	0.09	54
CAS036	44.48	0.09	0.10	60
CAS041-1	75.15	0.12	0.13	54
CAS041-2	67.09	0.13	0.14	53
CAS054	114.46	0.05	0.05	58
CAS067	62.64	0.08	0.09	56
CAS090	138.43	0.06	0.06	56
CAS095	93.62	0.06	0.06	56
CAST001	2726.52	0.18	0.21	54
CAST002	83.19	0.10	0.11	55
MAR001	67.65	0.09	0.09	55
MAR003	73.55	0.17	0.18	48
MAR006	68.29	0.08	0.09	55
MAR012	44.80	0.10	0.12	55
MAR017	88.22	0.09	0.10	54
MAR018	45.18	0.08	0.08	56
MAR020	60.40	0.11	0.11	52
MAR021	58.00	0.09	0.09	55
TAB010	51.54	0.11	0.12	52
TAB011	52.84	0.05	0.06	56
TAB012	43.91	0.05	0.05	53
TAB014	70.75	0.10	0.10	54
TAB016	62.33	0.07	0.07	57
VISA001	36.51	0.10	0.14	63

ID campioni	MT/NUC ratio	% Deaminazioni 3'	% Deaminazioni 5'	Dimensione media frammenti
MOTE001	55.35	0.08	0.08	61
MOTE002	51.65	0.09	0.09	63
MOTE003	5754.95	0.10	0.09	59
MOTE004	56.16	0.08	0.08	59
MOTE005	3463.32	0.06	0.07	60
MOTE006	1786.91	0.10	0.10	58
MOTE007	38.28	0.13	0.13	56
MOTE008	127.37	0.19	0.19	45
MOTE009	187.80	0.10	0.10	60
MOTE010	448.78	0.10	0.10	66
MOTE011	41.46	0.09	0.09	62
MOTE012	1350.48	0.09	0.09	57
MOTE013	2581.46	0.08	0.08	54
MOTE014	51.11	0.10	0.10	61
MOTE015	2175.57	0.09	0.09	59
MOVE001	27.11	0.09	0.09	67
MOVE002	61.27	0.09	0.09	63
MOVE003	28.24	0.11	0.11	63
MOVE004	35.85	0.10	0.10	65
MOVE005	39980.71	0.07	0.07	61
MOVE006	15757.60	0.05	0.04	73
MOVE007	51.04	0.07	0.07	67
MOVE008	31.46	0.07	0.07	68
MOVE009	1373.25	0.08	0.08	60
MOVE010	161.78	0.10	0.10	56
MOVE011	5491.39	0.10	0.10	56
SILV001	8.69	0.11	0.11	59
SILV002	0.00	0.04	0.02	57
SILV003	85.36	0.04	0.02	50
MARZ001	142.95	0.36	0.38	64

**Tabella 6.4** Statistiche ottenute da MapDamage sui dati di cattura nucleare: percentuali di transizioni G-A e C-T rilevate ai primi nucleotidi rispettivamente delle estremità 3' e 5' dei filamenti e dimensione media delle molecole di DNA. È riportato, inoltre, il valore di mt/nuclear ratio restituito da EAGER in seguito al mappaggio sull'intero genoma umano (comprensivo anche dell'mtDNA).

ID campioni	SNPs coperti HO	SNPs coperti 1240K	ID campioni	SNPs coperti HO	SNPs coperti 1240K
AGOS001	401070	742215	MOTE001	360905	672650
AGOS002	296884	546967	MOTE002	385020	722600
AGOS003	415324	776353	MOTE003	26355	50150
AGOS004	310079	572527	MOTE004	334115	615397
AGOS005	28494	52861	MOTE005	129124	243851
AGOS006	26071	48373	MOTE006	133893	252548
AGOS007	175823	320179	MOTE007	276188	515502
AGOS009	250840	458300	<b>MOTE008</b>	<b>4008</b>	<b>7779</b>
AGOS010	169898	307347	MOTE009	253620	473766
AZZA001	196821	358039	MOTE010	150483	282866
AZZA002	144767	262694	MOTE011	409351	760974
AZZA003	375460	690134	MOTE012	137910	258682
AZZA004	324904	591390	MOTE013	108153	203740
AZZA005	298743	545286	MOTE014	361820	679621
AZZA006	67986	123110	MOTE015	67721	128696
AZZA007	258948	471030	MOVE001	348885	642691
AZZA008	325023	595164	MOVE002	360407	674063
AZZA009	26521	48031	MOVE003	347696	641784
AZZA010	323932	599715	MOVE004	299374	557779
AZZA011	102066	184747	<b>MOVE005</b>	<b>5767</b>	<b>11124</b>
<b>AZZA012</b>	<b>19686</b>	<b>36042</b>	<b>MOVE006</b>	<b>12356</b>	<b>23366</b>
AZZA013	309428	564607	MOVE007	382011	714900
AZZA014	346058	637620	MOVE008	409143	765505
CAS001	205070	372582	MOVE009	49730	94561
CAS005-ETR	397374	724786	MOVE010	157740	299358
CAS005-VILL	380980	700491	<b>MOVE011</b>	<b>15057</b>	<b>29106</b>
CAS020	360097	654038	<b>SILV001</b>	<b>4947</b>	<b>9304</b>
CAS022	435444	809231	<b>SILV002</b>	<b>317</b>	<b>586</b>
CAS028	372398	685459	<b>SILV003</b>	<b>273</b>	<b>544</b>
CAS036	237994	430348	MARZ001	68746	129707
CAS041-1	424566	793809			
CAS041-2	406388	748850			
CAS054	377556	684337			
CAS067	427898	790199			
CAS090	390871	721386			
CAS095	360657	656774			
CAST001	25814	47437			
CAST002	424723	786408			
MAR001	393472	729180			
MAR003	305778	560318			
MAR006	443874	820554			
MAR012	430899	804559			
MAR017	428437	793700			
MAR018	416896	768089			
MAR020	392935	722684			
MAR021	294607	540763			
TAB010	383280	709534			
TAB011	62342	113571			
TAB012	25255	45630			
TAB014	178787	323921			
TAB016	398579	741114			
VISA001	337705	622279			

**Tabella 6.5** SNPs coperti in seguito alla genotipizzazione effettuata sia sul pannello Human Origins (HO, 597.573 SNPs totali) che 1240K (1.233.013 SNPs totali). In rosso sono indicati gli individui con meno di 20.000 SNPs coperti sul pannello HO.

### 6.3 Attribuzione del sesso biologico e valutazione della contaminazione

Tutti i campioni per i quali sono stati prodotti i dati di cattura nucleare sono stati sottoposti ad una fase preliminare di analisi volta all'identificazione, dal punto di vista genetico, del sesso e alla stima della contaminazione nucleare (per i soli individui di sesso maschile) e mitocondriale.

I risultati relativi alla determinazione del sesso sono riportati nella seguente tabella (*Tabella 6.6*).

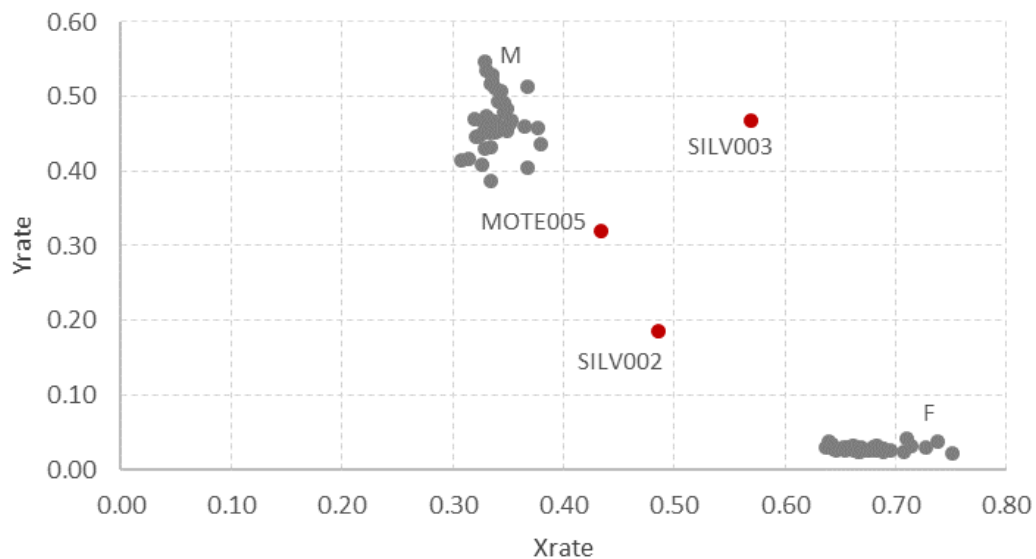
ID campioni	xCov	yCov	autCov	Xrate	Yrate	Sesso	Contaminazione X-chr $\pm$ SE	SNPs X-chr (cop>2)
AGOS001	1.0930	0.0398	1.5729	0.6949	0.0253	F	-	-
AGOS002	0.2866	0.3961	0.8271	0.3464	0.4789	M	0.005261 $\pm$ 0.003230	178
AGOS003	0.6073	0.8059	1.7243	0.3522	0.4674	M	0.010342 $\pm$ 0.002273	2759
AGOS004	0.3124	0.4145	0.9084	0.3439	0.4563	M	0.004882 $\pm$ 0.002838	1025
<b>AGOS005</b>	<b>0.0156</b>	<b>0.0217</b>	<b>0.0464</b>	<b>0.3356</b>	<b>0.4670</b>	<b>M</b>	<b>NA</b>	<b>1</b>
AGOS006	0.0283	0.0013	0.0426	0.6652	0.0302	F	-	-
AGOS007	0.2495	0.0100	0.3623	0.6887	0.0275	F	-	-
AGOS009	0.2202	0.2889	0.6704	0.3285	0.4310	M	0.01156 $\pm$ 0.00518	694
AGOS010	0.1103	0.1463	0.3511	0.3142	0.4167	M	0.007569 $\pm$ 0.008515	219
<hr/>								
AZZA001	0.3005	0.0144	0.4548	0.6607	0.0318	F	-	-
AZZA002	0.0950	0.1349	0.2894	0.3282	0.4661	M	0.012644 $\pm$ 0.011276	178
AZZA003	0.9040	0.0368	1.3284	0.6806	0.0277	F	-	-
AZZA004	0.6795	0.0309	1.0405	0.6530	0.0297	F	-	-
AZZA005	0.2810	0.3921	0.8423	0.3336	0.4655	M	0.005221 $\pm$ 0.003455	889
<b>AZZA006</b>	<b>0.0386</b>	<b>0.0545</b>	<b>0.1168</b>	<b>0.3303</b>	<b>0.4672</b>	<b>M</b>	<b>NA</b>	<b>25</b>
AZZA007	0.4410	0.0173	0.6662	0.6619	0.0259	F	-	-
AZZA008	0.3284	0.4701	0.9936	0.3305	0.4731	M	0.000364 $\pm$ 0.001788	1162
<b>AZZA009</b>	<b>0.0142</b>	<b>0.0164</b>	<b>0.0424</b>	<b>0.3343</b>	<b>0.3865</b>	<b>M</b>	<b>NA</b>	<b>5</b>
AZZA010	0.3342	0.4637	0.9574	0.3491	0.4843	M	0.000764 $\pm$ 0.001919	1192
AZZA011	0.1222	0.0055	1.1864	0.6557	0.0294	F	-	-
<b>AZZA012</b>	<b>0.0119</b>	<b>0.0137</b>	<b>0.0314</b>	<b>0.3799</b>	<b>0.4366</b>	<b>M</b>	<b>NA</b>	<b>3</b>
AZZA013	0.6072	0.0221	0.9097	0.6675	0.0243	F	-	-
AZZA014	0.3918	0.5369	1.1532	0.3398	0.4656	M	0.006273 $\pm$ 0.002697	1469
<hr/>								
CAS001	0.1575	0.2239	0.4801	0.3280	0.4663	M	0.001279 $\pm$ 0.003932	376
CAS005-ETR	1.0634	0.0426	1.6452	0.6464	0.0259	F	-	-
CAS005-VILL	0.4927	0.6865	1.5388	0.3202	0.4461	M	0.005313 $\pm$ 0.002109	2177
CAS020	0.8701	0.0404	1.3531	0.6430	0.0298	F	-	-
CAS022	1.3788	0.0624	2.0171	0.6835	0.0310	F	-	-
CAS028	0.4831	0.6567	1.4535	0.3324	0.4518	M	0.003997 $\pm$ 0.001880	1981
CAS036	0.1962	0.2703	0.6017	0.3261	0.4492	M	0.007367 $\pm$ 0.004102	624
CAS041-1	0.6492	0.8643	1.9033	0.3411	0.4541	M	0.006648 $\pm$ 0.001840	2905
CAS041-2	1.1376	0.0456	1.7045	0.6674	0.0267	F	-	-
CAS054	0.9712	0.0457	1.5253	0.6367	0.0300	F	-	-
CAS067	0.6875	0.9510	2.1354	0.3219	0.4453	M	0.006177 $\pm$ 0.001590	3151
CAS090	0.5384	0.7362	1.6110	0.3342	0.4570	M	0.005878 $\pm$ 0.001859	2410
CAS095	0.8680	0.0373	1.2780	0.6792	0.0292	F	-	-
<hr/>								
CAST001	0.0269	0.0014	0.0418	0.6427	0.0344	F	-	-
CAST002	1.2588	0.0486	1.8639	0.6754	0.0261	F	-	-
<hr/>								
MAR001	0.5240	0.7023	1.5517	0.3377	0.4526	M	0.005878 $\pm$ 0.002010	2299
MAR003	0.6396	0.0306	0.9657	0.6624	0.0317	F	-	-
MAR006	1.5192	0.0558	2.2450	0.6767	0.0249	F	-	-
MAR012	0.6824	0.8853	2.0447	0.3337	0.4330	M	0.00731 $\pm$ 0.00178	3161
MAR017	1.3159	0.0508	1.9311	0.6814	0.0263	F	-	-
MAR018	1.2768	0.0488	1.9006	0.6718	0.0257	F	-	-
MAR020	1.1036	0.0487	1.6520	0.6680	0.0295	F	-	-
MAR021	0.5885	0.0243	0.8553	0.6880	0.0284	F	-	-

ID campioni	xCov	yCov	autCov	Xrate	Yrate	Sesso	Contaminazione X-chr $\pm$ SE	SNPs X-chr (cop>2)
TAB010	0.5154	0.7053	1.5458	0.3334	0.4563	M	0.004279 $\pm$ 0.001954	2173
<b>TAB011</b>	<b>0.0347</b>	<b>0.0435</b>	<b>0.1062</b>	<b>0.3263</b>	<b>0.4095</b>	<b>M</b>	<b>NA</b>	<b>24</b>
TAB012	0.0258	0.0015	0.0403	0.6395	0.0380	F	-	-
TAB014	0.2567	0.0119	0.3766	0.6817	0.0315	F	-	-
TAB016	0.5869	0.7614	1.6804	0.3493	0.4531	M	0.00427 $\pm$ 0.00169	2655
VISA001	0.3708	0.5482	1.0788	0.3437	0.5082	M	0.003024 $\pm$ 0.002232	1410
MOTE001	0.7351	1.1887	2.2260	0.3302	0.5340	M	0.006646 $\pm$ 0.001761	2891
MOTE002	0.7810	1.1769	2.3060	0.3387	0.5104	M	0.002746 $\pm$ 0.001105	3214
MOTE003	0.0315	0.0018	0.0443	0.7101	0.0408	F	-	-
MOTE004	1.1574	0.0592	1.8021	0.6422	0.0328	F	-	-
<b>MOTE005</b>	<b>0.1154</b>	<b>0.0849</b>	<b>0.2660</b>	<b>0.4337</b>	<b>0.3191</b>	<b>M?</b>	<b>0.321741 <math>\pm</math> 0.032747</b>	<b>237</b>
MOTE006	0.1928	0.0087	0.2701	0.7138	0.0322	F	-	-
MOTE007	0.6623	0.0231	0.9612	0.6890	0.0240	F	-	-
<b>MOTE008</b>	<b>0.0024</b>	<b>0.0034</b>	<b>0.0066</b>	<b>0.3676</b>	<b>0.5130</b>	<b>M</b>	<b>NA</b>	<b>0</b>
MOTE009	0.5339	0.0225	0.7813	0.6834	0.0288	F	-	-
MOTE010	0.2352	0.0070	0.3133	0.7506	0.0224	F	-	-
MOTE011	1.8240	0.0639	2.7361	0.6666	0.0234	F	-	-
MOTE012	0.0987	0.1087	0.2682	0.3679	0.4054	M	0.031871 $\pm$ 0.019592	109
MOTE013	0.0749	0.0946	0.2059	0.3640	0.4595	M	0.001022 $\pm$ 0.001021	1313
MOTE014	0.5954	0.8630	1.7505	0.3401	0.4930	M	0.00734 $\pm$ 0.00203	2554
MOTE015	0.0892	0.0037	0.1227	0.7268	0.0299	F	-	-
MOVE001	1.4228	0.0624	2.2134	0.6428	0.0282	F	-	-
MOVE002	0.6864	1.0665	2.0476	0.3352	0.5209	M	0.005353 $\pm$ 0.001602	2745
MOVE003	1.2516	0.0499	1.9124	0.6544	0.0261	F	-	-
MOVE004	0.4599	0.7649	1.3973	0.3291	0.5474	M	0.003709 $\pm$ 0.001640	1798
<b>MOVE005</b>	<b>0.0033</b>	<b>0.0044</b>	<b>0.0095</b>	<b>0.3509</b>	<b>0.4630</b>	<b>M</b>	<b>NA</b>	<b>0</b>
MOVE006	0.0148	0.0008	0.0201	0.7375	0.0381	F	-	-
MOVE007	0.7443	1.1538	2.2290	0.3339	0.5176	M	0.002547 $\pm$ 0.001126	3019
MOVE008	0.9546	1.5092	2.8504	0.3349	0.5294	M	0.004684 $\pm$ 0.001231	3924
<b>MOVE009</b>	<b>0.0326</b>	<b>0.0396</b>	<b>0.0867</b>	<b>0.3760</b>	<b>0.4570</b>	<b>M</b>	<b>NA</b>	<b>8</b>
MOVE010	0.1301	0.1911	0.4066	0.3200	0.4700	M	0.01296 $\pm$ 0.00834	293
<b>MOVE011</b>	<b>0.0079</b>	<b>0.0106</b>	<b>0.0256</b>	<b>0.3076</b>	<b>0.4153</b>	<b>M</b>	<b>NA</b>	<b>2</b>
SILV001	0.0056	0.0002	0.0079	0.7066	0.0232	F	-	-
<b>SILV002</b>	<b>0.0002</b>	<b>0.0001</b>	<b>0.0005</b>	<b>0.4857</b>	<b>0.1847</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>
<b>SILV003</b>	<b>0.0003</b>	<b>0.0002</b>	<b>0.0005</b>	<b>0.5700</b>	<b>0.4669</b>	<b>M?</b>	<b>NA</b>	<b>0</b>
MARZ001	0.0516	0.0733	0.1492	0.3457	0.4912	M	-0.009639 $\pm$ 0.029111	30

**Tabella 6.6** Risultati ottenuti dalla determinazione del sesso biologico degli individui in esame (M=sesso maschile, F=sesso femminile, NA=sesso indeterminato) e dalla stima della contaminazione nucleare per gli individui di sesso maschile (con relativo errore standard). L'ultima colonna mostra il numero di posizioni sul cromosoma X coinvolte nell'analisi, coperte almeno 2 volte. In rosso sono evidenziati i campioni per i quali non è stato possibile determinare in modo accurato il sesso e/o la contaminazione nucleare o per i quali vi è evidenza di contaminazione.

Per la maggior parte degli individui è stato possibile effettuare con successo la determinazione del sesso biologico calcolando i parametri Xrate ( $x\text{Cov}/\text{autCov}$ ) ed Yrate ( $y\text{Cov}/\text{autCov}$ ) secondo i criteri descritti nel *Paragrafo 5.3.2*. In linea generale, è stato assegnato sesso femminile agli individui con Xrate intorno all'1 e Yrate prossima allo 0, e sesso maschile agli individui che mostravano sia Xrate che Yrate approssimativamente intorno allo 0.5.

Visualizzando in un grafico (*Figura 6.2*) i valori di Xrate ed Yrate ottenuti per ciascun individuo, si osserva come i campioni di sesso differente (M e F) si distribuiscano formando due raggruppamenti ben distinti (il primo all'incirca al centro del grafico ed il secondo lungo l'asse Xrate), ad eccezione di 3 individui che risultano cadere in una posizione intermedia e corrispondono a quelli indicati in rosso in *Tabella 6.6*. Se per SILV002 e SILV003 questo risultato è verosimilmente determinato dalla scarsa conservazione del materiale genetico endogeno e, di conseguenza, dalla bassa copertura che si osserva su tutto il genoma umano, al contrario, la posizione di MOTE005 suggerisce la probabile presenza di contaminazione. Confrontando i valori di Xrate ed Yrate del campione MOTE005 con quelli del gruppo M, è possibile ipotizzare che si tratti di un individuo di sesso maschile (in accordo anche con la determinazione morfometrica [156]) con presenza di DNA contaminante femminile.



**Figura 6.2** Visualizzazione grafica della distribuzione dei campioni sulla base dei rispettivi valori di Xrate ed Yrate: M rappresenta il gruppo degli individui di sesso maschile e F quello degli individui di sesso femminile. La posizione intermedia osservata per SILV002 e SILV003 dipende dalla scarsa preservazione del materiale genetico, mentre per l'individuo MOTE005 sembra essere indicativa della presenza di DNA contaminante.



La determinazione del sesso ha permesso, inoltre, di evidenziare alcune incongruenze tra i risultati ottenuti dall'analisi genetica e dalla valutazione morfometrica per gli individui AGOS004, AGOS009, AZZA005, AZZA006, CAST002 e TAB016 inclusi nel primo set e per i campioni MOVE008, MOTE001 e MOTE002 compresi nel secondo (per questi le informazioni erano state estrapolate da [156, 157]). Per gli individui maschili è stato possibile procedere con la verifica della contaminazione nucleare, basata sulla valutazione del livello di eterozigosi riscontrato in specifiche posizioni localizzate sul cromosoma X. I valori riportati in *Tabella 6.6* mostrano le stime ottenute, accompagnate dai relativi errori standard, e il numero di SNPs con copertura pari o superiore a 2 coinvolti nell'analisi.

Tralasciando SILV002 e SILV003, per i quali non è stato possibile ottenere in primo luogo una stima accurata del sesso, sono in totale 9 i restanti individui che, per mancanza di un numero sufficiente di SNPs coperti, non hanno restituito informazioni in merito al grado di contaminazione nucleare.

In tutti gli altri casi i campioni mostrano valori di contaminazione decisamente trascurabili (< 0.3%), ad eccezione di MOVE005. La stima di contaminazione ottenuta, pari al 32% ( $\pm 3$ ), conferma i segnali già precedentemente osservati in fase di determinazione del sesso. In situazioni come questa è di norma consigliabile effettuare un filtraggio delle *reads* con il programma PMDtools [195], andando a trattenere nelle seguenti fasi di analisi solo quelle caratterizzate dalla presenza di deaminazioni.

In questo caso, l'esistenza di un legame di secondo grado (confermato dall'aplotipo mitocondriale) tra MOVE005 e un altro individuo proveniente dello stesso sito e con una migliore preservazione di DNA endogeno, ha permesso direttamente di escludere MOVE005 dalle analisi popolazionistiche.

Infine, è importante sottolineare che, sebbene la stima effettuata non evidenzia la presenza di una significativa frazione contaminante per l'individuo MARZ001, tuttavia il numero di SNPs considerati nell'analisi (pari a 30) risulta piuttosto esiguo e potrebbe influenzare l'attendibilità del risultato.

Inoltre, per tutti gli individui è stato stimato il valore di contaminazione a livello mitocondriale con il pacchetto schmutzi. In *Tabella 6.7* sono indicate le stime di contaminazione ottenute a partire dai dati di cattura nucleare (1240K, in presenza di sufficiente copertura sull'mtDNA) o mitocondriale (MT, per il momento condotta solo sui campioni del primo set, ad eccezione di MARZ001 per il quale era già stata effettuata in precedenza). In tabella viene mostrata la copertura media sull'mtDNA ottenuta dal mappaggio sulla rCRS, la percentuale di misincorporazioni rilevata da MapDamage ai primi nucleotidi terminali dei frammenti di DNA, il valore medio di contaminazione e, tra parentesi, l'intervallo stimato.

Per la maggior parte degli individui sono stati ottenuti valori di contaminazione trascurabili, che si attestano prevalentemente sotto la soglia del 4% (solo per uno e tre campioni rispettivamente è stata stimata una contaminazione pari al 4 e 5%). Fanno eccezione AGOS004, AGOS009, AZZA001 e AZZA013, i quali presentano valori superiori al 5%. Trattandosi di campioni di sesso maschile senza

evidenza di contaminazione nucleare, AGOS004 ed AGOS009 (7% di contaminazione per entrambi) sono stati comunque inclusi nelle successive analisi, prestando particolare attenzione a possibili alterazioni o incongruenze nei risultati; AZZA001 e AZZA013 (8% e 6% di contaminazione), entrambi di sesso femminile, sono stati invece esclusi dalle analisi popolazionistiche.

Nonostante sia stata effettuato l'arricchimento del genoma mitocondriale, per i campioni AGOS006 (F) e AZZA012 (M) non è stato comunque possibile determinare il livello di contaminazione a causa della scarsa copertura sull'mtDNA (valore medio di 4,28 e 2,27 rispettivamente). Non essendo in grado di valutare l'autenticità del dato, l'individuo AGOS006 è stato escluso dalle successive analisi popolazionistiche; allo stesso modo è stato escluso anche AZZA012 data l'ulteriore mancanza sia della stima sul cromosoma X che, in generale, di un numero sufficiente di SNPs nucleari coperti.

Una situazione simile è stata osserva anche per i campioni MOTE008, SILV001, SILV002 e SILV003. Infine, non è stato possibile stimare la contaminazione mitocondriale per MOVE009 a causa della difficoltà del software nel rilevare i bassi tassi di deaminazione registrati per questo campione (M, contaminazione chr-X=NA), in combinazione ad un *coverage* non elevato.

In fase di determinazione delle parentele, però, MOVE009 e MOVE008 sono stati come riconosciuti come campioni identici e verosimilmente riferibili al medesimo individuo. Per la conduzione delle analisi popolazionistiche, al momento non ancora effettuate (ad eccezione della PCA), sarà possibile dunque unire i dati derivanti dai due campioni, senza escludere il contributo di MOVE009.

Valutando il parametro di mt/nuclear ratio (*Tabella 6.4* del precedente paragrafo) per i campioni processati a partire dalle rocche petrose, che si attesta in tutti i casi sotto il 200, è stato possibile estendere l'assenza di rilevante contaminazione mitocondriale anche al genoma nucleare per tutti gli individui di sesso femminile e per alcuni degli individui maschili per i quali non era stato ottenuto nessun risultato in relazione al cromosoma X (AGOS005, AZZA006 e TAB011).

Ricapitolando, in seguito alla valutazione del grado di contaminazione sono stati esclusi dalle analisi popolazionistiche AGOS006 (F, contaminazione mtDNA=NA), AZZA001 (F, contaminazione mtDNA), AZZA012 (M, contaminazione=NA e bassa copertura) e AZZA013 (F contaminazione mtDNA).

Tra i campioni del secondo set, verranno esclusi dalle analisi MOTE008 (M, contaminazione=NA e bassa copertura), MOVE005 (M, contaminazione chr-X), SILV001 (F, contaminazione mtDNA=NA), SILV002 e SILV003 (entrambi M, contaminazione=NA e bassa copertura).

ID campioni	Sesso	Tipologia di dato	Copertura media mtDNA	% Deaminazioni 3' mtDNA	% Deaminazioni 5' mtDNA	Contaminazione mtDNA
AGOS001	F	1240K	14.17	0.21	0.22	0.01 (0-0.02)
AGOS002	M	1240K	5.75	0.16	0.15	0.01 (0-0.02)
AGOS003	M	1240K	17.25	0.08	0.07	0.01 (0-0.02)
<b>AGOS004</b>	<b>M</b>	<b>MT</b>	<b>830.30</b>	<b>0.05</b>	<b>0.05</b>	<b>0.07 (0.06-0.08)</b>
AGOS005	M	MT	13.32	0.04	0.03	0.02 (0.01-0.03)
<b>AGOS006</b>	<b>F</b>	<b>MT</b>	<b>4.28</b>	0.03	0.04	<b>NA</b>
AGOS007	F	MT	77.91	0.06	0.07	0.03 (0.02-0.04)
<b>AGOS009</b>	<b>M</b>	<b>MT</b>	<b>152.84</b>	<b>0.18</b>	<b>0.18</b>	<b>0.07 (0.06-0.08)</b>
AGOS010	M	MT	227.39	0.07	0.07	0.03 (0.02-0.04)
<b>AZZA001</b>	<b>F</b>	<b>MT</b>	<b>508.99</b>	<b>0.23</b>	<b>0.24</b>	<b>0.08 (0.07-0.09)</b>
AZZA002	M	MT	549.72	0.22	0.22	0.02 (0.01-0.03)
AZZA003	F	MT	873.54	0.17	0.17	0.02 (0.01-0.03)
AZZA004	F	1240K	6.75	0.11	0.09	0.03 (0.02-0.04)
AZZA005	M	1240K	5.00	0.05	0.06	0.01 (0-0.02)
AZZA006	M	MT	281.45	0.07	0.07	0.02 (0.01-0.03)
AZZA007	F	MT	688.81	0.17	0.17	0.02 (0.01-0.03)
AZZA008	M	MT	1107.55	0.06	0.06	0.05 (0.04-0.06)
AZZA009	M	MT	5.04	0.04	0.05	0.04 (0.03-0.05)
AZZA010	M	MT	491.14	0.13	0.13	0.02 (0.01-0.03)
AZZA011	F	MT	263.57	0.24	0.24	0.02 (0.01-0.03)
<b>AZZA012</b>	<b>M</b>	<b>MT</b>	<b>2.27</b>	<b>0.08</b>	<b>0.06</b>	<b>NA</b>
<b>AZZA013</b>	<b>F</b>	<b>MT</b>	<b>897.70</b>	<b>0.21</b>	<b>0.21</b>	<b>0.06 (0.05-0.07)</b>
AZZA014	M	1240K	13.27	0.09	0.14	0.01 (0-0.02)
CAS001	M	1240K	9.39	0.06	0.07	0.01 (0-0.02)
CAS005-ETR	F	1240K	17.47	0.06	0.06	0.02 (0.01-0.03)
CAS005-VILL	M	1240K	18.76	0.11	0.12	0.02 (0.01-0.03)
CAS020	F	1240K	12.66	0.09	0.11	0.01 (0-0.02)
CAS022	F	1240K	28.83	0.11	0.12	0.01 (0-0.02)
CAS028	M	1240K	15.20	0.06	0.08	0.01 (0-0.02)
CAS036	M	MT	281.49	0.10	0.10	0.05 (0.04-0.06)
CAS041-1	M	1240K	22.82	0.10	0.12	0.01 (0-0.02)
CAS041-2	F	1240K	18.42	0.12	0.12	0.01 (0-0.02)
CAS054	F	1240K	24.01	0.05	0.04	0.01 (0-0.02)
CAS067	M	1240K	21.21	0.08	0.09	0.02 (0.01-0.03)
CAS090	M	1240K	29.69	0.05	0.04	0.02 (0.01-0.03)
CAS095	F	1240K	16.94	0.06	0.04	0.01 (0-0.02)
CAST001	F	1240K	16.54	0.18	0.18	0.01 (0-0.02)
CAST002	F	1240K	26.78	0.10	0.10	0.01 (0-0.02)
MAR001	M	1240K	15.48	0.07	0.09	0.02 (0.01-0.03)
MAR003	F	1240K	10.87	0.15	0.18	0.01 (0-0.02)
MAR006	F	1240K	23.88	0.08	0.10	0.01 (0-0.02)
MAR012	M	1240K	13.92	0.12	0.12	0.01 (0-0.02)
MAR017	F	1240K	26.92	0.09	0.09	0.03 (0.02-0.04)
MAR018	F	1240K	12.88	0.08	0.06	0.01 (0-0.02)
MAR020	F	1240K	15.18	0.09	0.11	0.01 (0-0.02)
MAR021	F	1240K	6.20	0.08	0.09	0.02 (0.01-0.03)
TAB010	M	1240K	11.09	0.13	0.11	0.01 (0-0.02)
TAB011	M	MT	281.49	0.05	0.04	0.05 (0.04-0.06)
TAB012	F	MT	75.56	0.04	0.04	0.03 (0.02-0.04)
TAB014	F	MT	359.76	0.09	0.09	0.02 (0.01-0.03)
TAB016	M	1240K	12.00	0.07	0.05	0.01 (0-0.02)
VISA001	M	MT	784.55	0.17	0.17	0.02 (0.01-0.03)

ID campioni	Sesso	Tipologia di dato	Copertura media mtDNA	% Deaminazioni 3' mtDNA	% Deaminazioni 5' mtDNA	Contaminazione mtDNA
MOTE001	M	1240K	10.97	0.07	0.10	0.02 (0.01-0.03)
MOTE002	M	1240K	11.08	0.06	0.08	0.01 (0-0.02)
MOTE003	F	1240K	23.09	0.06	0.06	0.01 (0-0.02)
MOTE004	F	1240K	8.67	0.06	0.06	0.01 (0-0.02)
MOTE005	M?	1240K	79.47	0.05	0.06	0.01 (0-0.02)
MOTE006	F	1240K	45.32	0.07	0.09	0.01 (0-0.02)
MOTE007	F	1240K	3.39	0.10	0.10	0.02 (0.01-0.03)
<b>MOTE008</b>	<b>M</b>	<b>1240K</b>	<b>0.11</b>	<b>0.13</b>	<b>0.00</b>	<b>NA</b>
MOTE009	F	1240K	13.26	0.06	0.04	0.01 (0-0.02)
MOTE010	F	1240K	13.37	0.07	0.08	0.01 (0-0.02)
MOTE011	F	1240K	11.21	0.10	0.08	0.03 (0.02-0.04)
MOTE012	M	1240K	40.75	0.07	0.06	0.01 (0-0.02)
MOTE013	M	1240K	54.49	0.05	0.05	0.01 (0-0.02)
MOTE014	M	1240K	8.74	0.09	0.09	0.02 (0.01-0.03)
MOTE015	F	1240K	26.29	0.05	0.05	0.01 (0-0.02)
MOVE001	F	1240K	4.54	0.08	0.10	0.01 (0-0.02)
MOVE002	M	1240K	11.35	0.08	0.09	0.01 (0-0.02)
MOVE003	F	1240K	4.16	0.11	0.09	0.02 (0.01-0.03)
MOVE004	M	1240K	3.95	0.10	0.08	0.01 (0-0.02)
MOVE005	M	1240K	31.34	0.03	0.04	0.01 (0-0.02)
MOVE006	F	1240K	28.78	0.05	0.05	0.01 (0-0.02)
MOVE007	M	1240K	10.09	0.06	0.06	0.01 (0-0.02)
MOVE008	M	1240K	9.08	0.06	0.06	0.02 (0.01-0.03)
<b>MOVE009</b>	<b>M</b>	<b>1240K</b>	<b>10.50</b>	<b>0.04</b>	<b>0.05</b>	<b>NA</b>
MOVE010	M	1240K	4.85	0.05	0.04	0.02 (0.01-0.03)
MOVE011	M	1240K	10.27	0.05	0.06	0.01 (0-0.02)
<b>SILV001</b>	<b>F</b>	<b>1240K</b>	<b>0.01</b>	<b>0.50</b>	<b>0.00</b>	<b>NA</b>
<b>SILV002</b>	<b>NA</b>	<b>1240K</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>
<b>SILV003</b>	<b>M?</b>	<b>1240K</b>	<b>0.01</b>	<b>0.00</b>	<b>0.00</b>	<b>NA</b>
MARZ001	M	MT	133.26	0.43	0.41	0.02 (0.01-0.03)

**Tabella 6.7** Risultati relativi alla stima della contaminazione mitocondriale effettuata con schmutzi. I valori di copertura media sull'mtDNA e i tassi di misincorporazione (calcolati sui primi nucleotidi terminali) sono stati estrapolati in seguito al mappaggio con EAGER sul genoma mitocondriale e alle statistiche calcolate da MapDamage; 1240K e MT In rosso sono marcati gli individui per i quali non è stato possibile ottenere risultati o che mostrano segnali di contaminazione.

## 6.4 Markers uniparentali e legami biologici

L'analisi dei *markers* uniparentali (mtDNA e chr-Y) ha previsto l'assegnazione, quando possibile, dell'aplogruppo mitocondriale e del cromosoma Y. Per quanto riguarda il primo, in *Tabella 6.8* sono riportati gli aplogruppi mitocondriali identificati con Haplogrep a partire dalle sequenze consenso ottenute mediante schmutzi applicando una qualità di filtraggio pari a 30 (q30). Gli aplogruppi del chr-Y sono stati determinati per gli individui di sesso maschile con Yleaf considerando inizialmente solo gli SNPs marker con copertura pari o superiore a 2 (-r 2); per gli individui indicati con asterisco in *Tabella 6.8* è stato possibile ottenere l'assegnazione dell'aplogruppo solo abbassando tale soglia ad 1 (-r 1) a causa del ristretto numero di posizioni rappresentate a coperture maggiori.

I campioni per i quali non è stato possibile determinare entrambi gli aplogruppi sono evidenziati in rosso in tabella; per questi individui sarà possibile cercare di recuperare tali informazioni mediante strategie di arricchimento mirate al genoma mitocondriale e a specifici marcatori del cromosoma Y (ad eccezione, per quanto riguarda il DNA mitocondriale, dei campioni AGOS006 e AZZA012 che, sebbene siano stati sottoposti alla cattura dell'mtDNA, non hanno restituito una copertura ideale per l'assegnazione dell'aplogruppo mitocondriale).

In generale, gli aplogruppi mitocondriali riscontrati ricadono tutti all'interno della tipica variabilità euroasiatica [196], eccetto l'aplogruppo L3e1a3b individuato per l'individuo di epoca imperiale CAS054. Si tratta, infatti, di un aplogruppo di origine africana appartenente al clade L3e originario dell'Africa centrale o orientale (intorno a 46.000 anni fa), la cui diffusione nel sud del continente è stata associata alle migrazioni Bantu [197, 198]. Nello specifico, il suo sub-clade L3e1, frequente nei territori dell'Africa subsahariana e particolarmente comune nel sud-est del continente, è originario dell'Africa centrale ed è molto raro nelle aree dell'Africa occidentale [197]. Nonostante precedenti studi abbiano già rilevato la presenza in Italia centrale di individui caratterizzati da *ancestry* africana durante l'epoca imperiale romana [117], si tratta della prima volta in cui viene riportato un aplotipo mitocondriale specificamente africano e, soprattutto, appartenente al clade L3e1.

La questione dell'origine africana (e forse subsahariana) del campione CAS054 verrà affrontata più approfonditamente nei paragrafi successivi.

I restanti campioni dell'età imperiale ricadono all'interno dei macro-aplogruppi U (N=1) e H (N=2), già diffusi nella penisola italiana durante i periodi precedenti e generalmente frequenti in Eurasia occidentale.

La maggior parte dei campioni riferibili all'età del Ferro (41%) presenta un aplotipo mitocondriale appartenente al macro-aplogruppo H; seguono, in ordine di frequenza, gli aplogruppi U (14%), T e J (12%), K (6%), W (4%) e, infine, V ed I (1%). Si tratta di aplogruppi mitocondriali già riscontrati con

frequenze simili in precedenti studi genetici incentrati su popolazioni italiane dell'età del Ferro e del periodo repubblicano [115, 117, 199] e già diffusi in Europa nei periodi precedenti.

Per il 9% degli individui non è stato possibile determinare l'aplogruppo mitocondriale (NA) a causa dell'insufficiente copertura sull'mtDNA.

Le frequenze degli aplogruppi mitocondriali identificati sono visibili nel grafico cumulativo di *Figura 6.3* (pagina 88), a fianco al quale vengono riportate anche le frequenze relative agli aplogruppi del chr-Y.

In merito ai campioni dell'età del Ferro, come già evidenziato da precedenti lavori [115, 117], si osserva la netta prevalenza dell'aplogruppo R1b (74%), rappresentato dai polimorfismi R1b-P312 (R1b1a1b1a1a2) e dal suo derivato R1b-L2 (R1b1a1b1a1a2b1), diffusi in Europa parallelamente alla componente della steppa in associazione alla cultura Bell-Beaker [106]. La loro elevata frequenza è coerente con l'arrivo della *steppe-related ancestry* nelle regioni dell'Italia peninsulare durante l'età del Bronzo.

In frequenza minore compaiono gli aplogruppi G2 (in particolare G2a, 10%) e I (I2a, 2%). Il primo, già riportato per individui italiani dell'età del Ferro [117], è l'aplogruppo prevalente in Italia durante il Neolitico; trattandosi dell'aplogruppo dominante nelle prime popolazioni di agricoltori-allevatori dell'Anatolia nord-occidentale, la sua diffusione nel continente europeo è associata alla rivoluzione neolitica [115]. I sub-cladi dell'aplogruppo I2a sono, dopo l'aplogruppo G, tra quelli più diffusi nelle popolazioni di agricoltori-allevatori dell'Europa centrale [200]; il sub-clade I2a1, a cui appartiene l'unico campione con macro-aplogruppo I (AZZA005), sembra essersi originato prima del LGM (*Last Glacial Maximum*) ed essere sopravvissuto grazie ai rifugi situati in Europa sud-orientale [201].

Gli aplogruppi del cromosoma Y dei due campioni di sesso maschile datati all'età imperiale romana ricadono all'interno dei macro-aplogruppi R1b e J (nello specifico nel sub-clade J2a, già attestato da precedenti studi in gruppi di cacciatori-raccoglitori in Italia, nei Balcani e in Anatolia [93, 94]).

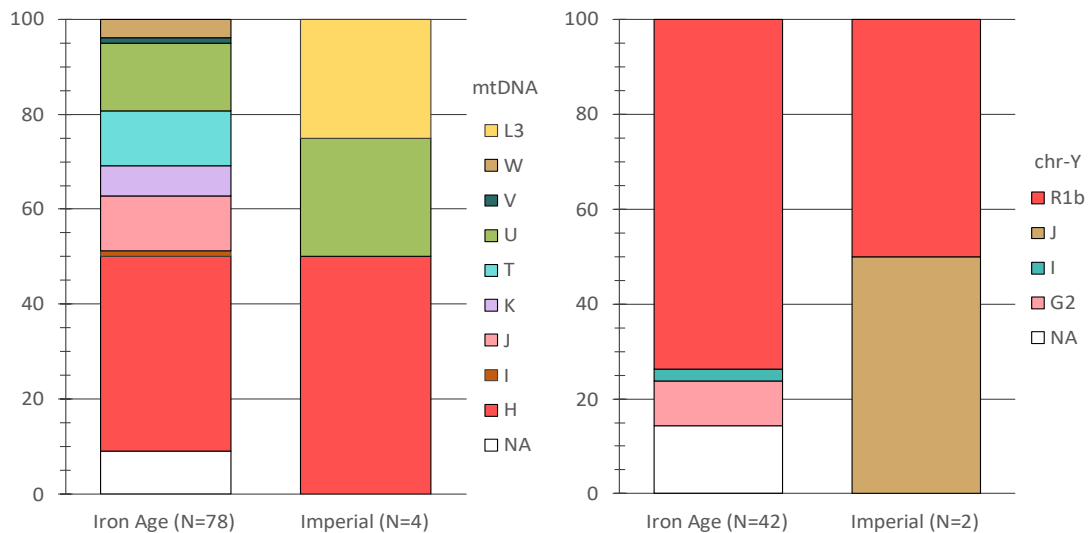
Infine, in assenza di un numero sufficiente di marcatori non è stato possibile assegnare l'aplogruppo del chr-Y al 14% degli individui in esame, tutti appartenenti all'età del Ferro.

ID campioni	Sesso	Aplogruppo mtDNA	Copertura media mtDNA	Aplogruppo chr-Y	Markers chr-Y (cop>2)
AGOS001	F	K1d	14.17	-	-
AGOS002	M	H83	5.75	R1b1a1b1a1a2	1596
AGOS003	M	H3+16311	17.25	R1b1a1b1a1a2b1	4247
AGOS004_mtDNA_contam	M	H46	830.30	G2a2b2a1a1b	1803
<b>AGOS005</b>	<b>M</b>	<b>H10d</b>	<b>13.32</b>	<b>NA*</b>	<b>457*</b>
<b>AGOS006</b>	<b>F</b>	<b>NA</b>	<b>4.28</b>	<b>-</b>	<b>-</b>
AGOS007	F	H3k	77.91	-	-
AGOS009_mtDNA_contam	M	U1a1a	152.84	R1b1a1b1a1a2	1068
AGOS010	M	U5a1b1	227.39	R1b1a1b1a	334
AZZA001_mtDNA_contam	F	H1n+146+195	508.99	-	-
AZZA002	M	HV0+195	549.72	R1b1a1b1a1a2*	2634*
AZZA003	F	T2b3	873.54	-	-
AZZA004	F	J1c3e2	6.75	-	-
AZZA005	M	J1c	5.00	I2a1b2	1679
AZZA006	M	W6a	281.45	R1b1a1b1a1a1*	1272*
AZZA007	F	T2b3	688.81	-	-
AZZA008	M	H60a	1107.55	R1b1a1b1a1a2b1	2109
AZZA009	M	J1c12b	5.04	R1*	423*
AZZA010	M	HV16	491.14	G2a2b2a1a1b	2036
AZZA011	F	H1e1a	263.57	-	-
<b>AZZA012_bassa_cop</b>	<b>M</b>	<b>NA</b>	<b>2.27</b>	<b>NA*</b>	<b>307*</b>
AZZA013_mtDNA_contam	F	H10h	897.70	-	-
AZZA014	M	HV0	13.27	R1b1a1b1a1a2b1	2531
CAS001	M	T1a1	9.39	R1b1a1b1a1a2b1	723
CAS005-ETR	F	W5	17.47	-	-
CAS005-VILL	M	U5b3a2	18.76	R1b1a1b1a1a2	3379
CAS020	F	T2a1b1a1	12.66	-	-
CAS022	F	HV0a	28.83	-	-
CAS028	M	W6a	15.20	R1b1a1b1a1a2	3239
CAS036	M	H5b1	281.49	R1b1a1b	981
CAS041-1	M	U5b2b4a	22.82	R1b1a1b1a1a2d1a	4480
CAS041-2	F	H1	18.42	-	-
CAS054	F	L3e1a3b	24.01	-	-
CAS067	M	H13a1a4	21.21	R1b1a1b1a1a2b3c	4962
CAS090	M	U1b2	29.69	J2a1a4	3842
CAS095	F	H15b1	16.94	-	-
CAST001	F	U4a	16.54	-	-
CAST002	F	U5a2a1	26.78	-	-
MAR001	M	J2b1a1	15.48	R1b1a1b1a1a2	3571
MAR003	F	H1u	10.87	-	-
MAR006	F	J2b1a1	23.88	-	-
MAR012	M	V17	13.92	R1b1a1b1a1a2b1	4618
MAR017	F	H3+16311	26.92	-	-
MAR018	F	HV0+195	12.88	-	-
MAR020	F	T2b3	15.18	-	-
MAR021	F	K1b1a1	6.20	-	-
TAB010	M	HV+16311	11.09	R1b1a1b1a1a2	3452
TAB011	M	K1a+195	281.49	R1b1a1b1a1*	1001*
TAB012	F	K1a4g	75.56	-	-
TAB014	F	HV0+195	359.76	-	-
TAB016	M	I1	12.00	G2a2b2a1a1b1a1a1a	3894
VISA001	M	T2c1d+152	784.55	R1b1a1b1a1a2	1026

ID campioni	Sesso	Aplogruppo mtDNA	Copertura media mtDNA	Aplogruppo chr-Y	Markers chr-Y (cop>2)
MOTE001	M	T2b3	10.97	R1b1a1b1a1a2b1	4836
MOTE002	M	K1a1b1g	11.08	R1b1a1b1b3a1a	5321
MOTE003	F	H+16311	23.09	-	-
MOTE004	F	HV0+195	8.67	-	-
MOTE005_X-chr_contam	M?	U5a2c3a	79.47	R1b1a1*	1791*
MOTE006	F	U5a2c3a	45.32	-	-
MOTE007	F	H4a1a1a	3.39	-	-
<b>MOTE008_bassa_cop</b>	<b>M</b>	<b>NA</b>	<b>0.11</b>	<b>NA*</b>	<b>72*</b>
MOTE009	F	H7	13.26	-	-
MOTE010	F	J1c1	13.37	-	-
MOTE011	F	HV0+195	11.21	-	-
MOTE012	M	H5b	40.75	R1b1a1b1a1a*	2517*
MOTE013	M	J1c1	54.49	R1b	176
MOTE014	M	H7	8.74	R1b1a1b1a1a2b1	4273
MOTE015	F	J1c3g	26.29	-	-
MOVE001	F	H+152	4.54	-	-
MOVE002	M	T2b+16362	11.35	R1b1a1b1a	4716
MOVE003	F	H6b	4.16	-	-
MOVE004	M	J1c3	3.95	R1b1a1b1a1a2	3347
<b>MOVE005_bassa_cop</b>	<b>M</b>	<b>U5b1d1</b>	<b>31.34</b>	<b>NA*</b>	<b>94*</b>
MOVE006_bassa_cop	F	U5b1d1	28.78	-	-
MOVE007	M	H+152	10.09	R1b1a1b1a1a2	5202
MOVE008	M	T2b	9.08	R1b1a1b1a1a2	6407
<b>MOVE009</b>	<b>M</b>	<b>NA</b>	<b>10.50</b>	<b>R1b1a1b1a1a2*</b>	<b>862*</b>
MOVE010	M	H2a1a	4.85	G2a	615
<b>MOVE011_bassa_cop</b>	<b>M</b>	<b>H1</b>	<b>10.27</b>	<b>NA*</b>	<b>243*</b>
<b>SILV001_bassa_cop</b>	<b>F</b>	<b>NA</b>	<b>0.01</b>	<b>-</b>	<b>-</b>
<b>SILV002_bassa_cop</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>
<b>SILV003_bassa_cop</b>	<b>M?</b>	<b>NA</b>	<b>0.01</b>	<b>NA*</b>	<b>5*</b>
MARZ001	M	U5a1g	133.26	R1b1a1b*	1278*

**Tabella 6.8** Aplogruppi mitocondriali e del cromosoma Y identificati nei campioni in esame. Sono riportati anche i valori di copertura media sull' mtDNA ed il numero di SNPs marker disponibili per ciascun individuo maschile per l'assegnazione dell'aplogruppo del chr-Y. In rosso sono evidenziati i campioni per i quali non è stato possibile determinare entrambi gli aplogruppi. \* Identificazione dell'aplogruppo del chr-Y considerando anche *markers* con copertura minima di 1 (al posto di filtrare per coperture pari o superiori a 2).





**Figura 6.3** Grafici cumulativi rappresentanti le frequenze degli aplogruppi dell' mtDNA e del cromosoma Y identificati nei campioni in esame, suddivisi per periodo temporale di appartenenza.

Gli eventuali legami biologici esistenti tra gli individui in esame sono stati indagati al fine di ottenere informazioni circa la struttura sociale delle comunità analizzate e le possibili tradizioni funerarie ad essa associate, nonché minimizzare l'introduzione di *bias* nelle analisi di genetica di popolazione. All'interno del primo set di campioni è stato identificato mediante il software RAED un solo legame di parentela tra gli individui MAR001 (M, 17-20 anni) e MAR006 (F, 20-29 anni), provenienti dalla necropoli di Giardini Margherita (MAR006 dalla Tomba Grande menzionata nel *Paragrafo 5.1.5*). Si tratta di un rapporto di parentela di primo grado, confermato anche dal coefficiente di relazione pari  $\alpha=0,496$  (ottenuto dal calcolo della *pairwise mismatch rate*, come descritto nel *Paragrafo 5.3.5*) e dal fatto che i due individui condividono lo stesso aplogruppo del genoma mitocondriale J2b1a1. Alla luce di tali evidenze e considerando l'età alla morte, i due individui potrebbero essere madre e figlio oppure fratelli. In *Figura 6.4* è riportato un estratto del *plot* ottenuto da READ in cui è possibile osservare il risultato restituito per la coppia MAR001-MAR006; per tutte le altre coppie di individui non è stata evidenziata l'esistenza di nessun legame biologico stretto.

Data la parentela di primo grado, solo il campione caratterizzato da un numero maggiore di SNPs coperti (MAR006) è stato incluso delle analisi di genetica di popolazione.

Più numerosi risultano, invece, i legami biologici identificati all'interno del secondo set di campioni. Gli individui MOTE008, MOVE005, SILV001, SILV002 e SILV003 sono stati esclusi da questa analisi in quanto l'estremamente limitata copertura sul genoma umano (pari o inferiore allo 0,01) avrebbe potuto influenzare l'attendibilità del dato finale.

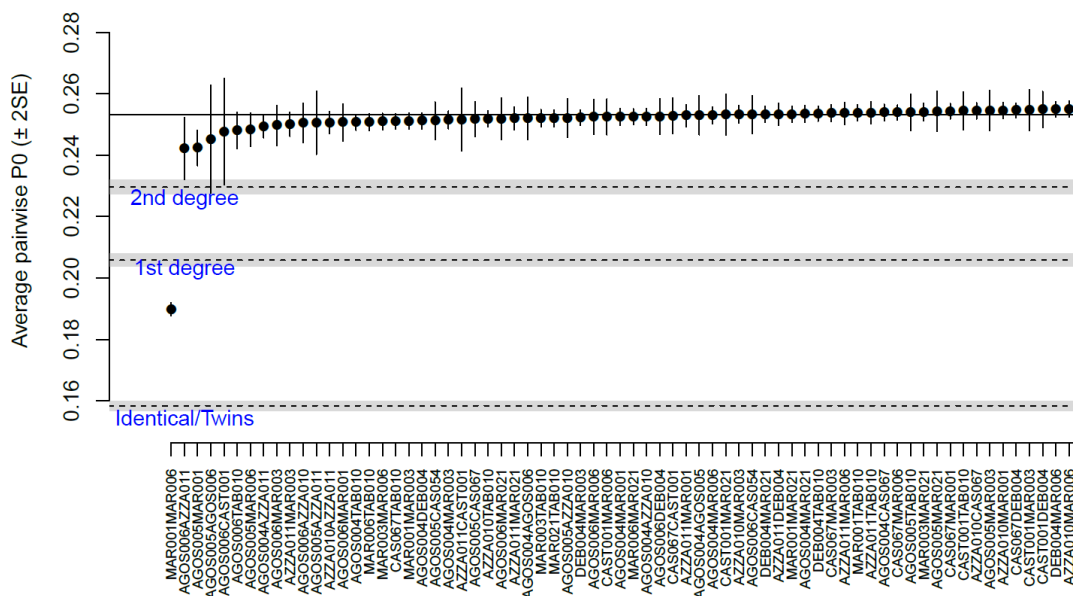
Come illustrato dal grafico in *Figura 6.5*, MOVE008 e MOVE009 (derivanti dallo stesso sito) risultano essere due individui gemelli o due campioni appartenenti ad un unico individuo. Dal momento che

i dati generati derivano da due elementi scheletrici diversi (rocca petrosa per MOVE008 e dente per MOVE009), è verosimile ipotizzare che si tratti di campioni riferibili ad uno stesso individuo. Ad ogni modo, i risultati di READ sono supportati dal valore del coefficiente  $\alpha$  (0,956), dalla determinazione del sesso (entrambi M) e dall'aplogruppo del chr-Y (entrambi R1b1a1b1a1a2); è assente il confronto dell'aplogruppo mitocondriale in quanto, per il momento, non è stato possibile determinarlo per il campione MOVE009. Inoltre, sulla base dell'analisi morfometrica i campioni sono stati associati allo stesso intervallo di età alla morte (20-35 anni) [156]. Per tali motivi, i dati generati dai due campioni saranno uniti prima di procedere con le analisi di genetica di popolazione.

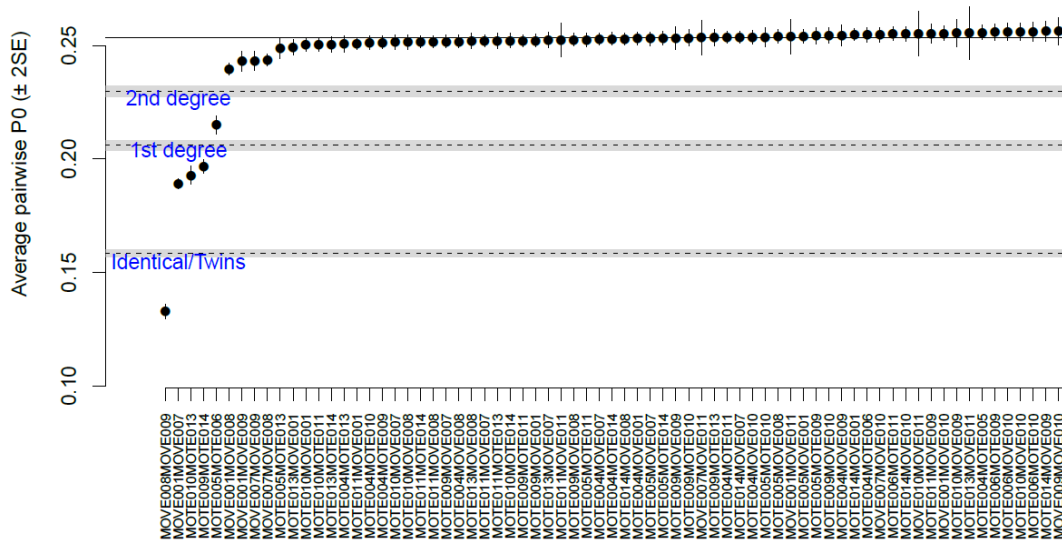
In aggiunta, per le seguenti tre coppie di individui è stata identificata una relazione primo grado: MOVE001-MOVE007 ( $\alpha=0,503$ ), MOTE010-MOTE013 ( $\alpha=0,449$ ) e MOTE009-MOTE014 ( $\alpha=0,458$ ). Trattandosi in tutti e tre i casi di un individuo di sesso maschile e uno di sesso femminile, è stato possibile confermare il legame osservato confrontando gli aplogruppi mitocondriali determinati in precedenza (per entrambi gli individui della coppia rispettivamente H+152, J1c1 e H7).

Tramite READ è stata individuata anche una relazione di secondo grado tra il campione maschile MOTE005 e quello femminile MOTE006 ( $\alpha=0,274$ , per entrambi aplogruppo mtDNA U5a2c3a).

Infine, il calcolo delle *pairwise mismatch rates* (PMR) ha evidenziato un possibile legame di terzo grado tra i campioni MOVE002 (M) e MOVE007 (F), suggerito dal valore del coefficiente di relazione  $\alpha$  pari a 0,136 ma non confermato da altre evidenze.



**Figura 6.4** Visualizzazione parziale del grafico restituito dal *tool*/READ in cui si evidenzia il rapporto di parentela di primo grado individuato per la coppia di campioni MAR001-MAR006.



**Figura 6.5** Visualizzazione parziale del grafico restituito dal *tool* READ in cui si evidenziano i legami di parentela riscontrati nel set di campioni provenienti dalle necropoli di Monterenzio Vecchio e Monte Tamburino.

Sulla base dei risultati appena discussi, per ciascuna coppia di individui con un legame di primo o secondo grado, solamente il campione con il maggior numero di SNPs coperti (e non contaminato) verrà trattenuto per le analisi popolazionistiche: MOVE006, MOVE007, MOTE010 e MOTE014.

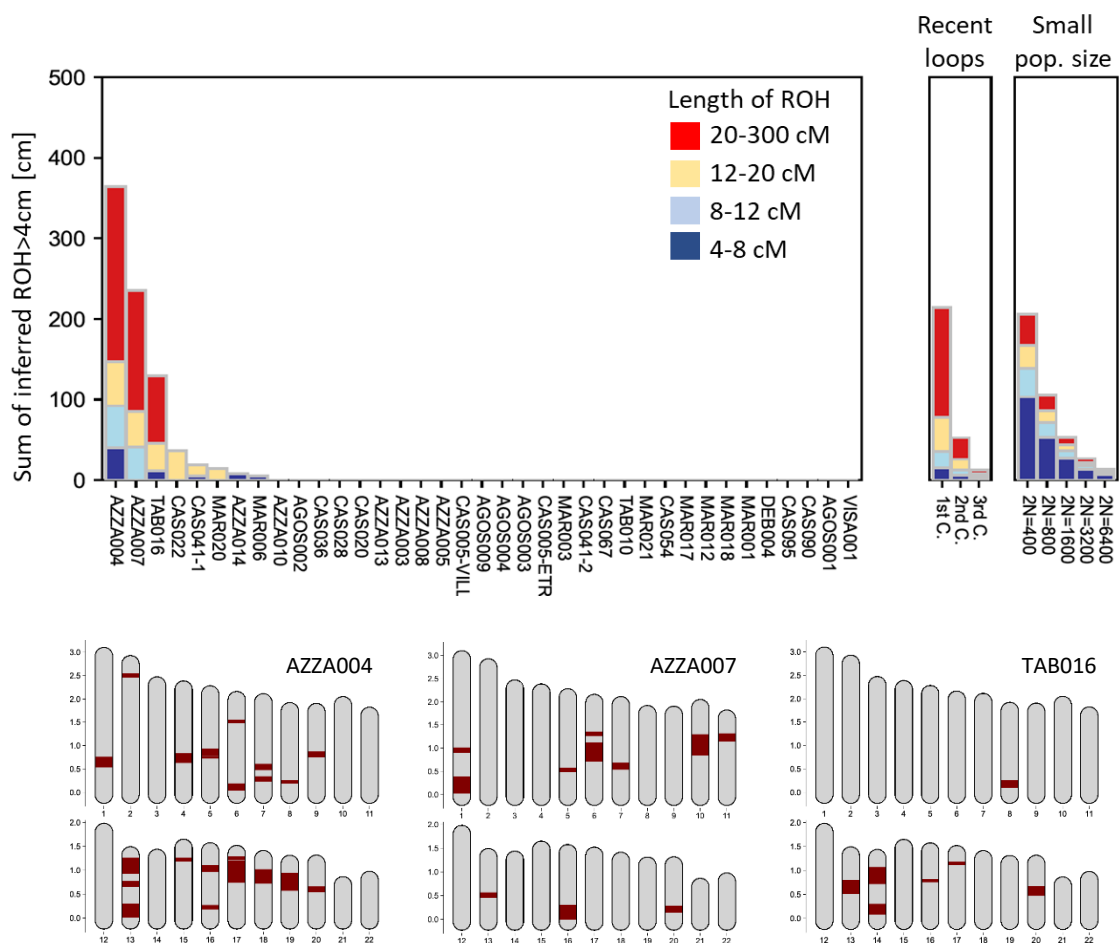
L'analisi delle *runs of homozygosity* (ROH) è stata al momento condotta solo sui campioni del primo set caratterizzati da una copertura minima di almeno 400.000 SNPs. Nel grafico di *Figura 6.6* sono mostrati i risultati ottenuti dal software hapROH: nel pannello di sinistra è rappresentata la somma delle *runs of homozygosity* identificate negli individui in esame e distinte, sulla base della lunghezza, nei quattro colori visibili in legenda, mentre il pannello a destra rappresenta i profili attesi nel caso di recente *inbreeding* (*Recent loops*), suggerito dall'elevata frequenza di ROH di lunghe dimensioni (>20 cM), o di una popolazione caratterizzata da una ristretta dimensione effettiva (*Small pop. size*), suggerita dall'elevata presenza di ROH di corte dimensioni (4-8 cM).

Ciò che si osserva per gli individui AZZA004, AZZA007 e TAB016 è la significativa presenza di ROH di lunghe dimensioni (rappresentate in rosso), indicative del fatto che i rispettivi genitori erano legati da stretti rapporti di parentela biologica. Confrontando i risultati restituiti dai campioni etruschi con i profili di riferimento forniti dal programma (dove la C. indicata nelle etichette significa "*cousins*"), si può dedurre che il profilo di AZZA007 potrebbe corrispondere a quello atteso per un individuo i cui genitori risultano essere primi cugini. Si può dedurre, invece, l'esistenza di un rapporto meno stretto tra i genitori di TAB016 ed ancora più stretto, invece, nel caso di AZZA004.

L'evidenza di *inbreeding* in due campioni provenienti dalla necropoli di Piazza Azzarita ha permesso di ipotizzare l'esistenza di un'élite o, comunque, di una comunità ristretta associata a tale sito.

Dal momento che solo alcuni degli individui provenienti da questa necropoli sono stati sottoposti a studio paleogenetico, sarebbe interessante poter approfondire l'argomento ampliando il numero di campioni coinvolti nelle analisi; potrebbero trasparire, in aggiunta, probabili legami di parentela attualmente non ancora riscontrati.

Lo stesso discorso può essere esteso anche al campione TAB016, soprattutto alla luce delle ipotesi avanzate sull'utilizzo della necropoli dell'area ex Manifattura Tabacchi da parte di un singolo nucleo familiare o, comunque, di una comunità ristretta.



**Figura 6.6** In alto: risultati della stima delle *runs of homozygosity* (ROH) effettuata con il software hapROH; nel pannello di sinistra sono riportati i risultati ottenuti sui campioni in esame, mentre a destra sono mostrati i profili di ROH attesi nel caso di recenti eventi di inbreeding (prevalenza di ROH lunghe) o di ristretta dimensione effettiva della popolazione (prevalenza di ROH corte). In basso: distribuzione delle ROH lungo i 22 autosomi negli individui AZZA004, AZZA007 e TAB016.

## 6.5 PCA e ADMIXTURE

L'analisi delle componenti principali (PCA) ha consentito di ottenere una prima visualizzazione della variabilità genetica che caratterizza il dataset in esame. I risultati relativi alla PCA sono stati integrati con quelli ottenuti dall'analisi con ADMIXTURE [186], grazie alla quale sono stati stimati i contributi genetici che, nel loro insieme, meglio descrivono il profilo degli individui analizzati.

Entrambe le analisi sono state condotte sul pannello Human Origins includendo gli individui non contaminati con almeno 20.000 SNPs coperti; l'analisi di ADMIXTURE, al momento effettuata solo sul primo set di campioni, ha previsto la rimozione preliminare, per ciascuna coppia di individui tra loro relazioni, dell'individuo con minore copertura in termini di SNPs nucleari.

In *Figura 6.7* (pagina 94) è visibile una PCA dell'Eurasia occidentale in cui sono indicate in grigio le popolazioni moderne utilizzate per la costruzione dello spazio di variabilità genetica (con l'aggiunta dei moderni toscani [TSI.SG], solo proiettati a causa della bassa copertura); con triangoli colorati sono indicate, invece, le componenti genetiche distali utilizzate anche per le successive analisi.

Oltre ai campioni in esame, sono stati proiettati sull'inverso della PC1 e sulla PC2 genomi antichi di confronto disponibili in letteratura, secondo quanto descritto in legenda [110, 115, 117, 118, 142].

I nuovi campioni dell'area bolognese sono riportati con simboli differenti sulla base del periodo di appartenenza (età del ferro ed epoca imperiale) e con il simbolo senza riempimento per indicare gli individui per il quali si dispone della datazione al radiocarbonio. In *Figura 6.7* vengono riportati esclusivamente i campioni del primo set, mentre quelli del secondo sono trattati separatamente a fine paragrafo.

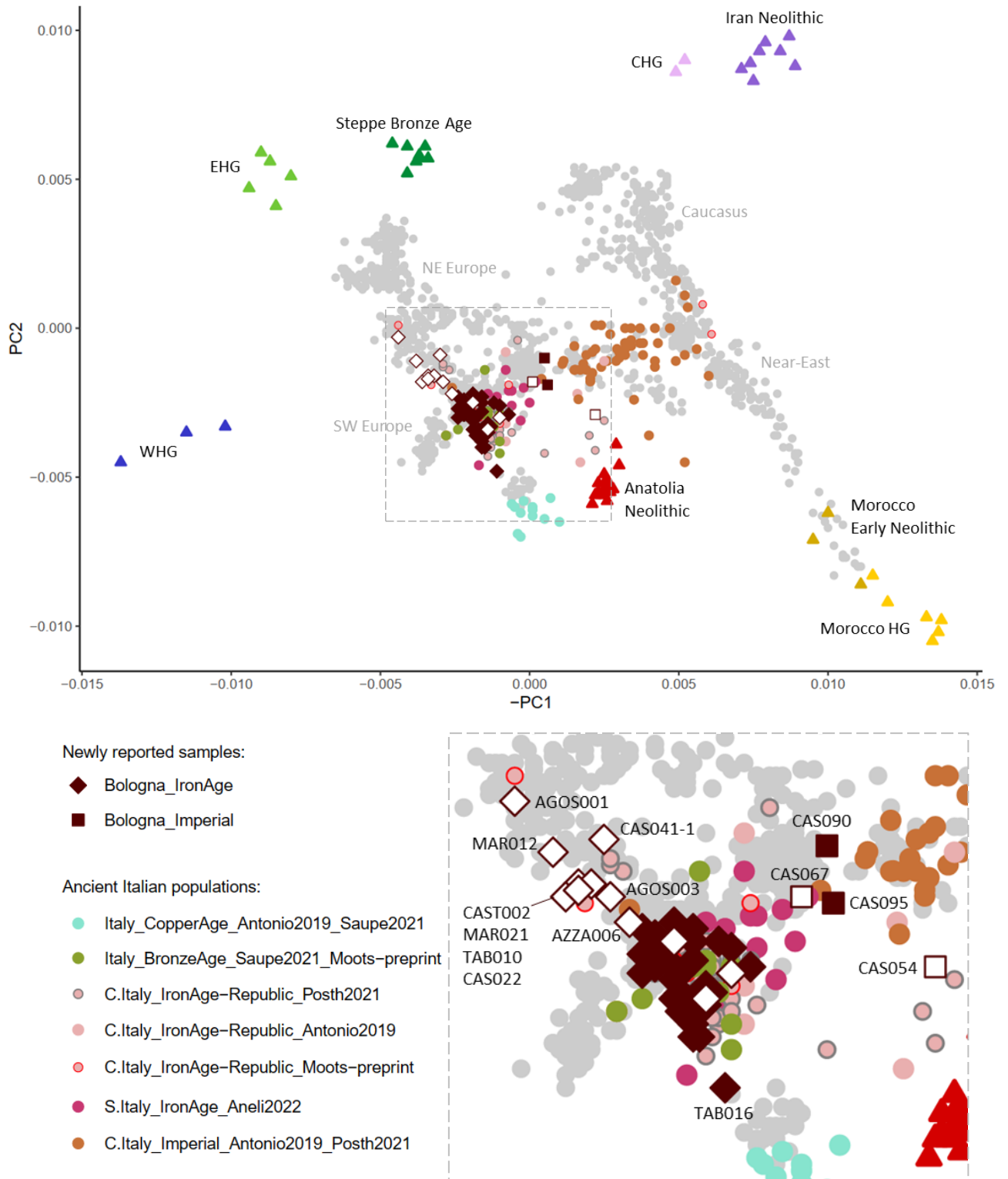
La maggior parte dei campioni etruschi dell'area bolognese forma un cluster in corrispondenza dei moderni Spagnoli, sovrapponendosi al segnale genetico già osservato per altre popolazioni italiane dell'età del Ferro (quali individui etruschi provenienti dall'Etruria classica, Latini e in parte Dauni [110, 115, 117, 118]) e per i precedenti gruppi dell'età del Bronzo [110, 142].

Oltre all'individuo TAB016, che si allontana sensibilmente dal cluster principale in direzione delle popolazioni calcolitiche (vedi zoom di *Figura 6.7*), si evidenzia la presenza di un numero significativo di campioni spostati verso il nord del cline europeo. Quest'ultimi cadono in prossimità di campioni, disponibili in letteratura, provenienti da necropoli etrusche e geneticamente associati a popolazioni dell'Europa centrale di probabile estrazione celtica (cluster *C.Italy\_Etruscan.Ceu* descritto in [117], R10339.SG e R10342.SG analizzati in [110]). Diversamente da quanto riscontrato in Italia centrale, questi individui, disposti secondo un cline che parte dal cluster principale (forse da AZZA006 o dagli adiacenti AZZA005 e CAST001) in direzione delle popolazioni del centro Europa, sembrano formare un gruppo piuttosto eterogeneo e costituiscono quasi il 17% del dataset padano dell'età del Ferro.

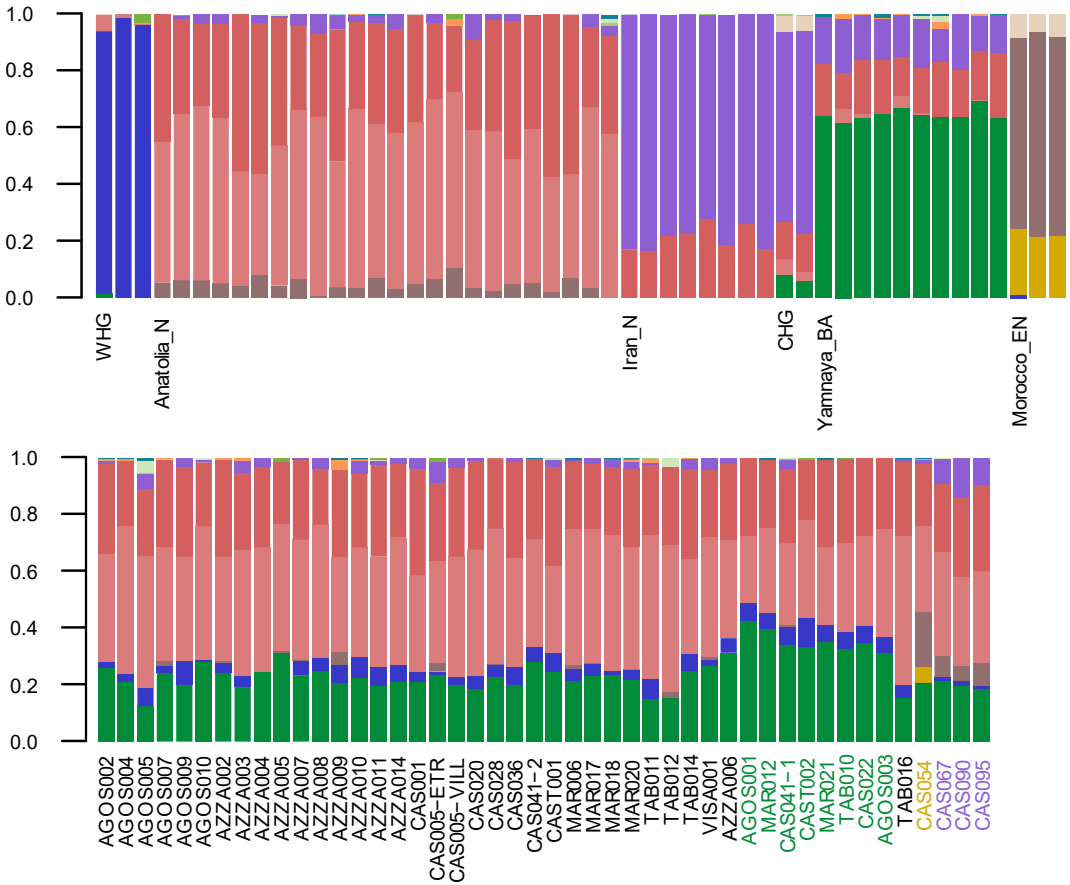
È importante sottolineare, inoltre, che provengono da necropoli differenti e sono tutti datati tra l'VIII e il V secolo a.C. (intervallo in cui si collocano anche i tre campioni datati del cluster principale), ad eccezione di MAR012 che risulta essere più recente (V-III secolo a.C.). Tra questi, l'individuo maggiormente spostato verso il nord del cline europeo (in prossimità del campione R10339.SG) proviene dalla necropoli di Piazza VIII Agosto ed è datato al periodo 769-516 a.C. (AGOS001).

I due campioni datati al radiocarbonio al periodo imperiale (CAS067 e CAS054, Casalecchio di Reno) si collocano in punti differenti della PCA: il primo cade in una posizione intermedia tra la variabilità genetica del periodo precedente e quella osservata per campioni coevi provenienti dal centro Italia [115, 117], mentre il secondo si posiziona in prossimità di campioni dell'età del Ferro caratterizzati da un'*ancestry* africana (cluster *C.Italy\_Etruscan.Afr* riportato in [117]). Sulla base della posizione occupata in PCA e della vicinanza all'individuo CAS067, anche i due campioni non ancora datati provenienti da Casalecchio di Reno (CAS090 e CAS095) possono essere verosimilmente associati al periodo imperiale.

Rispetto al cluster dell'età del Ferro, gli individui imperiali dell'area bolognese sono interessati da uno shift in direzione del cline del Vicinio Oriente che, come già riportato in studi precedenti [115, 117], comporta un incremento dell'affinità genetica con le popolazioni del Mediterraneo orientale. La posizione intermedia occupata in PCA dai campioni analizzati suggerisce che questo processo, determinato dalle intense dinamiche di mobilità che caratterizzano il periodo imperiale, potrebbe aver influenzato in maniera più lieve l'area padana. Considerato però il limitato numero di individui in esame riferibili a questo periodo (tutti provenienti dal medesimo sito), è necessario sottolineare che quanto appena osservato potrebbe non essere rappresentativo delle dinamiche complessive che caratterizzano l'area padana in questo momento storico.



**Figura 6.7** PCA costruita considerando la variabilità genetica di 62 popolazioni moderne dell’Eurasia occidentale (punti grigi) sulla quale sono stati proiettati gli individui del primo set con l’aggiunta di altri genomi antichi di riferimento (vedi legenda). I rombi e i quadrati senza riempimento rappresentano i campioni dell’area bolognese per i quali è disponibile la datazione al radiocarbonio; i triangoli colorati indicano, invece, la posizione in PCA di individui associati alle componenti genetiche distali utilizzate nelle successive analisi. In basso a destra è riportato un ingrandimento dell’area indicata in PCA. SW e NE si riferiscono rispettivamente a *southwestern* e *northeastern Europe*.



**Figura 6.8** Plot generato a partire dai risultati di ADMIXTURE, lanciato in modalità *unsupervised* (K=13). Nel grafico in alto sono riportati alcuni individui antichi di confronto, mentre in quello sottostante i nuovi campioni dell'area bolognese (con l'esclusione di MAR001 in quanto legato a MAR006 da una relazione di primo grado). BA indica Bronze Age, mentre EN Early Neolithic.



L'analisi con ADMIXTURE, effettuata in modalità *unsupervised*, ha permesso di confermare alcuni *patterns* già visibili in PCA. Gli individui riferibili all'età del Ferro (tutti quelli riportati in *Figura 6.8* ad esclusione dei campioni indicati in viola e giallo) risultano caratterizzati prevalentemente da una componente caratteristica dei cacciatori-raccoglitori europei (WHG, blu), da due componenti neolitiche associate alle migrazioni dei primi agricoltori-allevatori dall'Anatolia (Anatolia\_N, due sfumature di rosa) e massimizzate in popolazioni Europee datate tra il Neolitico ed il Calcolitico e, infine, dalla componente della steppa, correlata all'espansione della popolazione Yamnaya avuta luogo durante la prima età del Bronzo (Yamnaya\_BA, verde). L'*ancestry* associata alle popolazioni neolitiche della regione dell'Iran (Iran\_N, viola) è presente in proporzione trascurabile nella maggior parte dei campioni dell'età del Ferro; risulta leggermente più elevata nel campione CAS005-ETR che, tra gli individui del cluster principale etrusco, è uno di quelli maggiormente spostati in PCA in direzione del cline intermedio costituito dalle moderne popolazioni del bacino del Mediterraneo. I campioni indicati in verde in *Figura 6.8* presentano una maggior frazione di *steppe-related ancestry* rispetto agli altri individui dell'età del Ferro e corrispondono esattamente al gruppo di campioni che in PCA si distacca dal cluster principale, spostandosi verso il nord del cline europeo; tale *pattern* era già stato osservato in Etruria classica per il gruppo *C.Italy\_Etruscan.Ceu* descritto in [117]. Tra questi campioni, AGOS001 è l'individuo caratterizzato dalla percentuale maggiore di componente della steppa.

I risultati finora descritti sono in linea con quanto ottenuto per i campioni etruschi toscani e laziali riportati in Posth *et al.* [117].

Il campione TAB016, che in PCA risultava sensibilmente spostato verso i gruppi italiani del periodo calcolitico, si caratterizza, invece, per una minore frazione di *steppe ancestry*, accompagnata da un incremento, nel loro insieme, delle componenti Anatolia\_N + WHG. La carenza di *steppe ancestry* potrebbe suggerire l'appartenenza di questo campione ad un periodo precedente, durante il quale il processo di diffusione di tale componente era ancora in corso. Tuttavia, sarà possibile verificare questa ipotesi solo in presenza di una datazione diretta al radiocarbonio.

Proporzioni di componente della steppa paragonabili a quelle riscontrate in TAB016 sono osservate solo per il campione TAB012, che presenta una copertura mediamente bassa sul pannello utilizzato (~25.000 SNPs), e per AGOS005 e TAB011 che, pur non distinguendosi dal cluster principale, sono tra i campioni più spostati in direzione di TAB016.

Analizzando i campioni imperiali CAS067, CAS090 e CAS095, riportati in viola nel grafico di *Figura 6.8*, si evidenzia un incremento della componente Iran\_N rispetto al periodo precedente, in accordo allo *shift* osservato in PCA e a quanto già descritto relativamente al periodo imperiale in [115, 117]. L'*outlier* CAS054 presenta, invece, una componente associata ad individui provenienti dal Marocco

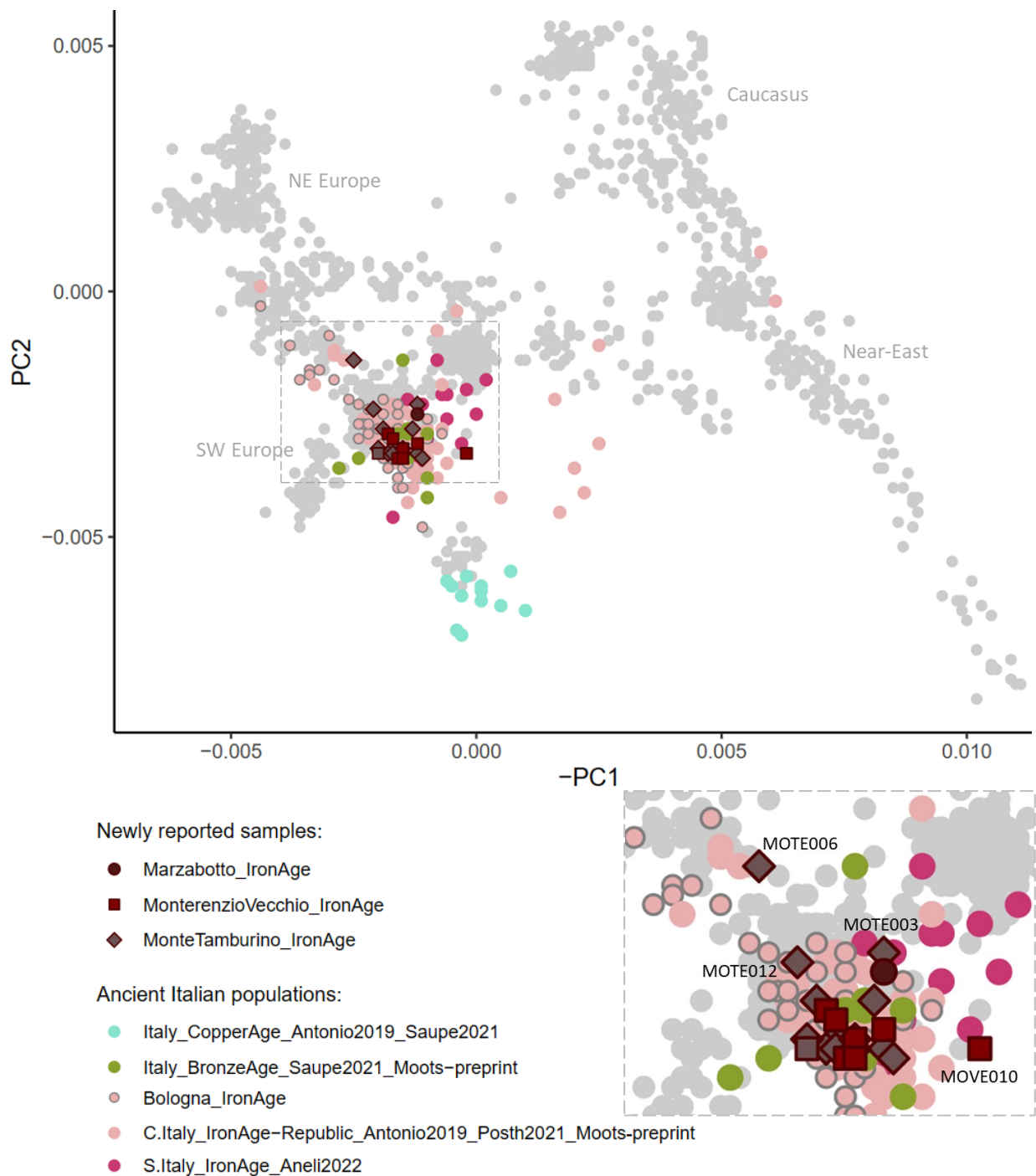
e riconducibili ad una fase iniziale del Neolitico (Morocco\_EN, giallo), che risulta massimizzata nel gruppo etno-linguistico Yoruba (Africa occidentale). Tale segnale conferma le evidenze osservate in PCA e supporta l'identificazione per questo individuo di un aplogruppo mitocondriale africano.

Oltre a ciò, si può riscontrare la presenza in CAS054 di un'ulteriore *ancestry* associata a Morocco\_EN (in marrone) e presente in percentuale minore anche negli altri tre individui del periodo imperiale; si tratta di una componente che, dai risultati di ADMIXTURE, risulta essere massimizzata in gruppi moderni che abitano le regioni desertiche mediorientali (denominati Yemeni\_Desert e BeduinB nel dataset HO).

Per quanto riguarda, invece, il secondo set di campioni, in *Figura 6.9* viene mostrata la relativa PCA dell'Eurasia occidentale sopra la quale sono stati proiettati per confronto anche i campioni del set precedente (punti rosa con contorno grigio). Diversamente da quanto atteso sulla base del periodo di appartenenza dei campioni provenienti da Monterenzio Vecchio e Monte Tamburino (V-II secolo a.C.), la maggior parte di quest'ultimi cade all'interno della variabilità genetica locale già osservata per la penisola italiana durante l'età del Ferro. I campioni di Monterenzio Vecchio, in particolare, costituiscono un gruppo apparentemente molto omogeneo, con la sola eccezione di MOVE010 che risulta leggermente distaccato dal cluster principale e si sovrappone parzialmente ad un campione pugliese dell'età del Ferro. Seppur distribuiti in uno spazio più ampio, anche i campioni di Monte Tamburino rientrano all'interno del grande cluster dell'età del Ferro, con l'esclusione di MOTE006 che sembra avere un profilo genetico più affine al cluster *C.Italy\_Etruscan.Ceu* descritto in Posth *et al.* [117].

Trattandosi di campioni riferibili al periodo La Tène, che corrisponde al momento delle migrazioni storiche di gruppi di cultura celtica a sud del Po e della loro occupazione dei centri etruschi dell'area padana, la quasi totale assenza di individui geneticamente affini a popolazioni del centro Europa è un dato sicuramente inatteso, soprattutto considerando gli influssi celtici rilevati nelle necropoli in questione. Tuttavia, l'analisi degli isotopi pubblicata in [155, 157], che ha interessato anche alcuni degli individui inclusi nel presente studio, ha rilevato per quest'ultimi un'origine prevalentemente locale. Per quanto riguarda Monte Tamburino, soltanto i campioni MOTE002, MOTE003 e MOTE013 risultano non essere originari del luogo sulla base dell'analisi degli isotopi; in PCA tali individui non si distinguono però in modo particolare dagli altri dello stesso sito. Tra questi, MOTE012, per il quale non sono disponibili dati isotopici, è l'unico che, pur sovrapponendosi alla variabilità genetica del cluster principale dell'età del Ferro (punti rosa in PCA), sembra essere leggermente più spostato in direzione del campione MOTE006. Per Monterenzio, invece, i campioni MOVE004 e MOVE008/009, identificati come non autoctoni dall'analisi degli isotopi, sembrano sovrapporsi completamente in PCA ai restanti individui provenienti dalla medesima necropoli.

Infine, MARZ001, ritrovato presso il tempio di Uni a Marzabotto e datato al radiocarbonio tra l'VIII e l'inizio del IV secolo a.C., cade anche esso all'interno del cluster dell'età del Ferro.



**Figura 6.9** PCA costruita considerando la variabilità genetica di 62 popolazioni moderne dell'Eurasia occidentale (punti grigi) sulla quale sono stati proiettati gli individui del secondo set con l'aggiunta di altri genomi antichi di riferimento (vedi legenda). In basso a destra è riportato un ingrandimento dell'area indicata in PCA.

Ulteriori analisi genomiche saranno necessarie per meglio indagare le prime evidenze riscontrate in PCA, unitamente a datazioni al radiocarbonio che permetterebbero di confermare la collocazione temporale dei campioni in esame.

## 6.6 Definizione dei clusters genetici

Le analisi riportate dal presente paragrafo in poi riguardano esclusivamente il primo set di campioni. Al fine di verificare la presenza di alcuni *outliers* rispetto ai clusters principali dell'età del Ferro e del periodo imperiale osservati in PCA, sono state condotte delle statistiche  $F_4$  nella forma  $f_4(\text{Onge.DG}, X; \text{Individuo}, \text{Cluster})$  dove  $X$  è rappresentato dalle componenti distali già precedentemente citate e riportate in *Tabella 6.9*, *Individuo* indica il potenziale *outlier* testato mentre *Cluster* include gli individui riconducibili allo stesso periodo temporale che in PCA risultano strettamente relazionati dal punto di vista genetico (ovvero i clusters che definiremo Bologna\_Etruscan e Bologna\_Imperial). Sono stati inizialmente inclusi nel cluster Bologna\_Etruscan tutti i campioni dell'età del Ferro ad eccezione del gruppo di individui spostati in PCA verso il nord del cline Europeo (AGOS001, CAS041-1, MAR012, CAST002, MAR021, TAB010 e CAS022), TAB016, AZZA006 e i due campioni adiacenti ad esso AZZA005 e CAST001. L'individuo AZZA006 è stato testato per la sua posizione intermedia tra il cluster principale etrusco ed il gruppo spostato verso componenti centro-europee; per identificare il punto iniziale del cline visibile tra il cluster principale e AGOS001 (che segue la direzione del cline formato dalle popolazioni europee moderne), sono stati sottoposti al test anche i due campioni AZZA005 e CAST001.

Per quanto riguarda i campioni imperiali, invece, il cluster Bologna\_Imperial comprende gli individui CAS067, CAS090, CAS095, mentre è stato testato come possibile *outlier* CAS054.

Ciascun individuo testato è stato classificato come *outlier* rispetto al cluster principale di riferimento in presenza di almeno una statistica significativa, cioè accompagnata da  $Z > |3|$  (marcata in rosso in *Tabella 6.9*).

Sulla base dell'assenza di statistiche significative per i campioni AZZA005 e CAST001 (vedi *Tabella 6.9*), quest'ultimi sono stati inclusi nel cluster Bologna\_Etruscan prima di procedere con i successivi test. Tutti i restanti campioni dell'età del Ferro hanno restituito almeno un paragone significativo e per questo motivo sono stati considerati degli *outliers* rispetto a Bologna\_Etruscan.

Analizzando il gruppo di campioni spostato verso nord, ciò che si rileva in tutti i casi è una maggiore affinità tra Turkey\_N (che corrisponde ad Anatolia\_N) e Bologna\_Etruscan rispetto all'individuo testato, evidenziata dalla positività della statistica  $F_4$ ; tale risultato è accompagnato nella maggior

parte dei confronti da una maggiore affinità dei gruppi EHG e/o WHG e gli individui testati rispetto al cluster principale, segnalata dalla negatività della statistica. È importante sottolineare che queste due componenti (EHG e WHG) caratterizzano, insieme a CHG, il profilo genetico della popolazione Yamnaya; in particolare, EHG rappresenta più della metà dell'*ancestry* associata ad essa [88, 101]. Per quanto riguarda AZZA006, il confronto con la componente Russia\_Samara\_EBA\_Yamnaya (che corrisponde alla *steppe-related ancestry*) ha restituito una statistica negativa con Z score appena sotto la soglia finora considerata (-2,91). Data la vicinanza alla significatività e considerato il numero di SNPs coinvolti nell'analisi, minore rispetto a quanto si osserva in media per gli altri test a causa di una più bassa copertura di AZZA006 sul pannello 1240K, il risultato ottenuto è stato considerato significativo e indicativo di una distinzione genetica tra questo campione e Bologna\_Etruscan.

Anche TAB016 può essere considerato un *outlier* rispetto al cluster etrusco dal momento che risulta essere più strettamente correlato alla componente Turkey\_N di Bologna\_Etruscan, come previsto sulla base della posizione occupata in PCA, spostata rispetto al cluster etrusco in direzione dei sardi moderni e dei gruppi calcolitici, entrambi caratterizzati da un'elevata componente di derivazione neolitica [142, 202].

Infine, il campione imperiale CAS054 mostra maggiore affinità genetica con la componente africana Morocco\_Iberomaurusian (associata ai cacciatori-raccoglitori del Marocco) di quanto non si osservi per Bologna\_Imperial che, rispetto a CAS054, risulta invece più strettamente relazionato alla *steppe ancestry*. È possibile notare, inoltre, che nel confronto con Morocco\_EN, la statistica risulta negativa e associata ad uno Z score superiore a  $|2,50|$ , ulteriore segnale della presenza di una componente di origine africana nel profilo genetico di CAS054.

Per confermare i risultati ottenuti e per verificare se alcuni degli *outliers* dell'età del Ferro potessero formare a loro volta un cluster genetico, coppie di individui sono state testate mediante il software qpWave e di p-values restituiti dall'analisi sono stati convertiti in scala colore nella heatmap visibile in *Figura 6.10*. Confronti con p-value  $> 0,01$  sono stati considerati indicativi del fatto che gli individui testati costituiscono un clade rispetto al pannello di popolazioni di riferimento e che possono essere quindi raggruppati in un unico cluster genetico. In *Figura 6.10*, i confronti con p-value  $< 0,01$  sono stati rappresentati in grigio, mentre quelli con p-value superiore alla soglia definita sono raffigurati, come mostrato in legenda, secondo un gradiente di colore che va dal rosso al giallo.

Considerando i campioni dell'età del Ferro, dal grafico si evidenziano due raggruppamenti distinti (marcati da quadrati bianchi) che racchiudono al loro interno rispettivamente i campioni del cluster etrusco principale (azzurro), che sembrano essere in continuità genetica con C.Italy\_Etruscan [117], e alcuni degli *outliers* caratterizzati, secondo le statistiche  $F_4$ , da una maggiore affinità genetica ai gruppi EHG/WHG. Sulla base di tali risultati, questi campioni, che nella prevalenza dei casi mostrano

confronti con  $p\text{-value} > 0,01$  con il gruppo C.Italy\_Etruscan.Ceu [117], sono stati raggruppati nel cluster Bologna\_Etruscan.Ceu (verde). Quest'ultimo include, quindi, i campioni CAS41-1, CAST002, MAR021, CAS022, TAB10, AGOS003, datati tra l'VIII e il V secolo a.C. (coevi al cluster principale), e MAR012 datato tra il V e il III secolo a.C.

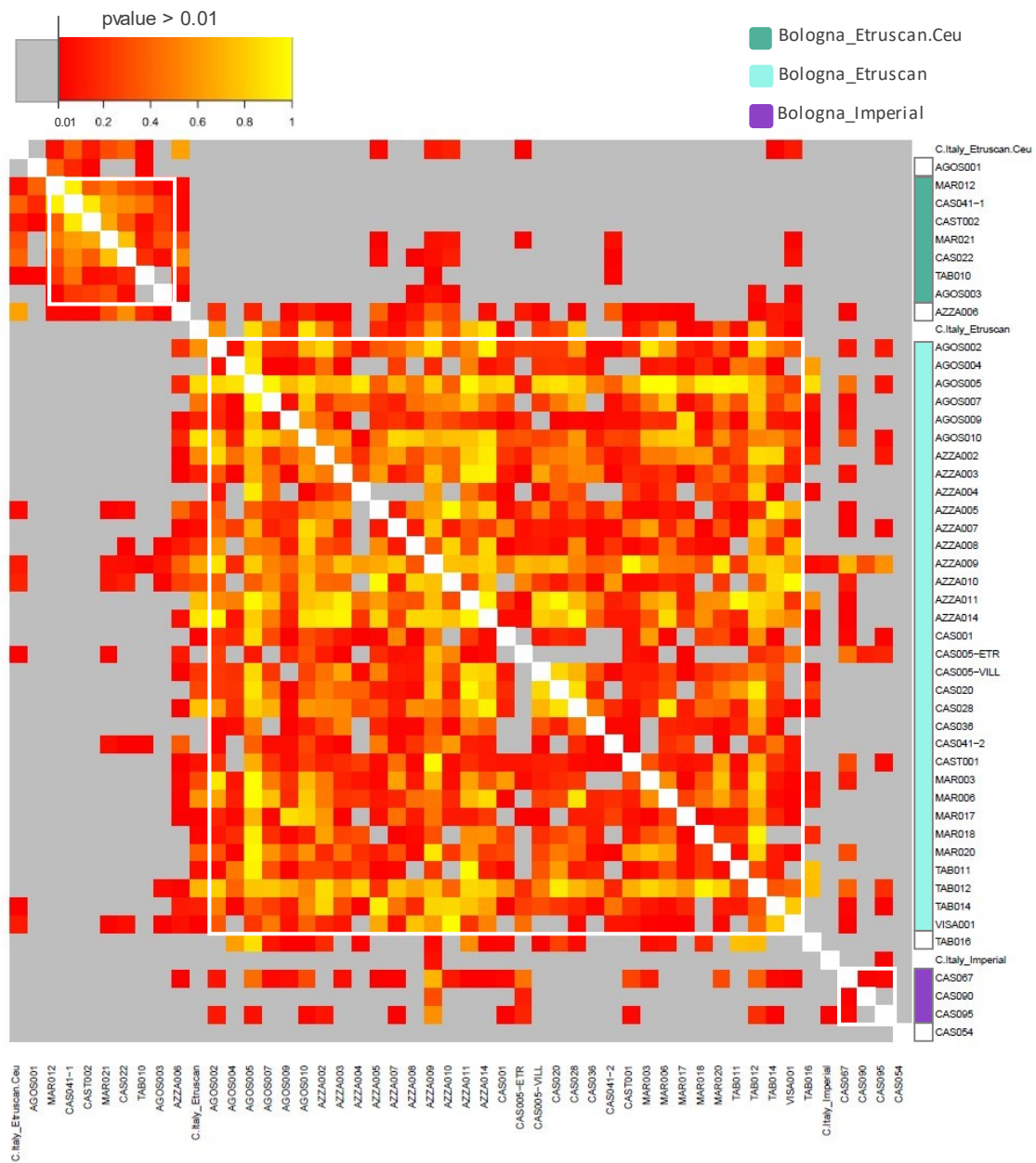
Il campione AGOS001 non è stato incluso in questo gruppo in quanto i confronti a coppie hanno restituito 3 modelli non validi ( $p\text{-value} < 0,01$ ) su 7.

Sebbene tutti i test tra l'individuo AZZA006 e i campioni del cluster Bologna\_Etruscan.Ceu mostrino  $p\text{-value}$  superiore a 0,01, AZZA006 è stato considerato un *outlier* sulla base del fatto che presenta un elevato numero di modelli funzionanti anche quando confrontato con gli individui del cluster principale. Questo segnale potrebbe essere dovuto ad una copertura minore della media (~120.000 SNPs sul pannello 1240K rispetto ad una media osservata di ~500.000 SNPs) o potrebbe rispecchiare effettivamente la posizione intermedia di AZZA006 tra i due cluster etruschi. Allo scopo di verificare questi risultati, per AGOS001 e AZZA006 sono state ripetute le statistiche descritte sopra, inserendo nella posizione *Cluster* Bologna\_Etruscan.Ceu. I risultati, riportati nella parte finale di *Tabella 6.9*, non raggiungono la significatività nel caso di AGOS001 ma si osserva comunque uno Z score di 2,53 associato ad una statistica positiva nel caso in cui X è rappresentato da Turkey\_N, ad indicare una maggiore affinità genetica, seppur lieve, tra tale componente e Bologna\_Etruscan.Ceu rispetto ad AGOS001. Nel caso di AZZA006 si osserva, invece, la situazione opposta, evidenziata per il confronto con Turkey\_N da una statistica negativa accompagnata da Z score significativo.

Andando a testare mediante qpWave la relazione tra AGOS001 ed il cluster Bologna\_Etruscan.Ceu preso nel suo insieme si ottiene un  $p\text{-value}$  di 0.0006, a conferma del fatto che AGOS001 costituisce un *outlier* e per questo motivo verrà analizzato successivamente come singolo individuo. Ripetendo il test per AZZA006 si ottiene un modello funzionante con  $p\text{-value} = 0.063$ . Quando confrontato con Bologna\_Etruscan, invece, AZZA006 restituisce un  $p\text{-value}$  non significativo pari a 0.002; tale valore cresce però fino alla soglia di significatività (0.018) nel momento in cui il test viene ristretto ai soli campioni del cluster Bologna\_Etruscan effettivamente datati al  $^{14}\text{C}$  alla prima età del Ferro (N=3). Considerati i segnali contrastanti e le evidenze ottenute dalle statistiche  $F_4$ , si è deciso di analizzare AZZA006 separatamente rispetto ai due *clusters* dell'età del Ferro. Anche l'individuo TAB010 è stato considerato un *outlier* rispetto a Bologna\_Etruscan poiché, oltre alle evidenze riscontrate dalla PCA e dalle statistiche  $F_4$ , presenta un elevato numero di modelli non funzionanti (14 su 33) quando comparato ai campioni di questo cluster.

Infine, all'interno del riquadro Bologna\_Imperial si osserva la presenza di un  $p\text{-value} < 0,01$  per il modello che include CAS090 e CAS095; dal momento che entrambi risultano geneticamente simili a CAS067 e che i test con qpWave mostrano  $p\text{-value} > 0,01$  quando vengono confrontati CAS090 e

CAS095 individualmente con Bologna\_Imperial (comprensivo di CAS067 e alternativamente del campione non testato), è stata confermata l'inclusione di tutti e tre gli individui all'interno dello stesso cluster, rispetto al quale CAS054 costituisce sicuramente un *outlier*. Infine, la presenza di un solo modello valido nel confronto tra il cluster C.Italy\_Imperial [117] ed i singoli individui del cluster Bologna\_Imperial, suggerisce l'assenza di un'elevata affinità genetica tra i due gruppi imperiali.



**Figura 6.10** Heatmap generata a partire dai P-values ottenuto nei diversi test a coppie effettuati con qpWave. In legenda è riportata la scala colore associata ai P-values (valori minori di 0,01 sono stati riportati in grigio), mentre i riquadri bianchi evidenziano i clusters determinati che sono stati indicati con colori diversi anche a lato del grafico.

Outgroup	Pop1	Individual	Pop2	f4-stats	StdErr	Zscore	ABBA	BABA	SNPs
Onge.DG	EHG	AZZA005	Bologna_Etruscan	-0.000386	0.000423	-0.91	25326	25520	504392
Onge.DG	Morocco_EN.SG	AZZA005	Bologna_Etruscan	0.000112	0.000536	0.21	14758	14726	292259
Onge.DG	Russia_Samara_EBA_Yamnaya	AZZA005	Bologna_Etruscan	0.000151	0.000348	0.43	26048	25971	514037
Onge.DG	WHG	AZZA005	Bologna_Etruscan	0.000225	0.000426	0.53	26170	26054	514428
Onge.DG	Iran_GanjDareh_N	AZZA005	Bologna_Etruscan	0.000282	0.000373	0.76	25653	25509	508626
Onge.DG	Morocco_Iberomaurusian	AZZA005	Bologna_Etruscan	0.000571	0.000407	1.40	26253	25960	512026
Onge.DG	Turkey_N	AZZA005	Bologna_Etruscan	0.000611	0.000336	1.82	26774	26460	514316
Onge.DG	Iran_GanjDareh_N	CAST001	Bologna_Etruscan	-0.001674	0.000745	-2.25	2221	2296	44779
Onge.DG	Russia_Samara_EBA_Yamnaya	CAST001	Bologna_Etruscan	-0.000413	0.000692	-0.60	2266	2284	45070
Onge.DG	Morocco_EN.SG	CAST001	Bologna_Etruscan	-0.000608	0.001256	-0.48	1270	1286	25375
Onge.DG	WHG	CAST001	Bologna_Etruscan	-0.000358	0.000836	-0.43	2289	2305	45101
Onge.DG	Morocco_Iberomaurusian	CAST001	Bologna_Etruscan	-0.000311	0.000821	-0.38	2290	2304	44936
Onge.DG	Turkey_N	CAST001	Bologna_Etruscan	-0.000256	0.000690	-0.37	2325	2337	45088
Onge.DG	EHG	CAST001	Bologna_Etruscan	0.000984	0.000862	1.14	2244	2200	44458
<b>Onge.DG</b>	<b>Russia_Samara_EBA_Yamnaya</b>	<b>AZZA006</b>	<b>Bologna_Etruscan</b>	<b>-0.001400</b>	<b>0.000481</b>	<b>-2.91</b>	<b>5852</b>	<b>6016</b>	<b>117118</b>
Onge.DG	WHG	AZZA006	Bologna_Etruscan	-0.000622	0.000574	-1.08	5927	6000	117164
Onge.DG	EHG	AZZA006	Bologna_Etruscan	-0.001459	0.000585	-2.49	5709	5877	115560
Onge.DG	Iran_GanjDareh_N	AZZA006	Bologna_Etruscan	-0.000368	0.000541	-0.68	5831	5873	116394
Onge.DG	Morocco_Iberomaurusian	AZZA006	Bologna_Etruscan	-0.000298	0.000579	-0.52	5927	5962	116829
Onge.DG	Turkey_N	AZZA006	Bologna_Etruscan	-0.000158	0.000488	-0.32	6052	6070	117128
Onge.DG	Morocco_EN.SG	AZZA006	Bologna_Etruscan	-0.000224	0.000830	-0.27	3329	3343	65996
<b>Onge.DG</b>	<b>EHG</b>	<b>AGOS001</b>	<b>Bologna_Etruscan</b>	<b>-0.001780</b>	<b>0.000395</b>	<b>-4.51</b>	<b>34113</b>	<b>35327</b>	<b>682218</b>
Onge.DG	WHG	AGOS001	Bologna_Etruscan	-0.000722	0.000406	-1.78	35737	36243	700050
Onge.DG	Russia_Samara_EBA_Yamnaya	AGOS001	Bologna_Etruscan	-0.000709	0.000338	-2.10	35508	36004	699188
Onge.DG	Iran_GanjDareh_N	AGOS001	Bologna_Etruscan	0.000631	0.000356	1.77	35162	34728	689190
Onge.DG	Morocco_EN.SG	AGOS001	Bologna_Etruscan	0.001021	0.000459	2.22	20608	20200	400338
Onge.DG	Morocco_Iberomaurusian	AGOS001	Bologna_Etruscan	0.001069	0.000376	2.85	36092	35348	695453
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>AGOS001</b>	<b>Bologna_Etruscan</b>	<b>0.002407</b>	<b>0.000339</b>	<b>7.10</b>	<b>37303</b>	<b>35620</b>	<b>699404</b>
<b>Onge.DG</b>	<b>EHG</b>	<b>CAS041-1</b>	<b>Bologna_Etruscan</b>	<b>-0.001283</b>	<b>0.000401</b>	<b>-3.20</b>	<b>36153</b>	<b>37079</b>	<b>721162</b>
Onge.DG	WHG	CAS041-1	Bologna_Etruscan	-0.000848	0.000408	-2.08	37765	38395	742350
Onge.DG	Russia_Samara_EBA_Yamnaya	CAS041-1	Bologna_Etruscan	-0.000503	0.000335	-1.50	37603	37976	741107
Onge.DG	Iran_GanjDareh_N	CAS041-1	Bologna_Etruscan	0.000400	0.000345	1.16	37018	36727	728788
Onge.DG	Morocco_Iberomaurusian	CAS041-1	Bologna_Etruscan	0.000669	0.000391	1.71	37999	37507	736525
Onge.DG	Morocco_EN.SG	CAS041-1	Bologna_Etruscan	0.000945	0.000470	2.01	21824	21424	423449
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>CAS041-1</b>	<b>Bologna_Etruscan</b>	<b>0.001646</b>	<b>0.000333</b>	<b>4.94</b>	<b>39213</b>	<b>37993</b>	<b>741389</b>
<b>Onge.DG</b>	<b>EHG</b>	<b>MAR012</b>	<b>Bologna_Etruscan</b>	<b>-0.001707</b>	<b>0.000394</b>	<b>-4.33</b>	<b>36515</b>	<b>37764</b>	<b>731650</b>
<b>Onge.DG</b>	<b>WHG</b>	<b>MAR012</b>	<b>Bologna_Etruscan</b>	<b>-0.001307</b>	<b>0.000395</b>	<b>-3.31</b>	<b>38056</b>	<b>39040</b>	<b>752620</b>
Onge.DG	Russia_Samara_EBA_Yamnaya	MAR012	Bologna_Etruscan	-0.000995	0.000333	-2.99	37977	38725	751527
Onge.DG	Iran_GanjDareh_N	MAR012	Bologna_Etruscan	0.000052	0.000359	0.15	37392	37353	739519
Onge.DG	Morocco_EN.SG	MAR012	Bologna_Etruscan	0.000532	0.000473	1.13	22008	21780	429215
Onge.DG	Morocco_Iberomaurusian	MAR012	Bologna_Etruscan	0.001037	0.000399	2.60	38662	37888	746869
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>MAR012</b>	<b>Bologna_Etruscan</b>	<b>0.001400</b>	<b>0.000345</b>	<b>4.05</b>	<b>39670</b>	<b>38618</b>	<b>751701</b>
Onge.DG	EHG	CAST002	Bologna_Etruscan	-0.000860	0.000410	-2.10	36284	36903	719400
Onge.DG	WHG	CAST002	Bologna_Etruscan	-0.000704	0.000425	-1.66	37851	38372	740552
Onge.DG	Russia_Samara_EBA_Yamnaya	CAST002	Bologna_Etruscan	0.000161	0.000344	0.47	37853	37735	739329
Onge.DG	Iran_GanjDareh_N	CAST002	Bologna_Etruscan	0.000851	0.000366	2.33	37170	36551	726953
Onge.DG	Morocco_Iberomaurusian	CAST002	Bologna_Etruscan	0.001018	0.000395	2.58	38207	37458	734616
Onge.DG	Morocco_EN.SG	CAST002	Bologna_Etruscan	0.001364	0.000486	2.81	21914	21338	422246
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>CAST002</b>	<b>Bologna_Etruscan</b>	<b>0.002011</b>	<b>0.000350</b>	<b>5.74</b>	<b>39356</b>	<b>37868</b>	<b>739541</b>
Onge.DG	EHG	MAR021	Bologna_Etruscan	-0.000475	0.000398	-1.19	25370	25608	501987
Onge.DG	WHG	MAR021	Bologna_Etruscan	-0.000296	0.000431	-0.69	26207	26359	511954
Onge.DG	Russia_Samara_EBA_Yamnaya	MAR021	Bologna_Etruscan	-0.000202	0.000346	-0.58	26048	26151	511596
Onge.DG	Morocco_Iberomaurusian	MAR021	Bologna_Etruscan	0.000534	0.000404	1.32	26316	26044	509646
Onge.DG	Iran_GanjDareh_N	MAR021	Bologna_Etruscan	0.000554	0.000377	1.47	25727	25446	506660
Onge.DG	Morocco_EN.SG	MAR021	Bologna_Etruscan	0.000733	0.000535	1.37	14946	14733	290742
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>MAR021</b>	<b>Bologna_Etruscan</b>	<b>0.001671</b>	<b>0.000358</b>	<b>4.67</b>	<b>27081</b>	<b>26226</b>	<b>511670</b>



Outgroup	Pop1	Individual	Pop2	f4-stats	StdErr	Zscore	ABBA	BABA	SNPs
<b>Onge.DG</b>	<b>EHG</b>	<b>TAB010</b>	<b>Bologna_Etruscan</b>	<b>-0.001323</b>	<b>0.000423</b>	<b>-3.12</b>	<b>32487</b>	<b>33348</b>	<b>650266</b>
Onge.DG	WHG	TAB010	Bologna_Etruscan	-0.001113	0.000434	-2.56	33755	34497	666293
Onge.DG	Russia_Samara_EBA_Yamnaya	TAB010	Bologna_Etruscan	-0.000703	0.000335	-2.10	33629	34097	665621
Onge.DG	Morocco_Iberomaurusian	TAB010	Bologna_Etruscan	0.000938	0.000405	2.32	34195	33574	662371
Onge.DG	Iran_GanjDareh_N	TAB010	Bologna_Etruscan	0.001035	0.000367	2.82	33488	32808	657276
Onge.DG	Morocco_EN.SG	TAB010	Bologna_Etruscan	0.001193	0.000500	2.39	19529	19077	379203
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>TAB010</b>	<b>Bologna_Etruscan</b>	<b>0.001254</b>	<b>0.000371</b>	<b>3.38</b>	<b>35041</b>	<b>34206</b>	<b>665735</b>
<b>Onge.DG</b>	<b>EHG</b>	<b>CAS022</b>	<b>Bologna_Etruscan</b>	<b>-0.001164</b>	<b>0.000384</b>	<b>-3.04</b>	<b>37013</b>	<b>37874</b>	<b>738987</b>
Onge.DG	WHG	CAS022	Bologna_Etruscan	-0.000860	0.000402	-2.14	38673	39328	761479
Onge.DG	Russia_Samara_EBA_Yamnaya	CAS022	Bologna_Etruscan	-0.000518	0.000325	-1.59	38487	38880	760150
Onge.DG	Morocco_Iberomaurusian	CAS022	Bologna_Etruscan	0.000410	0.000392	1.05	38840	38530	755225
Onge.DG	Morocco_EN.SG	CAS022	Bologna_Etruscan	0.000481	0.000472	1.02	22228	22019	434563
Onge.DG	Iran_GanjDareh_N	CAS022	Bologna_Etruscan	0.000550	0.000355	1.55	37907	37496	747082
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>CAS022</b>	<b>Bologna_Etruscan</b>	<b>0.001143</b>	<b>0.000355</b>	<b>3.22</b>	<b>39992</b>	<b>39123</b>	<b>760371</b>
<b>Onge.DG</b>	<b>WHG</b>	<b>AGOS003</b>	<b>Bologna_Etruscan</b>	<b>-0.001324</b>	<b>0.000400</b>	<b>-3.31</b>	<b>36837</b>	<b>37800</b>	<b>726906</b>
Onge.DG	EHG	AGOS003	Bologna_Etruscan	-0.000652	0.000396	-1.65	35648	36109	707288
Onge.DG	Russia_Samara_EBA_Yamnaya	AGOS003	Bologna_Etruscan	-0.000584	0.000331	-1.76	36829	37254	725966
Onge.DG	Iran_GanjDareh_N	AGOS003	Bologna_Etruscan	-0.000368	0.000364	-1.01	36032	36295	714764
Onge.DG	Morocco_Iberomaurusian	AGOS003	Bologna_Etruscan	0.000264	0.000387	0.68	37144	36953	721990
Onge.DG	Morocco_EN.SG	AGOS003	Bologna_Etruscan	0.000631	0.000464	1.36	21290	21028	415236
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>AGOS003</b>	<b>Bologna_Etruscan</b>	<b>0.001088</b>	<b>0.000350</b>	<b>3.11</b>	<b>38272</b>	<b>37482</b>	<b>726195</b>
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>TAB016</b>	<b>Bologna_Etruscan</b>	<b>-0.001140</b>	<b>0.000350</b>	<b>-3.26</b>	<b>35641</b>	<b>36433</b>	<b>694839</b>
Onge.DG	WHG	TAB016	Bologna_Etruscan	-0.000707	0.000412	-1.72	35111	35603	695379
Onge.DG	Morocco_EN.SG	TAB016	Bologna_Etruscan	-0.000472	0.000489	-0.97	20016	20203	396434
Onge.DG	Morocco_Iberomaurusian	TAB016	Bologna_Etruscan	-0.000466	0.000406	-1.15	35081	35403	691338
Onge.DG	Iran_GanjDareh_N	TAB016	Bologna_Etruscan	-0.000140	0.000347	-0.40	34368	34464	685903
Onge.DG	Russia_Samara_EBA_Yamnaya	TAB016	Bologna_Etruscan	0.000049	0.000318	0.15	35096	35062	694701
Onge.DG	EHG	TAB016	Bologna_Etruscan	0.000093	0.000384	0.24	34156	34093	679080
Onge.DG	Morocco_EN.SG	CAS054	Bologna_Imperial	-0.001468	0.000576	-2.55	17734	18244	347833
<b>Onge.DG</b>	<b>Morocco_Iberomaurusian</b>	<b>CAS054</b>	<b>Bologna_Imperial</b>	<b>-0.001457</b>	<b>0.000455</b>	<b>-3.20</b>	<b>31445</b>	<b>32336</b>	<b>611926</b>
Onge.DG	WHG	CAS054	Bologna_Imperial	0.000559	0.000460	1.22	31830	31486	613852
Onge.DG	EHG	CAS054	Bologna_Imperial	0.000767	0.000437	1.76	31039	30576	603820
Onge.DG	Turkey_N	CAS054	Bologna_Imperial	0.000777	0.000394	1.97	32528	32051	61383
Onge.DG	Iran_GanjDareh_N	CAS054	Bologna_Imperial	0.001157	0.000404	2.87	31516	30811	609428
<b>Onge.DG</b>	<b>Russia_Samara_EBA_Yamnaya</b>	<b>CAS054</b>	<b>Bologna_Imperial</b>	<b>0.001178</b>	<b>0.000370</b>	<b>3.19</b>	<b>31839</b>	<b>31116</b>	<b>613697</b>
Onge.DG	EHG	AGOS001	Bologna_Etruscan.Ceu	-0.000672	0.000404	-1.66	34203	34658	676253
Onge.DG	Russia_Samara_EBA_Yamnaya	AGOS001	Bologna_Etruscan.Ceu	-0.000233	0.000351	-0.66	35322	35483	691920
Onge.DG	Morocco_EN.SG	AGOS001	Bologna_Etruscan.Ceu	0.000055	0.000495	0.11	20114	20092	396028
Onge.DG	WHG	AGOS001	Bologna_Etruscan.Ceu	0.000163	0.000424	0.38	35675	35562	692554
Onge.DG	Iran_GanjDareh_N	AGOS001	Bologna_Etruscan.Ceu	0.000264	0.000367	0.72	34672	34492	683444
Onge.DG	Morocco_Iberomaurusian	AGOS001	Bologna_Etruscan.Ceu	0.000347	0.000393	0.88	35371	35132	688714
Onge.DG	Turkey_N	AGOS001	Bologna_Etruscan.Ceu	0.000868	0.000343	2.53	36280	35679	692377
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>AZZA006</b>	<b>Bologna_Etruscan.Ceu</b>	<b>-0.001814</b>	<b>0.000514</b>	<b>-3.53</b>	<b>5920</b>	<b>6132</b>	<b>116631</b>
Onge.DG	Russia_Samara_EBA_Yamnaya	AZZA006	Bologna_Etruscan.Ceu	-0.001194	0.000500	-2.39	5841	5980	116611
Onge.DG	Morocco_Iberomaurusian	AZZA006	Bologna_Etruscan.Ceu	-0.001141	0.000614	-1.86	5866	5998	116354
Onge.DG	Morocco_EN.SG	AZZA006	Bologna_Etruscan.Ceu	-0.000966	0.000882	-1.10	3305	3369	65712
Onge.DG	Iran_GanjDareh_N	AZZA006	Bologna_Etruscan.Ceu	-0.000893	0.000552	-1.62	5782	5885	115968
Onge.DG	EHG	AZZA006	Bologna_Etruscan.Ceu	-0.000683	0.000608	-1.12	5756	5835	115143
Onge.DG	WHG	AZZA006	Bologna_Etruscan.Ceu	0.000303	0.000608	0.50	5969	5933	116646

**Tabella 6.9** Risultati delle statistiche distali  $F_4$  effettuate per identificare la presenza di *outliers* rispetto ai clusters genetici principali dell'età del Ferro e del periodo imperiale. Viene riportato il valore della statistica  $F_4$ , il relativo errore standard, lo Z score, il numero di SNPs associati agli schemi ABBA/BABA e, infine, il numero totale di SNPs coinvolti nei confronti. Sono marcate in rosso le statistiche significative, ovvero con Z score  $> |3|$  (o molto vicino a tale soglia). In ciascun blocco i valori di  $f_4$  sono stati ordinati in modo crescente.

## 6.7 Evidenze popolazionistiche emerse dalle statistiche F

Per indagare il grado di affinità genetica tra la comunità etrusca padana (Bologna\_Etruscan) e quella dell'Etruria classica (C.Italy\_Etruscan) sono state effettuate delle statistiche  $F_4$  distali formulate come indicato in *Tabella 6.10*. Nonostante nessuna delle statistiche calcolate raggiunga pienamente la significatività ( $Z \text{ score} < |3|$ ), suggerendo in linea generale una certa continuità genetica tra i due gruppi (che risultavano sovrapposti anche in PCA), si osserva però uno Z score di 2,61 associato al confronto con la componente Iran\_N. Allo scopo di approfondire tale aspetto, la statistica è stata ripetuta considerando solamente gli individui dei due clusters datati al radiocarbonio (N=21 e N=3 rispettivamente per C.Italy\_Etruscan e Bologna\_Etruscan), ottenendo in questo caso tutti risultati lontani dalla soglia di significatività. Sebbene in Etruria classica si denoti il mantenimento di un *pool* genico altamente omogeneo nel corso dell'intero intervallo temporale considerato nello studio (800-1 a.C.) [117], è importante sottolineare che le datazioni disponibili per il cluster Bologna\_Etruscan sono al momento ridotte e tutte riconducibili tra l'VIII ed il V secolo a.C. L'ottenimento di ulteriori datazioni potrebbe rivelare nel contesto padano eventuali *patterns* genetici correlati alla variabile temporale che in mancanza di questo dato non possono essere apprezzati.

La presenza di possibili evidenze di *admixture* tra Bologna\_Etruscan e 562 gruppi antichi dell'Eurasia occidentale e dell'Africa settentrionale, datati tra il Mesolitico e l'età del Ferro, è stata indagata per il cluster Bologna\_Etruscan.Ceu mediante statistiche  $F_3$  con la seguente costruzione:  $f_3(\text{Onge.DG}, \text{Pop}, \text{Bologna\_Etruscan.Ceu}, \text{Bologna\_Etruscan})$ .

In *Figura 6.11* sono riportati i primi 20 risultati ottenuti, ordinati in modo crescente; i confronti significativi, con  $Z \text{ score} < -3$ , sono indicati con punti rossi, mentre i punti con solamente il contorno rosso rappresentano confronti associati a  $Z \text{ score}$  compreso tra -2,5 e -3.

Sulla base dei risultati riportati in *Figura 6.11*, il cluster Bologna\_Etruscan.Ceu può essere definito come derivante da un fenomeno di mescolanza genetica (suggerito dalle negatività delle statistiche) tra Bologna\_Etruscan e gruppi antichi caratterizzati da un'elevata proporzione di EHG *ancestry*; si evidenziano statistiche negative significative anche per i gruppi WHG e SHG (ovvero Scandinavian Hunter-Gatherers, derivanti da *admixture* tra EHG e WHG [203, 204]). Andando a sostituire il cluster Bologna\_Etruscan con Italy\_Broion\_BA (che racchiude individui provenienti dal nord Italia e datati all'età del Bronzo) [142], testando in questo modo uno scenario di *admixture* meno recente, si osservano gli stessi segnali rilevati precedentemente con l'aggiunta di un maggior numero di gruppi dell'età del Bronzo contraddistinti dalla componente della steppa (provenienti da Russia e Ucraina). In accordo a questo, anche nel paragone tra Bologna\_Etruscan e Bologna\_Etruscan.Ceu effettuato con la statistica  $F_4$  distale riportata in *Tabella 6.10*, emerge chiaramente una maggiore affinità tra

EHG, WHG, Russia\_Samara\_EBA\_Yamnaya ed il gruppo Bologna\_Etruscan.Ceu rispetto al cluster etrusco locale; la situazione opposta si verifica quando vengono considerate Turkey\_N e le componenti africane Morocco\_Iberomaurusian e Morocco\_EN.SG. La stessa tipologia di analisi non evidenzia differenze significative a livello genetico tra il gruppo in esame e C.Italy\_Etruscan.Ceu [117]. Segnali coerenti a quelli discussi finora si osservano anche dalla statistica  $f_4(\text{Onge.DG, Pop; Bologna\_Etruscan.Ceu, Bologna\_Etruscan})$  visibile in *Figura 6.12*, dove vengono riportati anche in questo caso i primi 20 risultati ordinati in senso crescente; questi suggeriscono ancora una volta flusso genico tra il cluster Bologna\_Etruscan.Ceu e popolazioni associate alle componenti WHG e, in particolare, EHG. Tutte queste evidenze puntano verso un profilo genetico degli individui del cluster Bologna\_Etruscan.Ceu verosimilmente associato a popolazioni del centro e del nord Europa, caratterizzate da particolare affinità con la componente della steppa.

In aggiunta, la presenza all'interno di questo gruppo di un individuo più recente rispetto agli altri (MAR012) suggerisce una certa continuità nell'origine e arrivo di tale componente caratteristica del centro-nord Europa, attestata in Etruria padana tra l'VIII e il III secolo a.C. (dalla fase Hallstatt al periodo La Tène).

Evidenze simili a quelle appena osservate per Bologna\_Etruscan.Ceu sono state individuate anche per il campione AGOS001 nelle statistiche  $f_4(\text{Onge.DG, Pop, AGOS001, Bologna\_Etruscan})$  visibili in *Figura 6.12*. Quest'ultime sono significativamente negative per popolazioni mesolitiche e neolitiche il cui profilo genetico è in prevalenza caratterizzato dalla componente EHG o per gruppi dell'età del rame e del bronzo contraddistinti dalla componente della steppa (Russia\_Khvalynsk\_Eneolithic [94] e Russia\_North\_Caucasus [104]). Andando a sostituire nell'analisi il gruppo Bologna\_Etruscan con Bologna\_Etruscan.Ceu, le statistiche negative non raggiungono la significatività mentre tra i risultati positivi compaiono Z scores significative per Italy\_Sardinia\_MBA [116], Italy\_N.SG, Italy\_C.SG [115] e Greece\_Peloponnese\_N [93], ad indicare una maggiore affinità del cluster Bologna\_Etruscan.Ceu, rispetto ad AGOS001, alla componente neolitica tipica degli *Early European Farmers* (EEF); ciò è in linea con quando osservato in PCA e dai risultati di ADMIXTURE, dai quali AGOS001 risulta essere il campione più strettamente relazionato alla componente della steppa. Ulteriore conferma di questo deriva dall' $f_4(\text{Onge.DG, Pop; AGOS001; Italy\_North\_BellBeaker\_2})$  dove Italy\_North\_BellBeaker\_2 rappresenta un campione del nord Italia datato all'età del bronzo che non presenta *admixture* con popolazioni associate alla *steppe ancestry* [106]. Emerge ancora più chiaramente in questo caso il legame tra AGOS001 e la componente della steppa, rappresentata dai confronti significativamente negativi che si osservano non solo con EHG ma anche con gruppi riferibili al calcolitico e all'età del Bronzo provenienti dalla Russia e da regioni limitrofe, tra loro geneticamente molto simili (quali

Russia\_Steppe\_Maikop, Russia\_North\_Caucasus [104], Russia\_Afanasio, Russia\_MBA\_Poltavka [94], Russia\_BA\_Okunevo.SG [205], Ukraine\_EBA\_Yamnaya [93]).

Passando all'individuo AZZA006 i test  $f_4(\text{Onge.DG, Pop; AZZA006, Bologna\_Etruscan})$  e  $f_4(\text{Onge.DG, Pop; AZZA006, Bologna\_Etruscan.Ceu})$  confermano la posizione intermedia osservata per questo campione dalle analisi precedenti e, in particolare, dalle  $F_4$  distali trattate nel *Paragrafo 6.6*.

Nonostante l'analisi sia influenzata dalla copertura non elevata del campione sul pannello 1240K (circa 120.000 SNPs), che si traduce in errori standard mediamente elevati, nel primo test si osserva una maggior condivisione di alleli tra popolazioni del centro Europa riferibili all'età del Bronzo e AZZA006 rispetto al cluster etrusco principale (risultati significativamente negativi); sebbene non si denoti più la forte attrazione precedentemente mostrata da AGOS001 e da Bologna\_Etruscan.Ceu verso popolazioni tipicamente associate alla *steppe ancestry*, questo risultato conferma comunque una maggiore affinità di AZZA006, rispetto al cluster etrusco, a componenti genetiche dell'Europa centrale. Situazione che si inverte nel confronto con Bologna\_Etruscan.Ceu (*Figura 6.12*), dove si rilevano per AZZA006 statistiche negative e significative principalmente per gruppi caratterizzati da componenti di derivazione anatolica/levantina del periodo neolitico e calcolitico.

Rispetto al cluster etrusco principale, l'*outlier* TAB016 mostra maggiore deriva genetica condivisa con gruppi europei neolitici e calcolitici (accumunati dalla componente tipica degli EEF), nonché con campioni appartenenti all'età del Bronzo che non mostrano significativa evidenza di *steppe-related ancestry* (come nel caso degli individui North\_Italy\_BellBeaker\_2, France\_BellBeaker\_NoSteppe [106] e Greece\_BA\_Mycenaean\_Pylos [206]); si notano, inoltre, statistiche negative e significative anche per gruppi/individui dell'età del Ferro contraddistinti da una consistente percentuale di EEF *ancestry* [107]. Come già ipotizzato in precedenza, sulla base delle evidenze genetiche il campione TAB016 potrebbe appartenere ad un periodo precedente rispetto a quanto supposto, oppure aver conservato una particolare affinità ai gruppi neolitici e calcolitici italiani, superiore a quella rilevata nel complesso nel *pool* genetico etrusco (e, più in generale, dell'età del Ferro).

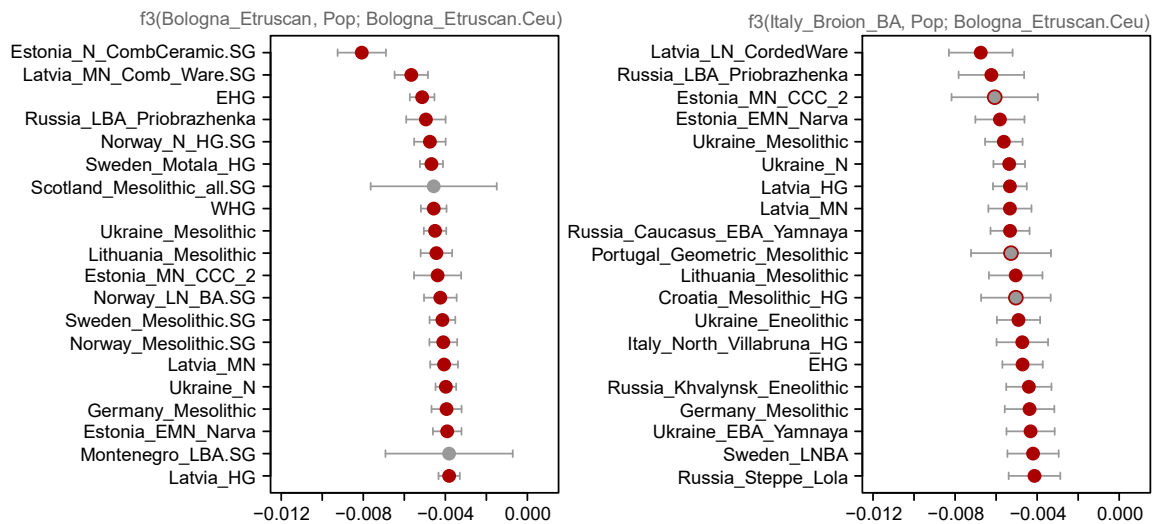
Per quanto riguarda gli individui riferibili al periodo imperiale, la statistica  $f_3(\text{Bologna\_Etruscan, Pop; Bologna\_Imperial})$  ha rilevato solo due test significativi che prevedono in entrambi i casi *admixture* con popolazioni associate alla componente Iran\_N (Iran\_TepeAbdulHosein\_N.SG con Z score pari a -3,70 e Iran\_Wezmeh\_N.SG con Z score pari a -2.68). Tale segnale è in accordo con l'incremento, già osservato in studi precedenti [115, 117], della frazione di Iran\_N nella popolazione imperiale della penisola italiana rispetto al periodo antecedente, ben rappresentato in PCA da uno *shift* in direzione del cline del Vicino Oriente. Tra i primi risultati negativi ottenuti del test  $f_4(\text{Onge.DG, Pop; Bologna\_Imperial, Bologna\_Etruscan})$ , disposti in ordine crescente in *Figura 6.12*, compaiono anche

in questo caso i confronti con i gruppi Iran\_TepeAbdulHosein\_N.SG e Iran\_Wezmeh\_N.SG, che però non raggiungono la soglia di significatività (Z score rispettivamente di -2,36 e -1,26).

Se da una parte il cluster etrusco e quello imperiale non costituiscono un clade, data la presenza di un elevato numero di test significativamente positivi che corrispondono principalmente a confronti con popolazioni antiche del continente europeo, anche Bologna\_Imperial e C.Italy\_Imperial [117] non possono essere considerati un unico cluster genetico, come mostrato dall' $F_4$  distale riportata in *Tabella 6.10* dove l'affinità con le componenti africane Morocco\_EN.SG e Morocco\_Iberomaussian e con Iran\_GanjDareh\_N (Z score = -2.70) distingue il cluster padano da quello del centro Italia (ad esse più affine). Per indagare con maggiore risoluzione questa differenziazione, già visibile anche in PCA (dove gli individui imperiali di Casalecchio di Reno cadono in una posizione intermedia tra il cluster dell'età del Ferro e l'eterogeneo gruppo imperiale dell'Italia centrale [115, 117]), è stata effettuata la statistica  $F_4$  nella forma  $f_4(\text{Onge.DG}, \text{Pop}, \text{Bologna\_Imperial}, \text{C.Italy\_Imperial})$ . Sulla base dei risultati ottenuti, C.Italy\_Imperial risulta essere più affine rispetto a Bologna\_Imperial a componenti levantine rappresentate dalle popolazioni Israel\_Ashkelon\_LBA (Z score = 3,91) [207], Lebanon\_MBA.SG (Lebanon\_MBA.SG = 3,18) [208], Jordan\_LBA\_IA (Z score = 4,03) e Israel\_MLBA (Z score = 3,26) [209] in corrispondenza delle quali si osservano risultati significativamente positivi. Il campione imperiale CAS054, invece, mostra evidenti segnali di *admixture* con popolazioni antiche africane caratterizzate da elevate proporzioni di *ancestry* nordafricana o subsahariana (vedi *Figura 6.12*), come già anticipato dai risultati discussi nei precedenti paragrafi. È necessario sottolineare in questo contesto che la maggioranza di confronti significativi con popolazioni/individui subsahariani può dipendere dalla limitata disponibilità di genomi antichi provenienti dall'Africa settentrionale. Per questa analisi, infatti, il dataset africano è stato arricchito con l'aggiunta di 34 gruppi/individui antichi provenienti dall'Africa subsahariana come indicato nel *Paragrafo 5.3.9*.

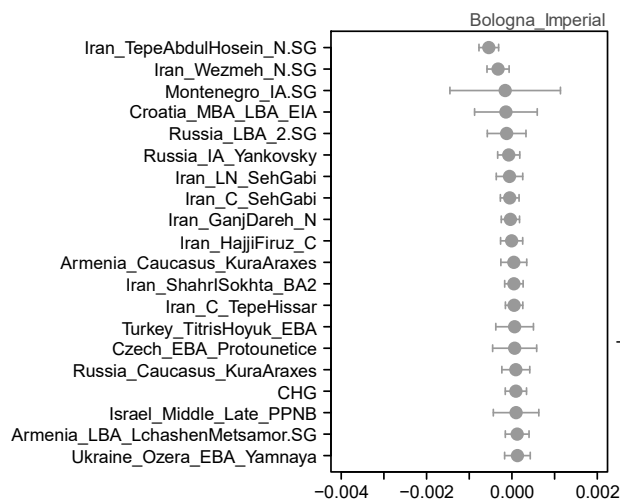
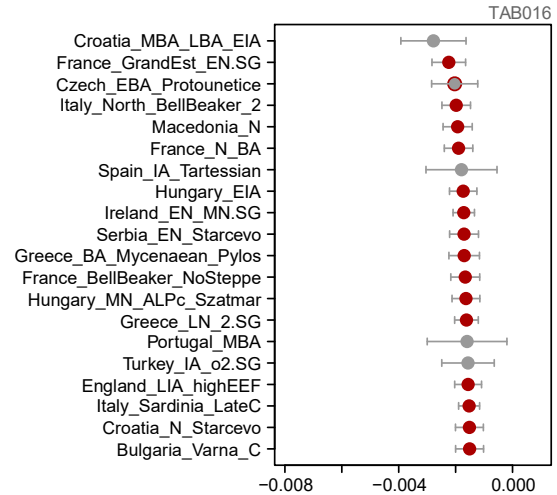
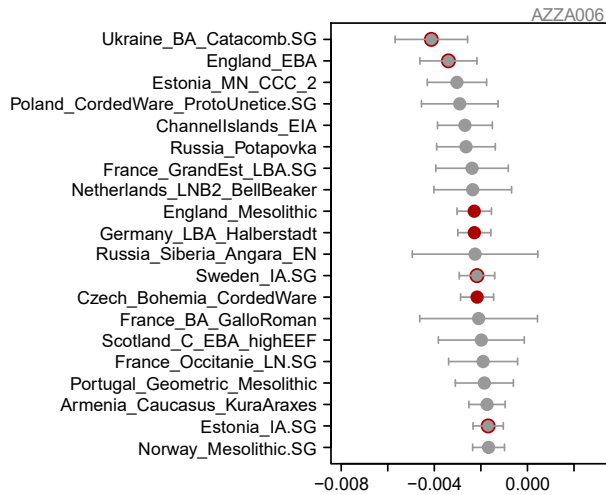
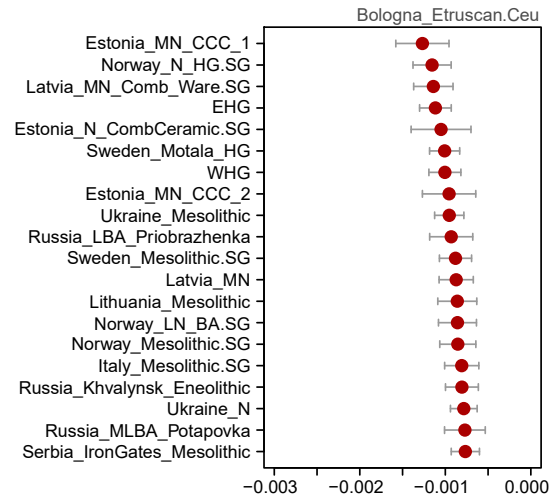
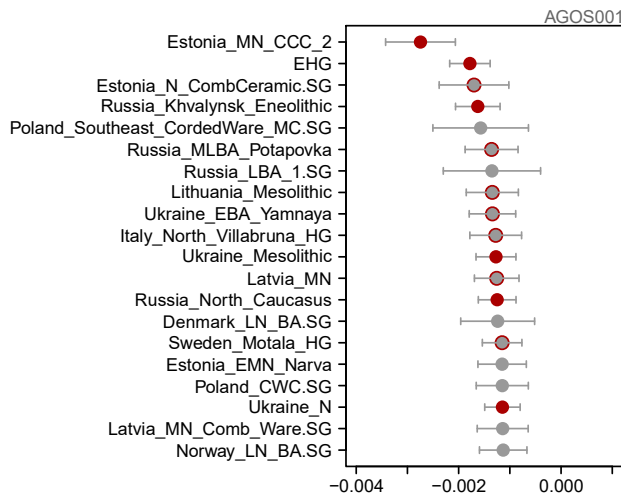
Outgroup	Pop1	Pop2 (*only dated)	Pop3 (*only dated)	f4-stats	StdErr	Zscore	ABBA	BABA	SNPs
Onge.DG	Morocco_EN.SG	C.Italy_Etruscan	Bologna_Etruscan	0.000035	0.000154	0.23	31804	31783	621712
Onge.DG	WHG	C.Italy_Etruscan	Bologna_Etruscan	0.000091	0.000118	0.77	55946	55848	1087433
Onge.DG	Turkey_N	C.Italy_Etruscan	Bologna_Etruscan	0.000127	0.000105	1.21	56590	56453	1079489
Onge.DG	Russia_Samara_EBA_Yamnaya	C.Italy_Etruscan	Bologna_Etruscan	0.000169	0.000105	1.60	55175	54993	1078018
Onge.DG	Morocco_Iberomaurusian	C.Italy_Etruscan	Bologna_Etruscan	0.000190	0.000123	1.54	54535	54334	1058108
Onge.DG	EHG	C.Italy_Etruscan	Bologna_Etruscan	0.000274	0.000122	2.24	51551	51273	1016369
Onge.DG	Iran_GanjDareh_N	C.Italy_Etruscan	Bologna_Etruscan	0.000280	0.000107	2.61	52033	51746	1023535
Onge.DG	Morocco_EN.SG	C.Italy_Etruscan*	Bologna_Etruscan*	0.000007	0.000292	0.02	26900	26896	528776
Onge.DG	Morocco_Iberomaurusian	C.Italy_Etruscan*	Bologna_Etruscan*	0.000132	0.000259	0.51	46830	46710	913427
Onge.DG	WHG	C.Italy_Etruscan*	Bologna_Etruscan*	0.000160	0.000266	0.60	47314	47166	924416
Onge.DG	Turkey_N	C.Italy_Etruscan*	Bologna_Etruscan*	0.000209	0.000224	0.93	48182	47989	923413
Onge.DG	Iran_GanjDareh_N	C.Italy_Etruscan*	Bologna_Etruscan*	0.000278	0.000231	1.21	45483	45233	898991
Onge.DG	Russia_Samara_EBA_Yamnaya	C.Italy_Etruscan*	Bologna_Etruscan*	0.000319	0.000225	1.42	47005	46711	921956
Onge.DG	EHG	C.Italy_Etruscan*	Bologna_Etruscan*	0.000327	0.000258	1.27	44967	44677	889049
<b>Onge.DG</b>	<b>EHG</b>	<b>Bologna_Etruscan.Ceu</b>	<b>Bologna_Etruscan</b>	<b>-0.001116</b>	<b>0.000185</b>	<b>-6.04</b>	<b>49442</b>	<b>50538</b>	<b>982206</b>
<b>Onge.DG</b>	<b>WHG</b>	<b>Bologna_Etruscan.Ceu</b>	<b>Bologna_Etruscan</b>	<b>-0.001005</b>	<b>0.000187</b>	<b>-5.39</b>	<b>52878</b>	<b>53918</b>	<b>1035894</b>
<b>Onge.DG</b>	<b>Russia_Samara_EBA_Yamnaya</b>	<b>Bologna_Etruscan.Ceu</b>	<b>Bologna_Etruscan</b>	<b>-0.000537</b>	<b>0.000150</b>	<b>-3.59</b>	<b>52590</b>	<b>53144</b>	<b>1030973</b>
Onge.DG	Iran_GanjDareh_N	Bologna_Etruscan.Ceu	Bologna_Etruscan	0.000345	0.000168	2.06	50600	50259	992333
<b>Onge.DG</b>	<b>Morocco_Iberomaurusian</b>	<b>Bologna_Etruscan.Ceu</b>	<b>Bologna_Etruscan</b>	<b>0.000696</b>	<b>0.000180</b>	<b>3.88</b>	<b>52765</b>	<b>52056</b>	<b>1016932</b>
<b>Onge.DG</b>	<b>Morocco_EN.SG</b>	<b>Bologna_Etruscan.Ceu</b>	<b>Bologna_Etruscan</b>	<b>0.000715</b>	<b>0.000232</b>	<b>3.08</b>	<b>30585</b>	<b>30161</b>	<b>592760</b>
<b>Onge.DG</b>	<b>Turkey_N</b>	<b>Bologna_Etruscan.Ceu</b>	<b>Bologna_Etruscan</b>	<b>0.001444</b>	<b>0.000154</b>	<b>9.38</b>	<b>54899</b>	<b>53406</b>	<b>1033792</b>
Onge.DG	Morocco_EN.SG	C.Italy_Etruscan.Ceu	Bologna_Etruscan.Ceu	-0.000439	0.000387	-1.13	25064	25281	492605
Onge.DG	Iran_GanjDareh_N	C.Italy_Etruscan.Ceu	Bologna_Etruscan.Ceu	-0.000421	0.000304	-1.38	42729	43084	843368
Onge.DG	Morocco_Iberomaurusian	C.Italy_Etruscan.Ceu	Bologna_Etruscan.Ceu	-0.000375	0.000322	-1.17	43842	44162	853516
Onge.DG	Turkey_N	C.Italy_Etruscan.Ceu	Bologna_Etruscan.Ceu	-0.000344	0.000277	-1.24	44754	45050	859082
Onge.DG	EHG	C.Italy_Etruscan.Ceu	Bologna_Etruscan.Ceu	-0.000012	0.000308	-0.04	42450	42460	83266
Onge.DG	Russia_Samara_EBA_Yamnaya	C.Italy_Etruscan.Ceu	Bologna_Etruscan.Ceu	0.000005	0.000250	0.02	44030	44026	858396
Onge.DG	WHG	C.Italy_Etruscan.Ceu	Bologna_Etruscan.Ceu	0.000721	0.000311	2.32	44630	44011	859435
<b>Onge.DG</b>	<b>Morocco_EN.SG</b>	<b>C.Italy_Imperial</b>	<b>Bologna_Imperial</b>	<b>-0.001422</b>	<b>0.000395</b>	<b>-3.60</b>	<b>20495</b>	<b>21083</b>	<b>412916</b>
<b>Onge.DG</b>	<b>Morocco_Iberomaurusian</b>	<b>C.Italy_Imperial</b>	<b>Bologna_Imperial</b>	<b>-0.001058</b>	<b>0.000304</b>	<b>-3.48</b>	<b>36121</b>	<b>36882</b>	<b>718774</b>
Onge.DG	Iran_GanjDareh_N	C.Italy_Imperial	Bologna_Imperial	-0.000780	0.000289	-2.70	35369	35924	711651
Onge.DG	Turkey_N	C.Italy_Imperial	Bologna_Imperial	-0.000636	0.000263	-2.42	36799	37257	721149
Onge.DG	Russia_Samara_EBA_Yamnaya	C.Italy_Imperial	Bologna_Imperial	0.000162	0.000259	0.63	36179	36063	720927
Onge.DG	WHG	C.Italy_Imperial	Bologna_Imperial	0.000417	0.000315	1.32	36296	35995	721171
Onge.DG	EHG	C.Italy_Imperial	Bologna_Imperial	0.000557	0.000316	1.76	35493	35099	708189

**Tabella 6.10** Risultati delle statistiche F<sub>4</sub> distali effettuate per indagare le caratteristiche genetiche dei principali clusters individuati. In rosso sono evidenziati i confronti che hanno restituito Z score significativo (< |3|).

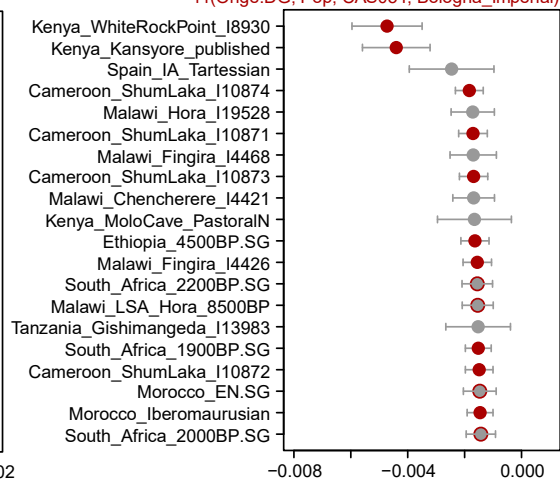


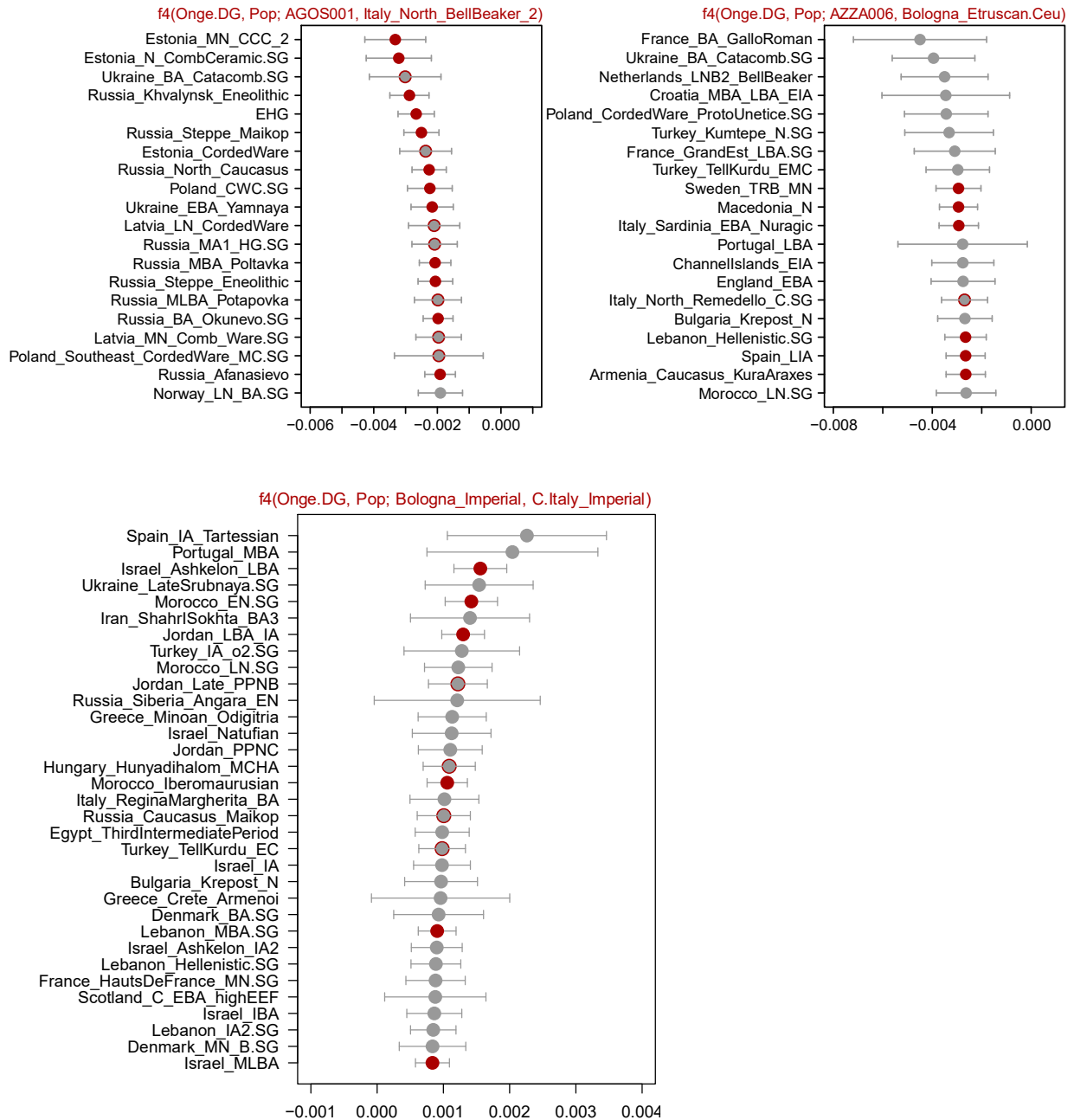
**Figura 6.11** Primi 20 risultati, ordinati in senso crescente, delle statistiche F<sub>3</sub> effettuate per testare ipotesi di *admixture* relativamente al cluster Bologna\_Etruscan.Ceu. I punti rossi rappresentano test significativi, con Z score < -3, mentre quelli con il solo contorno rosso test con Z score compreso tra 2,50 e 3; i punti grigi indicano statistiche non significative.

f4(Onge.DG, Pop; Test, Bologna\_Etruscan)



f4(Onge.DG, Pop; CAS054, Bologna\_Imperial)





**Figura 6.12** Primi 20 risultati, riportati in ordine crescente, delle principali statistiche F4 effettuate per meglio definire il profilo genetico associato ai clusters e relativi *outliers* del dataset padano. I punti con riempimento rosso rappresentano test significativi, con Z score minore di -3, mentre i punti con il solo contorno rosso evidenziano test con Z score compreso tra 2,50 e 3; i punti grigi indicano statistiche non significative. Per l'ultima statistica sono stati riportati tutti i risultati a partire dal valore di f4 positivo più elevato fino all'ultimo confronto significativo.



## 6.8 Modelli di *admixture* ottenuti con qpAdm

Mediante il software qpAdm sono stati testati possibili modelli per verificare se due popolazioni potessero essere equamente relazionate rispetto ad un set di campioni antichi di riferimento o per modellare un gruppo come derivante dall'*admixture* di due popolazioni sorgente (rispettivamente *1-way* e *2-way models*). Questo ha permesso di approfondire i segnali osservati in precedenza, in particolare dalle statistiche F.

L'analisi è stata condotta utilizzando 365 popolazioni antiche provenienti dell'Eurasia occidentale e dell'Africa settentrionale, datate tra il Neolitico medio e l'età del Ferro, i cui dati sono disponibili in letteratura o inclusi nel database pubblico 1240K [176]. Sono stati considerati validi i modelli con p-value > 0,05 e, nel caso dei test *two-way*, selezionati quelli che presentavano bassi errori standard (< 10) e proporzioni verosimili dei due contributi (comprese tra 0 e 1, escludendo valori troppo prossimi agli estremi).

In *Tabella 6.11* vengono riportati i modelli validi ottenuti dal preliminare test *one-way*.

Per quanto riguarda il cluster etrusco principale, si evidenzia una generale affinità con popolazioni dell'età del Bronzo e del Ferro provenienti dall'Europa meridionale (tra cui la penisola balcanica), nonché segnali di continuità genetica con individui o gruppi italiani dell'età del Bronzo e del Ferro caratterizzati dalla componente della steppa, tra i quali compaiono Italy\_North\_BellBeaker\_3 [106] e Italy\_Broion\_EBA [142] dall'Italia settentrionale. L'ipotesi che i clusters etruschi Bologna\_Etruscan e C.Italy\_Etruscan formano un clade è supportata solo nel momento in cui vengono trattenuti nel test esclusivamente gli individui datati al radiocarbonio.

Si osservano modelli validi anche per Turkey\_IA\_o2.SG [103] ed England\_MIA\_LIA\_highEEF [107], individui che presentano profili peculiari: il primo si sovrappone alla variabilità genetica osservata in sud Europa durante l'età del Bronzo, mentre il secondo costituisce un *outlier* rispetto al cluster di riferimento per una maggiore affinità alla componente EEF. Tuttavia, il segnale ottenuto nel caso di Turkey\_IA\_o2.SG potrebbe essere influenzato dalla copertura non elevata di questo campione sul pannello 1240K (meno di 100.000 SNPs); tale ipotesi è supportata anche dal fatto che è possibile ottenere test validi con Turkey\_IA\_o2.SG come popolazione sorgente anche per Bologna\_Imperial e CAS054, geneticamente diversi tra loro e dal cluster etrusco.

A differenza di quanto appena riscontrato per il cluster principale etrusco, Bologna\_Etruscan.Ceu e AGOS001 restituiscono p-value > 0,05 prevalentemente quando i modelli coinvolgono popolazioni (dell'età del Bronzo e del Ferro) provenienti dall'Europa centrale e settentrionale (con l'aggiunta di Spain\_EBA\_Mallorca che si caratterizza per una consistente frazione di *steppe ancestry* [116]).

È interessante notare come Bologna\_Etruscan.Ceu ed AGOS001 mostrino rispettivamente affinità genetica con C.Italy\_Tarquini IA\_CentralEuropean\_10342 e 10339, campioni etruschi di Tarquinia caratterizzati da un profilo genetico non locale associato a popolazioni del centro Europa [110]. In particolare, C.Italy\_Tarquini IA\_CentralEuropean\_10339 è l'individuo in prossimità del quale cade AGOS001 in PCA.

Nonostante segnali simili siano visibili anche per l'individuo AZZA006, rappresentati dai test con p-value significativo nel caso in cui la popolazione sorgente è rappresentata da C.Italy\_Etruscan.Ceu, C.Italy\_Tarquini IA\_CentralEuropean\_10342, gruppi celtici e numerose popolazioni del centro e nord Europa, emergono anche affinità con un individuo etrusco caratterizzato da profilo genetico locale (C.Italy\_Etruscan\_UDC\_P [117]), Italy\_North\_BellBeaker\_3 e popolazioni dei Balcani (come osservato per Bologna\_Etruscan). Ancora una volta, questo campione manifesta caratteri genetici intermedi rispetto al *pool* locale e al gruppo di individui probabilmente più strettamente relazionati ai territori d'oltralpe. È importante tenere in considerazione, però, che l'elevato numero di modelli validi ottenuti per AZZA006 potrebbe dipendere anche dalla copertura non elevata del campione sul pannello utilizzato.

Infine, nonostante l'outlier TAB016 e Bologna\_Etruscan risultino in alcuni casi essere descritti dal medesimo modello (come nel caso in cui Italy\_Broion\_EBA rappresenta la popolazione sorgente), il segnale generale che si osserva per TAB016 è una certa vicinanza genetica a gruppi accomunati da una ridotta proporzione di componente della steppa (come, per esempio, Italy\_Sicily\_BellBeaker, Italy\_Sardinia\_IA\_2, Ukraine\_Eneolithic\_Trypillia e France\_BellBeaker\_LowSteppe).

Per quanto riguarda Bologna\_Imperial, l'unico modello *1-way* valido (escludendo Turkey\_IA\_o2.SG) è quello in cui la popolazione sorgente è rappresentata da un individuo etrusco del centro Italia che presenta un componente genetica di derivazione vicino-orientale (C.Italy\_Etruscan\_MAS001 [117]), mentre CAS054 risulta affine ad un campione dell'età del Ferro, ritrovato ad Ibiza, caratterizzato da un'*ancestry* nordafricana [116, 210].

Test	Source	p-value
Bologna_Etruscan	Spain_MBA_Formentera_all	0.473598
<b>Bologna_Etruscan (dated)</b>	<b>C.Italy_Etruscan (dated)</b>	<b>0.407883</b>
<b>Bologna_Etruscan</b>	<b>Italy_North_BellBeaker_3</b>	<b>0.206135</b>
Bologna_Etruscan	Slovenia_LBA_EIA	0.203774
Bologna_Etruscan	Croatia_MBA_LBA_EIA	0.192288
Bologna_Etruscan	England_MIA_LIA_highEEF	0.180648
<b>Bologna_Etruscan</b>	<b>Italy_Broion_EBA</b>	<b>0.119596</b>
Bologna_Etruscan	Turkey_IA_o2.SG	0.115247
Bologna_Etruscan	Bulgaria_EBA_Ezero	0.072968
<b>Bologna_Etruscan</b>	<b>Italy_Sardinia_IA_2</b>	<b>0.055651</b>
Bologna_Etruscan.Ceu	ChannellIslands_IA	0.923523
Bologna_Etruscan.Ceu	France_BA	0.860491
Bologna_Etruscan.Ceu	England_MIA_LIA_highEEF	0.846479
Bologna_Etruscan.Ceu	England_BellBeaker_highEEF	0.739740
Bologna_Etruscan.Ceu	England_MIA_highEEF	0.688143
Bologna_Etruscan.Ceu	Czech_MBA_Tumulus	0.666263
Bologna_Etruscan.Ceu	IsleofMan_EBA_lowsteppe	0.665477
Bologna_Etruscan.Ceu	France_Occitanie_MBA.SG	0.640150
Bologna_Etruscan.Ceu	Spain_Hellenistic_oLocal	0.589817
Bologna_Etruscan.Ceu	England_LIA_highEEF	0.574780
Bologna_Etruscan.Ceu	Spain_EBA_Mallorca	0.441059
Bologna_Etruscan.Ceu	France_EBA	0.399260
Bologna_Etruscan.Ceu	Sweden_BA.SG	0.352902
Bologna_Etruscan.Ceu	Czech_IA_LaTene_Hallstatt	0.309581
Bologna_Etruscan.Ceu	France_BA_GalloRoman	0.299738
Bologna_Etruscan.Ceu	England_EIA_highEEF	0.264120
Bologna_Etruscan.Ceu	France_Occitanie_EBA.SG	0.214849
<b>Bologna_Etruscan.Ceu</b>	<b>C.Italy_Tarquinia_IA_CentralEuropean_10342</b>	<b>0.197133</b>
Bologna_Etruscan.Ceu	France_GrandEst_LBA.SG	0.184605
Bologna_Etruscan.Ceu	Hungary_LBA_Kyjatice	0.183672
Bologna_Etruscan.Ceu	Switzerland_BellBeaker	0.175436
Bologna_Etruscan.Ceu	France_HautsDeFrance_IA2.SG	0.108650
Bologna_Etruscan.Ceu	Croatia_MBA_LBA_EIA	0.101440
Bologna_Etruscan.Ceu	French_EIA_Alsace	0.092748
Bologna_Etruscan.Ceu	France_GrandEst_IA1.SG	0.090774
Bologna_Etruscan.Ceu	France_GrandEst_IA2	0.081275
Bologna_Etruscan.Ceu	Hungary_EIA_Hallstatt	0.075071
Bologna_Etruscan.Ceu	England_LBA_highEEF	0.073061
Bologna_Etruscan.Ceu	ChannellIslands_EIA	0.055722
Bologna_Etruscan.Ceu	England_MBA_highEEF	0.051001
AGOS001	IsleofMan_EBA_lowsteppe	0.857425
AGOS001	Germany_LBA_Halberstadt	0.367190
AGOS001	Netherlands_BA	0.354360
AGOS001	Czech_EBA_Protounetice	0.340712
AGOS001	Scotland_MIA	0.297201
AGOS001	France_BA_GalloRoman	0.295442
AGOS001	Scotland_C_EBA_mediumhighEEF	0.220802
AGOS001	England_EastYorkshire_LIA	0.218579
AGOS001	Czech_BA_Veterov_1	0.215672
AGOS001	Netherlands_MIA_LIA	0.205111
AGOS001	England_EBA	0.189644
AGOS001	Lithuania_LBA.SG	0.180533
AGOS001	Scotland_MBA	0.179604
AGOS001	Netherlands_LIA	0.161827
AGOS001	French_IA_Normandy	0.145670
AGOS001	Netherlands_EIA	0.141686
AGOS001	England_EastYorkshire_EIA	0.129424
<b>AGOS001</b>	<b>C.Italy_Tarquinia_IA_CentralEuropean_10339</b>	<b>0.121079</b>

Target	Source	p-value
AGOS001	Poland_BellBeaker	0.112349
AGOS001	Sweden_BA.SG	0.104248
AGOS001	England_IA	0.094847
AGOS001	Netherlands_MBA_LBA	0.075690
AGOS001	ChannellIslands_LIA_LaTene	0.061255
AGOS001	England_EastYorkshire_IA	0.058626
AGOS001	England_BellBeaker_mediumEEF	0.057914
AGOS001	France_BA	0.057906
AGOS001	Scotland_LBA	0.050199
TAB016	Bulgaria_EBA	0.687587
TAB016	Bulgaria_BeliBreyag_EBA	0.568714
<b>TAB016</b>	<b>Italy_Sardinia_IA_2</b>	<b>0.564591</b>
TAB016	Bulgaria_EBA_Ezero	0.532925
<b>TAB016</b>	<b>Italy_Sicily_BellBeaker</b>	<b>0.345897</b>
<b>TAB016</b>	<b>Italy_Sardinia_IA_Punic_2</b>	<b>0.275274</b>
TAB016	Ukraine_Eneolithic_Trypillia	0.067554
<b>TAB016</b>	<b>Italy_Broion_EBA</b>	<b>0.063676</b>
TAB016	France_BellBeaker_LowSteppe	0.063318
TAB016	Spain_EIA_Tartessian	0.060009
AZZA006	England_MIA_LIA_highEEF	0.913172
AZZA006	Hungary_EIA_Hallstatt	0.800715
AZZA006	Serbia_BA_Maros	0.762860
AZZA006	Hungary_IA_LaTene	0.731362
AZZA006	Germany_Lech_MBA	0.696789
<b>AZZA006</b>	<b>C.Italy_Etruscan.Ceu</b>	<b>0.637649</b>
<b>AZZA006</b>	<b>Italy_North_BellBeaker_3</b>	<b>0.618361</b>
AZZA006	Slovakia_IA_Vekerzug	0.588637
AZZA006	Hungary_EBA_Protonagyrev	0.579377
AZZA006	Slovenia_LBA	0.544260
AZZA006	Slovakia_LIA_LaTene	0.538191
AZZA006	Hungary_EIA_Prescythian_Mezocsat	0.534862
AZZA006	Hungary_EBA_BellBeaker	0.525187
AZZA006	Hungary_LBA	0.453510
AZZA006	Hungary_LaTene	0.439127
AZZA006	England_LBA_highEEF	0.403333
AZZA006	Czech_MBA_Tumulus	0.372357
AZZA006	Germany_SouthernGermany_Singen_IA	0.368369
AZZA006	Slovenia_LBA_EIA	0.311039
AZZA006	Turkey_IA_o2.SG	0.310838
AZZA006	Croatia_MBA_LBA	0.305960
AZZA006	Spain_Greek_oLocal	0.305919
AZZA006	France_SouthEast_IA2	0.287547
AZZA006	Germany_Roman.SG	0.280440
AZZA006	England_LIA_highEEF	0.279957
AZZA006	Croatia_MBA	0.275328
AZZA006	Germany_Lech_EBA	0.251663
AZZA006	ChannellIslands_EIA	0.249856
AZZA006	Serbia_Mokrin_EBA_Maros	0.236190
AZZA006	France_GrandEst_IA1.SG	0.211963
AZZA006	French_EIA_Alsace	0.206193
AZZA006	Austria_IA_LaTene	0.199325
AZZA006	England_EIA_highEEF	0.198198
AZZA006	France_BA_GalloRoman	0.197054
AZZA006	France_Occitanie_MBA.SG	0.195220
AZZA006	France_BA	0.191700
AZZA006	Slovenia_MBA	0.191238
AZZA006	Czech_IA_Hallstatt	0.177026
AZZA006	France_HautsDeFrance_IA2.SG	0.172982

Target	Source	p-value
AZZA006	Czech_IA_LaTene	0.170013
AZZA006	Slovenia_EIA	0.144652
AZZA006	ChannellIslands_IA	0.135837
AZZA006	Spain_Hellenistic_oLocal	0.135293
AZZA006	Hungary_LBA_Kyjatice	0.134243
AZZA006	England_MIA_highEEF	0.125119
AZZA006	IsleofMan_EBA_lowsteppe	0.119823
<b>AZZA006</b>	<b>C.Italy_Etruscan_UDC_P</b>	<b>0.113688</b>
AZZA006	Czech_IA_LaTene_Hallstatt	0.110198
AZZA006	Germany_SouthernGermany_Singen_EBA	0.101738
AZZA006	Croatia_LIA_LaTene	0.101033
AZZA006	Croatia_LateC_EBA_Vucedol	0.100055
AZZA006	France_GrandEst_IA2	0.099575
AZZA006	France_Occitanie_EBA.SG	0.096962
AZZA006	Switzerland_BellBeaker	0.089230
AZZA006	Spain_EBA_Mallorca	0.088660
<b>AZZA006</b>	<b>C.Italy_Tarquini IA_CentralEuropean_10342</b>	<b>0.086071</b>
AZZA006	Spain_IA_Celt	0.084769
AZZA006	England_C_EBA_highEEF	0.084111
AZZA006	French_BA_Paris_Basin	0.082251
AZZA006	England_MBA_highEEF	0.081453
AZZA006	Spain_MBA_Formentera_all	0.070721
AZZA006	Germany_SouthernGermany_Anselfingen_EBA	0.070434
AZZA006	Slovakia_LIA	0.067554
Bologna_Imperial	Turkey_IA_o2.SG	0.726644
<b>Bologna_Imperial</b>	<b>C.Italy_Etruscan_MAS001</b>	<b>0.085332</b>
CAS054	Ibiza_Punic.SG	0.324100
CAS054	Turkey_IA_o2.SG	0.061600

**Tabella 6.11** Modelli 1-way validi (p-value > 0,05) ottenuti con il software qpAdm. In rosso sono evidenziati i modelli che mostrano segnali di affinità genetica tra gli individui/cluster in esame ed altri campioni del contesto italiano.

Il passaggio successivo ha riguardato la verifica, per i campioni dell'età del Ferro, di modelli 2-way che prevedono *admixture* tra la *steppe ancestry*, rappresentata dai gruppi di pastori della prima età del Bronzo provenienti dalle regioni della Russia occidentale (Russia\_Samara\_EBA\_Yamnaya) o da popolazioni di cultura *Bell Beaker* del centro Europa (Germany\_BellBeaker, più vicini nel tempo e nello spazio), e popolazioni del contesto italiano datate tra il Neolitico e l'età del Bronzo (sia delle regioni peninsulari che insulari).

Nel grafico di *Figura 6.13*, che riassume i risultati ottenuti, i modelli statisticamente significativi per ogni gruppo/individuo sono indicati con una casella verde.



cluster etrusco principale dell'Etruria classica, descritto come derivante da *admixture* tra gruppi neolitici/calcolitici del contesto italiano e Russia\_Samara\_EBA\_Yamnaya/Germany\_BellBeaker, quest'ultimi rappresentati da percentuali intorno al 25 e 50% [117].

Anche per AZZA006 e TAB016 i modelli funzionano quando la componente locale è rappresentata da un gruppo antico (neolitico o calcolitico) oppure non caratterizzata da un'elevata componente della steppa. In media, la seconda componente è presente in proporzione più elevata per l'individuo AZZA006 rispetto a Bologna\_Etruscan (35 e 74% rispettivamente per Russia\_Samara\_EBA\_Yamnaya e Germany\_BellBeaker) e più bassa per TAB016 (14% e 30%), a conferma delle ipotesi già avanzate in precedenza circa la diversa affinità di questi campioni alla *steppe-related ancestry*. In generale, si osservano percentuali maggiori di componente associata alla steppa quando questa risulta essere meno abbondante nel gruppo locale considerato dal modello. Infatti, l'unico modello funzionante per AZZA006 che comprende un individuo con una più abbondante frazione di componente della steppa (Italy\_North\_BellBeaker\_1) è associato al valore più basso di Russia\_Samara\_EBA\_Yamnaya e Germany\_BellBeaker riscontrato per il campione in questione (24 e 60%).

Solo un test risulta valido per il gruppo Bologna\_Etruscan.Ceu (Italy\_North\_Remedello\_EBA.SG con 40% Russia\_Samara\_EBA\_Yamnaya e 80% Germany\_BellBeaker), probabilmente a causa del fatto che le popolazioni sorgente selezionate non costituiscono delle buone *proxy* per la descrizione del profilo genetico del cluster; bisogna considerare, infatti, che entrambe le popolazioni associate alla steppa sono decisamente più antiche dei campioni in esame (il gruppo Yamnaya utilizzato li precede di oltre due millenni) e per questo motivo potrebbero non rappresentare delle popolazioni sorgente ideali, soprattutto nell'eventualità di un *admixture* più recente.

Sebbene per l'individuo AGOS001 non vi siano modelli funzionanti, quelli con Germany\_BellBeaker mostrano in tutti i casi una percentuale associata a quest'ultimo gruppo superiore al 100%; ciò suggerisce che AGOS001 potrebbe derivare totalmente (o comunque in gran parte) la sua *ancestry* da popolazioni dell'Europa centrale o settentrionale geneticamente simili a Germany\_BellBeaker.

Procedendo con i modelli 2-way testati selezionando come popolazioni sorgente Italy\_Broion\_BA (rappresentativo del *pool* genetico esistente in Italia settentrionale nel periodo precedente a quello trattato) e gli stessi gruppi precedentemente utilizzati per il test 1-way, emergono alcuni modelli statisticamente significativi anche per AGOS001 e Bologna\_Etruscan.Ceu.

Come già accennato nel *Paragrafo 5.3.9*, non verranno riportati in questa sezione i risultati relativi a TAB016 poiché la selezione della popolazione sorgente locale non può essere effettuata in modo adeguato in assenza di una datazione diretta del campione, anche alla luce del fatto che si tratta di un *outlier*.

In *Tabella 6.12* sono visibili i modelli validi ottenuti per AGOS001, Bologna\_Etruscan.Ceu e AZZA006.

Test	Pop1	Pop2	p-value	% Pop1	% Pop2	SE
AGOS001	Italy_Broion_BA	Poland_EBA	0.690298	0.28	0.73	0.06
AGOS001	Italy_Broion_BA	England_EBA	0.679432	0.22	0.78	0.08
AGOS001	Italy_Broion_BA	Netherlands_EIA	0.523033	0.21	0.79	0.07
AGOS001	Italy_Broion_BA	Netherlands_BellBeaker	0.120333	0.27	0.73	0.05
AGOS001	Italy_Broion_BA	Ukraine_Eneolithic	0.108198	0.40	0.60	0.05
AGOS001	Italy_Broion_BA	Netherlands_EBA	0.081151	0.24	0.76	0.08
AGOS001	Italy_Broion_BA	Netherlands_MBA	0.080617	0.16	0.84	0.06
AGOS001	Italy_Broion_BA	Germany_CordedWare-Tauber	0.063225	0.36	0.64	0.07
AGOS001	Italy_Broion_BA	Latvia_BA	0.054511	0.37	0.63	0.05
AGOS001	Italy_Broion_BA	Lithuania_BA	0.054045	0.31	0.70	0.06
Bologna_Etruscan.Ceu	Italy_Broion_BA	England_EBA	0.764069	0.48	0.52	0.05
Bologna_Etruscan.Ceu	Italy_Broion_BA	Lithuania_LBA.SG	0.550037	0.52	0.48	0.09
Bologna_Etruscan.Ceu	Italy_Broion_BA	Netherlands_EIA	0.434004	0.49	0.51	0.04
Bologna_Etruscan.Ceu	Italy_Broion_BA	Czech_EBA_Protounetice	0.429478	0.38	0.62	0.08
Bologna_Etruscan.Ceu	Italy_Broion_BA	Czech_BA_Veterov_1	0.350417	0.39	0.61	0.06
Bologna_Etruscan.Ceu	Italy_Broion_BA	Netherlands_LIA	0.284278	0.35	0.65	0.06
Bologna_Etruscan.Ceu	Italy_Broion_BA	England_EastYorkshire_EIA	0.145123	0.28	0.72	0.07
Bologna_Etruscan.Ceu	Italy_Broion_BA	Ukraine_Eneolithic	0.129934	0.60	0.40	0.03
Bologna_Etruscan.Ceu	Italy_Broion_BA	Poland_EBA	0.083891	0.52	0.48	0.04
Bologna_Etruscan.Ceu	Italy_Broion_BA	Lithuania_BA	0.083502	0.54	0.46	0.03
Bologna_Etruscan.Ceu	Italy_Broion_BA	England_EastYorkshire_IA	0.080639	0.22	0.78	0.09
Bologna_Etruscan.Ceu	Italy_Broion_BA	Denmark_BA.SG	0.078239	0.38	0.62	0.07
Bologna_Etruscan.Ceu	Italy_Broion_BA	Netherlands_MIA_LIA	0.054101	0.34	0.66	0.05
Bologna_Etruscan.Ceu	Italy_Broion_BA	Latvia_BA	0.051591	0.58	0.42	0.03
AZZA006	Italy_Broion_BA	Germany_CordedWare-Tauber	0.739061	0.71	0.29	0.07
AZZA006	Italy_Broion_BA	Czech_Bohemia_CordedWare	0.642559	0.73	0.27	0.06
AZZA006	Italy_Broion_BA	Netherlands_LBA	0.540690	0.64	0.37	0.08
AZZA006	Italy_Broion_BA	Bulgaria_MLBA	0.540319	0.73	0.27	0.06
AZZA006	Italy_Broion_BA	Russia_MLBA_Potapovka	0.534222	0.83	0.17	0.04
AZZA006	Italy_Broion_BA	England_EBA	0.519500	0.66	0.34	0.09
AZZA006	Italy_Broion_BA	Russia_Potapovka	0.509160	0.80	0.20	0.05
AZZA006	Italy_Broion_BA	England_C_EBA	0.503921	0.62	0.39	0.09
AZZA006	Italy_Broion_BA	Poland_ChopiceVeseleCulture	0.494079	0.68	0.32	0.08
AZZA006	Italy_Broion_BA	Russia_Srubnaya	0.474440	0.72	0.28	0.06
AZZA006	Italy_Broion_BA	Russia_MLBA_Krasnoyarsk	0.473110	0.71	0.29	0.07
AZZA006	Italy_Broion_BA	Netherlands_MBA	0.468798	0.61	0.39	0.09
AZZA006	Italy_Broion_BA	Czech_EBA_Unetice	0.467130	0.62	0.38	0.09
AZZA006	Italy_Broion_BA	Scotland_MIA_LIA	0.461444	0.60	0.40	0.09
AZZA006	Italy_Broion_BA	England_BellBeaker_lowEEF	0.452527	0.70	0.30	0.07
AZZA006	Italy_Broion_BA	Russia_Steppe_Catacomb	0.452119	0.83	0.17	0.04
AZZA006	Italy_Broion_BA	Ukraine_EBA_Yamnaya	0.448631	0.82	0.18	0.04
AZZA006	Italy_Broion_BA	England_MBA	0.446163	0.58	0.42	0.10
AZZA006	Italy_Broion_BA	Netherlands_BellBeaker	0.443891	0.66	0.34	0.08
AZZA006	Italy_Broion_BA	Scotland_LIA	0.443262	0.58	0.42	0.10
AZZA006	Italy_Broion_BA	Scotland_LBA	0.441842	0.57	0.43	0.10
AZZA006	Italy_Broion_BA	Russia_Steppe_Maikop	0.431752	0.88	0.12	0.03
AZZA006	Italy_Broion_BA	Russia_MLBA_Sintashta	0.422713	0.70	0.30	0.07
AZZA006	Italy_Broion_BA	Scotland_MBA	0.419648	0.59	0.41	0.10
AZZA006	Italy_Broion_BA	Netherlands_BA	0.414632	0.61	0.39	0.10
AZZA006	Italy_Broion_BA	Russia_Andronovo.SG	0.413958	0.75	0.25	0.06
AZZA006	Italy_Broion_BA	Czech_CordedWare	0.413047	0.71	0.29	0.07
AZZA006	Italy_Broion_BA	England_BellBeaker	0.411559	0.65	0.35	0.08
AZZA006	Italy_Broion_BA	Russia_North_Caucasus	0.407861	0.83	0.17	0.04
AZZA006	Italy_Broion_BA	Sweden_IA.SG	0.406798	0.64	0.36	0.10
AZZA006	Italy_Broion_BA	Russia_MBA_Poltavka	0.406660	0.84	0.16	0.04
AZZA006	Italy_Broion_BA	Russia_Afanasiovo	0.390624	0.83	0.17	0.04
AZZA006	Italy_Broion_BA	Czech_EBA	0.390109	0.65	0.35	0.08



Test	Pop1	Pop2	p-value	% Pop1	% Pop2	SE
AZZA006	Italy_Broion_BA	Poland_CWC_3.SG	0.387505	0.74	0.26	0.06
AZZA006	Italy_Broion_BA	Poland_EBA	0.387298	0.70	0.30	0.08
AZZA006	Italy_Broion_BA	Poland_EBA_Unetice.SG	0.387037	0.70	0.30	0.08
AZZA006	Italy_Broion_BA	Germany_EBA_Unetice	0.382566	0.62	0.38	0.09
AZZA006	Italy_Broion_BA	England_EBA_BellBeaker	0.378037	0.66	0.34	0.09
AZZA006	Italy_Broion_BA	Germany_CordedWare	0.365823	0.75	0.25	0.06
AZZA006	Italy_Broion_BA	Scotland_BellBeaker	0.363423	0.66	0.34	0.09
AZZA006	Italy_Broion_BA	France_Alsace_Lingolsheim_EBA	0.353186	0.67	0.33	0.09
AZZA006	Italy_Broion_BA	Ukraine_MBA	0.350734	0.70	0.30	0.08
AZZA006	Italy_Broion_BA	Russia_Caucasus_EBA_Yamnaya	0.347466	0.83	0.17	0.04
AZZA006	Italy_Broion_BA	Netherlands_LNB_EBA_BellBeaker	0.346985	0.69	0.32	0.08
AZZA006	Italy_Broion_BA	Russia_Srubnaya_Alakul.SG	0.333040	0.76	0.24	0.06
AZZA006	Italy_Broion_BA	Poland_CWC.SG	0.326513	0.83	0.17	0.04
AZZA006	Italy_Broion_BA	Ireland_EBA.SG	0.323871	0.66	0.34	0.09
AZZA006	Italy_Broion_BA	Scotland_C_EBA	0.320822	0.64	0.36	0.09
AZZA006	Italy_Broion_BA	Russia_Kalmykia_EBA_Yamnaya.SG	0.311303	0.84	0.16	0.04
AZZA006	Italy_Broion_BA	Sweden_BAC.SG	0.309328	0.66	0.34	0.09
AZZA006	Italy_Broion_BA	Russia_Petrovka	0.295781	0.73	0.27	0.07
AZZA006	Italy_Broion_BA	Netherlands_EIA	0.294556	0.66	0.34	0.10
AZZA006	Italy_Broion_BA	Netherlands_EBA	0.291956	0.70	0.30	0.09
AZZA006	Italy_Broion_BA	Poland_Southeast_CordedWare.SG	0.279320	0.78	0.23	0.06
AZZA006	Italy_Broion_BA	Sweden_IA_1.SG	0.277215	0.71	0.29	0.09
AZZA006	Italy_Broion_BA	Lithuania_BA	0.272776	0.72	0.28	0.07
AZZA006	Italy_Broion_BA	England_C_EBA_lowEEF	0.237253	0.71	0.30	0.09
AZZA006	Italy_Broion_BA	Russia_Mezhovskaya.SG	0.225744	0.79	0.21	0.06
AZZA006	Italy_Broion_BA	Ukraine_Eneolithic	0.220297	0.76	0.24	0.07
AZZA006	Italy_Broion_BA	England_LBA_lowEEF	0.216149	0.66	0.34	0.09
AZZA006	Italy_Broion_BA	Estonia_CordedWare	0.212540	0.76	0.25	0.07
AZZA006	Italy_Broion_BA	Hungary_IA_Prescythian.SG	0.211334	0.71	0.30	0.09
AZZA006	Italy_Broion_BA	Russia_EarlySarmatian_SouthernUrals.SG	0.210458	0.81	0.19	0.05
AZZA006	Italy_Broion_BA	Russia_Siberia_Tenisei_EBA	0.208964	0.90	0.10	0.03
AZZA006	Italy_Broion_BA	Russia_IA_EarlySarmatian	0.203090	0.79	0.21	0.06
AZZA006	Italy_Broion_BA	Latvia_BA	0.201008	0.75	0.26	0.07
AZZA006	Italy_Broion_BA	Russia_Sarmatian.SG	0.194224	0.80	0.20	0.06
AZZA006	Italy_Broion_BA	Russia_Karasuk_oRISE.SG	0.191631	0.83	0.17	0.05
AZZA006	Italy_Broion_BA	Russia_Bolshoy	0.190058	0.87	0.13	0.03
AZZA006	Italy_Broion_BA	Ukraine_Ozera_EBA_Yamnaya	0.186208	0.80	0.20	0.06
AZZA006	Italy_Broion_BA	Russia_BA_Okunevo.SG	0.184634	0.89	0.11	0.03
AZZA006	Italy_Broion_BA	Russia_IA_Ingria.SG	0.176906	0.74	0.26	0.08
AZZA006	Italy_Broion_BA	Russia_Late_Sarmatian.SG	0.176483	0.81	0.19	0.06
AZZA006	Italy_Broion_BA	Russia_Chalmny_Varre	0.172422	0.79	0.21	0.06
AZZA006	Italy_Broion_BA	Russia_Tagar.SG	0.168027	0.81	0.19	0.05
AZZA006	Italy_Broion_BA	Russia_Tuva_IA_AldyBel	0.158076	0.85	0.15	0.05
AZZA006	Italy_Broion_BA	Estonia_BA.SG	0.145088	0.77	0.23	0.07
AZZA006	Italy_Broion_BA	Moldova_Cimmerian.SG	0.138417	0.83	0.17	0.05
AZZA006	Italy_Broion_BA	Norway_IA.SG	0.120956	0.68	0.32	0.10
AZZA006	Italy_Broion_BA	Russia_IA_3.SG	0.119720	0.87	0.13	0.04
AZZA006	Italy_Broion_BA	Iran_BA_HajjiFiruz	0.119401	0.77	0.23	0.08
AZZA006	Italy_Broion_BA	Poland_CWC_1.SG	0.115164	0.76	0.24	0.07
AZZA006	Italy_Broion_BA	Ukraine_IA_WesternScythian.SG	0.103851	0.72	0.28	0.08
AZZA006	Italy_Broion_BA	Ukraine_EBA	0.087868	0.83	0.17	0.06
AZZA006	Italy_Broion_BA	Russia_EasternScythian_SouthernUrals.SG	0.077364	0.86	0.15	0.05
Bologna_Imperial	Bologna_Etruscan	Iran_ShahrIsokhta_BA3	0.698241	0.72	0.28	0.04
Bologna_Imperial	Bologna_Etruscan	Iran_Hasanlu_IA.SG	0.300360	0.75	0.25	0.03
Bologna_Imperial	Bologna_Etruscan	Armenia_Caucasus_EBA_KuraAraxes	0.267061	0.74	0.26	0.03
Bologna_Imperial	Bologna_Etruscan	Armenia_MBA	0.212782	0.72	0.28	0.03
Bologna_Imperial	Bologna_Etruscan	Russia_Caucasus_LateMaikop	0.134669	0.74	0.26	0.03

Test	Pop1	Pop2	p-value	% Pop1	% Pop2	SE
Bologna_Imperial	Bologna_Etruscan	Russia_Caucasus_Maikop	0.123718	0.70	0.30	0.04
Bologna_Imperial	Bologna_Etruscan	Armenia_LBA.SG	0.115553	0.73	0.28	0.03
Bologna_Imperial	Bologna_Etruscan	Turkey_IA_o3.SG	0.102367	0.70	0.30	0.04
Bologna_Imperial	Bologna_Etruscan	Russia_Caucasus_Maikop_Novosvobodnaya	0.087354	0.77	0.23	0.03
Bologna_Imperial	Bologna_Etruscan	Iran_BA1_ShahrISokhta	0.071687	0.86	0.14	0.02
Bologna_Imperial	Bologna_Etruscan	Armenia_EBA	0.067092	0.75	0.26	0.03
Bologna_Imperial	Bologna_Etruscan	Armenia_C	0.066660	0.66	0.34	0.04
CAS054	Bologna_Etruscan	Tunisia_Kerkouane_IA_SubSaharan	0.081959	0.70	0.30	0.03

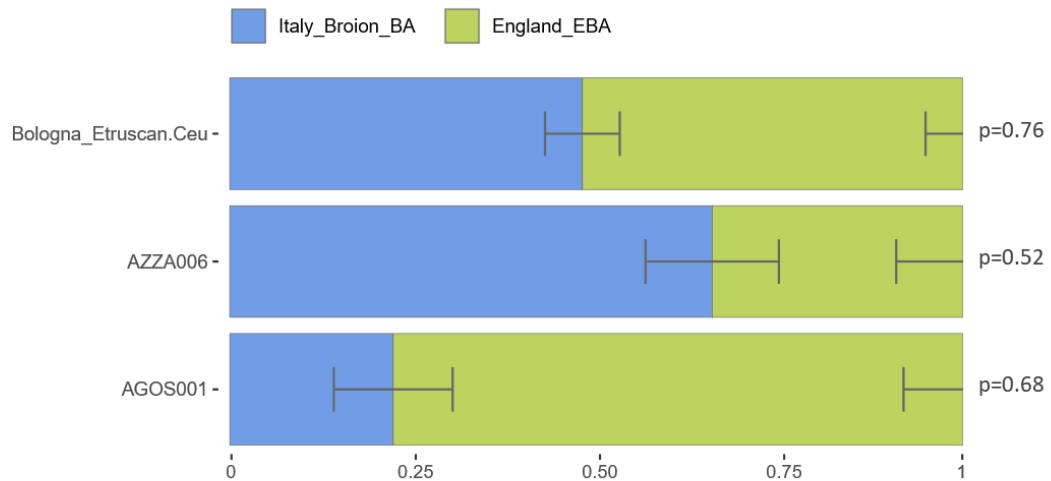
**Tabella 6.12** Modelli di *admixture 2-way* validi (p-value > 0,05) ottenuti con il software qpAdm.

Come anticipato dai test *1-way*, anche in questa analisi i modelli di *admixture* validi per AGOS001 e Bologna\_Etruscan.Ceu coinvolgono principalmente popolazioni del centro e nord Europa riferibili all'età del Bronzo e del Ferro, caratterizzate da un'elevata *steppe ancestry*; in generale si osservano proporzioni associate a tali popolazioni molto elevate nel caso di AGOS001 (60-84%) e leggermente più basse per Bologna\_Etruscan.Ceu (40-78%). L'elevato numero di test funzionanti per AZZA006 potrebbe dipendere, anche in questo caso, dalla copertura limitata sul pannello 1240K, dalla quale derivano anche errori standard mediamente più alti. A differenza dei precedenti campioni, i modelli risultano validi non solo nel caso di *admixture* tra Italy\_Broion\_BA e gruppi del centro e nord Europa ma anche quando la seconda componente è costituita da popolazioni più strettamente associate alla diffusione della *steppe ancestry* (quali Yamnaya provenienti dalle regioni dell'attuale Russia ed Ucraina e culture successive associate alle medesime aree), seppur rappresentate da percentuali minime rispetto a quanto si osserva per gli altri gruppi non locali.

In generale, la componente non locale stimata per AZA006 non supera mai il 43%, coerentemente alla posizione occupata in PCA dal campione, che cade nelle vicinanze del cluster etrusco principale. In *Figura 6.14* è riportato, come esempio, un modello di *admixture* (Italy\_Broion\_BA+ England\_EBA) risultato valido sia per il cluster Bologna\_Etruscan.Ceu che per gli *outliers* AGOS001 e AZZA006.

Come atteso, AGOS001 risulta quasi interamente rappresentato dall'*ancestry* di derivazione non locale (78%), motivo per cui non è possibile escludere per questo campione l'ipotesi alternativa di un'origine del tutto alloctona (vedi modelli *1-way*); percentuali minori di componente England\_EBA sono invece sufficienti per modellare Bologna\_Etruscan.Ceu (52%) e AZZA006 (34%), per i quali, in generale, uno scenario di *admixture* sembra essere quello che meglio descrive il profilo genetico osservato.

Per quanto non sia possibile al momento determinare con maggior risoluzione l'area di provenienza della componente alloctona, i segnali genetici risultano del tutto compatibili con l'ipotesi di contatti diretti con popolazioni di estrazione celtica (anche d'oltralpe), in accordo alle evidenze di carattere storico e archeologico attualmente note.



**Figura 6.14** Modelli di *admixture 2-way* ottenuti dalla combinazione delle componenti Italy\_Broion\_BA ed England\_EBA. A lato di ciascun modello è riportato il p-value ad esso associato.

Allo scopo di esplorare tutti gli scenari possibili, per il campione AZZA006 è stato testato anche un modello di *admixture* tra Bologna\_Etruscan e Bologna\_Etruscan.Ceu. Il test risulta statisticamente significativo (p-value = 0.39, 40% Bologna\_Etruscan e 60% Bologna\_Etruscan.Ceu) ma il modello regge, seppur con p-value limite (0.05), anche considerando unicamente Bologna\_Etruscan.Ceu (*1-way*) e spostando il gruppo etrusco principale tra le *right populations*. Probabilmente l'ottenimento di un dato più consistente per AZZA006 (attraverso un sequenziamento più profondo, per esempio) consentirebbe di chiarire e meglio interpretare la posizione di questo campione dei confronti dei due cluster dell'età del Ferro.

Analogamente a quanto effettuato per i campioni dell'età del Ferro, test di *admixture 2-way* sono stati condotti anche per i campioni imperiali, sostituendo Italy\_Broion\_BA con Bologna\_Etruscan (rappresentativo del *pool* genico esistente nell'area nel periodo precedente a quello imperiale).

Per quanto riguarda Bologna\_Imperial si osservano chiari segnali di *admixture* con componenti di provenienza caucasica (caratteristiche delle regioni dell'Armenia e dell'Iran). A differenza di quanto osservato per i campioni imperiali del centro Italia negli studi di Antonio *et al.* e Posth *et al.* [115, 117], sembrano mancare nell'area padana evidenze di *admixture* con popolazioni caratterizzate da componenti anatoliche (Turkey\_IA\_o3.SG si colloca infatti all'interno della variabilità armena [103]) e, soprattutto levantine, come anticipato dalle statistiche  $F_4$  illustrate nel precedente paragrafo.

L'assenza di uno *shift* verso la variabilità genetica tipica di popolazioni del Mediterraneo orientale andrà ulteriormente indagata in futuro per identificare possibili spiegazioni alla base della diversità osservata (già a partire dalla PCA) tra la popolazione imperiale padana e quella del centro Italia. È importante sottolineare nuovamente che il limitato numero di campioni imperiali, tutti provenienti

dallo stesso sito, potrebbe non offrire un quadro completo della variabilità genetica esistente nella regione padana durante quel periodo.

Infine, l'*outlier* CAS054 può essere modellato con 70% di *ancestry* associata a Bologna\_Etruscan in combinazione a 30% di *ancestry* rappresentata da un individuo dell'età del Ferro proveniente dalla città di Kerkouane (Tunisia), caratterizzato da una componente genetica subsahariana [110].

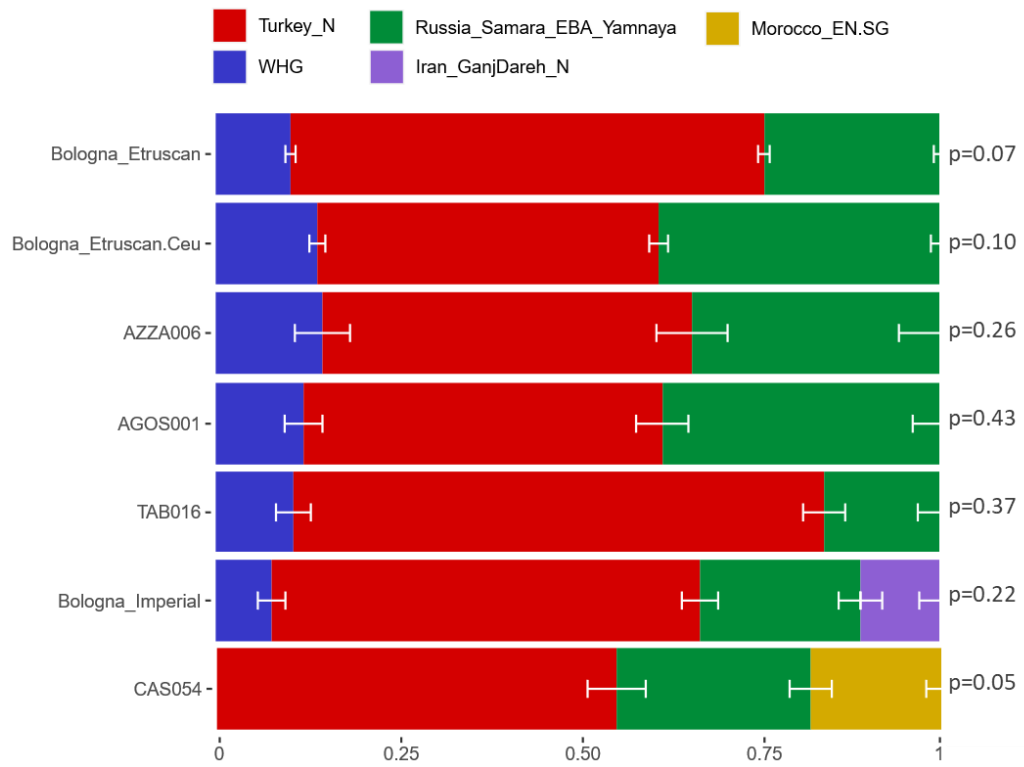
Il ritrovamento di tale componente può essere correlato all'intensa rete commerciale sviluppata dall'Impero romano che si estendeva ben oltre i suoi confini, spingendosi verso il nord dell'Europa, l'Africa subsahariana, l'India e, in generale, il continente asiatico [115].

L'ultima analisi riportata in questo paragrafo, che conferma e sintetizza i risultati finora esposti, ha permesso di modellare i campioni/clusters come derivanti dalla combinazione delle 5 componenti distali riportate in *Figura 6.15*, dove si possono osservare i modelli ottenuti per ciascuno.

Le tre componenti Turkey\_N (che equivale ad Anatolia\_N), WHG e Yamnaya sono sufficienti per descrivere il profilo genetico degli individui dell'età del Ferro, come anticipato da ADMIXTURE (vedi *Paragrafo 6.5*); la componente associata alla steppa è minima nell'individuo TAB016 mentre risulta maggiore in Bologna\_Etruscan e ancora di più in AZZA006, Bologna\_Etruscan.Ceu e AGOS001.

Sebbene il profilo ottenuto per AGOS001 risulti molto simile a quello di Bologna\_Etruscan.Ceu, per questo campione è stato possibile ottenere un modello funzionante solo escludendo il gruppo EHG dal set di *right populations*, probabilmente perché troppo strettamente relazionato all'*ancestry* che caratterizza AGOS001. È importante sottolineare, inoltre, che i modelli ottenuti per tutti i campioni dell'età del Ferro reggono anche nel momento in cui la componente Iran\_N viene spostata nel set di popolazioni di riferimento, a conferma del fatto che tale *ancestry*, come riscontrato anche nei campioni dell'Etruria classica, non è significativamente presente nell'area padana prima del periodo imperiale.

Come atteso, invece, un contributo di componente Iran\_N è necessario per modellare i campioni del cluster imperiale, mentre l'individuo CAS054 necessita di un'*ancestry* africana rappresentata da Morocco\_EN.



**Figura 6.15** Modelli di *admixture* distali ottenuti con qpAdm. A lato di ciascun modello lato è indicato il p-value ad esso associato.

## 7. Conclusioni

Il presente studio ha riguardato la caratterizzazione dal punto di vista genomico di oltre 50 individui antichi associati alla cultura etrusca provenienti da necropoli situate nella città di Bologna e dintorni e datate principalmente tra l'VIII ed il V secolo a.C.

Sebbene precedenti studi [115, 117] abbiano fornito evidenze paleogenomiche circa l'origine della popolazione etrusca e abbiano contribuito a delinearne il profilo genetico, le analisi condotte hanno riguardato in maniera esclusiva resti umani antichi riferibili ai territori dell'Etruria classica (Toscana e Lazio settentrionale). L'estensione dell'analisi genomica alle necropoli della regione padana, e in particolare a quelle associate all'antica città di Felsina, ha permesso di rilevare per la prima volta i connotati della struttura genetica che caratterizzava la popolazione di una delle principali aree di espansione etrusca già a partire dalla prima età del Ferro.

I risultati ottenuti dimostrano che la popolazione di Felsina e delle aree limitrofe si caratterizza in prevalenza per un profilo genetico locale, che si sovrappone in linea generale ai segnali già osservati per la popolazione etrusca dell'Etruria classica e, complessivamente, per il contesto italiano durante l'età del Bronzo e l'età del Ferro. Evidenze di chiara affinità genetica tra i gruppi etruschi dell'area padana e della zona tirrenica emergono quando le analisi popolazionistiche coinvolgono solamente i campioni che dispongono di una datazione al radiocarbonio a conferma della loro appartenenza al periodo trattato. Datazioni dirette aggiuntive potrebbero consentire di rilevare possibili *patterns* temporali alla base delle sottili differenze tra il cluster locale padano e quello dell'Etruria tirrenica osservate nel momento in cui le analisi non vengono ristrette ai soli campioni datati.

Accanto ad un *pool* genetico prevalentemente omogeneo e duraturo, i dati genomici disponibili in letteratura sulla penisola italiana durante l'età del Ferro mostrano anche un incremento generale, rispetto al periodo precedente, della variabilità genetica della popolazione, che si manifesta con la comparsa occasionale di *ancestries* di origine non locale. Numerose fonti storiche e testimonianze archeologiche documentano intensi contatti tra l'Etruria padana e i popoli adiacenti, tra cui i Veneti e, in particolare, i gruppi celtici dell'area golasecchiana (e d'oltralpe) già a partire dalla fase Hallstatt che precede il periodo delle ben note migrazioni storiche del IV-III secolo a.C. (fase La Tène).

L'analisi genetica ha consentito di indagare la natura delle dinamiche relazionali esistenti tra gruppi etruschi e celtici, nonché l'impatto a livello biologico di tali contatti, finora concretamente attestati dalla sola presenza di evidenze di cultura materiale non locale (come, ad esempio, corredi funebri con caratteri hallstattiani).

I dati genomici confermano la presenza di un significativo numero di individui con profilo genetico non locale compatibile con l'arrivo in Etruria padana di componenti di derivazione nordica associate a popolazioni verosimilmente del centro e nord Europa. Le informazioni di carattere archeologico disponibili per alcuni degli individui con questo profilo genetico attestano la presenza nelle relative sepolture di elementi di corredo contraddistinti da influssi culturali non locali, in particolare tipici dell'area golasecchiana e renano-ticinese (oltre che isontina).

Rispetto a quanto osservato precedentemente per i siti dell'Etruria classica [117], a Felsina il gruppo di individui caratterizzati da quest'*ancestry* non locale risulta più eterogeneo e numeroso (17% degli individui riferibili all'età del Ferro contro il 6% riscontrato in Italia centrale), probabilmente come conseguenza della posizione strategica che garantiva ai centri urbani padani contatti e scambi più agevoli con le popolazioni celtiche stanziato più a nord e oltre i confini delle Alpi. Inoltre, la presenza di questi individui è attestata in tutte le necropoli esaminate e copre un intervallo temporale che si estende dall'VIII al V-III secolo a.C., periodo a cui è datato ( $^{14}\text{C}$ ) l'unico campione dell'età del Ferro riferibile ad una fase più recente (MAR012). Questo suggerisce una certa continuità nell'arrivo o comunque nella diffusione di questa componente alloctona nei centri urbani dell'area bolognese.

Sebbene le indagini condotte finora non abbiano permesso di individuare in modo specifico la fonte primaria dell'*ancestry* riscontrata, lo stretto legame che traspare con gruppi dell'Europa centrale e settentrionale datati all'età del Bronzo e del Ferro, suggerito anche da una particolare affinità alla componente della steppa (rispetto a quanto osservato per il *pool* genetico etrusco locale), è in linea con le evidenze materiali di contatti diretti tra l'Etruria padana e popolazioni di estrazione celtica.

Gli individui etruschi recentemente aggiunti al dataset di analisi, provenienti dai siti di Monterenzio Vecchio e Monte Bibele e datati tra il V e il II secolo a.C., ricadono all'interno del periodo delle grandi migrazioni storiche e derivano da necropoli in cui la presenza celtica è attestata dai corredi funebri e supportata (per Monterenzio) dallo studio dei tratti non metrici dentali [155]. A livello genetico, nonostante la peculiarità del periodo considerato, non si evidenzia un'elevata frequenza di individui caratterizzati da componenti genetiche non locali. Nonostante si tratti di indagini preliminari e dal carattere descrittivo (PCA), i segnali attualmente emersi sono nella maggioranza dei casi concordi a quanto riscontrato da precedenti lavori incentrati sullo studio della mobilità individuale a partire dall'esame degli isotopi dello stronzio [155, 157].

Per quanto riguarda i campioni dell'età imperiale provenienti dalla necropoli di Casalecchio di Reno, si osserva, in generale, una maggiore affinità alla componente Iran\_N rispetto a quanto rilevato nel corso della precedente periodo e si registrano evidenze di *admixture* che coinvolgono componenti tipiche della regione caucasica (nello specifico dell'attuale Iran ed Armenia). Per quanto si tratti di caratteristiche già riscontrate in altre aree della penisola italiana, derivanti da cambiamenti genetici

correlati ad una maggiore interconnessione e mobilità umana e allo sviluppo da parte dell'Impero di una fitta rete di relazioni commerciali, il piccolo cluster padano si differenzia dall'eterogeneo *pool* genetico dell'Italia centrale per l'assenza di componenti associate al Mediterraneo orientale e, nello specifico, alle regioni levantine. Dato il ridotto numero di individui imperiali esaminati nel presente studio, la significatività della tendenza appena osservata potrà essere verificata solo in presenza di un campione statisticamente più rilevante.

Infine, tra i campioni imperiali compare un individuo (CAS054) contraddistinto da una componente di derivazione africana, probabilmente subsahariana; questa ipotesi è rafforzata dall'identificazione per questo campione di un aplotipo mitocondriale che rientra all'interno del clade L3e1.

Oltre ai risultati più strettamente popolazionistici, l'identificazione del sesso e la determinazione dei legami di parentela ha permesso di ottenere informazioni circa il profilo biologico degli individui in esame e la composizione della struttura sociale. Sebbene all'interno del primo set di campioni non siano stati rilevati frequenti rapporti di parentela diretta, l'analisi delle *runs of homozygosity* ha mostrato per alcuni individui delle necropoli di ex Manifattura Tabacchi e Piazza Azzarita evidenze di stretti rapporti di consanguineità tra i genitori, che supportano l'ipotesi della presenza di un'élite aristocratica (la cui esistenza a Felsina è attestata da testimonianze archeologiche) o, comunque, di una comunità ristretta.

Tra gli individui della necropoli di Monte Bibele, invece, sono stati identificati numerosi rapporti di parentela: 6 individui su 15 sono caratterizzati infatti da legami biologici fino al secondo grado.

Come già accennato, lo studio paleogenomico sul dataset presentato in questo elaborato è ancora in corso. Ulteriori strategie di analisi verranno messe a punto per gli individui del primo set al fine di affrontare questioni al momento irrisolte che necessitano sia di maggiore approfondimento dal punto di vista genomico che di un riferimento temporale più accurato (come nel caso dell'individuo TAB016). Sarà di particolare interesse la ricerca di un approccio di analisi utile ad identificare con maggiore risoluzione la regione di provenienza delle componenti associate all'area transalpina; ciò permetterebbe, inoltre, di effettuare una datazione molecolare mirata del processo di *admixture* da cui verosimilmente deriva il profilo genetico riscontrato in Bologna\_Etruscan.Ceu e in AZZA006, andando ad utilizzare come popolazioni sorgente i gruppi identificati come migliore *proxy*.

In aggiunta, le analisi effettuate fino ad ora sul primo set di campioni verranno condotte anche sul secondo gruppo, possibilmente incrementando il dato genetico per gli individui che mostrano una copertura buona ma non ottimale ai fini delle indagini popolazionistiche.

Complessivamente, i risultati finora ottenuti evidenziano le potenzialità del DNA antico nel fornire informazioni sia di carattere individuale che popolazionistico nell'ambito di contesti antichi dove il



dato genetico ricopre un ruolo di fondamentale importanza nell'arricchire e completare il quadro generale offerto dalle evidenze storiche, archeologiche ed antropologiche.

In particolare, l'analisi genomica dell'intero dataset descritto in questo studio potrà contribuire ad ampliare le conoscenze disponibili sulla storia di Bologna durante la fase di dominazione etrusca e ad approfondire le questioni relative ai rapporti esistenti tra la comunità etrusca e le popolazioni di cultura celtica sia durante la prima età del Ferro che nel corso del periodo lateniano.

## Riferimenti

1. Higuchi, R., et al., *DNA sequences from the quagga, an extinct member of the horse family*. Nature, 1984. **312**(5991): p. 282-284.
2. Meyer, M., et al., *Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins*. Nature, 2016. **531**(7595): p. 504-507.
3. van der Valk, T., et al., *Million-year-old DNA sheds light on the genomic history of mammoths*. Nature, 2021. **591**(7849): p. 265-269.
4. Palkopoulou, E., et al., *Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth*. Current Biology, 2015. **25**(10): p. 1395-1400.
5. Barlow, A., et al., *Partial genomic survival of cave bears in living brown bears*. Nat Ecol Evol, 2018. **2**(10): p. 1563-1570.
6. Green, R.E., et al., *A draft sequence of the Neandertal genome*. Science, 2010. **328**(5979): p. 710-722.
7. Mafessoni, F., et al., *A high-coverage Neandertal genome from Chagyrskaya Cave*. Proc Natl Acad Sci U S A, 2020. **117**(26): p. 15132-15136.
8. Prüfer, K., et al., *A high-coverage Neandertal genome from Vindija Cave in Croatia*. Science, 2017. **358**(6363): p. 655-658.
9. Prüfer, K., et al., *The complete genome sequence of a Neanderthal from the Altai Mountains*. Nature, 2014. **505**(7481): p. 43-49.
10. Meyer, M., et al., *A high-coverage genome sequence from an archaic Denisovan individual*. Science, 2012. **338**(6104): p. 222-6.
11. Reich, D., et al., *Genetic history of an archaic hominin group from Denisova Cave in Siberia*. Nature, 2010. **468**(7327): p. 1053-1060.
12. Dabney, J., M. Meyer, and S. Pääbo, *Ancient DNA damage*. Cold Spring Harb Perspect Biol, 2013. **5**(7).
13. Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction*. Cold Spring Harbor symposia on quantitative biology, 1986. **51 Pt 1**: p. 263-273.
14. Pääbo, S., *Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification*. Proc Natl Acad Sci U S A, 1989. **86**(6): p. 1939-43.
15. Orlando, L., et al., *Ancient DNA analysis*. Nature Reviews Methods Primers, 2021. **1**(1): p. 14.
16. Ávila-Arcos, M.C., et al., *Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA*. Scientific Reports, 2011. **1**(1): p. 74.
17. Fu, Q., et al., *DNA analysis of an early modern human from Tianyuan Cave, China*. Proceedings of the National Academy of Sciences, 2013. **110**(6): p. 2223-2227.
18. Maricic, T., M. Whitten, and S. Pääbo, *Multiplexed DNA sequence capture of mitochondrial genomes using PCR products*. PLoS One, 2010. **5**(11): p. e14004.
19. Carpenter, M.L., et al., *Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries*. Am J Hum Genet, 2013. **93**(5): p. 852-64.
20. Fu, Q., et al., *An early modern human from Romania with a recent Neanderthal ancestor*. Nature, 2015. **524**(7564): p. 216-219.
21. Boessenkool, S., et al., *Combining bleach and mild predigestion improves ancient DNA recovery from bones*. Molecular Ecology Resources, 2017. **17**(4): p. 742-751.
22. Dabney, J., et al., *Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments*. Proceedings of the National Academy of Sciences, 2013. **110**(39): p. 15758-15763.

23. Gamba, C., et al., *Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing*. *Molecular Ecology Resources*, 2016. **16**(2): p. 459-469.
24. Korlević, P., et al., *Reducing microbial and human contamination in DNA extractions from ancient bones and teeth*. *BioTechniques*, 2015. **59**(2): p. 87-93.
25. Gansauge, M.-T. and M. Meyer, *Selective enrichment of damaged DNA molecules for ancient genome sequencing*. *Genome research*, 2014. **24**(9): p. 1543-9.
26. Rohland, N., et al., *Partial uracil–DNA–glycosylase treatment for screening of ancient DNA*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2015. **370**(1660): p. 20130624.
27. Damgaard, P.B., et al., *Improving access to endogenous DNA in ancient bones and teeth*. *Scientific Reports*, 2015. **5**(1): p. 11184.
28. Gamba, C., et al., *Genome flux and stasis in a five millennium transect of European prehistory*. *Nature Communications*, 2014. **5**(1): p. 5257.
29. Parker, C., et al., *A systematic investigation of human DNA preservation in medieval skeletons*. *Scientific Reports*, 2020. **10**(1): p. 18225.
30. Sirak, K., et al., *Human auditory ossicles as an alternative optimal source of ancient DNA*. *Genome Res*, 2020. **30**(3): p. 427-436.
31. Krause, J., et al., *A complete mtDNA genome of an early modern human from Kostenki, Russia*. *Current biology : CB*, 2010. **20**(3): p. 231-236.
32. Renaud, G., et al., *Authentication and Assessment of Contamination in Ancient DNA*, in *Ancient DNA: Methods and Protocols*, B. Shapiro, et al., Editors. 2019, Springer New York: New York, NY. p. 163-194.
33. Skoglund, P., et al., *Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal*. *Proc Natl Acad Sci U S A*, 2014. **111**(6): p. 2229-34.
34. Jónsson, H., et al., *mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters*. *Bioinformatics*, 2013. **29**(13): p. 1682-1684.
35. Leonardi, M., et al., *Evolutionary Patterns and Processes: Lessons from Ancient DNA*. *Systematic Biology*, 2016. **66**(1): p. e1-e29.
36. Gansauge, M.-T., et al., *Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA*. *Nature Protocols*, 2020. **15**(8): p. 2279-2300.
37. Slon, V., et al., *Neandertal and Denisovan DNA from Pleistocene sediments*. *Science*, 2017. **356**(6338): p. 605-608.
38. Cassidy, L.M., et al., *A dynastic elite in monumental Neolithic society*. *Nature*, 2020. **582**(7812): p. 384-388.
39. Sirak, K.A., et al., *A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis*. *BioTechniques*, 2017. **62**(6): p. 283-289.
40. Harney, É., et al., *A minimally destructive protocol for DNA extraction from ancient teeth*. *Genome Res*, 2021. **31**(3): p. 472-483.
41. Modi, A., et al., *Combined methodologies for gaining much information from ancient dental calculus: testing experimental strategies for simultaneously analysing DNA and food residues*. *Archaeological and Anthropological Sciences*, 2020. **12**(1): p. 10.
42. Liu, Y., et al., *Insights into human history from the first decade of ancient human genomics*. *Science*, 2021. **373**(6562): p. 1479-1484.
43. Olalde, I. and C. Posth, *Latest trends in archaeogenetic research of west Eurasians*. *Current Opinion in Genetics & Development*, 2020. **62**: p. 36-43.
44. 09/11/2022; Available from: <https://www.nobelprize.org/prizes/medicine/2022/press-release/>.
45. Sawyer, S., et al., *Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA*. *PLOS ONE*, 2012. **7**(3): p. e34131.
46. Hofreiter, M., et al., *Ancient DNA*. *Nat Rev Genet*, 2001. **2**(5): p. 353-9.

47. Brotherton, P., et al., *Novel high-resolution characterization of ancient DNA reveals C &gt; U-type base modification events as the sole cause of post mortem miscoding lesions*. *Nucleic Acids Research*, 2007. **35**(17): p. 5717-5728.
48. Briggs, A.W., et al., *Patterns of damage in genomic DNA sequences from a Neandertal*. *Proceedings of the National Academy of Sciences*, 2007. **104**(37): p. 14616-14621.
49. Cooper, A. and H.N. Poinar, *Ancient DNA: do it right or not at all*. *Science*, 2000. **289**(5482): p. 1139.
50. Pilli, E., et al., *Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains*. *PLoS One*, 2013. **8**(1): p. e52524.
51. Spyrou, M.A., et al., *Ancient pathogen genomics as an emerging tool for infectious disease research*. *Nature Reviews Genetics*, 2019. **20**(6): p. 323-340.
52. Warinner, C., et al., *A Robust Framework for Microbial Archaeology*. *Annu Rev Genomics Hum Genet*, 2017. **18**: p. 321-356.
53. Nielsen, R., et al., *Tracing the peopling of the world through genomics*. *Nature*, 2017. **541**(7637): p. 302-310.
54. Skov, L., et al., *Genetic insights into the social organization of Neanderthals*. *Nature*, 2022. **610**(7932): p. 519-525.
55. Slon, V., et al., *The genome of the offspring of a Neanderthal mother and a Denisovan father*. *Nature*, 2018. **561**(7721): p. 113-116.
56. Seguin-Orlando, A., et al., *Genomic structure in Europeans dating back at least 36,200 years*. *Science*, 2014. **346**(6213): p. 1113-1118.
57. Fu, Q., et al., *Genome sequence of a 45,000-year-old modern human from western Siberia*. *Nature*, 2014. **514**(7523): p. 445-449.
58. Sikora, M., et al., *Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers*. *Science*, 2017. **358**(6363): p. 659-662.
59. Yang, M.A., et al., *40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia*. *Current Biology*, 2017. **27**(20): p. 3202-3208.e9.
60. Hajdinjak, M., et al., *Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry*. *Nature*, 2021. **592**(7853): p. 253-257.
61. Prüfer, K., et al., *A genome sequence from a modern human skull over 45,000 years old from Zlatý kůň in Czechia*. *Nature Ecology & Evolution*, 2021. **5**(6): p. 820-825.
62. Wang, K., et al., *Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa*. *Sci Adv*, 2020. **6**(24): p. eaaz0183.
63. Lipson, M., et al., *Ancient DNA and deep population structure in sub-Saharan African foragers*. *Nature*, 2022. **603**(7900): p. 290-296.
64. Willerslev, E. and D.J. Meltzer, *Peopling of the Americas as inferred from ancient genomics*. *Nature*, 2021. **594**(7863): p. 356-364.
65. Lipson, M., et al., *Population Turnover in Remote Oceania Shortly after Initial Settlement*. *Curr Biol*, 2018. **28**(7): p. 1157-1165.e7.
66. Yang, M.A., et al., *Ancient DNA indicates human population shifts and admixture in northern and southern China*. *Science*, 2020. **369**(6501): p. 282-288.
67. Lipson, M., et al., *Ancient genomes document multiple waves of migration in Southeast Asian prehistory*. *Science*, 2018. **361**(6397): p. 92-95.
68. McColl, H., et al., *The prehistoric peopling of Southeast Asia*. *Science*, 2018. **361**(6397): p. 88-92.
69. Fowler, C., et al., *A high-resolution picture of kinship practices in an Early Neolithic tomb*. *Nature*, 2022. **601**(7894): p. 584-587.
70. Vai, S., et al., *Kinship Determination in Archeological Contexts Through DNA Analysis*. *Frontiers in Ecology and Evolution*, 2020. **8**.
71. Yaka, R., et al., *Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes*. *Current Biology*, 2021. **31**(11): p. 2455-2468.e18.

72. Zeberg, H. and S. Pääbo, *The major genetic risk factor for severe COVID-19 is inherited from Neanderthals*. *Nature*, 2020. **587**(7835): p. 610-612.
73. Salo, W.L., et al., *Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy*. *Proceedings of the National Academy of Sciences*, 1994. **91**(6): p. 2091-2094.
74. Sabin, S., et al., *Estimating molecular preservation of the intestinal microbiome via metagenomic analyses of latrine sediments from two medieval cities*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2020. **375**(1812): p. 20190576.
75. Shillito, L.-M., et al., *The what, how and why of archaeological coprolite analysis*. *Earth-Science Reviews*, 2020. **207**: p. 103196.
76. Warinner, C., C. Speller, and M.J. Collins, *A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome*. *Philos Trans R Soc Lond B Biol Sci*, 2015. **370**(1660): p. 20130376.
77. Frantz, L.A.F., et al., *Animal domestication in the era of ancient genomics*. *Nature Reviews Genetics*, 2020. **21**(8): p. 449-460.
78. Kistler, L., et al., *Ancient Plant Genomics in Archaeology, Herbaria, and the Environment*. *Annual Review of Plant Biology*, 2020. **71**(1): p. 605-629.
79. Lorenzen, E.D., et al., *Species-specific responses of Late Quaternary megafauna to climate and humans*. *Nature*, 2011. **479**(7373): p. 359-64.
80. Fortes, G.G., et al., *Ancient DNA reveals differences in behaviour and sociality between brown bears and extinct cave bears*. *Molecular Ecology*, 2016. **25**(19): p. 4907-4918.
81. Gretzinger, J., et al., *Large-scale mitogenomic analysis of the phylogeography of the Late Pleistocene cave bear*. *Scientific Reports*, 2019. **9**(1): p. 10700.
82. Murchie, T.J., et al., *Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set*. *Quaternary Research*, 2021. **99**: p. 305-328.
83. Seersholm, F.V., et al., *Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change*. *Nature Communications*, 2020. **11**(1): p. 2770.
84. Günther, T. and M. Jakobsson, *Genes mirror migrations and cultures in prehistoric Europe—a population genomic perspective*. *Current Opinion in Genetics & Development*, 2016. **41**: p. 115-123.
85. Lazaridis, I., et al., *Ancient human genomes suggest three ancestral populations for present-day Europeans*. *Nature*, 2014. **513**(7518): p. 409-413.
86. Skoglund, P., et al., *Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe*. *Science*, 2012. **336**(6080): p. 466-9.
87. Lipson, M., et al., *Parallel palaeogenomic transects reveal complex genetic history of early European farmers*. *Nature*, 2017. **551**(7680): p. 368-372.
88. Haak, W., et al., *Massive migration from the steppe was a source for Indo-European languages in Europe*. *Nature*, 2015. **522**(7555): p. 207-211.
89. Reich, D., *Who We are and how We Got Here: Ancient DNA and the New Science of the Human Past*. 2018: Oxford University Press.
90. Lazaridis, I., et al., *Ancient DNA from Mesopotamia suggests distinct Pre-Pottery and Pottery Neolithic migrations into Anatolia*. *Science*, 2022. **377**(6609): p. 982-987.
91. Harris, E., *Demic and cultural diffusion in prehistoric Europe in the age of ancient genomes*. *Evolutionary Anthropology: Issues, News, and Reviews*, 2017. **26**: p. 228-241.
92. Bramanti, B., et al., *Genetic discontinuity between local hunter-gatherers and central Europe's first farmers*. *Science (New York, N.Y.)*, 2009. **326**(5949): p. 137-140.
93. Mathieson, I., et al., *The genomic history of southeastern Europe*. *Nature*, 2018. **555**(7695): p. 197-203.
94. Mathieson, I., et al., *Genome-wide patterns of selection in 230 ancient Eurasians*. *Nature*, 2015. **528**(7583): p. 499-503.

95. Hofmanová, Z., et al., *Early farmers from across Europe directly descended from Neolithic Aegeans*. Proceedings of the National Academy of Sciences, 2016. **113**(25): p. 6886-6891.
96. Valdiosera, C., et al., *Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia*. Proceedings of the National Academy of Sciences, 2018. **115**(13): p. 3428-3433.
97. Olalde, I., et al., *The genomic history of the Iberian Peninsula over the past 8000 years*. Science, 2019. **363**(6432): p. 1230-1234.
98. Rivollat, M., et al., *Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers*. Science Advances, 2020. **6**(22): p. eaaz5344.
99. 15/11/2022; Available from: <https://www.shh.mpg.de/1713184/haak-french-dna>.
100. Brace, S., et al., *Ancient genomes indicate population replacement in Early Neolithic Britain*. Nat Ecol Evol, 2019. **3**(5): p. 765-771.
101. Allentoft, M.E., et al., *Population genomics of Bronze Age Eurasia*. Nature, 2015. **522**(7555): p. 167-172.
102. Jones, E.R., et al., *Upper Palaeolithic genomes reveal deep roots of modern Eurasians*. Nature Communications, 2015. **6**(1): p. 8912.
103. de Barros Damgaard, P., et al., *The first horse herders and the impact of early Bronze Age steppe expansions into Asia*. Science (New York, N.Y.), 2018. **360**(6396): p. eaar7711.
104. Wang, C.-C., et al., *Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions*. Nature Communications, 2019. **10**(1): p. 590.
105. Narasimhan, V.M., et al., *The formation of human populations in South and Central Asia*. Science, 2019. **365**(6457): p. eaat7487.
106. Olalde, I., et al., *The Beaker phenomenon and the genomic transformation of northwest Europe*. Nature, 2018. **555**(7695): p. 190-196.
107. Patterson, N., et al., *Large-scale migration into Britain during the Middle to Late Bronze Age*. Nature, 2022. **601**(7894): p. 588-594.
108. Amorim, C., et al., *Understanding 6th-century barbarian social organization and migration through paleogenomics*. Nature Communications, 2018. **9**.
109. Antonio, M.L., et al., *Stable population structure in Europe since the Iron Age, despite high mobility*. bioRxiv, 2022: p. 2022.05.15.491973.
110. Moots, H.M., et al., *A Genetic History of Continuity and Mobility in the Iron Age Central Mediterranean*. bioRxiv, 2022: p. 2022.03.13.483276.
111. Martiniano, R., et al., *Genomic signals of migration and continuity in Britain before the Anglo-Saxons*. Nature Communications, 2016. **7**(1): p. 10326.
112. Schiffels, S., et al., *Iron Age and Anglo-Saxon genomes from East England reveal British migration history*. Nature Communications, 2016. **7**(1): p. 10408.
113. Brunel, S., et al., *Ancient genomes from present-day France unveil 7,000 years of its demographic history*. Proceedings of the National Academy of Sciences, 2020. **117**(23): p. 12791-12798.
114. Fischer, C.-E., et al., *Origin and mobility of Iron Age Gaulish groups in present-day France revealed through archaeogenomics*. iScience, 2022. **25**(4): p. 104094.
115. Antonio, M.L., et al., *Ancient Rome: A genetic crossroads of Europe and the Mediterranean*. Science (New York, N.Y.), 2019. **366**(6466): p. 708-714.
116. Fernandes, D.M., et al., *The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean*. Nature Ecology & Evolution, 2020. **4**(3): p. 334-345.
117. Posth, C., et al., *The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect*. Science Advances, 2021. **7**(39): p. eabi7673.
118. Aneli, S., et al., *The Genetic Origin of Daunians and the Pan-Mediterranean Southern Italian Iron Age Context*. Molecular Biology and Evolution, 2022. **39**(2).

119. Reitsema, L.J., et al., *The diverse genetic origins of a Classical period Greek army*. Proceedings of the National Academy of Sciences, 2022. **119**(41): p. e2205272119.
120. Pesando, F., *L'Italia antica. Culture e forme del popolamento nel I millennio a.C.* 2005: Carocci editore.
121. Benelli, E., *Etruschi, Breve introduzione storica*. 2021: Idea Libri.
122. Pallottino, M., *Etruscologia*. 1984: Hoepli.
123. Rendelli, M., *I "viaggi" degli Etruschi*, in *Viaggio nelle terre dei Rasna. Etruschi. Viaggio nelle terre dei Rasna. Catalogo della mostra (Bologna, 7 dicembre 2019-24 maggio 2020)*. Electa, Editor. 2019. p. 39-45.
124. Govi, E., *L'Etruria padana*, in *Etruschi. Viaggio nelle terre dei Rasna. Etruschi. Viaggio nelle terre dei Rasna. Catalogo della mostra (Bologna, 7 dicembre 2019-24 maggio 2020)*. 2019, Electa. p. 357-361.
125. Sassatelli G. and Govi E., *Etruria on the Po and the Adriatic Sea*, in *The Etruscan World*. 2013, London and New York, Routledge Taylor & Francis Group. p. 281-300.
126. Guidi F. and Marchesi M., *Bologna, metropoli dei Rasna*, in *Etruschi. Viaggio nelle terre dei Rasna. Etruschi. Viaggio nelle terre dei Rasna. Catalogo della mostra (Bologna, 7 dicembre 2019-24 maggio 2020)*. 2019, Electa. p. 377-383.
127. Sassatelli, G., *Gli Etruschi nella pianura padana*, in *Gli Etruschi fuori d'Etruria*, G. Camporeale, Editor. 2001, Arsenale: Verona. p. 168-191.
128. Sassatelli, G., *Celti ed Etruschi nell'Etruria Padana e nell'Italia settentrional*, in *Tra mondo celtico e mondo italico. La necropoli di Monte Bibele. Atti della Tavola Rotonda, Roma 1997*, D. Vitali and S. Verger, Editors. 2008: Bologna. p. 323–348.
129. Locatelli, D., *Stranieri a Felsina e forse nella pianura occidentale. Dinamiche di mobilità in Emilia nel VI secolo a.C.*, in *Mobilità geografica e mercenariato nell'Italia preromana. Atti del XX Convegno Internazionale di Studi sulla Storia e l'Archeologia dell'Etruria*. 2013, Quasar: Roma. p. 361-395.
130. Dondarini, R., *Breve storia di Bologna*. 2007, Pisa: Pacini Editore.
131. Francalacci, P., et al., *Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe*. American Journal of Physical Anthropology, 1996. **100**(4): p. 443-460.
132. Veronesi, U., et al., *A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer*. N Engl J Med, 2003. **349**(6): p. 546-53.
133. Ghirotto, S., et al., *Origins and Evolution of the Etruscans' mtDNA*. PLOS ONE, 2013. **8**(2): p. e55519.
134. Tassi, F., et al., *Genetic evidence does not support an etruscan origin in Anatolia*. American Journal of Physical Anthropology, 2013. **152**(1): p. 11-18.
135. Achilli, A., et al., *Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans*. Am J Hum Genet, 2007. **80**(4): p. 759-68.
136. Pellecchia, M., et al., *The mystery of Etruscan origins: novel clues from *Bos taurus* mitochondrial DNA*. Proceedings of the Royal Society B: Biological Sciences, 2007. **274**(1614): p. 1175-1179.
137. Brisighelli, F., et al., *The Etruscan timeline: a recent Anatolian connection*. Eur J Hum Genet, 2009. **17**(5): p. 693-6.
138. Gómez-Carballa, A., et al., *Mitogenomes from The 1000 Genome Project reveal new Near Eastern features in present-day Tuscans*. PLoS One, 2015. **10**(3): p. e0119242.
139. Pardo-Seco, J., et al., *A genome-wide study of modern-day Tuscans: revisiting Herodotus's theory on the origin of the Etruscans*. PLoS One, 2014. **9**(9): p. e105920.
140. Belle, E.M.S., et al., *Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans*. Proceedings of the National Academy of Sciences, 2006. **103**(21): p. 8012-8017.

141. Guimaraes, S., et al., *Genealogical Discontinuities among Etruscan, Medieval, and Contemporary Tuscans*. *Molecular biology and evolution*, 2009. **26**: p. 2157-66.
142. Saupe, T., et al., *Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula*. *Current Biology*, 2021. **31**(12): p. 2576-2591.e12.
143. Locatelli D. and Malnati L., *Nuovi dati sulla fase orientalizzante delle necropoli felsinee.*, in *Les necròpolis d'incineració entre l'Ebre i el Tiber (segles IX-VI a.C.): metodologia, pràctiques funeràries i societat, Monografies 14*. 2012, MAC: Barcelona. p. 321-340.
144. Gambari F.M. and Venturino M., *La medio-tarda età del Ferro (V-II secolo a.C.) nella Liguria interna*, in *Ligures Celeberrimi. La Liguria interna nella seconda età del Ferro. Atti del Congresso Internazionale. Mondovì, 26-28 Aprile 2002*. 2004: Bordighera. p. 29-48.
145. *La Fenice, archeologia e restauro s.r.l., Relazione di scavo*. 1999.
146. Buoite C., Campagnari S., and Locatelli D., *Bologna nell'età del Ferro. Lo studio delle necropoli come indice delle dinamiche evolutive degli insediamenti*, in *Incontri annuali di preistoria e protostoria. Le età del Bronzo e del Ferro in Italia: contesti protostorici in scavi urbani. Abstract book*. 2017, Istituto Italiano di Preistoria e Protostoria.
147. *La Fenice, archeologia e restauro s.r.l., Relazione di scavo*. 2004-2005.
148. Desantis, P. *Un caso di committenza funeraria a Felsina : la tomba con stele della necropoli di via Saffi*. 2014.
149. 14/12/2022; Available from: [http://www.archeobologna.beniculturali.it/bologna/necropoli\\_via\\_saffi/scavo\\_restauero.htm](http://www.archeobologna.beniculturali.it/bologna/necropoli_via_saffi/scavo_restauero.htm).
150. Brasili Gualandi P., Belcastro M.G., and Giusberti G., *I più recenti ritrovamenti scheletrici umani dell'età del Ferro nel Bolognese: Casalecchio (VII-VI sec. a.C.), Giardini Margherita (VI-IV sec. a.C.), Casteldebole (V sec. a.C.)*. *Bullettino di Paleontologia Italiana*, 1994. **85**: p. 475-492.
151. 13/12/2022; Available from: <http://www.comune.bologna.it/archeologico/percorsi/89814/id/89812/oggetto/89828/>.
152. Ortalli J., *La necropoli celtica della zona "A" di Casalecchio di Reno (Bologna). Note preliminari sullo scavo del complesso sepolcrale e dell'area di culto*, in *L'Europe celtique du Ve au IIIe siècle avant J. -C., Actes du deuxième Symposium International d'Hautvillers, 1992*. 1995: Sceaux. p. 189-238.
153. Ortalli J., *Archeologia topografica: la ricostruzione dell'ambiente e dell'insediamento antico nell'esperienza di Casalecchio di Reno*, in *XLIII Corso di Cultura sull'Arte Ravennate e Bizantina, Ravenna 1997*. 1998: Ravenna. p. 565-606.
154. Pancaldi P., et al., *La tomba 23 della necropoli celtica (IV-III sec. a.C.) di Casalecchio di Reno (BO). Analisi antropologiche*, in *Sepulture anomale. Indagini archeologiche e antropologiche dall'epoca classica al Medioevo in Emilia Romagna. Giornata di Studi (Castelfranco Emilia, 19 dicembre 2009)*, Belcastro M.G. and Ortalli J., Editors. 2010, Edizioni All'Insegna del Giglio: Borgo S. Lorenzo (FI). p. 76-78.
155. Sorrentino, R., et al., *Unravelling biocultural population structure in 4th/3rd century BC Monterenzio Vecchio (Bologna, Italy) through a comparative analysis of strontium isotopes, non-metric dental evidence, and funerary practices*. *PLOS ONE*, 2018. **13**(3): p. e0193796.
156. Vitali D., *La necropoli di Monte Tamburino a Monte Bibeale*. 2003, Bologna: Gedit.
157. Scheeres, M., et al., *Evidence for "Celtic migrations"? Strontium isotope analysis at the early La Tène (LT B) cemeteries of Nebringen (Germany) and Monte Bibeale (Italy)*. *Journal of Archaeological Science*, 2013. **40**(10): p. 3614-3625.
158. Modi A., et al., *Datazione al radiocarbonio e analisi genetica esplorativa del reperto umano perinatale del Tempio di Uni a Marzabotto (BO) in BIRTH. Archeologia dell'infanzia nell'Italia preromana*, Elisabetta G., Editor. 2021, Bononia University Press. p. 107-120.
159. Pinhasi, R., et al., *Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone*. *PLOS ONE*, 2015. **10**(6): p. e0129102.



160. Korlević, P., et al., *Reducing microbial and human contamination in DNA extractions from ancient bones and teeth*. Biotechniques, 2015. **59**(2): p. 87-93.
161. Meyer, M. and M. Kircher, *Illumina sequencing library preparation for highly multiplexed target capture and sequencing*. Cold Spring Harb Protoc, 2010. **2010**(6): p. pdb.prot5448.
162. Rohland, N., et al., *Partial uracil-DNA-glycosylase treatment for screening of ancient DNA*. Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1660): p. 20130624.
163. Peltzer, A., et al., *EAGER: efficient ancient genome reconstruction*. Genome Biology, 2016. **17**(1): p. 60.
164. Schubert, M., S. Lindgreen, and L. Orlando, *AdapterRemoval v2: rapid adapter trimming, identification, and read merging*. BMC Research Notes, 2016. **9**(1): p. 88.
165. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics, 2009. **25**(14): p. 1754-60.
166. 20/12/2022; Available from: <https://github.com/statgen/bamUtil>.
167. Korneliusen, T.S., A. Albrechtsen, and R. Nielsen, *ANGSD: Analysis of Next Generation Sequencing Data*. BMC Bioinformatics, 2014. **15**(1): p. 356.
168. Renaud, G., et al., *Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA*. Genome Biology, 2015. **16**(1): p. 224.
169. Furtwängler, A., et al., *Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis*. Sci Rep, 2018. **8**(1): p. 14075.
170. Weissensteiner, H., et al., *HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing*. Nucleic Acids Res, 2016. **44**(W1): p. W58-63.
171. van Oven, M., *PhyloTree Build 17: Growing the human mitochondrial DNA tree*. Forensic Science International: Genetics Supplement Series, 2015. **5**: p. e392-e394.
172. Ralf, A., et al., *Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data*. Molecular Biology and Evolution, 2018. **35**(5): p. 1291-1294.
173. 21/12/2022; Available from: <https://isogg.org/tree/index.html>.
174. Martiniano, R., et al., *Placing Ancient DNA Sequences into Reference Phylogenies*. Molecular Biology and Evolution, 2022. **39**(2).
175. 22/12/2022; Available from: <https://github.com/stschiff/sequenceTools>.
176. 19/12/2022; Available from: <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>.
177. 23/12/2022; Available from: <https://github.com/DReichLab/EIG>.
178. Patterson, N., A.L. Price, and D. Reich, *Population Structure and Eigenanalysis*. PLOS Genetics, 2006. **2**(12): p. e190.
179. Novak, M., et al., *Genome-wide analysis of nearly all the victims of a 6200 year old massacre*. PLOS ONE, 2021. **16**(3): p. e0247332.
180. Clemente, F., et al., *The genomic history of the Aegean palatial civilizations*. Cell, 2021. **184**(10): p. 2565-2586.e21.
181. Monroy Kuhn, J.M., M. Jakobsson, and T. Günther, *Estimating genetic kin relationships in prehistoric populations*. PLOS ONE, 2018. **13**(4): p. e0195491.
182. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses*. Am J Hum Genet, 2007. **81**(3): p. 559-75.
183. Kennett, D.J., et al. *Archaeogenomic evidence reveals prehistoric matrilineal dynasty*. Nature communications, 2017. **8**, 14115 DOI: 10.1038/ncomms14115.
184. Ringbauer, H., J. Novembre, and M. Steinrücken, *Parental relatedness through time revealed by runs of homozygosity in ancient DNA*. Nature Communications, 2021. **12**(1): p. 5425.
185. Auton, A., et al., *A global reference for human genetic variation*. Nature, 2015. **526**(7571): p. 68-74.
186. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals*. Genome Res, 2009. **19**(9): p. 1655-64.

187. Marcus, J.H., et al. *Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia*. Nature communications, 2020. **11**, 939 DOI: 10.1038/s41467-020-14523-6.
188. Patterson, N., et al., *Ancient admixture in human history*. Genetics, 2012. **192**(3): p. 1065-1093.
189. 27/12/2022; Available from: <https://github.com/DReichLab/AdmixTools>.
190. Harney, É., et al., *Assessing the performance of qpAdm: a statistical tool for studying population admixture*. Genetics, 2021. **217**(4).
191. Skoglund, P., et al., *Reconstructing Prehistoric African Population Structure*. Cell, 2017. **171**(1): p. 59-71.e21.
192. Schlebusch, C.M., et al., *Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago*. Science, 2017. **358**(6363): p. 652-655.
193. Prendergast, M.E., et al., *Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa*. Science, 2019. **365**(6448).
194. Keller, A., et al., *New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing*. Nature Communications, 2012. **3**(1): p. 698.
195. Skoglund, P., et al., *Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**.
196. Torroni, A., et al., *Classification of European mtDNAs from an analysis of three European populations*. Genetics, 1996. **144**(4): p. 1835-50.
197. Salas, A., et al., *The making of the African mtDNA landscape*. Am J Hum Genet, 2002. **71**(5): p. 1082-111.
198. Bandelt, H.J., et al., *Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade*. Annals of Human Genetics, 2001. **65**(6): p. 549-563.
199. Emery, M., et al., *Ancient Roman mitochondrial genomes and isotopes reveal relationships and geographic origins at the local and pan-Mediterranean scales*. Journal of Archaeological Science: Reports, 2018. **20**: p. 200-209.
200. Kivisild, T., *The study of human Y chromosome variation through ancient DNA*. Hum Genet, 2017. **136**(5): p. 529-546.
201. Szécsényi-Nagy, A., et al., *Tracing the genetic origin of Europe's first farmers reveals insights into their social organization*. Proc Biol Sci, 2015. **282**(1805).
202. Chiang, C.W.K., et al., *Genomic history of the Sardinian population*. Nature Genetics, 2018. **50**(10): p. 1426-1434.
203. Lazaridis, I., et al., *Genomic insights into the origin of farming in the ancient Near East*. Nature, 2016. **536**(7617): p. 419-424.
204. Mittnik, A., et al., *The genetic prehistory of the Baltic Sea region*. Nature Communications, 2018. **9**(1): p. 442.
205. de Barros Damgaard, P., et al., *The first horse herders and the impact of early Bronze Age steppe expansions into Asia*. Science, 2018. **360**(6396): p. eaar7711.
206. Lazaridis, I., et al., *Genetic origins of the Minoans and Mycenaeans*. Nature, 2017. **548**(7666): p. 214-218.
207. Feldman, M., et al., *Ancient DNA sheds light on the genetic origins of early Iron Age Philistines*. Science Advances, 2019. **5**(7): p. eaax0061.
208. Haber, M., et al., *Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences*. Am J Hum Genet, 2017. **101**(2): p. 274-282.
209. Agranat-Tamir, L., et al., *The Genomic History of the Bronze Age Southern Levant*. Cell, 2020. **181**(5): p. 1146-1157.e11.

210. Zalloua, P., et al., *Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza*. Sci Rep, 2018. **8**(1): p. 17567.