J. Biol. Res. 2012 n.1 - Vol. LXXXV - Rubbettino - Soveria Mannelli

Genomic Selection by Oligo-Capture and Next Generation Sequencing for Genetic Characterization of Ancient Human Remains from Italy

M. Lari[1], E. Rizzi[2], C. Balsamo[1], S. Ghirotto[3], F. Tassi[3], A. Ronchitelli[4], A. Fischetti[5], V. Grugni[6], O. Semino[6], G. De Bellis[2], G. Barbujani[3], D. Caramelli[1]

[1] Dipartimento di Biologia Evoluzionistica, Laboratori di Antropologia, Università di Firenze, Via del Proconsolo 12, 50122 Firenze, Italy. E-mail: martina.lari@unifi.it
[2] Istituto di Tecnologie Biomediche (ITB), CNR, Via Fantoli 16/15, 20138 Milano, Italy.
[3] Dipartimento di Biologia ed Evoluzione Via Borsari 46, Università di Ferrara, 44100 Ferrara, Italy
[4] Dipartimento di Scienze Ambientali "G. Sarfatti", UR: Ecologia Preistorica, Università di Siena, Via T. Pendola 62, 53100, Siena, Italy
[5] TELTEC srl, via Lecco 4, 20041 Agrate Brianza (MI), Italy
[6] Dipartimento di Biologia e Biotecnologie, Laboratori di Genetica e Microbiologia, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy

KEY WORDS: ancient DNA authentication, target selection, next generation sequencing, human evolution.

Introduction

In the last two decades, it has become clear that DNA survival is not unlimited and that, even under ideal conditions such as a cold and dry environment, DNA does not survive for more than a million years in fossil samples (Paabo *et al.*, 2004). Strict guidelines were established to ensure the authenticity of obtained results. Nevertheless, ancient DNA research remained limited due to several reasons: first, the small amount of available material, second, the highly degraded and fragmented nature of the ancient DNA, and third, problems of contamination with contemporary DNA. Only recently, when new technologies and methodological approaches became available, the field of ancient DNA started experiencing a new era. Studies once thought to be impossible are now feasible: large-scale population genetic surveys (Shapiro *et al.*, 2004), functional analysis of extinct genes (Campbell *et al.*, 2004), whole genome sequencing of extinct organisms (Green *et al.*, 2010; Rasmussen *et al.*, 2010) and genetic investigations on early modern humans (Krause *et al.*, 2010). High throughput sequencing can also help overcoming the contamination problem in ancient DNA studies and eliminating possible doubts on the authenticity of the obtained results as often occurred (Handt *et al.*, 2004a; Paabo *et al.*, 2004; Krause *et al.*, 2010). Only well-preserved samples from the permafrost such as the ice man (Handt *et al.*, 2004b; Ermini *et al.*, 2008) or Eskimos from Greenland (Gilbert *et al.*, 2008; Rasmussen *et al.*, 2010) have enough DNA to minimize the risk of contamination which could alter sequencing results. Nevertheless this restriction limits research to exceptional cases and circumvents the usage of thousands of modern human remains. It would therefore be quite useful to consider other properties of the DNA that would allow distinguishing ancient endogenous molecules from modern human exogenous contamination. In this regard an important contribution has been provided by Briggs *et al.* (2007) who for the first time were able to investigate the DNA sequence context around strand breaks in ancient DNA by using the 454 sequencing technology. This analysis has not been previously possible because, using PCR technique, primers that target particular DNA sequences are generally used, and thus the ends of the ancient DNA molecules are not revealed. Therefore, the 454 sequencing process is of crucial importance for these analyses. The high frequency of C to T misincorporations at the 5'-ends of ancient DNA sequences and the correspondingly high frequency of G to A misincorporations at the 3'-ends imply that deamination of cytosine residues is significantly elevated at the 5'-ends of ancient DNA molecules (Briggs *et al.*, 2007). These changes could be caused either by a tendency of cytosine residues at the ends of molecules to undergo deamination or by a tendency of strand breaks to occur near deaminated cytosine residues. The authors proposed that this pattern consists in a preferential fragmentation of ancient DNA at purine bases, while modern human contaminants show no increase in frequency of purines or pyrimidines on either side of the fragment (Fig. 1), and that cytosine residues close to the ends of ancient DNA molecules are more susceptible to deamination than the internal cytosine residues. In conclusion, fragment length, deamination-induced sequence errors at the ends of molecules and purine-associated fragmentation represent features by which endogenous and contaminating populations of DNA molecules can be distinguished in at least some late Pleistocene specimens (Briggs *et al.*, 2007). The combination of ancient DNA properties and

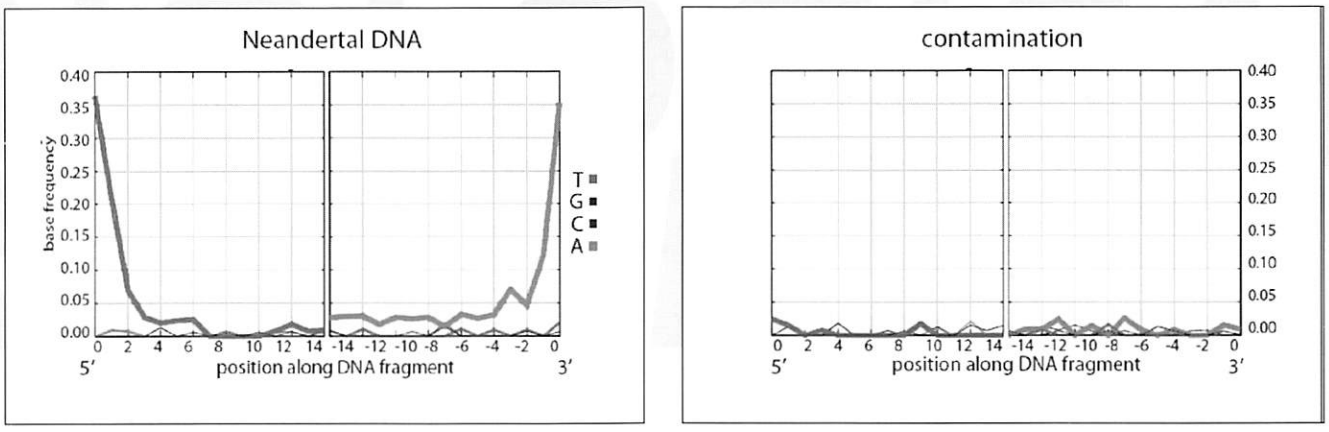

Fig. 1. *Fragmentation pattern obtained in endogenous ancient DNA sequences (Neanderthal DNA) and in modern human contaminants (from Krause et al. 2010, modified).*

identification of a single biological source to study early modern human DNA was first used on a Cro-Magnon sample from the Kostenki site in Western Russia (Krause *et al.*, 2010).

In this study we applied recently developed methodologies, namely genomic selection by oligo-capture and next generation sequencing, to the analysis of a human specimen of the Upper Palaeolithic (Paglicci of the layer 23) previously successfully characterized by PCR and standard ancient DNA procedure for the mitochondrial DNA HVR-I region and for some nuclear *loci*. Enrichment capture probes were designed as in Fig.2 and synthesized as oligonucleotides using an in-situ customized synthesis (Customarray -TM- platform).

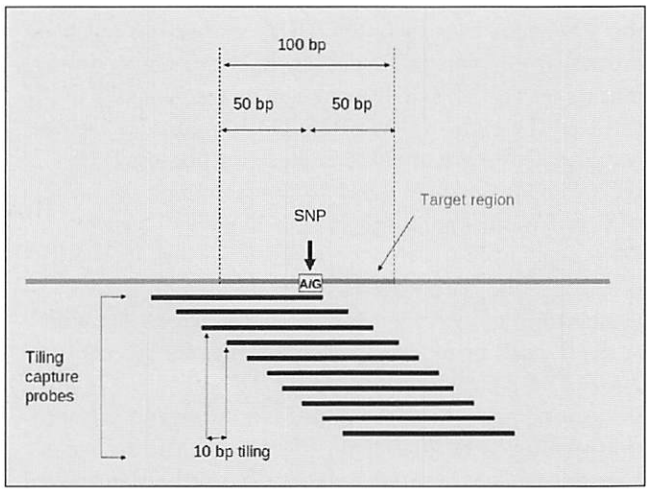

Fig. 2. *Capture probes design.*

The principal aims of this work are to: i. test probe design and synthesis by converting oligonucleotide probes in a 454 library and sequencing it; ii. perform a capture control of modern DNA target capture following the scheme summarized in Fig.3; iii. enrich and sequence libraries from the ancient DNA sample.

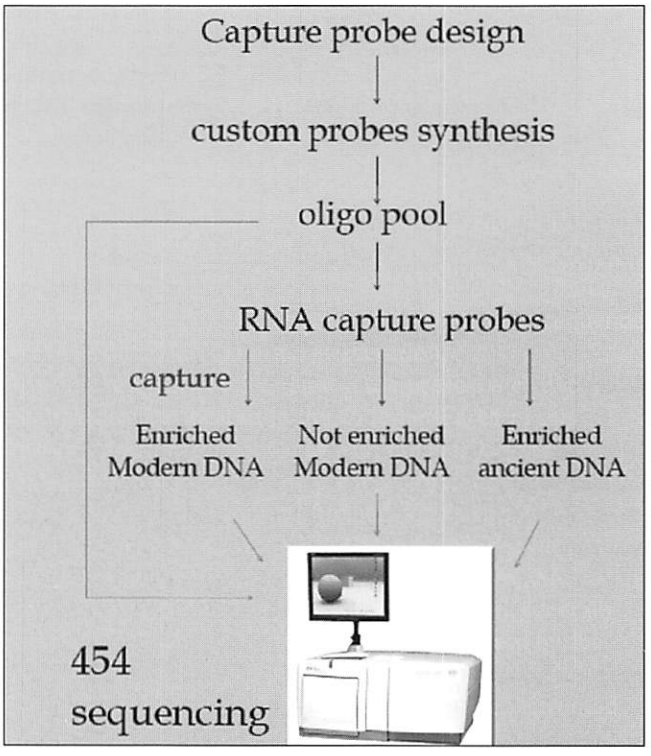

Fig. 3. *Experimental design.*

Materials and Methods

Modern DNA isolated from human blood was extracted with QIAmp DNA Blood Mini QIAcube Kit. Ancient DNA was extracted by a silica based method as described in Rohland and Hofreiter (2007). Preparation of 454 libraries was performed following the procedure described in Briggs *et al.* (2009). The capture approach consists in a "in-liquido" strategy that occurs using the hybridization of an adaptor-ligated DNA library onto RNA biotinylated probes. In liquido hybridization needs a specific probe design, hybridization of adaptor-ligated DNA library, washing of contaminant non-specific DNA and recovery of enriched sample. After capture, the RNA probes will be then digested so that the only remaining nucleotide is the targeted DNA of interest. Target region for probe design encompassed the entire human mitochondrial genome and

several nuclear loci (Tab.1 and Tab.2) in order to provide a phylogenetic classification. Genomic enrichment targeted also several genes involved in the expression of features that may have played an important role in the evolution of humans such as skin and hair pigmentation, cognitive, cardiac and respiratory function, perception of different tastes as reported in Tab. 2.

Target class	Target description	Number of target	Target length (bp)	Number of probes	Number of base captured
1	Single nucleotide (SN)	233	100	2330	46600
2	Coding region (CDS)	1	954	96	1054
3	Short Tandem Repeat (STR)	1	362	37	462
4	Complete mitochondrial genome (mtDNA)	1	16569	1657	16569
	Total	**236**	**17985**	**4120**	**64685**

Tab. 1. *Target regions for probes design.*

n° target	Target description
62	Autosomal SNPs related to evolutionary important physiological and metabolic functions (*skin and hair pigmentation, cognitive, cardiac, respiratory and muscular function, taste perception)*
33	Autosomal loci fixed and derivate in human genome but ancestral in Neanderthals
120	Y chromosome SNPs
46	X chromosome SNPs

Tab. 2. *Target selection.*

	454 PTP lanes	total reads (after nrdb)	mapping reads	% mapping reads	n° matched references	average mapping reads length (bp)	average coverage
Oligo pool	2/16	40477	36471	90.1%	220	106.66	76.36
Captured modern DNA (enriched)	1/8	114068	48782	42.7%	81	245.17	457.87
Captured modern DNA (not enriched)	1/8	111191	53199	47.8%	85	256.3	515.12

Tab. 3. *Results from oligo pool and captured modern DNA sequencing.*

Results and Discussion

Reads obtained by oligopool sequencing matched on almost all reference sequences (220 out of 236) reflecting a homogeneous composition of oligo probe mixture (Tab.3). Synthesis specificity was high, 90.1% of reads mapped on reference sequences. This value allowed performing the capture test on modern DNA. Two libraries were considered, one obtained by amplifying the 454 library (enriched in Tab. 3) with 14 PCR cycles and one using the library not enriched. Comparison of results obtained by these two modern DNA libraries allowed excluding bias in the amplification step. The obtained results are just below the expected value of mapping reads percentage (set at 50%), indicating a heterogeneous composition of captured fragments. Figs. 4 and 5 show examples of reads from oligopool and modern DNA libraries mapping on mtDNA. Modern DNA results allowed us to perform aDNA capture and sequencing, data analysis is still in progress.

The present data show the feasibility of the *in-situ* customized synthesis by Customarray (TM) platform to produce capture oligonucleotide probes. In contrast to other technologies, Customarray (TM) platform can synthesize probes up to 100 bp long; in addition, modification of the experimental design and probes re-synthesis after control experiment are fast and cheap. This versatility of the platform makes Customarray (TM) oligonucleotide probes a promising tool for capture studies on ancient samples.

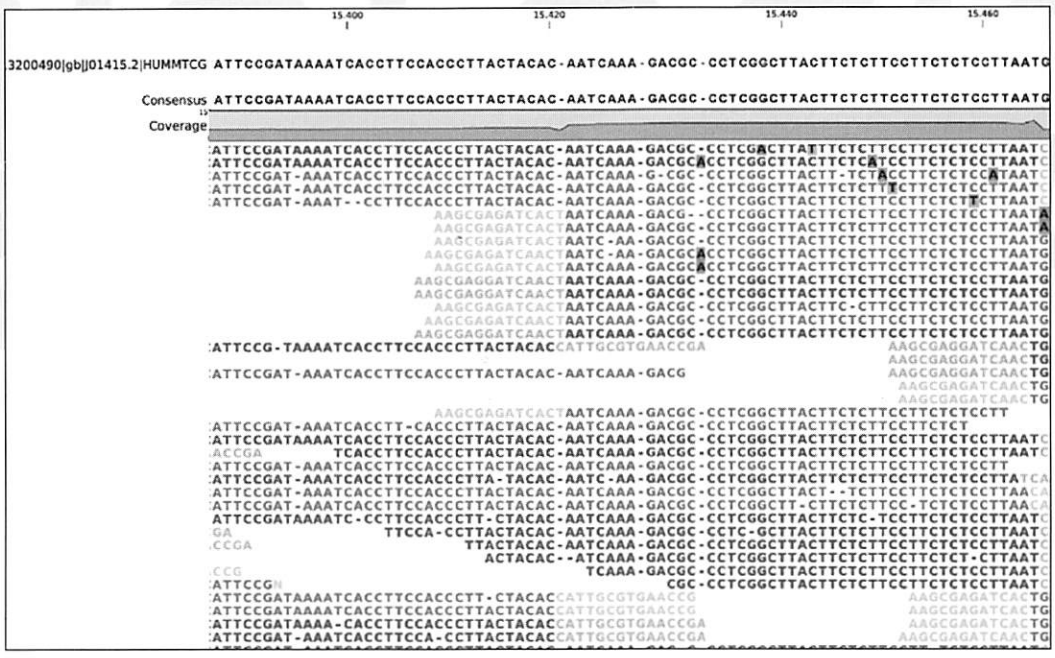

Fig. 4. *Reads from oligo pool mapping on mtDNA, an example.*

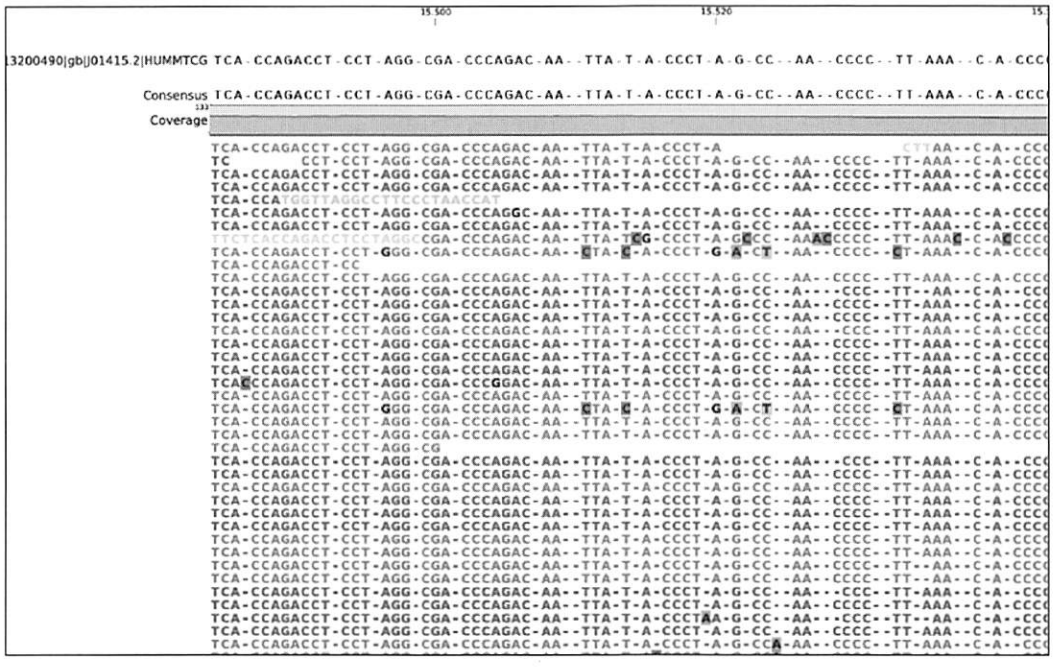

Fig. 5. *Reads from modern DNA libraries mapping on mtDNA, an example.*

Acknowledgements

This study was supported by MIUR grant n 2008TE2B8S_002 assigned to DC and FIRB grant n. RBFR08U07M "Futuro in Ricerca" assigned to ER.

References

Briggs A.W., Good J.M., Green R.E., Krause J., Maricic T., Stenzel U., Lalueza-Fox C., Rudan P., Brajkovic D., Kucan Z., Gusic I., Schmitz R., Doronichev V.B., Golovanova L.V., de la Rasilla M., Fortea J., Rosas A., Pääbo S. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science*, 325: 318-321.

Briggs A.W., Stenzel U., Johnson P.L., Green R.E., Kelso J., Prüfer K., Meyer M., Krause J., Ronan M.T., Lachmann M., Pääbo S. 2007. Patterns of damage in genomic DNA sequences from a NeandertalNeanderthal. *Proc. Natl. Acad. Sci U S A,* 104:14616-14621. Epub 2007, Aug 21.

Campbell K.L., Roberts J.E., Watson L.N., Stetefeld J., Sloan A.M., Signore A.V., Howatt J.W., Tame J.R., Rohland N., Shen T.J., Austin J.J., Hofreiter M., Ho C., Weber R.E., 2010. Cooper A. Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nat. Genet.*, 42: 536-540. Epub 2010, May 2.

Ermini L., Olivieri C., Rizzi E., Corti G., Bonnal R., Soares P., Luciani S., Marota I., De Bellis G., Richards M.B., Rollo F. 2008. Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr. Biol.*, 18: 1687-1693. Epub 2008, Oct 30.

Gilbert M.T., Kivisild T., Grønnow B., Andersen P.K., Metspalu E., Reidla M., Tamm E., Axelsson E., Götherström A., Campos P.F., Rasmussen M., Metspalu M., Higham T.F., Schwenninger J.L., Nathan R., De Hoog C.J., Koch A., Møller L.N., Andreasen C., Meldgaard M., Villems R., Bendixen C., Willerslev E. 2008. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science*, 320: 1787-1789.

Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H., Hansen N.F., Durand E.Y., Malaspinas A.S., Jensen J.D., Marques-Bonet T., Alkan C., Prüfer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Höber B., Höffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Z., Gusic I., Doronichev V.B., Golovanova L.V., Lalueza-Fox C., de la Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Pääbo S. 2010. A draft sequence of the Neandertal genome. *Science*, 328: 710-722.

Handt O., Höss M., Krings M., Pääbo S. 1994. Ancient DNA: methodological challenges. *Experientia*, 50: 524-529.

Handt O., Richards M., Trommsdorff M., Kilger C., Simanainen J., Georgiev O., Bauer K., Stone A., Hedges R., Schaffner W., Uttermann G., Sykes B., Pääbo S. *et al.*, 1994.. Molecular genetic analyses of the Tyrolean Ice Man. *Science*, 264: 1775-1778.

Krause J., Briggs A.W., Kircher M., Maricic T., Zwyns N., Derevianko A., Pääbo S. 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Curr. Biol.*, 20: 231-236. Epub. 2009, Dec. 31.

Pääbo S., Poinar H., Serre D., Jaenicke-Despres V., Hebler J., Rohland N., Kuch M., Krause J., Vigilant L., Hofreiter M. 2004. Genetic analyses from ancient DNA. *Annu. Rev. Genet.*, 38: 645-679.

Rasmussen M., Li Y., Lindgreen S., Pedersen J.S., Albrechtsen A., Moltke I., Metspalu M., Metspalu E., Kivisild T., Gupta R., Bertalan M., Nielsen K., Gilbert M.T., Wang Y., Raghavan M., Campos P.F., Kamp H.M., Wilson A.S., Gledhill A., Tridico S., Bunce M., Lorenzen E.D., Binladen J., Guo X., Zhao J., Zhang X., Zhang H., Li Z., Chen M., Orlando L., Kristiansen K., Bak M., Tommerup N., Bendixen C., Pierre T.L., Grønnow B., Meldgaard M., Andreasen C., Fedorova S.A., Osipova L.P., Higham T.F., Ramsey C.B., Hansen T.V., Nielsen F.C., Crawford M.H., Brunak S., Sicheritz-Pontén T., Villems R., Nielsen R., Krogh A., Wang J., Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463: 757-762.

Rohland N., Hofreiter M. 2007. Ancient DNA extraction from bones and teeth. *Nat. Protoc.*, 2: 1756-1762.

Shapiro B., Drummond A.J., Rambaut A., Wilson M.C., Matheus P.E., Sher A.V., Pybus O.G., Gilbert M.T., Barnes I., Binladen J., Willerslev E., Hansen A.J., Baryshnikov G.F., Burns J.A., Davydov S., Driver J.C., Froese D.G., Harington C.R., Keddie G., Kosintsev P., Kunz M.L., Martin L.D., Stephenson R.O., Storer J., Tedford R., Zimov S., Cooper A. 2004. Rise and fall of the Beringian steppe bison. *Science*, 306: 1561-1565.