Sequencing ancient RNA in bioarchaeology - Review

Gabriel Dorado [1], Fernando Luque [2], Plácido Pascual [3], Inmaculada Jiménez [4], Francisco Javier S. Sánchez-Cañete [5], Margarita Pérez-Jiménez [1], Patricia Raya [6], Jesús Sáiz [7], Adela Sánchez [7], Jesús Martín [8] Teresa E. Rosales [9], Víctor F. Vásquez [10], Pilar Hernández [11]

[1] Author for correspondence, Dep. Bioquímica y Biología Molecular, Campus Rabanales C6-1-E17, Campus de Excelencia Internacional Agroalimentario (ceiA3), Universidad de Córdoba, 14071 Córdoba (Spain), eMail: <bb1dopeg@uco.es>; [2] Laboratorio de Producción y Sanidad Animal de Córdoba, Ctra. Madrid-Cádiz km 395, 14071 Córdoba; [3] Laboratorio Agroalimentario de Córdoba, Consejería de Agricultura y Pesca, Junta de Andalucía, 14004 Córdoba; [4] IES Puertas del Campo, Avda. San Juan de Dios 1, 51001 Ceuta; [5] EE.PP. Sagrada Familia de Baena, Avda. Padre Villoslada 22, 14850 Baena (Córdoba); [6] Jefe de Servicio de Protección Radiológica, Colonia San José Nº 4 Planta 2a, Campus Rabanales, Universidad de Córdoba, 14071 Córdoba; [7] Dep. Farmacología, Toxicología y Medicina Legal y Forense, Facultad de Medicina, Avda. Menéndez Pidal, s/n, Universidad de Córdoba, 14071 Córdoba; [8] Dep. Microbiología, Campus Rabanales C6-B-O, Universidad de Córdoba, 14071 Córdoba; [9]Centro de Investigaciones Arqueobiológicas y Paleoecológicas Andinas Arqueobios, Apartado Postal 595, Trujillo (Peru); [10] Centro de Investigaciones Arqueobiológicas y Paleoecológicas Andinas Arqueobios, Apartado Postal 595, Trujillo (Peru); [11] Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Alameda del Obispo s/n, 14080 Córdoba

Abstract

First-Generation Sequencing (FGS) revolutionized molecular biology in general, and bioarchaeology in particular. That allowed to study ancient DNA (aDNA) remains for the first time. Second-Generation Sequencing (SGS) further improved throughput, reducing time and cost per sequenced nucleotide base. Most importantly, it allowed to sequence full ancient-genomes, like the ones of Neanderthals. Third-Generation Sequencing (TGS) improved even more SGS throughput, also allowing to sequence single molecules, without requiring previous amplification or modification steps. That is revolutionary, overcoming putative biases associated to such steps. This opens the door for true-direct RNA sequencing [without retrotranscription into complementary DNA (cDNA)], including ancient RNA (aRNA). Different TGS platforms are arising, promising exciting new discoveries on bioarchaeology for both aDNA and aRNA.
Key words: functional genomics, epigenetics, messenger RNA, ancient transcriptomes, paleotranscriptomics.

Resumen

La secuenciación de primera generación (FGS; del inglés, “First-Generation Sequencing”) revolucionó la biología molecular en general, y la bioarqueología en particular. Ello permitió estudiar restos de ADN antiguo (ADNa) por primera vez. La secuenciación de segunda generation (SGS; del inglés, “Second-Generation Sequencing”) mejoró aún más el rendimiento, reduciendo el tiempo y coste por base nucleotídica secuenciada. Además, permitió secuenciar genomas antiguos completos, como los de neandertales. La secuenciación de tercera generación (TGS; del inglés, “Third-Generation Sequencing”) mejoró aún más el rendimiento de SGS, permitiendo además secuenciar moléculas individuales, sin requerir pasos previos de amplificación o modificación. Ello es revolucionario, evitando posibles sesgos asociados a tales pasos previos. Esto abre la puerta para la secuenciación directa verdadera del ARN [sin retrotranscripción en ADN complementario (ADNc)], incluyendo ARN antiguo (ARNa). Diferentes plataformas de TGS están surgiendo, prometiendo nuevos e interesantes descubrimientos en bioarqueología, tanto para ADNa como ARNa.
Palabras clave: genómica funcional, epigenética, ARN mensajero, transcriptomas antiguos, paleotranscriptómica.

Introduction

First-Generation Sequencing (FGS) revolutionized bioarchaeology, allowing to sequence ancient DNA (aDNA). The breakthrough was extended to ancient genomes with Second-Generation Sequencing (SGS), sporting higher throughput, faster turnaround and cheaper cost per read nucleotide base. Among others, that included sequencing of Neanderthal genome (Green et al. 2006; Noonan et al. 2006; Dorado et al. 2007-2015). Third-Generation Sequencing (TGS) further improves previous methodologies, allowing to sequence single-molecules, without previous modification or amplification steps, generating longer (kilobase; kb) reads.

Such advancements avoid biases associated to previous approaches, further increasing productivity, reducing time and cost (Bleidorn 2015). Therefore, SGS and TGS are considered High-Throughput Sequencing (HTS) technologies, being sometimes referred with the ambiguous “Next-Generation” Sequencing (NGS) terminology. Additionally, these methodologies can also be applied to study DNA epigenetics, including aDNA (Gokhman et al. 2014; Pedersen et al. 2014; Orlando et al. 2015; Seguin-Orlando et al. 2015; Smith et al. 2015).

Besides aDNA, it would be also interesting to sequence ancient RNA (aRNA), allowing to take functional snapshots of particular tissues at ancient times. Yet, RNA is more labile than DNA, and usually degrades after physiological activity supporting-life ceases. It is known that aDNA is usually scarce, physically broken (degraded) and may be also chemically modified. Indeed, aDNA may be hard to isolate and sequence, being prone to generate false-positive results due to modern DNA contamination. Thus, it is expected that aRNA should be harder to isolate and sequence, albeit messenger RNA (mRNA) is much shorter than genomic DNA, which is certainly an advantage for sequencing. But, is it possible to actually sequence ancient RNA? The answer is yes –to some extent, so far–, as described below (Fordyce et al. 2013; Guy 2013, 2014; Ng et al. 2014; Smith et al. 2014).

Sequencing ancient RNA

There are two basic approaches to sequence RNA: retrotranscription and direct sequencing. The first approach can be accomplished with FGS and SGS, to generate complementary DNA (cDNA) (Fig. 1) and further sequencing. Such methodology is sometimes known with the misleading name of RNA-seq, instead of the more suitable cDNA-seq terminology. Thus, different aRNA have been sequenced (Guy et al. 2014), including 50-year-old RNA peach latent-mosaic viroid (Guy 2013), white-clover cryptic virus-1 (Guy and Gerard 2016), 700-year-old viral genomes from caribou feces (Ng et al, 2014) and 750 year-old barley stripe-mosaic virus (Smith et al. 2014). Even the transcriptome of ancient (723 year-old) maize kernels has been sequenced (Fordyce et al. 2013).

Curiously, since retroviruses insert into host genomes as Endogenous RetroViruses (ERV), their presence has been also detected sequencing modern genomes, as well as ancient ones. That is the case of 1,400 year-old

aDNA from coprolites (Rivera-Perez et al. 2015). Obviously, in this latter case, retrotranscription was carried out in living cells, being included here as a research curiosity.

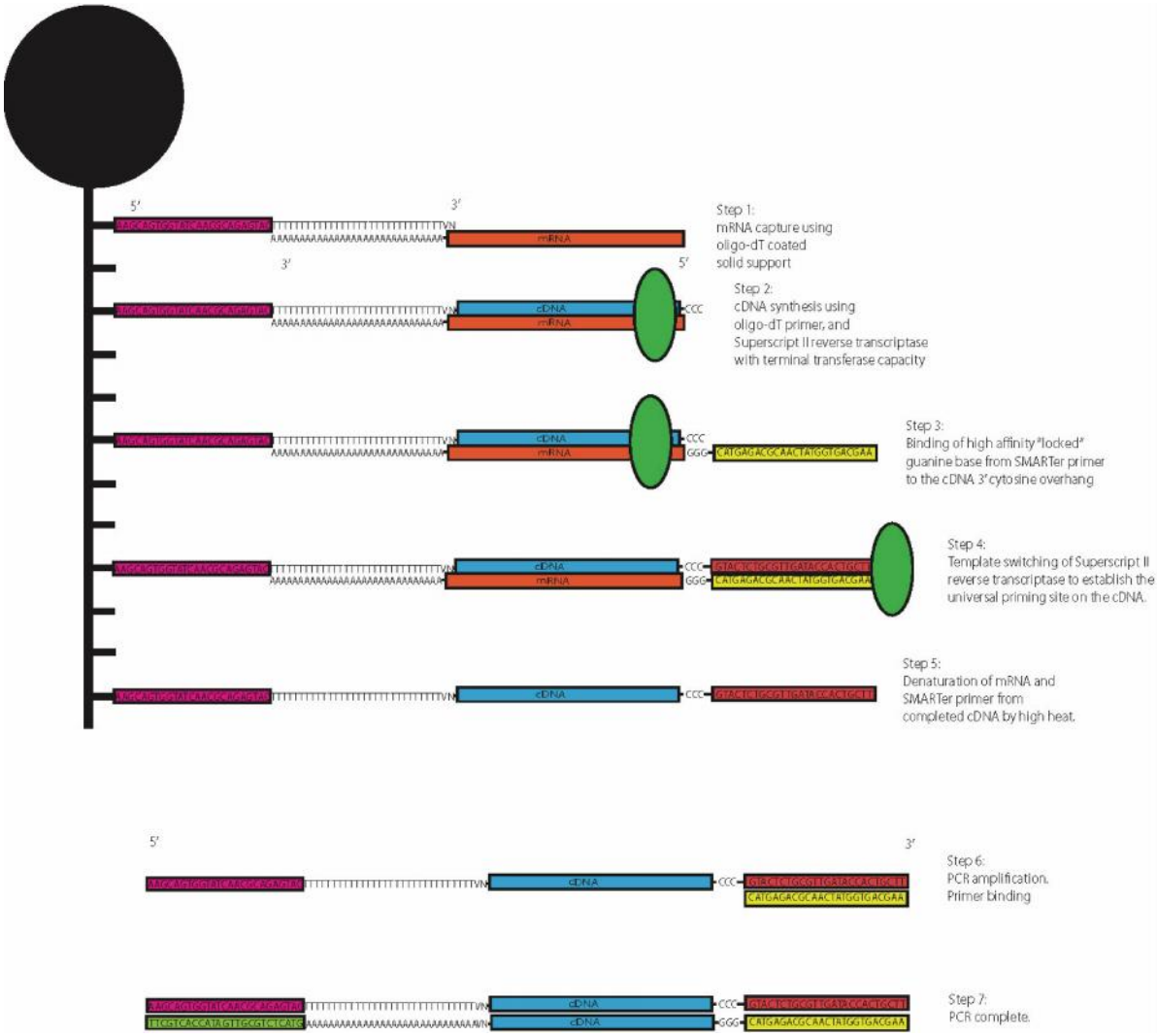

Figure N° 1: RNA retrotranscription into cDNA. Example of retrotranscription, in which RNA is captured with paramagnetic beads (top-left black ball) and retrotranscribed into cDNA. Terminal-transferase activity of used retrotranscriptase allows to generate CC tails, to further hybridize single-stranded adapters, which become double-stranded after polymerization. RNA-cDNA hybrid is denatured, allowing to amplify the first-strand cDNA into double-stranded DNA, via Polymerase Chain Reaction (PCR). © 2016 Danandmike, Wikimedia Commons <http://commons.wikimedia.org> and Creative Commons <http://creativecommons.org>.

Fortunately, TGS platforms allow to directly sequence nucleic-acids, as indicated above, without previous retrotranscription or amplification steps, effectively opening to door to truly sequence RNA in general, as well as aRNA in particular. Among them are: i) Helicos BioSciences <http://www.helicosbio.com> true Single-Molecule Sequencing (tSMS), based on Sequencing-by-Synthesis (SbS). It was the first single-molecule sequencing technology available (Harris et al., 2008), being later on offered as a service by "Sequence the Lower Limit" (SeqLL) <http://seqll.com>; ii) Pacific Biosciences

(PacBio) <http://www.pacb.com> Single-Molecule Real-Time (SMRT) sequencing, derived from former Nanofluidics <http://nanofluidics.com>, via SbS; iii) Complete Genomics <http://www.completegenomics.com> combinatorial Probe-Anchor Ligation (cPAL) sequencing, based on Sequencing by Hybridization (SbH) and Sequencing by Ligation (SbL), which is offered as AllSeq <http://allseq.com> service (Fig 2); and iv) Oxford Nanopore Technologies <http://nanoporetech.com> sequencing, based on ionic (electrical)-current alterations, as nucleotides pass though biological or synthetic solid-state nanopores. It generates long reads of 230 to 300 kb, at a speed of 280 to 500 bases per second (bps). Interestingly, a miniature handheld MinION Mk 1B sequencer (87 g weight) has been successfully used by the National Aeronautics and Space Administration (NASA) <https://www.nasa.gov> for astrobiology. It has exciting potential for crew-disease diagnostics and nucleic-acid-based life research (Rezzonico 2014; John et al. 2016).

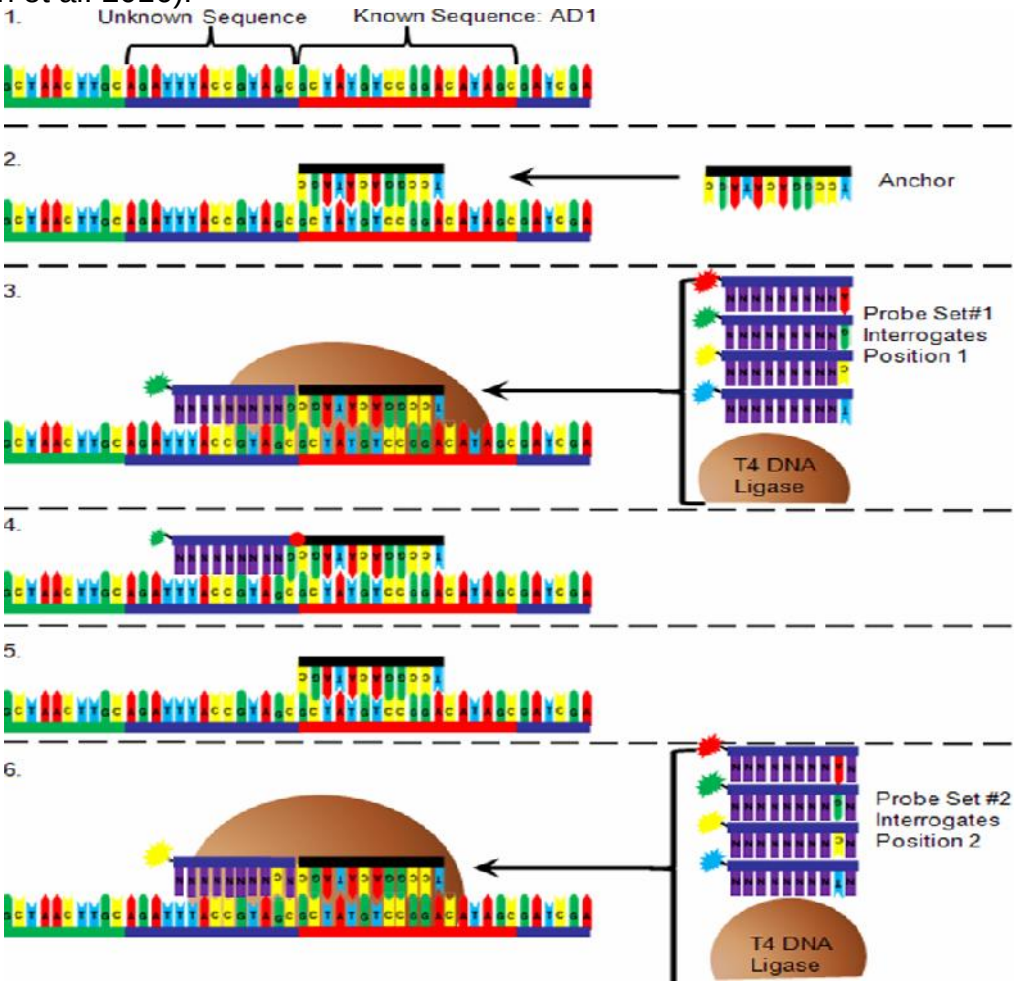

Figure N° 2: Single-molecule sequencing. Example of TGS (cPAL), in which an oligonucleotide anchor is hybridized to a known region of single-molecule to be sequenced. Further, a fluorescent interrogating-probe (containing A, C, G or T nitrogenous base at the interrogating position) is hybridized and ligated to the anchor. Non-hybridized probes are removed, and bound fluorescence is detected. The process is repeated for the rest of interrogating nucleotides. Then, the probe-anchor is removed and the process is repeated for other probes-anchors, until the target nucleic-acid is sequenced. © 2016 Suspencewl, Wikimedia Commons <http://commons.wikimedia.org> and Creative Commons <http://creativecommons.org>.

TGS has been used to sequence not only DNA, but also RNA by Direct RNA Sequencing (DRS) (Ozsolak et al. 2009). It has also been used to sequence ancient genomes, including Pleistocene horse (Orlando et al. 2011; Ginolhac et al. 2012; Schubert et al. 2012).

Future prospects and concluding remarks

TGS is an active research and development area, with promising new technologies in the works, including the following (alphabetical order by manufacturer): i) Electron Optica <http://www.electronoptica.com>, based on Low-Energy Electron Microscopy (LEEM); ii) Electronic BioSciences <http://electronicbio.com>, based on nanopore sequencing; iii) GenapSys <http://www.genapsys.com> Gene Electronic Nano-Integrated Ultra-Sensitive (GENIUS) sequencer, capable of measuring both pH and temperature changes during nucleic-acid synthesis; iv) Genia <http://www.geniachip.com>, purchased in 2014 by Roche <http://sequencing.roche.com>, based on SbS via PolyEthylene Glycol (PEG) NanoTags on active analog-to-digital sensors; v) GnuBIO <http://gnubio.com>, that belongs to Bio-Rad <http://www.bio-rad.com> since 2014, based on SbH hexamers in nanodrops, promising error-free runs (100% accuracy) after filtering-out data, in just ~3'5 h; vi) IBM <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/dnatransistor> sequencer, based on DNA Transistor solid-state nanopores; vii) LaserGen <http://lasergen.com>, using Lightning Terminators, based on photocleavable dyes and terminating groups (Wu et al. 2007); viii) Solametrix <http://www.solametrix.com>, that became Lightspeed Genomics <http://www.lsgen.com>, based on sub-pixel Synthetic Aperture Optics (SAO) with LightSpeed sensors, allowing 4x miniaturization and 16x reagent savings; ix) NABsys <http://nabsys.com>, based on hexamer probes and SbS with solid-state nanopores, sporting an amazing speed of more than one million bps (~100,000 bps per detector), allowing to sequence the ~3 mega-base pairs (Mbp) human genome with 25-fold depth for 100 USD, in less than an hour; and x) NobleGen Biosciences <http://noblegenbio.com> Optipore optical-nanopore, using fluorescent molecular-beacons and Circular DNA Conversion (CDC) to convent target nucleic-acid into Expanded Synthetic Representation (ESR) oligonucleotide chain (McNally et al. 2010).

Feasibility of direct RNA sequencing has been demonstrated (Ozsolak et al. 2009), albeit the age limit of aRNA suitable for sequencing remains to be determined, as previously done for aDNA (Allentoft et al, 2012; Millar and Lambert 2013). The possibility to sequence ancient transcriptomes (paleotranscriptomics) is certainly exciting, mostly when using TGS. Indeed, the direct and longer reads of new TGS approaches (like those using nanopores) are quite convenient for such a goal. Yet, special assembly algorithms may be also needed for aRNA (as in the case of aDNA), since such material is usually degraded, as previously pointed out. Besides, increasing computing power will be required to analyze and compare the exponentially-growing data generated from genomics, transcriptomics and proteomics studies.

***Acknowledgements*.** Supported by "Ministerio de Economía y Competitividad" (MINECO grants AGL2010-17316, BIO2011-15237-E and BIO2015-64737-R) and "Instituto Nacional de

Investigación y Tecnología Agraria y Alimentaria" (MINECO and INIA RF2012-00002-C02-02); "Consejería de Agricultura y Pesca" (041/C/2007, 75/C/2009 and 56/C/2010), "Consejería de Economía, Innovación y Ciencia" (P11-AGR-7322 and P12-AGR-0482) and "Grupo PAI" (AGR-248) of "Junta de Andalucía"; and "Universidad de Córdoba" ("Ayuda a Grupos"), Spain.

References

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MT, Willerslev E, Zhang G, Scofield RP, Holdaway RN, Bunce M (2012): The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc Biol Sci 279: 4724-4733.

Bleidorn C (2015): Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. Systematics and Biodiversity 14: 1-8.

Dorado G, Jiménez I, Rey I, Sánchez-Cañete FJS, Luque F, Morales A, Gálvez M, Sáiz J, Sánchez A, Rosales TE, Vásquez VF, Hernández P (2013): Genomics and proteomics in bioarchaeology - Review. Archaeobios 7: 47-63.

Dorado G, Luque F, Pascual P, Jiménez I, Sánchez-Cañete FJS, Pérez-Jiménez M, Raya P, Gálvez M, Sáiz J, Sánchez A, Rosales TE, Vásquez VF, Hernández P (2015): Second-generation nucleic-acid sequencing and bioarchaeology - Review. Archaeobios 9: 216-230.

Dorado G, Rey I, Rosales TE, Sánchez-Cañete FJS, Luque F, Jiménez I, Gálvez M, Sáiz J, Sánchez A, Vásquez VF (2009): Ancient DNA to decipher the domestication of dog (REVIEW). Archaeobios 3: 127-132.

Dorado G, Rey I, Rosales TE, Sánchez-Cañete FJS, Luque F, Jiménez I, Morales A, Gálvez M, Sáiz J, Sánchez A, Hernández P, Vásquez VF (2010): Biological mass extinctions on planet Earth (REVIEW). Archaeobios 4: 53-64.

Dorado G, Rosales TE, Luque F, Sánchez-Cañete FJS, Rey I, Jiménez I, Morales A, Gálvez M, Sáiz J, Sánchez A, Vásquez VF, Hernández P (2011): Ancient nucleic acids from maize - A review. Archaeobios 5: 21-28.

Dorado G, Rosales TE, Luque F, Sánchez-Cañete FJS, Rey I, Jiménez I, Morales A, Gálvez M, Sáiz J, Sánchez A, Vásquez VF, Hernández P (2012): Isotopes in bioarchaeology - Review. Archaeobios 6: 79-91.

Dorado G, Sánchez-Cañete FJS, Pascual P, Jiménez I, Luque F, Pérez-Jiménez M, Raya P, Gálvez M, Sáiz J, Sánchez A, Rosales TE, Vásquez VF, Hernández P (2014): Starch genomics and bioarchaeology - Review. Archaeobios 8: 41-50.

Dorado G, Vásquez V, Rey I, Luque F, Jiménez I, Morales A, Gálvez M, Sáiz J, Sánchez A, Hernández P (2008): Sequencing ancient and modern genomes (REVIEW). Archaeobios 2: 75-80.

Dorado G, Vásquez V, Rey I, Vega JL (2007): Archaeology meets Molecular Biology (REVIEW). Archaeobios 1: 1-2.

Fordyce SL, Ávila-Arcos MC, Rasmussen M, Cappellini E, Romero-Navarro JA, Wales N, Alquezar-Planas DE, Penfield S, Brown TA, Vielle-Calzada JP, Montiel R, Jorgensen T, Odegaard N, Jacobs M, Arriaza B, Higham TF, Ramsey CB, Willerslev E, Gilbert MT (2013): Deep sequencing of RNA from ancient maize kernels. PLoS One: e50961 (9 pp).

Ginolhac A, Vilstrup J, Stenderup J, Rasmussen M, Stiller M, Shapiro B, Zazula G, Froese D, Steinmann KE, Thompson JF, Al-Rasheid KA, Gilbert TM, Willerslev E, Orlando L (2012): Improving the performance of true single molecule sequencing for ancient DNA. BMC Genomics 13: 177 (14 pp).

Gokhman D, Lavi E, Prufer K, Fraga MF, Riancho JA, Kelso J, Paabo S, Meshorer E, Carmel L (2014): Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. Science 344:523-527.

Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, Paabo S (2006): Analysis of one million base pairs of Neanderthal DNA. Nature 444: 330-336.

Guy PL (2013): Ancient RNA? RT-PCR of 50-year-old RNA identifies peach latent mosaic viroid. Arch Virol 158: 691-694.

Guy PL (2014): Prospects for analyzing ancient RNA in preserved materials. Wiley Interdiscip Rev RNA 5: 87-94.

Guy PL, Gerard PJ (2016): White clover cryptic virus-1 in New Zealand and eastern Australia. Ann Appl Biol 168: 225-231.

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, and Xie Z (2008): Single-molecule DNA sequencing of a viral genome. Science 320: 106-109.

John KK, Botkin DS, Burton AS, Castro-Wallace SL, Chaput JD, Dworkin JP, Lehman N, Lupisella ML, Mason CE, Smith DJ, Stahl S, Switzer C (2016): The Biomolecule Sequencer Project: Nanopore sequencing as a dual-use tool for crew health and astrobiology investigations. 47th Lunar and Planetary Science Conference (21-25 March 2016; The Woodlands, TX, USA). Web <http://ntrs.nasa.gov/search.jsp?R=20160002394>.

McNally B, Singer A, Yu Z, Sun Y, Weng Z, Meller A (2010): Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays. Nano Lett 10: 2237-2244.

Millar CD, Lambert DM (2013): Ancient DNA: Towards a million-year-old genome. Nature 499: 34-35.

Ng TF, Chen LF, Zhou Y, Shapiro B, Stiller M, Heintzman PD, Varsani A, Kondov NO, Wong W, Deng X, Andrews TD, Moorman BJ, Meulendyk T, MacKay G, Gilbertson RL, Delwart E (2014): Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch. Proc Natl Acad Sci USA 111: 16842-16847.

Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Paabo S, Pritchard JK, Rubin EM (2006): Sequencing and analysis of Neanderthal genomic DNA. Science 314: 1113-1118.

Orlando L, Gilbert MTP, Willerslev E (2015): Applications of next-generation sequencing -Reconstructing ancient genomes and epigenomes. Nat Rev Genet 16: 395-408.

Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M, Magnussen K, Steinmann KE, Kapranov P, Thompson JF, Zazula G, Froese D, Moltke I, Shapiro B, Hofreiter M, Al-Rasheid KA, Gilbert MT, Willerslev E (2011): True single-molecule DNA sequencing of a Pleistocene horse bone. Genome Res 21: 1705-1719.

Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, and Milos PM (2009): Direct RNA sequencing. Nature 461: 814-818.

Pedersen JS, Valen E, Velazquez AM, Parker BJ, Rasmussenforlando M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, Andersson R, Jones PA, Hoover CA, Tikhonov A, Prokhortchouk E, Rubin EM, Sandelin A, Gilbert MT, Krogh A, Willerslev E, Orlando L (2014): Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res 24: 454-466.

Rezzonico F (2014): Nanopore-based instruments as biosensors for future planetary missions. Astrobiology 14: 344-351.

Rivera-Perez JI, Cano RJ, Narganes-Storde Y, Chanlatte-Baik L, Toranzos GA (2015): Retroviral DNA Sequences as a Means for Determining Ancient Diets. PLoS One 10: e0144951 (10 pp).

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KA, Willerslev E, Krogh A, Orlando L (2012): Improving ancient DNA read mapping against modern reference genomes. BMC Genomics 13: 178 (15 pp).

Seguin-Orlando A, Gamba C, Sarkissian CD, Ermini L, Louvel G, Boulygina E, Sokolov A, Nedoluzhko A, Lorenzen ED, Lopez P, McDonald HG, Scott E, Tikhonov A, Stafford TWJr, Alfarhan AH, Alquraishi SA, Al-Rasheid KA, Shapiro B, Willerslev E, Prokhortchouk E, Orlando L (2015): Pros and cons of methylation-based enrichment methods for ancient DNA. Sci Rep 5: 11826 (15 pp).

Smith O, Clapham A, Rose P, Liu Y, Wang J, Allaby RG (2014): A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. Sci Rep 4: 4003 (6 pp).

Smith RW, Monroe C, Bolnick DA (2015): Detection of Cytosine methylation in ancient DNA from five native American populations using bisulfite sequencing. PLoS One 10: e0125344 (23 pp).

Wu W, Stupi BP, Litosh VA, Mansouri D, Farley D, Morris S, Metzker S, Metzker ML (2007): Termination of DNA synthesis by N6-alkylated, not 3'-O-alkylated, photocleavable 2'-deoxyadenosine triphosphates. Nucleic Acids Res. 35:6339-6349.