

Review

수생태계 생물다양성 연구를 위한 환경유전자(environmental DNA) 기술의 적용과 활용

곽인실* · 박영석¹ · 장광현²

전남대학교 해양융합과학과, ¹경희대학교 생물학과, ²경희대학교 환경학및환경공학과

Application and Utilization of Environmental DNA Technology for Biodiversity in Water Ecosystems. *Ihn-Sil Kwak** (0000-0002-1010-3965), *Young-Seuk Park*¹ (0000-0001-7025-8945) and *Kwang-Hyeon Chang*² (0000-0002-7952-4047), (Department of Ocean Integrated Science, Chonnam National University, Yeosu 59626, Republic of Korea; ¹Department of Biology, Kyung Hee University, Seoul 02447, Republic of Korea; ²Department of Environmental Science and Engineering, Kyung Hee University, Yongin 17104, Republic of Korea)

Abstract The application of environmental DNA in the domestic ecosystem is also accelerating, but the processing and analysis of the produced data is limited, and doubts are raised about the reliability of the analyzed and produced biological taxa identification data, and the sample medium (target sample, water, air, sediment, Gastric contents, feces, etc.) and quantification and improvement of analysis methods are also needed. Therefore, in order to secure the reliability and accuracy of biodiversity research using the environmental DNA of the domestic ecosystem, it is a process of actively using the database accumulated through ecological taxonomy and undergoing verification procedures, and experts verifying the resolution of the data increased by gene sequence analysis. This is absolutely necessary. Environmental DNA research cannot be solved only by applying molecular biology technology, and interdisciplinary research cooperation such as ecology-taxa identification-genetics-informatics is important to secure the reliability of the produced data, and researchers dealing with various media can approach it together. It is an area in desperate need of an information sharing platform that can do this, and the speed of development will proceed rapidly, and the accumulated data is expected to grow as big data within a few years.

Key words: environmental DNA, DNA barcoding, metabarcoding

환경유전자(environmental DNA, eDNA)는 생태학과 환경관리에 떠오르는 중요한 도구가 되고 있다. 최근 10여 년에 걸쳐 차세대 염기서열분석(next-generation sequencers, NGS)의 급격한 발달로 연구자들의 관심과 다양성 연구 및

희귀종 탐색의 접근론적 방법으로 고려가 집중되고 있다. 2019년 젊은 수생태 연구자들이 모여, eDNA연구회를 만들어 연구 교류가 시작되었고, 2020년 한국생태환경과학협회의 학술발표대회 주제로 “차세대 생태환경과학 연구를 위한 첨단기법”에서 eDNA 기술의 활용과 적용을 알렸으며, 이후 2021년 7월 생태학 및 분류학을 기반으로 eDNA 연구에 관심을 가진 연구자들이 모여 한국환경유전자학회(Korean

Manuscript received 15 September 2021, revised 24 September 2021, revision accepted 24 September 2021

* Corresponding author: Tel: +82-61-659-7148, Fax: +82-61-653-6620
E-mail: iskwak@chonnam.ac.kr

© The Korean Society of Limnology. All rights reserved.

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provide the original work is properly cited.

Society of Environmental DNA: KSED)를 창립하였다. 다양한 환경 매체를 다루는 연구자들이 시료 채집과 처리방법, 데이터 처리 기술과 현안 문제들을 공유하고 해결할 수 있는 플랫폼으로서의 역할을 KSED가 할 수 있기를 기대한다. 현재, 시료로부터 유전자 추출과 시퀀싱 기술은 많은 발전을 이루어져 있으며, eDNA에 적용되는 기본 기술은 동일하고 시퀀싱으로 얻어진 생물 데이터의 신뢰성과 정확성을 보장받기 위한 강력한 데이터베이스의 축적과 파이프라인이 구축되어야 하는 필요성이 있으며, 또한 이는 국내 실정에 맞는 국제적인 틀을 기반으로 이루어져야 한다. 환경유전자에 많은 관심을 가지고 역할과 기능을 만들어 참여하실 수 있도록, 유연한 조직을 갖추어 연구 실무진에 속하는 미래 학문 후속세대인 젊은 연구자와 대학원생들이 참여하는 운영진으로 구성하고자 하였으며, 원하는 연구자들이라면 누구나 참여할 수 있도록 하였다. 앞으로 많은 연구자들이 협력하고 연구를 진행할 수 있는 연구 플랫폼으로 성장하리라 기대한다.

본 총설에서는 환경유전자의 용어 정의와 연구가 이루어진 진행과정을 살펴보고자 한다. 미생물 분야의 eDNA 연구는 first NGS (next-generation sequencers) 시대인 2005년 이전에는 새로운 sequence의 발생과 분석에 필요한 추가 연구의 고비용 문제로 연구가 매우 소수 이루어졌다. 또한 다양한 생물이 복합 혼합되어 있는 DNA의 경우에는 분해되어진 상태나 낮은 농도로 신선한 단일 생물의 조직으로부터 추출하고 분석하는 것보다 어려움이 많아 쉽지 않았다. 그리고 eDNA 분석은 전통적인 생태학, 바이오인포메틱스, 분자생물학적인 기술에 대한 지식을 필요하기에 분야 간의 이해가 이루어져야 한다는 점에서는 어려움이 있다.

먼저, 환경유전자(environmental DNA, eDNA)란 환경으로부터 수집된 시료에 포함되어져 있는 많은 생물로부터 추출된 염색체로 이루어진 genomic DNA의 복합체라고 정의되기도 하였다(Taberlet *et al.*, 2012). 현재는 물, 눈, 흙, 퇴적물, 공기 등 다양한 환경에 잔존하는 DNA 복합체로 생물 유래의 DNA를 총칭하는 의미로 확장되어 사용되고 있다. 또한 DNA 기반의 분류학적 또는 기능적인 정보 분석으로 생태계를 이해하기 위한 목적으로 사용하기 위한 시도도 이루어지고 있다. 총 DNA는 세포 내와 세포 외의 DNA 둘 다 포함한 것이다(Levy-Booth *et al.*, 2007). 세포 내 DNA는 살아있는 세포 또는 환경 시료로부터 얻어진 살아있는 다양한 생물 유기체로부터 추출한 것이고, 세포 외 DNA는 죽은 세포 또는 파괴된 세포로 물리적, 화학적, 생물학적 과정을 거쳐 분해되고 있는 상태로 구분할 수 있다. DNA 분자는 뉴클레아제(nucleases)의 작용으로 작은 절편으로 잘라지고, 이 절편들은 진흙, 모래, 점토 그리고 축축한 기질의 표면에 부착되어진 상태의 세포 외 eDNA로 존재하게 될 것이다.

eDNA 시료에서 분류군을 확인하고 싶다면, PCR (Polymerase Chain Reaction)을 기반으로 2가지 접근법을 고려할 수 있는데, 1) 단일종(single species)의 존재 유무 파악을 목표로 할 경우, 종 특이적 정량적인 PCR 분석방법(Logan *et al.*, 2009)과 2) target PCR (Saiki *et al.*, 1988, White *et al.*, 1989)이나 shotgun sequencing (산탄 염기서열 결정법: Deininger, 1983)으로 고도로 잘 보존된 짧은 염기서열을 이용하여 잠재 가능한 모든 종을 제시하는 방법이 있는데, 보통 두 번째 방법을 “DNA metabarcoding”이라고 한다. “DNA metabarcoding”은 동일한 eDNA 샘플 내에서 많은 분류군을 동시에 식별할 수 있는 방식의 DNA 바코딩하는 기술이다(Pompanon *et al.*, 2011; Riaz *et al.*, 2011). 바코딩은 잘 보존된 비교적 짧은 분류학적인 정보가 담긴 유전자를 이용하여 프라이머를 제작하고 PCR를 진행할 수 있게 되는데, 주로 단일 종을 파악하기 위해 DNA 추출, PCR 증폭, 시퀀싱, 데이터 분석의 단계이다. 반면에, 메타바코딩은 바코드와 동일하게 DNA 추출, 증폭, 시퀀싱, 분석을 거치지만, shotgun sequencing을 수행하여 동일 샘플에서 광범위하게 얻은 짧은 염기서열로 분류학적인 동정을 파악하게 됨으로 정확한 종 동정을 위한 데이터베이스 또는 파이프라인을 사용하는 방식이다. “metagenomics”라는 용어가 사용되기도 하지만 genome의 기능적인 측면과 조합을 다루게 될 때 적절하다는 의견이 있으며, shotgun sequencing와 16S rDNA의 분류학적 동정에 주로 사용하고 있는 방법이다(Tringe *et al.*, 2005). “metatranscriptomics”은 환경시료로부터 RNA를 추출하여, 그 시점의 유전자 발현(gene expression)과 조절(regulation)을 연구하는 방법으로 분석 대상 시료의 분류학적 기능적 요인을 분석하는 데, mRNA의 반감기는 매우 짧아 유전자 발현을 측정하는 데 극복할 문제들이 있으며, 총 RNA는 주로 rRNA이기에 기능적 정보를 직접적으로 해석하기 어려운 한계를 해결하기 위한 연구가 진행 중이다(Selinger *et al.*, 2003).

eDNA 연구의 발전에 대해 살펴보면, Fig. 1과 같이 간략하게 소개되어졌는데, Ogram 등에 의해 1987년 퇴적물에서 eDNA를 추출하는 프로토콜이 보고되었고, 그 후 불과 3년만에 획기적인 DNA metabarcoding이 소개되었다(Giovannoni *et al.*, 1990). Sargasso Sea에서 조사된 박테리아플랑크톤 16S rRNA 유전자의 다양성을 보여주었다(Ogram *et al.*, 1987). Handelsman은 1998년 볼특정 미생물 내 활성 분자의 합성 경로를 규명하기 위해 토양 eDNA의 시퀀싱 서열을 무작위로 잘게 자르고 클로닝하는 metagenomics를 처음으로 시도하였다. 2000년대는 DNA metabarcoding과 metagenomics가 미생물학 분야에서 보편성을 갖추게 되었고, 2003년에는 연구대상 생물의 범위가 넓

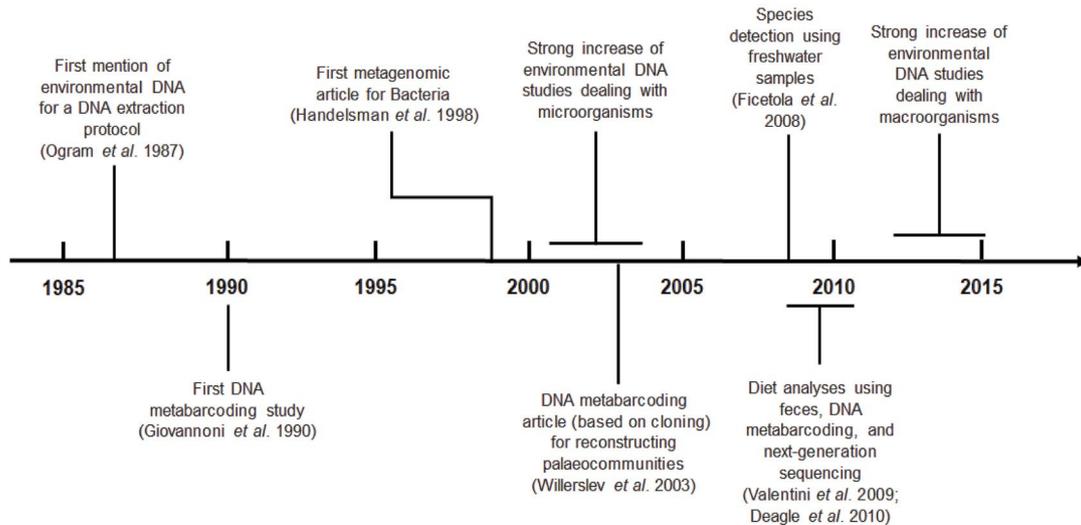


Fig. 1. Overview of the emergence of eDNA studies (Taberlet *et al.*, 2018).

어져서 동굴 퇴적물로부터 추출된 DNA를 이용하여 크기가 큰 맘모스, 말 그리고 고식물 연구에 DNA metabarcoding을 적용하였다(Willerslev *et al.*, 2003). 처음의 metabarcoding과 metagenomics는 단일 DNA 절편을 분리하고 클로닝하는 기술에 의존하였는데, 이를 진행하기 기반 기술은 분리 복제된 DNA 절편을 클론 벡터(플라스미드, 박테리오파이지 등)에 주입하고, 주입시킬 수 있는 적합한 숙주세포(주로, *Escherichia coli*)가 필요하였다. 이 방법은 클로닝 단계의 많은 비용과 시간 소모가 들어가게 되었고, 이를 개선하여 불필요한 클로닝 단계를 획기적으로 줄여준 기술이 2005년 이후에 사용된 NGS다. 2000년대 후반부터 DNA 메타바코딩은 미생물의 범위를 벗어나 대형생물로 확장되었고, 연구대상도 다양해지면서, 분변을 이용한 먹이원을 분석하는 시도가 있었다(Daegle *et al.*, 2010; Pompanon *et al.*, 2012). 이후 환경유전자의 연구는 물과 토양(Ficetola *et al.*, 2008), 담수의 대형생물(Dejean *et al.*, 2012)로 연구 논문들이 보고되었고, 메타바코딩을 적용하여 많은 분류군의 동정에 사용되었다(Thomsen *et al.*, 2012; Valentini *et al.*, 2016). 현재, 환경유전자 연구는 다양한 분류군 연구에 쉽게 적용되어지는 장점으로 연구가 급격히 늘어나고 있으나 주로 기술적인 방법의 개선과 바이오인포매틱스에 의존하는 수준으로 진행되고 있다.

환경유전자 연구를 위해 연구자가 접근할 방법을 선택하기 전에 파악할 흐름도는 Fig. 2에 개요적으로 담겨져 있는데, 크게 3가지 접근법으로 정리할 수 있다. 1) 종 특이 DNA를 이용한 targeting single species를 대상으로 접근하는 연구는 이미 알려진 분류군 연구에 적용되고 있다(Ficetola

et al., 2008; Thomsen *et al.*, 2012). 2) PCR 분석을 기반으로 하여, 대상 분류군에 포함된 모든 분류군을 탐색하는 것(group specific primers)을 목적으로 하는 “메타바코딩”을 사용하는 방법으로, 박테리아, 균류, 식물, 진행생물, 지렁이, 물고기 등을 대상으로 작용이 되었다. 3) 메타제노믹스로 특정 targeting PCR이 없이 eDNA의 shotgun sequencing을 진행하는 방식으로, 계놈의 기능적 연구, 주로 미생물에 적용되고 있다(Simon and Daniel, 2011).

바이오 기술의 환경유전자로의 적용이 진행되면서, 단일 종의 존재 유무 파악을 위한 연구 방법으로 Taq-man probe 및 SYBR green을 이용한 quantitative PCR (qPCR)과 digital droplet PCR (ddPCR)을 중심으로 연구가 활발하게 진행되고 있다(Dejean *et al.*, 2011; Davy *et al.*, 2015; Carlsson *et al.*, 2017; Currier *et al.*, 2017). 또한, 최근에는 CRISPR-Cas 기술을 이용하여 타겟 종을 확인하는 연구가 제안되는 등 기술적으로 다양한 방법이 진행되고 있다(Williams *et al.*, 2019). 다양한 환경 샘플(물, 퇴적물, 토양, 배설물 등)에 존재하는 eDNA는 상업적으로 판매하는 키트(DNeasy Powerwater kit, DNeasy Powersoil kit, DNeasy Blood & Tissue kit 등)와 매뉴얼 방법(high salt, CTAB 등)으로 DNA 추출방법이 소개되고 있으나, 환경유전자 연구에 처음으로 접근하는 사용자들은 제품화된 키트를 이용하는 경우가 대다수이다(Lear *et al.*, 2017). 성공적으로 추출된 eDNA는 분석 목적에 맞추어 12S rRNA, 16S rRNA, 18S rRNA와 COI (Cytochrome oxidase subunit I)과 같이 분류학적 보전성이 높은 유전자를 기반으로 제작된 프라이머를 이용하여 PCR을 적용하여 DNA 복합체 내 타겟 DNA만 선



Fig. 2. The main steps of eDNA researches, following targeting purpose (Taberlet *et al.*, 2018).

별적으로 증폭이 가능하게 되었고, 이후 PCR로 타겟 영역이 증폭된 DNA는 차세대 염기서열 분석을 진행하고, 얻은 데이터를 파이프라인(Qiime2, dada2, PEMA 등)을 이용하여 생태학적 분석과 해석을 가능한 기반 자료가 도출이 가능하게 되었다.

환경유전자의 활용과 전망은 가속화되고 있으며, 다양한 분야에서 활용이 가능하다. 희귀종 존재 유무 탐색, 외래종 유입 여부, 공기 속에 유해생물, 수생태계 내 생물의 서식 여부 등 알지 못하는 원인 생물에 대해 신속하게 스크리닝을 할 수 있는 기술이다. 다만 알려진 유전자 정보가 있는 경우에만 가능하다는 크고 중요한 단점을 가지고 있다. 특히 기존의 수생태계 생물 모니터링에 소모되는 시간과 비용, 노동력 절감의 대안으로 시도를 하고 있거나 계획을 하고 있다면, 시행착오와 방법정립에 연구자들의 협력이 필요하다. 시민참여와 비전문가 그룹에 대한 교육과 이해를 기반으로 시료 수집의 정확성을 담보할 선제 기술을 요하며, 이 방안이 수립되지 않은 상태에서 진행될 경우 산출된 자료의 신뢰성 문제로 데이터는 사용되기 힘들게 된다. 분류학적 지식을 기반으로 한

전문가들의 현미경 사용에 의한 종 동정의 기술이 유전자 분석 장비의 발달로 기계화 대량화가 예고되는 시점이다. 더욱 중요한 것은 대량화된 생물 빅데이터를 다룰 정보 기술과 짧은 서열정보에서 기인한 오동정을 걸러내기 위한 기초학문에 대한 지원과 노력이 없다면, 허상 속에 놓인 다량의 데이터를 분석하지만 효용가치가 없는 자료의 홍수 속에 갇히게 될 수도 있음을 인지하여야 할 것이다.

적 요

국내 생태계의 환경유전자 적용도 가속화되고 있으나 생산된 데이터를 처리하고 분석하는 데 한계를 느끼고 있으며, 분석 생산된 생물데이터의 신뢰성에 대한 의구심이 제기되기도 하고 시료 매체(대상 시료, 물, 공기, 퇴적물, 위내용물, 분변 등) 간의 방법에 따른 차이와 분석 방법의 정량화와 개선 등이 필요한 상태이기도 하다. 따라서 국내 생태계의 환경유전자를 활용한 생물다양성 연구의 신뢰성과 정확성을 확

보하기 위해서는 생태분류학으로 축적된 데이터베이스를 적극적으로 활용하여 검증 절차를 거치고, 유전자 서열 분석으로 높아진 데이터의 해상력을 전문가들이 검증하는 과정이 반드시 필요하다. 환경유전자 연구는 분자생물학적인 기술 적용으로만 해결될 수 없으며, 생산된 데이터의 신뢰성을 확보하기 위한 생태-분류-유전학-인포메틱스 등 학제 간의 연구 협력이 중요하며, 다양한 매체를 다루는 연구자들이 함께 접근할 수 있는 정보 공유 플랫폼이 절실한 분야이며, 발전의 속도도 급격하게 진행될 것이고 축적되는 데이터는 수년 내에 빅데이터로서 성장할 수 있을 것으로 기대된다.

저자정보 곽인실 (전남대학교 교수), 장광현 (경희대학교 교수), 박영석 (경희대학교 교수)

저자기여도 개념설정: 곽인실 & 장광현 & 박영석, 방법론: 곽인실 & 장광현 & 박영석, 분석: 곽인실 & 장광현 & 박영석, 자료제공: 곽인실 & 장광현 & 박영석, 자료관리: 곽인실 & 장광현 & 박영석, 원고 초안작성: 곽인실, 원고 교정: 곽인실 & 장광현 & 박영석, 원고 편집 및 검토: 곽인실 & 장광현 & 박영석, 과제관리: 곽인실, 연구비 수주: 곽인실. 모든 저자는 논문의 결과에 동의하였고, 출판될 최종본을 검토하고 동의하였습니다.

이해관계 이 논문에는 이해관계 충돌의 여지가 없음.

연구비 이 논문은 한국연구재단 중점연구사업 [NRF-2018-R1A6A1A-03024314]과 한국환경산업기술원 수생태계 건강성 확보 기술개발사업[2021003050001]의 지원을 받았습니다.

REFERENCES

- Deagle, B.E., A. Chiaradia, J. McInnes and S.N. Jarman. 2010. Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conservation Genetics* **11**: 2039-2048.
- Deininger, P.L. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Analytical Biochemistry* **129**: 216-223.
- Dejean, T., A. Valentini, C. Miquel, P. Taberlet, E. Bellemain and C. Miaud. 2012. Improved detection of and alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology* **49**: 953-959.
- Ficetola, G.F., C. Miaud, F. Pompanon and P. Taberlet. 2008. Species detection using environmental DNA from water samples. *Biology Letters* **4**: 423-425.
- Giovannoni, S.J., T.B. Britschgi, C.L. Moyer and K.G. Field. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**: 60-63.
- Levy-Booth, D.J., R.G. Campbell, R.H. Gulden, M.M. Hart, J.R. Powell, J.N. Klironomos and K.E. Dunfield. 2007. Cycling of extracellular DNA in the soil environment. *Soil Biology & Biochemistry* **39**: 2977-2991.
- Logan, J.M.J., K.J. Edwards and N.A. Saunders. 2009. Real-time PCR: current technology and applications. Horizon Scientific Press Co., Norwich, UK.
- Ogram, A., G.S. Sayler and T. Barkay. 1987. The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods* **7**: 57-66.
- Pompanon, F., B.E. Deagle, W.O.C. Symondson, D.S. Brown, S.N. Jarman and P. Taberlet. 2012. Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology* **21**: 1931-1950.
- Pompanon, F., E. Coissac and P. Taberlet. 2011. Metabarcoding a new way to analyze biodiversity. *Biofutur* **319**: 30-32.
- Riaz, T., W. Shehzad, A. Viari, F. Pompanon, P. Taberlet and E. Coissac. 2011. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* **39**: e145.
- Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, T.G. Horn and H.A. Erlich. 1998. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.
- Selinger, D.W., R.M. Saxena, K.J. Cheung, G.M. Church and C. Rosenow. 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Research* **13**: 216-223.
- Simon, C. and R. Daniel. 2011. Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology* **77**: 1153-1161.
- Taberlet, P., E. Coissac, M. Hajibabaei and L.H. Rieseberg. 2012. Environmental DNA. *Molecular Ecology* **21**: 1789-1793.
- Taberlet, P., A. Bonin, L. Zinger and E. Coissac. 2018. Environmental DNA for biodiversity research and monitoring. Oxford University Press.
- Thomsen, P.F., J. Kielgast, L.L. Iversen, P.R. Moller, M. Rasmussen and E. Willerslev. 2012. Detection of a diverse marine fish fauna using environmental DNA from seawater samplers. *Plos One* **7**: e41732.
- Tringe, S.G., C. Von Mering, A. Kobayashi, A.A. Salamov, K. Chen, H.W. Chang and E.M. Rubin. 2005. Comparative metagenomics of microbial communities. *Science* **308**: 554-557.
- Valentini, A., C. Miquel and P. Taberlet. 2010. DNA barcoding for honey biodiversity. *Diversity* **2**: 610-617.
- White, T.J., N. Arnheim and H.A. Erlich. 1989. The polymerase chain reaction. *Trends in Genetics* **5**: 185-189.
- Willerslev, E., A.J. Hansen, J. Binladen, T.B. Brand, M.T.P. Gilbert, B. Shapiro and A. Cooper. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**: 791-795.