Visualization of international environmental DNA research

Li Jiang[1,2] **and Yongqing Yang**[3,*]

[1]Library and [2]College of History and Society, Chongqing Normal University, Chongqing 401331, China
[3]College of Life Sciences, Chongqing Normal University, Chongqing 401331, China

Environmental DNA (eDNA) analysis is increasingly being used in research fields of archaeology, biology and environmental science. In this study, scientometric methods have been used to quantitatively assess the current global research status in the eDNA field based on SCI-EXPANDED and Social Sciences Citation Index databases during the period 1992–2016. CiteSpace software was used to visualize the eDNA knowledge domains. The most productive category and journal are microbiology and *Applied and Environmental Microbiology* respectively. USA is the leading country, Rockefeller University is the prominent institution and Brady, S. F. is the most productive author. Document co-citation analysis demonstrated that the most recent domain is focusing on using eDNA as a tool to detect species in aquatic environments. These findings may help researchers better understand the current progress as well as identify the latest frontiers in the field of eDNA.

Keywords: Citation burst, environmental DNA, scientometrics, visualization.

ENVIRONMENTAL DNA (eDNA) refers to the DNA obtained from environmental samples such as sediments, ice, water and even air. The term ‘environmental DNA’ first occurred in an article that characterized microbial communities according to DNA extracted directly from marine sediments[1]. Several fields like archaeology, biology and environmental science use eDNA to detect taxa across various types of environments. In recent years, the rapidly expanding study of eDNA has provided unprecedented ability to detect species and conduct genetic analyses for conservation, management and research in diverse systems, including terrestrial, freshwater and marine, particularly in scenarios where collection of whole organisms is impractical or impossible[2]. eDNA analysis has been widely used in the detection of rare, invasive or elusive species, and has also been applied to eDNA persistence studies, estimation of biodiversity, and species biomass and distribution.

Scientometric studies are important to make comprehensive evaluation of the development of certain research fields and have been widely used in many natural and social sciences. CiteSpace, a free Java application, has become one of the most popular citation analysis and visualization software for its priorities in clarity and interpretability of visualization[3]. CiteSpace not only supports a co-citation network analysis which particularly focuses on nodes that play critical roles in the evolution of a network over time, but also provides a burst detection that can identify abrupt changes in the scientific literature during any given time period[3–7]. Using Citespace, intellectual basis, landmarks and hotspots of a group of papers in the literature in a field can be easily identified. Several articles about trend analysis using CiteSpace have been published. However, few studies have been made on eDNA using scientometric methods.

In this article, we deal with the co-citation network analysis and burst detection functions supplied by CiteSpace to detect and visualize the emerging trends and evolutions in the field of eDNA from 1992 to 2016. We generated two types of visual maps: a mixed map of country/institution/author (Figure 1), and a co-citation timeline map (Figure 2). These visual maps can assist in understanding the distribution of research contributors, research evolution and frontiers of eDNA area. In these maps, each node represents one entity (country/institution/author or cited reference). Larger node sizes imply that the entity is an important one within the knowledge domain. The thickness of each ring in a node is proportional to the frequency of occurrence or the number of citations received in a given time slice. In this study, each slice represents a time length of two years. The blue, green, yellow and orange colours in the figures represent different year ranges, whereas red node indicates a citation burst of this article in a given time period.

Databases and methods

The dataset used in this study was obtained from Thompson Reuters Web of Science (WoS), the on-line version of the SCI-EXPANDED and Social Sciences Citation Index, on 5 August 2016 using the topic =‘environmental DNA’ with timspan = ‘1900–2016’, which retrieved a total of 948 documents. The retrieved results consist of article (800), review (99), meeting abstract (20), proceedings paper (19), editorial material (16), news item (5), letter (44) and correction addition (4). As article and

---

*For correspondence. (e-mail: yangyqcq@126.com)

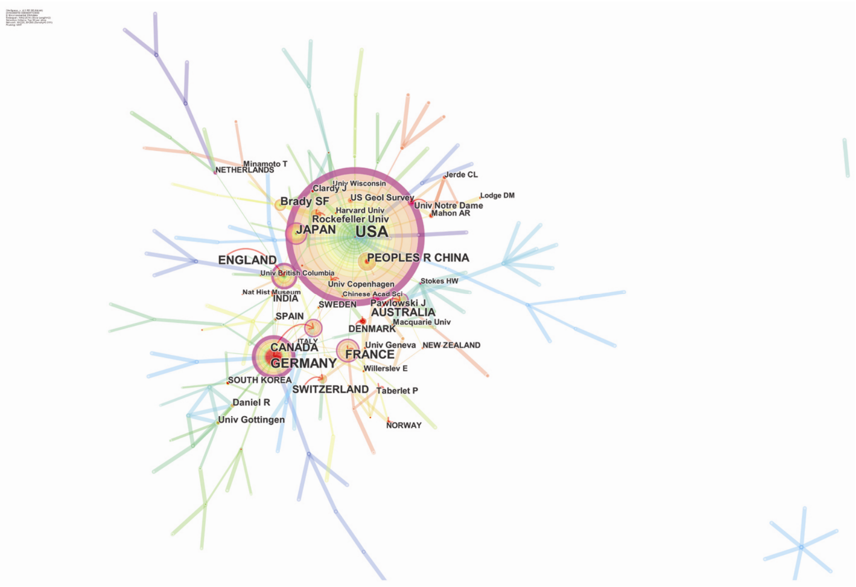

Figure 1. Country–institution–author map.

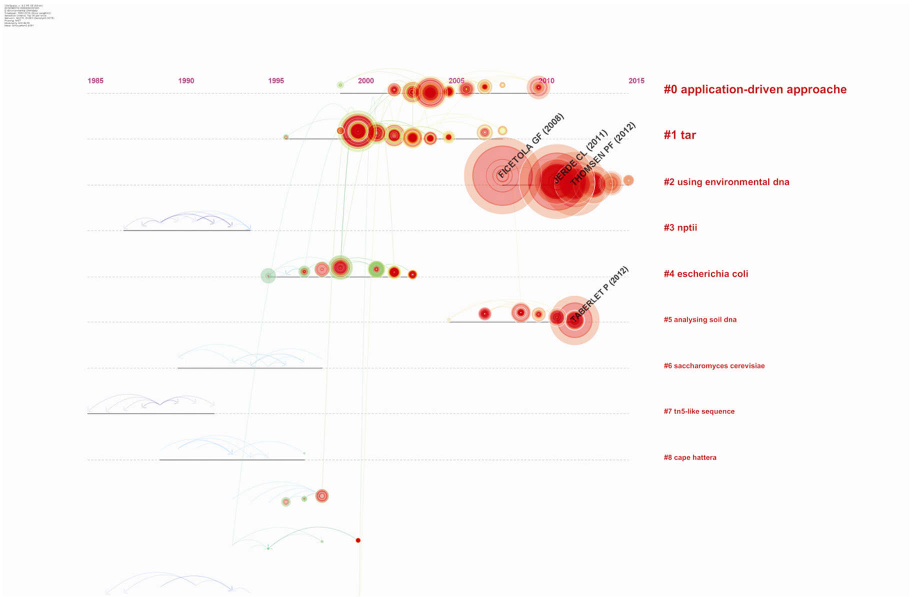

Figure 2. Co-citation timeline map.

review represented the majority of document types, data were refined by document type = (article and review), and the full bibliographic records (author, title, abstract and reference) for 899 papers were retrieved and downloaded for further analysis by CiteSpace III (version 4.0 R5 SE), to analyse and visualize co-citation networks.

Results and discussions

Yearly distribution

According to the data, research in the eDNA field has existed for more than two decades, from 1992 to 2016. The publication outputs increased steadily with time. However, a clear interest in eDNA research did not emerge until 2000, and the period from 2001 to 2015 has been the rapidly developing phase. The annual cumulative publications fit very well by exponential growth model $y = 3\text{E-}231e^{0.2671x}$ ($R^2 = 0.9569$), where y is the number of annual cumulative publications and x is the year, thus predicting that the research interest in eDNA will continue to increase in the future.

Categories and journals

Research on eDNA covers a wide range of subject categories, as much as 72 WoS categories. The top 12 productive categories are microbiology (269 articles),

Table 1. Top 12 categories and core journals on environmental DNA

Rank	Subject category	Frequency	Percentage	Journal	Frequency	Percentage
1	Microbiology	269	29.922	*Applied and Environmental Microbiology*	66	7.341
2	Biotechnology applied microbiology	186	20.690	*PLoS ONE*	63	7.008
3	Ecology	153	17.019	*Molecular Ecology Resources*	30	3.337
4	Biochemistry molecular biology	134	14.905	*Environmental Microbiology*	29	3.226
5	Multidisciplinary sciences	101	11.235	*Molecular Ecology*	24	2.670
6	Evolutionary biology	81	9.010	*Applied Microbiology and Biotechnology*	22	2.447
7	Marine freshwater biology	63	7.008	*FEMS Microbiology Ecology*	16	1.780
8	Environmental Sciences	55	6.118	*Proceedings of the National Academy of Sciences of the United States of America*	15	1.669
9	Genetics heredity	51	5.673	*Biological Conservation*	14	1.557
10	Biodiversity conservation	43	4.783	*FEMS Microbiology Letters*	13	1.446
11	Biochemical research methods	40	4.449	*Frontiers in Microbiology*	13	1.446
12	Biology	24	2.670	*Journal of Microbiological Methods*	13	1.446

biotechnology applied microbiology (186), ecology (153), biochemistry molecular biology (134), multidisciplinary sciences (101), evolutionary biology (81), marine freshwater biology (63), environmental sciences (55), genetics heredity (51), biodiversity conservation (43), biochemical research methods (40) and biology (24) (Table 1).

Articles on eDNA appeared in 333 journals. A concentration of eDNA articles within major journals could be observed, with the top 10 journals publishing 318 (35.37%) of the total articles. Among these top 10 journals, *Applied and Environmental Microbiology* ranked first with 66 publications (7.341% of the total), followed by *PLoS ONE* (7.008%); journals ranked from 3 to 10 were: *Molecular Ecology Resources*, *Environmental Microbiology*, *Molecular Ecology*, *Applied Microbiology and Biotechnology*, *FEMS Microbiology Ecology*, *Proceedings of the National Academy of Sciences USA (PNAS)*, *Biological Conservation*, *FEMS Microbiology Letters*, *Frontiers in Microbiology* and *Journal of Microbiological Methods* (Table 1). It should be noted that most of these journals generally have an high impact factor, and one of the world's most cited comprehensive journal, *PNAS*, is among the top 10 core journals, indicating the dominant academic influence of the eDNA field of research.

Countries, institutions and authors

Countries, institutions and authors contributed more than 10 articles, as shown in Figure 1. In this visualization, the node size represents the overall frequency of occurrence of countries, institutions and authors, and the coloured rings of the nodes represent yearly time slices. The top 10 most productive countries were USA (293), Germany (100), England (67), France (62), Japan (59), Austalia (54), China (51), Canada (49), Switzerland (38) and India (25). Most of the above are developed countries, while two developing countries, China and India, also did well in this field. The important research institutions included Rockefeller University, University of Gottingen, University of Geneva, University of Notre Dame, US Geological Survey, University of Copenhagen, Macquarie University, Harvard University, University of British Columbia and Chinese Academy of Sciences. Brady, S. F. was the leading author who contributed 43 articles related to eDNA, followed by Daniel, R. and Pawlowski, J.

Document co-citation network

A timeline co-citation cluster visualization was generated to reveal the evolution process of scientific activity of eDNA from 1992 to 2016 (Figure 2). There were eight main co-citation clusters labelled by index terms from their own citers on the right. Log-likelihood test (LLR) was chosen, because it usually gives the best result in terms of uniqueness and coverage of themes associated with a cluster. In CiteSpace, the homogeneity of each cluster is measured by a silhouette; a high silhouette value indicates greater homogeneity of the cluster. The quality of the grouping in this map is considered high due to greater high homogeneity of these clusters in the range 0.816–1, which represents a perfect differentiation among the clusters in the network[3–5]. In this map, larger node sizes imply higher co-cited frequency, and highly co-cited article indicates a landmark of the domain. In addition, nodes with red colour indicate strong citation burst, implying that the citations of these articles increased rapidly in a given time period.

As shown in Figure 2, cluster #3 is the oldest cluster with mean year 1990 and research on application-driven approaches is the largest cluster (#0), having 30 members. While using environmental DNA (#2) is the youngest cluster by mean year 2012 and articles in this cluster represent the major efforts of the research in the most recent years. Many red nodes in cluster #0, cluster #1, cluster #2, cluster #4 and cluster #5, indicate that these articles have experienced or are experiencing the strongest citation bursts, which consist of the knowledge basis or research frontiers in eDNA field.

Table 2. Top 10 highly co-cited papers with co-citation frequency

Frequency	First author	Title of document	Reference
123	Ficetola, G. F.	*Species detection using environmental DNA from water samples*	8
121	Jerde, C. L.	*'Sight-unseen' detection of rare aquatic species using environmental DNA*	9
105	Thomsen, P. F.	*Monitoring endangered freshwater biodiversity using environmental DNA*	10
82	Taberlet, P.	Environmental DNA	11
74	Goldberg, C. S.	Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders	12
71	Rondon, M. R.	Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms	13
71	Dejean, T.	Persistence of environmental DNA in freshwater ecosystems	14
65	Darling, J. A.	From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments	15
65	Thomsen, P. F.	Detection of a diverse marine fish fauna using environmental DNA from seawater samples	16
64	Dejean, T.	Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*	17

Landmark papers

Table 2 shows the top 10 highly co-cited papers. The paper by Ficetola *et al.*[8] is ranked first with the most co-citations (frequency of 123) suggesting that it is the most influential paper in the eDNA field. This article is the first study of eDNA on freshwater samples and presents a novel approach to detect the presence of a species in freshwater based on limited eDNA. The paper by Jerde *et al.*[9] has 121 co-cited counts and is ranked second. It demonstrates the efficacy of eDNA as a detection tool in freshwater environments. The third paper in the list is by Thomsen *et al.*[10], which established eDNA as a tool for monitoring rare and threatened species across a wide range of taxonomic groups. The fourth paper is a review by Taberlet *et al.*[11]. Goldberg *et al.*[12] successfully used eDNA techniques to detect aquatic vertebrates (ranked fifth). The sixth paper in the list is by Rondon *et al.*[13], who constructed libraries of genomic DNA isolated directly from soil (termed metagenomic libraries). The seventh paper by Dejean *et al.*[14] deals with eDNA persistence in freshwater ecosystem. The next paper in the list by Darling and Mahon[15] reviews the application of eDNA in monitoring biological invasion. The ninth paper by Thomsen *et al.*[16] studies the potential of using metabarcoding of eDNA obtained directly from sea-water samples to account for marine fish biodiversity. The last paper by Dejean *et al.*[17] on improved detection of an alien invasive species through eDNA barcoding, has reported early detection of alien invasive species at very low densities and at any life stage.

Almost all the top most highly co-cited documents were published after 2008 (except Rondon's paper that was published in 2000), while 8 of the 10 most highly co-cited papers were published in 2011 or 2012, indicating rapidly increasing interest in eDNA research in recent years. Among the highly co-cited papers, two are reviews, one is related to soil microorganism, while the other seven research articles and a review are related to aquatic ecosystem. Furthermore, these eight highly co-cited articles are included in cluster #2, and according to their research content, we find that eDNA methods are mainly used in aquatic environment, focusing on monitoring of biodiversity, rare species, biological invasions and so on. These research directions represent the current hot topics of interest. Also, Thomsen, P. F. and Dejean, P. T. have two papers among the top 10 highly co-cited articles, which shows their scientific influence in the eDNA field.

References with strongest citation bursts

Citation burst is an important parameter to indicate whether a certain research topic is hot or not. Generally, significant increases in research interests within a specific knowledge domain are characterized by publications that have experienced citation bursts. Table 3 lists top 10 references with strongest citation bursts.

The paper by Rondon[13] has the strongest citation burst references (burst strength of 25.58) followed by Thomsen[10] (24.00) and Henne[18] (22.64), the fourth to tenth are Jerde's paper[9] (21.78), Ficetola's paper[8] (20.30), Venter's paper[19] (19.52), Henne's paper[20] (19.37), Pilliod's paper[21] (18.60), Taberlet's paper[11] (18.14) and Thomsen's paper[16] (17.40). Interestingly, among the top 10 strongest burst references, three were published around the year 2000, while five were published post 2011, demonstrating that research interest in the eDNA field has been steadily increasing in the new millennium. On the other hand, comparing Tables 2 and 3, we find that the top 10 highly co-cited articles share 7 articles with the top 10 citation burst ones, which indicates that papers with strongest citation burst also have highly cited counts in eDNA field throughout the research history. In addition, five of the 10 strongest citation burst references are included in cluster #2, reflecting the abrupt surge of interest in the subjects of this cluster.

Table 3. Top 10 references with strongest citation bursts and the most recent strongest burst-cited references per year from 2011

Rank	First author	Title of document	Reference	Strength	Begin	End	Cluster #
1	Rondon, M. R.	Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms	13	25.58	2001	2008	1
2	Thomsen, P. F.	**Monitoring endangered freshwater biodiversity using environmental DNA**	10	24.00	2013	2016	2
3	Henne, A.	Screening of environmental DNA Libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*	18	22.64	2001	2008	1
4	Jerde, C. L.	**'Sight-unseen' detection of rare aquatic species using environmental DNA**	9	21.78	2012	2016	2
5	Ficetola, G. F.	**Species detection using environmental DNA from water samples**	8	20.30	2011	2016	2
6	Venter, J. C.	Environmental genome shotgun sequencing of the Sargasso Sea	19	19.52	2005	2012	0
7	Henne, A.	Construction of environmental DNA Libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-Hydroxybutyrate	20	19.37	2000	2007	4
8	Pilliod, D. S.	**Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples**	21	18.60	2014	2016	2
9	Taberlet, P.	Environmental DNA	11	18.14	2013	2016	5
10	Thomsen, P. F.	Detection of a diverse marine fish fauna using environmental DNA from seawater samples	16	17.40	2014	2016	2

Table 4. Countries/institutions/authors/keywords with the most bursts post 2011

Countries/institutions/authors/keywords	Strength	Begin	End
Country			
Norway	3.5858	2012	2014
Institutions			
US Geological Survey	4.7095	2013	2016
University of Notre Dame	4.6086	2013	2016
Authors			
Willerslev, E.	4.1074	2012	2013
Mahon, A. R.	4.331	2013	2016
Jerde, C. L.	4.331	2013	2016
Pawlowski, J.	3.6245	2014	2016
Keywords			
Diet	3.4669	2011	2016
DNA barcoding	4.6937	2012	2016
Barcode	3.686	2012	2016
Sequence	3.9112	2012	2013
Ancient DNA	5.4127	2012	2016
Real time PCR	3.7935	2013	2016
Conservation	8.9273	2014	2016
Ecology	5.0131	2014	2016
Occupancy	5.116	2014	2016
Biodiversity	15.3536	2014	2016
Water sample	12.1088	2014	2016
Persistence	3.7208	2014	2016
Metabarcoding	7.1187	2014	2016
Extracellular DNA	5.116	2014	2016
Biodiversity assessment	5.4481	2014	2016
Abundance	6.7506	2014	2016

References with most recent citation bursts

The most recent citation burst documents might become new intellectual turning point documents in near future. The most recent citation bursts each year from 2011 onwards are summarized in Table 3 (see table 3 with bold titles). It can be observed that one article published in 2008 started its citation bursts in 2011, and one article published in 2011 started its citation bursts in the same year. Citation burst starting from 2013 is associated with a 2012 article; while citation burst starting from 2014 is associated with a 2013 article. These findings indicate that the citation frequencies of these documents have increased rapidly in recent years and that their research contents are hot topics and therefore worthy of much more interest and attention.

It can be seen that three out of the four articles with the most recent citation bursts are among the top ten highly co-cited papers (see Table 2 with italic titles). These are the articles by Ficetola *et al.*[8], Dejean *et al.*[9] and Thomsen *et al.*[10], ranked 1 to 3. Moreover, all the four articles with the most recent citation bursts are in cluster #2. Since a cluster includes many nodes with stiff citation bursts usually implying a new active area of research, undoubtedly, these articles represent not only the landmarks in eDNA domain, but also pivotal nodes and the most recent hotspots in the near future.

Most recent active countries, institutions, authors and topics

Burst detection using CiteSpace can also effectively identify an active entity by investigating whether the frequency of an node (country, institution, author, cited reference, keyword, etc.) increases abruptly during a given time. A country, institution, author or keyword with burst shows a strong surge of frequency in a number of articles. Here, we are particularly interested in those bursts starting from 2011 onwards, which can reveal the most recent active countries, institutions, authors, and can also identify emerging topics through keywords and cited references. As shown in Table 4, the country with the most recent strongest burst from 2011 onwards is Norway, with burst strength of 3.5858, from 2012 to 2014, while US Geological Survey and University of Notre Dame are the most recent burst institutions, both from 2013 to 2016, while Willerslev, E., Mahon, A. R. and Jerde, C. L. are the most productive authors in recent years, indicating they are the most active contributors with an abrupt increase in publications related to eDNA. Keywords with burst since 2011 include diet, DNA barcoding and so on. Biodiversity (starting from 2014) shows maximum strength followed by water sample. For all these keywords, the burst lasts till 2016, and they could represent the most recent research trends in the eDNA field.

Conclusion

In this article, we have used scientometric methods to assess the progress achieved in eDNA research based on SCI-EXPANDED and Social Sciences Citation Index databases for the period 1992–2016. According to the network visualization, document co-citation analysis and burst detection supported by CiteSpace, we explored the key clusters of articles and identified research state and emerging trends in the literature.

Results indicate that papers related to eDNA have increased steadily with time and will continue to grow. Publications are widely distributed in a large number of source journals, covering various subject categories, and published by authors around the world, which indicates that eDNA research has attracted increasing attention.

---

1. Ogram, A., Sayler, G. S. and Barkay, T., The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods*, 1987, **7**, 57–66.
2. Barnes, M. A. and Turner, C. R., The ecology of environmental DNA and implications for conservation genetics. *Conserv. Genet.*, 2015, **17**(1), 1–17.
3. Chen, C., Searching for intellectual turning points: progressive knowledge domain visualization. *Proc. Natl. Acad. Sci., USA*, 2004, **101** (Suppl), 5303–5310.
4. Chen, C., CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Technol.*, 2006, **57**(3), 359–377.
5. Chen, C. *et al.*, Towards an explanatory and computational theory of scientific discovery. *J. Informetr.*, 2009, **3**(3), 191–209.
6. Chen, C. *et al.*, Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. *Expert Opin. Biol. Ther.*, 2012, **12**(5), 593–608.
7. Chen, C., Dubin, R. and Kim, M. C., Emerging trends and new developments in regenerative medicine: a scientometric update (2000–2014). *Expert Opin. Biol. Ther.*, 2014, **14**(9), 1295–1317.
8. Ficetola, G. F. *et al.*, Species detection using environmental DNA from water samples. *Biol. Lett.*, 2008, **4**(4), 423–425.
9. Jerde, C. L. *et al.*, 'Sight-unseen' detection of rare aquatic species using environmental DNA. *Conserv. Lett.*, 2011, **4**(2), 150–157.
10. Thomsen, P. F. *et al.*, Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.*, 2012, **21**(11), 2565–2573.
11. Taberlet, P. *et al.*, Environmental DNA. *Mol. Ecol.*, 2012, **21**(8), 1789–1793.
12. Goldberg, C. S. *et al.*, Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PLoS ONE*, 2011, **6**(7), 3007–3024.
13. Rondon, M. R. *et al.*, Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.*, 2000, **66**(6), 2541–2547.
14. Dejean, T. *et al.*, Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE*, 2011, **6**(8), e23398.
15. Darling, J. A. and Mahon, A. R., From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environ. Res.*, 2011, **111**(7), 978–988.
16. Thomsen, P. F. *et al.*, Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, 2012, **7**(8), e41732.
17. Dejean, T. *et al.*, Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *J. Appl. Ecol.*, 2012, **49**(4), 953–959.
18. Henne, A. *et al.*, Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.*, 2000, **66**(7), 3113–3116.
19. Venter, J. C. *et al.*, Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 2004, **304**(5667), 66–74.
20. Henne, A. *et al.*, Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.*, 1999, **65**(9), 3901–3907.
21. Pilliod, D. S. *et al.*, Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Can. J. Fish Aquat. Sci.*, 2013, **70**(8), 1123–1130.

ACKNOWLEDGEMENTS. This study was partially supported by the Natural Science Foundation Project of CQ CSTC (cstc2012jjA00011). L.J. conceived the research and retrieved the data, and prepared the tables and figures. L.J. and Y.Y. wrote the manuscript.

Received 15 August 2016; revised accepted 15 November 2016

doi: 10.18520/cs/v112/i08/1659-1664