

Highly Cited Articles in the Field of Dna Barcode: A Bibliometric Analysis

Fugang Ren¹, Jing Huang² & Yongqing Yang^{3*}

¹College of Medical Technology, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

²Department of Vocal Performance, Sichuan Conservatory of Music, Chengdu 610021, China

³College of Life Science, Chongqing Normal University, Chongqing 401331, China

***Correspondence to:** Dr. Yongqing Yang, College of Life Science, Chongqing Normal University, Chongqing 401331, China.

Copyright

© 2022 Dr. Yongqing Yang, *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 21 June 2022

Published: 19 August 2022

Keywords: *Citation Analysis; DNA Barcode; Y-Index*

Abstract

Identification of species by short orthologous DNA sequences was known as DNA barcodes. In this study, identification and analysis were performed to get the properties of the top-cited articles in the field of DNA barcodes based on the Science Citation Index Expanded. The parameters include journals, publication years, Web of Science categories, authors, institutions, countries, and citation life cycle curve. Y-index is applied to evaluate the publication properties of authors of top-cited articles. As for the citations, citations in 2018, total citations, and citations in the publication year were concerned to analyze the top-cited articles in DNA barcodes. The top three journals were Molecular Ecology Resources, Proceedings of the National Academy of Sciences of the United States of America, and Plos One. The three most active authors in DNA barcode research were P.D.N. Hebert, M. Hajibabaei, and M.A. Smith. University of Guelph of Canada was the most productive institution in DNA barcode research. The mainstream countries have controlled

this research field with several publications. Some keywords such as “barcodes”, “identification”, “taxonomy”, “evolution”, “diversity”, and “mitochondrial” appeared on the top keywords in all three types of keywords analysis: Words in titles, author keywords, and KeyWords Plus.

Introduction

DNA barcode is defined as one or few relatively short DNA sequences taken from a standardized portion of the genome. The term DNA barcode was first applied in 1993 by Arnot *et al.* [1], wherein the circumsporozoite gene was employed as a barcode to distinguish isolates of *Plasmodium falciparum*. In 2003, Hebert *et al.* first proposed “barcoding DNA” as a way to identify species, and they established the mitochondrial gene cytochrome c oxidase I (COI) as the core of a global bio-identification system for animals [2]. Since then, DNA barcoding approach has been successful used for identification of algae [3,4], fungi [5,6], land plants [7-9] and many animal groups, such as Nematodes [10], insect [11], spider [12], fish [13,14], birds [15] and mammals [16]. Nowadays, DNA barcode is considered the best tool for species taxonomy and identification. In particular, due to its universality and versatility, DNA barcode is becoming an increasingly powerful tool for other branches of biology such as conservation biology, ecology, medicine, pharmaceuticals, food science and systems biology [17,18]. Research on DNA barcode has achieved significant progress and lots of research articles have been published by authors from all over the world in a large number of journals during the recent decades. More than three hundred of reviews related to DNA barcode have been published in order to provide the readers with different kinds of information. However, the vast majority of these research focuses on just one subject, for instance, DNA barcode application on species identification [19-25], on conservation biology [26,27], on food science [28], on ecology [18] and on pharmacy [25,29], but with few research papers considering the whole field by using the bibliometrics methods. Bibliometrics studies are important approaches to make comprehensive evaluation of the development of regarding research area and have been widely conducted in many natural and social sciences [30-35]. In this study, we focus only on the bibliometric analysis of DNA barcodes based on highly cited articles. Highly cited articles generally represent the most significant developments in a specific field, thus providing a substantial and valuable insight into which authors, articles, and topics are motivating the research field over time [39]. Previous bibliometrics researchers revealed the characteristics of highly cited articles on specific topics, such as ecosystem services [37], biomass [38], and electrocardiogram [38]. However, to our knowledge, no study has attempted to analyze the highly cited articles in the field of DNA barcode research.

Therefore, the objective of this paper is to make a comprehensive study of highly cited papers in the web of science (WOS) database in the DNA barcode field with respect to the most influential scientists, institutions, and countries. Based on these results, we can reveal the profile of scientific advancement and give a historic perspective on scientific progress, thus help scientific researchers better understand the high impact research in the DNA barcode field.

Methods

Articles

Initially, we searched a total of 9,711 documents comprising 13 document types, including article, review, meeting abstract, editorial material, proceedings paper, letter, correction, early access, news item, book chapter, data paper, reprint, and retracted publication. But only the document type “article” was accepted for this study.

Citations

Two additional filters, total citations since their publication to September 24 of 2019, TC2019, and the “front page” [39], were employed to select the top-cited articles. The articles selected by TC2019 ≥ 100 were regarded as top-cited articles. The total number of times an article was cited from its publication until the September of 2019 was marked as TC2019.

Front Page

The other filter, the “front page” [39], was used to distinguish articles with the indicated keywords (DNA barcod*) on their “front page” only, including the article title, abstract, and keyword parts. Articles found only by KeyWords Plus were removed. At last 217 articles (2.48% of the 8,759 total articles) were screened as top-cited articles. These records were downloaded into spreadsheet software and additional coding was manually applied by Microsoft Excel 2016 [40].

UK

The items of country downloaded from WoS need to be organized before analysis. Articles originating from England, Scotland, Northern Ireland, and Wales belong to the United Kingdom (UK) [41].

VOSviewer software (version 1.6.11, www.vosviewer.com/) was employed to map Co-citation of authors and journals references, as well as co-occurrence of authors’ keywords. Co-citation means that two authors/ journals /references are cited together in a citing reference. In these maps, each node represents one entity (author/ journal or keyword). Larger node sizes imply that the entity is an important one within the knowledge domain. Nodes share the same color belonged to one cluster. The links between nodes represented the co-citation or co-occurrence relationships, a shorter distance and/or a thicker link between nodes reveals a stronger relation, which is also numerically represented as relative link strength. The total link strength attribute indicates the total strength of the co-citation/co-occurrence links of a given item with other items.

Results and Discussion

1. Distribution of Top-Cited Articles

A total of 217 articles are selected as top-cited articles (articles obtained greater than or equal to 100 citations from their time of publication to September of 2019) which were published between 2000 and 2016. Figure 1 illustrates the number of top-cited articles and citations per publication by year which illustrated that most of the top-cited articles were published between 2005 and 2012. Most top-cited articles were published in 2009. There were no top-cited articles in the years 2001, 2002, 2017, 2018, and 2019.

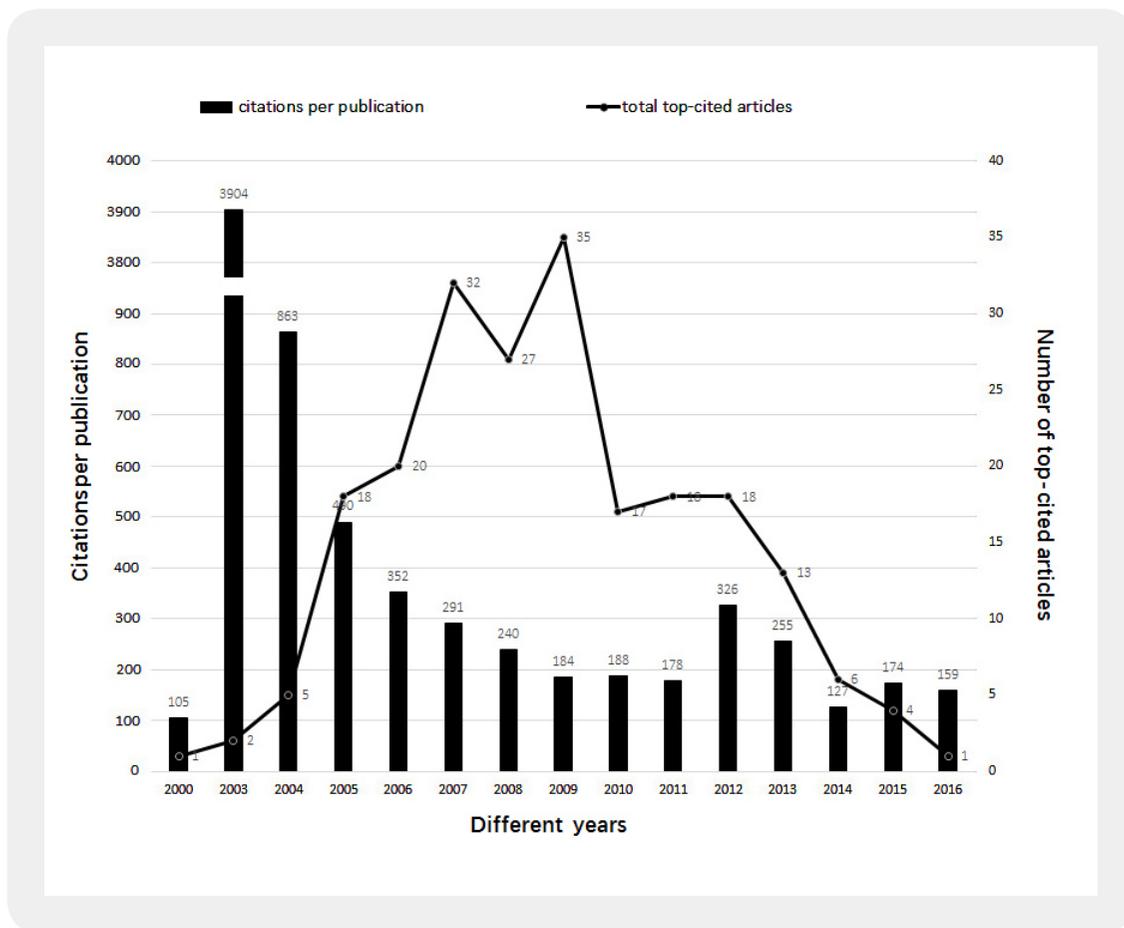


Figure 1: Number of articles and citations per publication by years.

In all, 1557 authorships contributed to the 217 top-cited articles and the average number of authors in every top-cited article were 7.2 Small groups of authors appear in highly cited articles, for example 1 in Environmental Sciences and 3 in Biodiversity Conservation A large group of authors were also found in highly cited articles in Mycology with 27.8. Totally 11,481 references were cited by the 217 top-cited

articles. An average of 13.2 pages in every top-cited article were applied to publish the research findings. Furthermore, there was a wide variation in pages from 2 to 379 and in references from 5 to 547 in every article.

2. Journals and Co-citation Journals

These top-cited articles were published in 73 journals. Only one top-cited article was published per journal by 32(63%) Journals and two articles were published per journal by 11 (15%) journals and 120 (55.3%) top-cited articles were published by nine core journals (Table 1). The impact factor (IF2019) of the top nine journals exists between 2.776 and 10.266.

Table 1: Top nine core journals on DNA barcode

Journal	TP (%)	IF2019	Web of Science category
Molecular Ecology Resources	28(12.9)	7.049	Biochemistry & Molecular Biology; Ecology; Evolutionary Biology
Proceedings Of The National Academy Of Sciences Of The United States Of America	21(9.7)	9.580	Multidisciplinary Sciences
Plos One	19(8.8)	2.776	Multidisciplinary Sciences
Philosophical Transactions Of The Royal Society B-Biological Sciences	12(5.5)	6.139	Biology
Molecular Ecology Notes	9(4.1)	NA	Biochemistry & Molecular Biology; Ecology; Evolutionary Biology
Systematic Biology	9(4.1)	10.266	Evolutionary Biology
Molecular Ecology	8(3.7)	5.855	Biochemistry & Molecular Biology; Ecology; Evolutionary Biology
Proceedings Of The Royal Society B-Biological Sciences	7(3.2)	4.304	Biology; Ecology; Evolutionary Biology
New Phytologist	7(3.2)	7.299	Plant Sciences

TP: total number of articles; IF2019: impact factor in 2019

The Journal *Molecular Ecology Resources* (IF2019=7.049) published the most top-cited articles with 28 articles (12.9% of 217 articles), followed by *Proceedings of The National Academy of Sciences of The United States of America* (IF2019=9.580) with 21 articles (9.7%), *Plos One* (IF2019=2.776) with 19 articles (8.8%), *Philosophical Transactions of The Royal Society B-Biological Sciences* (IF2019=6.139) 12 articles(5.5%), *Molecular Ecology Notes* (IF2019=not available) with 9 articles(4.1%), *Systematic Biology* (IF2019=10.266) with 9 articles(4.1%), *Molecular Ecology* (IF2019=5.855) with 8 articles (3.7%), *Proceedings of The Royal Society B-Biological Sciences* (IF2019=4.304) and *New Phytologist* (IF2019=7.299) with seven articles (3.2%) respectively.

However, top-cited articles were also published in journals with low-impact factors, for example, *Zootaxa* (IF2019=0.99), *Journal of Crustacean Biology* (IF2019=1.069), and *Journal of Aoac International* (IF2019=1.201). Of course, three high-impact factor journals also published the top-cited articles, such as *Science* (IF2019 =41.037) with one article, *Cell* (IF2019 =36.216) with one article, and *Nature Biotechnology* (IF2013 =31.864) with one article.

The top-cited articles were published in the 30 Web of Science subject categories in the science edition. Among these, one-fifth of the categories published each one article and half categories published the top-cited articles between 2 and 6. The top 9 (30%) categories published the top-cited articles between 7 and 84. The five top categories contain *Evolutionary Biology* (84 articles; 38.7% of 217 articles), *Ecology* (64; 29.5%), *Biochemistry & Molecular Biology* (62; 28.6%), *Multidisciplinary Sciences* (42; 19.4%), and *Biology* (34; 15.7%).

These five categories published a majority of the total top-cited articles (159 articles; 73.2% of 217 articles). One thing should be noticed that some journals may be sorted into two or more categories in WoS, for instance, Molecular Ecology exists in both categories of “*Biochemistry & Molecular Biology*”, “*Ecology*” and “*Evolutionary Biology*”. It illustrates the multidisciplinary character of this research field.

VOSviewer uses cluster algorithms based on the strengths of the connections among items to analyze the network. In these maps, the total link strength value denotes the importance of an item in the field since a higher value means that it has been linked with others many times. A map of the journal co-citation demonstrates that the journals are divided into four clusters (Figure 2.) The blue cluster contains *Proceedings of the National Academy of Sciences of the United States of America*, *Plos One*, *American Journal of Botany*, *Bioinformatics*, *BMC Evolutionary Biology*, *Molecular Ecology Resources*, *Taxon*, representing journals in evolution and taxa. The green cluster contains *Applied and Environmental Microbiology*, *Canadian Journal of Botany*, *Journal of Phycology*, *Mycological Research*, *Mycologia*, *Studies in Mycology*, *New Phytologist*, and *Nucleic Acids Research*, representing journals in algae and fungi. The yellow cluster contains *Nature* and *Science* representing journals of high impact factor and *Multidisciplinary Sciences*. The red cluster contains half of the total journals, including *Annual Review of Ecology Evolution and Systematics*, *Biological Journal of the Linnean Society*, *Canadian Journal of Zoology*, *Cladistics*, *Conservation Biology*, *Evolution*, *Journal of Molecular Evolution*, *Molecular Biology and Evolution*, *Molecular Ecology*, *Molecular Ecology Notes*, *Molecular Phylogenetics and Evolution*, *Proceedings of The Royal Society B-Biological Sciences*, *Philosophical Transactions of the Royal Society B-Biological Sciences*, *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *Plose Biology*, *Systems Biology and Trends in Ecology & Evolution*, represents journals in evolution and ecology.



Figure 2: Cluster density map of co-citation analysis for journals. Journals with minimum 50 co-citations were included where only 34 met the threshold. Journals share the same color belongs to a cluster.

Obviously, among these journals, *Proceedings of the National Academy of Sciences of the United States of America* had the highest total link strength of 434.15, and the highest citations of 484 indicating that this journal was co-cited with most other journals, followed by *Systems Biology* (total link strength of 340.16), and *Molecular Ecology* (total link strength of 330.48).

3. Most Productive and Co-citation Authors

Among the 1105 authors responsible for 217 top-cited articles, (176 first authors) 929 authors (84.1% of 1105 authors) had no first author articles, (164 corresponding authors) 941 (85.2%) authors had no corresponding author articles and only 130 (11.8%) authors had both first author articles and corresponding author articles. Table 2 illustrates the total top-cited articles (TP), first-author top-cited articles (FP), corresponding-author top-cited articles (RP); and single-author top-cited articles (SP) of leading authors with at least six top-cited articles.

Table 2: Characteristics of the 17 most – productive authors (≥ 6 articles)

Author	Rank (TP)	Rank (FP)	Rank (RP)	Rank (SP)
P.D.N. Hebert	1 (43)	1 (5)	1 (8)	N/A
M. Hajibabaei	2 (15)	1 (5)	2 (6)	N/A
D.H. Janzen	3 (11)	9 (2)	6 (3)	N/A
P. Taberlet	4 (10)	27 (1)	13 (2)	N/A
W. Hallwachs	5 (9)	N/A	N/A	N/A
E. Coissac	6 (8)	N/A	30 (1)	N/A
M.A. Smith	7 (7)	1 (5)	3 (5)	N/A
S.G. Newmaster	7 (7)	4 (4)	4 (4)	N/A
S. Ratnasingham	9 (6)	27 (1)	N/A	N/A
R.D. Ward	9 (6)	5 (3)	4 (4)	2 (1)
N.V. Ivanova	9 (6)	9 (2)	13 (2)	N/A
S. Chen	9 (6)	27 (1)	6 (3)	N/A
G.W. Saunders	9 (6)	5 (3)	6 (3)	1 (2)
A.P. Vogler	9 (6)	27 (1)	13 (2)	N/A
J.Z. Groenewald	9 (6)	27 (1)	30 (1)	N/A
P.W. Crous	9 (6)	9 (2)	13 (2)	N/A
T.G. Barraclough	9 (6)	N/A	30 (1)	N/A

TP: total top - cited articles; FP: first - author top - cited articles; RP: corresponding - author top - cited articles; SP: single - author top - cited articles; N/A: not available

The top three authors of top-cited articles are P.D.N. Hebert from the University of Guelph in Canada with 43 articles, M. Hajibabaei from the University of Guelph in Canada with 15 articles, and D.H. Janzen from the University of Pennsylvania in the USA with 11 articles. G.W. Saunders and R.D. Ward published single-author articles with two and one respectively.

The definition of the first and corresponding author provides a powerful method to take the multiple authorships into account [42]. The first author conducts the research work and writes the research paper so that this author makes the most contribution to the paper [43,44]. The corresponding author supervises the planning and executes the research work and write the research paper [45,46]. The corresponding author contributes to the study by increasing the author's credit [47]. The country or institution of the corresponding author is usually the origin of the selected field study [48].

The recently developed Y-index [48] has been applied in many studies to assess the performance of highly cited authors, such as the highly cited articles in the research field of volatile organic compounds [49]. The Y-index (j, h) was employed to assess authors of the top-cited articles [48]. This index is concerned to the

numbers of first-author publications (FP) and corresponding-author publications (RP), the formulae are as follows [50,51]:

$$j = FP+RP \quad (1)$$

$$h = \tan^{-1}(RP/FP) \quad (2)$$

An author with a higher j means more first- or corresponding-author articles, contributing to leadership. j illustrates that one author published more articles as the important author [51]. h is a publication property constant that describes the proportion of FP to RP [42].

Values of $h > 0.7854$ mean more corresponding-author articles, and values of $h < 0.7854$ mean more first-author articles. When $h=0.7854$, it illustrates the author has the same number of first-author articles and corresponding-author articles. If $h=0$, j =the number of first-author articles, and if $h=\pi/2$, j =the number of corresponding-author articles [51].

Figure 3 indicates the distribution of the top 20 authors ($j \geq 4$) with a Y-index from the top-cited 217 articles.

These 20 authors were deemed to be the main contributors to the top-cited articles in DNA barcode research. Their contribution included conception and design, analysis and interpretation of data and the drafting or reviewing of the article [52,53].

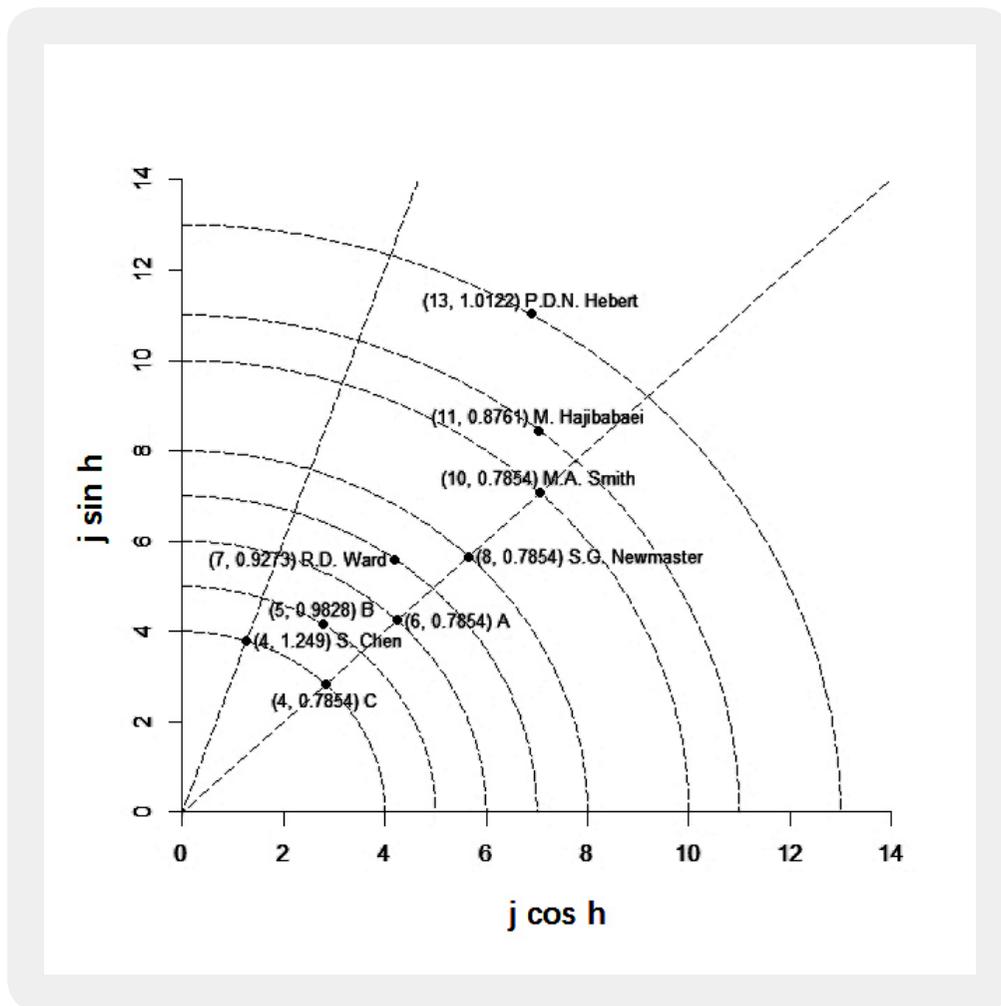


Figure 3: Distribution of the top 20 authors with their Y-index values ($j \geq 4$).

In Figure 3, each dot stands for one value that could be a single author or several authors when they had the same publication intensity and properties [51]. The three most productive authors in DNA barcode research were P.D.N. Hebert ($j=13$), M. Hajibabaei ($j=11$), and M.A. Smith ($j=10$). Publication property h indicates the different proportion of first author articles to corresponding author articles. It is useful especially if j of authors is similar to recognize the different contributions of authors. Among these 20 authors, six authors: P.D.N. Hebert ($h=1.0122$), M. Hajibabaei ($h=0.8761$), R.D. Ward ($h=0.9273$), P.M. Hollingsworth ($h=0.9828$), D.H. Janzen ($h=0.9828$) and S. Chen ($h=1.2490$) had more corresponding author articles than the first author articles. No authors had more first author articles than corresponding author articles. Fourteen authors exist on the boundary line because of the same numbers of first author articles and corresponding author articles with $h=0.7854$. Although the fourteen authors had the same h values of 0.7854, the higher values of j indicate that M.A. Smith ($j=10$) and S.G. Newmaster ($j=8$) had higher publication intensity.

In addition, the results of Figure 4 show that P.D.N. Hebert is the most co-cited author with the highest total link strength (333.52). The second and third most co-cited authors were M. Hajibabaei and R.D. Ward, with total link strength of 115.47 and 88.43, respectively. These authors could be considered the core authors of this network because they developed research co-cited with nearly all other top 31 authors.

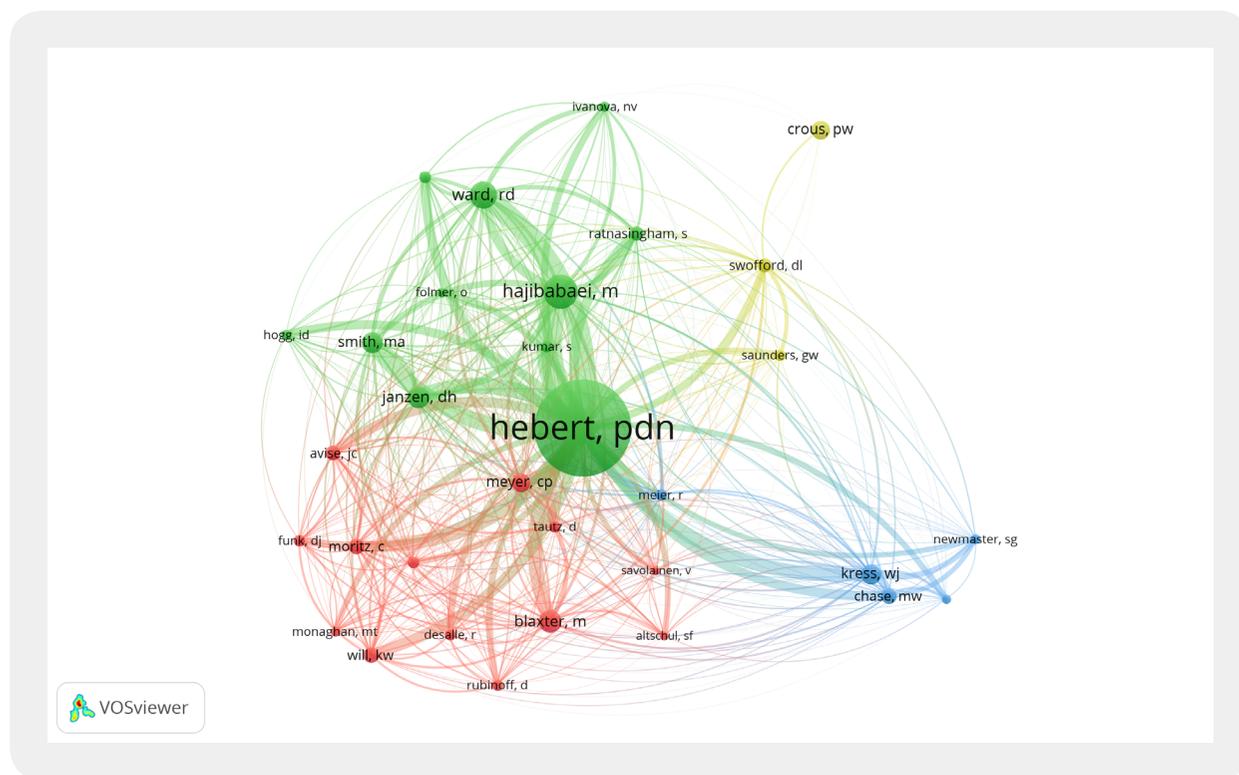


Figure 4: Network visualization of co-citation analysis for authors. Authors with minimum 30 co-citations were included where only 32 met the threshold. Authors share the same color belongs to a cluster.

4. Distribution of Institutions

Of the 217 top-cited articles, only one article had no corresponding author affiliation information and three articles had no affiliation information on the Web of Science. In the 214 top-cited articles with author affiliations, 31 (14.5% of 214 articles) were single institution articles and 183 (85.5%) were inter-institutionally collaborative articles.

A small proportion of institutions published a high proportion of top-cited articles which was similar to the research field of volatile organic compounds [49]. Table 3 indicates the properties of the top 11 most productive institutions by six indicators: total number of top-cited articles (TP), single institution top-cited articles (IP), inter-institutionally collaborative top-cited articles (CP), first author top-cited articles (FP), corresponding author top-cited articles (RP), and single author top-cited articles (SP).

Table 3: Characteristics of the 11 most productive institutions

Institution	TP	TP R (%)	IP R (%)	CP R (%)	FP R (%)	RP R (%)	SP R (%)
Univ Guelph, Canada	63	1 (29.4)	1 (3.7)	1 (25.2)	1 (15.9)	1 (16.4)	N/A
Smithsonian Inst, USA	14	2 (6.5)	N/A	3 (5.6)	3 (2.3)	2 (2.3)	N/A
Nat Hist Museum, UK	13	3 (6.1)	N/A	2 (6.1)	8 (1.4)	10 (1.4)	N/A
Royal Bot Gardens, UK	11	4 (5.1)	4 (0.5)	6 (4.7)	5 (1.9)	5 (1.9)	N/A
Univ Penn, USA	11	4 (5.1)	N/A	4 (5.1)	19 (0.9)	10 (1.4)	N/A
Agr & Agri Food Canada, Canada	11	4 (5.1)	4 (0.5)	6 (4.7)	8 (1.4)	10 (1.4)	2 (0.5)
Univ London Imperial Coll Sci Technol & Med, UK	11	4 (5.1)	N/A	4 (5.1)	8 (1.4)	5 (1.9)	N/A
Univ Calif Berkeley, USA	10	8 (4.7)	4 (0.5)	8 (4.2)	3 (2.3)	2 (2.3)	2 (0.5)
Univ New Brunswick, Canada	9	9 (4.2)	2 (0.9)	15 (2.3)	2 (3.3)	2 (2.3)	1 (0.9)
Univ Grenoble 1, France	9	9 (4.2)	N/A	8 (4.2)	5 (1.9)	5 (1.9)	N/A
Univ Florida, USA	8	11(3.7)	4 (0.5)	10 (3.3)	19 (0.9)	21 (0.9)	N/A

TP: total top – cited articles; IP: single - institution top - cited articles; CP: inter - institutionally collaborative top - cited articles; FP: first - author top - cited articles; RP: corresponding - author top - cited articles; SP: single - author top - cited articles; R: rank; N/A: not available

University of Guelph of Canada (63 articles) was ranked first on the list, Smithsonian Institution of the USA (14 articles) and Natural History Museum of the UK (13 articles) ranked second and third. Of the top 11 most productive institutions, four were situated in the USA followed by Canada (3 institutions) and the UK (3 institutions), and France (1 institution). University of Guelph of Canada was ranked No.1 in all the indicators except for single-author top-cited articles. University of New Brunswick of Canada was ranked 1st in SP and 2nd in IP, FP, and RP, and 15th in CP, which implied that the scientists in this organization were willing to contribute top-cited articles without collaboration. Smithsonian Institution of the USA and Natural History Museum of the UK were ranked 2nd and 3rd in CP, however, these two institutions were not available in IP, which suggested that the scientists from these institutions published articles in the way of inter-institutional collaboration.

5. Distribution of Countries

The 214 top-cited articles containing author affiliations were published by 55 countries or regions, of which 102 (48% of 214 articles) were single country articles and 112 (52%) were internationally cooperative articles. The countries were ranked based on the number of total top articles published as indicated in Table 4. The Rank comprised the total number of articles (TP), single country articles (IP), internationally collaborative articles (CP), first author articles (FP), corresponding author's articles (RP), and single author articles (SP).

Table 4: Characteristics of the contributing countries

Country	TP	TP R (%)	IP R (%)	CP R (%)	FP R (%)	RP R (%)	SP R (%)
Canada	88	1 (41)	1 (15)	1 (26)	1 (29)	1 (28)	1 (6.5)
USA	80	2 (37)	2 (12)	2 (25)	2 (19)	2 (20)	2 (3.3)
UK	39	3 (18)	3 (4.2)	3 (14)	3 (10)	3 (11)	3 (1.9)
Germany	27	4 (13)	8 (1.9)	4 (11)	4 (6.5)	4 (6.5)	5 (0.47)
France	25	5 (12)	4 (2.8)	5 (8.9)	6 (4.2)	5 (4.7)	N/A
Australia	18	6 (8.4)	4 (2.8)	7 (5.6)	9 (3.7)	9 (3.7)	5 (0.47)
Netherlands	15	7 (7.0)	9 (0.47)	6 (6.5)	6 (4.2)	7 (4.2)	N/A
China	13	8 (6.1)	7 (2.3)	10 (3.7)	5 (4.7)	5 (4.7)	N/A
Italy	12	9 (5.6)	9 (0.47)	9 (5.1)	13 (0.93)	12 (1.4)	N/A
Spain	12	9 (5.6)	N/A	7 (5.6)	10 (1.9)	10 (1.9)	N/A
New Zealand	9	11 (4.2)	4 (2.8)	19 (1.4)	6 (4.2)	7 (4.2)	4 (0.93)
Sweden	8	12 (3.7)	N/A	10 (3.7)	17 (0.47)	17 (0.47)	N/A
Norway	7	13 (3.3)	9 (0.47)	12 (2.8)	10 (1.9)	10 (1.9)	N/A
South Africa	6	14 (2.8)	N/A	12 (2.8)	17 (0.47)	N/A	N/A

TP: total top – cited articles; IP: single – country top – cited articles; CP: internationally collaborative top – cited articles; FP: first – author top – cited articles; RP: corresponding – author top – cited articles; SP: single– author top – cited articles; R: rank; N/A: not available.

Canada published the most number of articles, with 88 articles, occupying 41%, followed closely by the USA with 80 articles, the UK with 39 articles, Germany with 27 articles, and France with 25 articles. Control by core countries in the publication is not surprising, because this phenomenon has taken place in many research fields such as membrane science [54], cancer research [55], and ammonia oxidation research [56].

As far as internationally cooperative articles were concerned, 26% were published by Canada; and in terms of non-cooperative independent articles, 15% were contributed by Canada. Canada was ranked top in all fields, which illustrated Canada is the most productive country in independent and cooperative research top-cited articles. The USA and the UK were ranked second and third in all 6 indicators. Germany and France were ranked fourth and fifth in most indicators respectively. However, France had no single author top-cited articles.

6. Citation Life Cycles

Total citations of articles were extensively used in most studies [57]. Three kinds of citations: C2018, TC2019, and C0 were applied to evaluate the top-cited articles in this DNA barcode research [36].

The articles with the highest TC2019 were deemed as the most popular articles in this research field. The total number of times (TC2019) meant an article was cited from its time of publication to the September of 2019 [58].

The citation lives of the top seven articles (TC2019 >1500) are illustrated in Figure 5. Six of the seven articles were published before 2007.

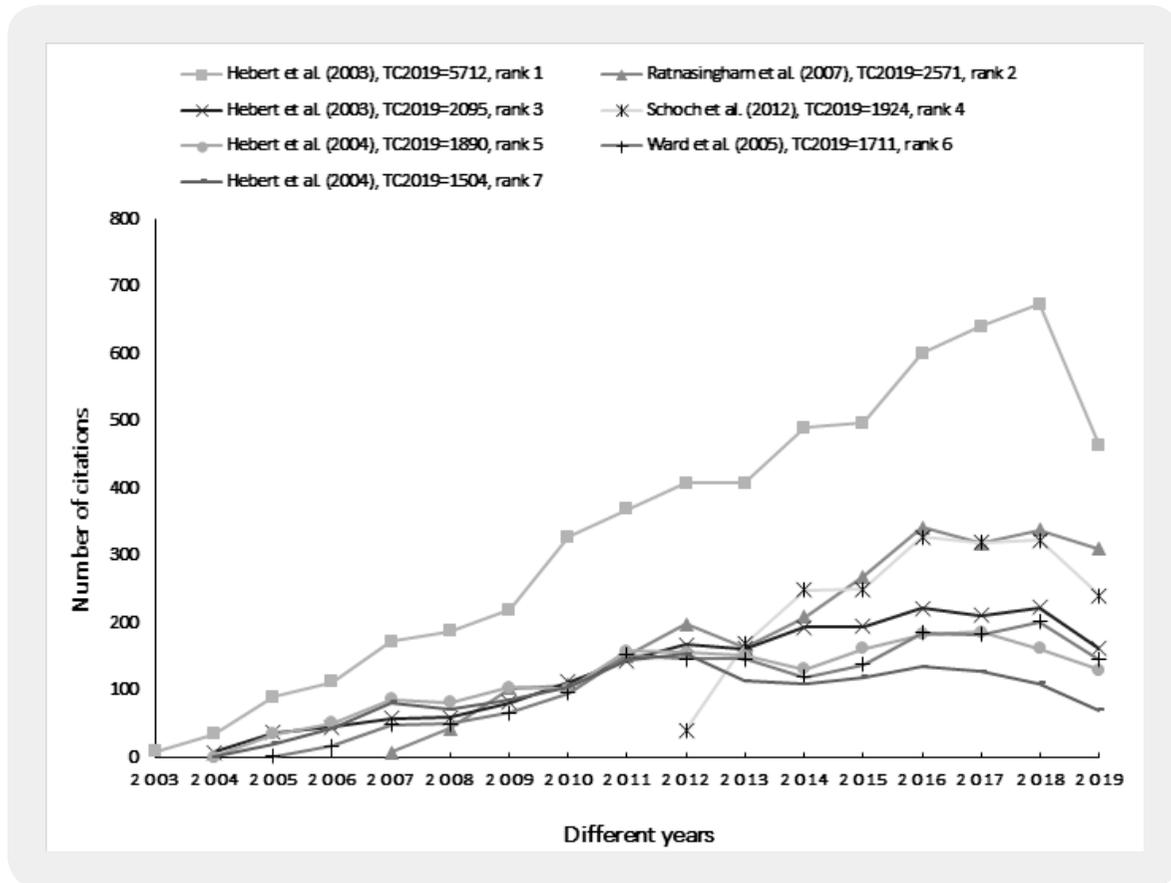


Figure 5: Citation life cycles of the top seven articles (TC2019 >1500).

Six of the top seven articles in TC2019 also appeared in the top seven in C2018 such as Hebert *et al.* [2], Ratnasingham *et al.* [59], Schoch *et al.* [6] and Ward *et al.* [60]. Of the seven articles, the top article “Biological identifications through DNA barcodes” by Hebert *et al.* published in 2003 illustrated a continual and sharp increase in citations in all years since its time of publication (Figure 5). This article is also top-ranked one in C2018 =674.

Since TC2019 is an accumulative number, it would become a large value as time flew. We needed to study the citations of an article every single year (2018) to explain the research highlight transfer in recent years. The citation life curves of the top eight articles (C2018 >160) are illustrated in Figure 6. All these articles were published after 2003. The No.1 article in the C2018 is still the paper “Biological identifications

through DNA barcodes” by Hebert *et al.* published in 2003, which demonstrate the COI identification system would supply a reliable, cost-effective, and accessible solution to the current problem of species identification and molecular evolution [2].

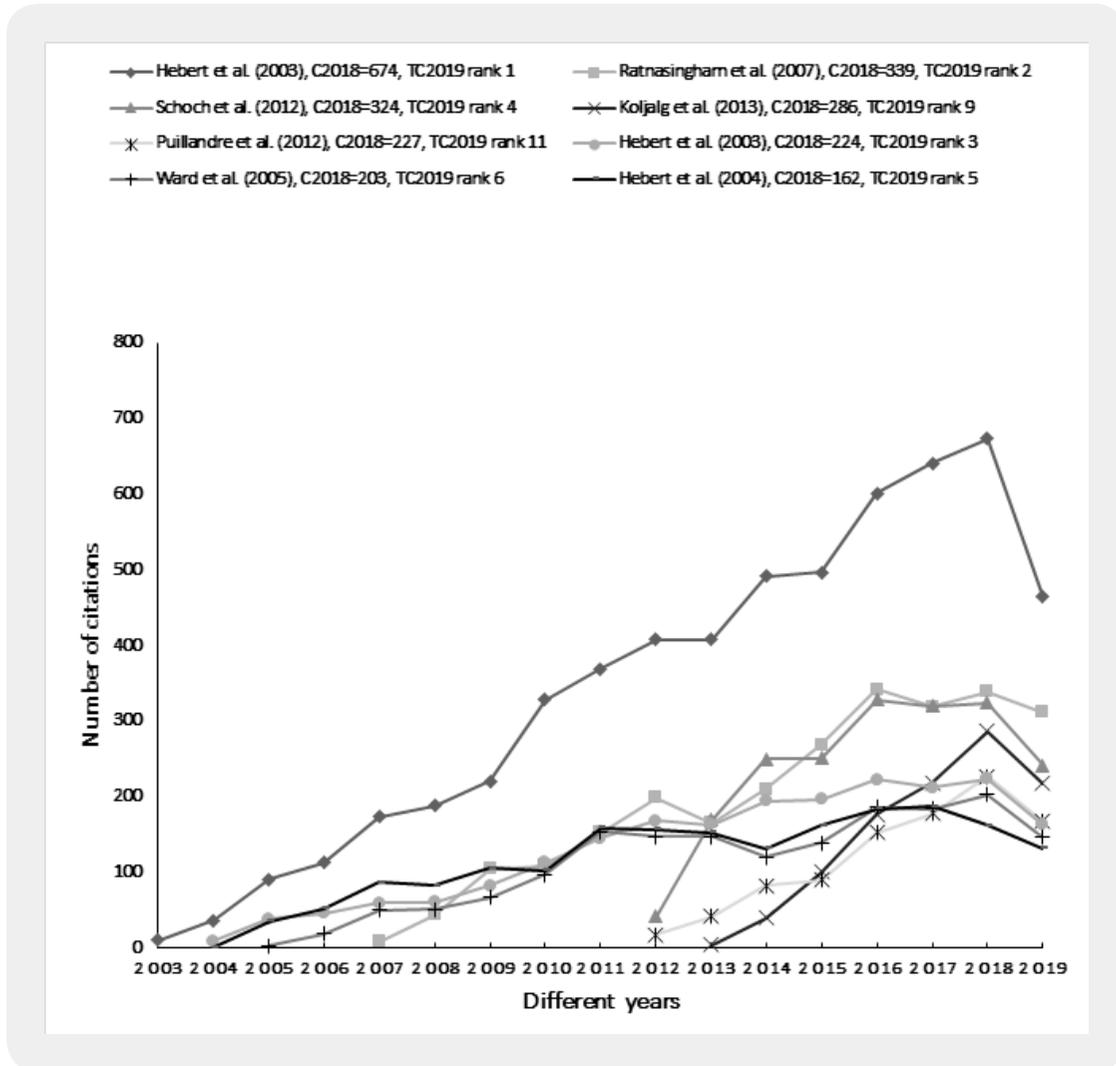


Figure 6: Citation life cycles of the top eight articles (C2018 >160)

Articles with higher citations in the publication year (C0) were applied to characterize the top-cited articles in recent years [51]. Figure 7 indicated the citation life of the top six most cited articles in their publication year (C0 ≥15). Among them, three articles were published in 2012, one article was published in 2006, 2010, and 2015 respectively.

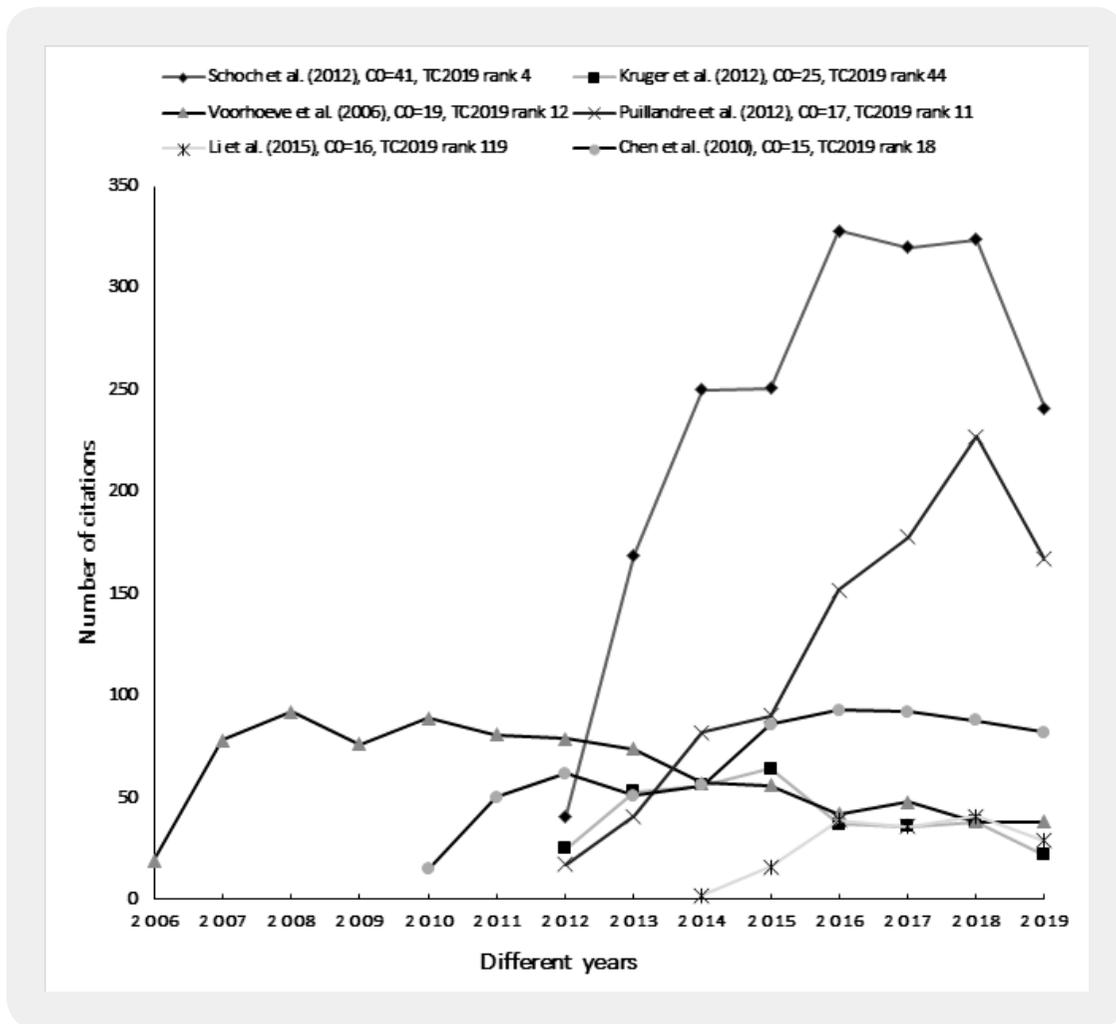


Figure 7: Citation life cycles of the top six articles (CO ≥ 15)

The No. 1 Top-cited article in the C0 list was “Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi” by Schoch *et al.* [6], which was also ranked fourth top-cited article in TC2019. Schoch *et al.* suggested the internal transcribed spacer (ITS) region as the standard barcode for fungi identification. All the articles had an increasing trend of citations after their publications then showed decreasing trends in the following years.

7. Distribution of words in titles, author keywords, and KeyWords Plus

In the past decade, distribution of words in article titles, author keywords, and KeyWords Plus were employed to assess trends of research fields [40,61].

The words in titles and author keywords provided useful information about the article subjects and contained the main message the author would describe to their readers [62]. Only 161 top-cited articles comprised

author keyword information on the Web of Science. Author keywords appearing more than six times in top-cited articles were ranked in Table 5. The most frequently author keywords were “DNA barcoding”, “COI” and “species identification”. We also investigated all the single words in the titles of top-cited articles on DNA barcode research. Some prepositions for example “of” and “with” and articles such as “A” and “the” were removed, as they meant nothing for the next study [63]. The most employed single words in titles were “DNA” and “barcoding” which existed in 148 articles followed by “species” (63), “identification” (41), “plant” (26), “fungi” (17), “biodiversity” (17), “diversity” (16) and “mitochondrial” (16). Some words and abbreviation such as ITS, taxonomy, evaluation, sequencing, phylogenetic, cytochrome, oxidase and evolution were frequently applied in the titles, which indicated that the research fields focus on identification and assessment. Among 217 top-cited articles, 10 articles had no KeyWords Plus information in Web of Science. Analysis results indicated that the keywords “identification”, “diversity”, “taxonomy”, “sequences”, “evolution”, “mitochondrial” and “barcodes” appeared highly frequently. Distribution of words illustrated that some keywords existed among the top ranks in all the forms of word analysis such as “barcodes”, “identification”, “taxonomy”, “evolution”, “diversity”, and “mitochondrial”.

Table 5: *Distribution of author keywords*

Author keyword	TP	rank	%
DNA barcoding	102	1	63
COI	64	2	40
species identification	23	3	14
biodiversity	18	4	11
taxonomy	15	5	9.3
cryptic species	12	6	7.5
molecular taxonomy	9	7	5.6
phylogeny	9	7	5.6
barcoding	8	9	5
Area de Conservacion Guanacaste	6	10	3.7
mitochondrial DNA	6	10	3.7
rbcL	6	10	3.7
LSU	6	10	3.7
Lepidoptera	6	10	3.7

TP: total number of articles

In addition, DNA barcoding-related author keywords with minimum co-occurrences of 6 were visualized and presented in Figure.5. Higher the co-occurrence frequency of two keywords, the closer the relationship.

Most articles were published in six main journals: *Molecular Ecology Resources*, *Proceedings of the National Academy of Sciences of The United States of America*, *Plos One*, *Philosophical Transactions of the Royal Society B-Biological Sciences*, *Molecular Ecology Notes*, and *Systematic Biology*. Most articles focused on five WOS categories: *Evolutionary Biology*, *Ecology*, *Biochemistry & Molecular Biology*, *Multidisciplinary Sciences*, and *Biology*.

According to the distribution of Y-index results, P.D.N. Hebert possessed the most publication potential. M. Hajibabaei and M.A. Smith published the same account article in the first authored and corresponding authored articles. University of Guelph of Canada was the most productive on top-cited articles in the DNA barcode fields. The USA published most of the top articles followed closely by Canada.

Citation life cycles (an article's publication year, in the most recent recorded year, and total citations from publication to September of 2019) indicated different rankings of top-cited articles. Results of keywords analysis illustrated that this research field was related to taxonomy and evolution.

Highly cited articles is a good indicator to characterize the high impact research, and generally represent the most significant developments in a specific field, therefore, our analysis and results revealed the profile of scientific advancement and give a historic perspective on scientific progress in the DNA barcode field.

Acknowledgements

This work was supported by the Key project of Sichuan Provincial Human Resources and Social Security Department (00809501).

Conflict of Interest

The authors report no conflict of interest

Contributions

Fugang Ren and Yongqing Yang conceived the research, prepared the tables and figures and analyzed the results, Jing Huang retrieved the data. All authors participated in writing the manuscript.

Bibliography

1. Arnot, D. E., Roper, C. & Bayoumi, R. A. L. (1993) Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates. *Mol Biochem Parasitol.*, 61(1), 15-24.
2. Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc R Soc B-Biol Sci.*, 270(1512), 313-321.

3. Saunders, G. W. (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos Trans R Soc B-Biol Sci.*, 360(1462), 1879-1888.
4. Saunders, G. W. & Moore, T. E. (2013). Refinements for the amplification and sequencing of red algal DNA barcode and RedToL phylogenetic markers: a summary of current primers, profiles and strategies. *Algae*, 28(1), 31-43.
5. Kruger, M., Stockinger, H., Kruger, C. & Schussler, A. (2009). DNA-based species level detection of Glomeromycota: one PCR primer set for all arbuscular mycorrhizal fungi. *New Phytol.*, 183(1), 212-223.
6. Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., *et al.* (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A.*, 109(16), 6241-6246.
7. Kress, W. J. & Erickson, D. L. (2007). A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One.*, 2(6), 10.
8. Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci U S A.*, 102(23), 8369-8374.
9. Starr, J. R., Naczi, R. F. C. & Chouinard, B. N. (2009). Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Mol Ecol Resour.*, 9(s1), 151-163
10. Bhadury, P., Austen, M. C., Bilton, D. T., Lamshead, P. J. D., Rogers, A. D. & Smerdon, G. R. (2006). Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. *Mar Ecol-Prog Ser.*, 320, 1-9.
11. Ball, S. L. & Armstrong, K. F. (2006). DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). *Can J For Res-Rev Can Rech For.*, 36(2), 337-350.
12. Barrett, R. D. H. & Hebert, P. D. N. (2005). Identifying spiders through DNA barcodes. *Can J Zool-Rev Can Zool.*, 83(3), 481-491.
13. Lakra, W. S., Verma, M. S., Goswami, M., Lal, K. K., Mohindra, V., Punia, P., *et al.* (2011). DNA barcoding Indian marine fishes. *Mol Ecol Resour.*, 11(1), 60-71.
14. Zemlak, T. S., Ward, R. D., Connell, A. D., Holmes, B. H. & Hebert, P. D. N. (2009). DNA barcoding reveals overlooked marine fishes. *Mol Ecol Resour.*, 9(s1), 237-242.
15. Hubert, N., Meyer, C. P., Bruggemann, H. J., Guerin, F., Komeno, R. J. L., Espiau, B., *et al.* (2012). Cryptic diversity in indo-pacific coral-reef fishes revealed by dna-barcoding provides new support to the centre-of-overlap hypothesis. *PLoS One.*, 7(3), 8.

16. Francis, C. M., Borisenko, A. V., Ivanova, N. V., Eger, J. L., Lim, B. K., Guillen-Servent, A., *et al.* (2010). The Role of DNA Barcodes in Understanding and Conservation of Mammal Diversity in Southeast Asia. *PLoS One.*, 5(9), 12.
17. Campanaro, A., Tommasi, N., Guzzetti, L., Galimberti, A., Bruni, I. & Labra, M. (2019). DNA barcoding to promote social awareness and identity of neglected, underutilized plant species having valuable nutritional properties. *Food Research International*, 115, 1-9.
18. Kress, W. J., Garcia-Robledo, C., Uriarte, M. & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends Ecol Evol.*, 30(1), 25-35.
19. Adams, E. R. & Hamilton, P. B. (2008). New molecular tools for the identification of trypanosome species. *Future Microbiol.*, 3(2), 167-176.
20. Beebe, N. W. (2018). DNA barcoding mosquitoes: advice for potential prospectors. *Parasitology*, 145(5), 622-633.
21. Bhargava, M. & Sharma, A. (2013). DNA barcoding in plants: Evolution and applications of in silico approaches and resources. *Mol Phylogenet Evol.*, 67(3), 631-641.
22. Bhattacharya, M., Sharma, A. R., Patra, B. C., Sharma, G., Seo, E. M., Nam, J. S., *et al.* (2016). DNA barcoding to fishes: current status and future directions. *Mitochondrial DNA Part A.*, 27(4), 2744-2752.
23. Bucklin, A., Steinke, D. & Blanco-Bercial, L. (2011). DNA Barcoding of Marine Metazoa Annual Review of Marine. *Science*, 3(3), 471-508.
24. Chakraborty, C., Doss, C. G. P., Patra, B. C. & Bandyopadhyay, S. (2014). DNA barcoding to map the microbial communities: current advances and future directions. *Appl Microbiol Biotechnol.*, 98(8), 3425-3436.
25. Mishra, P., Kumar, A., Nagireddy, A., Mani, D. N., Shukla, A. K., Tiwari, R. & Sundaresan, V. (2016). DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnol J.*, 14(1), 8-21.
26. Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K. & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends Ecol Evol.*, 22(3), 148-155.
27. Crawford, A. J., Cruz, C., Griffith, E., Ross, H., Ibanez, R., Lips, K. R., *et al.* (2013). DNA barcoding applied to ex situ tropical amphibian conservation programme reveals cryptic diversity in captive populations. *Mol Ecol Resour.*, 13(6), 1005-1018.
28. Galimberti, A., De Mattia, F., Losa, A., Bruni, I., Federici, S., Casiraghi, M., *et al.* (2013). DNA barcoding as a new tool for food traceability. *Food Res Int.*, 50(1), 55-63.

29. Narayana, D. B. A. & Johnson, S. T. (2019). DNA barcoding in authentication of herbal raw materials, extracts and dietary supplements: a perspective. *Plant Biotechnology Reports*, 13(3), 201-210.
30. Chen, H., Feng, Y., Li, S. S., Zhang, Y. Y. & Yang, X. X. (2019a). Bibliometric analysis of theme evolution and future research trends of the type a personality. *Personality and Individual Differences*, 150.
31. Chen, J. Y., Ye, X. D., Chen, M. & Liang, Y. (2019b). Bibliometric analysis of the papers on urban education. *Library Hi Tech.*, 37(4), 894-905.
32. Merediz-Sola, I. & Bariviera, A. F. (2019). A bibliometric analysis of bitcoin scientific production. *Research in International Business and Finance*, 50, 294-305.
33. Ospina-Mateus, H., Jimnez, L. A. Q., Lopez-Valdes, F. J. & Salas-Navarro, K. (2019). Bibliometric analysis in motorcycle accident research: a global overview. *Scientometrics*, 121(2), 793-815.
34. Raparelli, E. & Bajocco, S. (2019). A bibliometric analysis on the use of unmanned aerial vehicles in agricultural and forestry studies. *International Journal of Remote Sensing*, 40(24), 9070-9083.
35. Zhang, L. M., Zhong, Y. G. & Geng, Y. (2019a). A bibliometric and visual study on urban mining. *Journal of Cleaner Production*, 239.
36. Hsu, Y. H. E. & Ho, Y. S. (2014). Highly cited articles in health care sciences and services field in science citation index expanded a bibliometric analysis for 1958-2012. *Methods of Information in Medicine*, 53(6), 446-458.
37. Zhang, X. M., Estoque, R. C., Xie, H. L., Murayama, Y. & Ranagalage, M. (2019b). Bibliometric analysis of highly cited articles on ecosystem services. *PLoS One.*, 14(2), 16.
38. Garris, H. W., Baldwin, S. A., Van Hamme, J. D., Gardner, W. C. & Fraser, L. H. (2016). Genomics to assist mine reclamation: a review. *Restoration Ecology*, 24(2), 165-173.
39. Fu, H. Z., Wang, M. H. & Ho, Y. S. (2012). The most frequently cited adsorption research articles in the Science Citation Index (Expanded). *Journal of Colloid and Interface Science*, 379(1), 148-156.
40. Li, Z. & Ho, Y. S. (2008). Use of citation per publication as an indicator to evaluate contingent valuation research. *Scientometrics*, 75(1), 97-110.
41. Yeung, A. W. K., El-Demerdash, A., Berindan-Neagoe, I., Atanasov, A. G. & Ho, Y. S. (2018). Molecular responses of cancers by natural products: modifications of autophagy revealed by literature analysis. *Critical reviews in oncogenesis*, 23(5-6), 347-370.
42. Fu, H. Z. & Ho, Y. S. (2014). Top cited articles in adsorption research using Y-index. *Research Evaluation*, 23(1), 12-20.

43. Marusic, M., Bozиков, J., Katavic, V., Hren, D., Kljakovic-Gaspic, M. & Marusic, A. (2004). Authorship in a small medical journal: A study of contributorship statements by corresponding authors. *Science and Engineering Ethics*, 10(3), 493-502.
44. Riesenberг, D. & Lundberg, G. D. (1990). The order of authorship: Who's on first. *Jama-Journal of the American Medical Association*, 264(14), 1857-1857.
45. Burman, K. D. (1982). Hanging from the masthead: Reflections on authorship. *Annals of Internal Medicine*, 97(4), 602-605.
46. Mattsson, P., Sundberg, C. J. & Laget, P. (2011). Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics*, 87(1), 99-105.
47. Bhandari, M., Busse, J. W., Kulkarni, A. V., Devereaux, P. J., Leece, P. & Guyatt, G. H. (2004). Interpreting authorship order and corresponding authorship. *Epidemiology*, 15(1), 125-126.
48. Ho, Y. S. (2014b). Classic articles on social work field in Social Science Citation Index: a bibliometric analysis. *Scientometrics*, 98(1), 137-155.
49. Mo, Z. W., Fu, H. Z. & Ho, Y. S. (2018). Bibliometrics of highly cited articles in the research field of volatile organic compounds. *Collnet Journal of Scientometrics and Information Management*, 12(2), 309-326.
50. Elango, B. & Ho, Y. S. (2018). Top-cited articles in the field of tribology: A bibliometric analysis. *Collnet Journal of Scientometrics and Information Management*, 12(2), 289-307.
51. Ho, Y. S. (2014a). A bibliometric analysis of highly cited articles in materials science. *Current Science*, 107(9), 1565-1572.
52. Davidson, F., Smith, R., Squires, B. P., Lundberg, G., Glass, R., Horton, R., et al. (1997). Uniform requirements for manuscripts submitted to biomedical journals. *Jama-Journal of the American Medical Association*, 277(11), 927-934.
53. Wren, J. D., Kozak, K. Z., Johnson, K. R., Deakyne, S. J., Schilling, L. M. & Dellavalle, R. P. (2007). The write position - A survey of perceived contributions to papers based on byline position and number of authors. *Embo Reports.*, 8(11), 988-991.
54. Fu, H. Z. & Ho, Y. S. (2015). A bibliometric analysis of the Journal of Membrane Science (1976-2010). *Electronic Library*, 33(4), 698-713.
55. Chuang, K. Y., Chuang, Y. C. & Ho, Y. S. (2015). Global influence of Cancer Statistics articles. *Current Science*, 109(9), 1552-1554.

-
56. Zheng, M. S., Fu, H. Z. & Ho, Y. S. (2017). Research trends and hotspots related to ammonia oxidation based on bibliometric analysis. *Environmental Science and Pollution Research*, 24(25), 20409-20421.
57. Khan, M. A. & Ho, Y. S. (2012). Top-cited articles in environmental sciences: Merits and demerits of citation analysis. *Science of the Total Environment*, 431, 122-127.
58. Garfield, E. (1990). KeyWords Plus: ISI's breakthrough retrieval method. Part 1. Expanding your searching power on Current Contents on Diskette. *Current Contents*, 32, 5-9.
59. Ratnasingham, S. & Hebert, P. D. N. (2007). BOLD: the barcode of life data system. *Mol Ecol Notes*, 7(3), 355-364.
60. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. & Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philos Trans R Soc B-Biol Sci*, 360(1462), 1847-1857.
61. Ho, Y. S., Satoh, H. & Lin, S. Y. (2010). Japanese Lung Cancer Research Trends and Performance in Science Citation Index. *Internal Medicine*, 49(20), 2219-2228.
62. Suk, F. M., Lien, G. S., Yu, T. C. & Ho, Y. S. (2011). Global trends in *Helicobacter pylori* research from 1991 to 2008 analyzed with the Science Citation Index Expanded. *European Journal of Gastroenterology & Hepatology*, 23(4), 295-301.
63. Zhang, G., Xie, S. & Ho, Y. S. (2010). A bibliometric analysis of world volatile organic compounds research trends. *Scientometrics*, 83(2), 477-492.
64. John Kress, W. (2005). Use of DNA barcodes to identify flowering plants. *PNAS*, 102(23), 8369-8374.