

Tracing information through the scientific literature

Dean Pentcheff <pentcheff@gmail.com> July 2016

Imagine you are reading a scientific paper and you find a statement that interests you. You would like to learn more about it. If it describes work other than the research done specifically for the paper you are reading, there will be a *citation* of someone else's published paper. In the text, these are links to the **References** section at the end of the paper.

You want to find out more about "heteroplasmy in the mtDNA"; the citation for that is "Rubinoff 2006". Furthermore, you want to find out more about "NUMTs"; that citation is "Song et al. 2008".

We'll start with the "Song et al. 2008" paper.

Turn to the end of the paper, and you'll see the **References** section. Usually, this is alphabetically arranged by first author's name, so look through it to find the paper by Song and other authors published in 2008. The "Song et al. 2008" *citation* in the text refers to this particular *reference*. It turns out to be a paper by four authors. The title of the paper begins "Many species in one...". The first key piece of information is to find the title of the journal in which it was published. In this case, it's abbreviated: "Proc Natl Acad Sci U S A", which stands for "Proceedings of the National Academy of Sciences of the USA". (If you don't know the journal by its abbreviation, you may be able to just use the abbreviation to search for it.)

So, let's find that journal. Google is your friend. Searching for "proceedings national academy usa" leads directly to "Proceedings of the National Academy of Sciences", which is the website for the journal we're looking for.

The trick now is to find where they keep all old the journal issues. Look for something like "Archive", "Past issues", "All issues", *etc.* In this case, there's a menu item for "ARCHIVE" near the top.

Nevertheless, these molecular approaches also have limitations. The selection of a barcode locus is complicated by the trade-off that arises between the need for universal application in a wide range of taxa and sequence substitution saturation (Kress et al. 2005a, b). In some cases, identical chloroplast or mitochondrial sequences are present in related species due to introgression, rendering these sources of DNA less useful or redundant for species discrimination. The heteroplasmy in the mtDNA genome (Rubinoff 2006) and the presence of nuclear pseudogenes of mitochondrial origin (NUMTs; non-functional copies of mtDNA in the nucleus) may also lead to misidentifications (Song et al. 2008). Although barcoding can serve as an important aid for taxonomic workflow, it cannot replace comprehensive taxonomic analyses and molecular phylogenetics. The primary role of this technique is not to build phylogenetic trees, but to provide rapid and accurate identifications of unidentified organisms with the use of verified reference material (Erickson and Kress 2012).

Global DNA barcoding efforts have resulted in the formation of the Consortium for the Barcode of Life (CBOL). In

Simion MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos Trans R Soc Lond B Biol Sci* 360:1825–1834

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc Natl Acad Sci U S A* 105(36):13486–13491

Sucher NJ, Charles MC (2008) Genome-based approaches to the authentication of medicinal plants. *Planta Med* 74(6):603–623

Sweeney BW, Battle JM, Jackson JK, Dapkey T (2011) Can DNA barcodes of stream macroinvertebrates improve descriptions of

The screenshot shows the homepage of the Proceedings of the National Academy of Sciences (PNAS). The navigation bar at the top includes links for "CURRENT ISSUE", "ARCHIVE", "NEWS", "MULTIMEDIA", "AUTHORS", "ABOUT", "COLLECTED ARTICLES", "BROWSE BY TOPIC", "EARLY EDITION", and "FRONT MATTER". The "ARCHIVE" link is circled in red. Below the navigation bar is a large image of a bison skull. To the right of the image, there is a "Current Issue" section for July 19, 2016, issue 113 (29), with a "From the Cover" section listing topics like "Bison phylogeography and paleoecological history", "Human protein structure", "Breast cancer and T-cell receptors", and "Interbacterial signaling". Below the image is a "TABLE OF CONTENTS" button. At the bottom, there is an "Early Edition" section with "New papers posted daily" and a "PNAS Direct to Your Inbox" section with options to "Customize your PNAS emails" by "Subject specific", "Keyword specific", or "All Early Edition".

Selecting that gets you to a page that lists all the published issues of that journal. Looking back at the reference, we see that our article was published in 2008, so click that.

Full Text and Abstracts: January 1915 - Present										
2010s	2010	2011	2012	2013	2014	2015	2016	-	-	-
2000s	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
1990s	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
1980s	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
1970s	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
1960s	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969

Journals are published as periodic “volumes”, each of which is split into a number of “issues”. Looking back at the reference, the volume and issue number come right after the journal title. In this case the volume is 105 and the issue is 36. Scanning the publications for 2008, we can find volume 105, issue 36. The page numbers for the article are 13486–13491, so we click on the issue that includes those page numbers.

September	
1033 - 11032	September 2; 105 (35) : 12637 - 13184, E56 - E57
1033 - 11451, E51 - E57	September 9; 105 (36) : 13185 - 13696
1033 - 11451, E51 - E57	September 16; 105 (37) : 13697 - 14234, E58 - E59
1033 - 11451, E51 - E57	September 23; 105 (38) : 14235 - 14744
1033 - 11451, E51 - E57	September 30; 105 (39) : 14745 - 15220, E60 - E61

Clicking on that brings us to a page listing all the articles in that issue. Scan down, and you'll find the Song *et al.* 2008 article.

PNAS

Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified
 Hojun Song, Jennifer E. Buhay, Michael F. Whiting, and Keith A. Crandall
 PNAS 2008 105 (36) 13486-13491; published ahead of print August 29, 2008, doi:10.1073/pnas.0803076105

[Abstract](#) | [Full Text \(HTML\)](#) | [Full Text \(PDF\)](#) | [Figures Only](#) | [Supporting Information](#)

Bergmann's rule and climate change revisited: Disentangling environmental and genetic responses in a wild bird population
 Céline Tenllisky, James A. Mills, Jussi S. Alho, John W. Yarrall, and Juha Merilä

Click on “Full Text (PDF)”, and you'll get a copy of the paper. Yay!

PNAS

Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified

Hojun Song^{*,†}, Jennifer E. Buhay^{*,†}, Michael F. Whiting^{*}, and Keith A. Crandall^{*}

^{*}Department of Biology, Brigham Young University, Provo, UT 84602; and [†]Belle W. Baruch Institute for Marine Sciences, University of South Carolina, Columbia, SC 29208

Edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved July 14, 2008 (received for review March 28, 2008)

Nuclear mitochondrial pseudogenes (numts) are nonfunctional copies of mtDNA in the nucleus that have been found in major plasmids. Plasmid is the presence of a mixture of more than one type of mitochondrial genome within a single individual, and the coamplification of numts and plasmids during DNA barcoding can lead to overestimation of species diversity.

The tougher case

Sometimes papers are not publicly available — publishers restrict their availability for commercial reasons. There are ways to deal with that.

Colleges and universities pay publishers large annual subscription fees to get access to published articles. If you are a student or staff at one of those institutions, you will be able to access the resources that your institution pays for.

Many of us don't have that kind of access, though. In the past (and still in the present), scientists would respond directly to requests for a copy of their paper. You can email one or more of the authors of a paper and request a copy. Very often (for recent papers), they'll reply with a copy of the paper.

However, a recent development has made that less necessary. An enterprising graduate student in Khazakstan devised a computer system that could access journal websites through university

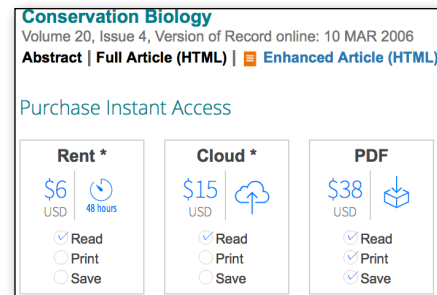
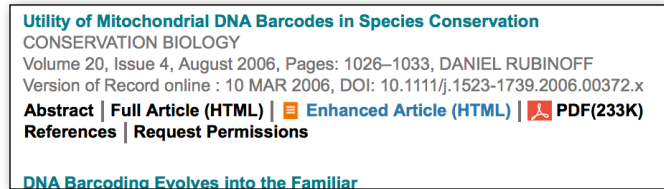
services, then make the papers available to anyone. That service is called “Sci-Hub”, and currently has at least 40,000,000 scientific papers in its archive.

For example, imagine you wanted a copy of the Rubinioff (2006) paper. In the **References** section it is listed as:

Rubinioff D (2006) Utility of mitochondrial DNA barcodes in species conservation.
 Conserv Biol 20:1026–1033

If you tracked down through the journals (similar to the way we did for the Song paper), you’d find the listing for the article on the website of the journal *Conservation Biology*.

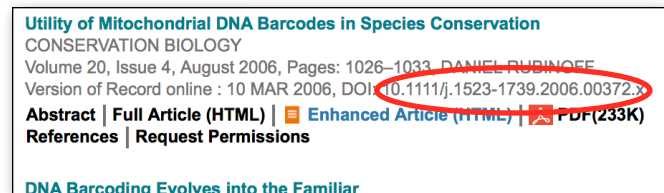
However, clicking on the “PDF(233K)” link leads you to a page requesting payment to read the article. The term for this is the “paywall” to access. The publisher would like you to pay \$38 for the article.



You can retrieve it, however, from Sci-Hub.

First, find the “DOI” (Document Object Identifier) for the article. Nearly all modern articles will have this as a unique identifier. For this article, check back at the article link to get the DOI. Mouse over the DOI and copy it. Don’t include the “DOI:” part, but just the numbers, slash, and letters of the DOI itself:

10.1111/j.1523-1739.2006.00372.x



Now visit Sci-Hub in your browser. Sci-Hub occasionally changes its “home” URL. If the URL given here doesn’t work, you can find alternate URLs by looking up “Sci-Hub” on Wikipedia. Today, the following URL works:

http://sci-hub.bz



Paste the DOI into the search blank at Sci-Hub and click “open”.

Next, you may have to type a short set of letters to prove you’re human.

After that, you’ll (nearly always) have access to the paper identified by the DOI. The Russian bits are parts of the Sci-Hub system, and will not appear in the copy you download. Now you can download and read the paper. Yay!

