

FINDING REFERENCES [GENETICS POSTGRADS]

Save time and effort - Make use of targeted resources!

You can actually make your search for information much *easier* if you use resources which allow you more control over searches, and help filter out the junk you don't want. Being 'search-savvy' is a skill you'll need, to keep up with and sift through the daily flood of information.

Obviously in many ways Google makes life easy, and we rely on it a lot. However, it works best for the 'known unknown' – information you know exists, and can *recognise* when you see it. So for reading lists, or if someone recommends a specific paper, Google will probably find it, or at least a reference to it. Google does such a fantastic job with searches like this, that we don't necessarily realise it is not the best option for finding scientific papers by *subject*. We want results quickly, and Google is scrabbling around in a vast amount of unregulated 'stuff' - so it would be expecting a lot for it to come up with exactly what you need, near the top of all those 'hits' it usually presents to you. We tend to lose interest after the first few hits - some of which were probably published some time ago – so we stop checking and assessing results. So although Google *appears* to make things easy, if *you are looking for written research by subject*, *Google actually forces you to do MORE work and takes up MORE of your time*.

To obtain a PhD you must undertake original research. But you also have to satisfy the Examiners that your dissertation '...takes due account of previously published work on the subject'. [Source: University Graduate Exams web pages] You are no longer just trying to find identified papers – you need to find what has been produced relating to a given subject – but you don't really *know* what is out there [the 'unknown unknown']. You want to be confident you have retrieved all the relevant information, but without having huge amounts to sift through, attempting to assess the content. You also need to keep up to date with research *being done* in your subject. Many PhD students only ever check out research a supervisor or colleague has helpfully suggested. But it is far better not to have to rely on others. Furthermore, busy scientists do quick scans through search results, applying their own 'filters' based on their experience, looking for names they know, for publication in high impact journals, etc. You may not have those 'filters' in place yet - but that's a good thing. It means you can be open and unbiased!

Tools to help you find information are getting better all the time at Cambridge. So even if you think you already know what the system provides, you should occasionally re-check the possibilities. For example, in summer 2013, a version of PubMed was made available with links to Cambridge resources, but many users still don't realise this facility exists, and search on the generic 'public' version, which doesn't 'know' what full-text papers we have access to. In July 2015, Scopus broke its 1997 barrier [the most common cut-off date for online resources] and is adding lots of older papers. In 2014, Web of Science rebranded [again] and upgraded, and Zetoc began to provide links to accessible papers via institutional log-in.

You could search in publishers' websites [large scientific ones are Elsevier, Blackwell, Wiley, Springer] but this obviously immediately limits coverage. Or you could search in a more independent, discrete database with wide, specialist, high-quality, reliable content – a bibliographic database.....

BIBLIOGRAPHIC DATABASES

Often, you search for references to useful information using one resource - for example a general search like Google, a bibliographic database, a metasearch such as Search+, a library catalogue - whilst the papers themselves are stored in *another* resource - for example ScienceDirect or SpringerLink [publishers' databases], Nature [journal or journal website], even a book.

Bibliographic databases basically consist of references to papers, usually with limited additional information, eg keywords or an abstract. Bibliographic databases will not always provide a link to the full text of articles cited, because these are *stored separately*. This is why you may have to go through several stages to get to a paper. This separation is not usually apparent when you retrieve a reference, so it may cause people to become confused and frustrated when trying to access full-text papers.

Searching subject-appropriate bibliographic databases saves you time and effort. The information they contain is more specific to begin with, it is of higher quality, the science is verified, there's less 'junk' to sort through, no repeats, links to similar research are provided, plus you can formulate a more precise search to get better results. So they have actually *done* some of the filtering work for you already.

You'll find your options for bibliographic databases listed in the e-resources [as opposed to e-journals and e-books] section of the University Library website : choose your general subject on <http://www.lib.cam.ac.uk/eresources/index.php>

The 3 bibliographic databases most likely to be of use to you are:



These are the initial preferred options for bioscience searches because of the large amount, and wide scope, of appropriate high-quality scholarly material covered. The choice might depend on your subject specialism, or simply personal preference for the interface.

Points you might consider when choosing:

- All of these databases contain millions of high-quality, scientifically-sound references from thousands of publications: Sources searched are listed and easily checked. They will all cover the major scientific literature, but will include different resources at the 'periphery', which is where suitability for a particular search comes in. Scopus supposedly searches in about 19,800 sources, WoS 12,300 and PubMed 44,300
- Web of Science [ex: Web of Knowledge] and PubMed are trusted by scientists because they are not controlled by a publisher pushing their product. Google does not disclose what it searches, and is believed to have secret agreements with some publishers, so retrieval is biased
- If you search 'all databases' in WoS, you are also searching in PubMed. In fact Google Scholar also relies heavily on PubMed, but results are 'muddied' by the inclusion of many other sources
- It is a good idea to avoid always relying on one database to find refs. Most of us go to old favourite sources from habit, but there will be other resources you know of but have ignored, and probably some you are unaware of. For any given search, using different resources than your usual go-to search will probably give different, and possibly better, results. Likewise, the same

search terms used in different databases should give a 'core' of the same references, but with surprisingly different 'extra' refs.

- PubMed facilitates the tightest searches because of its indexing scheme, but all these databases have indexing schemes, which are better for science than free-language searches
- All these databases provide useful [but underused] tools for improving and tightening searches, such as guided boxes that insert Boolean logic for you
- The search terms you used are highlighted in both the initial retrieval lists and the abstracts in WoS, and in bold in Scopus, allowing quick assessment of results. PubMed only displays the terms in bold in the initial retrieval list
- Both WoS and Scopus [and Google Scholar as well] display the number of times a paper has been cited in the initial retrieval list, another means of quickly assessing the value of research. In PubMed this information is harder to dig out
- If you find a 'ideal match' paper, these services allow you to find other papers where similar indexing terms were applied, or where some of the same papers were cited
- Default display of search results in all of these is most-recent-first, rather than best-match [which is what Google uses]. This ensures you are getting the recent developments before you get tired of checking! You can of course resort by 'relevance' [mechanically defined] if required

A few further points to bear in mind:

- **Never assume that the first set of references you are presented with is the best that you can get, even if you are in a hurry.** Even people whose job it is to search know they must refine their searches before they get best-fit results. Similarly, try important searches in different databases : One database may generally be best for you, but for some searches may not.
- You want the useful references you find to lead to the full-text of the paper. Usually they will do so, though sometimes not in one click. You sometimes need to know the work-around. But also be aware that the full text of everything you might need may *not* be available online at all, or we may not have access to it, because the University of Cambridge does not pay for it. We have mostly moved over to electronic subscriptions, but occasionally you may have to use a paper copy. If we don't have online access to a paper, don't write it off. We could have a paper copy in the Genetics Library; and if not, there is often a paper copy somewhere in the University. If it isn't, the library can usually obtain a copy for you.
- You will be told about funded scientific research being required to be **Open Access** : This is by no means fully in place yet, and will not apply retrospectively, so you cannot assume you will be able to access the papers you need because of this upcoming legislation.

THE MAJOR GENERAL SCIENCE BIBLIOGRAPHIC DATABASES

PUBMED - <http://tinyurl.com/campubmed> provides links to Cambridge resources [see below]

<http://www.ncbi.nlm.nih.gov/PubMed> [standard free-access version] does not

Note that you can also include PubMed in a search via other gateways, such as Web of Science & SciFinder Scholar [see below]

Key points about PubMed :

- Free-access, so no password ever needed
- Good life sciences coverage - by no means just medical

- **Tightest indexing** of all the databases = most precise retrieval, provided you define your search precisely. MeSH linking of descriptors improves retrieval by linking the words you input to similar concepts, as well as to wider subject areas, if you choose. Also helps you build very precise searches with Boolean logic, and combining good elements of a search with new concepts
- **Great for picking up recent research** Updated daily, with pre-publication and 'in-process' records. However, some of these refs remain unpublished even though dated months ago. If retrievable, the information from these will not be scientifically verifiable
- Through PubMed Central, one initiative to provide free full-text access to research on the web, you can see the full text of articles in many journals, though often not recent papers [which we should have access to via paid subscriptions - see next point]
- **Links to the papers we have access to are available on-campus** and can be provided off-campus. 'Standard' PubMed does not 'know' what Cambridge has access to, and so automatic links to free-access initiatives may mean our [paid] access is missed

Training and help : Many Cambridge scientists are expert users, and the Medical Library runs courses on PubMed and its useful 'bolt-ons'. Postgrads from Genetics and Plant Scis are welcome to attend. See end of this handout. Excellent help and training from PubMed is easily accessible – see <http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/> or for short YouTube videos, <http://www.nlm.nih.gov/bsd/disted/pubmed.html#qt>

Update info : There are quite a few blogs related to PubMed, but they are quite 'niche', eg this terrifying one on search strategies - <http://pubmedsearches.blogspot.co.uk/> The Twitter feed - <http://twitter.com/NCBI> - is no longer PubMed specific

Best Reference Management System match: No direct link but all RMSs should work with PubMed

WEB OF SCIENCE - <http://wok.mimas.ac.uk>

Key points about WoS :

- Useful if you need to search beyond scientific literature into social sciences / humanities; conversely you can limit to science [by unticking others under 'more settings' for a more precise search]
- **Genuinely searches the 'hidden web'**, since it is a subscription service linking to subscribed titles
- **Display of hits shows accurate-for-Cambridge links to full-text access, including off-campus.** This does mean you must log-in to WoS even on-campus, but once logged-in no subsequent publisher log-ins should be needed
- Immediately identifies – and links to - all the papers cited in a given paper and all the papers that cite it, so you **can follow research back in the past and forward too**. Hence 'Web' of Science ; this is such a useful feature other databases subsequently copied it, so it is now less unique to WoS
- Keywords are displayed with abstracts, so you can check and 'steal' them, to improve your search – because the database recognises those key words possibly better than your terms
- This is also the database used to check 'standard' impact factors [Journal Citation Reports facility]. In fact you can immediately check this from a retrieved reference, to assess its importance, although there are now of course alternatives to Impact Factor, such as Altmetrics

Training and help : Cambridge academics are the heaviest user worldwide. Good training documentation is easily accessible – see http://wokinfo.com/training_support/training/web-of-knowledge/ The University Information Services occasionally runs courses, but not often – see end of this handout.

Update info : Twitter feed - <http://twitter.com/thomsonreuters> is not specific to WoS & best avoided!

Best Reference Management System match : Direct link to EndNote Basic/EndNote Web [but note you will lose access to this when you complete your course]

SCIVERSE SCOPUS - <http://www.scopus.com/scopus/home.url>

SciVerse as a brand consists of two 'arms' : Scopus - where you search - and ScienceDirect, where the papers from the electronic journals published by mega-publisher Elsevier are stored. You can search for papers in ScienceDirect [*see later*] but Scopus has *much* wider coverage, and is therefore preferable.

Key points about Scopus :

- Has an easy tick-box system for **bulk-downloading** papers, except that although all appear to be available, they aren't - some are left off for no obvious reason, whilst others are labelled 'excluded'
- Details of each article has an Altmetric and Mendeley box in the right hand panel side so you can see who is referencing the paper, who is Tweeting etc
- Good for keeping you updated on developments. You can even recommend a journal for inclusion!

Training and help : Good help is available at http://help.scopus.com/Content/tutorials/sc_menu.html. The University may occasionally run course, but they are rare – see end of handout.

Update info : Scopus has by far the best update info of the big databases, both as a blog - <http://blog.scopus.com/> and via Twitter - <http://twitter.com/Scopus>

Best Reference Management System match : Direct export to Mendeley [both controlled by Elsevier]

OTHER BIBLIOGRAPHIC DATABASES WORTH INVESTIGATING

SciVerse-Science Direct – <http://www.sciencedirect.com/>

Because this is a separate site, where the papers themselves are stored, it is treated separately from SciVerse-Scopus [*above*]. You can search within ScienceDirect itself, which means you will only be searching Elsevier/Academic Press journals - but there are almost 2000 of them, including many key titles such as the 'Cell' stable. This is the largest publisher's site in existence, but be aware that the University has subscribed to 'bundles' of journal titles, so we don't have access to everything in the database – which is sometimes annoyingly not immediately obvious; in fact quite difficult to determine.

IngentaConnect - <http://www.ingentaconnect.com/>

Like Scopus, this is a search engine devised by a publisher [Springer] so naturally it favours their own output. Springer is another large publisher of scientific journals, but we are much less likely to have an on-line subscription to them than to Elsevier titles. Ingenta also searches in Springer books – which may seem great, especially as many look very good, until you find we have access to only a very few. This is because Springer titles are very expensive, and this being a German company, the University Library does not get access via its status as a Copyright Library. That also means it is hard to borrow these items or get scans from other libraries in the UK. In **addition, Ingenta** often doesn't 'know' if we have access. However, experts report that its retrieval of appropriate references is excellent.

SciFinder Scholar – follow instructions at <http://www-library.ch.cam.ac.uk/scifinder/>

Originally acquired for members of the Department of Chemistry, so there's a subject bias. Cross-searches on PubMed. Training sometimes available, but meanwhile see the Dept of Chemistry page above for links to help guides and online tutorials.

FlyBase - <http://flybase.org>

Specialist resource, obviously, but curators are working within the Dept of Genetics, indexing papers and providing comprehensive support resources. Does not appear in UL subject lists mentioned below.

Zetoc – <http://zetoc.mimas.ac.uk/>

A British Library database covering all subjects (including 13,000 scientific serials). A possible advantage over some other databases is the speed the references appear – because no sophisticated indexing is done, turnaround time is fast. But having no index/keywords means that the title of a paper retrieved must contain the *exact terms* you input. Unbiased, and covers 'niche' and foreign serials the big databases don't bother with. Also helpful are the spin-off services – Zetoc RSS and Zetoc Alerts – to keep you up to date. Guides to setting these up can be found on <http://zetoc.mimas.ac.uk/support.html> A recent development is the requirement to identify your organisation and log in through it: This may seem a nuisance, but it means Zetoc can provide you with a direct link to accessible papers online, including in alerts, instead of just giving you a possible lead to follow up, as it used to do.

COVERAGE COMPARISON

Academic Database Assessment Tool <http://adat.crl.edu/databases> allows you to compare titles covered by WoS, Scopus and Zetoc, plus other databases – though not PubMed

Aren't there any other specialised databases ...?

The University also has access to less well-known [and less-used] databases that may have subject coverage suited to your needs. Go to <http://www.lib.cam.ac.uk/eresources/index.php> and choose your subject to see a list of possibilities. At first glance, there appear to be many more databases listed by the UL than I have bothered to mention here, but many entries are hidden duplicates. Using LibrarySearch+, described in the next section, might also help you identify useful databases that you may not initially have considered.

METASEARCHES / GENERAL SEARCH ENGINES

It is *becoming* easier to search over more than a single database. You do this with a 'metasearch' or 'federated search'. But remember this is *not* the most efficient way to search - general search engines can never be as accurate and reliable as targeted ones. This is because the more sources of references that are being searched, the more difficult it becomes to process the search terms you type in, because they all have different rules. You will probably be presented with a large number of 'hits', including lots of inexplicable ones, which you will have to sift through. This is because each source operates with different rules. *This applies to Google*. We have not reached the stage where inputting free language into a single search box leads you to the precise full-text information you need.

LibrarySearch+ - <http://searchplus.lib.cam.ac.uk/>

This is Cambridge University Library's metasearch – as of 2015, replacements were being tested, so that tells you something! LibrarySearch+ not only searches the library catalogue [which is what LibrarySearch, without the +, is doing], but also in major journals databases, in newspaper databases, within e-books, and within [paper] books with online indexes. A possible advantage of LibrarySearch+ is that it searches in databases you may not have considered, or be aware of.

The fact that it searches within a huge range of books is useful, *but do not expect much of what is retrieved from books to be immediately available online*. This is because Cambridge currently provides very few e-books compared to other universities – and that is not a case of us lagging behind. First of all, teaching is increasingly being done from resources other than textbooks. Secondly, you might be surprised to learn that few recent high-quality science textbooks are even available as e-books. However, most of the books will be available in one of the Cambridge libraries as a paper copy.

Its most useful feature is perhaps the unmissable yellow sign which clearly indicates if a particular reference is available to Cambridge users online – you can also set the search to *only* retrieve items to which we have online access.

Google Scholar - <http://scholar.google.co.uk/>

Google Scholar will probably retrieve for you a large number of refs, but 'more' is not always 'better'. You waste time checking through a large number of hits, many of which are not appropriate, are duplicates, or consist of references in other papers. WoS, Scopus or PubMed will give you fewer hits overall, but more genuinely 'scholarly' hits, despite Google Scholar riding on the back of PubMed.

In the cam-domain, Google Scholar *should* automatically display a link to the e-journals interface next to references. If it does not, or if you are at home, you need to **set Scholar preferences**. Click on the cog at top right, then click on Library Links on the left of the next page. You should now be able to indicate University of Cambridge. Google Scholar checks what we have access to on a weekly basis - but there is a time-lag of up to 3 months, and it is not 100% reliable anyway. The most recently-published refs will be retrieved by Google itself before GS picks them up. *Note also that whilst Google Scholar can be set up to provide links to accessible material, Google itself cannot.*

Key points about Google Scholar

- Easily finds a paper where author, title or DoI are known, but is **less reliable for subject searches because it is harder to 'control'**
- The default simple Google search box is not efficient – use 'Advanced search' and filters for more precise results. Acceptance of 'free-language' input, plus 'best match' default retrieval, is rarely helpful in science
- Google's mechanical way of assessing what is scholarly from a huge mass of information includes the number of citations, which pushes older papers to the top. It also factors in how 'popular' it is, ie how many times accessed – not always a reliable indicator for serious research
- Searches full text of paper or significant part of it [if accessible]; not reliant on indexers, which may or may not be a good thing
- As mentioned above, you can set it up to show links to Cambridge e-access. But you may still not get to the paper itself in one or two clicks – links are inconsistent
- May find some full-text papers that other databases don't link to, eg notably on researchers' own websites, though these tend to be older material
- GS gives every appearance that references retrieved will lead to full-text information, and often you waste time trying to get to them, getting confused by password boxes, only to find they are on pay-only sources we don't have access to

World Wide Science - <http://worldwidescience.org/>

Not run by a publisher, so in that respect is unbiased. Its [listed and de-selectable] sources are relatively unknown, and more overseas research is retrieved – This is its possible main advantage, as the other services are heavily biased towards English-language US/UK sources. It searches for data and non-‘print’ resources as well as papers. It is certainly likely to give you different results than the standard databases.

COURSES INFO

The University Information Services runs some library-related courses, but they are not available very often. There is a combined Web of Science / Scopus course on 9 November, but nothing else currently advertised, so you are advised to book on that one.. See <http://training.cam.ac.uk/ucs/> for details.

The University Library occasionally runs short courses on subjects like finding refs or using databases. They don’t run all year, nor are there many repeats on some sessions. Provision seems to depend on staff available. See <http://training.cam.ac.uk/cul/>.

RECOMMENDED : The Medical Library on the Addenbrooke’s site has on its staff a search specialist [Isla Kuhn]. She runs courses on PubMed and ‘The Big Four’ databases [includes Embase as well as the three above] These courses are very useful in helping to discover the tools the databases provide to help your search. Even if you don’t end up using any of these databases, the search tips you will pick up will be useful. See <http://training.cam.ac.uk/cul/theme/med?providerId=119246> Currently the only courses on the Big Four and PubMed are in October, so book now.

IMPORTANT : Many courses only run in this early part of the academic year, but as they are usually only listed for the current term, you cannot determine if they will be available in future. Often they are not. Don’t put it off – sign up now!

The Medical Library also produces short teach yourself guides – see <http://library.medschl.cam.ac.uk/research-support/teach-yourself-guides/>

The University provides some links for referencing and study skills [quite a few go to other universities’ websites!] See <http://www.admin.cam.ac.uk/univ/plagiarism/students/referencing/>

And finally..... Please remember that library staff are employed specifically to help you. Do ASK if you have a question or problem – If I can’t help immediately, I should at least know someone who can!

If you can’t find me, you can always e-mail me on :
library@gen.cam.ac.uk or library@plantsci.cam.ac.uk

© Christine Alexander Genetics Library / Plant Sciences Library October 2015