Supplemental material for Bowman et al. "On applications of landscape genetics".

Conservation Genetics, 2016, doi:10.1007/s10592-016-0834-5.

Appendix 1. Web of Science title extraction, web scraping, and summary statistics methodologies.

General overview

We first extracted article titles for each research topic from Web of Science (WoS). We extracted 19610 publications from 25 different research fields but reduced the analyzed set to 18745. We removed 415 titles that were either within topic duplicates or without title or author. Of these 18745 titles 202 were duplicated between research topics (18543 were unique). We performed our analysis on these 18745 titles.

We used the Google Inc. search engine to find applications that were related to each title. For each title we created a search term that specifically searched for Portable Document Format (PDF) documents on government websites. We considered a title to be 'applied' if it had at least one search result. A title can also have more than 1 application. For instance, a title could have 5 results; it was therefore an 'applied' title with 5 applications.

To reduce difficulties caused by submitting large numbers of search terms to Google, we randomly sampled 5% of titles 9999 times within each research field, and repeatedly searched for these randomly sampled subsets. We calculated summary statistics for each research field (e.g., mean, median, mode, and standard deviation of the number of applied titles). Most of all subsequent methodology was performed in the R statistical language (version 3.2.1).

Extracting titles of research fields from Web of Science

In the WoS database, we performed searches of each topic using quotes around the search term (e.g., "landscape genetics"). The search was refined by excluding every document type that was not a peer-reviewed journal article. The total number of titles was then recorded. WoS limits extractions to more than 500 or fewer article titles at a time. Therefore, the titles were extracted in fragments of 500 titles. We downloaded the search list using the "Save to Other File Format" tool in WoS and save the data as 'Tab-delimited (Win, UTF-8)' files.

Building search terms

We structured our Google search engine queries to create specific search term boundary limits in the following ways. Individual titles were extracted by enclosing within quotations (e.g., "title"). We then added file type specification (filetype:document type) to identify only PDF documents, and we added a site specification (site:website) to limit searches to government websites. We considered all government websites that contained "*.gov.*", "*.gc.*" or

"*.gouv.*" in their URL. US government websites did not show up on these searches; thus we added a second indicator (".gov.") to not exclude US websites. For each title the search term was similar to the following string:

"Article title" filetype:pdf site:*.gov.* OR site:*.gc.* OR site:*.gouv.* OR site:.gov. (1)

This allowed us to search the web pages of 161 national governments (http://www.politicsresources.net/official.htm), along with numerous, additional subnational governments (e.g., US states).

Titles of 5 words or less caused some problems, by generating hundreds to thousands of hits. Therefore we added the last name of each author within the search term. For example the article titled "Landscape genetics" by Holderegger and Wagner (2008) produced multiple search results that were not related to this title. Special characters and the whitespaces found in article titles also caused some issues. Whitespaces were replaced with the additional character '+' which is equivalent to adding '%20' and double quotes were replaced with '%22', since white spaces and double quotes cannot be used in a URL context. We replaced the ampersand (&) with the wildcard asterisk character (*) and the period (.), the single quote (') and the backslash (\) characters with a zero-width space.

Titles with whitespace and names added were coded as the following:

%22article+title%22+%22author+last+name%22+filetype:pdf+site:*.gov.*+OR+site:*.gc.*+OR+site:*.gouv.*+OR+site:.gov. (2)

We added each title to the search engine URL ('http://www.google.ca'/search?q=') as follows:

http://www.google.ca/search?q=%22article+title%22+%22author+last+name%22+filetype:pdf+site:*.gov.*+OR+site:*.gc.*+OR+site:*.gouv.*+OR+site:.gov. (3)

Web Scraping

We used functions from the RCurl R package (version 1.95-4.7) (Lang 2015a) to connect to the Google search engine and perform our searches and we used functions from the XML R package (version 3.98-1.3) (Lang 2015b) to extract information from the search pages. We counted the number of possible applications by summing the search result URLs found on each page. We changed pages until the Google search engine indicated the following:

1) "No results found for"
2) "did not match any documents"
3) "In order to show you the most relevant results, we have omitted some"

Options 1 and 2 both indicated that there were no results found at all, thus the total number of applications was 0 and the title was not applied.

Summary statistics

For each research field we randomly sampled a number of titles that was equal to 5% of the total number of titles from WoS search of that field. We repeated this process 9999 times to produce a distribution of the number of applied titles. We then calculated the mean, mode, median and standard deviation.

References

Holderegger R, Wagner HH (2008) Landscape genetics. BioScience 58:199–207

Lang, DT (2015a) RCurl: General network (HTTP/FTP/…) client interface for R. R Package Version 1.95-4.7.

Lang, DT (2015b) XML: Tools for parsing and generating XML within R and S-Plus. R Package Version 3.98-1.3.

Appendix 2. Results of a multiple linear regression ($F = 102.02$, $df = 2,22$, $P < 0.0001$, $R^2 = 0.90$) comparing the number of applied titles in 25 different research fields to the age of the field and the number of peer-reviewed articles in the field. The number of applied titles was estimated via a web-scraping search of national and subnational governments. The age of each field was estimated using Google's Ngram viewer as the year the field first appeared in English literature, and the number of peer-reviewed articles was based on a search of Web of Science. All searches were conducted in November 2015. The number of applications and the number of articles were log_{10}-transformed.

Dependent variable Log_{10} (Number of applications)

Variable	**Coefficient**	**SE**	**t**	**P**
Constant	-2.06	2.34	-0.88	0.39
Log_{10}(Articles)	0.80	0.06	12.51	<0.0001
Year	0.00	0.00	0.16	0.874

Appendix 3. Peer-reviewed articles (N = 22) that were used in landscape genetics applications. The articles were identified as being applied by searching web sites of national or sub-national governments (*.gov.* OR site:*.gc.* OR site:*.gouv.* OR site:.gov.) for titles of 879 papers found on Web of Science using "landscape genetics" as a search term (searches conducted November 2015). Most of the applied articles are methods or theory development (32%), or studies of mammals (32%). The rest are studies of reptiles (18%), fish (9%), amphibians (5%) or plants (5%). There were no landscape genetics applications of studies involving birds.

Blanchong JA, Samuel MD, Scribner KT, Weckworth BV, Langenberg JA, Filcek KB (2008) Landscape genetics and the spatial distribution of chronic wasting disease. Biol Lett 4:130–133

DeYoung RW, Zamorano A, Mesenbrink BT, Campbell TA, LeLand BR, Moore GM, Honeycutt RL, Root JJ (2009) Landscape-genetic analysis of population structure in the Texas gray fox oral rabies vaccination zone. J Wildl Manage 73:1292–1299

Epps CW, Wehausen JD, Bleich VC, Torres SG, Brashares JS (2007) Optimizing dispersal and corridor models using landscape genetics. J Appl Ecol 44:714–724

Francois O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. Mol Ecol Resour 10:773–784

Funk CW, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27:489–496

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. Genetics 170:1261–1280

Guillot G, Mortier F, Estoup A (2005) GENELAND: A computer package for landscape genetics. Mol Ecol Notes 5:712–715

Hagerty BE, Nussear KE, Esque TC, Tracy RC (2010) Making molehills out of mountains: landscape genetics of the Mojave desert tortoise. Landscape Ecol 26:267–280

Hagerty BE, Tracy CR (2010) Defining population structure for the Mojave desert tortoise. Conserv Genet 11:1795–1807

Holderegger R, Wagner HH (2008) Landscape Genetics. BioScience 58:199–207

Hunter ME, Mignucci-Giannoni AA, Tucker KP, King TL, Bonde RK, Gray BA, McGuire PM (2012) Puerto Rico and Florida manatees represent genetically distinct groups. Conserv Genet 13:1623–1635

Landguth E L, Cushman SA, Schwartz MK, McKelvey KS, Murphy M, Luikart G (2010) Quantifying the lag time to detect barriers in landscape genetics. Mol Ecol 19:4170–4191

Latch EK, Boarman WI, Walde A, Fleischer RC (2011) Fine-Scale Analysis Reveals Cryptic Landscape Genetic Structure in Desert Tortoises. Plos One 6:e27794

McRae BH (2006) Isolation by resistance. Evolution 60:1551–1561

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. Proc Natl Acad Sci USA 204:19885–19890

Monzon-Arguello C, Lopez-Jurado LF, Rico C, Marco A, Lopez P, Hays GC, Lee PLM (2010) Evidence from genetic and Lagrangian drifter data for transatlantic transport of small juvenile green turtles. J Biogeogr 37:1752–1766

Muñoz-Fuentes V, Darimont CT, Wayne RK, Paquet PC, Leonard JA (2009) Ecological factors drive differentiation in wolves from British Columbia. J Biogeogr 36:1516–1531

Raeymaekers JAM, Maes GE, Geldof S, Hontis I, Nackaerts K. Volckaert FAM (2008) Modeling genetic connectivity in sticklebacks as a guideline for river restoration. Evol Appl 1:475–488

Row JR, Wilson PJ, Gomez C, Koen EL., Bowman J, Thornton D, Murray DL (2014) The subtle role of climate change on population genetic structure in Canada lynx. Glob Change Biol 20:2039-2378

Schwartz MK, Copeland JP, Anderson NJ, Squires JR, Inman RM, McKelvey KS, Pilgrim KL, Waits LP, Cushman SA (2009) Wolverine gene flow across a narrow climatic niche. Ecology 90:3222–3232

Spies, I (2012) Landscape Genetics Reveals Population Subdivision in Bering Sea and Aleutian Islands Pacific Cod. Trans Am Fish Soc 141:1557–1573

Stevens VM, Verkenne C, Vandewoestijne S, Wesselingh RA, Baguette M (2006) Gene flow and functional connectivity in the natterjack toad. Mol Ecol 15:2333–2344