Mark Ian Cooper
Department of Botany and
Zoology, University of
Stellenbosch, Private Bag X1,
Matieland 7602, South Africa.

Darwin not a better name than Wallace

Mark Ian Cooper

Abstract
The position of an author's surname in the alphabet is tested for randomization and double barrel names are tested in a bibliometric analysis to see which names occur more frequently in first and second position. Using an analysis of data from nine online journals (African Entomology, African Invertebrates, African Zoology, Arthropods, Journal of Entomology and Zoology Studies, South African Journal of Science, Behavioural Ecology and Sociobiology, Ostrich, Trends in Ecology and Evolution) I found no evidence for a random distribution of surnames based on first letter (t=-6.913024, $p<0.00001$, n=8333 pdf's). A table for comparing relative surname positions is given. Within double barrel names the D's where not more frequent than the W's in the second position (t = -0.36, $P = 0.719$, n=21 letters).

Keywords: alphabet; Darwin; random; Wallace

1. Introduction
Population sampling often requires a random sample to be obtained where the chances of obtaining each sample therein are equal [1]. In order to obtain a sample from the alphabet a surname bias arises [2]. Here I test the hypothesis that the observed position of an author's last/surname name does not fall randomly in the alphabet. I downloaded files from archives of journals in order to establish whether the position of an author's surname in the alphabet is random. I construct a frequency table for comparing surnames in different positions of the alphabet. The particular bias which was identified by Tregenza (1997) in double-barrel names is questioned [2].

2. Materials and Methods
Portable document files (pdf's) were downloaded from *African Entomology* (reference.sabinet.co.za/sa_epublication/ento; *ISSN* 1021-3589; 1993-2015), *African Invertebrates* (content.ajarchive.org/cdm4/browse.php?CISOROOT=%2F16815556; *ISSN* 2305-2562; 2006, 2007, 2011-2014), *African Zoology* (reference.sabinet.co.za/sa_epublication/afzoo; *ISSN* 2224-073X; 1965-2015), *Arthropods* (http://www.iaees.org/publications/journals/arthropods/arthropods.asp; *ISSN* 2224-4255; 2012-2015), *Journal of Entomology and Zoology Studies* (http://www.entomoljournal.com/archive.html; *ISSN* 2320-7078; 2013-2015), *South African Journal of Science* (http://www.sajs.co.za/archive; *ISSN* 1996-7489; 2001-2015), the South African Bird Atlas Project [3-4], and open access articles in the journals *Animal Behaviour* (http://www.journals.elsevier.com/animal-behaviour; *ISSN* 0003-3472; 2000-2015), *Behavioural Ecology & Sociobiology* (link.springer.com/journal/265; *ISSN* 1432-0762; 2015), *Ostrich* (http://www.tandfonline.com/loi/tost20; *ISSN* 1727-947X; 2014-2015), and *Trends in Ecology & Evolution* (http://www.sciencedirect.com/science/journal/01695347; *ISSN* 0169-5347; 2015). I performed a paired samples t-test (http://www.socscistatistics.com/about/Default.aspx) comparing observed percentages of files from a-z against random expectations which were calculated by dividing each of the observed frequencies by 26 and multiplying them by 100 to get random expected percentiles. Double barrel names were sampled from the population to compare the frequencies of first versus last position.

3. Results
I found a significant difference between observed position of an author's surname and the expected random percentages (t=-6.913024, $p<0.00001$, n=8333 pdf's). Table 1 illustrates the frequency distribution of author's surnames in the 26 letter alphabet (n=10049).

Correspondence
Mark Ian Cooper
Department of Botany and
Zoology, University of
Stellenbosch, Private Bag X1,
Matieland 7602, South Africa.

Surnames began most frequently with S (10.021%) and least frequently with X (0.090%). D's (5.155%) and W's (4.199%) have almost equal frequencies and in double barrel names the combined frequencies of 21 letters with double-barrels was not significantly higher in the second position (Table 1: 987.124 versus 1049.114; t = -0.36, N=21 letters, P = 0.719).

Table 1: Frequency of author's surnames obtained from a random sample of publications (n=10049) obtained from archived journals alongside their random frequencies and double barrel frequencies.

Position in the alphabet	Observed Frequency	Random Frequency	Cumulative Double barrel Frequency
A	4.697	0.180	5.0155
B	2.259	0.083	4.986
C	7.722	0.297	5.097
D	5.155	0.198	7.854
E	2.567	0.099	4.572
F	2.756	0.106	5.498
G	4.856	0.186	5.634
H	6.697	0.258	5.058
I	0.537	0.021	4.697
J	1.940	0.075	3.871
K	5.075	0.195	6.566
L	4.209	0.162	4.912
M	9.643	0.371	7.001
N	2.677	0.103	2.259
O	2.329	0.090	5.704
P	7.105	0.273	5.920
Q	0.239	0.009	0
R	4.787	0.184	4.894
S	10.021	0.385	5.085
T	3.871	0.149	5.332
U	0.567	0.022	0.557
V	4.597	0.177	3.677
W	4.199	0.162	3.612
X	0.090	0.003	0
Y	0.607	0.023	0
Z	0.896	0.034	0

4. Discussion

There are many different forms of documentation (systems of citations and referencing), varying across academic fields. The result of a non-random distribution of first author surnames in the western alphabet is given in this bioliometric analysis as a foundation for comparing surnames in different positions of the alphabet (Table 1). The data illustrates the importance of testing assumptions based on randomicity and illustrating cases where the null hypothesis is falsified for the purposes of explaining the trend in a non-random case. The data do not agree with Tregenza's (1997) [2] finding that Darwin is a better name than Wallace. There is a non-linear relationship between surnames in our alphabet. Each authors surname begins with a letter of the alphabet which has a unique and sometimes different frequency of publications associated with it. Rather than searching for an ultimate index of citation I objectively establish the relative importance of each position of an author's surname and account for any frequency dependence bias where this exists.

Insofaras double barrel names are concerned there is no difference between the relative frequencies of Darwin's compared to Wallace's. Thus I have falsified Tregenza'a (1997) [2] hypothesis or finding that Darwin is a better name than Wallace. Rather Darwin is just as frequent a name than Wallace as indicated by this data set of pdf files from a sample of nine (mainly southern African) journals. This is simply one result and should like any scientific study may be replicated.

This is one step closer to reaching a fair citation system and suggests treating each comparative case in order to examine whether there is no significant difference between author surnames or consider their unique frequencies based on surname initial letter.

5. Conclusion

Further investigation into the non-random alphabetic position of an author's surname considered a more accurate approach to determine whether there was structuring of surname position within double barrel names. There was no evidence that "Darwin [is] a better name than Wallace."

6. References

1. Winner L. Introduction to Biostatistics. Florida, United States, 2004, 204.
2. Tregenza T. Darwin a better name than Wallace? Nature 1997; 385(6616):480.
3. Harrison JA, Allan DA, Underhill LG, Herremans M, Tree AJ, Parker V, *et al*. The Atlas of Southern African Birds. Volume 1: Non-passerines, BirdLife South Africa, Johannesburg, 1997, 1-732.
4. Harrison JA, Allan DA, Underhill LG, Herremans M, Tree AJ, Parker V *et al*. The Atlas of Southern African Birds. Passerines, BirdLife South Africa, Johannesburg 1997, 2:1-785.