



This is a repository copy of *Advances in proteomics for production strain analysis*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/87279/>

Version: Supplemental Material

Article:

Landels, A., Evans, C., Noirel, J. et al. (1 more author) (2015) *Advances in proteomics for production strain analysis*. *Current Opinion in Biotechnology*, 35. 111 - 117. ISSN 0958-1669

<https://doi.org/10.1016/j.copbio.2015.05.001>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Proteomics Literature Assessment

Using bibliometric analysis methods

Import of the data

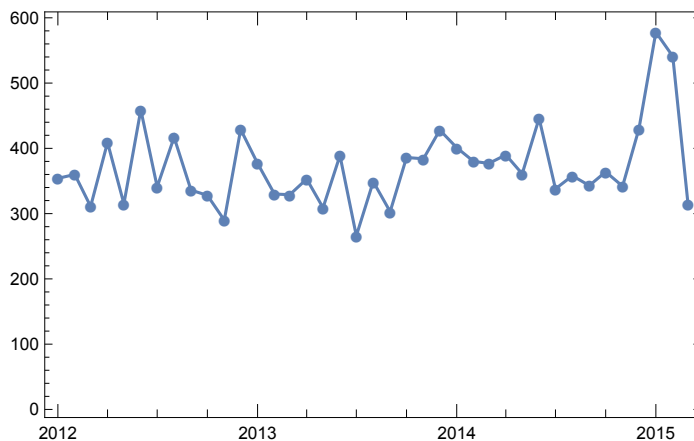
All data was imported from PubMed on 15/03/2015, using the following search parameters:

```
proteomics AND (("2012/01/01"[PDat] : "3000/12/31"[PDat]))
```

This produced 18588 hits, which were downloaded as an .xml file. We collected the fields "Year", "Month" and "Abstract"

Publication trends - proteomic publications by date

The data was then assessed for general publication trends. Using the PubMed meta-data "Year" and "Month", we plotted the frequency of proteomics publications over the review period on a month-by-month basis.



As a point of note, we highlight the up-trend in publication frequency this year. Both January and February 2015 have more proteomics publications associated with them than all previous years, and March 2015 (the final point on the graph) is showing a continuation of the same trend despite only having half a months data.

Publication trends - production-strain proteomics analysis

The abstracts of each paper were converted into a list of unique words, that were longer than 3 characters in length. Common words, such as "there" and "because" were removed from the investigation to enable focus on technical terms.

The list of publications was split into two groups:

- (1) the first group was made of the publications whose abstract matched the regular expression 'produc | isop | biofuel';
- (2) the second group was made of the remainder of the publications.

Frequency analysis

The 200 most frequent words used within these lists were tallied and ranked, producing two lists of the most common abstract terms.

The rankings were compared, and the change in rank is highlighted as follows: words in blue are higher-ranked in production-strain proteomics than in proteomics in general; words in red are lower ranked, words in black have not changed relative position. Faded words have changed rank by 5 places or fewer and words in bold-face are only present in the production-strain list. The solid line indicates the change in rank.

This work is not definitive as the selection terms are not exclusive to production strain analysis, however it does provide an indication of trends within the field.

