

LETTER • OPEN ACCESS

## Analysis of biased language in peer-reviewed scientific literature on genetically modified crops

To cite this article: Bo Maxwell Stevens *et al* 2021 *Environ. Res. Lett.* **16** 084035

View the [article online](#) for updates and enhancements.

ENVIRONMENTAL RESEARCH  
LETTERS

## LETTER

## Analysis of biased language in peer-reviewed scientific literature on genetically modified crops

## OPEN ACCESS

RECEIVED  
13 January 2021REVISED  
4 June 2021ACCEPTED FOR PUBLICATION  
14 July 2021PUBLISHED  
29 July 2021

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Bo Maxwell Stevens<sup>1,2,\*</sup> , Randi Reppen<sup>3</sup>, Mark Linhart<sup>7</sup>, Kara Gibson<sup>4</sup>, Adrah Parafiniuk<sup>5</sup>, Aradhana Roberts<sup>6</sup>, Robert Sanford<sup>2</sup> and Nancy Collins Johnson<sup>2,4</sup><sup>1</sup> Water Management and Systems Research, USDA ARS, Fort Collins, CO 80526, United States of America<sup>2</sup> School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ 86011, United States of America<sup>3</sup> Department of Applied Linguistics, Northern Arizona University, Flagstaff, AZ 86011, United States of America<sup>4</sup> Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, United States of America<sup>5</sup> Department of Politics and International Affairs, Northern Arizona University, Flagstaff, AZ 86011, United States of America<sup>6</sup> Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden<sup>7</sup> Deceased 13 May 16.

\* Author to whom any correspondence should be addressed.

E-mail: [bo\\_stevens@nau.edu](mailto:bo_stevens@nau.edu)**Keywords:** GMOs, linguistics, social, political, peer-reviewed articles, crops, stanceSupplementary material for this article is available [online](#)**Abstract**

Social, political, and economic forces may inadvertently influence the stance of scientific literature. Scientists strive for neutral language, but this may be challenging for controversial topics like genetically modified (GM) crops. We classified peer-reviewed journal articles and found that 40% had a positive or negative stance towards GM crops. Proportion of positive and negative stance varied with publication date, authors' country of origin, funding source, and type of genetic modification. Articles with a negative stance were more common at the beginning of the millennium. Authors from China had the highest positive:negative ratio (8:1), followed by authors from the USA (12:5) and the EU (5:7). Positive stance articles were six times more likely to be funded by private sources compared to those with a neutral or negative stance. Articles about glyphosate were more likely to be negative compared to articles about *Bacillus thuringiensis*. Linguistic features of articles with positive and negative stances were used to train a random forest classifier that predicts stance significantly better than random chance. This suggests the possibility of an automated tool to screen manuscripts for unintended biased language prior to publication.

**1. Introduction**

Public opinion on the safety of genetically modified (GM) food indicates a disconnect with the current state of scientific consensus. A 2015 poll by the Pew Research Center revealed that 37% of US adults believe GM foods are safe to consume while 88% of scientists within the American Association for the Advancement of Science believe them to be safe [1]. This disparity between the opinions of the general public and scientists is a concern for consumers, regulatory agencies, and producers of GM products. The internet is the main pipeline of information connecting published scientific research and public awareness through journalism [2]. Stance in favor of or against the use of GM crops can be subtly present in the

language of peer-reviewed journal articles, conveyed to media outlets and amplified into public opinion [3]. Articles with a negative stance on GM foods are more likely to become the source of rumors [4], while, a positive stance in scientific literature may shift public perception about GM foods toward a greater belief in their safety [5, 6]. On the other hand, confirmation bias leads to overconfidence in a particular opinion [7]. The purpose of this study was to quantify stance in published literature, and to identify linguistic patterns in positive and negative literature that could propagate from scientists into popular culture and may serve as potential sources of polarized public opinion about GM crops.

A wide variety of GM crops have been developed over the past 38 years [8]. The two most

common genetic modifications are resistance to the broad-spectrum herbicide glyphosate (aka Round-Up™), and incorporation of Cry protein genes from *Bacillus thuringiensis* (Bt) to make crops toxic to insect pests [3, 9]. Glyphosate resistant crops were first approved in 1996 and now account for more than 90% of cotton, corn, and soybean crops in the US [3]. Bt was isolated in 1901 and the first commercially successful Bt crops were introduced in 1996 [10]. Many GM crops contain both genetic modifications so that farmers can efficiently eliminate weeds with glyphosate and avoid the use of insecticides because the crop simultaneously withstands the herbicide and is no longer threatened by common insect pests. Other genetic modifications have been developed that help crops resist diseases [11] and increase their nutritional value such as the insertion of the genes required to synthesize vitamin A in golden rice [12].

There are many reasons for negative and positive attitudes about GM crops. A major concern held by the general public regards the safety and nutritional quality of GM foods [13], although recent reviews of scientific studies provide evidence that GM foods are substantially equivalent to non-GM foods [13–16]. The use of glyphosate-resistant crops has increased yields and profits for farmers [17–19], but there are potential negative impacts of the drastic increase in the use of glyphosate worldwide [3] such as the evolution of glyphosate resistant weeds [20]. Also, glyphosate is slow to degrade and can accumulate as residue on plants, and more research is needed to determine if chronic exposure results in deleterious effects [21]. Recent research linking cancer to glyphosate exposure is being scrutinized [22, 23]. Another serious concern is that GM crops can hybridize with wild relatives [24–26]. There is also the potential for Bt crops to adversely impact populations of non-target organisms [27], although evidence suggests this is not the case [28–30]. To the contrary, the use of Bt crops has substantially reduced the use of much more toxic insecticides, and this has helped protect populations of non-target insects [3, 31].

Despite reasons for positive and negative attitudes about GM foods, scientists strive for objectivity and are expected to convey a neutral stance in their writings. The intention of this study is not to criticize the use of any particular position or stance when conveying information regarding biotechnology, but rather to raise awareness of the subtle means in which authors can convey a position. A positive stance indicates that an author is optimistic about the potential for GM foods; a negative stance could indicate a concern for the environment. While both perspectives are equally important in the overall discussion of biotechnology, scientific literature is typically idealized as lacking opinion. Yet, many scientific articles convey a position or particular stance through subtle linguistic features that go beyond just the words used and include grammatical features.

Characteristics of language such as stance can be reflected in the writing style of authors. For example, in financial writing, reports that present lower earnings are harder to read (e.g. more syllables, more words) [32]. Compared to genuine scientific writing, fraudulent papers are characterized by fewer adjectives and more citations [33, 34]. Writing by students and poets with depression has distinct linguistic characteristics [35, 36]. We are not suggesting that authors of papers with a negative or positive stance toward GM crops are depressed, obfuscating results, or deceptive, but rather that stance may be conveyed with unique, subtle linguistic characteristics.

The purpose of this study was to quantify and analyze stance across peer-reviewed scientific literature published between 2000 and 2018 and test the hypotheses that (a) there exist subsets of peer-reviewed journal articles with detectable positive, neutral, or negative stance toward GM crops, and their proportions will vary with publication date, authors' country of origin, funding source and type of genetic modification; (b) characteristic linguistic features of papers with a positive or negative stance can be used to predict the stance of article; and (c) that articles with similar stances tended to cite similar articles.

## 2. Methods

### 2.1. Search strategy

The aim of our literature search was to gather published, peer-reviewed journal articles discussing GM crops within the context of the environment, as opposed to purely laboratory studies. Our goal was to quantify stance and analyze the linguistic characteristics of these articles, not to assess the actual results of individual papers. To quantify stance in published literature on GM crops, we first narrowed our focus to articles related to GM crops that were commercially available in 2014 and mentioned environmental impacts as a key term [37]. Our search was restricted to English language articles that were published after 2000. We conducted our first search in July 2014 and updated our database in August 2018 to include newer articles.

Search strings used to gather published articles were composed of three essential parts connected by an AND operator. Each of the three parts (crops, environmental impact, and genetic modification) included multiple key terms, some with wildcards to capture multiple variations of words, connected with an OR operator. Corn, rapeseed, cottonwoods, rice, wheat, soybean, tobacco, cotton, subspecies of *Brassica rapa* (e.g. turnips and bok choy), beets, beans, creeping bentgrass, field pumpkins, potatoes, alfalfa, sugarcane, peppers, tomatoes, Petunias, papayas, eggplants, chicory, plums, muskmelon, roses, flax, and carnations were included in our search. We searched eight databases:

CAB Abstracts, GreenFile, PubMed, BioSIS, Web of Science, Academic Search complete, and the Biological and Environmental Sciences collections within ProQuest, for published articles matching our search criteria. Each database was searched using keywords within a combination of titles, subjects, abstracts, and descriptions, and occasionally within the entire document (table S1 available online at [stacks.iop.org/ERL/16/084035/mmedia](https://stacks.iop.org/ERL/16/084035/mmedia)).

**Part 1:** ('*Medicago sativa*' OR '*Brassica napus*' OR '*Phaseolus vulgaris*' OR '*Dianthus caryophyllus*' OR '*Cichorium intybus*' OR '*Gossypium hirsutum*' OR '*Agrostis stolonifera*' OR '*Solanum melongena*' OR '*Linum usitatissimum*' OR '*Zea mays*' OR '*Cucumis melo*' OR '*Carica papaya*' OR '*Petunia hybrida*' OR '*Prunus domestica*' OR '*Brassica rapa*' OR '*Populus*' OR '*Solanum tuberosum*' OR '*Oryza sativa*' OR '*Rosa hybrida*' OR '*Glycine max*' OR '*Cucurbita pepo*' OR '*Beta vulgaris*' OR '*Saccharum*' OR '*Capsicum annuum*' OR '*Nicotiana tabacum*' OR '*Lycopersicon esculentum*' OR '*Triticum aestivum*') Limits our search to only articles which mention the genus and species name of GM crops that were commercially available in 2014.

**Part 2:** ('environmental assessment' OR 'environmental control' OR 'environmental degradation' OR 'environmental effects' OR 'environmental impact' OR 'environmental management' OR 'environmental protection' OR biosafety OR 'risk assessment' OR ecology OR ecological OR ecosystem OR biodiversity OR 'biological diversity' OR 'species diversity' OR 'species richness' OR 'biological input' OR 'biological output' OR landscape OR regional OR EIQ OR 'environmental impact quotient' OR 'ecosystem service' OR 'gene flow') Restricts our search to articles that mention terms related to environmental impacts.

**Part 3:** (GM OR GMO\* OR 'genetic engineering' OR 'genetic transformation' OR 'genetically engineered' OR 'genetic erosion' OR 'genetic contamination' OR 'genetic manipulation' OR 'genetically modified' OR 'genetic modification' OR transgenics OR transgenic OR transgenes OR introgression) Focuses our search on articles using genetic modification terminology.

Metadata was collected into a RefWorks account and downloaded into .csv files. All available PDF files were downloaded. For many unavailable files, a request for documents was made through interlibrary loan in an attempt to capture as many journal articles as possible. Files without selectable text were removed. Books and non-English PDFs were also removed from the database. Using the title of each article, duplicates and articles not matching metadata were removed.

## 2.2. Classification of stance

To classify journal articles as positive, neutral or negative stance, each article was randomly assigned a

number between 1 and 1873. Due to quality filtering, 114 of these articles were removed from subsequent classification and analyses because they did not match an article in our metadata or because they were a duplicate, resulting in a total of 1759 articles. Articles 1–500 were classified by at least four individuals. Articles were classified into one of five categories: 'Positive', 'Negative', 'Neutral', 'Does not discuss GM crops', and 'No consensus'. Individual examiners were trained to look for key terms suggesting stance in the context of genetic modification terminology (e.g. 'risks' or 'benefits' of GM foods or technology); full instructions for the examiners are in table S2. To ensure accountability, and to collect information on relevant sentences, when 'Positive' or 'Negative' was chosen, the examiners recorded the sentences that indicated the stance in their classification. In an effort to reduce the time required to classify articles, individuals were instructed to focus on the abstract, introduction, discussion, and conclusion, specifically on paragraphs that discussed GM crops. An article was classified as 'Does not discuss GM crops' if search terms did not match text in the main sections of the article (e.g. terms were exclusive to references). To verify that these disqualified articles did not discuss GM crops, a search was conducted within each article using three search terms ('transg', 'engin', 'modif'). Classification results were recorded in a Google spreadsheet using Google form submissions.

## 2.3. Data analysis

The stance of articles numbered 21 through 500 ( $n = 451$ ) was classified by four different individuals. We originally intended to analyze 500 articles, however some PDFs were removed because of poor quality and a subset was used for training purposes. Of the 451 journal articles analyzed, at least three of the four independent evaluators agreed on the stance of 230 articles. These 230 consensus articles were used in the final analyses.

The number of journal articles with a consensus stance of neutral, positive, or negative was standardized per year. A cumulative sum was calculated for each of the three stance categories, divided by the total articles for each year, and scaled with a minimum of zero and maximum of one, so that categories could be compared with one another. Metadata from RefWorks was used to determine the country of authorship for the first author and the primary source of funding, which we assumed was the first funding source listed. Funding was found searching for 'funding' or 'supported' or 'acknowledgements' or visually searching the end of the paper. Funding sources with 'national', 'federal', or 'ministry' were considered public. Non-profit funding was rare, so it was grouped with public funding (government agencies). Unknown funding sources and articles without funding sources were not included in the analysis

of funding sources, which occurred for 50 articles, leaving 180 articles for the funding analysis.

To examine the frequency that different topics were mentioned within the 230 articles with a consensus positive, negative, or neutral stance, we analyzed the occurrence of lowercase strings (sequences of characters) such as ' bt ' (with spaces between quotes), 'glyphosate', 'cancer' and Latin names from Part 1 of our search. Lowercase text was used for all articles. One-way analysis of variance (ANOVA; *scipy.stats* version 0.18.1) [38] was used to determine differences in the abundance of text strings in articles with different stances.

#### 2.4. Supervised classification

The same 230 articles that were used in the analyses described above were used in the supervised classification. All files were converted from PDF to plain text using AntFile Converter [39]. These files were then processed with the Biber tagger [40] to linguistically annotate each file, assigning each word in each file a part of speech (POS) label (table S4).

The accuracy of the Biber POS tagger has been established in several studies. Most recently, Gray, 2019 [41] evaluated the precision and recall of most of the linguistic features used in the present study. The majority of these features had an accuracy of between 95% and 100%. It is important to note that these rates include features such as relative clause structures, as in the text excerpt below, and different types of complement clauses (e.g. noun complements—'The idea that GM foods'; verb complements—'It is thought that GM foods'; Adjective complement clauses—'It is important that scientists'), and not just finite features such as definitive articles (a, an, the) or modals (will, can, should, might, etc).

These linguistically annotated files were then processed using a program that provides counts for the linguistic features in each text. This program also includes dictionaries to further annotate certain features. For example, verbs are labeled with semantic categories—e.g. as mental verbs (think, believe) or communication verbs (say, state)—and adjectives can be categorized as relating to size, time, color, evaluation. The counts for these features were normalized to allow for accurate comparisons across texts of unequal length. Each of the feature counts were normed to a rate of per 1000 words.

We used a random forest model to predict the stance of an article. Supervised classification of journal articles was performed using the python (version 3.6) package *sklearn* (version 0.18) [42]. One-way ANOVA was used to determine differences in the abundance of words beginning with 'transg' and to determine the best features for a random forest model. ANOVA results of each tag count of positive, negative, and neutral categories were ranked in order of increasing *p* value. The top 12 features were chosen to predict the stance of an article from

among the positive, negative, and neutral articles (figure 4, table S5). Articles were split, using the *train\_test\_split* function, into training (70%) and test (30%) to be used in a grid search to determine random forest parameters. The *GridSearchCV* function was used to find the ideal parameters among the number of estimators (200 or 500), number of features ('auto', 'sqrt', 'log2'), maximum depth (2–10), and split criterion ('gini' or 'entropy'). Using the best grid search parameters and a fixed random state (*RandomForestClassifier* function; *n\_estimators* = 200, *max\_features* = 'auto', *max\_depth* = 6, *criterion* = 'gini', *random\_state* = 42), we created a random forest for 50 random 70/30 training/test data splits (*train\_test\_split* function; *random\_state* = 0–49) to minimize the effect of any particular data split. Random chance for each training/test split was calculated using this function:  $(\text{Positive}/\text{Total})^2 + (\text{Negative}/\text{Total})^2 + (\text{Neutral}/\text{Total})^2$  where Positive, Negative, and Neutral refers to the number of respective labels in the testing data, and Total refers to the total number of testing labels. Percentages for random chance and accuracy (determined by the *accuracy\_score* function) were averaged for all 50 models.

#### 2.5. Analysis of references

We downloaded the number of times an article was cited and the citations for each article from Web of Science (<https://apps.webofknowledge.com>) on 7 November 2019 to determine the citation rates of positive and negative articles, and to test the hypothesis that articles tend to cite other articles with similar stance. Unfortunately, not all articles were in the WoS database and six of the eight databases did not provide citations. Consequently, only 186 articles could be included in the analysis of citation communities (i.e. collection of references within an article). Citation rate was calculated as the number of citations between the year of publication and 7 November 2019. These values were averaged for positive, negative, and neutral articles and ANOVA with Tukey's Honestly Significant Difference were performed (*MultiComparison* and *tukeyhsd* functions from *statsmodels*, version 0.9.0 [43]) to determine if there were differences for citation rates among the stances.

We performed an analysis to determine if any cited articles were significantly associated with a positive or negative stance. Separately, unique citations for positive and negative articles (*n* = 64; positive = 44, negative = 20) were used as column headers in a dataframe with article numbers representing rows. Indicator species analyses of citations were used to determine which references were associated with positive or negative stances using the 'multipatt' function in the 'indicspecies' package (version 1.7.6) in R (version 3.3.0) with default values [44]. We performed a post hoc classification of indicator

citations, however there were no indicator citations with a positive or negative stance consensus for these six articles.

Presence of specific citations within an article was represented with a 1, absence was denoted by a 0. The Jaccard dissimilarity index was used to compare the differences of the communities of citations. Values range from 0 to 1; a 0 indicates communities that are identical, a 1 indicates communities that are completely dissimilar. We performed alpha (i.e. number of citations per article) and beta diversity (e.g. change in citations between articles) calculations using the diversity functions within *skbio* (version 0.4.2). An ANOVA was used to compare alpha diversity among categories of stance while a permutational multivariate analysis of variance (PERMANOVA) was used to determine beta diversity differences. For beta diversity and PERMANOVA, citations that were only included in one article ( $n = 3805$ ) were removed, leaving 19 negative articles and 42 positive articles. Homogenous dispersion of variances was determined using 'betadisper' from the 'vegan' package (version 2.5-4) in R (version 3.3.0). The dispersion of variances were non-homogenous ( $F_{1,59} = 5.48, p = 0.03$ ). Because sample sizes were also different, we randomly sampled 19 positive articles 50 times and conducted a PERMANOVA, then averaged the  $F$  statistics and  $p$  values. Beta diversity was plotted using *seaborn* (version 0.9.0) and the *pcor* function from *skbio*.

### 3. Results

#### 3.1. Patterns in stance

Articles within our database discussed a total of 23 GM crops, the most common were: *Zea mays* (15%), *Brassica napus* (14%), *Populus* (14%), and *Oryza sativa* (13%) (table S3). From the 451 articles that were examined by four individuals, 65 (14.38%) did not discuss GM crops and were removed from further analyses. Consensus, where three or more individuals agreed on stance classification, was reached for 230 (50.88%) of the examined articles and of these, 139 (60.4%) were neutral, 60 (26.1%) were positive, and 31 (13.5%) were negative, indicating that 40% of the consensus articles contained a stance on GM crops (figure 1(a)). Articles with a negative stance were more common in the early 2000s and those with a neutral and positive stance had relatively steady rates of publication (figure 1(b)).

Data was summarized for countries and regions with the largest sample sizes, USA, China, and the EU, and all other countries were grouped together as 'other'. Authors from all countries and regions had a similar proportion of papers with a neutral stance (figure 1(c)). Authors from China had the highest positive:negative stance proportion (8:1), followed by authors from the USA (12:5) and other

countries (35:18). On the other hand, papers by authors from the EU were more negative than positive (5:7; figure 1(c)).

The relative abundance of neutral, positive and negative articles also differed among funding sources. The funding sources for 50 articles (22%) were unknown, 162 articles (70%) were supported by public funding and 18 (8%) were supported by private funding. Articles supported by public funding were 65% neutral, 20% positive, and 15% negative. In contrast, articles supported by private funding were 44% neutral, 50% positive, and 6% negative. Private funding sources were six times more likely in positive articles than negative articles (figure 1(d)).

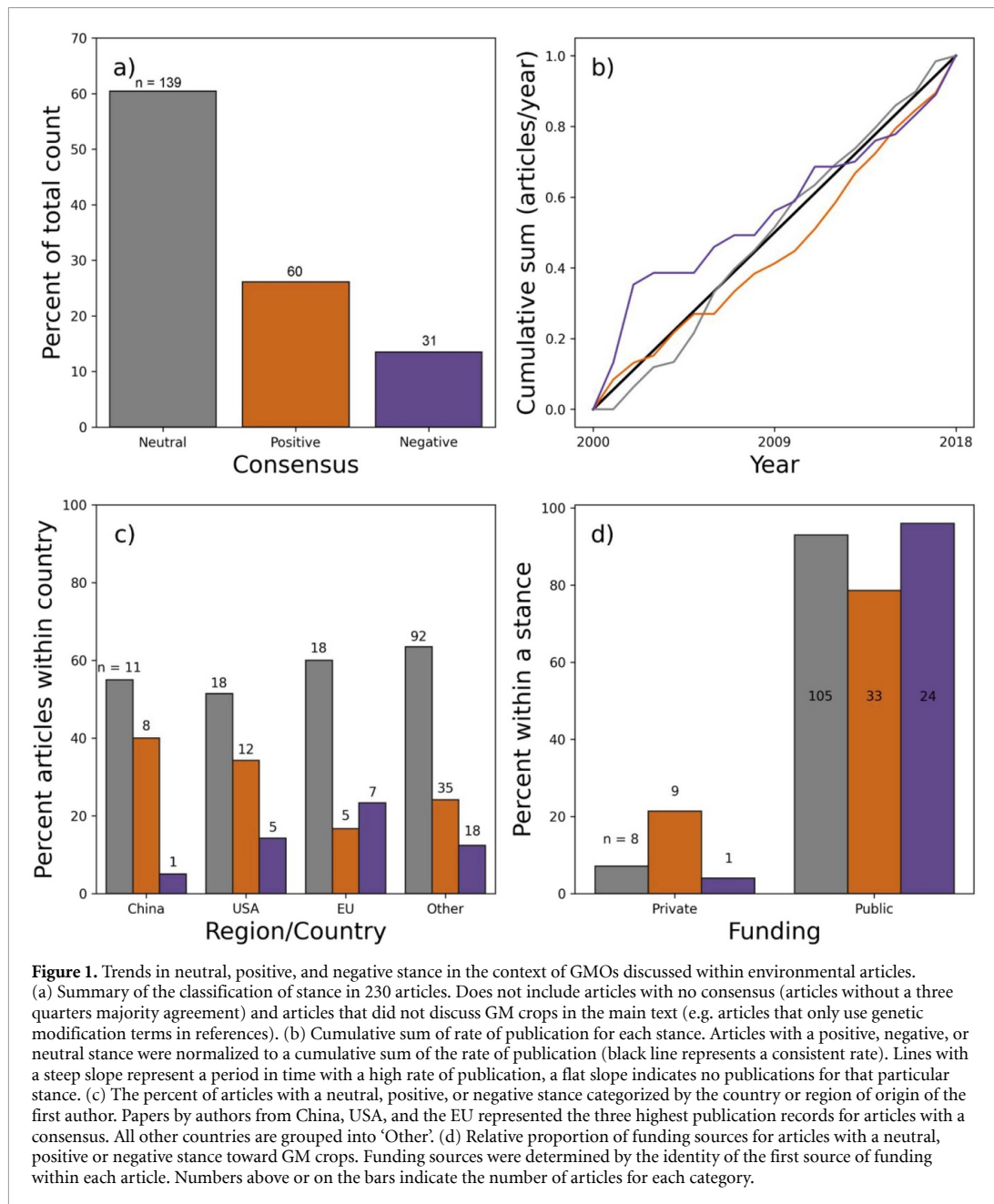
Within our database, the words 'gene flow' ( $F_{2,247} = 0.95, p = 0.39$ ) and 'cancer' ( $F_{2,247} = 1.35, p = 0.26$ ) were mentioned more frequently in articles with a negative stance than those with a neutral or positive stance, though not significantly more (figure 2). Bt crops were mentioned more frequently than glyphosate resistant crops and negative articles mentioned both Bt and glyphosate more often than neutral or positive articles ( $F_{2,247} = 2.00, p = 0.14$ ; figures 3(a) and (b)). The ratio of the number of times that Bt was mentioned relative to glyphosate was five times higher in positive articles than negative and neutral articles (figure 3(c)).

#### 3.2. Linguistic features

Characteristic linguistic features were identified in articles with a positive, negative, and neutral stance (figure 4). For example, some linguistic features that showed significant differences between positive and negative articles were *that* complement clauses, the definite article *the*, and modals of possibility (e.g. *may*, *might*, *could*). These constellations of features can be used to take a strong stand either for, or against a particular point of view—in this case they are used to cast doubt or raise possible problems with GM crops. The text excerpt from a negative article shows these features working together to cast doubt or caution against GM crops. *That* complements have been underlined, and possibility modals and *the* are bold.

However, some studies have shown that **the** host genotype **might** affect **the** community of microorganisms that establish a symbiotic relationship with **the** plant, and that any alteration in **the** microbial community diversity or activity **might** have significant effects on **the** plant's ability to grow and adapt [45].

In contrast, the excerpt below from a positive article reflects linguistic characteristics that were found to be significantly different from negative articles. For comparison of these two excerpts of similar length (46 and 47 words) that reflect the typical linguistic

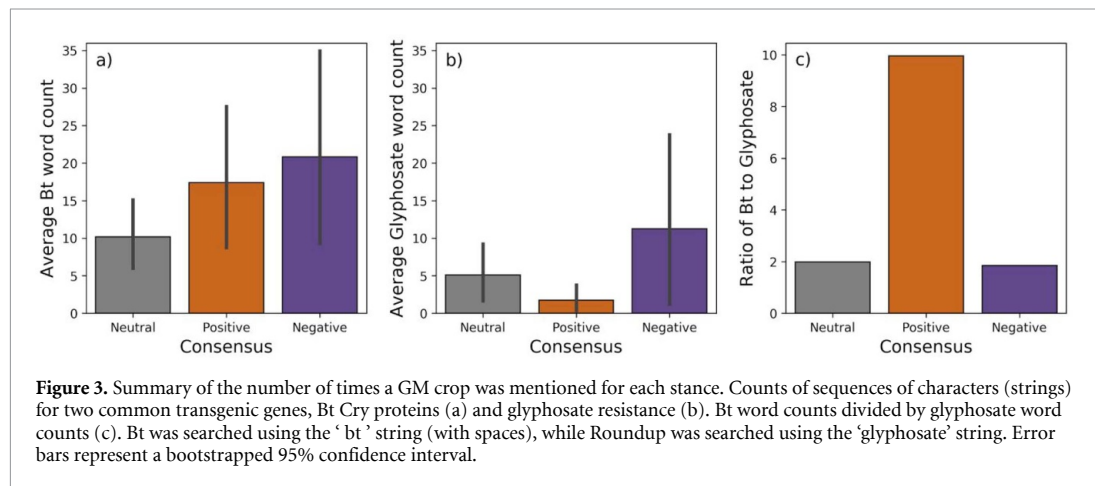
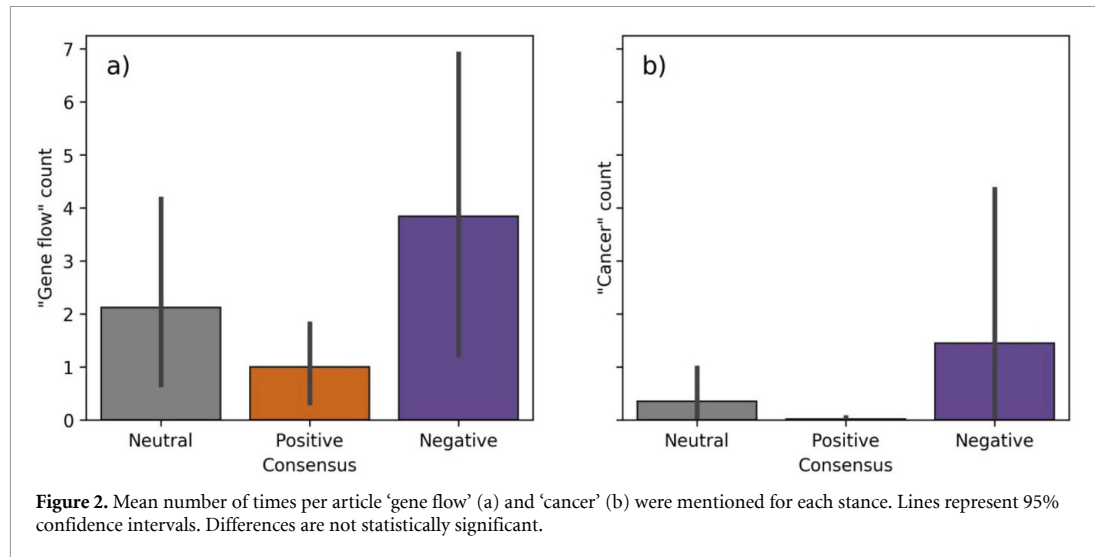


patterns found in negative and positive articles, the same features have been identified in both excerpts. In the excerpt below the proper nouns have been bolded and underlined. *The* is in bold.

**The** main pests of **poplar** (primarily **Lepidoptera**) were obviously inhibited in **the** transgenic **poplar**; however, **the** number of **Coleoptera** pests was generally low, and **the** inhibitory effect was not obvious. At **the** same time, there was little influence on **the** number of natural enemies and neutral insects [46].

Comparing features in these two samples we see that the use of the definite article *the* is the same in these two examples, five in each. However, the negative article sample has two *that* complement clauses while the positive article has none. Similarly, the positive article has four proper nouns while the negative article has none. The positive article does not have any possibility modals while the negative excerpt has two.

The linguistic analysis of sets of features helps to demonstrate how several linguistic features work together to craft points of view for or against a particular position, or to create a more neutral position. We identified a list of key linguistic features that can be used for automated classifications of



stance within peer-reviewed literature discussing GM crops (table S5 and figure 4). Supervised classification with random forest models using the top 12 linguistic features shown in figure 4 successfully predicted the stance of articles 64.2% of the time, which is significantly better than random chance (44%) for the top linguistic features reported in table S5 and visualized in figure 4. Therefore, using only 12 features, we predicted the stance of articles with a significantly better accuracy than random chance.

### 3.3. Citation patterns

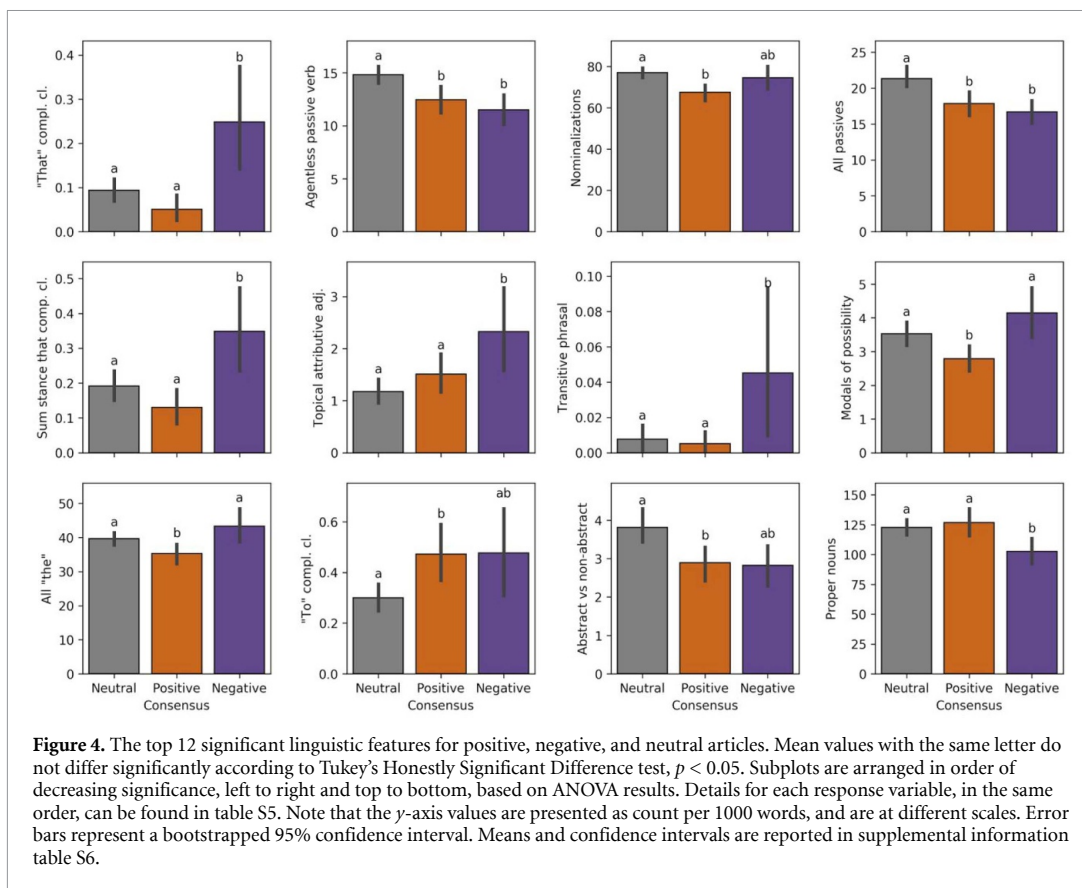
The reference sections of positive and negative articles were analyzed to determine if articles with a similar stance tend to have similar citations. A total of 3987 references were collected from 64 articles; positive articles contained 3205 citations (averaging 72.8 per article), while negative articles contained 999 (averaging 50.0 per article). There was no statistical difference in the number of citations per article (figure S1;  $F_{1,184} = 1.73$ ,  $p = 0.19$ ). Overall, 3805 references were cited once, 154 references were cited

twice, 20 references were cited three times, four references were cited four times, and three references were cited five times. Averaged PERMANOVA results for randomly sampled positive versus negative references indicate that there is a subtle yet significant difference between the bibliographies ( $F_{1,62} = 1.29$ ,  $p = 0.017$ ) (figure S2). Six articles were significantly associated with negative stance articles, according to an indicator species analysis, five of which discussed gene flow (table S7).

## 4. Discussion

Our results indicate that 40% of articles discussing GM crops and the environment have a positive or negative stance, with twice as many articles with a positive stance (26.1%) than a negative stance (13.5%; figure 1(a)). The proportions of peer-reviewed journal articles with detectable stance toward GM crops vary with year of publication, authors' country of origin, funding source and type of genetic modification. Distinctive linguistic features





within articles of a particular stance (figure 4) suggest that positive and negative articles can be detected by parsing the meaning and choice of words within an article, and we used these features to successfully predict the stance of an article.

Our survey reveals variation in scientific stance about GM crops over time. Articles with a negative stance toward GM crops were more likely to be published in the beginning of the millennium, while positive and neutral articles had relatively stable publication rates (figure 1(b)). Changes in stance over time could represent a shift in scientific consensus on the safety of GM crops. The trend in negative stance articles could be influenced by the negative public attitudes toward GM food in the late 90s that began to dissipate in the early 2000s [47–49].

The ratio of positive:negative articles about GM crops varies among countries and regions (figure 1(c)), and may or may not corroborate public opinion polls. The ratio for authors from the European Union (5:7) reflects the negative public opinion about GM products and political efforts to ban GM crops in many EU countries [50]. In contrast, there is a disparity between public opinion about GM crops and the stance of scientific publications by Chinese authors. Articles by Chinese authors had the highest ratio of positive to negative stance (8:1), yet a recent study indicates that 41.4% of the Chinese

public oppose and only 11.9% support GM crops [13, 51]. This incongruence could reflect the push for investment in GM crops from China's government, as China was the first country to approve a commercial GM crop [52, 53]. Alternatively, English is a second language for most Chinese authors, and this may influence subtle linguistic features which inadvertently makes the stance of their writing more positive. More research is needed to investigate potential linkages between the political environment and the stance of scientists.

The proportion of papers with positive and negative stance varied with funding sources and type of genetic modification. Positive articles were three and six times more likely to be funded by private sources than neutral and negative stance articles, respectively (figure 1(d)). Our study discovered some common themes of positive and negative stance toward GM crops. Early concerns over the potential for modified genes from GM crops to assimilate into populations of wild relatives could account for our finding that negative articles were four times more likely to mention gene flow (figure 2(a)). The ongoing debate about the long-term safety of glyphosate exposure [54] may be reflected in the observation that articles with a negative stance were 85 times more likely to mention cancer compared to positive articles (figure 2(b)). On the other hand, a positive stance

toward GM crops in our database could relate to generally more positive attitudes toward Bt crops, as opposed to glyphosate-resistant crops (figure 3). The genetic modification Bt is considered to be a relatively safe substitute for toxic pesticides [9], and positive articles mentioned Bt ten times more often than glyphosate, compared to negative and neutral articles which mention Bt only twice as often.

Machine learning has the potential to automatically screen manuscripts. The linguistic features identified here could be applied broadly to predict the stance of any body of English texts discussing GM crops. Together, the results of this analysis of bias can help us understand features behind the debate of the safety of GM crops, and also develop automated methods to screen manuscripts for unintended stance. Our machine learning algorithm could benefit from improved accuracy, possibly with more data, or neural network classifiers. During the classification, a significant portion of the journal articles did not have consensus partly because we skimmed articles and did not analyze each article in depth. Also, one person may not pick up on the same linguistic features as another person. Some articles contained linguistic features that indicated both positive and negative stances within the same article. The scope of this study was limited to a small subset of journal articles that discussed GM crops within the context of the environment. Our method could be also improved by including articles from other controversial topics, such as climate change. If similar linguistic features are used in other topics, an automated tool could be used to quickly pinpoint linguistic features that may have the unintended consequence of taking a stance rather than objectively reporting facts, if that is considered desirable. Whether stance on scientific topics are appropriate for peer-reviewed journal articles is beyond the scope of this study.

## 5. Conclusions

Our analysis of the stance toward GM crops within peer-reviewed scientific literature highlights characteristics of writing and the contextual environment. Despite the limitations of this study, we discovered a significant portion of peer-reviewed journal articles on GM crops contained a positive or negative stance, and that research funded by private corporations was six times more likely to have a positive stance. We also learned that articles discussing the Bt genetic modification corresponded to positive stances. Public opinion polls and the indirect opinions highlighted in our study suggest that political and social factors may play a role in the ratio of positive to negative stance articles from different regions. Though, just how much the political and social environment influences the stance of scientists remains a mystery.

## Data availability statement

All metadata for classified articles, the articles themselves, and random forest model code will be made available on the Dryad Digital Repository upon acceptance.

The data that support the findings of this study are available at <https://gitlab.com/usda-ars/gmo-stance/>.

## Acknowledgments

This research was initiated during a graduate seminar on Ecosystem Health and GMOs (BIO 698) when we noticed evidence of bias within peer reviewed scientific literature. Thanks to undergraduate interns, Parker Leber, Cat Watkins, Morgan Brion, Haley Harris, and Kayla Thompson, for their help downloading and classifying articles. Another thanks goes to David Minkler for his sage advice and classification of articles. We also thank Mary DeJong for her help in developing the journal article search strings and identifying articles in library databases. Special thanks to Doug Biber for his help tagging our texts. This research was partially funded by a grant to BMS from Northern Arizona University's School of Earth and Sustainability and support to NCJ through a Bullard Fellowship from Harvard Forest. Finally, BMS would like to thank the Skeptics' Guide to the Universe for planting the seed that grew into this publication.

## Author contributions

B M S and M L conceived the idea for this study. N C J and B S helped organize and facilitate the course. B M S, A R, A P, K G, and M L collected data. B M S and R R contributed to the linguistic analysis. B M S produced all figures and statistics. B M S, N C J, and R R interpreted the data. B M S and N C J wrote the first version of the manuscript. All authors but M L reviewed the manuscript.

## Conflict of interest

B M S and M L have a positive stance toward G M crops. N C J and A R have a negative stance toward G M crops. A P, K G, and R R have a neutral stance toward G M crops.

## Materials and correspondence

For linguistic methods, please contact Randi Reppen ([randi.reppen@nau.edu](mailto:randi.reppen@nau.edu)). For all other methods, contact Bo Maxwell Stevens ([bo\\_stevens@nau.edu](mailto:bo_stevens@nau.edu)).

## ORCID iD

Bo Maxwell Stevens  <https://orcid.org/0000-0002-6249-3459>

## References

- [1] Funk C and Rainie L 2015 Public and scientists' views on science and society *Pew Res. Cent.* **29**
- [2] Wunderlich S and Gatto K A 2015 Consumer perception of genetically modified organisms and sources of information *Adv. Nutr.* **6** 842–51
- [3] Benbrook C M 2016 Trends in glyphosate herbicide use in the United States and globally *Environ. Sci. Eur.* **28** 3
- [4] Ji J, Chao N and Ding J 2019 Rumormongering of genetically modified (GM) food on Chinese social network *Telemat. Inform.* **37** 1–12
- [5] Dixon G 2016 Applying the gateway belief model to genetically modified food perceptions: new insights and additional questions *J. Commun.* **66** 888–908
- [6] Kerr J R and Wilson M S 2018 Changes in perceived scientific consensus shift beliefs about climate change and GM food safety *PLoS One* **13** e0200295
- [7] Rabin M and Schrag J L 1999 First impressions matter: a model of confirmatory bias *Q. J. Econ.* **114** 37–82
- [8] FDA 2020 Science and history of GMOs and other food modification processes (FDA)
- [9] Bravo A, Likitvivatanavong S, Gill S S and Soberón M 2011 *Bacillus thuringiensis*: a story of a successful bioinsecticide *Insect Biochem. Mol. Biol.* **41** 423–31
- [10] Sanchis V 2011 From microbial sprays to insect-resistant transgenic plants: history of the biopesticide *Bacillus thuringiensis*. A review *Agron. Sustain. Dev.* **31** 217–31
- [11] Gonsalves D 1998 Control of papaya ringspot virus: a case study *Annu. Rev. Phytopathol.* **36** 415–37
- [12] Paine J A *et al* 2005 Improving the nutritional value of golden rice through increased pro-vitamin A content *Nat. Biotechnol.* **23** 482–7
- [13] Bawa A S and Anilakumar K R 2013 Genetically modified foods: safety, risks and public concerns—a review *J. Food Sci. Technol. Mysore* **50** 1035–46
- [14] European Commission 2010 A decade of EU-funded GMO research (2001–2010) *Dir. Res. Innov. Biotechnol. Agric. Food* (Brussels: European Union)
- [15] Koch M S, Ward J M, Levine S L, Baum J A, Vicini J L and Hammond B G 2015 The food and environmental safety of *Bt* crops *Front. Plant Sci.* **6** 283
- [16] Nicolia A, Manzo A, Veronesi F and Rosellini D 2014 An overview of the last 10 years of genetically engineered crop safety research *Crit. Rev. Biotechnol.* **34** 77–88
- [17] Klümper W and Qaim M 2014 A meta-analysis of the impacts of genetically modified crops *PLoS One* **9** e111629
- [18] Zhang C, Wohlhueter R and Zhang H 2016 Genetically modified foods: a critical review of their promise and problems *Food Sci. Hum. Wellness* **5** 116–23
- [19] Zilberman D, Holland T G and Trilnick I 2018 Agricultural GMOs—what we know and where scientists disagree *Sustainability* **10** 1514
- [20] Powles S B 2008 Evolved glyphosate-resistant weeds around the world: lessons to be learnt *Pest Manage. Sci.* **64** 360–5
- [21] van Bruggen A H C *et al* 2018 Environmental and health effects of the herbicide glyphosate *Sci. Total Environ.* **616–617** 255–68
- [22] Gastiazoro M P, Durando M, Milesi M M, Lorenz V, Vollmer G, Varayoud J and Zierau O 2020 Glyphosate induces epithelial mesenchymal transition-related changes in human endometrial Ishikawa cells via estrogen receptor pathway *Mol. Cell. Endocrinol.* **510** 110841
- [23] Tarone R E 2018 On the international agency for research on cancer classification of glyphosate as a probable human carcinogen *Eur. J. Cancer Prev.* **27** 82–7
- [24] Ellstrand N C, Prentice H C and Hancock J F 1999 Gene flow and introgression from domesticated plants into their wild relatives *Annu. Rev. Ecol. Syst.* **30** 539–63
- [25] Quist D and Chapela I H 2001 Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico *Nature* **414** 541–3
- [26] Snow A A, Andersen B and Jørgensen R B 1999 Costs of transgenic herbicide resistance introgressed from *Brassica napus* into weedy *B. rapa* *Mol. Ecol.* **8** 605–15
- [27] Losey J E, Rayor L S and Carter M E 1999 Transgenic pollen harms monarch larvae *Nature* **399** 214
- [28] de Vaulleury A, Kramarz P E, Binet P, Cortet J, Caul S, Andersen M N, Plumey E, Coeurdassier M and Krogh P H 2007 Exposure and effects assessments of *Bt*-maize on non-target organisms (gastropods, microarthropods, mycorrhizal fungi) in microcosms *Pedobiologia* **51** 185–94
- [29] Devos Y, de Schrijver A, de Clercq P, Kiss J and Romeis J 2012 *Bt*-maize event MON 88017 expressing Cry3Bb1 does not cause harm to non-target organisms *Trans. Res.* **21** 1191–214
- [30] Romeiras M M, Vieira A, Silva D N, Moura M, Santos-Guerra A, Batista D, Duarte M C and Paulo O S 2016 Evolutionary and biogeographic insights on the Macaronesian *Beta-patellifolia* species (Amaranthaceae) from a time-scaled molecular phylogeny *PLoS One* **11** e0152456
- [31] Kovach J, Petzoldt C, Degni J and Tette J 1992 A method to measure the environmental impact of pesticides
- [32] Li F 2008 Annual report readability, current earnings, and earnings persistence *J. Account. Econ.* **45** 221–47
- [33] Markowitz D M and Hancock J T 2014 Linguistic traces of a scientific fraud: the case of Diederik Stapel *PLoS One* **9** e105937
- [34] Markowitz D M and Hancock J T 2016 Linguistic obfuscation in fraudulent science *J. Lang. Soc. Psychol.* **35** 435–45
- [35] Rude S, Gortner E-M and Pennebaker J 2004 Language use of depressed and depression-vulnerable college students *Cogn. Emot.* **18** 1121–33
- [36] Wiltsey Stirman S and Pennebaker J W 2001 Word use in the poetry of suicidal and nonsuicidal poets *Psychosom. Med.* **63** 517–22
- [37] ISAAA 2014 GM crops list
- [38] Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in python *Nat. Methods* **17** 261–72
- [39] Anthony L 2015 *AntfileConverter*
- [40] Biber D 2006 *University Language: A Corpus-based Study of Spoken and Written Registers* (Amsterdam: John Benjamins Publishing)
- [41] Gray B 2019 *Linguistic Variation in Research Articles* (John Benjamins)
- [42] Pedregosa F *et al* 2011 Scikit-learn: machine learning in python *J. Mach. Learn. Res.* **12** 2825–30
- [43] Seabold S and Perktold J 2010 Statsmodels: econometric and statistical modeling with python *Proc. 9th Python Sci. Conf.* vol 57 p 61
- [44] de Cáceres M and Legendre P 2009 Associations between species and groups of sites: indices and statistical inference *Ecology* **90** 3566–74
- [45] da Silva K J, de Armas R D, Soares C R and Ogluari J B 2016 Communities of endophytic microorganisms in different developmental stages from a local variety as well as transgenic and conventional isogenic hybrids of maize *World J. Microbiol. Biotechnol.* **32** 189
- [46] Zuo L, Yang R, Zhen Z, Liu J, Huang L and Yang M 2018 A 5-year field study showed no apparent effect of the *Bt* transgenic 741 poplar on the arthropod community and soil bacterial diversity *Sci. Rep.* **8** 1–13
- [47] Cameron J 2007 Consumer reactions to products of new animal technologies presentation to: advisory committee on biotechnology and 21st century agriculture (AC21) November (available at: <https://slideplayer.com/slide/4702596/>)

- [48] Gaskell G *et al* 2003 Ambivalent GM nation? Public attitudes to biotechnology in the UK, 1991–2002 21
- [49] Vilella-Vila M and Costa-Font J 2008 Press media reporting effects on risk perceptions and attitudes towards genetically modified (GM) food *J. Socio-Econ.* **37** 2095–106
- [50] Tosun J and Debus M 2020 Right-wing populist parties and environmental politics: insights from the Austrian Freedom Party's support for the glyphosate ban *Environ. Polit.* **0** 1–21
- [51] Cui K and Shoemaker S P 2018 Public perception of genetically-modified (GM) food: a Nationwide Chinese Consumer study *Npj Sci. Food* **2** 1–8
- [52] Huang J, Hu R, Cai J and Wang X 2012 Human research capacity in Chinese agbiotech *Nat. Biotechnol.* **30** 1007
- [53] Jia S and Peng Y 2002 GMO biosafety research in China *Environ. Biosafety Res.* **1** 5–8
- [54] Mesnage R and Antoniou M N 2017 Facts and fallacies in the debate on glyphosate toxicity *Front. Public Health* **5** 316