Analysis of scientific society-awarded honors reveals disparities

A DOI-citable version of this manuscript is available at https://doi.org/10.1101/2020.04.14.927251.

This manuscript (permalink) was automatically generated from greenelab/iscb-diversity-manuscript@45c6d5b on August 5, 2021.

Authors

- **Trang T. Le**
  0000-0003-3737-6565 · trang1618 · trang1618
  Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, University of Pennsylvania

- **Daniel S. Himmelstein**
  0000-0002-3012-7446 · dhimmel · dhimmel
  Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania · Funded by the Gordon and Betty Moore Foundation (GBMF4552)

- **Ariel A. Hippen**
  0000-0001-9336-6543 · arielah · ariel_hippen
  Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania

- **Matthew R. Gazzara**
  0000-0001-7710-4551 · mrgazzara · MR_Gazzara
  Department of Genetics, Perelman School of Medicine, University of Pennsylvania

- **Casey S. Greene**
  0000-0001-8713-9213 · cgreene · greenescientist
  Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania; Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation · Funded by the Gordon and Betty Moore Foundation (GBMF4552)

Highlights

- proportion of women honorees was similar to that of the field
- scientists of East Asian origin have been underrepresented among honorees
- disparities arise partly, but not exclusively, from geography
- honorees with an affiliation in the US were overrepresented by a factor of 2.0

Summary

Delivering a keynote talk at a conference organized by a scientific society, or being named as a fellow by such a society, indicates that a scientist is held in high regard by their colleagues. To explore if the distribution of such indicators of esteem in the field of bioinformatics reflects the composition of this field, we compared the gender, name-origin and country of affiliation of 412 honorees from the *International Society for Computational Biology* (75 fellows and 337 keynote speakers) with over

170,000 last authorships on computational biology papers between 1993 and 2019. The proportion of honors bestowed on women was similar to that of the field's overall last authorship rate. However, names of East Asian origin have been persistently underrepresented among honorees. Moreover, there were roughly twice as many honors bestowed on scientists with an affiliation in the United States as expected based on literature authorship.

A record of this paper's Transparent Peer Review process is included in the Supplemental Information.

Introduction

Scientists' roles in society include identifying important topics of study, undertaking an investigation of those topics, and disseminating their findings broadly. The scientific enterprise is largely self-governing: scientists act as peer reviewers on papers and grants, comprise hiring committees in academia, make tenure decisions, and select which applicants will be admitted to doctoral programs. A lack of diversity in science could lead to pernicious biases that hamper the extent to which scientific findings are relevant to minoritized communities. Furthermore, even though minoritized groups innovate at higher rates, their novel contributions are discounted (Hofstra et al., 2020). One first step to address this systemic issue is to directly examine peer recognition in different scientific fields.

Gender bias among conference speakers has been recognized as an area that can be improved with targeted interventions (Klein et al., 2017; Langin, 2019; Martin, 2015; Martin, 2014). Having more female organizers on conference committees is associated with having more female speakers (Casadevall and Handelsman, 2014). At medical conferences in the US and Canada, the proportion of female speakers is increasing at a modest rate (Ruzycki et al., 2019). Gender bias appears to also influence funding decisions: an examination of scoring of proposals in Canada found that reviewers asked to assess the science produced a smaller gender gap in scoring than reviewers asked to assess the applicant (Witteman et al., 2019).

Challenges extend beyond gender: an analysis of awards at the NIH found that proposals by Asian, black or African-American applicants were less likely to be funded than those by White applicants (Ginther et al., 2011). There are also potential interaction effects between gender and race or ethnicity that may particularly affect women of color's efforts to gain NIH funding (Ginther et al., 2016). Another recent analysis found that minority scientists tend to apply for awards on topics with lower success rates (Hoppe et al., 2019). This finding might be the result of minority scientists selecting topics in more poorly funded areas. Alternatively, reviewing scientists may not recognize the scientific importance of these topics, which may be of particular interest to minority scientists.

We sought to understand the extent to which honors and high-profile speaking invitations were distributed equitably among gender and name origin groups by an international society and its associated meetings. As computational biologists, we focused on the International Society for Computational Biology (ISCB), its honorary Fellows as well as its affiliated international meetings that aim to have a global reach: Intelligent Systems for Molecular Biology (ISMB) and Research in Computational Molecular Biology (RECOMB).

Existing methods were relatively US-centric because most of the data was derived in whole or in part from the US Census. We scraped more than 700,000 entries from English-language Wikipedia that contained nationality information and built machine learning classifiers to predict name origin. We also examined last authorships for more than 170,000 computational biology publications to establish a field-specific baseline using the same metrics. We used methods to predict the gender and name origins of honor recipients and also examined the affiliations of authorships and honor recipients. Analysis of affiliations by country revealed disparities between authorships and honoree affiliations. We also observed fewer honors to scientists with East Asian name origin than expected from

authorship set, an effect that persisted even after we controlled for affiliation by restricting analysis to only US-affiliated scientists.

Results

We curated a dataset of ISCB honors that included 412 keynote speakers at international ISCB-associated conferences (ISMB and RECOMB) as well as ISCB Fellows. The ISCB Fellows set contained the complete set of Fellows named (2009–2019). Keynote speakers were available for ISMB for all years from 1993–2019. Keynote speakers for RECOMB were available for all years from 1997–2019. We included individuals who were honored multiple times as separate entries.

We sought to compare this dataset with a background distribution of potential speakers, which we considered to be last authors of bioinformatics and computational biology manuscripts (see Methods). We used authorships instead of authors as our metric for the field's composition under the expectation that honorees would be drawn in a manner weighted by the number of last author contributions. Using authorships also does not require accurate name disambiguation, which is an open challenge. We scraped PubMed for manuscripts written in English from 1993–2019 with the MeSH term "computational biology". We downloaded the metadata of manuscripts published in these journals from PubMed, which provided 176,110 articles for evaluation. For each article, we extracted its last author's fore name and last name for analysis.

Defining metrics and a field-specific background

We were faced with a number of choices as we set out to examine representation within the field. Here, we outline the choices that we made and the rationale for these choices.

For one step, we needed to determine whether the most appropriate unit for analysis was researchers or honor and authorship events. We elected to perform the analysis at the level of honors and authorships, not individuals. Our rationale was that, if a scientist was honored three times, these three honoree slots represented distinct selection processes and should be considered separately. We also elected to use authorships instead of authors because name disambiguation can be error prone, and because the overall design of the work was to examine authorship distributions, the precise linking of researchers was not necessary to address the research question. As such, we estimated the gender and origin of each last name of each given paper and honoree slot, but not how many papers an individual scientist has authored or how many times they were invited as speakers or named fellows. We also needed to determine whether or not authorships should be weighted by some property of the resulting manuscript, for example the number of citations. We elected not to weight based on the concern that certain honors - in particular keynotes - could increase the exposure of the work and consequently increase citations which would add circularity to the analysis.

At another point, we needed to determine what the appropriate set of publications were to define a field-specific background distribution. Defining this distribution is key to determining the extent to which representation diverges from that background. Practical options included selecting authors in key society-endorsed journals or selecting those based on some article metadata. We elected to perform the analysis to all 176,110 PubMed articles with a computational biology MeSH term to generate the largest possible set of relevant literature.

We assumed that, among the authors of a specific paper, the senior author (often research advisors) would be most likely to be invited for a keynote or honored as a fellow. Based on field-specific conventions, we could have selected the last or corresponding author. We found that corresponding authors were often first and/or last suggesting that they were highly involved with more corresponding authors being last authors than first authors. Given the convention and the

observation that corresponding authors were more often last, we selected the last author as the most appropriate in this setting.

We present the results with the selected strategy (Fig. 1); however, we performed this work through multiple iterations varying these parameters to examine the extent to which they influenced the results. Our findings across different combinations of the above choices were consistent with respect to the broad conclusions, though the numerical results differed (see Methods).

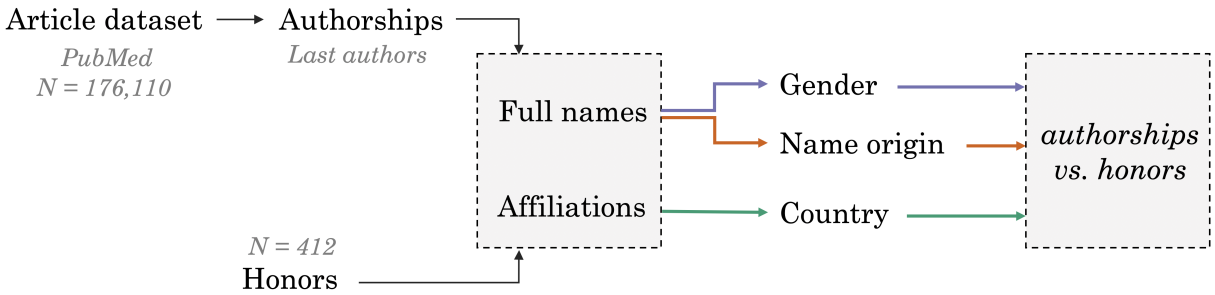

Figure 1: Study framework. We extracted full names and affiliations of the last authors of 176,110 computational biology PubMed articles and those of 412 honorees. We estimated the gender, name-origin group and country of affiliation of each scientist and compared the probability values between these two groups (see Methods).

Similar gender proportion between ISCB's honorees and the field

We predicted the gender of honorees and authors using the https://genderize.io API, which was trained on over 100 million name-gender pairings collected from the web (see the STAR Methods for more details) and is one of the three widely-used gender inference services (Santamaría and Mihaljević, 2018). The predictions represent the estimated probability of an honoree or author being male or female based on their first name; we did not convert probabilities to a hard group assignment. For example, a query to https://genderize.io on January 26, 2020 for "Casey" returns a probability of male of 0.74 and a probability of female of 0.26, which we would add for an author with this first name. Because of technical limitations, our analysis only considered two binary gender categories, and we used "male" and "female" to refer to the gender of the scientists. However, as described in the Discussion, we recognize the limitation of not accounting for non-binary gender categories and only considered predictions in aggregate and not as individual values for specific scientists.

We observed a gradual increase of the proportion of predicted female authorships, arriving at an average of approximately 28% in 2017-2019 (Fig. 2, left). In recent years, ISCB Fellows and keynote speakers appear to have similar gender proportions compared to the population of authors published in computational biology and bioinformatics journals (averaged around 30% in the last three years, Fig. 2, right). Examining each honor category, we observed in 10.visualize-gender an increasing trend of honorees who were women, especially in the group of ISCB Fellows, which markedly increased after 2015. Through 2019, there were a number of years when meetings or ISCB Fellow classes have a high proportion of honors for male honorees and none that appeared to have exclusively female honorees. We sought to examine whether or not there was a difference in the proportion of female names between authorships and honors. A multiple logistic regression of this proportion for the groups and year did not reveal a significant difference ($p = 0.19$). Interaction terms did not predict the group of scientists over and above the main effect of gender probability and year.

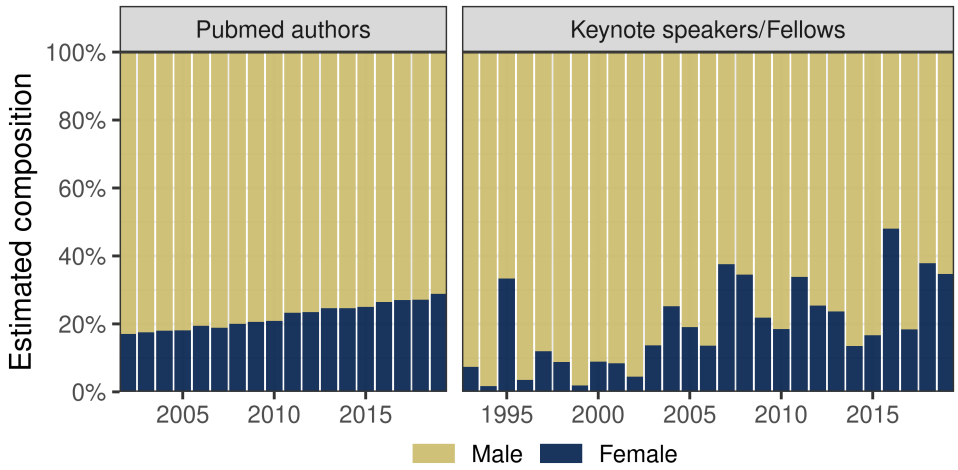

Figure 2: Estimated composition of gender prediction over the years of all PubMed computational biology and bioinformatics journal authorships (left), and all ISCB honors (right). Male proportion (yellow) was computed as the average of the probability of being male of last authors (weight accordingly) or ISCB honorees each year. Female proportion (blue) was the complement of the male proportion. ISCB honors appear to have similar gender proportions compared to that of PubMed authorships.

Honorees with Celtic/English names are overrepresented while honorees with East Asian names are underrepresented

We inferred the geographical region of origin of authors' names using a Long Short-Term Memory (LSTM) neural network trained on a dataset of 708,493 names called Wiki2019 (see the STAR Methods for details); the resulting model was called Wiki2019-LSTM. We found that the proportion of authorships with Celtic/English names had decreased (Fig. 3A, left). Among keynote speakers and fellows, we found that the majority were predicted to have Celtic/English or European names (Fig. 3A, right). When we directly compared honor composition with PubMed, we observed discrepancies between the two groups (Fig. 3B). Compared to other names, there was an overabundance of Celtic/English name among the honors ($OR_{Celtic/English}$ = 2.39, $\beta_{Celtic/English}$ = 0.86, $p < 10^{-5}$). Meanwhile, an East Asian name has significantly lower odds of being selected for an honor compared to other names ($OR_{East\ Asian}$ = 0.17, $\beta_{East\ Asian}$ = -1.75, $p < 10^{-5}$). The two groups of scientists did not have a significant association with names predicted to be European and in Other categories (p = 0.11 and p = 0.29, respectively). Interaction terms did not predict the group of scientists over and above the main effect of name origin probability and year.

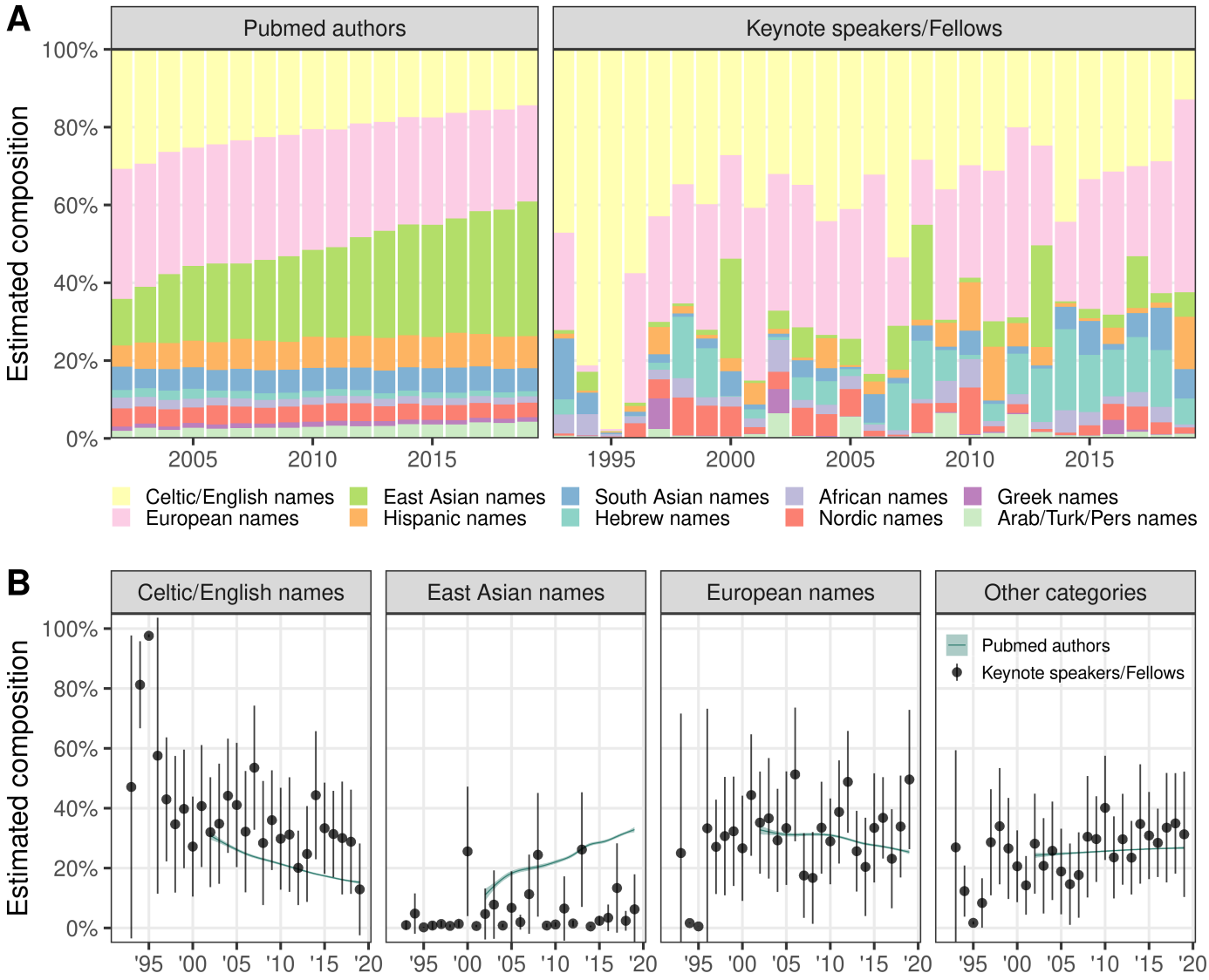

Figure 3: (A) Estimated composition of name origin prediction over the years of all PubMed computational biology and bioinformatics journal authorships (left), and all ISCB honors (right). (B) For each region, the mean predicted probability of PubMed authorships is shown as the teal generalized additive model curve. The mean probability and 95% confidence interval of the ISCB honor predictions are shown as dark circles and vertical lines. Compared to PubMed authorships, honors were more frequently associated with Celtic/English names and less frequently with East Asian names. We did not observe a statistically significant difference in other categories (see STAR Methods Table 1).

Consistent disparities were also apparent with an alternative approach that considered each pair of forename and last name was the unit of measurement. In addition, a time-lagged model, which might be appropriate if we assume that honors accrue ten years after an author's most prolifically cited year, results in a similar underrepresentation of East Asian scientists' names in the group of honorees, though the effect size is smaller. For example, the proportion of honor associated with East Asian name origins in 2019 is still substantially less than the proportion of senior authorships associated with East Asian names in 2009. The 11.visualize-name-origins analysis notebook for this portion provides the results in these scenarios.

We sought to disentangle geography from other factors by examining results for the country with the most affiliated scientists receiving honors, the US. When applying the Wiki2019-LSTM model to the name origins of only US-affiliated scientists, we found a similar underrepresentation of honors to scientists with East Asian names ($OR_{East\ Asian}$ = 0.15, $\beta_{East\ Asian}$ = -1.89, $p = 3.4\times10^{-5}$). We observed no statistically significant difference between the proportion of honors given to authors with Celtic/English names, European names, or names in Other categories (p = 0.15, p = 0.02, and p = 0.65, respectively). Please see the 14.us-name-origin notebook for more details of the US-specific analysis.

Overrepresentation of US-affiliated honorees

We analyzed the countries of affiliation between last authorships and ISCB honors. For each country, we report a value of $\log_2$ enrichment (LOE) and its 95% confidence intervals (see Methods). The full table with all countries and their corresponding enrichment can be browsed interactively in the corresponding 12.analyze-affiliation analysis notebook. A positive value of LOE indicates a higher proportion of honorees affiliated with that country compared to authors. A LOE value of 1 indicates that observed number of honors is twice as much as expected. In the 20 countries with the most publications, we found an overrepresentation of honorees affiliated with institutions and companies in the US (119 speakers more than expected, LOE = 1.0, 95% CI (0.8, 1.2)) and Israel (15 speakers more than expected, LOE = 2.7, 95% CI (1.9, 3.4)), and an underrepresentation of honorees affiliated with those in China, France, Italy, India, South Korea, and Brazil (Fig. 4).

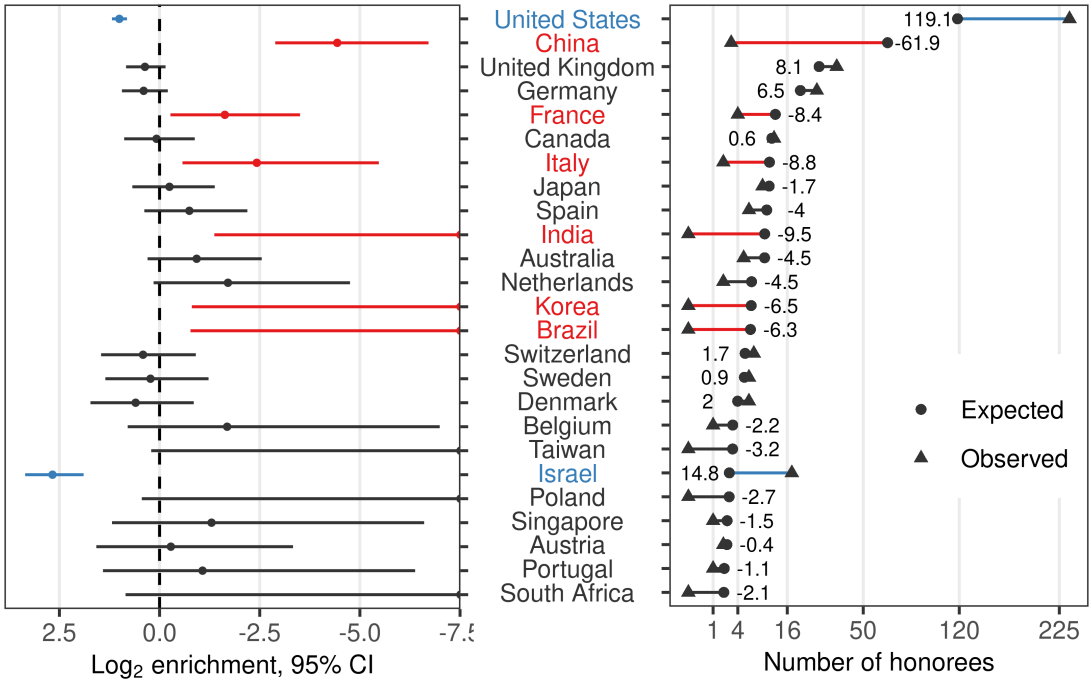

Figure 4: Each country's $\log_2$ enrichment (LOE) and its 95% confidence interval (left), and the absolute difference between observed (triangle) and expected (circle) number of honors (right). Positive value of LOE indicates a higher proportion of honorees affiliated with that country compared to authors. Countries are ordered based on the proportion of authorships in the field. The overrepresentation of honorees affiliated with institutions and companies in the US and Israel contrasts the underrepresentation of honorees affiliated with those in China, France, Italy, India, South Korea, and Brazil.

Improvements to Honoree Diversity Subsequent to Our Primary Analysis

While our study was primarily designed to assess the diversity of honor recipients, the findings raise important questions about what can be done to address the disparities. We examined changes subsequent to our initial report for suggestions that increased awareness may drive improvements in the practice of honoree suggestion. We released version 1.0 of our manuscript on 2020-01-30. Early indications suggested an increased the diversity of honorees. In 2020, among 12 ISCB Fellows and 5 ISMB keynote speakers, the mean predicted probability of each honoree having an East Asian name was 33%, higher than any estimate in previous years (see the 15.analyze-2020 notebook). The set of honorees also included the first ISCB Fellow from China. Compared to past years, the 2020 honorees appeared to better reflect the diversity of scientists in the computational biology field. These new results suggested: 1) deserving honorees who were members of under-recognized groups existed but had not been recognized, and 2) examining honoree distribution's alignment with the field may trigger

changes that begin to address issues of unequal representation. However, we note that this analysis dealt only with more senior scientists (the last authors on scientific manuscripts) in the context of honors and that many years of changed honoree distributions will be required for the set of honored scientists to better reflect the field's senior author contributions.

Discussion

There are significant technical and ethical challenges that one faces in carrying out retrospective work to examine the fairness of scientific practices. A major technical challenge was to narrow down geographic origins for some groups of names. Specific name origin groups, such as Hispanic names, are geographically disparate. We were unable to construct a classifier that could distinguish names from Iberian countries (Spain and Portugal) from those in Latin America in the group of Hispanic names. Discrepancies in representation between these groups are thus undetectable by our classifier. Honoree counts of those with Hispanic names are influenced from Spain as well as Latin America. In such cases, our analyses may understate the extent to which minoritized scientists are underrepresented among honorees and authors. Another technical challenge is that supervised machine learning approaches are neither error free nor bias free. By integrating different lines of evidence and preserving uncertainty by analyzing prediction probabilities rather than applying a hard assignment for each prediction, we aimed to alleviate method-specific biases and discover insightful findings that correctly reflect the current representation diversity at conferences.

A key ethical challenge with retrospective work to examine disparities is that algorithmic approaches to infer characteristics are often required. This leads to significant limitations such as considering gender as a binary variable. While this situation limits retroactive examination, there is a substantial opportunity for scientific societies, grant-making organizations, publishers, and others to proactively collect self-identified demographic information. It is equally crucial that this information is properly used to continuously evaluate inclusion practices.

A difficulty that straddles the ethical and technical divide for studies examining the representation of honorees is that the background that can be best assessed, which we assess here, is the current field composition. Scientific societies exist to promote the discipline, and many, including ISCB, include diversity as a value. In these cases, the ideal background distribution would be the population of senior computational biologists in the absence of systemic barriers to participation. The implication would be that scientific societies exist to reflect what the field could be, not just what it is. However, we are limited to measuring the field as it is. Furthermore, authorships, which we use to assess the field's composition, are also affected by systemic barriers to participation. We estimated the composition of the field using last author status, but in neuroscience (Shen et al., 2018) and other disciplines (Holman et al., 2018), women are underrepresented in this position. Such an effect would cause us to underestimate the number of women in the field. Similarly, other studies have showed that underrepresented groups are less likely to be last authors (Marschke et al., 2018), and Hispanic and Black scientists were underrepresented in academic publishing in general (Hopkins et al., 2012). Thus, systemic barriers that reduce representation within our estimation of the field would reduce apparent disparities in honor distributions as long as those systemic barriers did not also have a particular influence on the honoree selection process as well.

An important ethical question to ask when measuring representation is what the right level of representation is. Societies should examine their processes to determine whether the process of selecting honorees should be equal or equitable. For example, we found similar representation of women between authors and honorees, which suggests honoree diversity is similar to that of authors and that there may be equality during the honoree selection process. However, if fewer women are in the field because of systemic factors that inhibit their participation, reaching equality is not equivalent to reaching equity. In addition to holding fewer corresponding authorship positions, on average,

female scientists of different disciplines are cited less often Larivière et al. (2013), invited by journals to submit papers less often (Holman et al., 2018), suggested as reviewers less often (Lerback and Hanson, 2017), and receive significantly worse review scores (Fox and Paine, 2019). Meanwhile, a review of women's underrepresentation in math-intensive fields argued that today's underrepresentation is not explained by historic forms of discrimination but factors surrounding fertility decisions and lifestyle choices, whether freely made or constrained by biology and society (Ceci and Williams, 2011). A recent analysis of gender inequality across different disciplines showed that, although both gender groups have equivalent annual productivity, women scientists have higher dropout rates throughout their scientific careers (Huang et al., 2020). Therefore, although we found that ISCB's honorees and keynote speakers appear to have similar gender proportion to the field as a whole, the gender proportions have not reached parity. To the extent that this gap is due to systemic barriers, the process may have reached equality but not equity.

It is also possible to have processes that reach neither equality nor equity. We find that honorees include significantly fewer people of color than the field as a whole, and Asian scientists are dramatically underrepresented among honorees. Because invitation and honor patterns could be driven by biases associated with name groups, geography, or other factors, we cross-referenced name group predictions with author affiliations to disentangle the relationship between geographic regions, name groups and invitation probabilities. We found that disparities persisted even within the group of honorees with a US affiliation. Societies' honoree selection process failing to reflect the diversity of the field can play a part in why minoritized scientists' innovations are discounted (Hofstra et al., 2020). Although we estimate the fraction of non-White and non-Asian authors to be relatively similar to the estimated honoree rate, we note that both are represented at levels substantially lower than in the US population.

Societies, both through their honorees and the individuals who deliver keynotes at their meetings, can play a positive role in improving the presence of female STEM role models, which can boost young students' interests in STEM (Ceci and Williams, 2011) and, for example, lead to higher persistence for undergraduate women in geoscience (Hernandez et al., 2018). Efforts are underway to create Wikipedia entries for more female (Wade and Zaringhalam, 2018) and black, Asian, and minority scientists (O'Reilly, 2019), which can help early-career scientists identify role models. Societies can contribute toward equity if they design policies to honor scientists in ways that counter these biases such as ensuring diversity in the selection committees.

The central role that scientists play in evaluating each other and each other's findings makes equity critical. Even many nominally objective methods of assessing excellence (e.g., h-index, grant funding obtained, number of high-impact peer-reviewed publications, and total number of peer-reviewed publications) are subject to the bias of peers during review. These could be affected by explicit biases, implicit biases, or pernicious biases in which a reviewer might consider a path of inquiry, as opposed to an individual, to be more or less meritorious based on the reviewer's own background Murray et al. (2019). Our efforts to measure the diversity of honorees in an international society suggests that, while a focus on gender parity may be improving some aspects of diversity among honorees, contributions from scientists of color are underrecognized.

Acknowledgments

We would like to thank the Gordon and Betty Moore Foundation whose support makes the study possible (GBMF4552 to D.S.H. and GBMF4552 to C.S.G.). We thank the reviewers and editors for their insightful input to improve the quality of the study.

Author Contributions

Conceptualization, C.S.G.; Methodology, T.T.L., D.S.H., A.A.H., M.R.G. and C.S.G.; Software, T.T.L., D.S.H., A.A.H. and M.R.G.; Investigation T.T.L., D.S.H., A.A.H., M.R.G. and C.S.G.; Data Curation, T.T.L., D.S.H., A.A.H., M.R.G. and C.S.G.; Writing – Original Draft, T.T.L, D.S.H and C.S.G; Writing – Review & Editing, T.T.L and C.S.G; Visualization, T.T.L and D.S.H; Supervision, C.S.G; Project Administration, C.S.G; Funding Acquisition, D.S.H and C.S.G

Declaration of Interests

The authors declare no competing interests.

STAR Methods

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Data and code Zenodo deposit	This paper	https://doi.org/10.5281/zenodo.5014756
Source code on GitHub	This paper	https://github.com/greenelab/iscb-diversity
Browsable Python and R notebooks	This paper	https://greenelab.github.io/iscb-diversity/
PubmedPy	This paper	https://github.com/dhimmel/pubmedpy
Manubot	Himmelstein et al., 2019	https://doi.org/10.1371/journal.pcbi.1007128
Name origin prediction	This paper	https://github.com/greenelab/wiki-nationality-estimate
Geotext	Yaser Martinez Palenzuela yaser.martinez@gmail.com	https://github.com/elyase/geotext
Geopy Nominatim	https://github.com/geopy/geopy	https://geopy.readthedocs.io/en/stable/#nominatim
Genderize.io	https://genderize.io/	https://genderize.io/
Other		
Manuscript repository	This paper	https://github.com/greenelab/iscb-diversity-manuscript

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Casey Greene (casey.s.greene@cuanschutz.edu).

Materials Availability

This study did not generate new materials.

Data and Code Availability

All data have been deposited at https://github.com/greenelab/iscb-diversity and are publicly available as of the date of publication. DOIs are listed in the key resources table. Our Wikipedia name dataset is dedicated to the public domain under CC0 License at https://github.com/greenelab/wiki-nationality-estimate, with source code to construct the dataset available under a BSD 3-Clause License.

All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table. Specifically, our analysis of authors and ISCB-associated honorees is available under CC BY 4.0 at https://github.com/greenelab/iscb-diversity, with source code also distributed under a BSD 3-Clause License (Le et al., 2021). Rendered Python and R notebooks from this repository are browsable at greenelab.github.io/iscb-diversity. Our analysis of PubMed, PubMed Central, and author names relies on the Python pubmedpy package, developed as part of this project and available under a Blue Oak Model License 1.0 at https://github.com/dhimmel/pubmedpy and on PyPI. No additional information is required to reanalyze the data reported in this paper.

This manuscript was written openly on GitHub at github.com/greenelab/iscb-diversity-manuscript using Manubot (Himmelstein et al., 2019). The Manubot HTML version is available under a Creative Commons Attribution (CC BY 4.0) License at greenelab.github.io/iscb-diversity-manuscript.

Method details

Honoree Curation

From ISCB's webpage listing **ISCB Distinguished Fellows**, we found recipients listed by their full names for the years 2009–2019. We gleaned the full name of the Fellow as well as the year in which they received the honor. We identified major ISCB-associated conferences as those designated flagship (ISMB) or those that had been held on many continents (RECOMB). To identify **ISMB Keynote Speakers**, we examined the webpage for each ISMB meeting. The invited speakers at ISMB before 2001 were listed in the Preface pages of each year's proceedings, which were archived in the ISMB collection of the AAAI digital library. We found full names of all keynote speakers for the years 1993–2019.

For the RECOMB meeting, we found conference webpages with keynote speakers for 1999, 2000, 2001, 2004, 2007, 2008, and 2010–2019. We were able to fill in the missing years using information from the RECOMB 2016 proceedings, which summarizes the first 20 years of the RECOMB conference (2016). This volume has two tables of keynote speakers from 1997–2006 (Table 14, page XXVII) and 2007–2016 (Table 4, page 8). Using these tables to verify the conference speaker lists, we arrived at two special instances of inclusion/exclusion. Although Jun Wang was not included in these tables, we were able to confirm that he was a keynote speaker in 2011 with the RECOMB 2011 proceedings (2011), and thus we included this speaker in the dataset. Marian Walhout was invited as a keynote speaker but had to cancel the talk due to other obligations. Because her name was neither mentioned in the 2015 proceedings (2015) nor in the above-mentioned tables, we excluded this speaker from our dataset.

Name processing

When extracting honoree names, we began with the full name as provided on the site. Because our prediction methods required separated first and last names, we chose the first non-initial name as the first name and the final name as the last name. We did not consider a hyphen to be a name separator: for hyphenated names, all components were included. For metadata from PubMed and PMC where first (fore) and last names are coded separately, we applied the same cleaning steps. We created functions to simplify names in the pubmedpy Python package to support standardized fore and last name processing.

Last author extraction

We assumed that, in the list of authors for a specific paper, last authors (often research advisors) would be most likely to be invited for keynotes or to be honored as fellows. Therefore, we utilized PubMed to retrieve last author names to assess the composition of the field. PubMed is a search engine resource provided by the US National Library of Medicine and index scholarly articles. PubMed contains a record for every article published in journals it indexes (30 million records total circa 2020), within which we were able to extract author first and last names and their order using the E-Utilities APIs. To automate and generalize these tasks, we created the pubmedpy Python package.

From PubMed, we compiled a catalog of 176,773 journal articles that were published from 1993 through 2019 that were written in English and tagged with the MeSH term "computational biology", which is equivalent to "bioinformatics" and includes categories such as genomics and systems biology (via PubMed's term explosion to include subterms). Excluding 663 articles with no author information and years with less than 200 articles/year, we analyzed 176,110 articles from 1998–2019. We extracted the number of times an article has been cited by PubMed Central articles from the `PmcRefCount` of the PubMed DocSum XML records.

Countries of Affiliations

Publications often provide affiliation lists for authors, which generally associate authors with research organizations and their corresponding physical addresses. We implemented affiliation extraction in the pubmedpy Python package for both PubMed and PMC XML records. These methods extract a sequence of textual affiliations for each author.

We relied on two Python utilities to extract countries from text: geotext and `geopy.geocoders.Nominatim`. The first, geotext, used regular expressions to find mentions of places from the GeoNames database. To avoid mislabeling, we only mapped the affiliation to a country if geotext identified 2 or more mentions of that country. For example, in the affiliation string "Laboratory of Computational and Quantitative Biology, Paris, France", geotext detected 2 mentions of places in France: Paris, France. In this case, we assign France to this affiliation.

This country extraction method accommodates multiple countries. Although ideally each affiliation record would refer to one and only one research organization, sometimes journals deposit multiple affiliations in a single structured affiliation. In these cases, we assigned multiple countries to the article. For more details on this approach, please consult the accompanying 07.affiliations-to-countries notebook and label dataset.

When geotext did not return results, we use the geopy approach, which returns a single country for an affiliation when successful. Its `geocoders.Nominatim` function converts names / addresses to geographic coordinates using the OpenStreetMap's Nomatim service. With this method, we split a textual affiliation by punctuation into a list of strings and iterate backward through this list until we found a Nomatim search result. For the above affiliation, the search order would be "France", "Paris", and "Laboratory of Computational and Quantitative Biology". Since Nomatim would return a match for the first term "France" (matched to France), the search would stop before getting to "Paris", and "France" would be assigned to this affiliation.

Our ability to assign countries to authors was largely driven by the availability of affiliations. The country-assignment-rate for last authors from PubMed records was approximately 47%. This reflects the varying availability of affiliation metadata by journal.

For ISCB honorees, during the curation process, if an honoree was listed with their affiliation at the time, we recorded this affiliation for analysis. For ISCB Fellows, we used the affiliation listed on the ISCB page. Because we could not find affiliations for the 1997 and 1998 RECOMB keynote speakers' listed for these years, they were left blank. If an author or speaker had more than one affiliation, each was inversely weighted by the number of affiliations that individual had.

Estimation of Gender

We predicted the gender of honorees and authors using the https://genderize.io API, which was trained on over 100 million name-gender pairings collected from the web and is one of the three widely-used gender inference services (Santamaría and Mihaljević, 2018). We used author and honoree first names to retrieve predictions from genderize.io. The predictions represent the probability of an honoree or author being male or female. We used the estimated probabilities and did not convert to a hard group assignment. For example, a query to https://genderize.io on January 26, 2020 for "Casey" returns a probability of male of 0.74 and a probability of female of 0.26, which we would add for an author with this first name. Because of technical limitations, our analysis only considered two binary gender categories, and we used "male" and "female" to refer to the gender of the scientists. However, we recognized the limitation of not accounting for non-binary gender categories and only considered predictions in aggregate and not as individual values for specific scientists.

Of 412 ISCB honorees, genderize.io fails to provide gender predictions for one name. Of 176,110 last authors, 1,014 were missing a fore name in the raw paper metadata and 11,498 had fore names consisting of only initials. Specifically, the metadata for most papers before 2002 (2,566 out of 2,601 papers) only have initials for first and/or middle author names. Without gender predictions for these names, we consider only articles from 2002 on when comparing gender compositions between two groups. Of the remaining authors, genderize.io failed to predict gender for 10,003 of these fore names. We note that approximately 42% of these NA predictions are hyphenated names, which is likely because they are more unique and thus are more difficult to find predictions for. This bias of NA predictions toward non-English names has been previously observed (Wais, 2016) and may have a minor influence on the final estimate of gender compositions.

Estimation of Name Origin Groups

We developed a model to predict geographical origins of names. The existing Python package ethnicolr (Sood and Laohaprapanon, 2018) produces reasonable predictions, but its international representation in the data curated from Wikipedia in 2009 (Ambekar et al., 2009) is still limited. For instance, 76% of the names in ethnicolr's Wikipedia dataset are European in origin.

To address these limitations in ethnicolr, we built a similar classifier, a Long Short-term Memory (LSTM) neural network, to infer the region of origin from patterns in the sequences of letters in full names. We applied this model on an updated, approximately 4.5 times larger training dataset called Wiki2019 (described below). We tested multiple character sequence lengths and, based on this comparison, selected tri-characters for the primary results described in this work. We trained our prediction model on 80% of the Wiki2019 dataset and evaluated its performance using the remaining 20%. This model, which we term Wiki2019-LSTM, is available in the online file `LSTM.h5`.

To generate a training dataset for name origin prediction that reflects a modern naming landscape, we scraped the English Wikipedia's category of Living People. This category, which contained approximately 930,000 pages at the time of processing in November 2019, is regularly curated and allowed us to avoid pages related to non-persons. For each Wikipedia page, we used two strategies to find a full birth name and location context for that person. First, we looked for nationality mention in the first sentence in the body of the text. In most English-language biographical Wikipedia pages, the

first sentence usually begins with, for example, "John Edward Smith (born 1 January 1970) is an American novelist known for ..." This structure comes from editor guidance on biography articles and is designed to capture:

> ... the country of which the person is a citizen, national or permanent resident, or if the person is notable mainly for past events, the country where the person was a citizen, national or permanent resident when the person became notable.

Second, if this information is not available in the first sentence of the main text, we used information from the personal details sidebar; the information in this sidebar varied widely but often contained a full name and a place of birth.

We used regular expressions to parse out the person's name from this structure and checked that the expression after "is a" matched a list of nationalities. We were able to define a name and nationality for 708,493 people by using the union of these strategies. This process produced country labels that were more fine-grained than the broader patterns that we sought to examine among honorees and authors. We initially grouped names by continent, but later decided to model our categorization after the hierarchical taxonomy used by NamePrism (Ye et al., 2017). The NamePrism taxonomy was derived from name-country pairs by producing an embedding of names by Twitter contact patterns and then grouping countries using the similarity of names from those countries. The countries associated with each grouping are shown in supplementary figure S1. NamePrism excluded the US, Canada and Australia because these countries have been populated by a mix of immigrant groups (Ye et al., 2017).

In an earlier version of this manuscript, we also used category names derived from NamePrism, but a reader pointed out the titles of the groupings were problematic; therefore, in this version, we renamed these groupings to reflect that the NamePrism approach primarily identifies groups based on linguistic patterns from name etymology rather than religious or racial similarities. We note that our mapping from nationality to name origins was not without error. For example, a scientist of Israeli nationality may not bear a Hebrew name. These mismatches were assessed via the heatmap of the model performance (supplementary figure S2) and complemented by the affiliation analysis below. An alternative approach is to assign arbitrary names to these groups such as via letter coding (e.g., A, B, C, etc.), but we did not choose this strategy because ten arbitrary letters for ten groups can greatly reduce the paper's readibility.

Predicting Name Origin Groups with LSTM Neural Networks and Wikipedia

Table 1 shows the size of the training set for each of the name origin groups as well as a few examples of PubMed author names that had at least 90% prediction probability in that group. We refer to this dataset as Wiki2019 (available online in `annotated_names.tsv`).

Table 1: Predicting name-origin groups of names trained on Wikipedia's living people. The table lists the 10 groups and the number of living people for each region that the LSTM was trained on. Example names shows actual author names that received a high prediction for each region. Full information about which countries comprised each region can be found in the online dataset `country_to_region.tsv`.

Group	Training Size	Example Names
Celtic/English names	154,890	Adam O Hebb, Oliver G Pybus, David W Ritchie, James WJ Anderson, James W MacDonald, Robert Clarke
European names	78,157	Tracey M Filzen, Jos H Beijnen, Caroline Louis-Jeune, Christian Lorenzi, Boris Vassilev, Verena Heinrich
Hispanic names	66,931	Ramón Latorre, Antonio J Jimeno-Yepes, Felipe A Simão, Paulo S L de Oliveira, Juan Carlos Rodríguez-Manzaneque, Natalia Acevedo-Luna

Group	Training Size	Example Names
East Asian names	54,588	Heejoon Chae, Wenchao Jiang, Haizhou Liu, Miho Uchida, Wenxuan Zhang, Jiali Feng
Arabic/Turkic/Persian names	31,418	Hamidreza Chitsaz, Farzad Sangi, Habib Motieghader, Berke Ç Toptas, Ali Aliyari, Bülent Arman Aksoy
Nordic names	28,978	Cecilia M Lindgren, Ellen Larsen, Jesper R Gådin, Janne H Korhonen, Johan Åqvist, Jens Nilsson
South Asian names	20,025	Amitabh Chak, Matthew G Seetin, Matrika Gupta, Sumudu P Leelananda, VS Kumar Kolli, Swanand Gore
African names	17,826	Timothy Kinyanjui, Jammbe Z Musoro, Nyaradzo M Mgodi, Magambo Phillip Kimuda, Probhonjon Baruah, Adaoha E C Ihekwaba
Hebrew names	4,549	Alexander J Sadovsky, Boris Shraiman, Gil Goldshlager, Eytan Adar, Aviva Peleg, Nir Esterman
Greek names	4,138	Gianni Panagiotou, Themis Lazaridis, Eleni Mijalis, Nikolaos Tsiantis, Konstantinos A Kyritsis, Dimitris E Messinis

We next aimed to predict the name origin groups of honorees and authors. We constructed a training dataset with more than 700,000 name-nationality pairs by parsing the English-language Wikipedia. We trained a LSTM neural network on n-grams to predict name groups. We found similar performance across 1, 2, and 3-grams; however, because the classifier required fewer epochs to train with 3-grams, we used this length in the model that we term Wiki2019-LSTM. Our Wiki2019-LSTM returns, for each given name, a probability of that name originating from each of the specified 10 groups. We observed a multiclass area under the receiver operating characteristic curve (AUC) score of 95.9% for the classifier, indicating that the classifier can recapitulate name origins with high sensitivity and specificity. For each individual group, the high AUC (above 92%) suggests that our classifier was sufficient for use in a broad-scale examination of disparities. We also observed that the model was well calibrated. We also examined potential systematic errors between pairs of name origin groupings with a confusion heatmap and did not find off-diagonal enrichment for any pairing (supplementary figure S2).

Applying Wiki2019-LSTM on the author and honoree datasets, we obtained name origin estimates for all honorees' and authors' name, except the 12,512 that did not have fore names (see breakdown in the Estimation of Gender section above). Once again, because the large majority of author fore names prior to 2002 were recorded with initials only, predictions were not possible, and we excluded 1998–2001 when comparing name origin compositions between two groups.

Affiliation Analysis

For each country, we computed the expected number of honorees by multiplying the proportion of authors whose affiliations were in that country with the total number of honorees. We then performed an enrichment analysis to examine the difference in country affiliation proportions between ISCB honorees and field-specific last authors. We calculated each country's enrichment by dividing the observed proportion of honorees by the expected proportion of honorees. The 95% confidence interval of the $\log_2$ enrichment was estimated using the Poisson model method (Sahai and Khurshid, 1996).

Statistical Analysis

We estimated the levels of representation by performing the following logistic regression of the group of scientists on each name's prediction probability while controlling for year:

$$g = \beta_0 + \beta_1 prob + \beta_2 y + \epsilon.$$

The variable $prob$ is the prediction probablity of a demographic variable (gender and name origin) for names of scientists in group g (honoree or author) during year y. $\epsilon \sim N(0, \sigma^2)$ accounts for random variation. An effect was deemed statistically significant if its corresponding $p < \alpha = 0.01$.

We emphasize that the units in this analysis are honors and authorships. Therefore, each row of the input data frame represents either an honor or authorship with the scientist's name's probability value from the gender or name origin classifier ($prob$). We also performed an alternative approach in which each pair of forename and last name as the unit of measurement and the citations were totaled across different papers whose last author had that name. Controlling for each name's citations, this method tested the demographic effects but did not account for names honored more than once and may be affected by name collisions.

To reiterate, we only consider the prediction probabilities in aggregate and not as individual values for specific scientists. Moreover, although the average of the probabilities is not exactly "proportion", we use the phrase "estimated proportion" for readability. For example, the average of the probabilities of authors having an East Asian name origin is the estimate for the proportion of authors with East Asian names.

Iterative Research Process

Parallel analyses for the other versions are available in supplementary figure S3. In our first version of the analysis pipeline, we sought to characterize the distribution of authorships in the field using field-specific journals. This resulted in an analysis set of 29,755 authorships. We also examined corresponding authors, as we considered that senior authors may occasionally occupy a different position (Larivière et al., 2016), and only fell back to last authors in cases where corresponding author annotations were unavailable.

In the next version of the analysis, we extended the analysis to all 176,110 computational biology PubMed articles, substantially increasing the sample size. We also extracted names of last authors instead of other potential selections to better capture the honoree population. Our assumption was that, among the authors of a specific paper, the last author (often research advisors) would be most likely to be invited for a keynote or honored as a fellow. Also, the availability of information on corresponding authors was limited, and extracting last author became a more consistent approach.

In version 3, instead of weighting all articles equally as in the earlier versions, we used citation count to weight articles to control for the differential impact of research contributions. Using citation counts has key limitations: female scientists of different disciplines are cited less often than their male counterparts (Caplar et al., 2017; Dworkin et al., 2020; Fox and Paine, 2019). Furthermore, the act of being honored, particularly with a keynote at an international meeting, could lead to work being more recognized and cited, which would reverse the arrow of causality. In version 4, we returned to the equal weight for all articles.

Finally, in all versions of the analysis, rather than applying a hard assignment for each prediction, we analyzed the raw prediction probability values to capture the uncertainty of the prediction model. Although we expect our estimates of disparities for citation-weighted analyses to be conservative, through each analysis, the overall findings remained consistent. Examining the literature with and without citation weighting, we learned that disparities exist and these disparities are large enough to overcome existing disparities in citation patterns.

References

(2011). Research in Computational Molecular Biology (Springer Science and Business Media LLC).

(2015). Research in Computational Molecular Biology (Springer Science and Business Media LLC).

(2016). Research in Computational Molecular Biology (Springer Science and Business Media LLC).

Ambekar, A., Ward, C., Mohammed, J., Male, S., and Skiena, S. (2009). Name-ethnicity classification from open sources. Association for Computing Machinery (ACM).

Caplar, N., Tacchella, S., and Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. Nature Astronomy *1*, 0141.

Casadevall, A., and Handelsman, J. (2014). The Presence of Female Conveners Correlates with a Higher Proportion of Female Speakers at Scientific Symposia. mBio *5*.

Ceci, S.J., and Williams, W.M. (2011). Understanding current causes of women's underrepresentation in science. Proceedings of the National Academy of Sciences *108*, 3157–3162.

Dworkin, J.D., Linn, K.A., Teich, E.G., Zurn, P., Shinohara, R.T., and Bassett, D.S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists (arXiv).

Fox, C.W., and Paine, C.E.T. (2019). Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution. Ecology and Evolution *9*, 3599–3619.

Ginther, D.K., Schaffer, W.T., Schnell, J., Masimore, B., Liu, F., Haak, L.L., and Kington, R. (2011). Race, Ethnicity, and NIH Research Awards. Science *333*, 1015–1019.

Ginther, D.K., Kahn, S., and Schaffer, W.T. (2016). Gender, Race/Ethnicity, and National Institutes of Health R01 Research Awards. Academic Medicine *91*, 1098–1107.

Hernandez, P.R., Bloodhart, B., Adams, A.S., Barnes, R.T., Burt, M., Clinton, S.M., Du, W., Godfrey, E., Henderson, H., Pollack, I.B., et al. (2018). Role modeling is a viable retention strategy for undergraduate women in the geosciences. Geosphere.

Himmelstein, D.S., Rubinetti, V., Slochower, D.R., Hu, D., Malladi, V.S., Greene, C.S., and Gitter, A. (2019). Open collaborative writing with Manubot. PLOS Computational Biology *15*, e1007128.

Hofstra, B., Kulkarni, V.V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., and McFarland, D.A. (2020). The Diversity–Innovation Paradox in Science. Proceedings of the National Academy of Sciences *117*, 9284–9291.

Holman, L., Stuart-Fox, D., and Hauser, C.E. (2018). The gender gap in science: How long until women are equally represented? PLOS Biology *16*, e2004956.

Hopkins, A.L., Jawitz, J.W., McCarty, C., Goldman, A., and Basu, N.B. (2012). Disparities in publication patterns by gender, race and ethnicity based on a survey of a random sample of authors. Scientometrics *96*, 515–534.

Hoppe, T.A., Litovitz, A., Willis, K.A., Meseroll, R.A., Perkins, M.J., Hutchins, B.I., Davis, A.F., Lauer, M.S., Valantine, H.A., Anderson, J.M., et al. (2019). Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. Science Advances *5*, eaaw7238.

Huang, J., Gates, A.J., Sinatra, R., and Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. Proceedings of the National Academy of Sciences

117, 4609–4616.

Klein, R.S., Voskuhl, R., Segal, B.M., Dittel, B.N., Lane, T.E., Bethea, J.R., Carson, M.J., Colton, C., Rosi, S., Anderson, A., et al. (2017). Speaking out about gender imbalance in invited speakers improves diversity. Nature Immunology *18*, 475–478.

Langin, K. (2019). How scientists are fighting against gender bias in conference speaker lineups. Science.

Larivière, V., Ni, C., Gingras, Y., Cronin, B., and Sugimoto, C.R. (2013). Bibliometrics: global gender disparities in science. Nature *504*, 211–213.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., and Sugimoto, C.R. (2016). Contributorship and division of labor in knowledge production. Social Studies of Science *46*, 417–435.

Le, T., Himmelstein, D., and Greene, C. (2021). greenelab/iscb-diversity: ISCB analysis v4.0 release (Zenodo).

Lerback, J., and Hanson, B. (2017). Journals invite too few women to referee. Nature *541*, 455–457.

Marschke, G., Nunez, A., Weinberg, B.A., and Yu, H. (2018). Last Place? The Intersection of Ethnicity, Gender, and Race in Biomedical Authorship. AEA Papers and Proceedings *108*, 222–227.

Martin, G. (2015). Addressing the underrepresentation of women in mathematics conferences (arXiv).

Martin, J.L. (2014). Ten Simple Rules to Achieve Conference Speaker Gender Balance. PLoS Computational Biology *10*, e1003903.

Murray, D., Siler, K., Larivière, V., Chan, W.M., Collings, A.M., Raymond, J., and Sugimoto, C.R. (2019). Author-Reviewer Homophily in Peer Review. Cold Spring Harbor Laboratory.

O'Reilly, N. (2019). Why we're creating Wikipedia profiles for BAME scientists. Nature d41586-019-00812-8.

Ruzycki, S.M., Fletcher, S., Earp, M., Bharwani, A., and Lithgow, K.C. (2019). Trends in the Proportion of Female Speakers at Medical Conferences in the United States and in Canada, 2007 to 2017. JAMA Network Open *2*, e192103.

Sahai, H., and Khurshid, A. (1996). Statistics in epidemiology: methods, techniques, and applications (Boca Raton: CRC Press).

Santamaría, L., and Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. PeerJ Computer Science *4*, e156.

Shen, Y.A., Webster, J.M., Shoda, Y., and Fine, I. (2018). Persistent Underrepresentation of Women's Science in High Profile Journals. Cold Spring Harbor Laboratory.

Sood, G., and Laohaprapanon, S. (2018). Predicting Race and Ethnicity From the Sequence of Characters in a Name (arXiv).

Wade, J., and Zaringhalam, M. (2018). Why we're editing women scientists onto Wikipedia. Nature d41586-018-05947-8.

Wais, K. (2016). Gender Prediction Methods Based on First Names with genderizeR. The R Journal *8*, 17.

Witteman, H.O., Hendricks, M., Straus, S., and Tannenbaum, C. (2019). Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. The

Lancet *393*, 531–540.

Ye, J., Han, S., Hu, Y., Coskun, B., Liu, M., Qin, H., and Skiena, S. (2017). Nationality Classification Using Name Embeddings. Association for Computing Machinery (ACM).