Analysis of ISCB honors reveals disparities

Supplementary materials

Figure S1. NamePrism groups countries by name similarity. We used this grouping but renamed the groups to focus on the linguistic patterns based on name etymology identified by NamePrism.

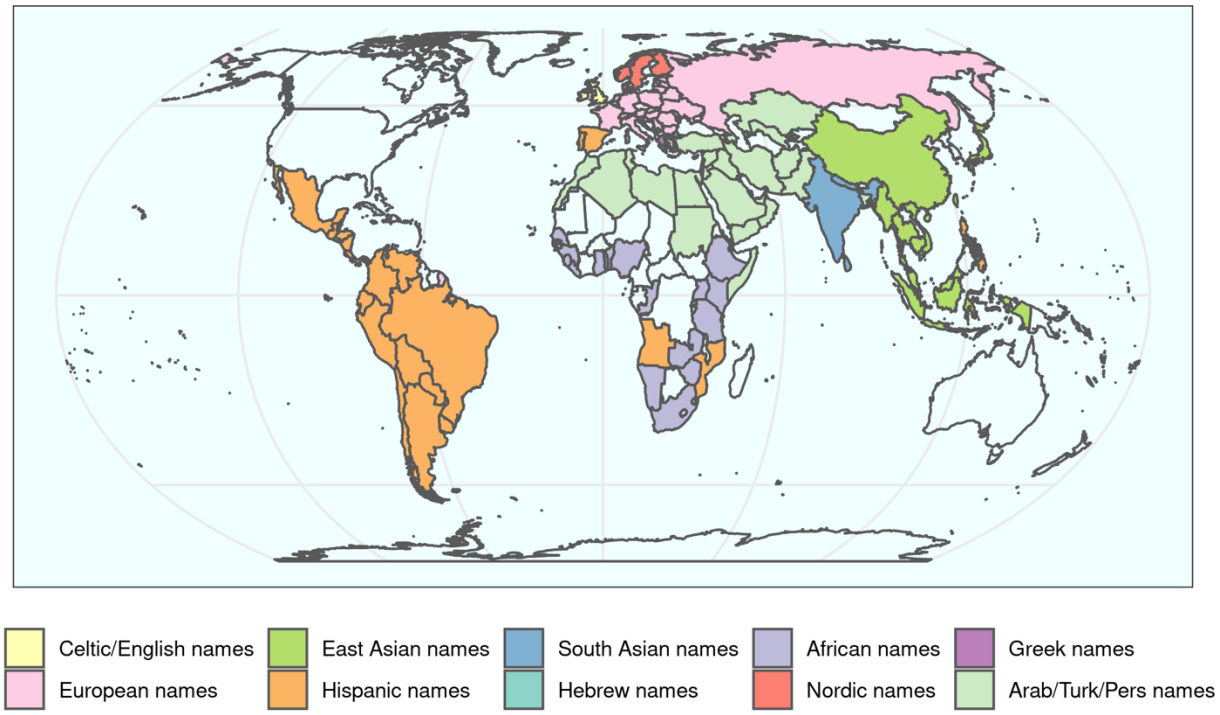


Figure S2. The Wiki2019-LSTM model performs well on the testing dataset. The area under the ROC curve is above 92% for each category, showing strong performance across origin categories (A). A calibration curve, computed with the caret R package, shows consistency between the predicted probabilities (midpoints of each fixed-width bin) and the observed fraction of names in each bin (B). Heatmap showing whether names from a given group (x-axis) received higher (purple) or lower (green) predictions for each group (y-axis) than would be expected by group prevalence alone (C). The values represent $\log_2$ fold change between the average predicted probability and the prevalence of the corresponding predicted group in the testing dataset (null). Scaling by group prevalence accounts for the imbalance of groups in the testing dataset. In all cases, the classifier predicts the true groups above the expected null probability (matrix diagonals are all purple). For off-diagonal cells, darker green indicates a lower mean prediction compared to the null. For example, the classifier does not often mistake East Asian names as Greek, but is more prone to mistaking South Asian names as Celtic/English

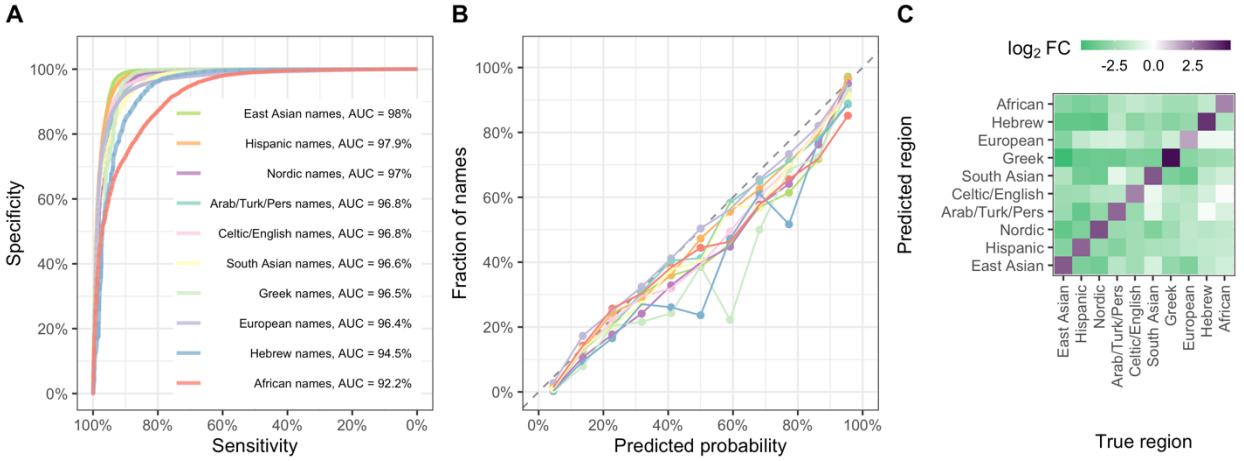
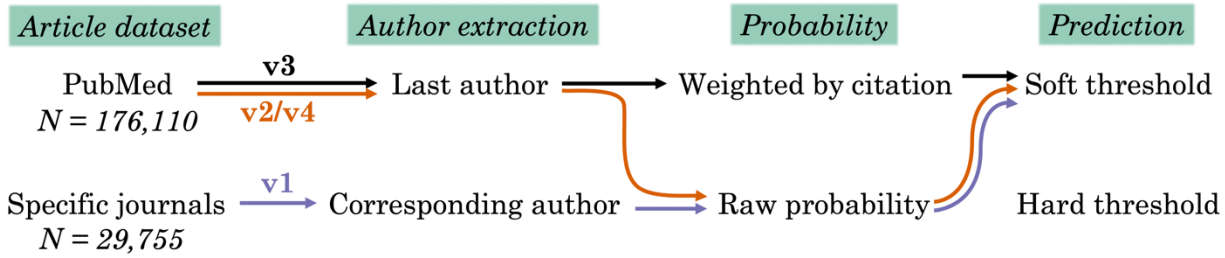


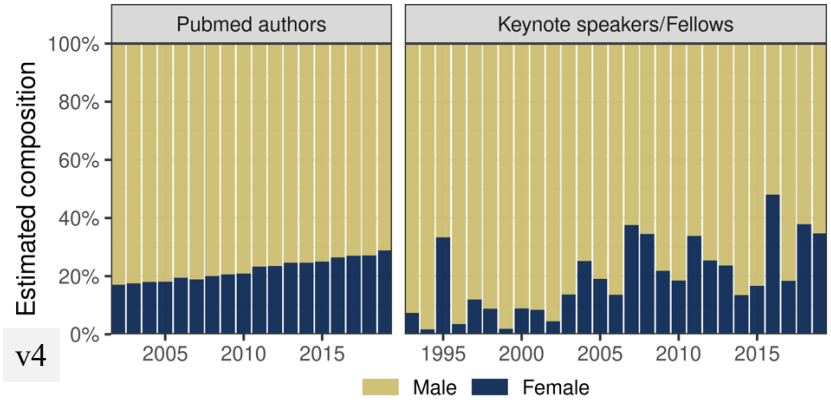
Figure S3. Comparison of results across different versions of the analysis. In v1, we only found webpages with full names for keynote speakers for the years 2002-2019. In v2-3, we found earlier records for ISMB, starting from 1993. In v3, we only considered predictions from 2002 on because, before 2002, most authors did not have gender or name origin predictions because of they only have initials for fore names. In v4, we return to equal weight for all articles. See our analysis notebook for further details.

Manuscript releases: https://github.com/greenelab/iscb-diversity-manuscript/releases

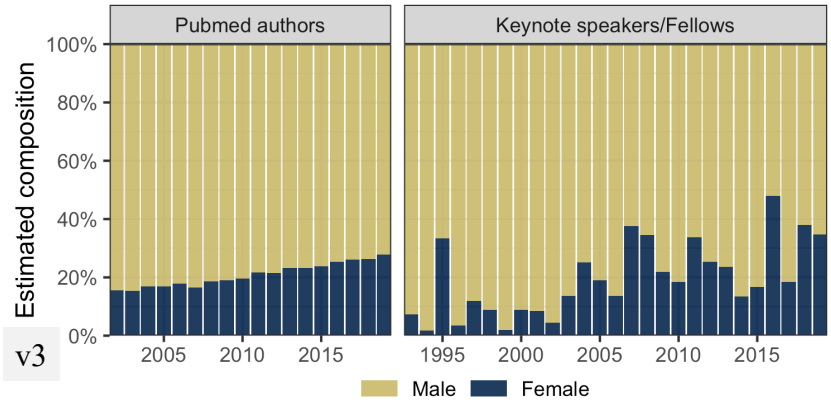
Corresponding analysis repository commit hashes:

	Manuscript repository	Analysis repository
v1	`c6e19bb21a2fa3b0dc21409343a9c3730cf8978b`	`c5a4a416fcb7e1e9bc0683fa0c5ba5ca30a3aa89`
v2	`b3729836b2476a73031a18a9a32d60364f21135e`	`1e53afc1c334af969c51dbdbab8c90b802343ccb`
v3	`45f778da125ac069b0e143f8172a5647cfbc3a39`	`abf6ceb147cc869d259ca3142c92d434ea204ab7`
v4	`f1742982f977553f78dea53f0f3bc5b878f73c52`	`3bd34a600e34c49816420207fc47458b61f3949d`

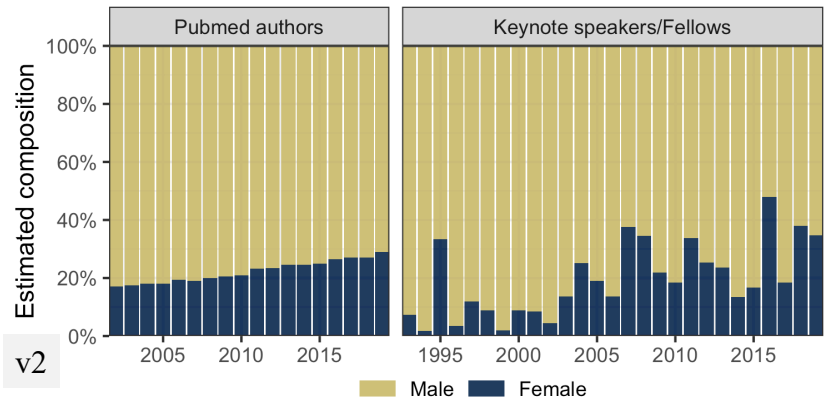



Pubmed authors
Keynote speakers/Fellows
Estimated composition
100%
80%
60%
40%
20%
0%
2005
2010
2015
1995
2000
2005
2010
2015
Male
Female

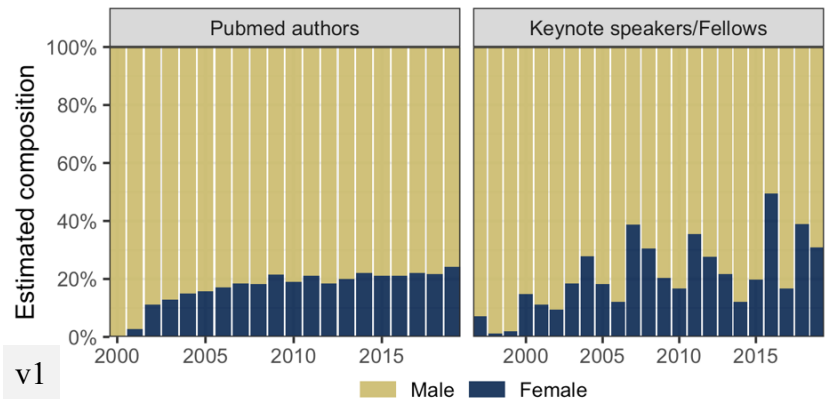
v4


Pubmed authors
Keynote speakers/Fellows
Estimated composition
100%
80%
60%
40%
20%
0%
2005
2010
2015
1995
2000
2005
2010
2015
Male
Female

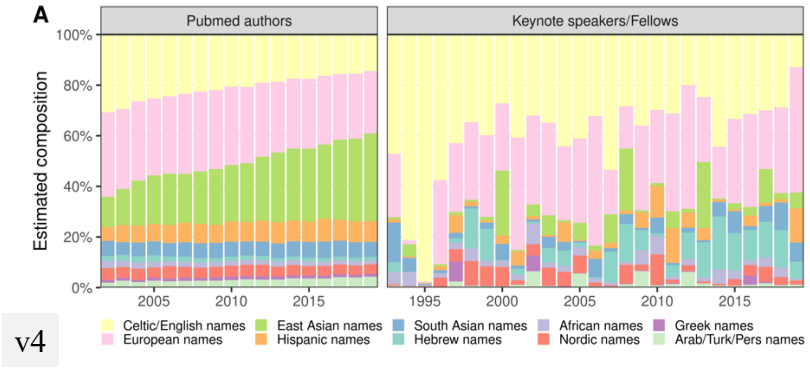
v3


Pubmed authors
Keynote speakers/Fellows
Estimated composition
100%
80%
60%
40%
20%
0%
2005
2010
2015
1995
2000
2005
2010
2015
Male
Female

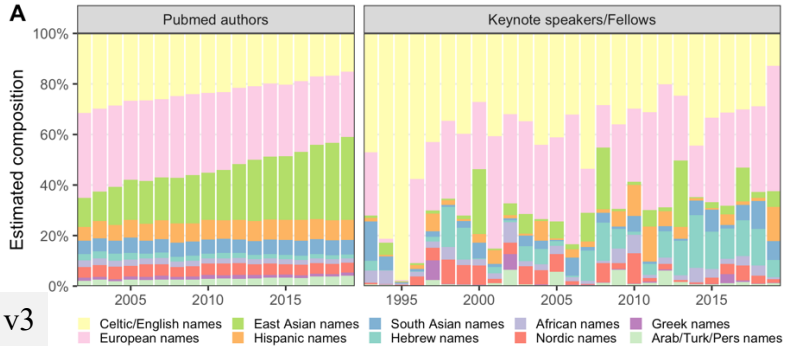
v2


Pubmed authors
Keynote speakers/Fellows
Estimated composition
100%
80%
60%
40%
20%
0%
2000
2005
2010
2015
2000
2005
2010
2015
Male
Female

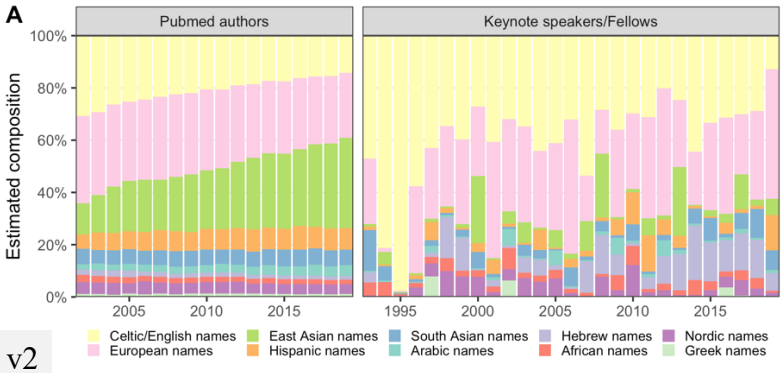
v1


A
Pubmed authors
Keynote speakers/Fellows
Estimated composition
100%
80%
60%
40%
20%
0%
2005
2010
2015
1995
2000
2005
2010
2015
Celtic/English names
European names
East Asian names
Hispanic names
South Asian names
Hebrew names
African names
Nordic names
Greek names
Arab/Turk/Pers names

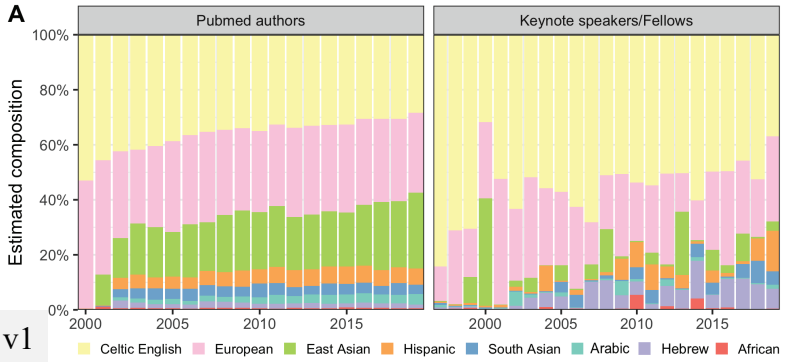
v4


A
Pubmed authors
Keynote speakers/Fellows
Estimated composition
100%
80%
60%
40%
20%
0%
2005
2010
2015
1995
2000
2005
2010
2015
Celtic/English names
European names
East Asian names
Hispanic names
South Asian names
Hebrew names
African names
Nordic names
Greek names
Arab/Turk/Pers names

v3


A
Pubmed authors
Keynote speakers/Fellows
Estimated composition
100%
80%
60%
40%
20%
0%
2005
2010
2015
1995
2000
2005
2010
2015
Celtic/English names
European names
East Asian names
Hispanic names
South Asian names
Arabic names
Hebrew names
African names
Nordic names
Greek names

v2


A
Pubmed authors
Keynote speakers/Fellows
Estimated composition
100%
80%
60%
40%
20%
0%
2000
2005
2010
2015
2000
2005
2010
2015
Celtic English
European
East Asian
Hispanic
South Asian
Arabic
Hebrew
African

v1

Figure S4. Estimated gender proportion of ISCB Fellows and keynote speakers compared to PubMed authors.