A critical analysis of plant science literature reveals ongoing inequities

Authors: Rose A. Marks[1,2,3,*], Erik J. Amézquita[4,*], Sarah Percival[5,*], Alejandra Rougon-Cardoso[6,7], Claudia Chibici-Revneanu[8], Shandry M. Tebele[3], Jill M. Farrant[3] , Daniel H. Chitwood[1,4], Robert VanBuren[1,2]
**Contributed equally*

1. Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA
2. Plant Resilience Institute, Michigan State University, East Lansing, MI 48824, USA
3. Department of Molecular and Cell Biology, University of Cape Town, Rondebosch 7701, South Africa
4. Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824, USA
5. Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA
6. Laboratory of Agrigenomic Sciences, Universidad Nacional Autónoma de México, ENES-León, León 37689, México
7. Plantec National Laboratory, ENES-León, León 37689, México
8. Intercultural Development and Management, Universidad Nacional Autónoma de México, ENES-León, León 37689, México

Keywords: colonialism, patriarchy, gender equality, eurocentrism, taxonomic gaps, diversity, inclusion

Running head: Critical analysis of plant science literature

Correspondence: Rose A. Marks, Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA; Email: marksros@msu.edu; Phone: (603) 852-3190; ORCID iD: https://orcid.org/0000-0001-7102-5959

Competing interests: The authors declare no competing interests.

ABSTRACT

The field of plant science has grown dramatically in the past two decades, but global disparities and systemic inequalities persist. Here, we analyzed ~300,000 papers published over the past two decades to quantify disparities across nations, genders, and taxonomy in the plant science literature. Our analyses reveal striking geographical biases—affluent nations dominate the publishing landscape and vast areas of the globe having virtually no footprint in the literature. Authors in Northern America are cited nearly twice as many times as authors based in Sub-Saharan Africa and Latin America, despite publishing in journals with similar impact factors. Gender imbalances are similarly stark and show remarkably little improvement over time. Some of the most affluent nations have extremely male biased publication records, despite supposed improvements in gender equality. In addition, we find that most studies focus on economically important crop and model species and a wealth of biodiversity is under-represented in the literature. Taken together, our analyses reveal a problematic system of publication, with persistent imbalances that poorly captures the global wealth of scientific knowledge and biological diversity. We conclude by highlighting disparities that can be addressed immediately and offer suggestions for long-term solutions to improve equity in the plant sciences.

SIGNIFICANCE STATEMENT

We analyzed ~300,000 papers published over the past two decades to quantify global, gender, and taxonomic disparities in plant science. Our analyses reveal striking geographical biases that are correlated with national affluence. Gender imbalances were also evident, with far more papers led by authors with masculine names than authors with feminine names. Lastly, we identified substantial taxonomic sampling gaps. The vast majority of surveyed studies focused on major crop and model species and the remaining biodiversity accounted for only a fraction of publications. Taken together, our analyses represent an important addition to the growing conversation about diversifying and decolonizing science.

INTRODUCTION

Plant science research is accelerating at a rapid pace. New technologies and expanding infrastructure have opened the door for cutting-edge research to be conducted at monumental scales. Despite this noteworthy growth, access to resources is not evenly distributed across the globe and recent studies have revealed striking participation gaps and longstanding disparities tied to colonialism, economic inequality, and systemic biases (1-6). Plant science, which, for the context of this study, we define broadly as *any research investigating an organism that performs photosynthesis*, suffers from acute historical exclusion and ongoing underrepresentation of marginalized identities (7). In Northern America, associations between plant science and agriculture with colonialism, slavery, and the exploitation of migrant workers (8) have contributed to a notable lack of diversity in the discipline. Global economic disparities, established under imperial colonialism and perpetuated through modern eurocentric frameworks, further exacerbate underrepresentation of diverse perspectives in plant science (9, 10, 3). Researchers working in low-income countries and under-resourced institutions face multiple barriers to participating in plant science research, including limited funding opportunities, reduced access to cutting-edge technologies and infrastructure, and exclusion from collaboration networks (5,11). In the field of plant genomics for instance, few projects have been led by researchers in the Global South, despite the striking biodiversity and extensive local botanical knowledge within these regions (3). These dynamics are reinforced by a eurocentric framework that centers English language standards, Latin binomial naming conventions, and reductionist thinking. Coupled with historical and ongoing expropriation of plant germplasm from the Global South, this has resulted in a system that unjustly benefits certain individuals and excludes others. A first step of addressing these inequalities is to quantify patterns of participation in plant science.

Both race and gender compound with global economic disparities to generate emergent barriers for people of color and individuals with marginalized gender identities (2, 12). For example, women of color are uniquely oppressed across multiple axes in ways that amount to more than the sum of their racial and gender identities (12). Although our analyses do not address race directly, we explore global patterns with links to imperial colonialism that cannot be understood without an acknowledgement of race and the persistent oppression faced by people of color, especially Black and Indigenous communities. Our analyses address patriarchy, sexism, and gender dynamics more directly. Patriarchy can be described as a way of living that privileges all men over women and some men over other men, and the politics of patriarchy can be understood as “the politics of domination – a politics that rationalizes inequality” (13). Systems of patriarchy vary in their manifestation and severity across the globe, but are pervasive and have infiltrated all levels of society including scientific research (14–17). While self-identified women are not excluded from the field of biology as a whole, they are often excluded from prestigious tenured and editorial positions as well as collaboration networks (7, 18–20). Studies suggest that gender biases also exist in hiring, publication, and funding decisions (6, 19, 21–25). These inequities impact academic currency on job and funding markets and further exacerbate imbalances in academia. Quantifying the extent and patterns of gender bias in plant science is an important step in creating a more equitable discipline.

Despite noteworthy efforts made towards cataloging all life, research attention has not been equally distributed across study systems and many species remain underexplored. In plant genomics, for example, there are substantial taxonomic gaps–multiple clades lack a reference genome assembly while other clades have dozens of sequenced species (3, 26, 27). These findings suggest that research attention has been disproportionately directed towards a few select species with agricultural and economic relevance to modern society. Focusing on these elite crop and model species has enabled noteworthy scientific breakthroughs and agricultural innovations, but it has come at the cost of exploring the rich biodiversity of wild plants and regionally important crops. With species extinction rates at an all-time high (28, 29), much of this biodiversity could be lost before it is understood scientifically. Participation gaps likely contribute to taxonomic sampling gaps in complex and context dependent ways. For example, the exclusion of Indigenous perspectives from science has removed valuable knowledge of local biodiversity and diverted resources away from regionally important plants (30). Together, these factors exacerbate the patriarchal and eurocentric system of publication, and result in a body of literature that poorly represents the global wealth of biological diversity and knowledge.

To better understand the changing global landscape of plant science research and quantify patterns of underrepresentation, we conducted a large-scale bibliometric analysis of nearly 300,000 papers published across the past two decades of plant science research. Our analyses are framed from the perspective of the first axiom of Ardila-Mantilla which states that scientific potential is “distributed equally among different groups, irrespective of geographic, demographic, and economic boundaries” (31). If we take such a statement to be the null hypothesis, then disparities in educational advancement and promotions, funding, or publication and citation rates indicate that other factors, like oppression, have created historical and contemporary biases in science. To test this hypothesis, we identified the demographic features (e.g., nationality and gender) associated with high publication and citation rates and quantified taxonomic sampling gaps and regional differences in focal organism choice to explore associations between participation gaps and study organisms. We examined how these dynamics change over time and space to identify areas that are improving, stagnant, or regressive. We close by discussing the need to dismantle oppressive systems in the plant sciences, improve equity, and how such changes will ultimately advance the field in the coming decades.

RESULTS

We compiled a database of 296,447 plant science papers published between 2000 to 2021. Papers were sourced from a representative set of 127 plant science journals based in 26 different nations across 5 continents, covering 21 different subspecialties. We included both society and for-profit journals in our analyses, with open access, hybrid, and subscription publishing models (see Supplementary Dataset S1 for journal information). The database we assembled does not capture the entire breadth of plant science and related fields as many regional and subject specific journals are not included here. As such, these analyses represent an important but non-exhaustive step towards quantifying inequities in plant science.

Geographic disparities in publication and citation rates

To gain insight into the global landscape of plant science research, we summarized geographic differences in publication and citation numbers. Vast areas of the world have virtually no footprint in the plant science literature over the past two decades (Figure 1A) and publication rate is tightly correlated with national affluence. On a continental level (Supplementary Figure 1A), nearly one third (27%) of all papers were led by authors based in Europe, another 18% were led by authors in Northern America, and 37% by authors in Asia. The remaining 17% of publications were led by authors distributed across Africa, Latin America, and Oceania. Within each continent, authors were further consolidated into distinct hubs of research activity, with the USA, China, and Western Europe dominating the plant science landscape (Figure 1A). National publication rates were highly correlated with Gross Domestic Product (GDP) (R^2=0.75, $F_{1,140}$=213, p=3.18e-43) (Figure 1B) and investment in research and development (R^2=0.83, $F_{2,117}$=295, p=2.08e-46) (Figure 1C and Supplementary Figure 2A). Both relationships follow a power law, as reflected by a linear relationship in logarithmic plots. However, some individual nations performed better (or worse) than expected. Many emerging economies such as India, South Africa, Mexico, Pakistan, Nigeria, Trinidad and Tobago, Iraq, and Madagascar produced far more publications than expected relative to the money invested in research and development. In contrast, some high income nations, particularly in Scandinavia, Northern Europe, and the Middle East produced far fewer publications than expected relative to the money invested in research and development (Figure 1C and Supplementary Figure 2A). There was very little correlation between publication rate and per capita income (R^2=0.23, $F_{2,136}$=20, p=1.68e-08) (Supplementary Figure 2B). Based on the United Nations' income classifications of high, upper-middle, lower-middle, and low income (Supplementary Figure 1B), we found that 61% of all papers published in the last 20 years were led by authors in high income nations. Another 32% were led by authors in upper-middle income nations, and the remaining ~7% of publications were distributed among lower-middle nations. Less than 1% of papers were led by authors in low income nations.

The plant science landscape has changed over the past 20 years. While research output in high income countries has remained relatively stable, there has been a 10-fold increase in the number of papers from upper-middle income nations in the past two decades. In fact, by 2021, there were more papers published by authors in upper-middle income nations than by authors in high income nations (Figure 2A and B). However, this increase was driven primarily by China, which accounted for more than 60% of the publication output from upper-middle income nations in 2020. Other emerging economies such as India, Brazil, Iran, South Africa, Mexico, and Argentina have also made noteworthy contributions to the increased research output of upper-middle income nations (Figure 2C). Publication rates in lower-middle and low income nations have also increased in the past two decades, but still lag far behind those of high and upper-middle income nations (Supplementary Figure 3). In some cases, noticeable decreases in research activity appear to correlate with national disasters and war (e.g. Syria's annual publications declined sharply in the past 10 years (Supplementary Figure 3A)). Despite noteworthy growth in plant science research, many countries remain underrepresented in the literature.

In general, productivity is expected to scale with population size following a power law, such that larger cities produce more research output than smaller ones (32–34), and this is what we observed in the plant science literature (Supplementary Figure 4). However, this scaling was variable across the globe. In general, cities in Northern America, Northern Europe, and Oceania had above average research output relative to population size. In contrast, cities in Asia, Africa, and Latin America had below average research output relative to their population size (Figure 2A). Taken together we find that high income nations produced a higher proportion of their research in rural areas, whereas lower income nations concentrated research activity in high-density, urban areas. This is noteworthy because high income nations in Northern America, Europe and Oceania account for less than 10% of the rural population globally (Supplementary Figure 5) but produce more than 64% of the plant science research.

International and intercontinental collaborations were strikingly uncommon in the past two decades of plant science research (Figure 4 and Supplementary Figure 6). More than two thirds (71%) of the publications in our database were written by authors based in a single nation. Just 22% of studies involved a collaboration between two nations, and only 5% of studies included three nations. Less than 1% of studies involved four nations even though 71% of papers have four or more authors, and just 0.04% included five nations despite the fact that 54% of papers had five or more authors. When international collaborations did occur, they tended to be across continents rather than within continents. Only Europe-based authors showed a high frequency of within-continent collaboration (Supplementary Figure 6). Collaborations across continents did occur but were not evenly distributed. Most nations preferred to collaborate with researchers in Europe, Northern America, or China (Figure 3), and were less likely to collaborate with authors in Latin America, Africa, or West Asia. A similar pattern is evident when considering income groupings–only the most affluent nations participated in within-group collaborations, and all other nations preferred to collaborate with high income nations (Supplementary Figure 6).

Despite striking differences in research output, the mean impact factor of the journals that papers were published in spanned just over one point across continents–ranging from 2.92 ± 0.017 for papers led by authors in Sub-Saharan Africa to 4.06 ± 0.011 in Northern America (Table 1). In contrast, citation rates were substantially more variable across continents. In general, papers from the Global South received dramatically fewer citations than those from the Global North, despite publishing in journals with similar impact factors. For example, mean cumulative citations ranged from 17.82 ± 0.304 for papers led by authors working in Sub-Saharan Africa to 36.75 ± 0.298 in Northern America (Table 1)--a twofold difference. This dynamic has remained relatively stable over the past 20 years, with persistent differences in annual citation rates between continents (Supplementary Figure 7). Some individual nations (e.g., China) have seen improvements in citation rates over time, but most have not.

Table 1. Continental averages and standard error for the impact factor of journals that authors published in and mean number of citations that papers received.

Continent	*Mean impact factor*	*Mean cumulative citations*
Northern America	*4.06 ± 0.011*	*36.75 ± 0.298*
Oceania	*3.67 ± 0.022*	*31.99 ± 0.621*
Europe	*3.94 ± 0.008*	*31.21 ± 0.193*
Asia (minus China and West Asia)	*3.53 ± 0.008*	*26.32 ± 0.210*
North Africa and West Asia	*3.05 ± 0.017*	*23.00 ± 0.409*
China	*4.13 ± 0.009*	*21.69 ± 0.159*
Latin America and the Caribbean	*3.13 ± 0.011*	*18.57± 0.240*
Sub-Saharan Africa	*2.92 ± 0.017*	*17.82 ± 0.304*

We also investigated how journal policies such as open access fee and society membership related to participation rates for authors with different identities. Of the 296,447 papers examined, only 14% were published Gold open access. Authors in Northern America and Asia published the highest proportion of open access papers (23% and 18% respectively). In contrast, only 10-15% of papers led by authors based in Africa, Latin America, and West Asia were published open access. Of the 16,641 papers published in “elite” journals (with impact factors above seven), 68% were led by authors in high income nations, compared to 61%

overall, and another 15% were led by authors based in China. The remaining 17% were distributed across authors in lower income nations. Citation rates were extremely skewed within these journals. For example, papers led by authors in high income nations were cited 82 ± 0.23 times whereas papers from low income nations were cited only 24 ± 0.86 times–a fourfold difference. In general, society journals did not exhibit any more geographic equality in publication and citation rates than the overall trend. Of the 158,711 papers published in society journals 63% were led by authors in high income nations and these received almost double the number of citations (38.9 ± 0.217) compared to papers led by authors in low income nations (20.3 ± 1.767).

Persistent gender inequalities in the plant sciences

We quantified the effects of patriarchy and gender discrimination in plant science publishing by associating author names with masculinity or femininity. We acknowledge that a binary gender division is an oppressive concept in itself and that true gender is self-identified (35), and we recognize the proximity of our approach to the harmful practice of gender inference. However, we find that we cannot discuss patriarchy and gender discrimination without employing the concept of gender. We purposefully seek to avoid inferring the gender identity of individuals and instead measure the oppressive effects of patriarchy associated with the names themselves. We focus on the normative association of names with masculinity or femininity to measure these effects, and do not presume to know the true gender identity of authors. We further acknowledge that biases in name-based gender inference can arise from the global diversity of cultural naming systems (36). The accuracy of name-based gender prediction varies considerably across ethnicities and is notably poor for East Asian names (37, 38). This is indicative of yet another layer of bias that has resulted in a eurocentric set of tools and analytical frameworks. Improved algorithms that can handle a diversity of naming conventions and accommodate non-binary gender classifications are needed (39). We have tried to adhere to the five principles for ethical gender inference articulated by (38) still, we struggled with the ethics of algorithmic gender inference within our working group and must acknowledge that in conducting such analyses, we too are culpable in propagating the gender binary anew. Given the caveats and obvious shortcomings of name-based gender inference, we urge all readers to interpret these findings with caution.

We hypothesized that individuals with marginalized gender identities (including women, non-binary, gender non-conforming, trans, and people of multiple sexes/genders) would face barriers to participation in plant science and that these would compound with socioeconomic disadvantages and/or historical oppression to further limit participation by intersectional individuals. We cannot test this hypothesis directly without knowing the gender identities of the authors in the paper, so we aimed to instead measure perceptive discrimination based on sexism, that disadvantages individuals with names normatively associated with femininity (40, 41). To test this prediction, names of corresponding authors for each paper were isolated and classified as either 1) names normatively associated with masculinity (NNMs) or 2) names normatively associated with femininity (NNFs) and used as a proxy for gender.

Globally, there were far more papers led by authors with NNMs than authors with NNFs (Figure 4A and B). However, the degree of gender imbalance varied considerably across continents and nations. Among the 20 nations with the highest publication rates, the most NNM biased nations were Japan (14% NNF), India (21% NNF), Netherlands (23% NNF), Switzerland (24% NNF), and Israel (25% NNF). In contrast, the least NNM biased nations were Poland (61% NNF), Argentina (57% NNF), Italy (41% NNF), Brazil (41% NNF), and Spain (38% NNF). On a continental level, Latin America and Europe had the highest proportions of papers led by authors with NNFs whereas Northern America, Asia, and Oceania had the lowest proportion of NNFs. There has been a modest increase in participation by individuals with NNFs over time, but gender ratios remain far from equal across much of the globe (Figure 5).

There was no correlation between national GDP and the proportion of papers led by NNFs (R^2=0.013, $F_{2,116}$=0.78, p=0.46) (Supplementary Figure 7A). In fact, some of the highest GDP nations had the lowest proportion of NNF authors. There was a similar lack of relationship between per capita income and the proportion of NNF authors (R^2=0.099, $F_{2,113}$=6.22, p=0.0027) (Supplementary Figure 8B).

There was no significant difference in the impact factor of the journals that authors with NNFs versus NNMs published in. However, there were noteworthy differences in the number of citations these papers received. Papers led by authors with NNMs were cited on average 5 more times than those led by authors with NNFs. This pattern has not improved over time and, if anything, the difference in annual citations for authors with NNFs versus NNMs has expanded (Supplementary Figure 8).

Taxonomic gaps in focal species studied in the plant sciences

Funding priorities and research activities have historically focused on a narrow subset of described plant species (3, 42, 43) and we expected to find notable taxonomic sampling gaps in the current dataset. To test this prediction, we identified all taxonomic entities mentioned in abstracts via natural language processing. We then summarized overall patterns and geographic differences in the choice of focal species to identify taxonomic sampling gaps and regional patterns.

There were 73,527 unique taxonomic entities represented in our publication database. While the majority of studies focused on plants, we also identified numerous non-plant species including pathogens, symbionts, and other interactors across animalia, fungi, and bacterial groups (Figure 6B). All the top 20 most studied plants represent economically important crop species or models developed by the plant research community (Figure 6A). The model plant *Arabidopsis thaliana* was by far the most studied plant in the past two decades, appearing in four times as many studies as the next most common species wheat (*Triticum aestivum*) (Figure 6A). Poales was the most studied order with over 50,000 mentions, followed by Brasicales, Fabales, and Solanales (Supplementary Figure 10). Many orders were statistically over- or under-represented in the dataset relative to their species richness. The most over-represented orders were Brasicales, Poales, Solanales, Fabales, and Cucurbitales. In contrast, the most under-represented clades were Asterales, Asparagales, Gentianales, Polypodiales, and Lamiales (Figure 6C).

We also identified regional differences in the choice of focal organisms. Most high income nations with high publication rates tended to focus on *A. thaliana*, grain crops, vegetables, fruits, and model species (Figure 7). In contrast, many of the nations underrepresented in publishing, tended to focus on lesser-known species and minor or regionally important crops. This finding exemplifies how underrepresentation at the human level impacts the diversity and breadth of focal organisms and research directions.

DISCUSSION

Our analyses reveal striking geographical biases in plant science research that are associated with national affluence. Global patterns of wealth distribution cannot be understood without an acknowledgement of the impact of imperial colonialism and the resulting consolidation of resources within select nations of the Global North (3, 9, 44). Not only was wealth redistributed during this process, but diverse perspectives and peoples were effectively erased from science as a eurocentric worldview was exported across the globe. Christian missionaries, European traders, and inquisitive researchers all helped to spread frameworks of capitalism, patriarchy, and white supremacy. In biology, these value systems are coupled with a precedence for the English language, reductionist thinking, Latin naming conventions, and biased standards of academic excellence that further exclude individuals from non-European backgrounds. We identified strong correlations between publication rates, GDP, and research and development expenditure. In general, high income nations spent a higher proportion of their GDP on research and development and this led to higher publication output. This finding highlights the privilege of being able to invest in research and development at all, and most lower income nations do not have the necessary funds to support a robust research sector. However, many lower income nations published far more papers than expected relative to their research and development expenditure, while many higher income nations published less than expected. Admittedly, some research output is not captured here because it does not flow into traditional plant science publishing channels and may, instead, be represented by growth in other academic, private, or governmental sectors. Still, this finding gives us reason to pause and recognize the noteworthy accomplishments of scientists from less affluent nations who are doing more with less–an impressive testament to the resourcefulness, creativity, and ingenuity in these regions.

Research output is also associated with increased population density. However, we detected regional differences in this pattern. High income nations (especially in Northern America and Oceania) generated a substantial proportion of their research in rural areas, which makes intuitive sense since plant science is inherently linked to agriculture and natural spaces, and numerous research centers and Land-grant Universities have been built in rural regions. However, lower income nations did not produce many papers in rural areas and instead concentrated research activity in urban centers. We suspect that this pattern is driven by the fact that rural areas are often the last places to be developed and still lack basic infrastructure across much of the globe (45). The differences in rural development impact where research is conducted and contribute to the exclusion of rural peoples and agricultural communities in less affluent nations from the scientific discussion. Only 8% of the world’s rural population lives in Europe, Northern America, and Oceania, but these areas produce more than 64% of the plant science papers. The remaining nations have a disproportionately small publication footprint and the knowledge of ecology, ethnobotany, and agriculture from local and indigenous communities within these areas is largely absent from the literature. These voices and perspectives are often co-opted by researchers from affluent nations through parachute science and other colonialist practices, with no acknowledgment, consultation, or compensation for the discoveries (46). Such gaps in participation undoubtedly translate into gaps in understanding and represent a lost opportunity. These harmful practices have perpetuated persistent inequity in the field.

International and intercontinental collaborations were notably uncommon in the past two decades of plant science research. Of the few international collaborations that we identified, the majority involved a collaborator from Europe, Northern America, and to a lesser degree, China. We suspect that differences in resources (both financial and infrastructural) contribute to these dynamics. Researchers working in high income nations have access to more funding for research, engaging collaborators, and traveling to conferences. Researchers in less affluent nations do not have the same funding opportunities and are therefore limited in the number and type of collaborations they can participate in, as well as the research activities they can undertake. There may also be more subtle and problematic factors driving the skewed collaborative networks we observed. Differences in institutional prestige, born out of eurocentric mindsets, have led some to believe that the best science is done in select institutions in the Global North and that working at or collaborating with those institutions is most desirable. We believe that this rationale is fundamentally flawed and should be dismantled. Affluent nations could do more to engage collaborators in less represented regions of the globe instead of following the well-established global network. Not only would this help to equalize the plant science landscape, but it would enrich our science by bringing in the wisdom of different perspectives.

We identified striking and persistent gender biases in plant science publishing. Given the caveats and shortcomings of name-based gender inference, making specific claims about the gender of individuals or small groups should be avoided, but the overarching patterns identified here are representative. Over 70% of publications in the past two decades were led by authors with masculine names. The extent of gender imbalance was variable across nations and continents but showed remarkably little change over time. In most regions, we detected only modest increases in the number of papers led by authors with feminine names over the past two decades. Interestingly, some of the most affluent nations (e.g., USA, Japan, Netherlands, Switzerland, Germany, Canada, and New Zealand) had extremely male biased publication records despite supposed improvements in women's rights in many of these nations. In contrast, some less affluent nations in the Global South (e.g., Argentina, Brazil, and Mexico) had among the highest proportions of authors with NNFs. This finding is similar to the “gender-equity paradox” detected in mathematics (47), and contradicts our prediction that individuals facing the intersecting barriers of economic constraints and marginalized gender identity would be more excluded from academic publishing. It suggests that other factors, like cultural differences, could be playing a role in gender inequity. For example, in regions where farming and agriculture are traditionally women's work, more women may choose to enter the plant sciences. In addition, differences in available support systems can drive career choice, with women sometimes pursuing higher paying jobs (often in STEM fields) when social support systems are limited (48). We looked at a variety of economic development indicators to try to understand what could be driving gender biases in plant science publishing. In contrast to geographical patterns, there was no association between national GDP, research and development expenditure, or per capita income with gender ratio. These findings suggest that the footprint of patriarchy in plant science is deeper than we acknowledge and does not align neatly with narratives about cultural differences in sexism. We also identified gender biases in citation rates that were independent of time, suggesting persistent and ongoing gender discrimination. Because individuals, not institutions drive citation rates, this suggests a deep and pervasive bias running through the discipline. It also means that we, as individuals, have the power to shift these patterns through our actions and choices.

In the past two decades, plant scientists have studied thousands of species spanning plants, animals, bacteria, and fungi. Despite the noteworthy diversity and volume of research, sampling effort has not been equally distributed across clades and taxa. The vast majority of studies have investigated major crop and model species, and the remaining biodiversity accounts for only a fraction of the research on plants. Our analyses identified a number of statistically overrepresented groups of plants, all of which included agriculturally and economically important plants. We also identified numerous underrepresented taxonomic groups, which were ecologically diverse, speciose, and generally of less economic relevance to modern society. These underexplored lineages could provide untold value to humans and ecosystems but have been largely overlooked by modern plant scientists (3, 30, 49). We found some evidence to indicate that taxonomic gaps are related to geographic and gender gaps and we suspect that limited diversity of authors is exacerbating biases in study organism choice. In general, affluent nations in Europe, Northern America, and Asia tended to focus on major crops associated with industrialized agriculture (e.g., wheat, rice, soybean, tobacco, tomato, etc.). In comparison, many of the nations with a smaller footprint in plant science focused their research on regionally important and underutilized crops such as cassava, yam, and millets or local plants with medicinal or historical importance. The disproportionate focus on major crops in the mainstream literature reinforces a homogenization of plant science and limits our ability to conserve and utilize biodiverse plants. It is possible that work on biodiverse species is disproportionately published in regional and subject-specific journals that are not included here, and future studies investigating parallel patterns in these sectors of plant science publishing would be worthwhile extensions of this work. We suspect that if more researchers from across the world were actively engaged in plant science research, there would be a natural diversification of study systems and a broadening of cumulative knowledge.

Conclusions

Our analyses provide evidence of deep disparities in plant science with links to colonialism, eurocentrism, and patriarchy. Despite the proliferation of statements, committees, workshops, and trainings aimed at increasing diversity, equity, and inclusion, little progress has been made towards actually diversifying plant science in the past two decades (51). These findings can be used as evidence in advocating for change at institutional and policy levels, while also motivating individuals to make positive change in their own research activities and philosophy. While many recognize that the current system is unfair, there are contrasting views on what changes should be made. Some advocate for reformation while others favor abolition, but both agree that there is a need to broaden science and embrace the diversity of knowledge acquisition systems that exist globally. We suggest that first steps towards improving the discipline should consist of a fundamental broadening of our definition of what science is and who can do it. By embracing a more nuanced and context dependent view of data, acknowledging that novelty is not the only source of scientific merit, and recognizing the value of qualitative research, we can begin to minimize colonial biases in academic culture, language, and institutions (30). Funding is another important component, and wealthy nations should take the lead in making efforts to equalize disparities in national affluence established through colonialism. Grants that specifically promote intercontinental collaborations coupled with direct funding to lower income nations could play an important role. Formal policies that provide guidelines and regulations for data ownership and benefit sharing can also help to ensure equitable research practices. The Nagoya protocol represents one such effort, but many nations lack the necessary infrastructure and institutional support to implement the policy effectively. Given the longstanding disparities that exist in plant science, it may be useful to employ concepts of restorative justice, truth and reconciliations practices (50, 51), and a more general shift away from gatekeeping policies and towards inclusive groundskeeping concepts (52). By expanding our definition of what constitutes scientific inquiry and who can take part in it, we begin to open the door to new sources of knowledge. After centuries of centering patriarchal ideals and eurocentric ways of knowing, it is time to make space for other systems of knowledge to rise to the forefront. We hope our analyses can be used to support these positive changes.

METHODS

Data acquisition and filtering

We assembled a large-scale database of plant science papers from 127 journals spanning a range of impact factors, nationalities, and sub-specialties (see Supplementary Dataset S1). We cross referenced plant science journals listed in the Journal Citation Reports Database (https://jcr.clarivate.com) with a list of plant science journals compiled by the American Society of Plant Biology (https://plantae.org/plant-biology-journal-database/). We then filtered journals on the following criteria: (1) the journal must have an impact factor, (2) it must be plant specific, and (3) it must include research articles. Metadata associated with all research papers from the resulting 127 journals across the last 20 years were included in the current study. Other metadata were incorporated by referencing JCR and journal webpages, the World Bank 2019 database, the UN Statistics Division, and the UN Department of Economic and Social Affairs (see Supplementary Appendix 1 for methodological details).

Geography based analyses

The location of authors was inferred from the addresses listed in the papers using an ad-hoc text processing script. Geographic coordinates (geocoordinates) for all these locations were obtained using the Google Maps Geocoding API (https://developers.google.com/maps/documentation/geocoding) with Python via GeoPy. We computed national summary stats, global patterns of author location, and associations with national development indicators using Python (v3.8.8) packages Pandas (v1.5.0) and Numpy (v1.22.4) and visualized data in Seaborn (v0.11.1) and Matplotlib (v3.6.1).

We quantified patterns of collaboration by identifying the location of the corresponding author relative to all other authors for each paper. We then determined if authors were from different countries, continents, or income brackets and summarized global patterns (see Supplementary Appendix 1 for methodological details).

Gender analyses

We quantified the effects of patriarchy and gender discrimination in plant science by associating author names with masculinity or femininity. The analyses presented here do not identify the true gender of authors. Rather, they show the assumed gender based on the association of first name with either masculinity or femininity. These analyses also likely mis-identify and fail to account for non-binary, gender neutral, and trans individuals, among others. Geographic biases in the performance of gender inference algorithms have also been documented, with most tools performing poorly on East Asian names. This is noteworthy since many of the papers in our dataset are led by individuals with East Asian heritage. Given these caveats, we selected the most robust tool available for this type of analysis (GenderAPI) based on the extensive benchmarking and comparative analyses presented in (36, 37). Summary stats, regional patterns, and changes over time in gender ratios were computed using Python (v3.8.8) packages Pandas (v1.5.0) and Numpy (v1.22.4) and visualized in Seaborn (v0.11.1) and Matplotlib (v3.6.1) (see Supplementary Appendix 1 for methodological details).

Study Species analyses

The species studied in each paper were identified from abstracts using the Python package TaxoNERD (53). Each biological entity was assigned to a NCBI taxonomy ID and higher-level taxonomic classifications were extracted with ETE Toolkit (54). We summarized the number of mentions for each species, genus, family, and order of land plants to identify sampling gaps in focal organisms and test for statistically over- and under-representation of focal organisms relative to the species richness of the order (see Supplementary Appendix 1 for methodological details).

DATA AVAILABILITY

Data associated with this study and a description of data acquisition and curation are deposited in Dryad at https://doi.org/10.5061/dryad.pg4f4qrtb.

AUTHOR CONTRIBUTIONS

RAM, EJA, SP, ARC, CCR, SMT, JMF, DHC and RV conceived of the study. RAM, EJA, SP, ARC, SMT, contributed to data acquisition and curation. RAM, EJA, and SP conducted data analyses. RAM, EJA, SP, ARC, CCR, SMT, JMF, DHC and RV contributed to data interpretation and conceptual framing of the manuscript. RAM, EJA, and SP drew the figures. RAM wrote the manuscript. All authors edited and reviewed the manuscript.

ACKNOWLEDGEMENTS

This work was supported by NSF Postdoctoral Research Fellowship in Biology (grant no. PRFB-1906094) to RAM and funding from the Michigan State University Department of

Biochemistry and Molecular Biology to SP. Other members of our working group were supported by the USDA National Institute of Food and Agriculture, Michigan State University AgBioResearch, the National Research Foundation of South Africa, Universidad Nacional Autónoma de México, and the Water and Life Interface Institute (NSF DBI grant # 2213983).

REFERENCES

1. V. Amarante, *et al.*, Underrepresentation of developing country researchers in development research. *Appl. Econ. Lett.*, 1–6 (2021).

2. B. Maas, *et al.*, Women and Global South strikingly underrepresented among top□publishing ecologists. *Conserv. Lett.* **14** (2021).

3. R. A. Marks, S. Hotaling, P. B. Frandsen, R. VanBuren, Representation and participation across 20 years of plant genome sequencing. *Nat Plants* (2021) https:/doi.org/10.1038/s41477-021-01031-8.

4. C. H. Trisos, J. Auerbach, M. Katti, Decoloniality and anti-oppressive practices for a more ethical ecology. *Nature Ecology & Evolution*, 1–8 (2021).

5. T. E. Ebenezer, *et al.*, Africa: sequence 100,000 species to safeguard biodiversity. *Nature* **603**, 388–392 (2022).

6. K. H. Wapman, S. Zhang, A. Clauset, D. B. Larremore, Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature* (2022) https:/doi.org/10.1038/s41586-022-05222-x.

7. T. F. Madzima, G. C. MacIntosh, Equity, diversity, and inclusion efforts in professional societies: intention versus reaction. *Plant Cell* **33**, 3189–3193 (2021).

8. A. L. Carter, A. Alexander, Soul food: [re]framing the African-American farming crisis using the culture-centered approach. *Front. Commun.* **5** (2020).

9. Z. Baber, The Plants of Empire: Botanic Gardens, Colonial Power and Botanical Knowledge. *J. Contemp. Asia* **46**, 659–679 (2016).

10. F.-W. Li, Decolonizing botanical genomics. *Nat Plants* **7**, 1542–1543 (2021).

11. C. N. Ibe, Democratizing plant genomics to accelerate global food production. *Nat. Genet.* **54**, 911–913 (2022).

12. K. Crenshaw, Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *Univ. Chic. Leg. Forum* **1989**, 8 (1989).

13. Gilligan, Carol, Snider, Naomi, *Why does patriarchy persist?* (2018).

14. E. Haghighat-Sordellini, "Patriarchy, Modernization, and the Global Economy" in *Women in the Middle East and North Africa: Change and Continuity*, E. Haghighat-Sordellini, Ed. (Palgrave Macmillan US, 2010), pp. 33–49.

15. T. Golash-Boza, M. D. Duenas, C. Xiong, White Supremacy, Patriarchy, and Global Capitalism in Migration Studies. *Am. Behav. Sci.* **63**, 1741–1759 (2019).

16. E. Uchendu, N. Edeagu, *Negotiating Patriarchy and Gender in Africa: Discourses, Practices, and Policies* (Rowman & Littlefield, 2021).

17. E. Kocabıçak, *The Political Economy of Patriarchy in the Global South* (Taylor & Francis, 2022).

18. J. D. West, J. Jacquet, M. M. King, S. J. Correll, C. T. Bergstrom, The role of gender in scholarly authorship. *PLoS One* **8**, e66212 (2013).

19. D. N. Frances, C. R. Fitzpatrick, J. Koprivnikar, S. J. McCauley, Effects of inferred gender on patterns of co□authorship in ecology and evolutionary biology publications. *Bull. Ecol. Soc. Am.* **101** (2020).

20. K. Lerman, Y. Yu, F. Morstatter, J. Pujara, Gendered citation patterns among the scientific elite. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2206070119 (2022).

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

22. C. W. Fox, C. S. Burns, A. D. Muncy, J. A. Meyer, Gender differences in patterns of authorship do not affect peer review outcomes at an ecology journal. *Funct. Ecol.* **30**, 126–139 (2016).

23. K. S. Bonham, M. I. Stefan, Women are underrepresented in computational biology: An analysis of the scholarly literature in biology, computer science and computational biology. *PLoS Comput. Biol.* **13**, e1005134 (2017).

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

25. H. O. Witteman, M. Hendricks, S. Straus, C. Tannenbaum, Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *Lancet* **393**, 531–540 (2019).

26. G. C. Vallée, D. S. Muñoz, D. Sankoff, Economic importance, taxonomic representation and scientific priority as drivers of genome sequencing projects. *BMC Genomics* **17**, 782 (2016).

27. P. Szövényi, A. Gunadi, F.-W. Li, Charting the genomic landscape of seed-free plants. *Nature Plants*, 1–12 (2021).

28. R. Dirzo, G. Ceballos, P. R. Ehrlich, Circling the drain: the extinction crisis and the future of humanity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20210378 (2022).

29. M. J. Munstermann, *et al.*, A global ecological signal of extinction risk in terrestrial vertebrates. *Conserv. Biol.* **36**, e13852 (2022).

30. W. Dwyer, C. N. Ibe, S. Y. Rhee, Renaming Indigenous crops and addressing colonial bias in scientific language. *Trends Plant Sci.* (2022) https:/doi.org/10.1016/j.tplants.2022.08.022.

31. F. Ardila-Mantilla, Todos Cuentan: Cultivating Diversity in Combinatorics. *Not. Am. Math. Soc.* **63**, 1164–1170 (2016).

32. L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G. B. West, Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7301–7306 (2007).

33. L. M. A. Bettencourt, J. Lobo, D. Strumsky, Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Res. Policy* **36**, 107–120 (2007).

34. L. M. A. Bettencourt, J. Lobo, D. Strumsky, G. B. West, Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. *PLoS One* **5**, e13541 (2010).

35. M. E. Castro-Peraza, *et al.*, Gender Identity: The Human Right of Depathologization. *Int. J. Environ. Res. Public Health* **16** (2019).

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

37. P. Sebo, Performance of gender detection tools: a comparative study of name-to-gender inference services. *J. Med. Libr. Assoc.* **109**, 414–421 (2021).

38. J. W. Lockhart, M. M. King, C. Munsch, What's in a Name? Name-Based Demographic Inference and the Unequal Distribution of Misrecognition. *https://osf.io/preprints/socarxiv/vedq6/* (2022) https:/doi.org/10.31235/osf.io/vedq6.

39. H. Mihaljević, M. Tullney, L. Santamaría, C. Steinfeldt, Reflections on Gender Analyses of Bibliographic Corpora. *Front Big Data* **2**, 29 (2019).

40. C. M. Topaz, *et al.*, Race- and gender-based under-representation of creative contributors: art, fashion, film, and music. *Humanit Soc Sci Commun* **9**, 221 (2022).

41. Maril, R. K. & Gill, A.M, "Discrimination Based on Perceived Characteristics" (Human Rights Campaign Foundation, 2018).

42. M. J. Trimble, R. J. van Aarde, Geographical and taxonomic biases in research on biodiversity in human-modified landscapes. *Ecosphere* **3**, art119 (2012).

43. M. Adamo, *et al.*, Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers. *Nat Plants* (2021) https:/doi.org/10.1038/s41477-021-00912-2.

44. M. Adas, "Colonialism and Science" in *Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures*, H. Selin, Ed. (Springer Netherlands, 2008), pp. 604–609.

45. D. Lagakos, Does Internal Migration Offer Opportunities? *J. Econ. Perspect.* **34**, 174–192 (2020).

46. P. V. Stefanoudis, *et al.*, Turning the tide of parachute science. *Curr. Biol.* **31**, R184–R185 (2021).

47. T. Breda, E. Jouini, C. Napp, G. Thebault, Gender stereotypes can explain the gender-equality paradox. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 31063–31069 (2020).

48. G. Stoet, D. C. Geary, The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. *Psychol. Sci.* **29**, 581–593 (2018).

49. P. S. Hendre, *et al.*, African Orphan Crops Consortium (AOCC): status of developing genomic resources for African orphan crops. *Planta* **250**, 989–1003 (2019).

50. C. Wong, K. Ballegooyen, L. Ignace, M. J. (gùdia) Johnson, H. Swanson, Towards reconciliation: 10 Calls to Action to natural scientists working in Canada. *Facets (Ott)* **5**, 769–783 (2020).

51. B. L. Montgomery, J. A. Whittaker, The roots of change: Cultivating equity and change across generations from healthy roots. *Plant Cell* **34**, 2588–2593 (2022).

52. B. L. Montgomery, Academic leadership: gatekeeping or groundskeeping? J Values-Based Leadersh 13: 135--151 (2020).

53. N. Le Guillarme, W. Thuiller, TaxoNERD: Deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature. Methods Ecol. Evol. 13, 625–641 (2022).

54. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. 33, 1635–1638 (2016)

FIGURE LEGENDS

Figure 1. Global patterns of plant science publishing. A) The global distribution of where authors are based, scaled by the number of publications from each location. B) The number of studies published by each nation relative to national Gross Domestic Product (GDP). C) The number of studies published by each nation relative to their research and development expenditure.

Figure 2. Publication output relative to national affluence and population size. A) The number of studies published each year by authors in high income nations. B) The number of studies published each year by authors in upper-middle income nations. C) Map of publication output relative to population size for locations with more than 300,000 inhabitants or with more than 100 papers produced during the period of 2000-2021. Locations are scaled and colored according to their research output relative to the global trend. Large green points correspond to locations that produce more research than expected based on the global population trend, while large pink circles represent regions that publish less than expected for a city of their size.

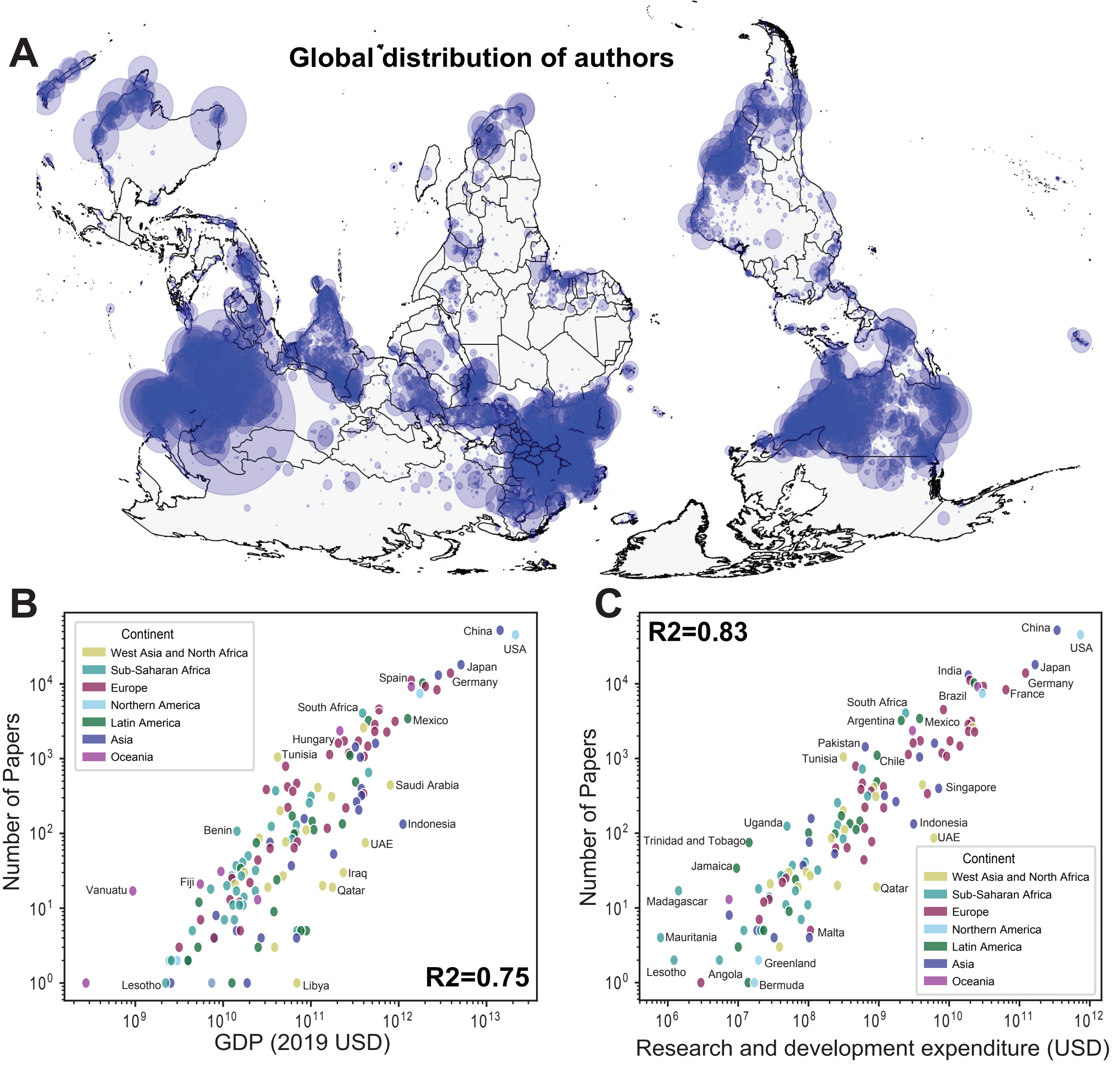
Figure 3. Disparities in global collaborations within plant science research. Circles represent publications that did not involve an intercontinental collaboration. Arrows represent cross-continental collaborations and are directed from corresponding author to co-author. Circles and arrows are scaled by the number of publications.

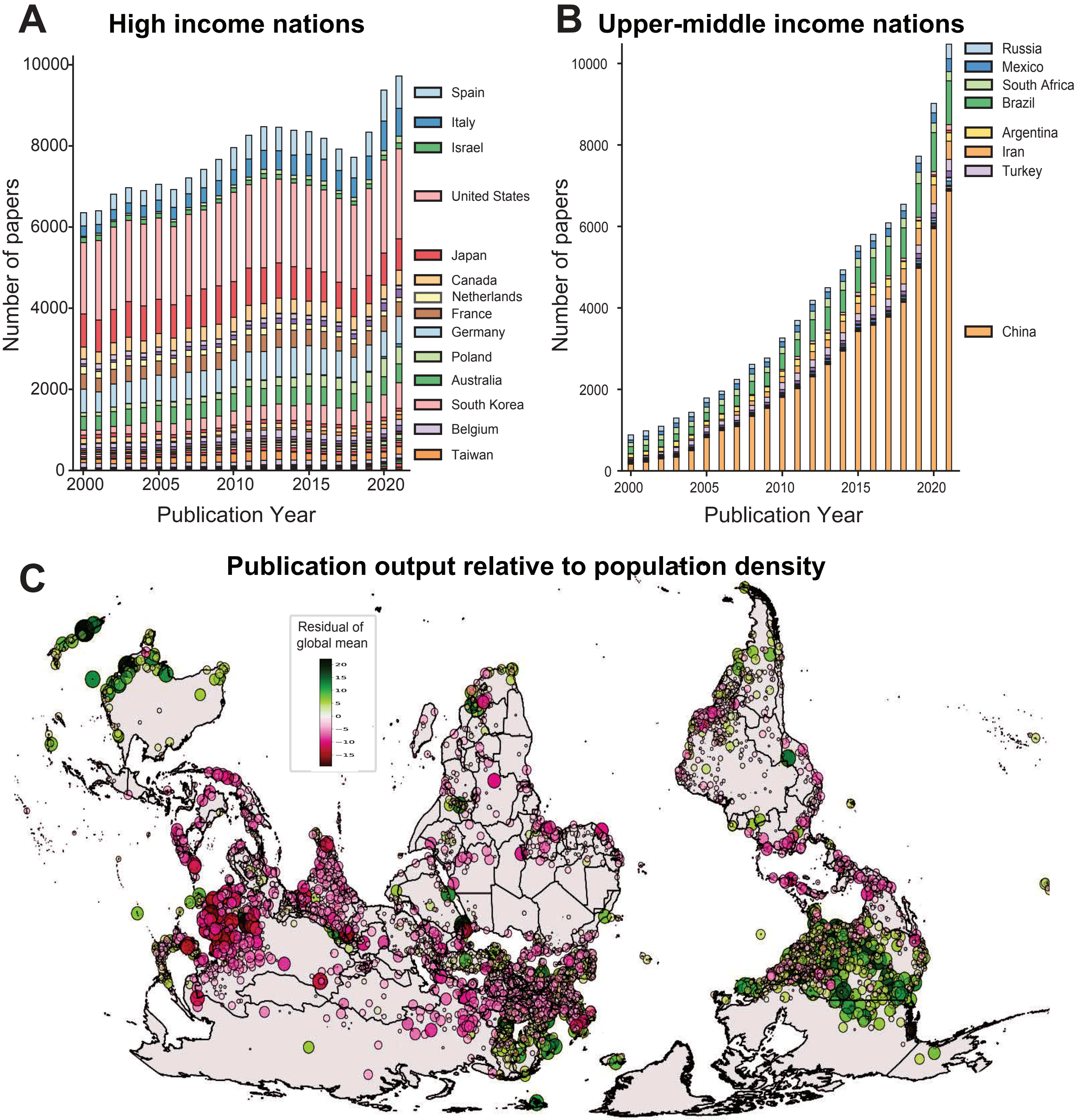
Figure 4. Disparities in the global distribution of corresponding authors by gender. A) Map showing the distribution of name-based gender ratio. All regions where accuracy of name-based gender prediction is less than 90% are faded out. There were many more papers led by authors who had names normatively associated with masculinity (NNMs) than by authors with names normatively associated with femininity (NNFs), but the extent of the imbalance was variable across the globe. B) The total number of publications led by authors with NNMs and NNFs, C) the impact factor of the journals that NNMs and NNFs published in, and D) the citation rates for papers led by authors with NNMs and NNFs.

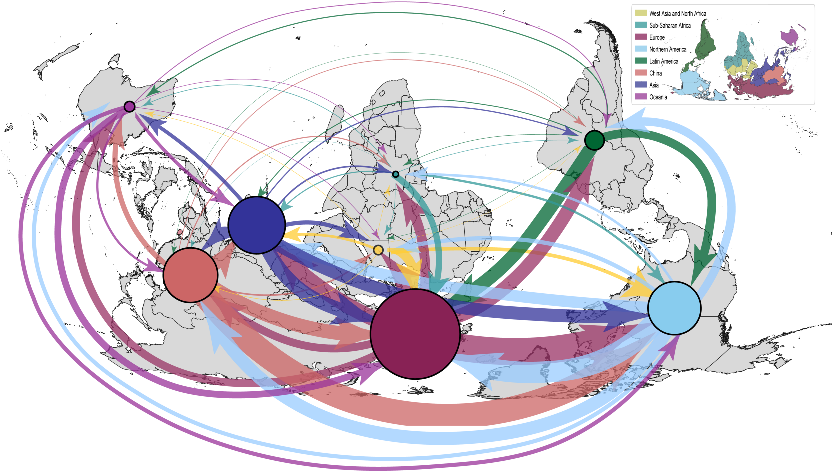
Figure 5. Stagnant gender bias over the past two decades. The proportion of authors with NNFs over the last 20 years is plotted for each of the eight geographical regions investigated.

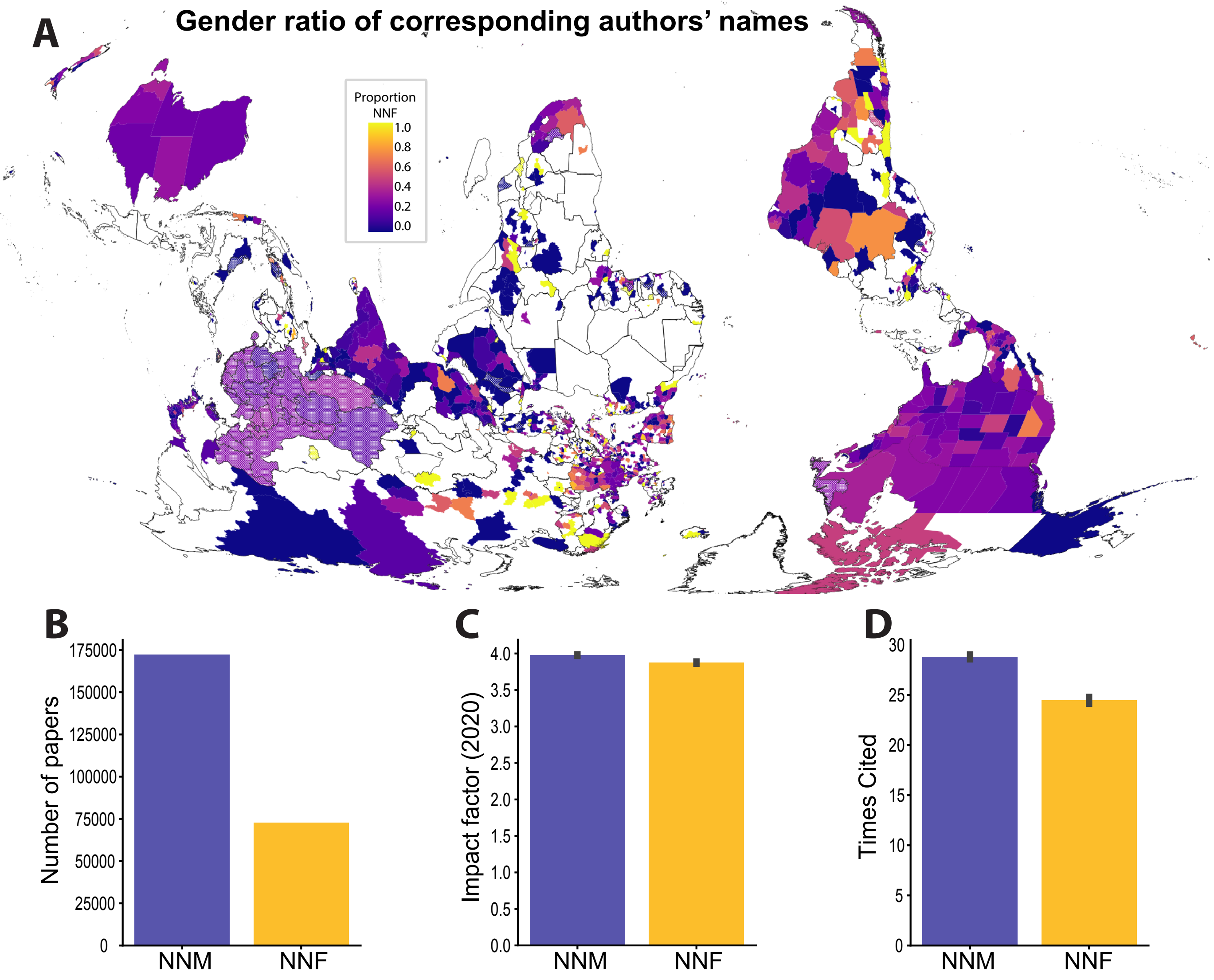
Figure 6. A) The top 20 most studied plant species across all studies. B) The top 9 most studied orders for non-plant groups (animalia, fungi, and bacteria). C) The observed number of studies investigating each order or land plants minus the number expected if sampling effort had been evenly distributed relative to species richness.

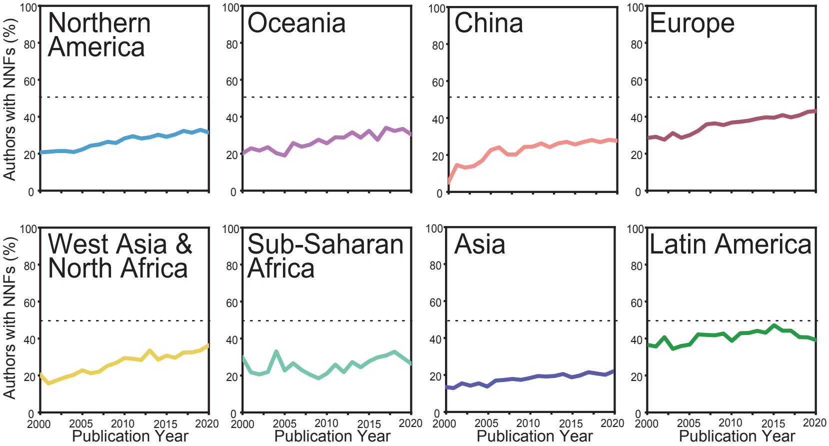
Figure 7. National differences in focal organism choice. The top 10 most studied species from the literature in select nations is plotted. Nations are organized from most prolific to least, and focal organisms are colored by generalized groupings of organism type. The x-axis shows the number of papers that focus on each focal organism.

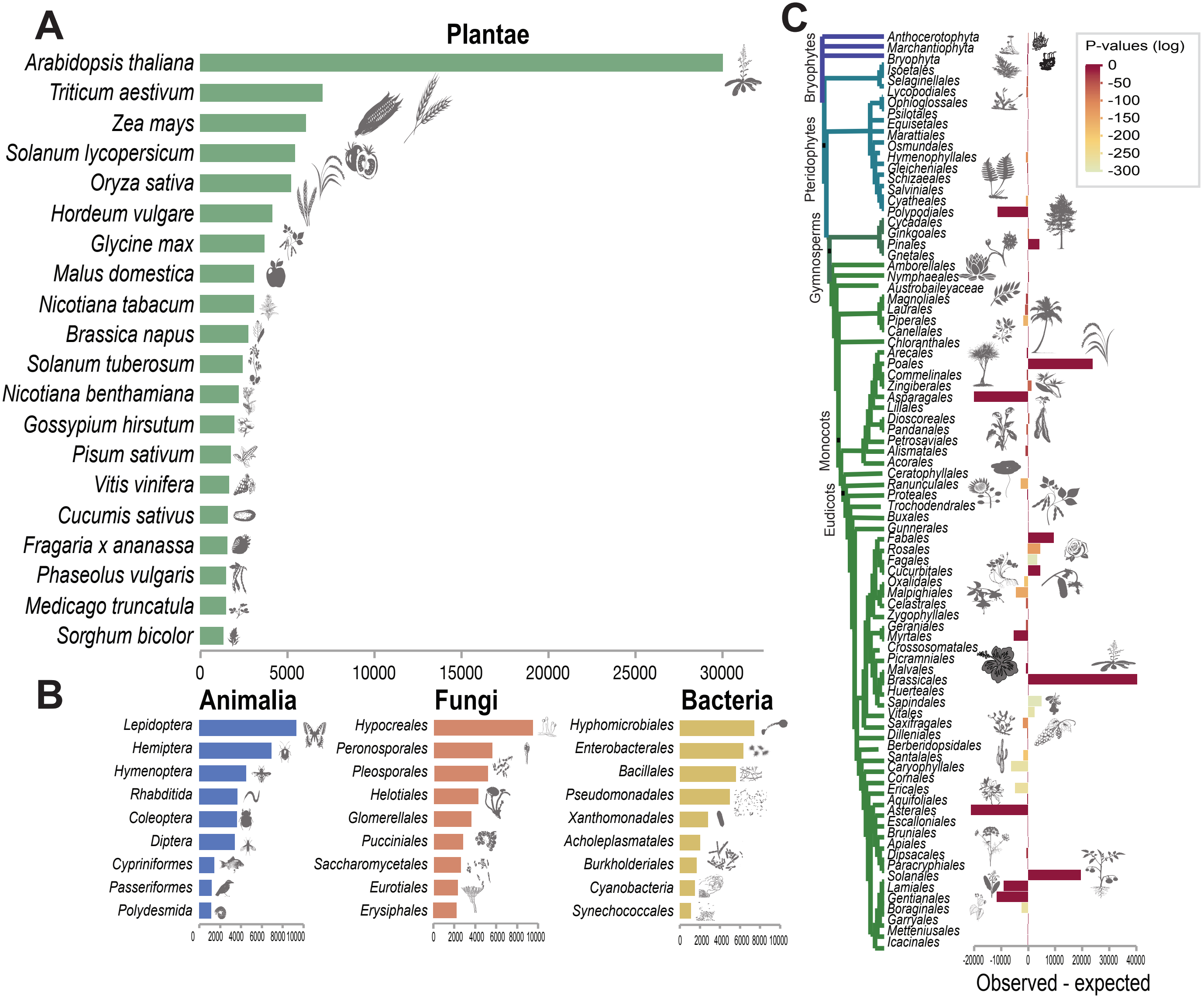

A
Global distribution of authors
B
Continent
West Asia and North Africa
Sub-Saharan Africa
Europe
Northern America
Latin America
Asia
Oceania
China
USA
Japan
Germany
Spain
South Africa
Mexico
Hungary
Tunisia
Saudi Arabia
Indonesia
Benin
UAE
Iraq
Qatar
Fiji
Vanuatu
Lesotho
Libya
R2=0.75
Number of Papers
10^0
10^1
10^2
10^3
10^4
10^9
10^{10}
10^{11}
10^{12}
10^{13}
GDP (2019 USD)
C
R2=0.83
China
USA
India
Japan
Germany
France
Brazil
South Africa
Argentina
Mexico
Pakistan
Tunisia
Chile
Singapore
Uganda
Indonesia
Trinidad and Tobago
UAE
Jamaica
Qatar
Madagascar
Malta
Mauritania
Greenland
Lesotho
Angola
Bermuda
Continent
West Asia and North Africa
Sub-Saharan Africa
Europe
Northern America
Latin America
Asia
Oceania
Number of Papers
10^0
10^1
10^2
10^3
10^4
10^6
10^7
10^8
10^9
10^{10}
10^{11}
10^{12}
Research and development expenditure (USD)

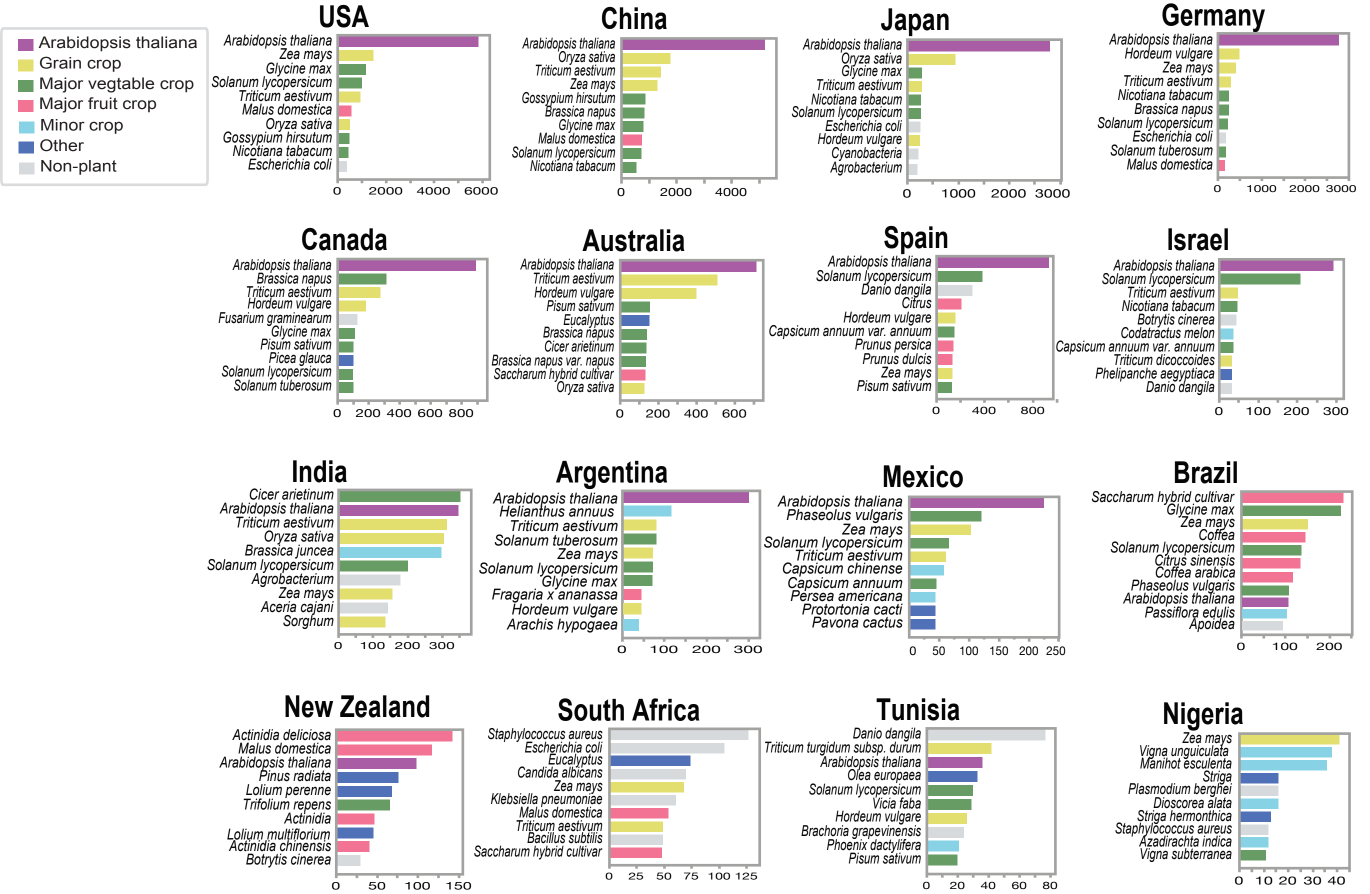

A
High income nations
Number of papers
0
2000
4000
6000
8000
10000
2000
2005
2010
2015
2020
Publication Year
Spain
Italy
Israel
United States
Japan
Canada
Netherlands
France
Germany
Poland
Australia
South Korea
Belgium
Taiwan
B
Upper-middle income nations
Number of papers
0
2000
4000
6000
8000
10000
2000
2005
2010
2015
2020
Publication Year
Russia
Mexico
South Africa
Brazil
Argentina
Iran
Turkey
China
C
Publication output relative to population density
Residual of global mean
20
15
10
5
0
−5
−10
−15


West Asia and North Africa
Sub-Saharan Africa
Europe
Northern America
Latin America
China
Asia
Oceania


A
Gender ratio of corresponding authors' names
Proportion NNF
1.0
0.8
0.6
0.4
0.2
0.0
B
Number of papers
175000
150000
125000
100000
75000
50000
25000
0
NNM
NNF
C
Impact factor (2020)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
NNM
NNF
D
Times Cited
30
25
20
15
10
5
0
NNM
NNF


Northern America
Oceania
China
Europe
West Asia & North Africa
Sub-Saharan Africa
Asia
Latin America
Authors with NNFs (%)
Publication Year
0
20
40
60
80
100
2000
2005
2010
2015
2020


A
Plantae
Arabidopsis thaliana
Triticum aestivum
Zea mays
Solanum lycopersicum
Oryza sativa
Hordeum vulgare
Glycine max
Malus domestica
Nicotiana tabacum
Brassica napus
Solanum tuberosum
Nicotiana benthamiana
Gossypium hirsutum
Pisum sativum
Vitis vinifera
Cucumis sativus
Fragaria x ananassa
Phaseolus vulgaris
Medicago truncatula
Sorghum bicolor
0
5000
10000
15000
20000
25000
30000
B
Animalia
Lepidoptera
Hemiptera
Hymenoptera
Rhabditida
Coleoptera
Diptera
Cypriniformes
Passeriformes
Polydesmida
0
2000
4000
6000
8000
10000
Fungi
Hypocreales
Peronosporales
Pleosporales
Helotiales
Glomerellales
Pucciniales
Saccharomycetales
Eurotiales
Erysiphales
0
2000
4000
6000
8000
10000
Bacteria
Hyphomicrobiales
Enterobacterales
Bacillales
Pseudomonadales
Xanthomonadales
Acholeplasmatales
Burkholderiales
Cyanobacteria
Synechococcales
0
2000
4000
6000
8000
10000
C
Bryophytes
Pteridophytes
Gymnosperms
Monocots
Eudicots
Anthocerotophyta
Marchantiophyta
Bryophyta
Isoetales
Selaginellales
Lycopodiales
Ophioglossales
Psilotales
Equisetales
Marattiales
Osmundales
Hymenophyllales
Gleicheniales
Schizaeales
Salviniales
Cyatheales
Polypodiales
Cycadales
Ginkgoales
Pinales
Gnetales
Amborellales
Nymphaeales
Austrobaileyaceae
Magnoliales
Laurales
Piperales
Canellales
Chloranthales
Arecales
Poales
Commelinales
Zingiberales
Asparagales
Liliales
Dioscoreales
Pandanales
Petrosaviales
Alismatales
Acorales
Ceratophyllales
Ranunculales
Proteales
Trochodendrales
Buxales
Gunnerales
Fabales
Rosales
Fagales
Cucurbitales
Oxalidales
Malpighiales
Celastrales
Zygophyllales
Geraniales
Myrtales
Crossosomatales
Picramniales
Malvales
Brassicales
Huerteales
Sapindales
Vitales
Saxifragales
Dilleniales
Berberidopsidales
Santalales
Caryophyllales
Cornales
Ericales
Aquifoliales
Asterales
Escalloniales
Bruniales
Apiales
Dipsacales
Paracryphiales
Solanales
Lamiales
Gentianales
Boraginales
Garryales
Metteniusales
Icacinales
P-values (log)
0
-50
-100
-150
-200
-250
-300
-20000
-10000
0
10000
20000
30000
40000
Observed - expected


Arabidopsis thaliana
Grain crop
Major vegtable crop
Major fruit crop
Minor crop
Other
Non-plant
USA
Arabidopsis thaliana
Zea mays
Glycine max
Solanum lycopersicum
Triticum aestivum
Malus domestica
Oryza sativa
Gossypium hirsutum
Nicotiana tabacum
Escherichia coli
0
2000
4000
6000
China
Arabidopsis thaliana
Oryza sativa
Triticum aestivum
Zea mays
Gossypium hirsutum
Brassica napus
Glycine max
Malus domestica
Solanum lycopersicum
Nicotiana tabacum
0
2000
4000
Japan
Arabidopsis thaliana
Oryza sativa
Glycine max
Triticum aestivum
Nicotiana tabacum
Solanum lycopersicum
Escherichia coli
Hordeum vulgare
Cyanobacteria
Agrobacterium
0
1000
2000
3000
Germany
Arabidopsis thaliana
Hordeum vulgare
Zea mays
Triticum aestivum
Nicotiana tabacum
Brassica napus
Solanum lycopersicum
Escherichia coli
Solanum tuberosum
Malus domestica
0
1000
2000
3000
Canada
Arabidopsis thaliana
Brassica napus
Triticum aestivum
Hordeum vulgare
Fusarium gramineaum
Glycine max
Pisum sativum
Picea glauca
Solanum lycopersicum
Solanum tuberosum
0
200
400
600
800
Australia
Arabidopsis thaliana
Triticum aestivum
Hordeum vulgare
Pisum sativum
Eucalyptus
Brassica napus
Cicer arietinum
Brassica napus var. napus
Saccharum hybrid cultivar
Oryza sativa
0
200
400
600
Spain
Arabidopsis thaliana
Solanum lycopersicum
Danio dangila
Citrus
Hordeum vulgare
Capsicum annuum var. annuum
Prunus persica
Prunus dulcis
Zea mays
Pisum sativum
0
400
800
Israel
Arabidopsis thaliana
Solanum lycopersicum
Triticum aestivum
Nicotiana tabacum
Botrytis cinerea
Codatractus melon
Capsicum annuum var. annuum
Triticum dicoccoides
Phelipanche aegyptiaca
Danio dangila
0
100
200
300
India
Cicer arietinum
Arabidopsis thaliana
Triticum aestivum
Oryza sativa
Brassica juncea
Solanum lycopersicum
Agrobacterium
Zea mays
Aceria cajani
Sorghum
0
100
200
300
Argentina
Arabidopsis thaliana
Helianthus annuus
Triticum aestivum
Solanum tuberosum
Zea mays
Solanum lycopersicum
Glycine max
Fragaria x ananassa
Hordeum vulgare
Arachis hypogaea
0
100
200
300
Mexico
Arabidopsis thaliana
Phaseolus vulgaris
Zea mays
Solanum lycopersicum
Triticum aestivum
Capsicum chinense
Capsicum annuum
Persea americana
Protortonia cacti
Pavona cactus
0
50
100
150
200
250
Brazil
Saccharum hybrid cultivar
Glycine max
Zea mays
Coffea
Solanum lycopersicum
Citrus sinensis
Coffea arabica
Phaseolus vulgaris
Arabidopsis thaliana
Passiflora edulis
Apoidea
0
100
200
New Zealand
Actinidia deliciosa
Malus domestica
Arabidopsis thaliana
Pinus radiata
Lolium perenne
Trifolium repens
Actinidia
Lolium multiflorium
Actinidia chinensis
Botrytis cinerea
0
50
100
150
South Africa
Staphylococcus aureus
Escherichia coli
Eucalyptus
Candida albicans
Zea mays
Klebsiella pneumoniae
Malus domestica
Triticum aestivum
Bacillus subtilis
Saccharum hybrid cultivar
0
25
50
75
100
125
Tunisia
Danio dangila
Triticum turgidum subsp. durum
Arabidopsis thaliana
Olea europaea
Solanum lycopersicum
Vicia faba
Hordeum vulgare
Brachoria grapevinensis
Phoenix dactylifera
Pisum sativum
0
20
40
60
80
Nigeria
Zea mays
Vigna unguiculata
Manihot esculenta
Striga
Plasmodium berghei
Dioscorea alata
Striga hermonthica
Staphylococcus aureus
Azadirachta indica
Vigna subterranea
0
10
20
30
40