

# Illuminating Women's Hidden Contribution to Historical Theoretical Population Genetics

Samantha Kristin Dung,<sup>\*1</sup> Andrea López,<sup>\*1</sup> Ezequiel Lopez Barragan,<sup>\*1</sup> Rochelle-Jan Reyes,<sup>\*1</sup> Ricky Thu,<sup>\*1</sup> Edgar Castellanos,<sup>\*</sup> Francisca Catalan,<sup>\*</sup> Emilia Huerta-Sánchez,<sup>†,\*,§,2,3</sup> and Rori V. Rohlf<sup>\*,2,3</sup>

<sup>\*</sup>Department of Biology, San Francisco State University, California 94132, <sup>†</sup>Department of Molecular and Cell Biology, University of California, Merced, California 95340, and <sup>‡</sup>Department of Ecology and Evolutionary Biology and <sup>§</sup>Center for Computational Molecular Biology, Brown University, Providence, Rhode Island 02912

ORCID IDs: 0000-0002-1506-5494 (E.H.-S.); 0000-0002-2625-192X (R.V.R.)

**ABSTRACT** While productivity in academia is measured through authorship, not all scientific contributors have been recognized as authors. We consider nonauthor “acknowledged programmers” (APs), who developed, ran, and sometimes analyzed the results of computer programs. We identified APs in *Theoretical Population Biology* articles published between 1970 and 1990, finding that APs were disproportionately women ( $P = 4.0 \times 10^{-10}$ ). We note recurrent APs who contributed to several highly-cited manuscripts. The occurrence of APs decreased over time, corresponding to the masculinization of computer programming and the shift of programming responsibilities to individuals credited as authors. We conclude that, while previously overlooked, historically, women have made substantial contributions to computational biology. For a video of this abstract, see: <https://vimeo.com/313424402>.

**KEYWORDS** acknowledged programmer; authorship; computational biology; population genetics; women in science

**P**LENTIFUL evidence shows a historic and continuing gender gap in participation and success in scientific research (Beede *et al.* 2011; Shen 2013). However, less attention has been directed at clarifying obscured contributions of women to science. The lack of visible women role models (particularly in quantitative fields) contributes to a reduced sense of belonging and retention among women (Steele 1997; Cheryan *et al.* 2017). We seek to counteract this cycle by illuminating the historical contribution of women programmers to our own fields, population and evolutionary genetics.

To do so, we consider computational population genetics in the 1970s. In this era, leaders in molecular evolution developed innovative methods to test evolutionary hypotheses (Ewens 1972; Felsenstein 1974; Watterson 1975). Many of

these methods were designed for protein variation data, and continue to be widely applied today on DNA sequence data. This work relied on extensive simulations and numerical approaches that were made possible by advances in computation. However, programming these computations required a detailed understanding of computational hardware, as well as a strong foundation in mathematics. Based on authorship at the time, it seems that this research was conducted by a relatively small number of independent individual scientists, nearly all of whom were men.

However, in some of the seminal papers from this time, we noticed that nonauthor computer programmers are thanked in the acknowledgments. Due to authorship norms at the time, these programmers were credited in the acknowledgments sections of manuscripts, rather than being recognized as authors. While this practice was typical at the time, these contributions might well have resulted in authorship today. For example, one acknowledgment reads “I thank Mrs. M. Wu for help with the numerical work, and in particular for computing table I.” (Watterson 1975).

We are now in a cultural moment when the historical scientific contributions of women and people of color are being increasingly revealed to popular audiences (*e.g.*, *Hidden Figures*) (Shetterly 2016; Evans 2018). In that

Copyright © 2019 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.118.301277>

Manuscript received June 25, 2018; accepted for publication October 11, 2018; published Early Online November 3, 2018.

Available freely online through the author-supported open access option.

Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.7280351>.

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>These senior authors contributed equally to this work.

<sup>3</sup>Corresponding authors: San Francisco State University, 1600 Holloway Ave., San Francisco, CA 94132. E-mail: rrohlf@sfsu.edu; and emilia\_huerta-sanchez@brown.edu

context, we performed a gender analysis of the scientific contribution of these “acknowledged programmers” (APs).

We identified APs in *Theoretical Population Biology (TPB)* articles published between 1970 and 1990. Using these data, we analyzed the gender representation among authors and APs, and trends over time. We use citation data to assess the impact of AP-supported manuscripts, relative to manuscripts without AP support. Finally, we note recurrent APs who contributed to several manuscripts.

### Quantifying the Contribution of APs

We selected the journal *TPB* because of its high density of population genetics articles that involved programming. We manually collected the author names, institutional affiliations, acknowledgments text, and APs for all articles published in *TPB* from 1970 to 1990. While work in computational population genetics began before 1970, our *TPB*-based analysis begins when the journal was first published. We classified both authors and APs into binary gender categories (men and women, see Supplemental Materials).

Cumulatively, over 883 articles, of individuals with classifiable binary gender, significantly more APs were women (43.2%) as compared to authors (7.4%) (Table 1) (two-tailed Fisher’s exact test,  $P = 4.0 \times 10^{-10}$ ). This difference is even more striking when considering just the 1970s, when 7.0% of authors were women and 58.6% of APs were women (Table 1).

The acknowledgment of women programmers peaks in the mid-1970s, after which the proportion of APs who are women decreases significantly (Figure 1 and Supplemental Material, Figures S1 and S2, and Table S1, one-tailed Fisher’s exact test,  $P = 4.3 \times 10^{-3}$ ). This parallels the broader cultural shift, which moved computer programming from pink collar work (where women workers are overrepresented and pay is typically low) to a respected male-dominated field (Vogel 2017). Between the 1970s and 1980s, the practice of acknowledging programmers declined as programming duties were likely transferred to graduate students, postdocs, and faculty who received authorship (Table S1, two-tailed Fisher’s exact test,  $P = 0.034$ ) (W. Hill, personal communication). Over the same time, we see a nonsignificant increase in the proportion of women authors (two-tailed Fisher’s exact test,  $P = 0.64$ ).

### AP Narratives and Contributions

In our data, three APs were acknowledged more than once over the years analyzed. When Barbara McCann worked as a research assistant at Brown University (Matter 1970), she was an AP for two articles in *TPB*, as well as an author of two papers (Tables S1 and S2). Jennifer Smith was acknowledged for programming and numerical analysis in three articles in *TPB*, as well as at least three additional articles in *Biometrics* when she was a computing assistant at the University of Edinburgh (Tables S1 and S2). Lastly, Margaret

**Table 1 Author and acknowledged programmer gender**

	Authors	Acknowledged programmers
Women	80 (38)	19 (17)
Men	998 (502)	25 (12)
Ambiguous	164 (24)	2 (2)

Number and binary gender of authors and APs contributing to *TPB* 1970-1990. Parenthetical numbers indicate contributions from 1970-1979.

Wu, a research assistant in the Department of Mathematics at Monash University, was acknowledged in two papers in *TPB*, one of which has been cited over 3400 times as it established a widely used estimator of genetic diversity (known as Watterson’s estimator) (Table S1). She was an AP in at least three additional manuscripts (Table S2). She went on to earn a PhD and hold a faculty position at the University of Melbourne, where she developed statistical methods to analyze educational data (Wu 2011).

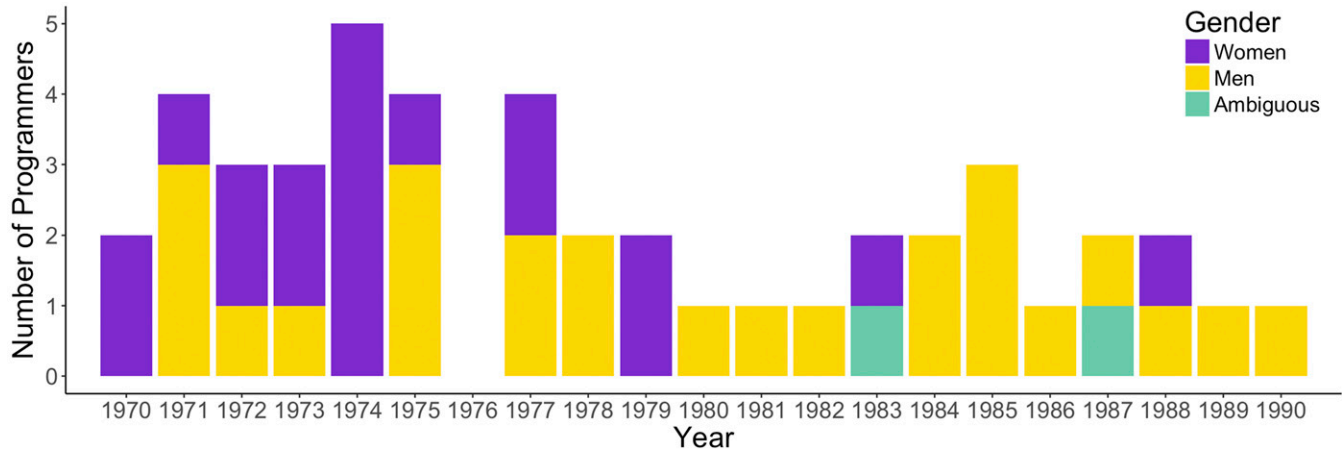
The specific technical contribution of an AP likely varied over projects. However, the fact that authors repeatedly chose to work with some APs suggests that these recurrent APs contributed particular expertise. Specifically, in addition to programming and numerical work, Jennifer Smith developed algorithms to carry out verbally specified analyses (W. Hill, personal communication). Margaret Wu performed a variety of statistical work including developing estimators for parameter values, devising algorithms for statistical tasks, and sometimes creating numerical methodology (M. Wu, personal communication). Overall, women’s contributions were substantial in terms of the high proportion of contributions from women APs, as well as in quality.

To begin to assess if papers with AP contributors had a disproportionate impact on the field, we compared the number of citations for AP-supported vs. non-AP papers. The support of an AP is nonsignificantly correlated with the number of citations (Pearson’s correlation test  $P = 0.06$ ). A more powerful analysis of broader data may clarify this relationship.

### Scientific Contributions and Authorship Norms

Our retrospective analysis has shed light on the contributions of women to computational genetics research. These women’s contributions were previously obscured by being relegated to footnote acknowledgments due to authorship norms. We showed that women’s contributions were substantial when measured by volume (the high proportion of contributions from women APs), as well as by quality when we consider that some women APs were involved in seminal papers and the development of cutting edge approaches.

Our findings raise questions about how our current norms of scientific credit may favor certain individuals or groups. For instance, the bibliometric *h* index (*h* such that a scholar authored *h* papers that have been cited at least *h* times) has gained popularity, in part due to its correlation with other indicators of academic success such as National Academy membership or Nobel prize laureateship (Hirsch 2005).



**Figure 1** Total number of programmers acknowledged per year in *Theoretical Population Biology*. Each bar indicates the total number of acknowledged programmers broken down by binary gender. Purple, women; yellow, men; and green, ambiguous.

However, this concordance may reflect that the  $h$  index is consistent with biases in scientific recognition processes (Kelly and Jennions 2007). Furthermore, in the present age of highly collaborative science, authorship can be difficult to interpret. Noncontributor authors (authors who may have provided data, materials, or funding, but made no intellectual or practical contribution) are presented in an estimated 35% of publications in biology, while nonauthors, particularly technicians, are estimated to contribute to 56% of publications in biology (Jabbekhadi and Walsh 2017). Because scientific roles (e.g., technician, student, or principal investigator) are related to social factors (e.g., gender, race, class background, and nationality), contributions from particular groups likely remain obscured.

In the two cases where we have specific information, without a clear path to persist in science, women APs left science to care for children and spouses (W. Hill, personal communication and M. Wu, personal communication). Today, the path for graduate students to advance still tends to favor men (Blickenstaff 2005; Martinez *et al.* 2007). It is fitting that our analysis is in computational population genetics because, while women's representation in evolutionary biology has dramatically improved in the past decades (Wellenreuther and Otto 2016), women are still underrepresented in population genetics (Telis 2017) and computational biology (Bonham and Stefan 2017).

The field of population genetics that we have studied here was chosen because it is our own field, not because we had a prior expectation that this particular field would reveal significant hidden contributions by women. Since we do find that women scientists' contributions to this field were often obscured, we speculate that, rather than being rare, this may be a general trend. It would therefore be interesting to test this further in other scientific fields. When contributions are hidden, this exaggerates perceptions that women are minor participants in work in STEM (science, technology, engineering, and mathematics) fields. Because perceiving underrepresentation in the field can impact a woman's

performance (Steele 1997; Cheryan *et al.* 2015), improving awareness of women's contributions may play a role in improving gender equity.

### Acknowledgments

We thank Michael DeGiorgio, Tracey Heath, Emily Jane McTavish, Fergal Casey, Rasmus Nielsen, and Pleuni Pennings for constructive comments on an earlier version of this manuscript; Michael Turelli, Jim Harner, and Michael Rose for helping identify the genders of ambiguously named acknowledged programmers; and Margaret Wu and Bill Hill for information that shaped our research. This work was funded by San Francisco Building Infrastructure Leading to Diversity National Institutes of Health grant number 1UL1-GM-118985, National Institutes of Health grant 1R35GM128946-01, and E.H.-S. was supported by National Science Foundation grant NSF-DEB-1557151. The authors declare no competing interests.

Author contributions: conceptualization: E.H.-S. and R.V.R.; methodology: E.H.-S. and R.V.R.; data curation: E.C., F.C., S.K.D., A.L., E.L.B., R.-J.R., and R.T.; investigation: S.K.D., A.L., E.L.B., R.-J.R., R.V.R., and R.T.; writing: S.K.D., E.H.-S., A.L., E.L.B., R.-J.R., R.V.R., and R.T.; and supervision: E.H.-S. and R.V.R.

### Literature Cited

- Beede, D. N., T. A. Julian, D. Langdon, G. McKittrick, B. Khan *et al.*, 2011 Women in STEM: a gender gap to innovation (August 1, 2011). Economics and Statistics Administration Issue Brief No. 04–11. Accessed June 2018. Available at SSRN: <https://ssrn.com/abstract=1964782> or <http://dx.doi.org/10.2139/ssrn.1964782>.
- Blickenstaff, J. C., 2005 Women and science careers: leaky pipeline or gender filter? *Gend. Educ.* 17: 369–386. <https://doi.org/10.1080/09540250500145072>
- Bonham, K. S., and M. I. Stefan, 2017 Women are underrepresented in computational biology: an analysis of the scholarly literature in biology, computer science and computational

- biology. *PLOS Comput. Biol.* 13: e1005134. <https://doi.org/10.1371/journal.pcbi.1005134>
- Cheryan, S., A. Master, and A. N. Meltzoff, 2015 Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Front. Psychol.* 6: 49. <https://doi.org/10.3389/fpsyg.2015.00049>
- Cheryan, S., S. A. Ziegler, A. K. Montoya, and L. Jiang, 2017 Why are some STEM fields more gender balanced than others? *Psychol. Bull.* 143: 1–35. <https://doi.org/10.1037/bul0000052>
- Evans, C. L., 2018 *Broad Band: The Untold Story of the Women Who Made the Internet*. Penguin, New York City.
- Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 3: 87–112. [https://doi.org/10.1016/0040-5809\(72\)90035-4](https://doi.org/10.1016/0040-5809(72)90035-4)
- Felsenstein, J., 1974 The evolutionary advantage of recombination. *Genetics* 78: 737–756.
- Hirsch, J. E., 2005 An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* 102: 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Jabbehdari, S., and J. P. Walsh, 2017 Authorship norms and project structures in science. *Sci. Technol. Human Values* 42: 872–900. <https://doi.org/10.1177/0162243917697192>
- Kelly, C. D., and M. D. Jennions, 2007 H-index: age and sex make it unreliable. *Nature* 449: 403. <https://doi.org/10.1038/449403c>
- Martinez, E. D., J. Botos, K. M. Dohoney, T. M. Geiman, S. S. Kolla *et al.*, 2007 Falling off the academic bandwagon: women are more likely to quit at the postdoc to principal investigator transition. *EMBO Rep.* 8: 977–981. <https://doi.org/10.1038/sj.embor.7401110>
- Matter, F., 1970 *Milbank Mem. Fund Q.* 48. Available at: <https://www.jstor.org/stable/pdf/3349285.pdf>.
- Shen, H., 2013 Inequality quantified: mind the gender gap. *Nature* 495: 22–24. <https://doi.org/10.1038/495022a>
- Shetterly, M. L., 2016 *Hidden Figures: The American Dream and the Untold Story of the Black Women Mathematicians Who Helped Win the Space Race*. William Morrow, New York City.
- Steele, C. M., 1997 A threat in the air: how stereotypes shape intellectual identity and performance. *Am. Psychol.* 52: 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>
- Telis, N., 2017 Questions and answers about questions. Accessed June 2018. Available at: <https://telis.blog/2017/11/06/questions-and-answers-about-questions>.
- Vogel, W., 2017 'The spitting image of a woman programmer': changing portrayals of women in the American computing industry, 1958–1985. *IEEE Ann. Hist. Comput.* 39: 49–64.
- Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7: 256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9)
- Wellenreuther, M., and S. Otto, 2016 Women in evolution – highlighting the changing face of evolutionary biology. *Evol. Appl.* 9: 3–16. <https://doi.org/10.1111/eva.12343>
- Wu, M., 2011 Curriculum vitae. Accessed June 2018. Available at: [www.cdhp.dn.edu.tw/ezfiles/53/1053/attach/35/pta\\_11174\\_2345834\\_04754.doc](http://www.cdhp.dn.edu.tw/ezfiles/53/1053/attach/35/pta_11174_2345834_04754.doc)

Communicating editor: A. S. Wilkins