A remote lecture series roadmap to equity, diversity, and inclusion in STEM

Evan A. Boyle[1,#], Gabriela Goldberg[1], Jonathan C. Schmok[1], Jillybeth Burgado[2], Fabiana Izidro Layng[3], Hannah A. Grunwald[4], Kylie M. Balotin[5], Michael S. Cuoco[6], Keng-Chi Chang[7], Gertrude Ecklu-Mensah[8], Aleena K. S. Arakaki[9], Noorsher Ahmed[1], Ximena Garcia Arceo[10], Pratibha Jagannatha[1], Jonathan Pekar[11], Mallika Iyer[12], DASL Alliance, Gene W. Yeo[1,13,14,#]

1 Department of Cellular & Molecular Medicine, University of California San Diego
2 Molecular Neurobiology Laboratory, The Salk Institute for Biological Studies
3 Cancer Center, Sanford Burnham Prebys Medical Discovery Institute
4 Department of Genetics, Harvard Medical School
5 Department of Biomedical Engineering, Vanderbilt University
6 Laboratory of Genetics, The Salk Institute for Biological Studies
7 Department of Political Science, University of California San Diego
8 Department of Pediatrics, University of California San Diego
9 Human Biology Division, Fred Hutchinson Cancer Research Center
10 Department of Chemistry and Biochemistry, University of California San Diego
11 Department of Biomedical Informatics, University of California San Diego
12 Graduate School of Biomedical Sciences, Sanford Burnham Prebys Medical Discovery Institute
13 Stem Cell Program, University of California San Diego
14 Institute for Genomic Medicine, University of California San Diego
Co-corresponding author: eboyle@ucsd.edu, geneyeo@ucsd.edu

Abstract

Within a year of the shutdowns caused by the COVID-19 pandemic, virtual meetings transformed from an auxiliary service to an essential work platform for hundreds of millions of people worldwide. Universities rapidly accelerated adoption of virtual platforms for remote conferences, classes, and seminars amidst a second crisis that started in May 2020 and has tested institutional commitment to the principles of equity, diversity, and inclusion. To address these concerns, we began the Diversity and Science Lecture series (DASL), a cross-institutional national platform where junior life scientists present personal stories, professional progress, and advice for their peers. We first demonstrate benchmarking of speaker diversity using public data and review scientific talk content by text mining of talk titles from DASL. We then review insight

into equity, diversity, and inclusion in STEM presented by DASL speakers. Notably, nearly every speaker highlighted the importance of social or interpersonal support to their career progression. Three-fifths of speakers commented on race or ethnicity. Other identities received considerable attention from a minority of speakers: immigrant identity, gender identity, mental illness, sexual minorities, disability, and rural or agricultural background. In addition, we report that ancillary small group sessions with speakers achieve greater engagement when held before formal seminars versus after. To facilitate future programming, we provide exhaustive guides that frame the goals of virtual programming in terms of benefits for executive team members, speakers, and attendees. In total, we have curated over 75 resources for personal and professional development cited by trainee life scientists, attracted over 4,000 unique website visitors, and exceeded 3,500 hours of engagement from attendees at weekly seminars. Our success with DASL confirms that community building can flourish under a remote lecture framework.

Introduction

We are trainee life scientists who created a platform where our peers in San Diego and beyond can share their research and comment on equity, diversity, and inclusion in STEM: the Diversity and Science Lecture series (DASL). Unlike existing series at UC San Diego that centered either research achievement or personal experience, DASL features integrated presentations on speakers' personal backgrounds, their scientific progress, and their advice for navigating a scientific career.

Hosting a new remote lecture series requires navigation of three overlapping phases – launch, execution, and maintenance – plus the tools needed to administer each phase. Recently published guidelines both for forming a new professional society [1] and for running a webinar series [2] provide high quality information on how to build community around a new remote lecture series. However, the COVID-19 pandemic has prompted reconsideration of the role of

virtual programming in educational settings. With the fraying of social bonds over months of social distancing, methods to foster a supportive culture amongst trainee scientists remotely are desperately needed. At the same time, widespread acclimatization to video conferencing has drastically lowered the burden of organizing virtual meetings.

For one year, we at DASL held weekly Zoom seminars. Each quarter, executive team members recruit speakers and scheduled times for dry-run practice sessions and formal seminars. Each seminar consists of either two 15-20 minute talks from trainee life scientists or one 1 hour talk from a junior life scientist or social science expert. Dry-runs are offered to all trainee speakers to address technical difficulties, provide feedback on talk content, and further discuss research and advocacy. Speakers are encouraged to speak freely about issues important to them and deliver a clear message advising their peers on how to advance equity, diversity, and inclusion in STEM.

Many other prominent virtual seminar series launched in 2020 hewed close to existing scientific communities [3–5]. By contrast, we chose to make DASL accessible to a broad cross-section of life scientists to cultivate a new community. Who claimed this opportunity and why? To answer this question, we have implemented methods to assess speaker diversity and trainee concerns and apply them to the first year of DASL. Informed by our experience, we offer new guidelines for expanding the role of virtual programming in academic institutions to light the way for future remote lecture series aimed at diverse audiences.

Results

In summer 2021, DASL concluded its first year of programming (Figure 1a). What once started as a community at UC San Diego grew to include scientists affiliated with institutions nationwide. After one year, our weekly seminar series featured 79 speakers from 14 institutions. Furthermore, participating institutions outside San Diego have since begun hosting remote lecture series of their own.

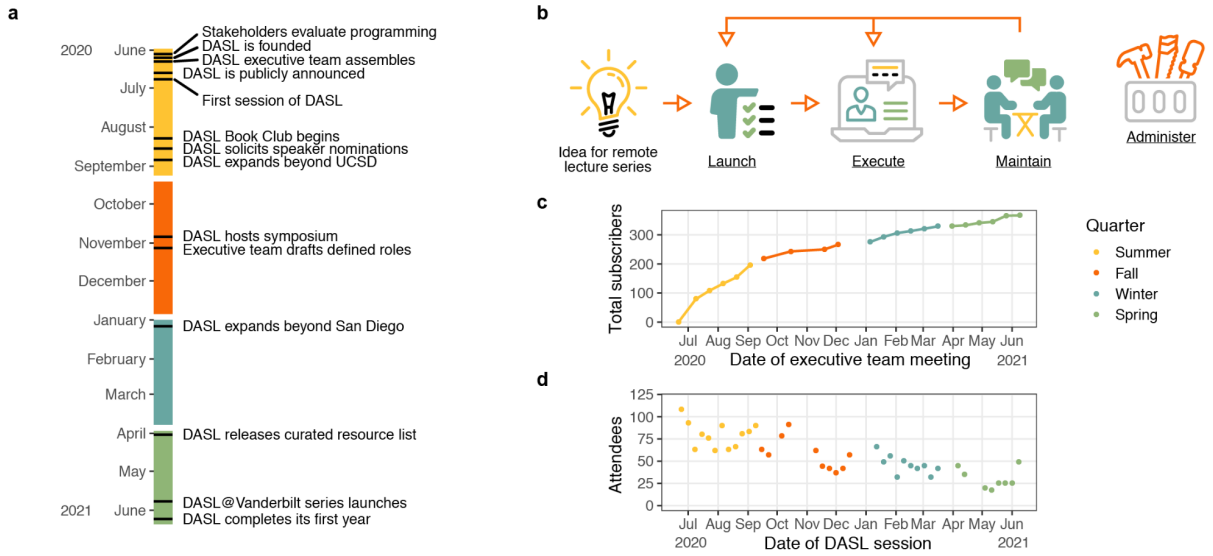

Figure 1: Organizing the first year of DASL. a) Timeline of DASL milestones separated by quarter. DASL was founded in June 2020 and completed its recent spring series in June 2021. b) Outline of running a remote lecture series. We provide guides detailing how to approach each underlined component (Supplementary Text). c) Total subscribers to the DASL mailing list and d) number of attendees per DASL session over the first year of programming.

Uniting trainee scientists in laboratories from across the country to organize a new initiative amidst a national reckoning over racial justice posed complex operational challenges. Foremost among them was how to sustain open and impactful conversation amongst graduate students, postdoctoral scholars, and faculty who had never met in person. Solving this challenge required recognizing how best to iterate and innovate through the phases of launching, executing, and maintaining a remote lecture series, specifically by leveraging tools for easier administration such as mass email lists and document merge functionality (Figure 1b). Over its first year, DASL continuously grew its subscriber base (Figure 1c) and accrued roughly 3,500 hours of engagement from attendees in aggregate (Figure 1d). From this experience, we drafted guides detailing how to approach each phase of a remote lecture series (Figure 1b, Supplementary Text). To further aid future efforts, we include example forms for scheduling and advertising sessions (Supplementary Data).

DASL speakers exhibit scientific, gender, and geographic diversity

In the first year of DASL's weekly series, we recruited 79 speakers from a variety of institutions and academic positions. Graduate students filled the majority of speaker slots (53%), followed by postdoctoral scholars (34%) and assistant professors (6%) (Figure 2a). Overall, a greater percentage of speakers were women (67%) relative to UCSD life scientists broadly (58% of graduate students, 48% of postdocs, and 28% of tenure-track faculty; Figure 2b). We also evaluated the remote lecture series hosted by UCSD's department of Cellular & Molecular Medicine and found that 42% of speakers in 2021 were women (n=31).

We summarized the research disciplines of DASL speakers by parsing keywords in talk titles. "Cell" was the most used term at nine mentions, followed by "protein" and "regulate" at seven mentions, and "cancer", "develop", and "metabolism" at six. In general, DASL speakers' research most frequently related to human cellular and molecular biology (Figure 2c).

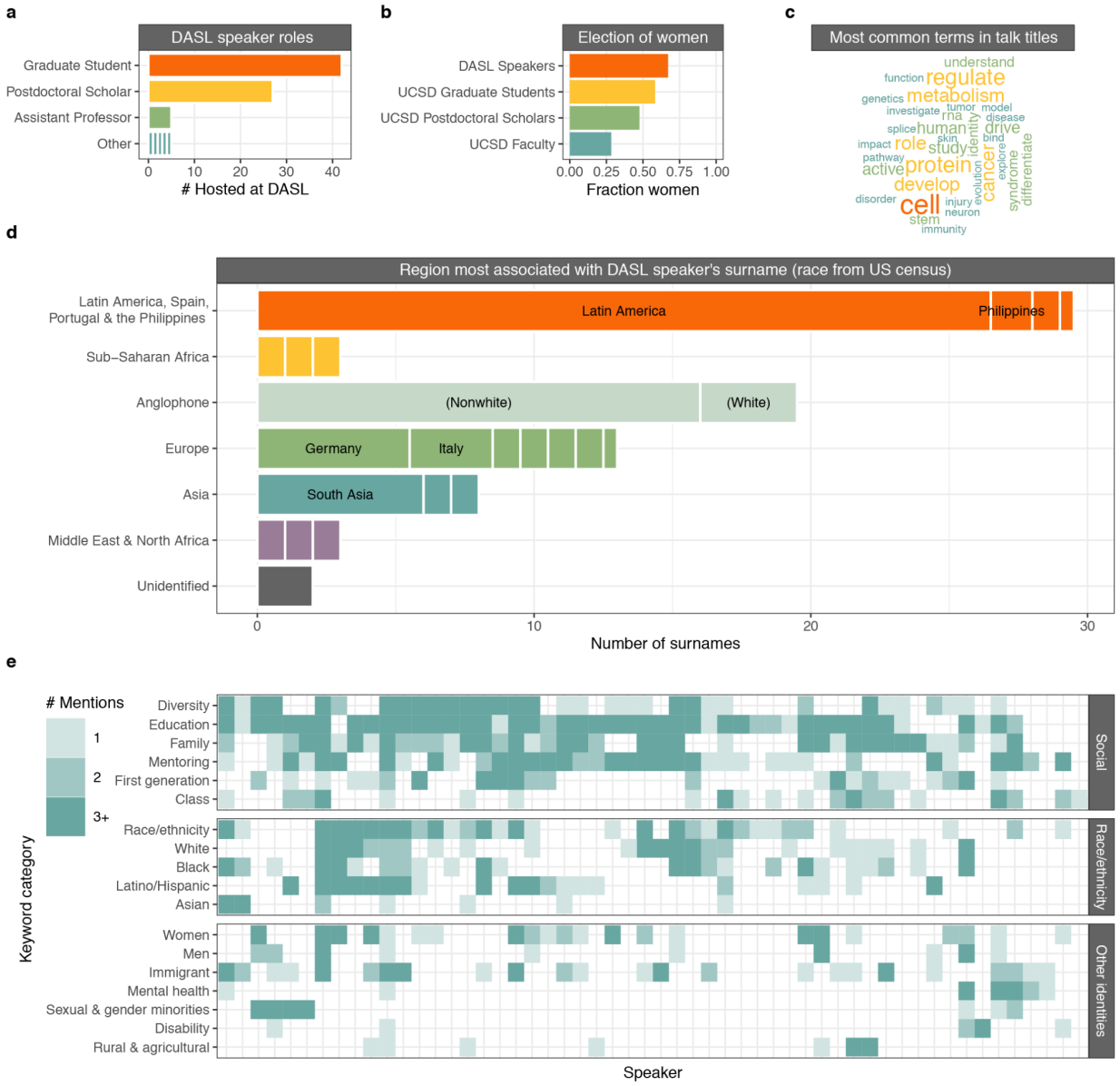

Figure 2: Reviewing the first year of DASL. a) Breakdown of the positions held by DASL's first 79 speakers. The 'Other' category includes Research scientist, Research Assistant Professor, Professor, Assistant Curator, and Administrator. b) Representation of women among DASL speakers compared to UCSD life science graduate students, postdoctoral scholars and faculty. c) Most common terms used in DASL talk titles. The size of the word reflects the frequency. d) Counts of most associated regions for DASL speakers' surnames using the geospatial name database Forebears. 'Anglophone' refers to names associated to Britain, the United States, Canada, Australia, New Zealand or a country in the Caribbean. e) Truncated counts of mentions of keywords across speakers (x-axis) organized by topic (y-axis) and broader type of concern (social, race/ethnicity, or other identities).

To further characterize the diversity of speakers, we determined the country or region in the world where each speaker's surname was most prevalent and abundant using the Forebears database (see Methods). For most names, prevalence and abundance pointed to the same region, enabling imperfect but scalable inference of where relatives of speakers currently reside. (Figure 2d). Names linked to Britain, the United States, Canada, Australia, New Zealand and the Caribbean were grouped in a separate Anglophone category.

Overall, Latin American (30 names) countries were the top associated regions for DASL speakers' surnames. Next most common were Anglophone (20 names) and European countries (13 names) (Figure 2d). We further annotated anglophone-associated surnames as "White" or "Nonwhite" using US Census data linking common US surnames to race [6] and labeled 82% of Anglophone surnames "Nonwhite."

We performed the same analysis for 63 speakers hosted by the Fragile Nucleosome forum, a concurrent international remote lecture series that has promoted community building (Supplementary Figure 1a) [5]. 30 names were associated to European countries and 14 to Anglophone countries. Most Anglophone surnames (64%, n=14) were labeled "White." Overall, 30% of speaker surnames were associated to regions other than Europe or Anglophone jurisdictions for the Fragile Nucleosome series compared to 58% for DASL.

We next compared direct lookup of associated regions in Forebears to imputation using a recently published long short-term memory recurrent neural network trained on names and nationalities scraped from Wikipedia ("Wiki2019-LSTM") [7]. For most geographic regions, predictions agreed well in aggregate, but Anglophone names and names associated to Latin America, Spain, Portugal, and the Philippines were discrepant (Supplementary Figure 1b). We re-examined predictions for 14 speakers who publicly identified their nationality and found a 71% success rate for direct lookup versus 36% for LSTM predictions, mostly because Latin American speakers' names were often associated to Anglophone regions (Supplementary figure 1c). The overrepresentation of Anglophone names in Wikipedia may bias LSTM performance.

Supplementary Figure 1: Support for surname associations. a) Speaker name associations from the Fragile Nucleosome seminar series. Counts are tallied as in Figure 2d. b) Aggregate counts for associated regions using Wiki2019-LSTM (x-axis) and the Forebears database (y-axis) across all DASL speakers. c) Confusion matrices for international DASL speaker names (19 surnames from 14 speakers) demonstrating accuracy of predictions for speaker nationality. Asterisks denote correct identifications. Names for which Wiki2019-STLM achieved less than 50% probability for a region were labeled as “Unidentified”.

DASL speakers' priorities for equity, diversity, and inclusion encompass both broad and narrow concerns

DASL speakers covered a broad range of topics: childhood familiarity with a career in STEM, the burden of fees in graduate admissions, immigrant identity, coming and being out as queer or trans in academia, navigating parenthood in STEM, mental health challenges in academia, health disparities for racial minorities, advocacy for people with disabilities, the complexity of mixed-race identity, cultural expectations clashing with career pressure, ethical research for indigenous communities, anti-science attitudes in rural America, and both subtle and overt racism against Black people and people of Middle East and North Africa descent.

By annotating the occurrences of keywords mentioned by each DASL speaker, we tracked common themes across talks (Figure 2e). Some topics, usually related to social factors, were broadly discussed. "Education" keywords were most frequently mentioned (46/54 talks), followed by "Family" keywords (39/54) and "Mentoring" keywords (37/54). Other topics were mentioned many times by a small number of speakers: "Sexual and gender minorities", "Disability", "Mental health", and "Rural & agricultural". The subject of race was intermediate: speakers broke roughly evenly between zero mentions, one or two mentions, or many mentions.

Five pairs of keyword groups co-occurred in the same speakers' presentations more often than expected by chance (Fisher's Exact Test, 20% FDR). The most significant co-occurring keywords by far were "Latino/Hispanic" and "Family" ($p = 3e\text{-}4$): 100% of talks mentioning "Latino/Hispanic" terms also mentioned family, linking discussion of Latino and Hispanic identity to family values.

"Latino/Hispanic" also co-occurred with "Women" (OR = 5.0; $p = 9e\text{-}3$), suggesting a special focus on Latina identity. "Women" co-occurred with "Men" (OR = 8.2; $p = 1e\text{-}2$), and "First-generation" co-occurred with "Family" (OR = 7.3; $p = 1e\text{-}2$) and "Men" (OR = 7.4; $p = 1e\text{-}2$), suggesting that speakers relate concepts of family and gender to narrower identities.

Flexible attendance options and small group sessions enhance stakeholders' virtual experience

Past work has suggested that remote formats have achieved greater equity and inclusion for junior career scientists in diverse cultural and institutional settings [8,9]. Similarly, in DASL, speakers were able to network with other scientists and receive feedback on their presentation skills; attendees had the opportunity to join in frank conversations about diversity and science; and executive team members tested their leadership capabilities while hosting.

We found that holding dry-runs for DASL sessions proved particularly valuable. The ability to troubleshoot technological issues in advance remained helpful throughout the year, and speakers who conversed with executive team members produced perceptibly more polished lectures and unanimously approved of the dry-run format when surveyed. Furthermore, the small group setting allowed in-depth conversations that facilitated durable relationships and knowledge sharing among participants. By comparison, meet-the-speaker sessions scheduled after formal sessions yielded mixed results.

With distance eliminated as a barrier, the pool of potential audience members increased dramatically. Beyond purely geographic constraints, the remote format allowed attendees to participate as much or as little as they wish from a location they choose without drawing attention to themselves. Scientists apprehensive about a new lecture series, especially one regarding diversity in science, had the option to attend from their bench or desk.

Building community around remote lecture series offers additional benefits to trainees

We have curated numerous resources that DASL speakers highlighted as helpful to their scientific careers. The entries, numbering more than 75, include opportunities for fellowships and research support, community building initiatives, mentorship groups, and professional

societies [10]. We noticed that mentoring and advising entries tend to be local entities, whereas funding opportunities tend to be from national organizations, hinting at differences in impact depending on organizational breadth.

In October 2020, we held our inaugural Diversity and Science Symposium virtually, featuring 20 speakers and 4 panels with over 1000 registered attendees from academic and industrial institutions across the country and concluded our second annual symposium in November 2021.

Conclusions

Embrace the advantages of remote formats

Prior to the COVID-19 pandemic, cross-institutional and inter-department lecture series in San Diego were made inconvenient by long transit times and difficulty navigating auditorium capacities. While some in-person activities have resumed at American universities, remote lecture series remain an attractive way to engage scientists nationally and internationally, especially amidst the ongoing uncertainty surrounding the evolution of the COVID-19 pandemic. Whereas traditional seminar series aim to appeal to the interests of a building or department, remote lecture series can address topics that are either very narrow or, crucially, very broad.

Remote lecture series can attain greater participation and more networking opportunities for trainee scientists relative to comparable in-person meetings. Moreover, these benefits do not impose additional organizing challenges: auditorium availability is a non-issue, the effort needed to prepare for 20 or 100 attendees on Zoom is equivalent, and there is no additional cost or technical support needed to record or share talks. Indeed, trainees can administer operations on their own.

Evaluate lecture series goals for all participants to adapt over time

Attention is a fiercely contested resource in professional settings. Lectures must be informative, entertaining, and topical to compete successfully for participants' time and attention. In the long term, speakers must feel heard and respected; attendees must highly esteem lecture content; and executive team members must feel intellectually engaged. Re-evaluate lecture frequency periodically — demand for speaking, attending, and organizing will wane over time.

Collecting feedback from participants is more difficult in a remote setting, but no less essential. Conducting exit surveys and tracking speaker permission for sharing talk content is recommended. Seek feedback from all participants and share it with executive team members without diminishing their ability to exercise independent judgment.

The skills that executive team members will acquire from organizing a particular format are ultimately limited. Waning interest can be combated by passing responsibility from senior team members to new recruits or compensating team members for their time to sustain organizing efforts.

Be mindful of tools to assess diversity

The first step to evaluating organizational diversity is surveying representation of relevant identities. Some identities, such as a nationality, geography, and gender are publicly searchable for most senior academics. Alternatively, programmatic and semi-automated onomastic methods have been developed to assess race, nationality, gender, and name origin [6,7,11]. The ability to eliminate response bias, ensure participant privacy, and mine datasets without self identification is an advantage that must be weighed against uncertain or inaccurate identifications for specific subpopulations. Furthermore, other identities, such as class, rural or agricultural background, first generation status, and sexual and gender minorities are inaccessible without self identification.

As with data curation, interpretation of demographic data requires care. University and professional society websites often publish demographic data useful for baseline comparisons. Tracking data over time can also indicate troubling trends that require intervention. Disparities for women [12–14], racial minorities [7,15,16], and sexual and gender minorities [17,18] in scientific disciplines have been well documented and can be assessed in a few questions. Also consider using disaggregated data for both racial and sexual and gender minority categories: “South Asian” instead of “Asian” and “Transgender” instead of “LGBT”.

Consistent and direct communication is key to a successful lecture series

No matter what level of quality a lecture series attains, one-on-one networking is the most effective way to recruit new speakers, attendees, and executive team members. To reach a wider pool of contacts, pursue existing trusted communications media: departmental email lists or institutional social media accounts. New communications medium will likely yield little engagement until the organizers amass meaningful momentum.

Regardless of the medium chosen, state succinctly what content to expect and how to access it. Speakers, attendees and executive team members should be able to complete a simple task to join or decline with as few clicks and textboxes as possible.

Do not fear sending polite reminders: trust that your efforts merit the recipients’ attention. Judicious messaging is an essential part of a successful remote lecture series.

Methods

Representation of women was computed from UCSD diversity dashboards. Graduate students in the Health Sciences and Biological Sciences who chose “Woman” as their gender were counted. For postdocs, academic personnel with the appointment title “Postdoctoral scholar/fellow” in the Health Sciences and Biological Sciences were counted. For faculty, the

“ladder-rank professor” appointment title was selected. Data was taken from fall 2019, the most recent data available.

Terms in talk titles were counted by parsing titles posted on the DASL website. Text was tokenized by tidytext::unnest_tokens and stop words removed using the tidytext’s stop_words dataset. Words were stemmed using SnowballC::wordstem, and the number of titles containing each stemmed word were counted. The frequency of retained stems that occurred in at least 3 titles were converted back into representative words and visualized using the wordcloud package.

Speaker surnames were noted from the DASL website and Fragile Nucleosome website over the same time period. Surnames were associated to geographic regions using Forebears.io, the “largest geospatial names database”, which purportedly aggregates records from over 27 million surnames and 4 billion individual records from 236 countries or jurisdictions [19]. Jurisdictions in Latin America were pooled into one region. Jurisdictions in Britain, the United States, Canada, Australia, New Zealand and the Caribbean were pooled into an Anglophone region because of their shared British surnames [20,21]. South Asian jurisdictions were pooled into a South Asian region. For each surname, the region with the highest incidence and frequency were noted. Names from speakers with two surnames were weighted accordingly. For surnames that matched multiple regions, the most frequent region only applied when the incidence was above 500.

Surnames were considered validated 1) if the most prevalent region was also the most frequent, 2) if the most prevalent region was at least tenfold greater than the most frequent region (most prevalent region chosen), 3) if the most prevalent region was the Anglophone region and less than tenfold greater than the most frequent region (most frequent region chosen). If these conditions were not met, or there was no match to the database, the surname was deemed unidentified. Surnames assigned to the Anglophone region were further annotated with race using the predictrace::predict_race command in R. Surnames assigned high

confidence for matching White US Census takers were labeled “(White)”. Surnames that were not high confidence for corresponding to White US Census takers were labeled “(Nonwhite)”. Note that these methods are not sensitive for Native American and Alaskan Native or multiracial identities reported to the US Census [22].

The Wiki2019-LSTM model was downloaded from its GitHub repository (https://github.com/greenelab/wiki-nationality-estimate). The 07.test-ismb-data.py script was modified to issue predictions on DASL speaker surnames. For aggregate counts, fractional assignments were summed across all speakers. For visualization of the confusion matrix, surnames that had no region above 50% probability were labeled as “Unidentified”. For the sake of comparison with Forebears lookup, the Wiki2019-LSTM’s Celtic/English category was deemed synonymous with our Anglophone category, and the Wiki2019-LSTM’s Israel category was pooled with our Middle East and North Africa category.

Transcripts from publicly viewable DASL talks were reviewed and keywords tallied in R. Exact keyword expressions are available in the Figure 2 supplementary file. Column order was determined by running the R command hclust on Pearson correlation distance for a matrix of the log10 of the count plus 1 for each keyword across 54 talks. The number of mentions of each keyword were truncated at 3 or more for data visualization purposes. Significantly co-occurring terms were determined using the R command fisher.test on the number of joint mentions, single mentions for each keyword and non-mentions for all pairs of keywords that appeared in at least ten talks. The false discovery rate was determined using the p.adjust command.

Icons for the DASL organizing schematic were downloaded from The Noun Project (nounproject.com). Lightbulb by Maxim Kulikov, Rules by Adrien Coquet, Webinar by ProSymbols, Whisper by ProSymbols, and Toolbox by WEBTECHOPS LLP.

Figure legends

Figure 1: Organizing the first year of DASL. a) Timeline of DASL milestones separated by quarter. DASL was founded in June 2020 and completed its recent spring series in June 2021. b) Outline of running a remote lecture series. We provide guides detailing how to approach each underlined component (Supplementary Text). c) Total subscribers to the DASL mailing list and d) number of attendees per DASL session over the first year of programming.

Figure 2: Reviewing the first year of DASL. a) Breakdown of the positions held by DASL's first 79 speakers. The 'Other' category includes Research scientist, Research Assistant Professor, Professor, Assistant Curator, and Administrator. b) Representation of women among DASL speakers compared to UCSD life science graduate students, postdoctoral scholars, and faculty. c) Most common terms used in DASL talk titles. The size of the word reflects the frequency: 'cell' occurred 9 times. d) Counts of most associated regions for DASL speakers' surnames using the geospatial name database Forebears. 'Anglophone' refers to names associated to Britain, the United States, Canada, Australia, New Zealand or a country in the Caribbean. e) Truncated counts of mentions of keywords across speakers (x-axis) organized by topic (y-axis) and broader type of concern (social, race/ethnicity, or other identities).

Supplementary Figure 1: Support for surname associations. a) Speaker name associations from the Fragile Nucleosome seminar series. Counts are tallied as in Figure 2d. b) Aggregate counts for associated regions using Wiki2019-LSTM (x-axis) and the Forebears database (y-axis) across all DASL speakers. c) Confusion matrices for international DASL speaker names (19 surnames from 14 speakers) demonstrating accuracy of predictions for speaker nationality. Asterisks denote correct identifications. Names for which Wiki2019-STLM achieved less than 50% probability for a region were labeled as "Unidentified".

Acknowledgements

DASL is funded in part by a grant from the Chan-Zuckerberg Initiative. E.A.B. acknowledges funding from the Helen Hay Whitney Foundation. G.W.Y. is an Allen Distinguished Investigator, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation. The 2020 DASL symposium was supported by HHMI Gilliam Fellows, the ThermoFisher Queer Working Group, the Huntsman Cancer Institute, the UCSD Biology Student Social Fund, UCSD Graduate Division, UCSD Biomedical Sciences Graduate Program, the UCSD Biology Diversity Committee, the UCSD Office for Equity, Diversity, and Inclusion, the UCSD Latinx/Chicanx Academic Excellence Initiative, the UCSD Black Academic Excellence Initiative, and the UCSD School of Medicine.

Supplementary materials

Supplementary Text: How to launch, execute, maintain, and administer a remote lecture series

Supplementary Data: Example forms from DASL

References

1. Gaëta BA, De Las Rivas J, Horton P, Meysman P, Mulder N, Romano P, et al. Ten simple rules for forming a scientific professional society. PLoS Comput Biol. 2017;13: e1005226.

2. Fadlelmola FM, Panji S, Ahmed AE, Ghouila A, Akurugu WA, Domelevo Entfellner J-B, et al. Ten simple rules for organizing a webinar series. PLoS Comput Biol. 2019;15: e1006671.

3. Bozelos PA, Vogels TP. Making (neuro) science accessible world-wide: Online seminars for the globe. Elife. 2020. Available: https://elifesciences.org/labs/d8f1d697/making-neuro-science-accessible-world-wide-online-seminars-for-the-globe

4. Bottanelli F, Cadot B, Campelo F, Curran S, Davidson PM, Dey G, et al. Science during lockdown - from virtual seminars to sustainable online communities. J Cell Sci. 2020;133. doi:10.1242/jcs.249607

5. Cucinotta CE, Martin BJE, Noé González M, Raman P, Teif VB, Vlaming H. Strength is in engagement: The rise of an online scientific community during the COVID-19 pandemic. EMBO Rep. 2021;22: e52612.

6. Tzioumis K. Demographic aspects of first names. Sci Data. 2018;5: 180025.

7. Le TT, Himmelstein DS, Hippen AA, Gazzara MR, Greene CS. Analysis of scientific society honors reveals disparities. Cell Syst. 2021;12: 900–906.e5.

8. Aguado BA, Porras AM. Building a virtual community to support and celebrate the success of Latinx scientists. Nat Rev Mater. 2020; 1–3.

9. Sarabipour S. Research Culture: Virtual conferences raise standards for accessibility and interactions. Elife. 2020;9: e62668.

10. DASL. Speakers' Highlighted Resources. In: DASL [Internet]. 2 Apr 2021 [cited 16 Oct 2021]. Available: https://www.ucsddasl.com/resources

11. Santamaría L, Mihaljević H. Comparison and benchmark of name-to-gender inference services. PeerJ Comput Sci. 2018;4: e156.

12. Cech EA, Blair-Loy M. The changing career trajectories of new parents in STEM. Proc Natl Acad Sci U S A. 2019;116: 4182–4187.

13. Huang J, Gates AJ, Sinatra R, Barabási A-L. Historical comparison of gender inequality in scientific careers across countries and disciplines. Proc Natl Acad Sci U S A. 2020;117: 4609–4616.

14. Holman L, Stuart-Fox D, Hauser CE. The gender gap in science: How long until women are equally represented? PLoS Biol. 2018;16: e2004956.

15. Ginther DK, Schaffer WT, Schnell J, Masimore B, Liu F, Haak LL, et al. Race, ethnicity, and NIH research awards. Science. 2011;333: 1015–1019.

16. Erosheva EA, Grant S, Chen M-C, Lindner MD, Nakamura RK, Lee CJ. NIH peer review: Criterion scores completely account for racial disparities in overall impact scores. Sci Adv. 2020;6: eaaz4868.

17. Cech EA, Waidzunas TJ. Systemic inequalities for LGBTQ professionals in STEM. Sci Adv. 2021;7. doi:10.1126/sciadv.abe0933

18. Freeman JB. Measuring and Resolving LGBTQ Disparities in STEM. Policy Insights from the Behavioral and Brain Sciences. 2020;7: 141–148.

19. Forebears. Forebears Names. In: Forebears [Internet]. 1 Sep 2019 [cited 16 Oct 2021]. Available: https://forebears.io/about/name-distribution-and-demographics

20. Elliott MN, Morrison PA, Fremont A, McCaffrey DF, Pantoja P, Lurie N. Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities. Health Services and Outcomes Research Methodology. 2009. pp. 69–83. doi:10.1007/s10742-009-0047-1

21. O'Boyle B. Tracing ancestry is not easy for African Americans. In: Times Leader [Internet].

16 Dec 2015 [cited 16 Oct 2021]. Available: https://www.timesleader.com/news/local/499109/tracing-ancestry-is-not-easy-for-african-americans

22. Allen Fremont, Joel S. Weissman, Emily Hoch, and Marc N. Elliott. When Race/Ethnicity Data Are Lacking. In: RAND Corporation [Internet]. 2016 [cited 16 Oct 2021]. Available: https://www.rand.org/pubs/periodicals/health-quarterly/issues/v6/n1/16.html