bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 testing institutional commitment to the principles of equity, diversity, and inclusion. Here we present thorough guidelines for drawing out hope from the Pandora's box of virtual programming now open to the world. We review milestones from our first year organizing the Diversity and Science Lecture series (DASL) and explore insights into equity, diversity, and inclusion in STEM gleaned from hosted speakers' talk content. Nearly every speaker highlighted the importance of social or interpersonal support to

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

their career progression, and three-fifths of speakers commented on race or ethnicity. Other recurring topics each received attention from a minority of speakers: immigrant identity, gender identity, mental health, sexual minorities, disability, and rural or agricultural background. We conclude with generalizable advice on creating new remote lecture series that benefit executive team members, speakers, and attendees. Our success with DASL demonstrates that community building and knowledge sharing 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 could 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 the academic policies that helped or hindered their scientific careers.

In summer 2021 we concluded our first year of programming (Figure 1a). What once started as a community at UC San Diego grew to include scientists from several institutions nationwide. In the end, our weekly seminar series featured 79 speakers from 14 institutions. Participating institutions outside San Diego have since begun hosting remote lecture series of their own.

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, impactful conversation amongst graduate students, postdoctoral scholars, and faculty who had never met in person. Nonetheless, DASL continuously grew its subscriber base (Figure 1b) and accrued roughly 3,500 hours of engagement from attendees (Figure 1c).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Unlike DASL, many prominent virtual seminar series launched in 2020 hewed close to existing scientific communities [1–3]. 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? Here, we analyze the slate of speakers from the first year of DASL. The results affirm the potential of remote formats for greater equity and inclusion for junior career scientists and minority groups [4,5]. The guidelines herein serve as a roadmap that can light the way for future remote lecture series aimed at new, diverse audiences.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

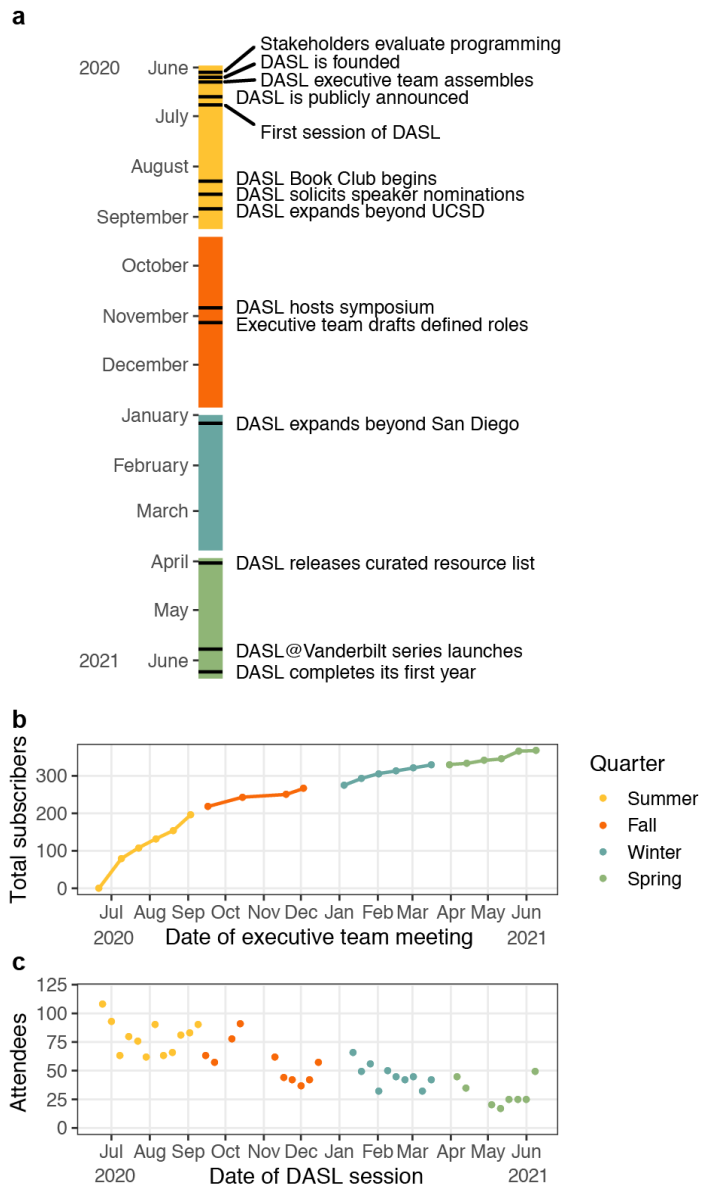

Figure 1: Organizing a year of DASL. a) Timeline of DASL milestones separated by quarter. DASL was founded in June 2020 and completed its recent spring series in June 2021. b) Total subscribers to the DASL mailing list and c) number of attendees per DASL session over the first year of programming.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

Who speaks at DASL?

We recruited speakers from a variety of institutions and academic positions. Graduate students filled the majority of speaker slots (42/79), followed by postdoctoral scholars (27/79) and assistant professors (5/79) (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 (18/31).

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).

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. For most names, prevalence and abundance pointed to the same region, enabling inference of where presumed relatives of speakers currently reside. (Figure 2d). Names linked to Britain, the United States, Canada, Australia, New Zealand and the Caribbean were grouped in an 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 speaker surnames "Nonwhite."

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

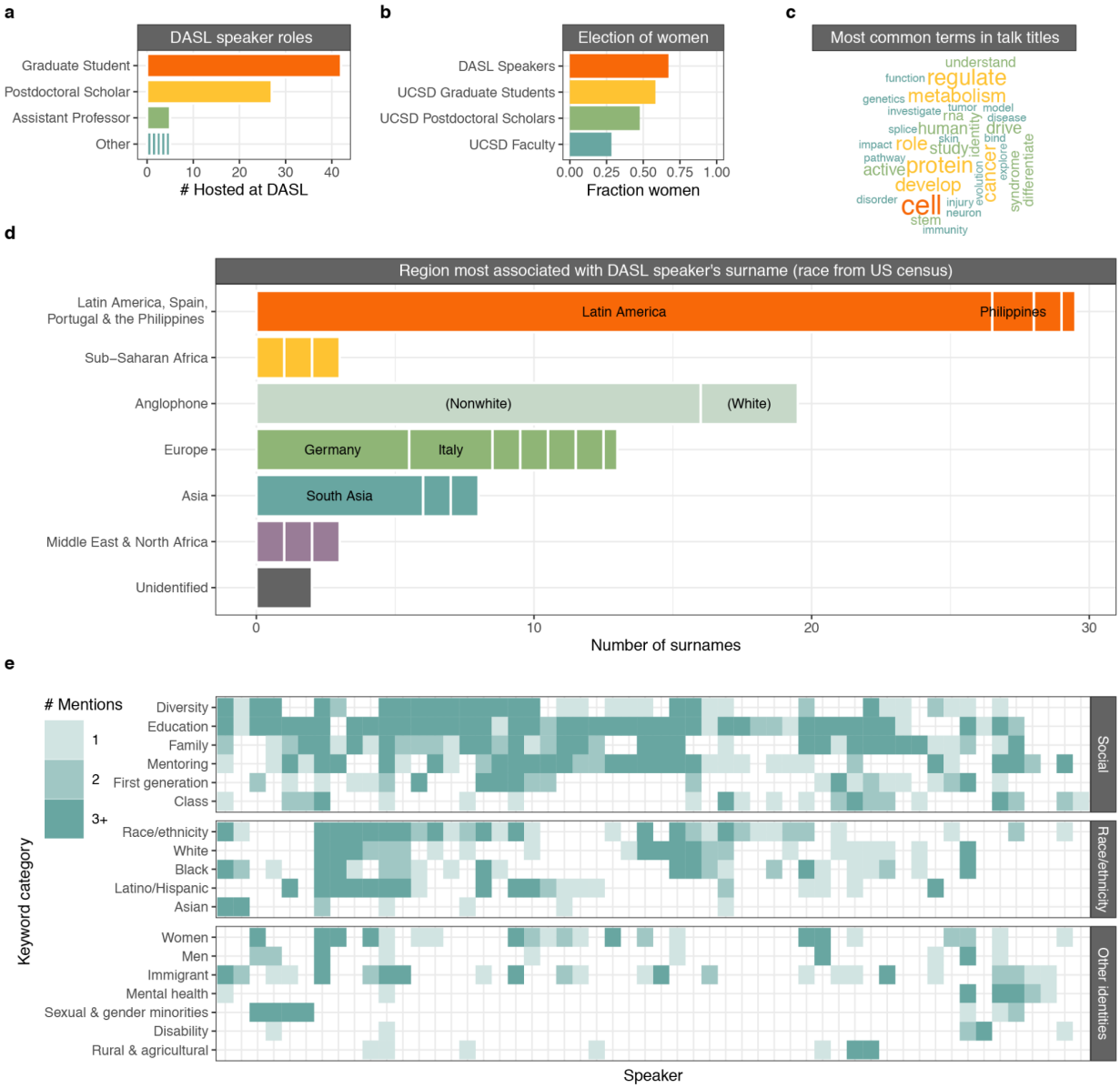

Figure 2: Reviewing one year of DASL. a) Breakdown of the positions held by DASL's first 79 speakers. The 'Other' category includes Research scientist, Research Assistant Professor, Professor, Assistant Curator, and Administrator. b) Representation of women among DASL speakers compared to UCSD life science 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

What equity, diversity, and inclusion-related topics do DASL speakers address?

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 anti-Black racism.

By annotating the occurrences of keywords in each DASL session, we tracked common themes across DASL speakers (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” ($p = 3e-4$): 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” ($OR = 5.0$; $p = 9e-3$), suggesting a special focus on Latina identity. “Women” co-occurred with “Men” ($OR = 8.2$; $p = 1e-2$), and “First-generation” co-occurred with “Family” ($OR = 7.3$; $p = 1e-2$) and “Men” ($OR = 7.4$; $p = 1e-2$), suggesting that speakers relate concepts of family and gender to narrower identities.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Takeaways from DASL's weekly lecture series

DASL has taught us that remote lecture series are uniquely suited to foster community in diverse cultural and institutional settings. Speakers are able to network with other scientists and receive feedback on their presentation skills; attendees have the opportunity to join in frank conversations about diversity and science; and executive team members can test their leadership capabilities as hosts to speakers.

Holding dry-runs for DASL sessions proved particularly valuable. In addition to making time to troubleshoot any technical considerations for the virtual meeting platform, we offered every trainee speaker the opportunity to present their talk to 2-4 executive team members and receive general or specific feedback. The ability to troubleshoot technological issues remained helpful throughout the year, and speakers who conversed with executive team members produced perceptibly more polished lectures. Furthermore, the small group setting led to more durable relationships and knowledge sharing among participants.

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

Other DASL accomplishments

We have curated numerous resources that DASL speakers highlighted as helpful to their scientific careers. The more than 75 entries include opportunities for fellowships and research support, community building initiatives, mentorship groups, and professional societies [7]. We noticed that mentoring and advising entries tend to be local entities, whereas funding

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. We concluded our second annual symposium in November 2021.

Finally, we have compiled guides on how to launch, execute, maintain, and administer remote lecture series (Supplementary Text) with example forms (Supplementary Data).

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 in-person activities have largely resumed at American universities, remote lecture series remain an attractive way to engage scientists nationally and internationally. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

Collecting feedback from participants is more difficult in a remote setting, but no less essential. 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 are ultimately limited: consider passing responsibility from senior team members to new recruits or paying team members for their time to sustain organizing efforts.

Consistent and direct communication is key

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. With judicious messaging, a successful remote lecture series is attainable.

Methods

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 surnames were associated to 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 [8]. 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 [9,10]. South Asian jurisdictions were pooled into a South Asian region. For each surname, the region with the highest incidence and frequency were noted. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 [11].

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.

Figure legends

Figure 1: Organizing a 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) Total subscribers to the DASL mailing list and c) number of attendees per DASL session over the first year of programming.

Figure 2: Reviewing one 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

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. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.05.471284; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2. 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

3. 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.

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

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

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

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

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

9. 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

10. 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

11. 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