A Comprehensive Bibliometric Analysis: Celebrating the Thirtieth Anniversary of the Pacific Symposium on Biocomputing

Rachit Kumar[1], Rasika Venkatesh[1], David Y. Zhang[1], Teri E. Klein[2], Marylyn D. Ritchie[3†]
[1]*Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA, USA*
[2]*Department of Biomedical Data Science, Medicine (BMIR) and Genetics (by courtesy), Stanford University, Stanford, CA, USA*
[3]*Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA*
Email: marylyn@pennmedicine.upenn.edu

The 2025 Pacific Symposium on Biocomputing (PSB) represents a remarkable milestone, as it is the thirtieth anniversary of PSB. We use this opportunity to analyze the bibliometric output of 30 years of PSB publications in a wide range of analyses with a focus on various eras that represent important disruptive breakpoints in the field of bioinformatics and biocomputing. These include an analysis of paper topics and keywords, flight emissions produced by travel to PSB by authors, citation and co-authorship networks and metrics, and a broad assessment of diversity and representation in PSB authors. We use the results of these analyses to identify insights that we can carry forward to the upcoming decades of PSB.

Keywords: Bibliometry; PSB Proceedings.

1. Introduction

1.1. *Overview*

The Pacific Symposium on Biocomputing (PSB) is an international conference where presentation and discussion of current research in the theory and application of computational methods in problems of biological significance take place. PSB has been held annually since 1996. PSB 2025 marks the 30th anniversary of this conference, a milestone that represents a critical opportunity to evaluate the impact the conference has had on the field of biocomputing, including the scientists in the field, and to find opportunities for growth for the future of PSB and other related conferences. This project was inspired by a similar initiative in 2015 that sought to commemorate the 20th anniversary of PSB[1].

We make use of bibliometric data available on all PSB proceedings from 1996 to 2024, which includes 1402 published papers, all of which are indexed on PubMed. Using these citations, we performed a variety of analyses, each focused on a different perspective or lens by which we reviewed the data. These analyses were inspired by the 20th anniversary review of the PSB proceedings[1] as well as some of the session topics for the PSB 2025 conference, as both a framework for the methodology as well as the topic of many of these analyses.

The session topics for PSB 2025 include:

- AI and Machine Learning in Clinical Medicine
- Earth Friendly Computation
- Precision Medicine: Multi-modal and multi-scale methods

† Corresponding author.

- Translating Big Data Imaging Genomics Findings to the Individual
- Overcoming health disparities in precision medicine

In the following sections, we discuss how we took inspiration from some of these session topics to formulate and perform a variety of analyses on the bibliometric data available from all previous PSB publications.

1.2. ***Session Topics and Mapping to Analyses***

1.2.1. *AI and Machine Learning in Clinical Medicine*

For this session topic, we used keyword analysis to explore how the topics of PSB publications have shifted from year to year along certain breakpoints or eras of PSB. This is described in more detail as the "Keywords and Topics Analysis" in Methods. It is worth noting that many of our other analyses as described below also indirectly make use of machine learning models.

1.2.2. *Earth Friendly Computation*

For this session topic, we sought to estimate the environmental impact of PSB in the form of carbon emissions, given that PSB is a conference that has been held in Hawaii every year (with the sole exception of PSB 2021, which was held virtually due to the COVID-19 pandemic). This is described as the "Emissions Analysis".

1.2.3. *Precision Medicine: Multi-modal and multi-scale methods*

For this session topic, we took inspiration from the idea of multiscale analyses in other fields such as social network analysis and applied them to citation analysis, where we look at the relationships of PSB papers and authors over time. We present this information in the form of a multimodal network that includes papers and authors as well as a co-authorship network. We further perform more traditional citation analyses. This is described as the "Citation and Authorship Analysis".

1.2.4. *Overcoming health disparities in precision medicine*

For this session topic, we wanted to perform analyses that are parallel to the growing understanding that health disparities are critical to acknowledge and address in precision medicine, specifically in the form of acknowledging the value of diverse perspectives in science and scientific discourse. Specifically, we explored the diversity of authors on PSB proceedings papers by exploring changes in representation and diversity along the axes of race and gender. This is described as the "Diversity Analysis".

2. Methods

2.1. ***Common Methods***

We acquired from PubMed the initial list of 1402 papers by using the search term " "Pac Symp Biocomput"[jour] " (outer quotes not included) and then exported a CSV of the results, which

contained information on paper titles, publication year, PMID, and authors (first initial and last names only). We spot-checked randomly-selected papers in this list against the online published conference proceedings to confirm concordance of papers and authors[2].

We then used the NCBI Entrez tools via Biopython[3] with each of the paper PMIDs to further acquire additional information from PubMed on citation PMIDs (only those to or from papers indexed in PubMed), author affiliations, paper abstract text, and full author names (where available). Notably, for papers from 1996-2004, a substantial number of authors only had first initials available.

Subsequently, we acquired dimensions.ai[4] data on PSB papers as of June 2024 to acquire overall citation counts, recent citation counts, and a machine-learning-based determination of the Australian and New Zealand Standard Research Classification 2020 Fields of Research (ANZSRC 2020 FoR)[5], which includes a hierarchical system that identifies broad categories such as "Health Sciences" (category #42) as well as more nuanced categories such as "Machine Learning" (category #4611). Papers can be assigned to multiple categories. Dimensions contained relevant information for 1367 papers out of 1402.

The analyses performed with this data are summarized in visual form in Figure 1 below and are further described in the subsequent sections.

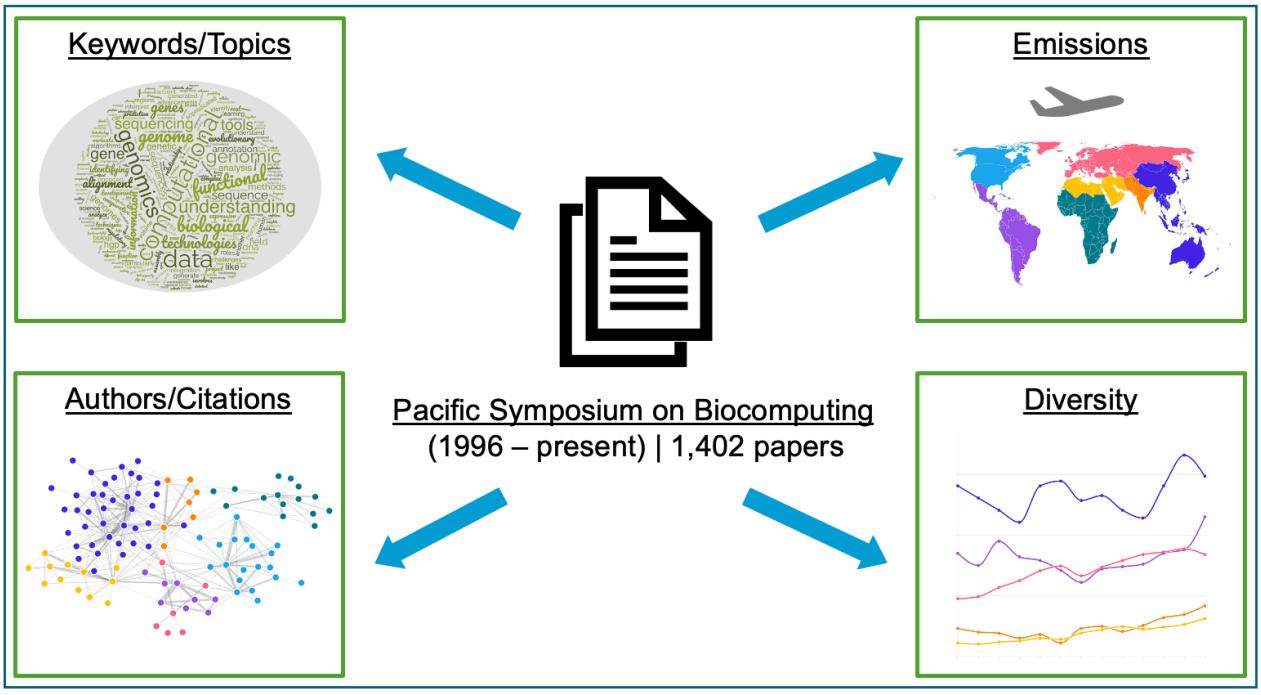

Figure 1: A graphical summary of the bibliometric analyses performed in this paper on PSB papers from 1996 to 2024, including: evaluation of keywords and topics; estimation of conference attendance emission costs; generation of author and citation co-networks; and review of author diversity.

2.2. *Keyword and Topics Analysis*

We used the assigned dimensions.ai ANZSRC 2020 FoR classifications as described in the Common Methods section as overarching paper topics. Additionally, we used KeyBERT, a tool that uses deep neural networks in the form of transformers[6]. KeyBERT generates BERT embeddings of papers and

keywords, and it then subsequently identifies the most relevant keywords for each paper. Specifically, we provide the abstract of each paper, and keywords are extracted with KeyBERT using the underlying "all-MiniLM-L6-v2" sentence transformer model. Each paper was given five keywords in this fashion, with an additional constraint added to try to make the keywords as distinct as possible using KeyBERT's MMR diversity parameter with a value of 0.7.

To interpret this information, we subdivided the papers in PSB into "eras" relating to various important milestones in the field as breakpoints:

- 1996-2003 (before the completion of the Human Genome Project[7])
- 2004-2016 (before the spike in popularity of deep learning in biomedicine, particularly transformers and LLMs)
- 2017-2024 (during the current era of an "AI" boom in biomedicine, including breakthroughs such as AlphaFold[8])

For each topic, the proportion of papers in each era as described above assigned to that topic was computed, and a bar plot was made showing the proportions for each era. Relevant topics were selected for presentation based on which topics were most nonredundant and had a critical number of papers assigned to them.

Separately, for each era, the keywords for papers in that era were lemmatized to combine singular and plural versions of the same word and then collated together to produce a word cloud using the `word_cloud` Python package[9] for each of the eras as a visualization of the most relevant keywords as aggregated across papers for each era.

2.3. *Emissions Analysis*

Using the PubMed information on author affiliations, we performed an analysis of the CO_2 equivalents that were produced as a result of flying to PSB. Specifically, we used the affiliations of the first authors of every paper and used the Google Maps Geocoding API[10] to programmatically and automatically identify the most likely latitude and longitude for each affiliation. We then used a set of data from OurAirports, an open-source and curated repository of airports around the world and their latitude and longitudes[11], to map each individual to their nearest "medium-sized" or "large-sized" airport (observing that "small-sized" airports tended to be regional or private airports) by calculating the Haversine distance[12] of each airport to each affiliation and identifying the closest such airport for each affiliation.

Once each affiliation was mapped to an airport, the Haversine distance of those airports to the Hawaii Kona airport (KOA) was computed to get a putative shortest-path flight distance. This distance was then multiplied by a constant scale factor of 0.148 kg CO_2e per passenger-kilometer (as reported by the UK Government's Department for Energy Security and Net Zero) to compute the carbon emissions of each flight (matched to each paper)[13,14].

In this analysis, we made several assumptions, some of which we recognize as unlikely (see discussion below): only the first authors fly to PSB (and they travel alone), authors fly from their closest (mapped) airport to their reported affiliation, all flights are direct to the KOA airport, all flights in the past have the same carbon efficiency as flights today, all flights take the shortest possible path according to the Haversine distance between airports, and the radius of the Earth is

generally constant at 6371 km for the purposes of computing the Haversine distance (modeling the Earth as a perfect sphere). Notably, PSB 2021 was online due to the COVID-19 pandemic, so emissions for that year were artificially zeroed out after the calculations above.

2.4. *Citation and Authorship Analysis*

Using the PubMed and dimensions.ai-acquired information on all PSB proceedings papers, we computed a variety of statistics for each paper and author in PSB as well as PSB-wide statistics. Additionally, we created an interactive network of papers and authors as well as an interactive network of coauthors. The paper-author network has edges connecting authors to the papers that they have written as well as edges connecting papers that have cited each other and has interactive nodes that allow one to see various statistics for each paper and author. This includes metrics such as the number of citations or the keywords of a paper, and the first year an author published in PSB or the total number of papers an author has published in PSB.

The coauthor network is a multigraph of nodes representing every author at PSB and edges representing their co-authorship in the three different eras of PSB as mentioned in the "Keywords and Topics Analysis" Methods section above. We identified authors uniquely by using their first initial and their last name due to limitations in the data from 1996 to 2004 (where only first initials were available). We performed some simple network analyses on the co-authorship graph: we used Louvain community detection[15,16] to identify communities of co-authors in each era; we used PageRank[17] to identify the most "central" authors for each era; we computed the "density" of "co-authorship ties" defined as a proportion of the number of co-authorship links for each era over the total number of possible links for that era (based on the authors in that era).

2.5. *Diversity Analysis*

Using the PubMed information on author first and last names, we performed an analysis of the likely genders and races of all authors to assess how the diversity of PSB has changed over time. In each of the below analyses, authors were *not* deduplicated within a given year or between years, as we sought to assess the overall diversity of published authors in PSB. For the gender analysis specifically, we took inspiration from prior work by Teich et al.[18]; however, they used a paid API that has since changed methodology to determine genders. To make our methodology more reproducible and to minimize costs, we used United States Social Security Administration data instead.

To identify gender probabilities for each name, we used available data from the United States Social Security Administration (SSA) on first names for children from 1900 to 2023[19] and the assigned gender at birth of those children, calculating a ratio of male/female for each name across all of those years. The proportions of gender probabilities for each author's first name were averaged across years and plotted, with first names not being present in the SSA data (representing names that occurred less than 5 times in every year) being dropped. Notably, PubMed and the original PSB proceedings are missing information on authors' first names for all years up to 2004 (with 100% of authors missing first names in every year up to 2004 except for 2002 and 2003, which are each missing over 50% of first names). As such, all data from the years prior to 2005 were dropped.

To identify race probabilities for each full name, we used a Python package called ethnicolr2[20], which uses deep learning models (long short-term memory models) trained on first and last names from a database of Florida voters in the United States to predict the likelihood of each name belonging to someone identifying as one of five categories: "Asian", "Hispanic", "Non-Hispanic Black", "Non-Hispanic White", and "Other"[21]. The proportions of race probabilities for each author's full name were averaged across years and plotted.

3. Results

3.1. *Keyword and Topics Analysis*

Figure 2 shows three bar plots, one for each of the broad ANZSRC 2020 FoR topics of "Biological Sciences", "Biomedical and Clinical Sciences", and "Information and Computing Sciences". We can see that the proportion of papers tagged as "Biological Sciences" decreased from ~73% in the first two eras to 60.6% in the third, the proportion of papers tagged as "Biomedical and Clinical Sciences" increased era-over-era from 3% of papers in the first era to 13.1% of papers in the second era and 24% of papers in the third, and the proportion of papers tagged as "Information and Computing Sciences" is 27.9% in the first era, 22.1% in the second era, and 33.1% in the third era.

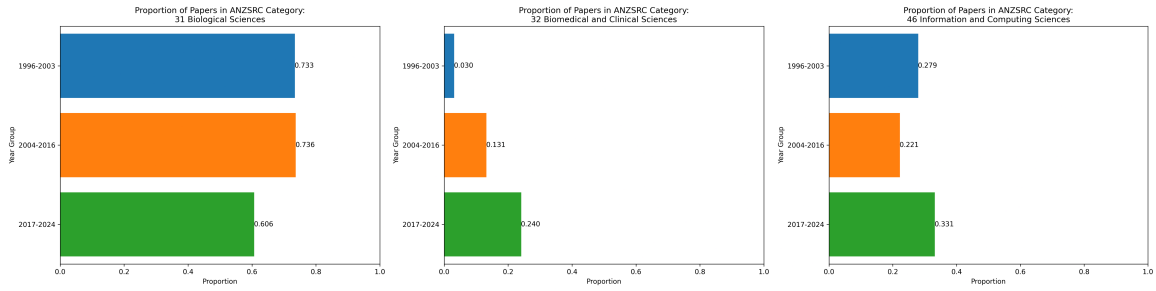

Figure 2: Proportion of papers in each of the three eras that were given the ANZSRC 2020 FoR broad categories of "Biological Sciences" (left), "Biomedical and Clinical Sciences" (middle), and "Information and Computing Sciences" (right).

Similarly to Figure 2, Figure 3 shows three bar plots for select ANZSRC subcategories - that is, categories one level lower than the broad topics as in Figure 2. The three categories shown are "Bioinformatics and Computational Biology", "Oncology and Carcinogenesis", and "Machine Learning", which are each (in order) a subcategory of the respective broad categories from above.

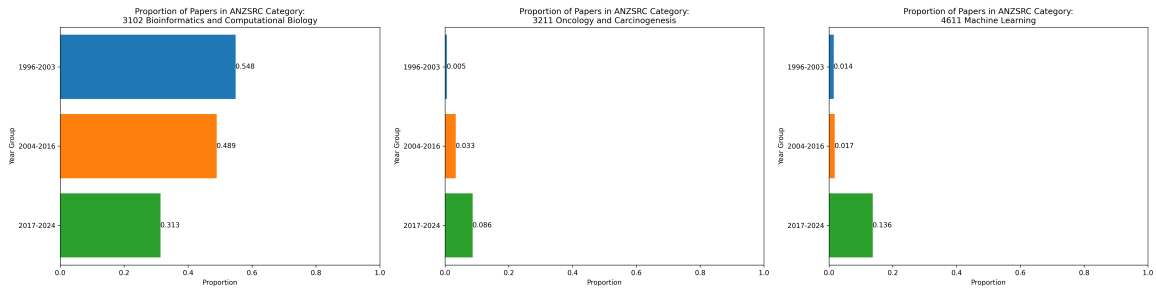

Figure 3: Proportion of papers in each of the three eras that were given the ANZSRC 2020 FoR subcategories of "Bioinformatics and Computational Biology" (left), "Oncology and Carcinogenesis" (middle), and "Machine Learning" (right).

Figure 4 shows the word clouds, one for each of the three eras. Notably, all three word clouds show many words related to genetics, genomics, and related topics with high prominence, such as "gene", "genomic", and "genome". Outside of these words, the first era word cloud shows a prominence of terms such as "alignment" and "sequence". The second era word cloud shows an increase in the prominence of "phenotype" and "annotation". The third era word cloud shows increased representation of research described by the words "neural" and "predicting". Also of note, the word "protein" was prominent in the first two word clouds, but significantly reduced in the most recent era.

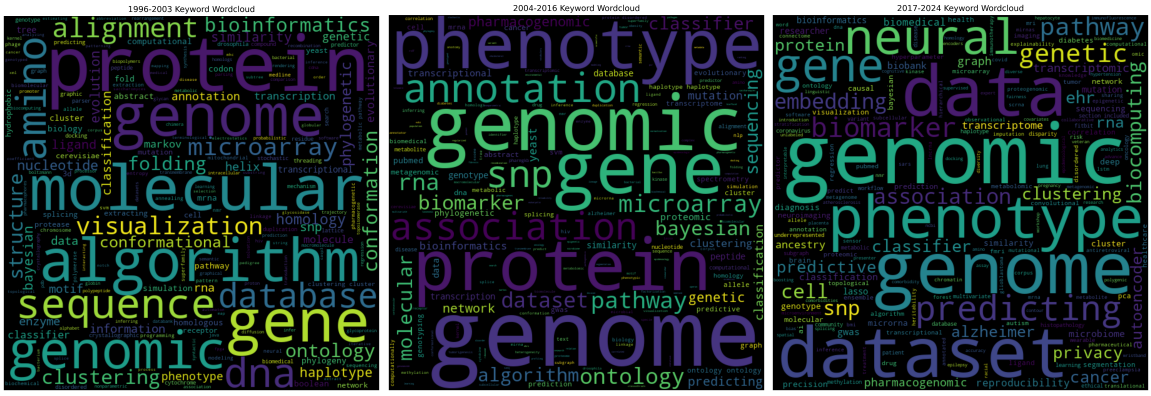

Figure 4: Word clouds for 1996-2003 (left), 2004-2016 (middle), and 2017-2024 (right).

3.2. *Emissions Analysis*

Given the assumptions and approach above, Figure 5 shows the estimated flight emissions for each year of PSB, showing the calculated average emissions per paper and the total emissions (across all papers). The total emissions for all 29 analyzed years of PSB was ~2,832,005 kg CO_2e (~2832 metric tons of CO_2e). Disregarding data from 2021 due to the online nature of PSB that year, this led to a computed average of ~2100 kg CO_2e per paper (~2.1 metric tons of CO_2e per paper) and an average of ~101 metric tons CO_2e per year of PSB.

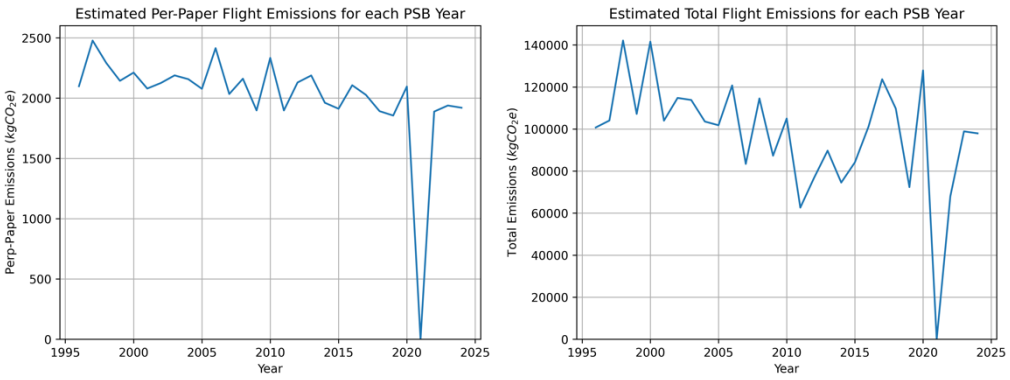

Figure 5: Emissions for each year of PSB from 1996 to 2024. (Left) The per-paper mean emissions for each year. (Right) The total emissions for each year. 2021 was held virtually due to the COVID-19 pandemic and as such had no estimated flight emissions.

3.3. *Citation and Authorship Analysis*

As of June 2024, 1367 papers out of 1402 in the PSB proceedings had citation information in dimensions.ai. The average number of citations across these papers was 20.91 (standard deviation 50.95; total 28579) with a median of 7.0 (max: 680). When normalizing by the number of years that a paper has been available up to 2025 (getting the number of times cited per year), the average is 1.42 citations/year, with a median of 0.54 citations/year. 1229 papers out of 1367 papers with citation data (~90%) published in PSB have been cited at least once.

Papers in PSB were cited, collectively, 3943 times in the past two years, with papers in the last decade receiving a larger proportion of those citations (Figure 6). PSB has, as of 2024, an h-index of 76 - that is, 76 papers have been published that received at least 76 citations. For papers in just the last 5 years (from 2020-2024), the corresponding h5-index is 13.

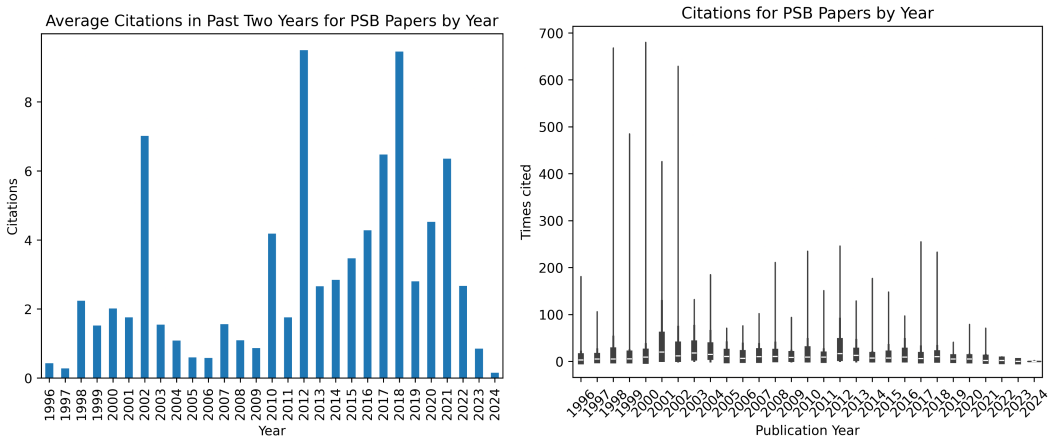

Figure 6: (Left) Average number of recent citations (in the past two years) for each paper per year of PSB from 1996 to 2024. (Right) Violin plot of the total number of citations for each year.

The interactive paper-author graph and the co-authorship graph are both available online here: https://ritchielab.org/publications/supplementary-data/psb-2025/psb-bibliometry. The best way to search for papers or authors is to select (1) “edge” (2) “from” (3) [PAPER/AUTHOR NAME] from the filter dropdowns, respectively.

From the co-authorship analysis done over eras, we find that the average size of communities of coauthors was 4.3 in 1996-2003, 7.8 in 2004-2016, and 12.2 in 2017-2024. For the same eras, the number of unique authors was 1065, 1815, and 1510, respectively. Across all years of PSB, we identified 4013 unique authors.

Based on PageRank centrality, the top 10 authors for each era are listed in Table 1:

Table 1. Top 10 authors for each era of PSB based on PageRank centrality of the co-authorship network.

1996-2003	2004-2016	2017-2024
Miyano, S	Altman, RB	Moore, JH
Altman, RB	Ritchie, MD	Crawford, DC
Takagi, T	Crawford, DC	Ritchie, MD
Hunter, L	Moore, JH	Zou, J
Dunker, AK	Cohen, KB	Tintle, N
Godzik, A	Liu, Y	Brenner, SE
Kohane, IS	Butte, AJ	Thompson, PM
Kitano, H	Chen, L	Chen, Y
Zimmer, R	Hartemink, AJ	Wall, DP
Huang, CC	Pendergrass, RA	Altman, RB

3.4. *Diversity Analysis*

Figure 7 has two line graphs, one for gender proportion and one for race and ethnicity proportions, for each year of PSB. For the gender analysis, ~32.5% of all listed authors across all years of PSB were estimated to be female. When taking the rolling mean of these proportions on a 5-year basis, we see that the earliest years of PSB of 1996-2000 had a gender proportion of ~25% while the most recent years of 2020-2024 had a gender proportion of ~35%.

For the race and ethnicity analysis, across all years of PSB, ~55.3% of all authors were estimated to be Non-Hispanic White, ~26.4% Asian, ~6.8% Non-Hispanic Black, ~5.3% Hispanic, and ~6.2% Other. When taking the rolling mean in a similar fashion to the gender analysis, we note the following changes (in the form of the mean proportion from 1996-2000 -> the mean proportion from 2020-2024): Non-Hispanic White ~62% -> ~51%; Asian ~22% -> ~30%; Non-Hispanic Black ~6.1% -> 6.5%; Hispanic ~3.4% -> 5.8%; and Other 5.7% -> 7.0%.

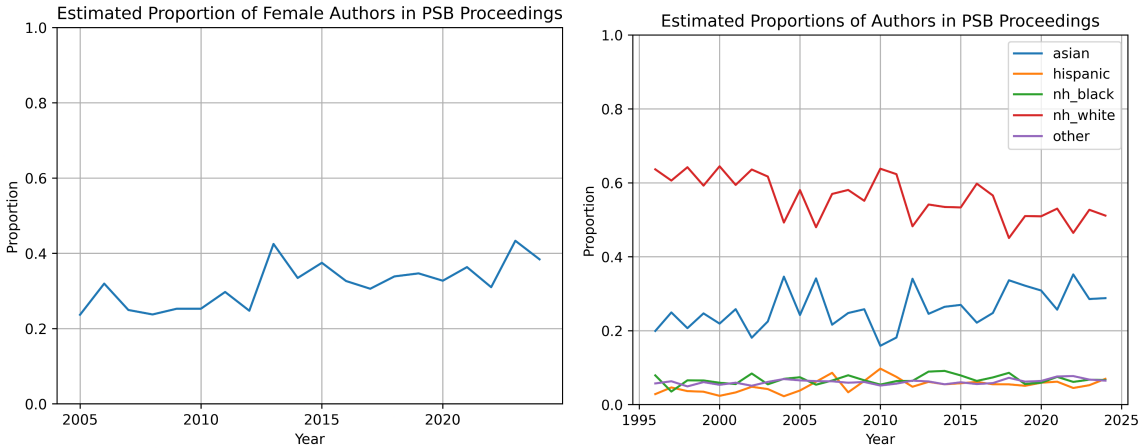

Figure 7: Line graphs of the estimated proportion of female authors (left) and the estimated proportion of authors belonging to one of the five indicated race and ethnic categories (Asian, Hispanic, Non-Hispanic Black, Non-Hispanic White, and Other).

4. Discussion

4.1. *Keyword and Topics Analysis*

As can be seen from the ANZSRC analysis in Figure 2, we see broad patterns across the eras in paper-broad topic assignments that align with the eras themselves. For example, we see a decrease in the number of papers identified as "Biological Sciences", which per the ANZSRC definitions includes more basic science and lab research with some focus on data analysis of wet-lab data[22]. This represents PSB's overall shift to more computational and big data approaches, as corroborated by the word clouds of keywords shifting slightly away from molecular data such as proteins, amino acids, and related, in favor of data and datasets. Similarly, we see an increase in research identified as "Biomedical and Clinical Sciences", which represents an increased tonal shift towards clinical data (as opposed to wet-lab or molecular data) over the years in the field of biocomputing (which encompasses computational biology, bioinformatics, biomedical informatics, and data science).

Interestingly, we see a slight decrease in the proportion of papers identified as "Information and Computing Science" from the first era to the second era, and then an increase from the second to the third era. This may reflect that, in the earliest iterations of PSB, there was a larger focus on developing methods to analyze sequencing data, as reflected in the keywords from the first word cloud showing "sequence", "amino", "alignment", and "algorithm" prominently. With the release of the Human Genome Project, such focuses became less critical and a shift occurred to more methodological *applications* such as genome-wide association studies (GWAS) and annotation analyses (which fall outside the realm of this topic, per ANZSRC), as reflected in the word cloud showing a disappearance of the aforementioned terms and the emergency of terms such as "phenotype" and "annotation". After the resurgence of machine learning and AI in biomedicine, a development of new approaches that leveraged these fields and made use of existing data became a larger focus once more - indicated in the word cloud by terms such as "data", "neural", "predicting", and "embedding" becoming more prominent.

These shifts are further reflected in the papers' subcategories assignments, as shown in Figure 3, where each plot is a subcategory of the broader categories from Figure 2, respectively. There is a consistent decrease in the number of papers described as "Bioinformatics and Computational Biology", which is curious at first for a Biocomputing conference until one recognizes that biomedical informatics is considered a distinct field that is included in the broad scope of Biocomputing. Similarly, there is an increase in clinical-adjacent research in the form of "Oncology and Carcinogenesis", which matches the broad trend of an increase in cancer research as we better understood phenotype data and with the emergence of GWAS, and these trends expectedly match the trends of their parent categories.

The final subcategory of "Machine Learning" has shown dramatic increases that align strongly with the defined eras, going from being a topic of less than 2% of papers in either of the first two eras to 13.6% of papers in the third era, reflecting the period of time in which machine learning and AI became much more strongly incorporated in biomedical research, as well as the transition of authors at PSB to more biomedical and clinical informatics research where big data allows for the training and application of more advanced and complex models.

4.2. *Emissions Analysis*

It is important to note that the assumptions that were made for the emissions analysis, as described in the methods, all lead to a likely underestimate of the true emissions produced. For example, most authors do not have access to airports that offer direct flights to Hawaii, and flights back in the 1990s produced more emissions per-passenger than flights today[23]. Additionally, PSB regularly sees approximately 200+ attendees per year, while this analysis only accounts for roughly ~40-50 of those attendees (the first authors of each accepted paper).

Despite these limitations, this analysis does highlight the fact that PSB does have a relatively high carbon footprint with an average total emission attributable to flights by just first authors of over 100 tons of CO_2e per year. Interestingly, PSB's average flight emissions per year has been decreasing, despite no notable change in the number of papers or attendees used in these calculations, which may indicate a consolidation in the number of traveling authors or an increase in the proportion of authors nearer to Hawaii. Over the years, PSB has contributed to the Hawaiian Legacy Reforestation Initiative[24] which plants koa and sandalwood trees. This is a step toward providing an offset for the carbon footprint[25].

4.3. *Citation and Authorship Analysis*

Overall, from a citation and research output perspective, PSB has been consistently impactful. With a total recorded citation count of 28579 and 90% of papers being cited at least once, PSB has contributed significantly to the body of scientific literature over the past 30 years, and continues to do so. With an h-index of 76 and an h5-index of 13, PSB remains competitive as a conference for biocomputing.

For example, the top papers by citations (Figure 6, right) are concentrated in the first decade of PSB, indicating that they have had a long and lasting impact over the years. However, the papers that have received the most citations in the last two years (Figure 6, left) are largely concentrated within the last decade of PSB, indicating PSB's consistency as a top conference in the field as time goes on, as well as its ability to best attract the cutting-edge ideas in the field of biocomputing.

Furthermore, we find that PSB has encouraged collaborations, with co-authorship networks increasing from 4.3 in its earliest years to 7.8 in the second decade and up to 12.2 in more recent years, indicating that larger groups of authors are working together in PSB. This increase occurs seemingly independently of the number of unique authors (going from 1065 to 1815 and then 1510), indicating that PSB fosters collaborations within its author network.

4.4. *Diversity Analysis*

It is important to note that this information cannot be considered definitive at any non-aggregate scale (that is, any individual level information) due to the use of computed probabilities based on machine learning models, and we recognize that the categories used do not conform to definitions outside or even inside of the USA (for race) or to nonbinary definitions (for gender). Furthermore, transgender individuals may not identify as the gender that they were assigned at birth (which is the information available from the SSA statistics used), and individuals can identify as members of multiple racial or ethnic groups. As such, we demur from drawing strong conclusions about any

individual authors and instead look primarily at population-level trends only with the caveat that this analysis is highly limited at best.

. With these considerations in mind, we do note a trend of an apparent increase in the estimated proportion of published authors that are female from ~25% in some of the first years of PSB to ~35% in recent years. This trend is relatively consistent with proportions of female authorship in other medical journals, with PSB having a slightly higher estimated representation of female authors overall[26–29], and PSB's apparent gender proportion aligns with the proportion of investigators funded by the NIH that identify as female (37% as of 2024)[30].

We note an apparent increase in the estimated proportion of authors that are one of Asian, Hispanic, Non-Hispanic Black, or Other. Correspondingly, we note an apparent decrease in the estimated proportion of authors that are Non-Hispanic White. When compared to the racial and ethnic makeup of NIH-funded investigators as of 2024, PSB has a recent estimated proportion of authors in the two subgroups that have been identified as underrepresented minorities by the NIH[31] that is similar or higher: Hispanic (NIH ~6.1%, PSB ~5.8%) and Non-Hispanic Black (NIH ~3.6%, PSB ~6.5%)[32].

5. Conclusion

Overall, PSB is a conference in the field of biocomputing that presents cutting edge research (Keyword and Topics Analysis) that is highly impactful and fosters collaboration (Citation and Authorship Analysis). Furthermore, PSB publishes papers from authors who represent a diverse range of perspectives and has improved in this regard over the years (Diversity Analysis). PSB remains committed to improving representation from a wider range of groups. We also recognize that these positive aspects of PSB do not come without an environmental cost in the form of flight emissions to travel to PSB (Emissions Analysis); however, the conference has made contributions back to the islands in the form of planting trees to offset this carbon footprint. These insights are useful as we continue to plan for PSB in coming years.

In conclusion, this paper highlights PSB's remarkable record as a leader in Biocomputing over the past thirty years, and we look forward to the future of PSB in fostering collaboration, publishing cutting edge research, and providing an avenue for continued discussions about how to best improve the landscape of biomedical research.

6. Acknowledgments

RK was partially supported by the Training Program in Computational Genomics grant from the National Human Genome Research Institute to the University of Pennsylvania (T32HG000046). DZ was supported by a fellowship from the National Heart, Lung, and Blood Institute (F30HL172382). The National Library of Medicine (R13LM006766) and the International Society for Computational Biology have provided continuous funding to PSB in support of travel awards to increase representation of broad diversity since 1996. This paper was written using data obtained on June 18, 2024, from Digital Science's Dimensions platform, available at https://app.dimensions.ai.

References

1. Hewett D, Whirl-Carrillo M, Hunter LE, Altman RB, Klein TE. A twentieth anniversary tribute to PSB. Pac Symp Biocomput Pac Symp Biocomput. 2015;1–7.

2. PSB Proceedings [Internet]. [cited 2024 Jul 31]. Available from: https://psb.stanford.edu/psb-online/

3. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinforma Oxf Engl. 2009 Jun 1;25(11):1422–3.

4. Hook DW, Porter SJ, Herzog C. Dimensions: Building Context for Search and Evaluation. Front Res Metr Anal [Internet]. 2018 Aug 23 [cited 2024 Jul 31];3. Available from: https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2018.00023/full

5. Porter SJ, Hawizy L, Hook DW. Recategorising research: Mapping from FoR 2008 to FoR 2020 in Dimensions. Quant Sci Stud. 2023 Mar 1;4(1):127–43.

6. Grootendorst M. KeyBERT: Minimal keyword extraction with BERT. [Internet]. Zenodo; 2020. Available from: https://doi.org/10.5281/zenodo.4461265

7. Austin CP. The impact of the completed human genome sequence on the development of novel therapeutics for human disease. Annu Rev Med. 2004;55:1–13.

8. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020 Jan;577(7792):706–10.

9. amueller/word_cloud: A little word cloud generator in Python [Internet]. [cited 2024 Jul 31]. Available from: https://github.com/amueller/word_cloud

10. googlemaps/google-maps-services-python [Internet]. Google Maps Platform; 2024 [cited 2024 Jul 31]. Available from: https://github.com/googlemaps/google-maps-services-python

11. Megginson D. davidmegginson/ourairports-data [Internet]. 2024 [cited 2024 Jul 31]. Available from: https://github.com/davidmegginson/ourairports-data

12. Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, et al. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. Nat Commun. 2014 Apr 29;5:3513.

13. Greenhouse gas reporting: conversion factors 2022 [Internet]. GOV.UK. 2022 [cited 2024 Jul 31]. Available from: https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2022

14. Ritchie H, Roser M. Which form of transport has the smallest carbon footprint? Our World Data [Internet]. 2024 Mar 18 [cited 2024 Jul 31]; Available from: https://ourworldindata.org/travel-carbon-footprint

15. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019 Mar 26;9(1):5233.

16. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008 Oct;2008(10):P10008.

17. Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. [Internet]. Stanford InfoLab; 1999 Nov. Report No.: 1999–66. Available from: http://ilpubs.stanford.edu:8090/422/

18. Teich EG, Kim JZ, Lynn CW, Simon SC, Klishin AA, Szymula KP, et al. Citation inequity and gendered citation practices in contemporary physics. Nat Phys. 2022 Oct;18(10):1161–70.

19. Popular Baby Names [Internet]. [cited 2024 Jul 31]. Available from: https://www.ssa.gov/oact/babynames/limits.html

20. Chintalapati R, Laohaprapanon S, Sood G. ethnicolr2: Predict Race and Ethnicity From Name [Internet]. 2023 [cited 2024 Jul 31]. Available from: https://github.com/appeler/ethnicolr2

21. Chintalapati R, Laohaprapanon S, Sood G. Predicting Race and Ethnicity From the Sequence of Characters in a Name [Internet]. arXiv; 2023 [cited 2024 Jul 31]. Available from: http://arxiv.org/abs/1805.02109

22. Australian and New Zealand Standard Research Classification (ANZSRC), 2020 | Australian Bureau of Statistics [Internet]. 2020 [cited 2024 Jul 31]. Available from: https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-research-classification-anzsrc/latest-release

23. Lee DS, Fahey DW, Skowron A, Allen MR, Burkhardt U, Chen Q, et al. The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. Atmospheric Environ Oxf Engl 1994. 2021 Jan 1;244:117834.

24. Hawaiian Legacy Reforestation Initiative [Internet]. [cited 2024 Jul 31]. Available from: https://legacyforest.org/

25. PSB Trees [Internet]. [cited 2024 Jul 31]. Available from: https://psb.stanford.edu/trees/

26. Brück O. A bibliometric analysis of the gender gap in the authorship of leading medical journals. Commun Med. 2023 Dec 11;3(1):1–7.

27. Krstacic JE, Carr BM, Yaligar AR, Kuruvilla AS, Helali JS, Saragossi J, et al. Academic medicine's glass ceiling: Author's gender in top three medical research journals impacts probability of future publication success. PloS One. 2022;17(4):e0261209.

28. Hart KL, Perlis RH. Trends in Proportion of Women as Authors of Medical Journal Articles, 2008-2018. JAMA Intern Med. 2019 Sep 1;179(9):1285–7.

29. Bernardi K, Lyons NB, Huang L, Holihan JL, Olavarria OA, Martin AC, et al. Gender Disparity in Authorship of Peer-Reviewed Medical Publications. Am J Med Sci. 2020 Nov;360(5):511–6.

30. NIH Data Book [Internet]. [cited 2024 Jul 31]. Available from: https://report.nih.gov/nihdatabook/report/218

31. Underrepresented Racial and Ethnic Groups | Diversity in Extramural Programs [Internet]. [cited 2024 Jul 31]. Available from: https://extramural-diversity.nih.gov/diversity-matters/underrepresented-groups

32. NIH Data Book - Data by Race [Internet]. [cited 2024 Jul 31]. Available from: https://report.nih.gov/nihdatabook/report/306