Major points of concern:

**IRB approval:**

B- As you present the findings of the self-study for publication, more details and discussion of human studies protocols should be discussed.

C - IRB exempt status or approval is not mentioned. Given this is research involving human subjects (e.g., hand-coding of proxy demographics), information should be given about whether this study was considered by an IRB, and if not, why consideration was not required…. Any interaction with IRB should be described. If an IRB did not consider the study, the reason this was not required should be explained in the Methods.

D- Would the self-identification tools DiversifyImmunology and DiversifyMicrobiology require an IRB approval? – only if used for research purposes

We did not obtain IRB approval, data were public

**Methodology:**

B- While this is a retrospective study and given the nature of what you aimed to categorize, I am not fully comfortable with your methods for proxy demographics. It would be extremely valuable, almost essential, that you dig into relevant literature to find methods to guide and describe/support what you did. Where did you get the criteria to assign people the way you did? What did you look for on a CV that might provide some objective insight? While you note that your proxy demographics have limitations, this isn't sufficient in my mind. Your approach would be a good preliminary study for internal use in your department, but not for a publication. Your study is interesting and important for us all to reflect on, but you need some literature support for your methods. Retrospective studies are done in many other fields and I would recommend looking there (e.g. epidemiology, biomedical studies, sociology, etc.). … There was no mention of IRB and I have concerns about some of the data collection regarding the trainees and faculty and the proxy demographics for external speakers.

D - How did you determine the demographics of the faculty hosts and trainees? You use the email addresses of the trainees for a list, but is this a similar approach where you hand-coded the data or do you have self-identified data? If you collected self-identified data for faculty and trainees, then this may require an IRB approval. Be sure to state if the faculty host and trainee demographic data were collected in the same way as the invited speaker data. In the background section, it is critical to provide evidence that your approach for determining demographics is sound or has relevance in previous work.

Editor - Percentages.

This leads me to my confusion on the percentages argument. As read here the implication seemed that if the invited percentages matched the US census percentage then equity was achieved. I'm not sure how I feel about that argument. What exactly would the sociological impacts of a 4 percent difference be versus say a 6 percent difference? Hard numbers I feel can blind those interested in change to true inclusions which have more to do more access to power and shaping narratives than matching percentages. While the manuscript is not written to discuss this, it doesn't address this at all, which for me limits its theoretical and sociological depth. This becomes problematic when solutions are discussed, but more on that later.

**Sufficient Literature Review:**

B- I believe so, but I am not well-read in social role theory. It would be helpful, and provide more strength for the rationale, if social role theory was explicitly defined around line 75. Something like 'According to social role theory, individuals make inferences....' and some more elaboration.

C - Social role theory should be described in more detail. A description of the constructs of science identity and belonging in science, which were both mentioned in the abstract, could be expanded. ... Social role theory is mentioned on lines 74 and 90 of the Introduction and line 157 of the Discussion, but not described. I recommend a short paragraph summarizing this theory and its major points. It is important to introduce the reader to this theory to help them appreciate the work. A description of the constructs of science identity and belonging in science, which were both mentioned in the abstract, could also be expanded.

D - There is a strong amount of literature reviewed in the manuscript. However, there is a gap that supports the methods used in your study. I would suggest including sources that use a similar retrospective study of demographics. You mention the limitations of your study, but are there other publications in the literature that use personal judgement to determine race and gender?   
  
D - I also suggest including a short discussion of the limitation of perceived identity with additional citations in the introduction section of the manuscript. Placement at the beginning helps explain the unconscious bias people have, establishing a framework for hand-coding the demographic data of the invited speakers and how that this can be misinformed as well as relevant to the social role theory.

Editor - A more comprehensive review in the introduction might better motivate the argument being made and the conclusions drawn from the data. Some of these reasons appeared in the discussion, but would make more sense being said earlier. The core issue is that barriers exist that result in under representation. The introduction would best serve to describe with literature support why those barriers exist. This would force the authors to at least touch on things beyond 'seeing people like them', and reference studies on bias in hiring, cultural tax experienced by faculty of color, funding biases and the list can go on and on. Biases associated with inner circle membership has everything to do with tribalism. There is a lot of literature on that. Also, the importance of increased representation should go beyond 'like seeing like', and emphasize the fact that inclusion is everyone having an equal opportunity to shape the future, including our scientific futures. Representation while important is still relatively low-level.

Social Role Theory

They may want to just take this out. As written here it is an inappropriate use of the framework (or even its cousin sociocultural theory). This framework focuses mostly on conventional gender roles, and I certainly have not seen examples of its applicability to ethnicity. If the authors have seen that, it should be made more explicit. Secondly, it is given as an example but it is not really being used to frame the study, which is an unusual way to incorporate theoretical frameworks into a paper like this. Framework use occurs more along the lines of 1) theory posits an explanation for an observed phenomenon, 2) authors believe it is applicable to their context and set about through some vetted method to figure that out. It is not an 'example' of why something happens.

**Other comments:**

B- Minor comments:  
  
Line 173-174. Do you have a citation to support the claim about 'a particular responsibility to ...'  
Line 181. If you don't have a citation to support this claim (some departments only invite tenured faculty) it would be better to say that 'it is possible' or 'may'.  
Line 216. Is naming the department head necessary?  
Lines 224-226. Is there a citation to support this? I don't fully understand  
Line 232: Should be Building Diversity (correct?)

C - Major comments   
  
2) Given that the data were collected from one department at one university, it is important to give the number of people in various categories. A good place for this would be on line 110 when referring to invited speakers (n=?) and line 111 faculty hosts (n=?) and line 112 (n=?). Line 135 was the one place I noticed a number of people given for a category.  
  
3) I appreciate the limitations mentioned in the Methods on lines 126-131 and I recommend expanding this to a separate Limitations section in the Discussion. In addition to acknowledging the limitations of hand-coding proxy demographics, a discussion of the sample size and the generalizability of the results, would be valuable. What future research could be done to address these limitations?  
  
4) Data analysis is mentioned on line 131 of the Methods, but no statistical analysis was done to determine whether the differences seen in Figure 1 are significant. I recommend including appropriate statistical analysis to strengthen the paper.   
  
C - Minor comments  
  
1) Lines 155-156: It would be helpful to frame the differences in invited seminar speakers and invited lectureship speakers and trainees in terms of the percent of women and HURM and NCNH scientists in the fields of microbiology and/or immunology. Are there data about the estimated percentage of microbiologists or immunologists you can use as a comparison or even scientists in general?  
  
2) Lines 197-199: The insights provided regarding assumptions made about potential speakers who are female was interesting. Can any insights be provided for Asian scientists?  
  
3) Line 216: Remove Dr. Mobley’s name, given they are not an author on the paper.  
  
4) Line 287: The title and description listed here do not match Table 1.

D -

For the title, I would avoid using the word "should". Maybe instead "Policy change to improve invited speaker diversity and reflect trainee diversity".   
  
How large is the faculty host proxy group versus the entire faculty? You list percentages, but it would also be useful to give total numbers for the groups being examined. If you use the faculty hosts as a proxy for the departmental faculty demographics, does this exclude some trainee's faculty advisors who would also contribute to the social role theory in other aspects? The faculty host proxy seems to simplistically exclude unrelated circumstances that may have prevented a faculty member from being a host. Assuming that the faculty hosts are more visible representatives of the department may or may not reflect the individual experience of a trainee. I would recommend comparing demographics of the faculty hosts with the entire faculty.   
  
Check your usage of White and Caucasian. I am not certain of the appropriate language here, but are you using them interchangeably? If so, it would be helpful to be consistent with the terminology. Also, it would be helpful to give an example of the NCNH grouping and how you made the distinction between HURM and NCNH. This is a delicate subject and it is important to be careful in your approach. There are a number of good points in this article and addressing the concern for invited speaker diversity is highly relevant.

Editor -

On the suggestions

The authors could probably do more here. As is, there is a list, and departments can go use it. Why do you suppose they may want to use the list? What level of cultural competency would one consider list-use to be? I like the suggestion of having the lab choose the speaker versus the professor. Any comments on how power structures in the lab might wind up with the students picking the same kinds of invited speakers? As a grad student I led a grad student group who did just that, and in too many cases, advisers were telling their grad students who to choose.