Null models for gene enrichment in plasmids

Liam Shaw[a,1]

[a]Department of Zoology, University of Oxford

The recent article by Che et al. (1) presents an analysis of a dataset of 14,029 plasmids. I enjoyed the paper but have some concerns about its analysis. While it is clear that more conjugative plasmids carry AMR genes, it is unclear whether this is simply because conjugative plasmids tend to carry more genes in general. Specifically, Che et al. aim to determine if the AMR gene pool is 'homogeneously distributed' or 'constrained' by plasmid categories. To do this they perform a Fisher's exact test of the proportion of AMR gene-bearing plasmids between plasmid categories. In my view this test is insufficient by itself for investigating 'enrichment'—a term which suggests differential processes of horizontal gene transfer (HGT) between categories.

To highlight the problem with an analogy, consider two types of dice trial: one where a die is rolled ten times and one where it is rolled only twice. Trials of the first type will have more 6s. It is wrong to conclude from this observation alone that trials-with-6s are 'enriched' in this type; one should also control for the number of rolls. (Che et al.'s subsequent analysis of enrichment of AMR gene types in sets of genes within plasmid categories is not subject to this problem.)

For the case in hand, assuming independence of association in Fisher's exact test is equivalent to assuming that plasmids of different sizes have equal proportions of AMR-gene-bearing plasmids. But there are better null models of homogeneous distribution by HGT. For example, one could assume AMR genes are Poisson-distributed, occurring at a mean rate λ per bp irrespective of plasmid category. Assuming 13,899 AMR genes distributed across 1.41 Gbp (total length of all plasmids) gives $\lambda = 9.8 \times 10^{-6}$ per bp. A simulation shows Fisher's exact test gives a highly significant result for conjugative plasmids (Table 1), which might lead to the conclusion that 'AMR gene-bearing plasmids were significantly enriched in this category'. However, we know the AMR genes have been homogeneously distributed; a greater proportion of conjugative plasmids are AMR gene-bearing simply due to their greater length.

	Category	No AMR genes	AMR gene-bearing
Poisson simulation	Non-conjugative	6881 (65.5%)	3611 (34.4%)
OR=[2.88-3.38]	Conjugative	1340 (37.9%)	2197 (62.1%)
Observed	Non-conjugative	8826 (84.1%)	1666 (15.9%)
OR=[5.38-5.85]	Conjugative	1681 (47.5%)	1856 (52.5%)

Table 1. Comparison of a simulation assuming Poisson-distributed AMR genes (top) against the observed results (bottom). OR=odds ratio for 'enrichment' of AMR gene-bearing plasmids from Fisher's exact test (both $p < 0.001$).

Table 1 shows the Poisson model is poor (probably a zero-inflated Poisson model would be a better choice). But we can still use it as a guide for thinking about odds ratios. Since the model produces an OR of around 3, that suggests OR>3 as a more appropriate threshold than OR>1. In fact, we observe OR>5, so there is a proportional difference for conjugative plasmids that remains unexplained. By this rough rule-of-thumb, Che et al.'s original conclusion stands.

Nevertheless, I raise this general point because I think language should match analysis. For example, elsewhere Che et al. describe detection of the same AMR gene on two plasmids as 'when an AMR gene was *transferred* between two plasmids' [my emphasis] without further analysis—I think this must be a mistake. But this is not just semantics. When investigating enrichment, considering null models carefully is important.

ACKNOWLEDGMENTS. I thank Tong Zhang for helpful correspondence about the original paper and Craig MacLean for feedback on a draft of this letter.

1. Y Che, et al., Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proc. Natl. Acad. Sci.* **118** (2021).

L.P.S. wrote the paper. L.P.S. is a Sir Henry Wellcome Postdoctoral Fellow funded by Wellcome (grant 220422/Z/20/Z).

No competing interests to declare.

[1]E-mail: liam.shaw@zoo.ox.ac.uk