

## Taxonomic assessment of two wild house mouse subspecies using whole genome sequencing

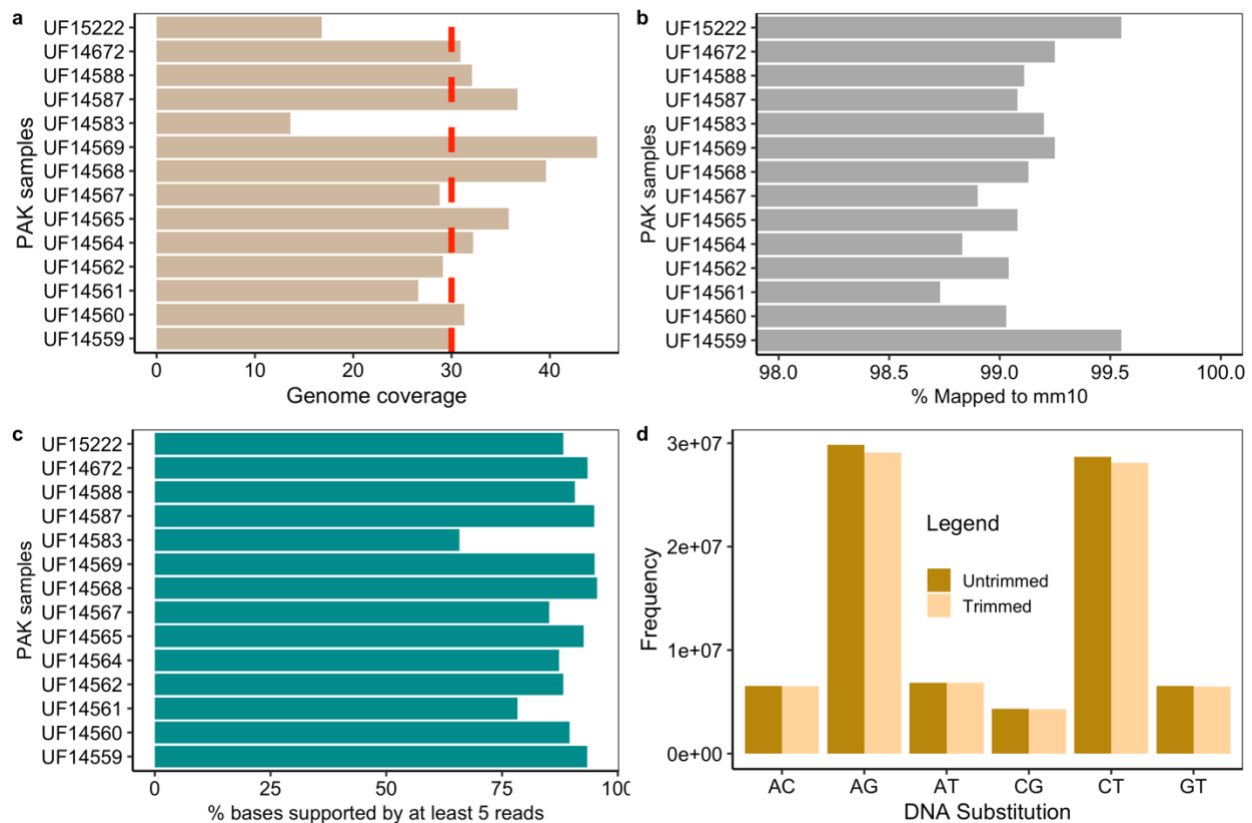
Raman Akinyanju Lawal<sup>1\*</sup>, Verity L. Mathis<sup>2</sup>, Mary E. Barter<sup>1</sup>, Jeremy R. Charette<sup>1</sup>, Alexis Garretson<sup>1,3</sup>, and Beth L. Dumont<sup>1,3\*</sup>

<sup>1</sup>The Jackson Laboratory, 600 Main Street, 04609. Bar Harbor, USA.

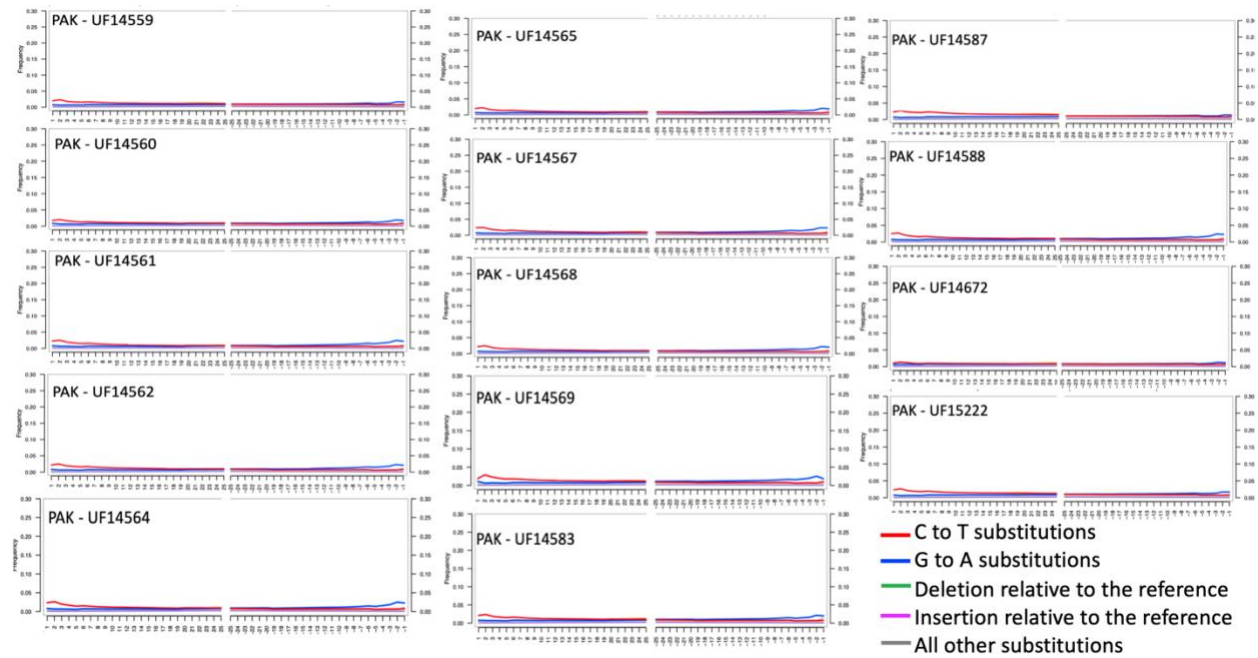
<sup>2</sup>Florida Museum of Natural History, University of Florida, 1659 Museum Road, Gainesville FL 32611, USA.

<sup>3</sup>Tufts University, Graduate School of Biomedical Sciences, 136 Harrison Ave, Boston, MA, 02111, USA

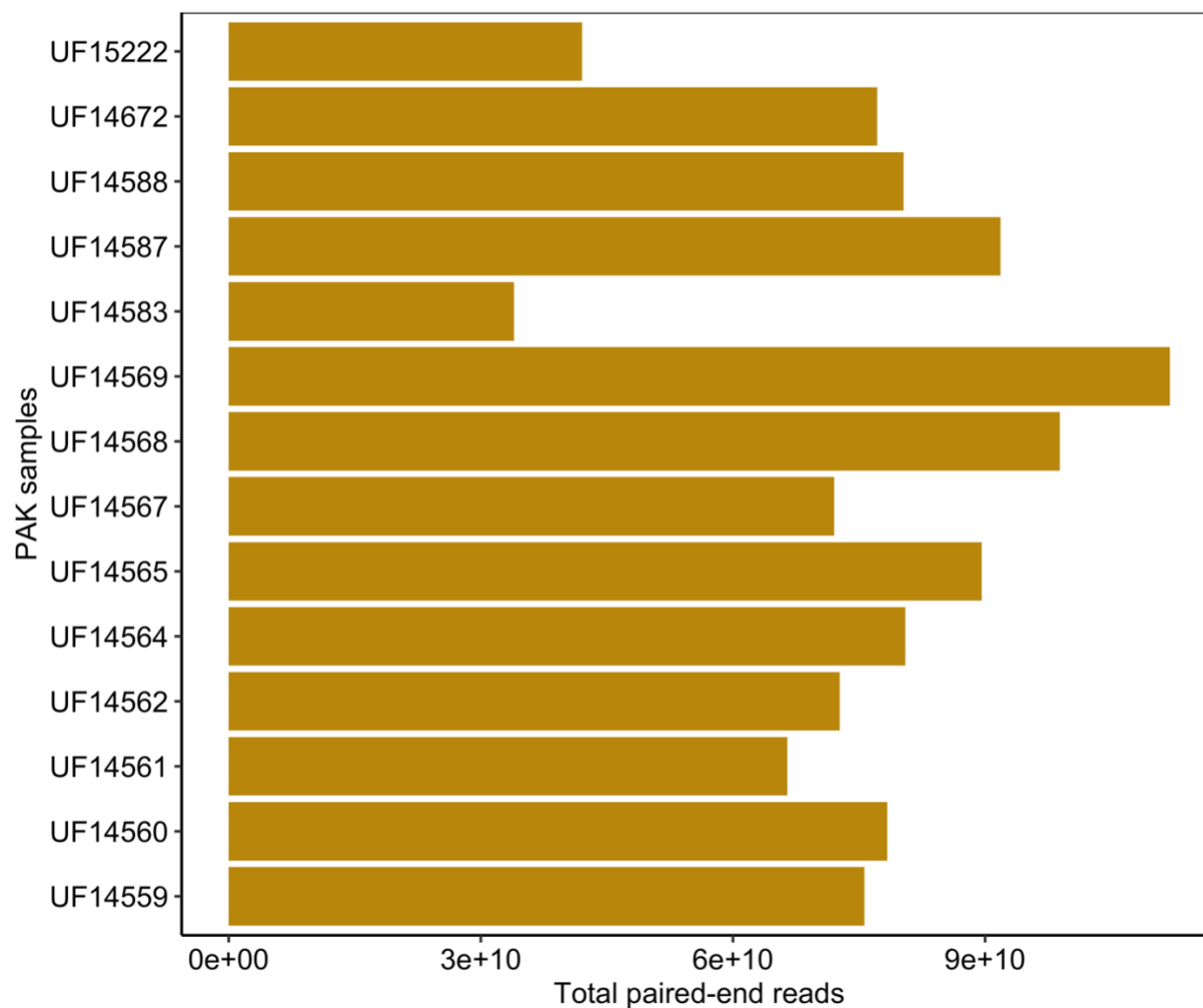
\*Correspondence: RAL ([raman.lawal@jax.org](mailto:raman.lawal@jax.org)) and BLD ([beth.dumont@jax.org](mailto:beth.dumont@jax.org))



**Supplementary Figure S1:** Sequence and mapping statistics for 14 PAK house mouse genomes. (a) We targeted an average of 30x coverage per sample (red line), with actual coverage values ranging from 17-45x. Bar charts showing (b) the percentage of sequenced reads mapping to the mm10 reference genome per sample and (c) the percentage of bases supported by at least 5 reads. (d) DNA substitution rates before and after trimming the 5bp at the 3' and 5' termini of each read.



**Supplementary Figure S2:** The distribution of DNA misincorporation errors at both the 5' and 3' termini of sequenced reads from PAK house mice. The y-axis shows the frequency of post-mortem DNA damage relative to the mm10 reference genome. Post-mortem DNA damage is expected to lead to an excess accumulation of C>T and G>A substitutions.



**Supplementary Figure S3:** Bar plot of the number of paired-end reads generated for each PAK sample.