

**S1 Table: Method of Calculation of sequence-based diversity measures**

**TABLE 1** Description of sequence-based diversity measures

Diversity measure	Method of calculation
% diversity	Average pairwise genetic distance was calculated using the maximum composite likelihood model. MEGA 7.0.14 ( <a href="http://www.megasoftware.net">www.megasoftware.net</a> ) was used to analyze gap-stripped sequence alignments. This value is reported as a percentage.
% complexity	$[(\text{no. of distinct variants})/(\text{total no. of reads})] \times 100$
Shannon entropy (S)	$S = - \left( \frac{1}{\log N} \right) \sum_{i=1}^n p_i \log p_i$ <p>where <math>p_i</math> is the proportion of reads consisting of each unique sequence pattern, <math>N</math> is the total no. of sequences, and <math>n</math> is the no. of distinct sequences</p>

*Sequence-based measures were calculated using the method previously described by Cousins MM, Ou SS, Wawer MJ, Munshaw S, Swan D, Magaret CA, Mullis CE, Serwadda D, Porcella SF, Gray RH, Quinn TC, Donnell D, Eshleman SH, Redd AD. Comparison of a high-resolution melting assay to next-generation sequencing for analysis of HIV diversity. J Clin Microbiol. 2012 Sep;50(9):3054-9. doi:10.1128/JCM.01460-12.*