



**Figure 5 - figure supplement 2.** (a) *In vivo* occupancy in *T. acidophilum* is poorly predicted by a Lasso model trained on a *T. acidophilum* naked DNA digest ( $\rho=0.07$ ,  $P < 2.2 \times 10^{-16}$ ). (b) In contrast, *in vivo* occupancy in *T. acidophilum* is well predicted by a Lasso model trained on digestion fragments from HTa-expressing *E. coli* ( $\rho=0.54$ ,  $P < 2.2 \times 10^{-16}$ ). All correlations/predictions are for short fragments.