This manuscript reports intriguing results illustrating the potential to use convergent evolution of sites in genes of HIV (namely, the *env* gene) as an indicator of the action of natural selection during acute HIV-1 infection. It illustrates that convergent mutations are frequent during infection of HIV-1; that they are more frequently located in the gp41 domain; that they don't occur preferentially in positions of high nucleotide diversity; and that they aren't significantly more likely to be synonymous in nature. Overall, the results are persuasive and interesting, and there is definitely utility to greater analysis of convergence as an indicator of selection in molecular evolution in general and the molecular evolution of HIV specifically. There are three major and a number of minor points that could be addressed in a revision to increase the impact of the manuscript.

First, additional contextualization of this work in comparison to other work (sometimes on other HIV genes) examining viral convergence (e.g. Xue et al. 2017) and the early evolution of HIV during acute infection (e.g. Herbeck et al. 2009, 2011a,b; Lee et al. 2009; Boutwell et al. 2010; Giorgi et al. 2013; Gounder et al. 2015; Garcia-Knight et al. 2016) is warranted. Particularly relevant to this manuscript are Yoshida et al. 2011; Henn et al. 2012; and Park et al. 2016.

Second, adherence to two guidelines would enhance readability and persuasiveness. First, a greater adherence to the guideline of dedicating the x-axis to the independent variable and the y-axis to the dependent variable would improve the figures. In Figure 1 it is OK, but using the HIV populations on the x-axis in Figure 1 sets up a precedent that when re-used in Figures 2, 3, and 5 means that the variable that is dependent on the hypothesis ends up depicted on the x-axis, which is confusing. Switching the axes would improve their comprehensibility. Additionally, Figure 4 would be clearer if proportion nonsynonymous (amino acid replacement) were illustrated instead of proportion synonymous, because the conventional null hypothesis would be that nonsynonymous changes would be more likely to be convergently selected—at least in comparison to synonymous changes, which are often and conventionally assumed to be neutral.

Third, for one of the main results of the manuscript (the clustering of convergent mutations within the discrete linear sequence of the env gene), there are much more powerful and persuasive approaches than are used here that can detect and illustrate the clustering of convergent mutations without a priori assumptions (e.g. Tang and Lewontin 1999; Zhang and Townsend 2009) that would represent a significant improvement over the ad hoc approach underlying Figure 2. It might be best to select an appropriate threshold of number of populations or (better) to perform clustering on all applicable thresholds (> x populations) to ensure that the result isn't a product of selecting a unique threshold and is not abrogated by saturation. Generally, it would improve the manuscript methodologically to more explicitly differentiate the utility of the methods used in this manuscript in comparison to other approaches such as molecular evolutionary models of selection (e.g. Rodrigo and Learn 2007; Kosakovsky Pond et al. 2008; Leitner 2012; Zhao et al. 2017).

Minor points:

1. Consider the points raised in Stayton (2015) regarding research on “convergent evolution”.
2. Throughout the manuscript, for clarity, follow “this” immediatly by a noun referent.
3. The writing in the abstract needs work: what is “it” (line 12)? This is missing a referent (line 13). Population(s) (line 15)? Use of negative cases (not...not) in lines 21–22 is awkward and confusing.
4. Run-on sentence lines 62–65.
5. Are differences in mutation rate across sites adequately accounted for (lines 79–89)?
6. Awkward / confusing phrasing, lines 147–148 of Fig. 3 legend.
7. Confusing writing on lines 165–168; break up & clarify sentence.
8. Test the assertion in line 190 by plotting diversity across sites and discrete linear clustering across sites in the same plot.
9. “Is adjust” line 203.
10. Writing requires greater clarity, lines 214–222.
11. Writing requires greater clarity, lines 258–262.
12. On line 346, nucleotide diversity is defined as Shannon entropy, which by definition has a range from 0–1. On line 354, nucleotide diversity is stated to range from 0–2. Clarify.

References cited:

Boutwell, C. L., M. M. Rolland, J. T. Herbeck, J. I. Mullins, and T. M. Allen. 2010. Viral evolution and escape during acute HIV-1 infection. J. Infect. Dis. 202 Suppl 2:S309–14.

Garcia-Knight, M. A., J. Slyker, B. L. Payne, S. L. K. Pond, T. I. de Silva, B. Chohan, B. Khasimwa, D. Mbori-Ngacha, G. John-Stewart, S. L. Rowland-Jones, and J. Esbjörnsson. 2016. Viral Evolution and Cytotoxic T Cell Restricted Selection in Acute Infant HIV-1 Infection. Sci. Rep. 6:29536.

Giorgi, E. E., B. T. Korber, A. S. Perelson, and T. Bhattacharya. 2013. Modeling sequence evolution in HIV-1 infection with recombination. J. Theor. Biol. 329:82–93.

Gounder, K., N. Padayachi, J. K. Mann, M. Radebe, M. Mokgoro, M. van der Stok, L. Mkhize, Z. Mncube, M. Jaggernath, T. Reddy, B. D. Walker, and T. Ndung'u. 2015. High frequency of transmitted HIV-1 Gag HLA class I-driven immune escape variants but minimal immune selection over the first year of clade C infection. PLoS One 10:e0119886.

Henn, M. R., C. L. Boutwell, P. Charlebois, N. J. Lennon, K. A. Power, A. R. Macalalad, A. M. Berlin, C. M. Malboeuf, E. M. Ryan, S. Gnerre, M. C. Zody, R. L. Erlich, L. M. Green, A. Berical, Y. Wang, M. Casali, H. Streeck, A. K. Bloom, T. Dudek, D. Tully, R. Newman, K. L. Axten, A. D. Gladden, L. Battis, M. Kemper, Q. Zeng, T. P. Shea, S. Gujja, C. Zedlack, O. Gasser, C. Brander, C. Hess, H. F. Günthard, Z. L. Brumme, C. J. Brumme, S. Bazner, J. Rychert, J. P. Tinsley, K. H. Mayer, E. Rosenberg, F. Pereyra, J. Z. Levin, S. K. Young, H. Jessen, M. Altfeld, B. W. Birren, B. D. Walker, and T. M. Allen. 2012. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. PLoS Pathog. 8:e1002529.

Herbeck, J. T., M. Rolland, Y. Liu, S. McLaughlin, J. McNevin, K. Diem, A. C. Collier, M. Juliana McElrath, and J. I. Mullins. 2011a. 175 HIV-1 Evolution in Primary Infection is Affected by Stochastic Followed by Selective Processes. JAIDS Journal of Acquired Immune Deficiency Syndromes 56:73.

Herbeck, J. T., M. Rolland, Y. Liu, S. McLaughlin, J. McNevin, H. Zhao, K. Wong, J. N. Stoddard, D. Raugi, S. Sorensen, I. Genowati, B. Birditt, A. McKay, K. Diem, B. S. Maust, W. Deng, A. C. Collier, J. D. Stekler, M. J. McElrath, and J. I. Mullins. 2011b. Demographic Processes Affect HIV-1 Evolution in Primary Infection before the Onset of Selective Processes. J. Virol. 85:7523–7534.

Herbeck, J. T., M. T. Rolland, W. C. Deng, A. C. Collier, and J. I. Mullins. 2009. P07-06. HIV-1 transmission and early evolution: whole genome analysis. Retrovirology 6:P104.

Kosakovsky Pond, S. L., A. F. Y. Poon, S. Zárate, D. M. Smith, S. J. Little, S. K. Pillai, R. J. Ellis, J. K. Wong, A. J. Leigh Brown, D. D. Richman, and S. D. W. Frost. 2008. Estimating selection pressures on HIV-1 using phylogenetic likelihood models. Stat. Med. 27:4779–4789.

Lee, H. Y., E. E. Giorgi, B. F. Keele, B. Gaschen, G. S. Athreya, J. F. Salazar-Gonzalez, K. T. Pham, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, B. H. Hahn, G. M. Shaw, B. T. Korber, T. Bhattacharya, and A. S. Perelson. 2009. Modeling sequence evolution in acute HIV-1 infection. J. Theor. Biol. 261:341–360.

Leitner, T. 2012. The Molecular Epidemiology of Human Viruses. Springer Science & Business Media.

Park, S. Y., T. M. T. Love, A. S. Perelson, W. J. Mack, and H. Y. Lee. 2016. Molecular clock of HIV-1 envelope genes under early immune selection. Retrovirology 13:38.

Rodrigo, A. G., and G. H. Learn Jr. 2007. Computational and Evolutionary Analysis of HIV Molecular Sequences. Springer Science & Business Media.

Stayton, C. T. 2015. What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. Interface Focus 5:20150039.

Tang, H., and R. C. Lewontin. 1999. Locating regions of differential variability in DNA and protein sequences. Genetics 153:485–495.

Xue, K. S., T. Stevens-Ayers, A. P. Campbell, J. A. Englund, S. A. Pergam, M. Boeckh, and J. D. Bloom.

2017. Parallel evolution of influenza across multiple spatiotemporal scales. Elife 6.

Yoshida, I., W. Sugiura, J. Shibata, F. Ren, Z. Yang, and H. Tanaka. 2011. Change of Positive Selection Pressure on HIV-1 Envelope Gene Inferred by Early and Recent Samples. PLoS One 6:e18630.

Zhang, Z., and J. P. Townsend. 2009. Maximum-likelihood model averaging to profile clustering of site types across discrete linear sequences. PLoS Comput. Biol. 5:e1000421.

Zhao, Z.-M., M. C. Campbell, N. Li, D. S. W. Lee, Z. Zhang, and J. P. Townsend. 2017. Detection of regional variation in selection intensity within protein-coding genes using DNA sequence polymorphism and divergence. Mol. Biol. Evol., doi: 10.1093/molbev/msx213.