1 We are very grateful to the reviewers for reading the manuscript in detail and providing helpful comments. Before
2 providing detailed response for each reviewer, we would like to address two common issues raised by the reviewers.

3 • We thank the reviewers for the suggestion of
4 reporting the experimental results regarding time
5 complexity of the method and its convergence. Re-
6 sults on DS1 are shown in the right figure. We see
7 that by the time PSP converges, the proposed flow-
8 based methods achieve comparable (if not better)
9 approximation quality and quickly surpass PSP as
10 the number of iterations increases. This means the
11 slower convergence of flow-based methods may
12 not affect the quality of approximation given sim-
13 ilar computation budget as that of PSP. We will
14 add direct comparison to MCMC in our revision.
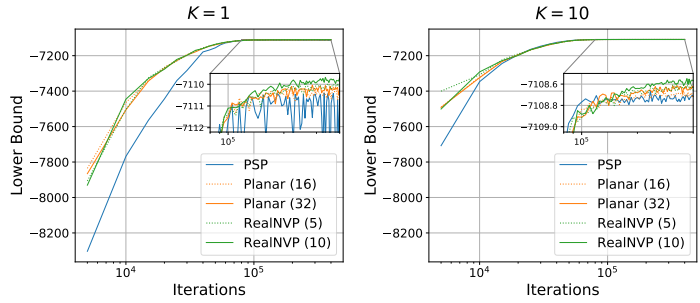
* The computation budget (10×400000) is the same as in Figure 4, Zhang
and Matsen 2019 (20×200000).

15 • Regarding marginal likelihood estimates, we want to clarify that all methods provide estimates for the same marginal
16 likelihood, and better approximation would lead to smaller variance. We will add comparisons to the stepping-stone
17 (SS) method (one of the state-of-the-art sampling-based methods for marginal likelihood estimation) in our revision.

18 Reviewer 1:

19 • We will add more details about the background to the appendix and clarify notation accordingly.

20 • The ELBO (K=1) for all methods are reported in the following table. The gain is more significant now.

	Data set	DS1	DS2	DS3	DS4	DS5	DS6	DS7	DS8
LB (K=1)	PSP	-7111.23(1.04)	-26369.63(0.69)	-33736.60(0.33)	-13332.37(0.54)	-8218.35(0.20)	-6729.27(0.50)	-37335.15(0.11)	-8655.48(0.38)
	Planar(16)	-7110.33(0.16)	-26368.80(0.27)	-33736.14(0.14)	-13331.92(0.11)	-8217.98(0.13)	-6728.89(0.18)	-37334.78(0.11)	-8655.15(0.17)
	Planar(32)	-7110.22(0.17)	-26368.69(0.23)	-33736.02(0.21)	-13331.73(0.12)	-8217.90(0.14)	-6728.68(0.19)	-37334.60(0.12)	-8654.97(0.16)
	RealNVP(5)	-7110.12(0.13)	-26368.75(0.24)	-33735.86(0.10)	-13331.71(0.11)	-8217.80(0.14)	-6728.54(0.15)	-37334.44(0.11)	-8654.62(0.13)
	RealNVP(10)	**-7109.80(0.11)**	**-26368.59(0.23)**	**-33735.81(0.12)**	**-13331.39(0.08)**	**-8217.56(0.12)**	**-6728.04(0.14)**	**-37333.94(0.09)**	**-8654.02(0.12)**

21 Reviewer 2:

22 • We want to clarify that the lower bounds are indeed significantly improved considering the marginal likelihoods they
23 approach. Moreover, the multi-sample lower bound (K=10) we reported is biased towards the marginal likelihood,
24 making improvements seem less significant. As suggested by Reviewer 1, we report the ELBO (K=1) in the table above,
25 which shows more significant gain of the proposed methods. Issues related to the marginal likelihood estimates are
26 discussed in the beginning. The structured parameterization of our flow-based models allows tree topologies to have
27 their own parameters for flows, while sharing similarities according to the local topological structures they possess.

28 • The existing flow models for unordered data require splitting the feature vector of each data point for RealNVP, which
29 does not work if these features are scalars (e.g., the branch lengths for edges of a phylogenetic tree). We will clarify this.

30 Reviewer 3:

31 • Phylogenetic analysis of viral genomes provides key insight into disease pathophysiology, spread and potential
32 control (e.g., the recent COVID-19 pandemic), offering invaluable information for public health decisions. Our methods
33 can perform Bayesian phylogenetic inference in a timely manner that is unlikely to be met by optimizing existing
34 MCMC strategies. Moreover, phylogenetic models are challenging for traditional methods and would inspire novel
35 statistical/machine learning approaches for models with complex, highly-structured, non-Euclidean parameter space.

36 • Yes, there are strong correlations among branch lengths. In fact, the significant lower bound improvement of RealNVP
37 (see Figure 3), which is based on coupling layers, is a strong evidence of the existence of such correlations.

38 • We thank the reviewer for introducing interesting recent work on equivariant normalizing flows. We will add them to
39 the related work section in our revision. We will add comparison to SS to better justify importance sampling using VI.

40 Reviewer 4:

41 • Issues regarding the marginal likelihood estimates are discussed in the beginning.

42 • Yes, $\psi^\sigma > 0$ and this constraint can be removed via the exponential transformation. γ_{x_i}, w_{x_i} are parameters associated
43 with the input variable x_i, and they do not depend on the value of x_i (echoed by Reviewer 1). For phylogenetic trees,
44 the branch lengths $\boldsymbol{q}$ are associated with the edges, and so are the parameters. We will clarify notation accordingly.