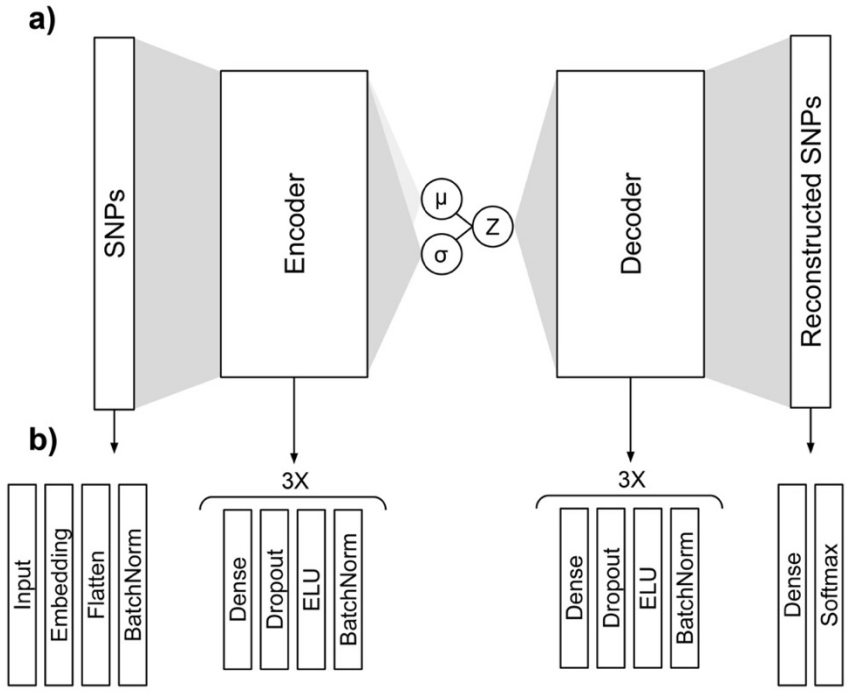

Fig. S1. Diagrammatic representation of the VAE used. a) The full VAE model. b) Detailed break-down of the layers used in the VAE model. A variational autoencoder (VAE) is used to learn a reduced representation of the SNP dataset by minimizing the difference between the input SNPs and the reconstructed SNPs. An embedding layer is used to reduce the one-hot encoded SNPs into a vector size of two. The vector is then flattened, normalized and fed to the encoder layers. The encoder and decoder are both composed of three stacks of layers, each consisting of a fully-connected layer of 100 neurons, with a dropout rate of 0.5, a non-linear ELU activation function and a normalization step. Dropout stochastically removes neurons in a given layer during training according to a dropout probability, which prevents co-dependencies from arising between neurons. This leads to an effectively smaller network that is combined together as an ensemble model during test phase. One drawback is that it is not straightforward when to stop training a VAE. Overtraining a VAE can lead to overfitting the data, which results in clusters that are still present, but the probability distribution over the data is less general, and hence cannot be used reliably for downstream analysis. The dropout (Srivastava et al. 2014) is used to prevent overfitting by deactivating half the neurons during each optimization cycle. Batch normalization (Ioffe and Szegedy 2014) is a technique that rescales values such that they are less sensitive to statistical fluctuations. For the reconstruction, a dense layer with a size proportional to number of SNPs multiplied by 2 is used. Then the Kullback-Leibler divergence (KL) and the difference (the categorical cross-entropy) between the input SNPs and the reconstructed SNPs are minimized. The KL acts as a regularization term to keep the μ at zero and σ at one. The vector is then expanded to four (or number of categories in the one-hot encoded data) followed by a softmax activation function. We trained the VAE by minimizing the variational lower bound objective function (Kingma and Welling 2013) using mini-batch stochastic gradient descent with ADAM updates (Kingma and Ba 2014) with default parameters for 4,000 epochs (generations). Unrelated samples with a higher percentage of missing data might be reconstructed in closer proximity, simply because they share high levels of missing data. This is particularly the case with data converted to one-hot format where a missing SNP was coded as “0,0,0,0”. As such, our implementation of VAE masks and ignores missing data.

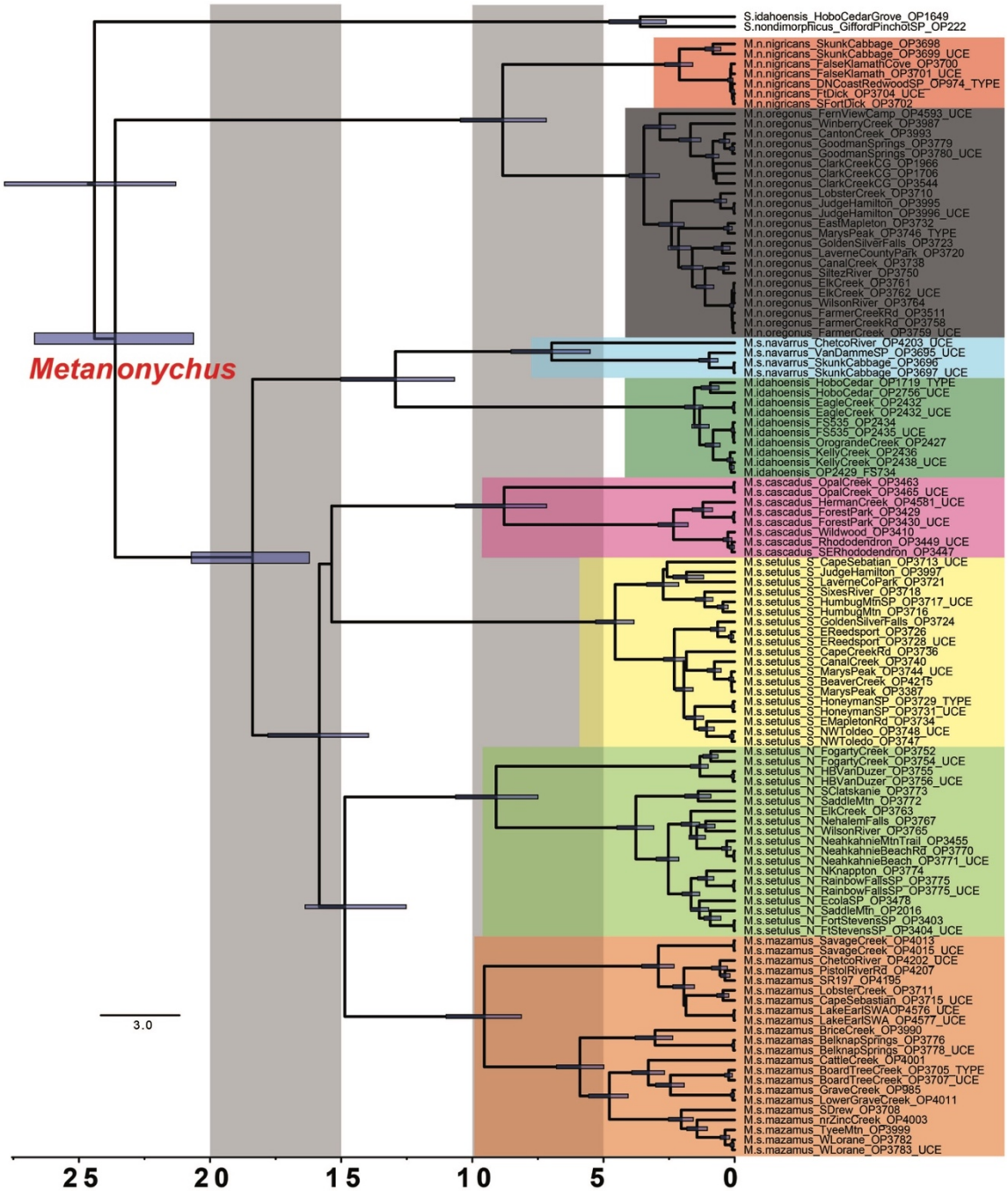

Fig. S2. BEAST COI phylogeny with estimated divergence times. Scale bar in millions of years.

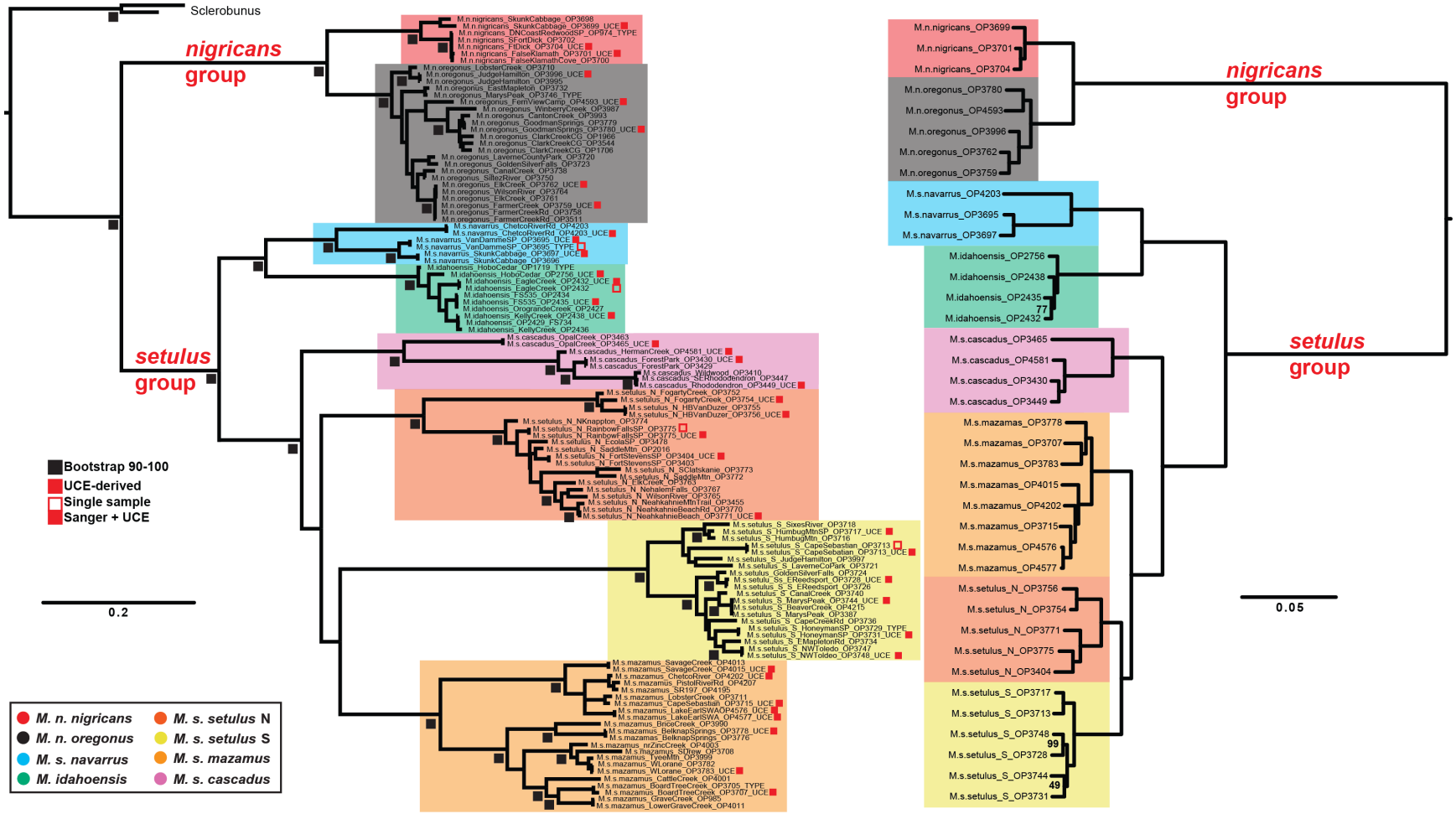

Fig. S3. Left: COI RAxML phylogeny, black boxes at nodes indicate bootstrap support of >90% (note: for simplicity, support is indicated at internal nodes only); terminal taxa with filled red box indicate sequences derived as UCE by-catch, open red boxes indicate sequences obtained via both Sanger + UCE sequencing. Right: UCE 70% RAxML phylogeny, nodes have 100% bootstrap support, unless indicated.

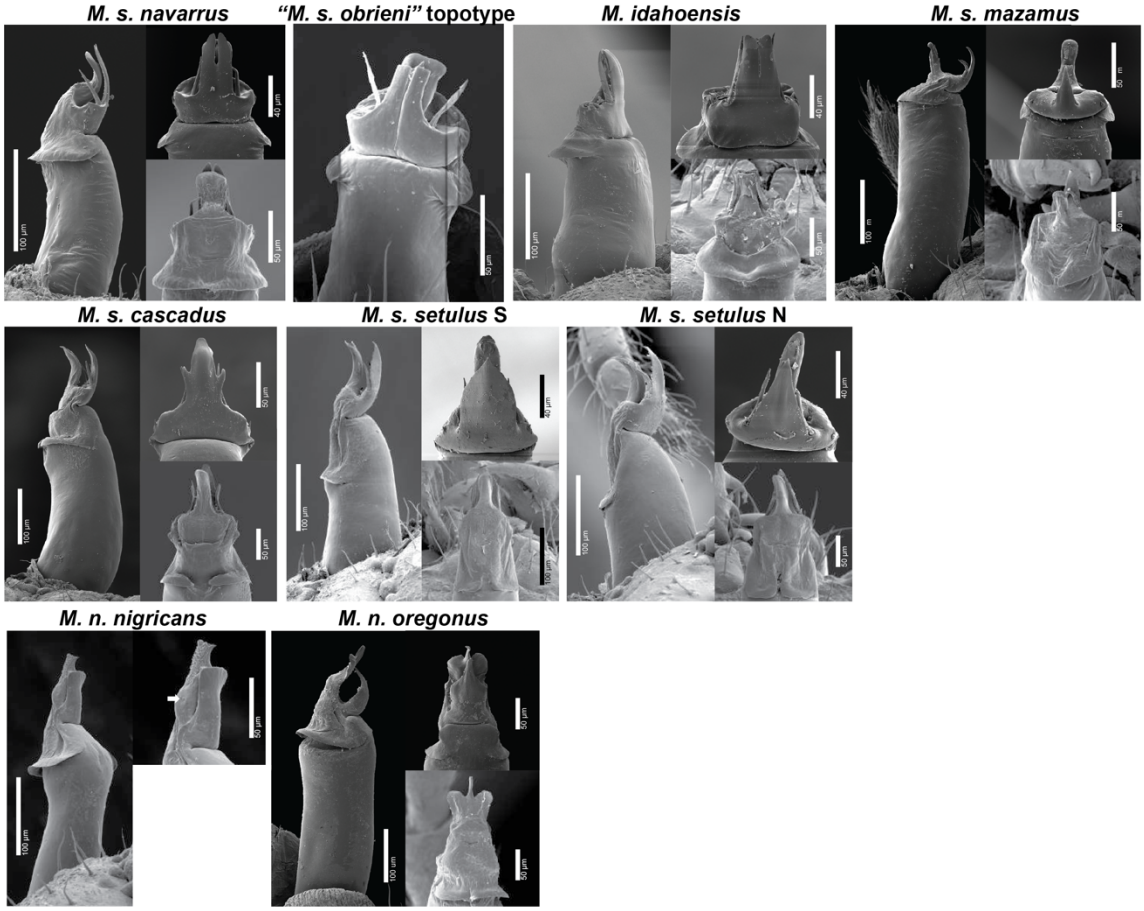

Fig. S4. Representative male genitalic morphology of *Metanonychus* taxa. For each set of images, left pane is lateral view, top right is dorsal view of dorsal plate, bottom right is ventral view of ventral plate. Note: *M. n. nigricans* images are only lateral views. Image of "*M. s. obrieni*" topotype is dorsal view.

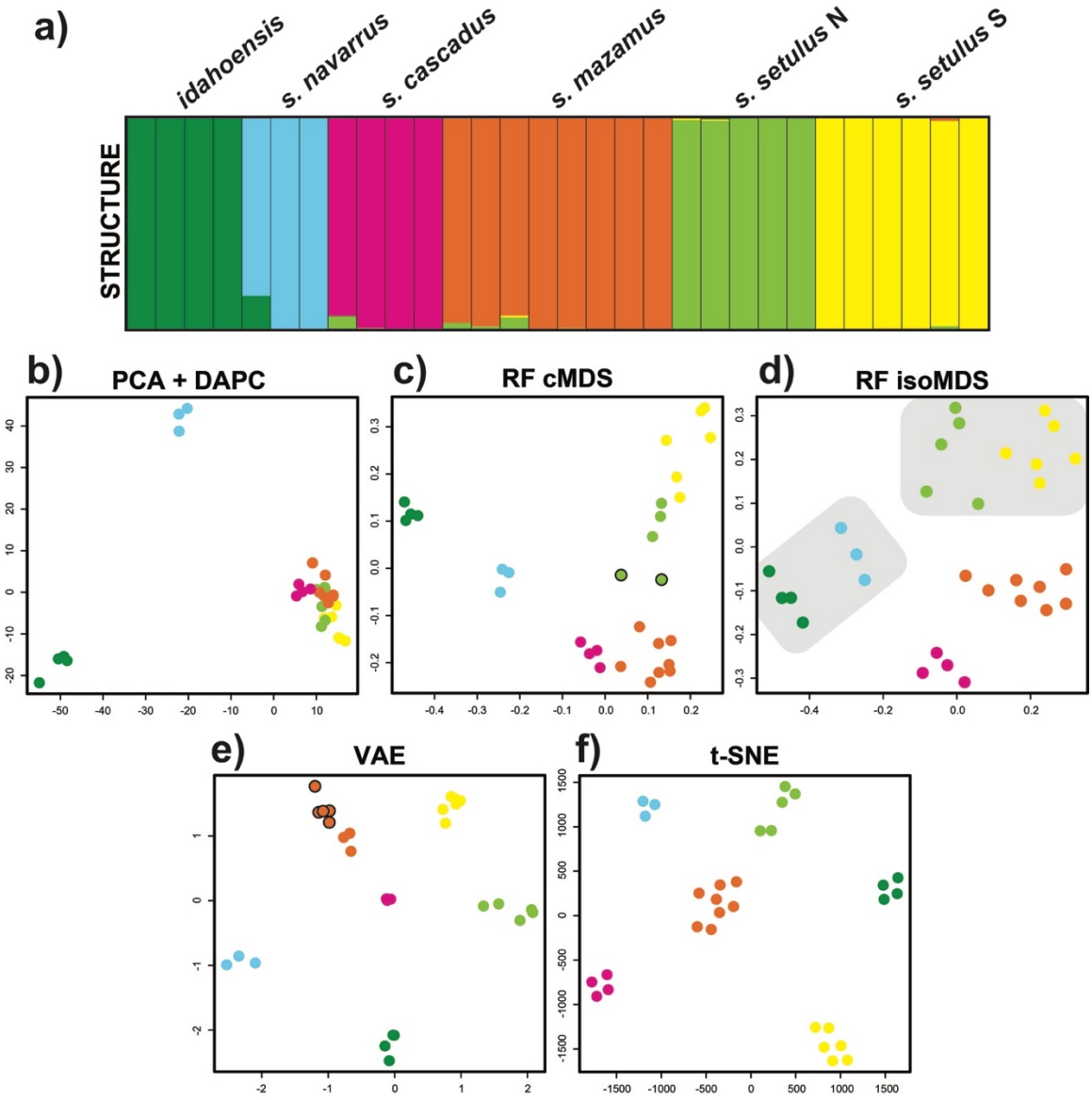

Fig. S5. Clustering results for the *Metanonychus* 50% SNP dataset. a) STRUCTURE plot. b) PCA plot with DAPC clusters. c) random forest cMDS plot, all clustering algorithms favored K=6, except hierarchical clustering with K=7 splitting the northern *setulus* subspecies (seventh cluster indicated with black outline). d) random forest isoMDS plot, PAM clustering favored K=6, except hierarchical clustering of RF output with K=4 (lumped clusters are indicated with grey shading) and K=1 for the gap statistic. e) VAE plot, all clustering algorithms favored K=6, except hierarchical clustering with K=7 splitting *mazamus* (seventh cluster indicated with black outline). f) t-SNE plot, all clustering algorithms favored K=6.

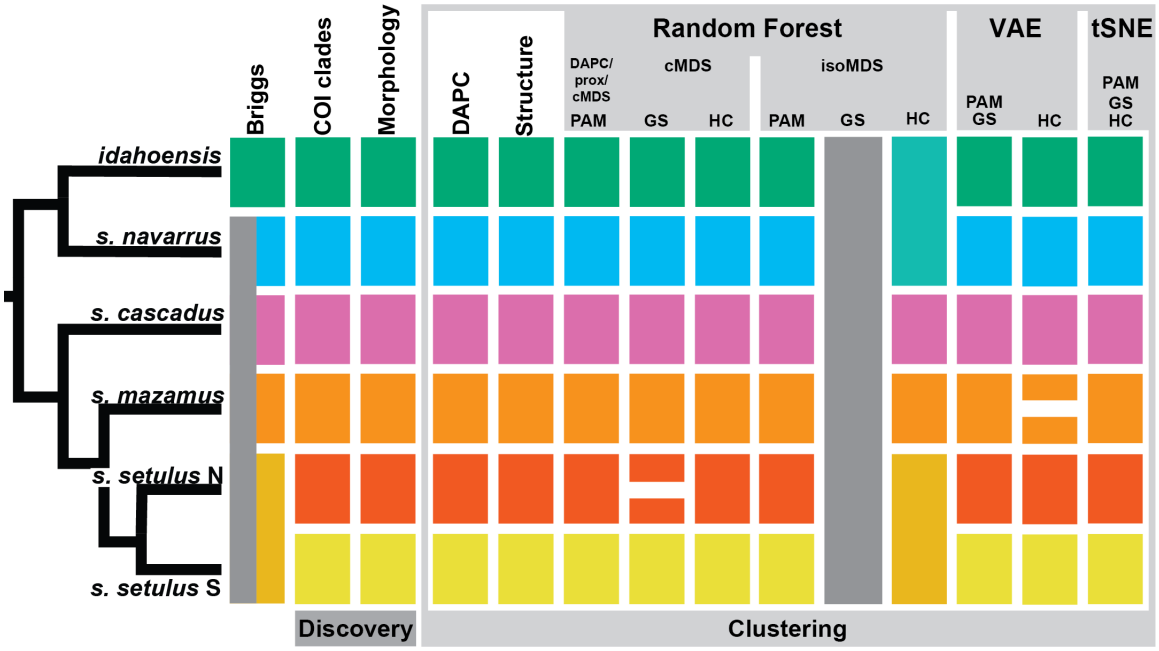

Fig. S6. Integrative species delimitation results for the *Metanonychus* 50% SNP dataset. Species tree at left adapted from RAxML analysis of 70% dataset of UCE loci. cMDS = classic multidimensional scaling, isoMDS = isotonic multidimensional scaling, GS = gap statistic, HC = hierarchical clustering.

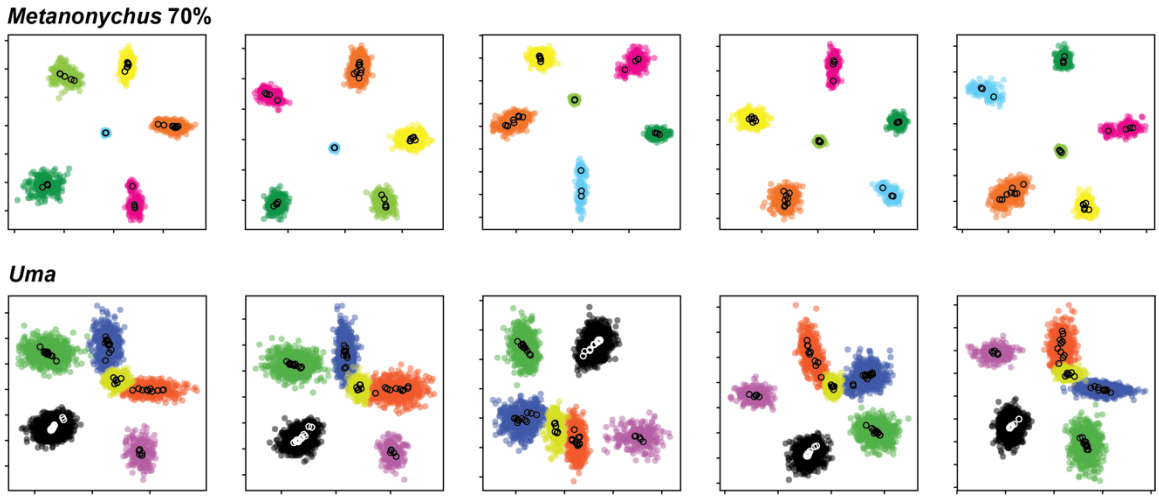

Fig. S7. VAE results with encoded mean (μ – open circles) and standard deviation (σ) for each sample across five replicate runs for the *Metanonychus* 70% SNP matrix (top) and the *Uma notata* dataset (Gottscho et al. 2017; bottom).

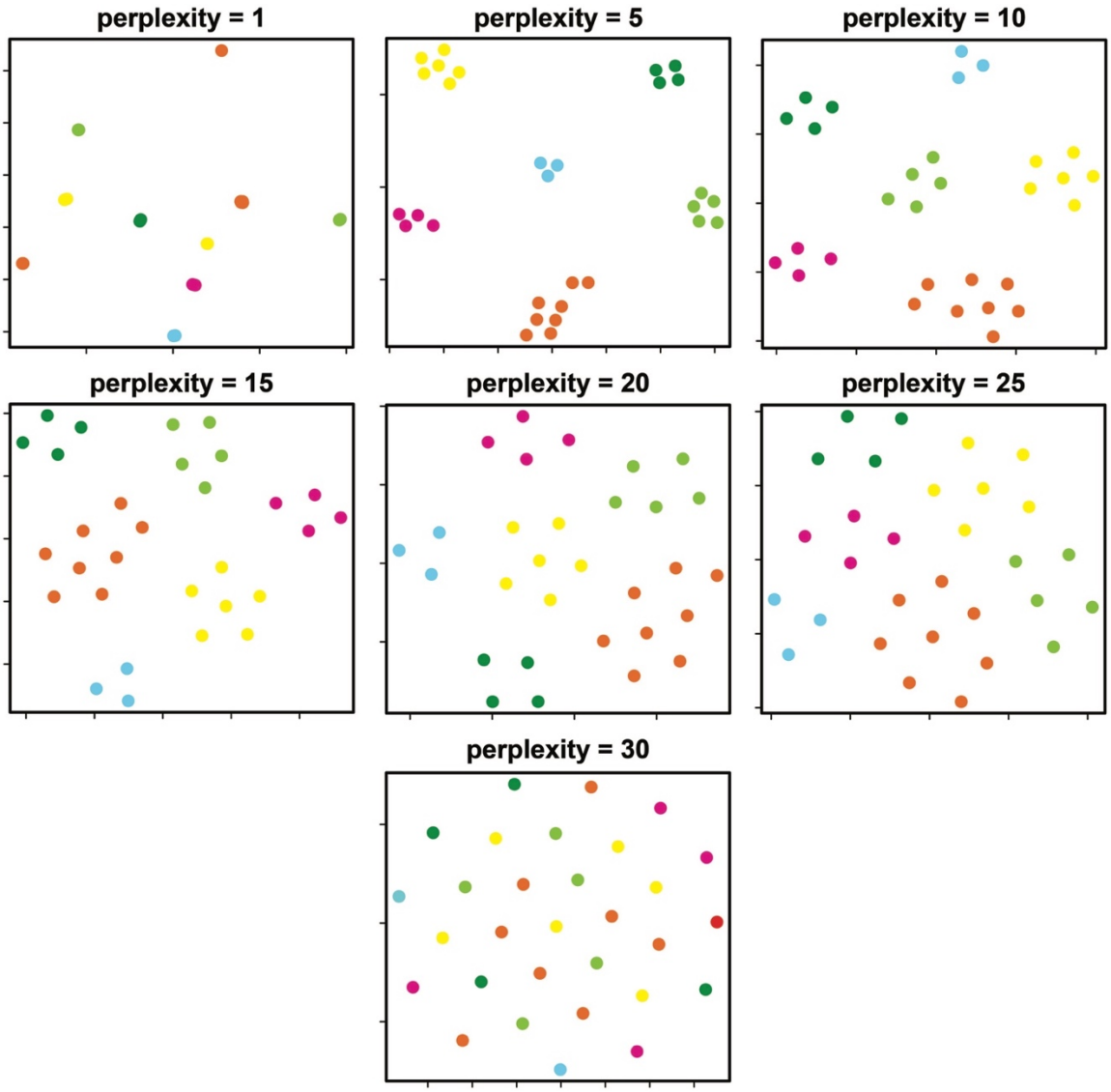

Fig. S8. t-SNE plots with varying perplexity values, using the *Metanonychus* 70% SNP dataset.

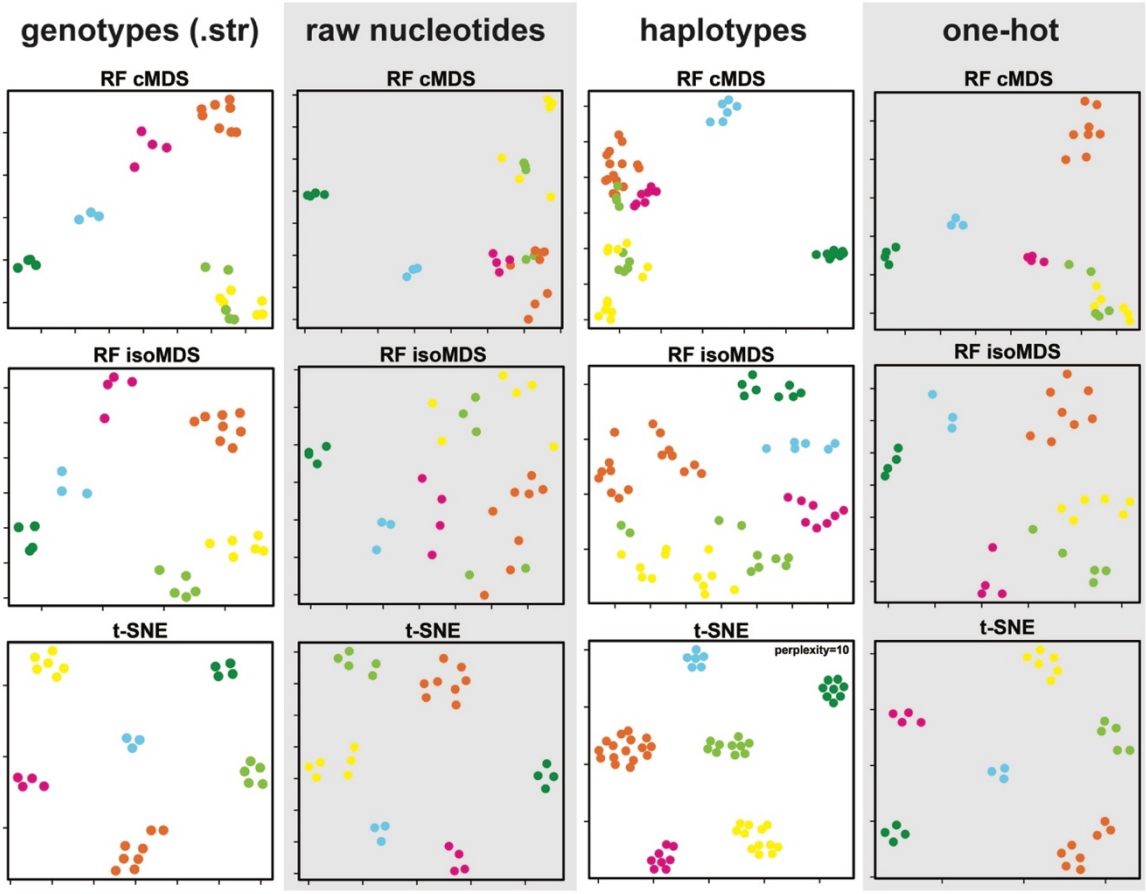

Fig. S9. Resulting plots of *Metanonychus* 70% SNP dataset with input data coded in multiple ways. From left to right: STRUCTURE-formatted file (as in main text), SNPs converted to raw nucleotides with ambiguities, SNP data converted to haplotypes, and SNPs converted via one-hot encoding.

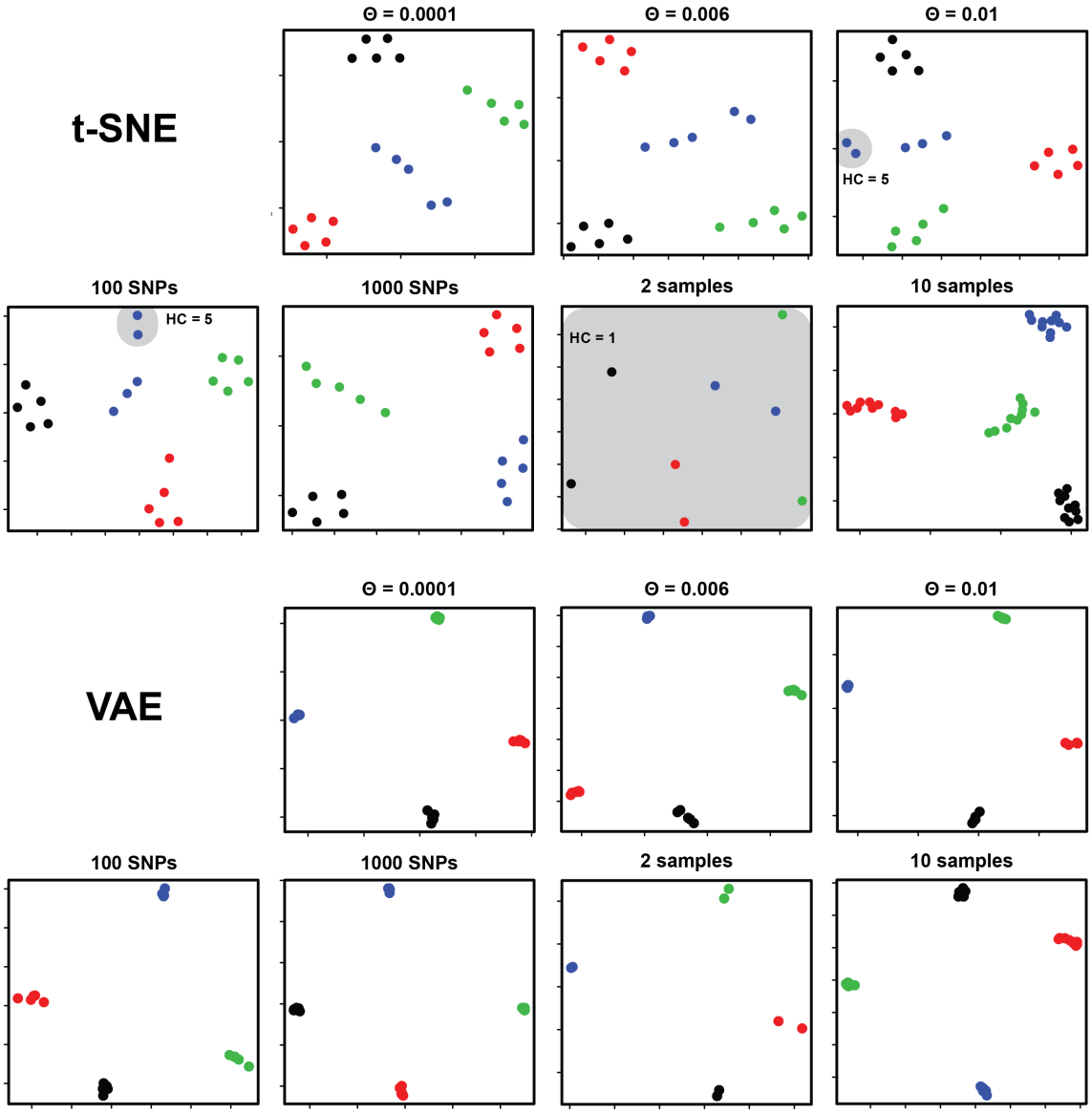

Fig. S10. Results of t-SNE and VAE for the simulated datasets of (Leaché et al. 2014). t-SNE plots with grey shading denote split or lumped clusters identified by hierarchical clustering.

References

Gottscho A.D. et al. Lineage diversification of fringe-toed lizards (Phrynosomatidae: Uma notata complex) in the Colorado Desert: Delimiting species in the presence of gene flow. Mol. Phylogenet. Evol. 106, 103-117 (2017).

Srivastava N., Hinton G., Krizhevsky A., Sutskever I. & Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929-1958 (2014).

Ioffe S. & Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167 [cs.LG] (2015).

Kingma D.P. & Welling M. Auto-encoding variational Bayes. In: Proceedings of the International Conference on Learning Representations (ICLR). arXiv:1312.6114v10 [stat.ML] (2013).

Kingma D.P. & Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980v9 [cs.LG] (2014).

Leaché A.D., Fujita M.K., Minin V.N. & Bouckaert R.R. Species delimitation using genome-wide SNP data. Syst. Biol. 63, 534-542 (2014).