Rove-Tree-11: Supplemental Material

Roberta Hunt[1]https://orcid.org/0000-0003-1963-3281 and Kim Steenstrup Pedersen[1,2]https://orcid.org/0000-0003-3713-0960

[1] Department of Computer Science,
University of Copenhagen, Universitetsparken 1, 2100, Copenhagen, Denmark
{r.hunt,kimstp}@di.ku.dk
[2] Natural History Museum of Denmark, Øster Voldgade 5 - 7, 1350, Copenhagen, Denmark kimstp@snm.ku.dk

1 Rove-Tree-11: Full Phylogenetic Trees

In fig. 1 we show the gold standard genus-level phylogeny for the train, validation and test sets. This phylogeny represents the current state of knowledge as it was pieced together from the most relevant recently published phylogenetic analyses, (from the main text): "such as [1] for sister-group relationships among all three subfamilies and the backbone topology of Xantholininae and Staphylininae, [2] for the subtribe Staphylinina, [3] for the subtribe Philonthina and [4] for the subfamily Paederinae. In this work we did not complete species-level phylogeny, so each species within the genera is assumed to be equally related."

In figs. 2 and 4 we show the species-level phylogenies produced by our model which performed best on the test set (multisimilarity, seed 2) on the validation set, and the test set, respectively. The best model gave an Align Score of 3.5 when compared against the gold standard (4.1 on average over 5 runs). The align score for each individual node is shown to provide some insight into the align score.

From figs. 2 and 4 we can easily see that most species and specimens in the dataset are mostly grouped close to those of the same genus, despite not telling the model these species are related. Which is encouraging. And using the align score as an indicator, we can see that many groups are well organized, however, there is still plenty of room for improvement.

With this visual comparison we can conclude that the model is learning some interesting phylogenetic features, but there is significant room for improvement, making this an interesting dataset for further research.

2 Species-level data distribution

We also show a species-level data distribution in fig. 6. From this we can see that the data is not uniformly distributed per species, with the largest group, rugilus orbiculatus having 262 specimens and the smallest, lathrobium castaneipenne, having 4. This of course may negatively affect our results and a uniform distribution would be preferred.

3 Expanded Results

Further results from experiments are shown in table 1 and expands on the results presented in the paper. Showing R1 scores and results on the Cars196 dataset for easy reference. It should be noted that the Cars196 data was taken directly from [5].

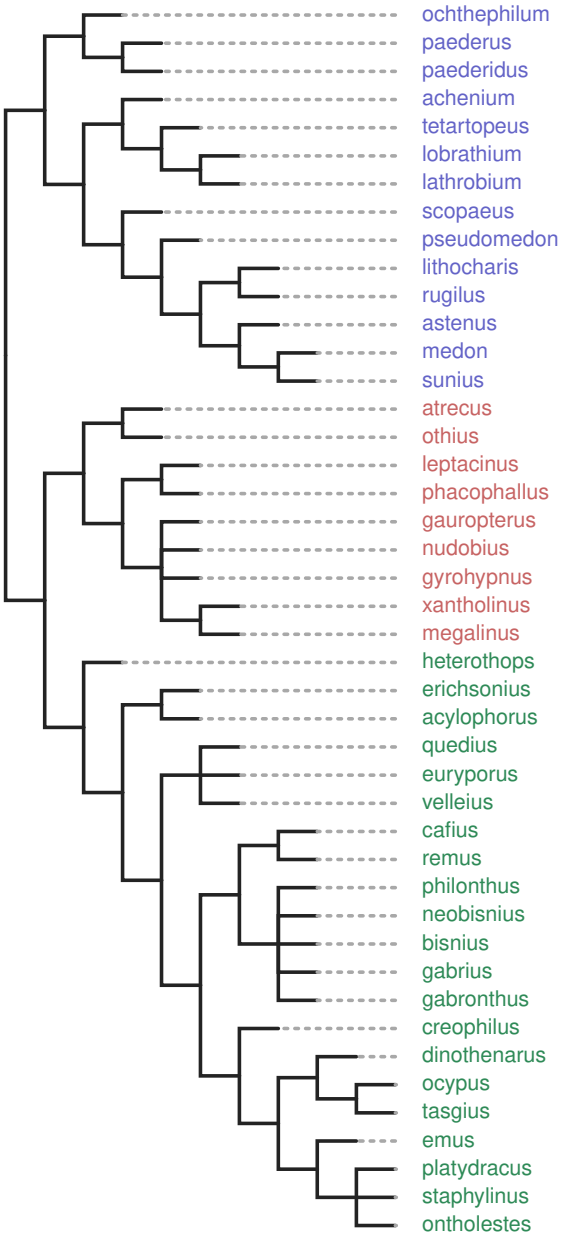

Fig. 1. Gold standard genera-level phylogenetic tree. Tree is split into training set (green), validation set (blue) and test set (red). Genera-level is used here instead of species level to make the tree more compact.

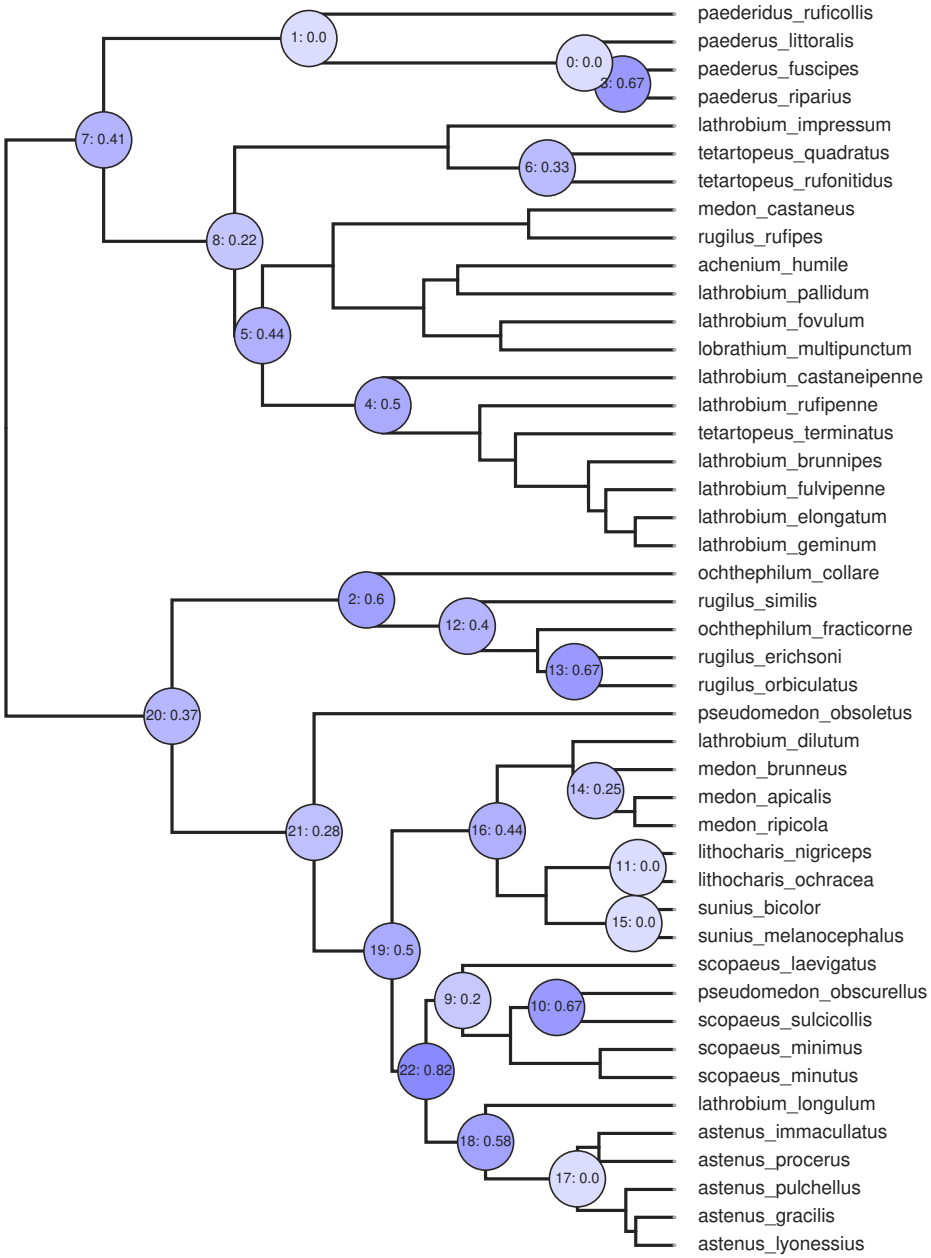

Fig. 2. Species-level phylogenetic tree produced by our best model for the validation set. The number on each node represents the align score of that node to it's matched node in the ground truth phylogeny (the format for this is [node number]:[align score]). Node numbers for the ground truth phylogeny on the validation set are provided in fig. 3. The shade of the node corresponds to the value of the align score - darker shades have higher (undesirable) align scores. The total score for this tree (from summing the score of each node) is 8.35.

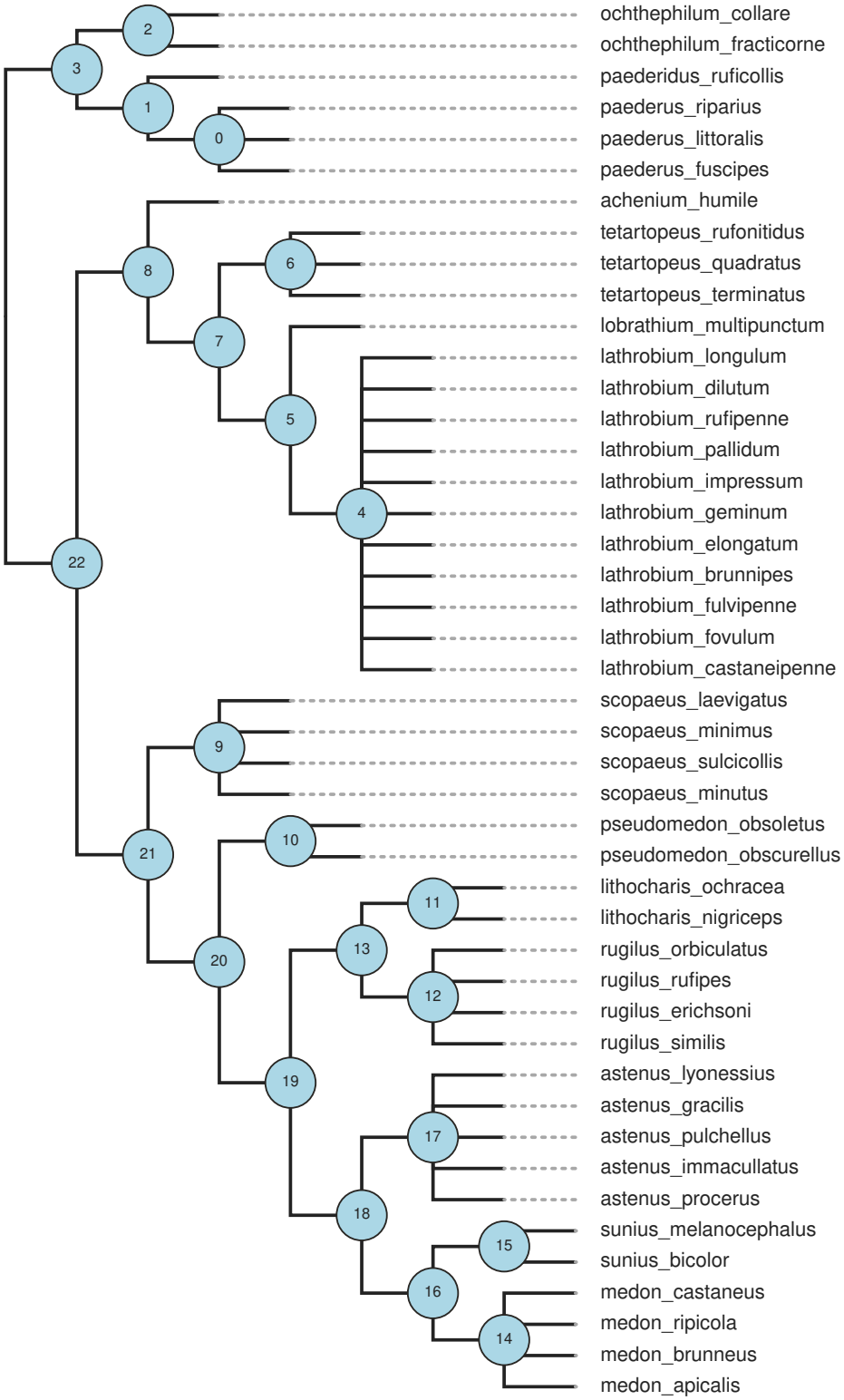

Fig. 3. Species-level ground truth phylogenetic tree for the validation set. The number on each node represents the number of that node, to link it to the align scores presented in fig. 2, which are somewhat synonymous to the most similar nodes in the other tree.

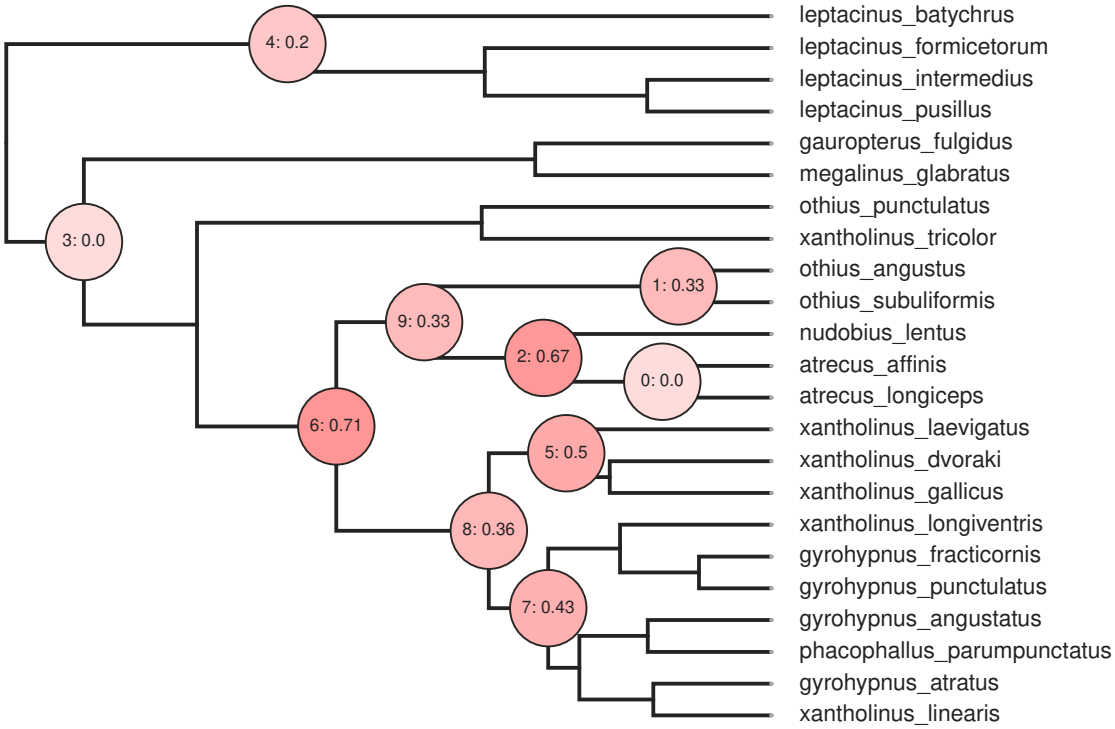

Fig. 4. Species-level phylogenetic tree produced by our best model for the test set. he number on each node represents the align score of that node to it's matched node in the ground truth phylogeny (the format for this is [node number]:[align score]). Node numbers for the ground truth phylogeny on the validation set are provided in fig. 5. The shade of the node corresponds to the value of the align score - darker shades have higher (undesirable) align scores. The total score for this tree (from summing the score of each node) is 8.35.

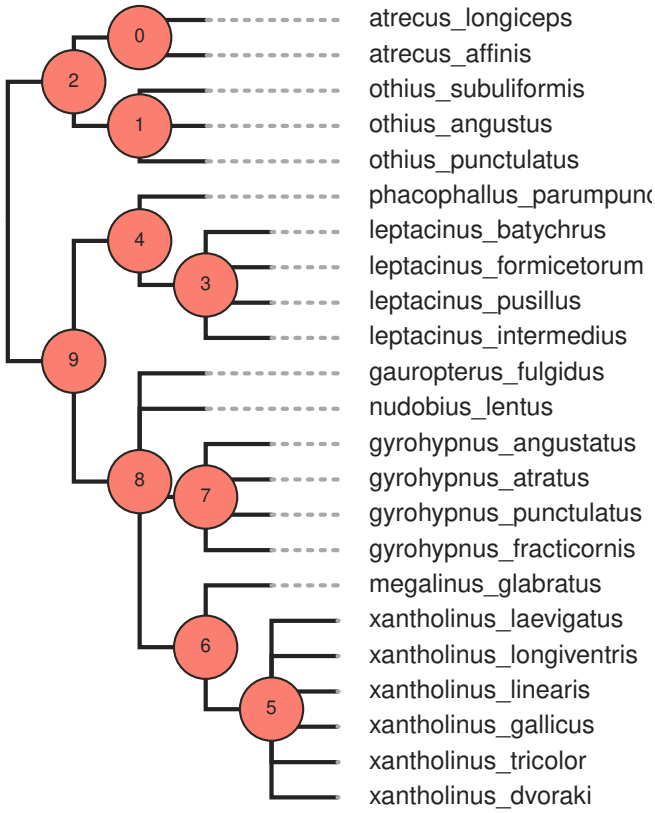

Fig. 5. Species-level ground truth phylogenetic tree for the test set. The number on each node represents the number of that node, to link it to the align scores presented in fig. 4, which are somewhat synonymous to the most similar nodes in the other tree.

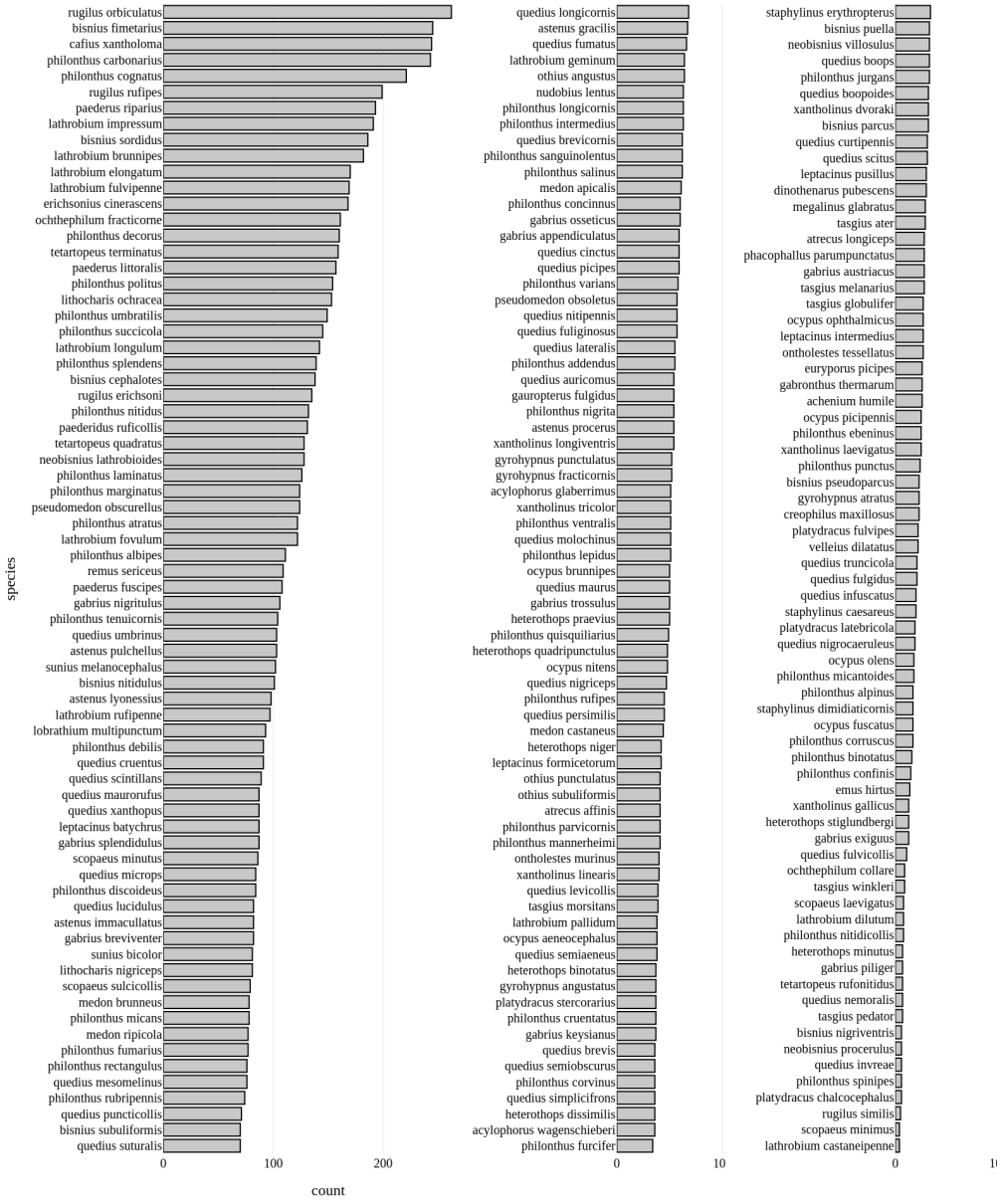

Fig. 6. Species-level data distribution. Here we can more clearly see that the images per species are not uniformly distributed, with Rugilus orbiculatus having 262 specimens and Lathrobium castaneipenne having 4.

Table 1. Benchmark clustering and Align-Score results on Rove-Tree-11 dataset. 'Random' represents the average align score of 5 randomly generated trees. This gives us a metric to compare our results with. A perfect align score would be 0. 95% confidence errors are provided based on 5 runs.

	Cars196 Test		CUB200 Test			Rove-Tree-11 Validation			Rove-Tree-11 Test		
Loss	NMI	R1	NMI	R1	Align	NMI	R1	Align	NMI	R1	Align
Random	-	-	-	-	21.9±0.2	-	-	15.8±0.9	-	-	6.6±0.5
Triplet	65.9±0.2	79.1±0.3	64.8±0.5	59.4±0.4	9.9±0.9	68.9±0.4	86.3±0.5	**7.8±1.1**	66.3±0.3	84.0±0.6	4.1±0.5
Margin	65.3±0.3	77.7±0.3	60.7±0.3	54.7±0.2	10.6±1.2	68.0±0.7	84.4±0.6	8.2±0.7	65.9±0.5	82.8±0.8	4.2±0.7
Lifted	72.2±0.4	63.8±0.4	34.8±3.0	23.5±5.2	15.9±2.0	55.0±0.6	63.3±1.2	10.5±0.7	56.0±1.1	58.8±0.8	4.9±0.8
Constrast.	64.0±0.1	75.8±0.4	59.0±1.0	53.6±0.8	11.0±1.2	66.7±0.5	82.3±0.6	8.5±1.0	65.4±0.5	80.8±1.4	4.5±0.6
Multisim.	**81.7±0.2**	69.4±0.4	**68.2±0.3**	62.9±0.6	**8.6±0.8**	**70.7±0.2**	**88.0±0.3**	8.2±0.4	**67.3±0.5**	**86.8±0.7**	**4.0±0.5**
ProxyNCA	78.5±0.6	65.8±0.2	66.8±0.4	**63.0±0.4**	9.8±0.8	67.5±0.7	83.7±0.8	9.0±0.8	65.5±0.3	82.1±0.9	4.2±0.4
Arcface	67.0±1.1	**79.2±1.0**	67.5±0.4	**63.0±0.7**	9.8±0.8	66.9±0.9	82.8±0.9	8.5±0.4	64.8±0.5	79.5±1.0	4.1±0.4

References

1. Żyła, D., Solodovnikov, A.: Multilocus phylogeny defines a new classification of staphylininae (coleoptera, staphylinidae), a rove beetle group with high lineage diversity. Systematic Entomology **45** (2020) 114–127
2. Brunke, A., Smetana, A.: A new genus of staphylinina and a review of major lineages (staphylinidae: Staphylininae: Staphylinini). Systematics and Biodiversity **17** (2019) 745–758
3. Chani-Posse, M.R., Brunke, A.J., Chatzimanolis, S., Schillhammer, H., Solodovnikov, A.: Phylogeny of the hyper-diverse rove beetle subtribe philonthina with implications for classification of the tribe staphylinini (coleoptera: Staphylinidae). Cladistics **34** (2018) 1–40
4. Żyła, D., Bogri, A., Hansen, A., Jenkins Shaw, J., Kypke, J., Solodovnikov, A.: A new termitophilous genus of paederinae rove beetles (coleoptera, staphylinidae) from the neotropics and its phylogenetic position. Neotropical Entomology (2022)
5. Roth, K., Milbich, T., Sinha, S., Gupta, P., Ommer, B., Cohen, J.P.: Revisiting training strategies and generalization performance in deep metric learning (2020)