

phruta: scrapping genbank and assembling phylogenetic trees *

Cristian Román-Palacios *School of Information, University of Arizona, Tucson, Arizona 85721, USA. ORCID : 0000-0003-1696-4886*

Current methodological practices for assembling phylogenetic trees often recur to sequence data stored in GenBank. However, understanding molecular and taxonomic availability in GenBank is generally not very straightforward. For instance, the genetic makeup of datasets available in GenBank can strongly differ between genera even within the same family. Similarly, the taxonomic information associated with sequence data in GenBank can be outdated, relative to other databases that mainly focus on the taxonomic side. *phruta*, a newly developed R package, is designed to improve the user experience and access to information to genetic data stored in GenBank. By using *phruta*, users are able to (1) quantitatively explore the molecular makeup of particular clades with information in GenBank, (2) assemble curated multi-gene molecular datasets with retrieved and local sequences, and (3) run basic phylogenetic tasks, all within R. The structure of the functions implemented in *phruta*, designed as a workflow, aim to allow users to assemble simple workflows for particular tasks, which are in turn expected to increase reproducibility when assembling phylogenies. This paper provides a brief overview on the performance and workflow associated with *phruta*.

Keywords: R package, Phylogenetics, Reproducibility, Workflow

Background

The phruta R package

The *phruta* package is designed to simplify the basic phylogenetic pipeline in R. *phruta* is expected to allow scientists from different backgrounds to assemble molecular databases or phylogenies for particular taxa with as minimal complexity and maximal reproducibility as possible. All the code in *phruta* is run within the same software (R) and data from intermediate steps are either stored to the environment or can be exported locally to different folders. In general, *phruta* is able to (1) find potentially (phylogenetically) relevant gene regions for a given set of taxa based on GenBank, (2) retrieve gene sequences and curate taxonomic information from the same database, (3) combine downloaded and local gene sequences, and (4) perform sequence alignment, phylogenetic inference, and basic tree dating tasks.

*Replication files are available on the author's Github account (<http://github.com/cromanpa>). **Current version:** September 03, 2022; **Corresponding author:** cromanpa94@arizona.edu.

Alternatives to phruta

Similar functionalities for assembling curated molecular datasets for phylogenetic analyses can be found in `phylotaR` and `SuperCRUNCH`. However, `phylotaR` is limited to downloading and curating sequences (e.g. it does not align sequences). Similarly, `SuperCRUNCH` only curates sequences that are already stored locally. In fact, `phruta` is closer to `SUPERSMART` and the associated R workflow `SUPERSMARTR`. However, most of the applications in the different packages that are part of `SUPERSMARTR` are simplified in `phruta`. Standalone applications that might resemble `phruta` could include `MEGA` and `geneious`. However, analyses in these two alternatives are either poorly reproducible (e.g. `MEGA`) or not all the functions are freely available to everyone (e.g. `geneious` has a paid version).

phruta in a nutshell

The current release of `phruta` includes a set of eight major functions. Running all eight major functions in `phruta` results in a time-calibrated phylogeny. However, users interested in using their own files at any stage can run each function independently. Note that all the functions for which their primary output are sequences (aligned or unaligned) are listed under `sq.*`. All the files that output phylogenies (time-calibrated or not) are listed under `tree.*`.

- First, the distribution of gene sampled for a given organism or set of taxa can be explored using the `acc.gene.sampling` function. This function will return a table that summarizes either the distribution of genes sampled for the search term in general or specifically across species.
- Second, given a list of target organisms, users can retrieve a list of accession numbers that are relevant to their search using `acc.table.retrieve()`. Instead of directly downloading sequences from `genbank` (see `sq.retrieve.direct()` below), retrieving accession numbers allow users to have more control over the sequences that are being used in the analyses. Note that users can also curate the content of the dataset obtained using `sq.retrieve.direct()`.
- Third, users should download gene sequences. Sequences can be download using the `sq.retrieve.indirect()` from the accession numbers retrieved before using the `acc.table.retrieve()` function. This is the preferred option within `phruta`. Additionally, users can directly download gene sequences using the `sq.retrieve.direct()` function. Both `sq.retrieve.indirect()` and `sq.retrieve.direct()` functions save gene sequences in `fasta` files that will be located in a new directory named `0.Sequences`.
- Fourth, `sq.add()` allows users to include local sequences to those retrieved from `genbank` in the previous step. This function saves all the resulting `fasta` files in two directories, combined sequences in `0.Sequences` and local sequences in `0.AdditionalSequences` (originally downloaded sequences are moved to `0.0.OriginalDownloaded` at this step). Note that `sq.add()` is optional.

- Fifth, the `sq.curate()` function filters out unreliable sequences based on information listed in genbank (e.g. PREDICTED) and on taxonomic information provided by the user. Specifically, this function retrieves taxonomic information from the Global Biodiversity Information Facility (GBIF) database's taxonomic backbone (see alternatives in the advanced vignette to `phruta`). If a given species belongs to a non-target group, this species is dropped from the analyses. This function automatically corrects taxonomy and renames sequences.
- Sixth, `sq.aln()` performs multiple sequence alignment on fasta files. Currently, `phruta` uses the DECIPHER R package, here. This package allows for adjusting sequence orientation and masking (removing ambiguous sites).
- Seventh, the `tree.raxml()` function allows users to perform tree inference under RAxML for sequences in a given folder. This is a wrapper to `ips::raxml()` and each of the arguments can be customized. The current release of `phruta` can manage both partitioned and unpartitioned analyses. Starting and constrained trees are allowed.
- Eight, `tree.dating()` enables users to perform time-calibrations of a given phylogeny using `geiger::congruify.phylo()`. `phruta` includes a basic set of comprehensively sampled, time-calibrated phylogenies that are used to extract secondary calibrations for the target phylogeny. Note that sampling in those phylogenies can be examined using `data(SW.phruta)`. Please make sure you have at least **two** groups in common with each of the phylogenies. Similarly, users can choose to run either `PATHd-8` or `treePL`.

Assembling a molecular dataset for target taxa in phruta

Let's learn how `phruta` works by assembling a molecular dataset at the species level for a few bird clades. Note that although this tutorial is based on a particular set of taxa, users can decide on what can choose their target clades in other families, orders, or even kingdoms.

Here we will focus on assembling a phylogeny for the new world Quails. Species in this group are classified in the family Odontophoridae, a clade including nearly 34 extant species classified in 10 genera (??fig:quail)). In general, the higher-level taxonomic information in GenBank for the Odontophoridae is largely congruent with recent studies on the systematics of the group. However, GenBank's the taxonomic backboneone classifies *Ptilopachus*, a genus commonly included under the Odontophoridae, as part of the Phasianidae. We will follow more recent studies suggesting that *Ptilopachus* is nested within the Odontophoridae. As our outgroup, we will select the Phasianidae. Within this clade, we will explicitly focus on the genus *Polyplectron*, a clade of eight extant species. Finally, given that the systematics of the Odontophoridae has been discussed before using morphological and molecular evidence, we will be able to compare the topology of our tree relative to recent studies (Crowe et al. 2006a, b; Cohen et al. 2012; Hosner et al. 2015).

```
library(phruta)
```

So far, we have decided the taxonomic makeup of our analyses. From this point, we could simply check the genetic sampling of previous studies and search for those genes in GenBank for the target taxa (Crowe et al. 2006a, b; Cohen et al. 2012; Hosner et al. 2015). For instance, [review sampling in each of those]. We could use these gene names to assemble a molecular dataset for the Odontophoridae and **Polyplectron** in phruta.

Alternatively, we could use phruta to figure out what genes are well sampled in GenBank for both the ingroup and outgroup. For simplicity, we will follow the later procedure in this paper implemented in `gene.sampling.retrieve()` function in phruta. The resulting `data.frame`, here named `gs.seqs`, will contain the list of full names for genes sampled in GenBank for the target taxa.

```
gs.seqs <- gene.sampling.retrieve(organism = c("Odontophoridae",  
      "Ptilopachus", "Polyplectron"), speciesSampling = TRUE)
```

For the search terms, phruta was able to retrieve the names for 79 gene regions from GenBank. The frequency estimates per gene are based on inter-specific sampling (??tab:topGenes)). Note that the `gene.sampling.retrieve()` function provides an estimate of the number of species in GenBank that match the taxonomic criteria of the search term and that have sequences for a given gene region. However, this estimate is only as good as the annotations for genes deposited in GenBank.

We will now generate a preliminary summary of the accession numbers retrieved for the combination of target taxa and gene regions. I call this a preliminary dataset because not all these accession numbers are expected to be in the final molecular dataset. For instance, some sequences may be removed after taxonomic synonyms are identified in the dataset.

Next, we will assemble a species-level summary of accession numbers using the `acc.table.retrieve` function in phruta (i.e. `speciesLevel = TRUE` argument). For simplicity, this tutorial will focus on analyzing gene regions that are sampled in >20% of the species (`targetGenes` `data.frame`). The `acc.table` object created below is a `data.frame` object that will later be used to download the relevant gene sequences from GenBank (??tab:AccN)).

```
targetGenes <- gs.seqs[gs.seqs$PercentOfSampledSpecies > 20,  
  ]  
acc.table <- acc.table.retrieve(clades = c("Odontophoridae",  
      "Ptilopachus", "Polyplectron"), genes = targetGenes$Gene,  
      speciesLevel = TRUE)
```

Since we are interested in retrieving sequences from GenBank using an existing preliminary accession numbers table, we will use the `sq.retrieve.indirect()` function in phruta. Please note that there are two versions of `sq.retrieve.*` in phruta. The one that we're using in this tutorial, `sq.retrieve.indirect()`, retrieves sequences "indirectly" because it requires a table of accession numbers to be pass as an argument (see the

`acc.table.retrieve()` function above). I present the information in this tutorial using `sq.retrieve.indirect()` instead of `sq.retrieve.direct()` because `sq.retrieve.indirect()` is simply more flexible and robust to mistakes. Specifically, `sq.retrieve.indirect()` allow users to correct issues *prior* to downloading/retrieving the sequences. For instance, you can add new sequences, species, populations to the resulting `data.frame` from `acc.table.retrieve()`. You could even manually assemble your own dataset of accession numbers to be retrieved using `sq.retrieve.indirect()`. Instead, `sq.retrieve.direct()` does its best to directly retrieve GenBank sequences for a target set of taxa and set of gene regions. In short, you should be able to catch errors using `sq.retrieve.indirect()` but mistakes will be harder to spot and fix if you're using `sq.retrieve.direct()`.

We still need to retrieve all the sequences from the accessions table generated using `acc.table`. Note that since we have specified `download.sqs = FALSE` in `sq.retrieve.indirect`, the sequences retrieved from GenBank are returned in a list. If we decide to download the sequences to our working directory using `download.sqs = TRUE`, `phruta` will write all the resulting fasta files into a newly created folder 0. Sequences located in our working directory.

```
sqs.downloaded <- sq.retrieve.indirect(acc.table = acc.table,  
  download.sqs = FALSE)
```

Now, let's make sure that we are only including sequences that are reliable and from species that we are actually interested in analyzing. We are going to use the `sq.curate()` function for this. We will provide a list of taxonomic names to filter out incorrect sequences (`filterTaxonomicCriteria` argument). For instance, we could simply provide a vector of the genera that we are interested in analyzing. This vector must have a length of 1, with all the target genera being separated with `|` (e.g. `"Callipepla|Colinus|Dendrortyx"` if we were interested in only those three genera). For now, we will assume that all of the species we downloaded are relevant to the analyses (i.e. `filterTaxonomicCriteria = [AZ]`). Finally, since we are not downloading anything to our working directory, we need to pass our downloaded sequences (`sqs.downloaded` object generated above using the `sq.retrieve.indirect()` function) to the `sqs.object` argument in `sq.curate()`.

```
sqs.curated <- sq.curate(filterTaxonomicCriteria = "[AZ]", kingdom = "animals",  
  sqs.object = sqs.downloaded, removeOutliers = FALSE)
```

Running the `sq.curate()` function will create an object of class `list` (i.e. `sqs.curated`) that includes (1) the curated sequences with original names, (2) the curated sequences with species-level names (`renamed_*` prefix), (3) the accession numbers table (`AccessionTable; ??tab:tw`)), and (4) a summary of taxonomic information for all the species sampled in the files (`??tab:tw2`, `??tab:tw3`)). From here, we will align the sequences that we just curated using `sq.aln()` with default parameters. We're again passing the output from `sq.curate()`, `sqs.curated`, using the `sqs.object` argument in `sq.aln()`.

```
sqs.aln <- sq.aln(sqs.object = sqs.curated)
```

The resulting multiple sequence alignments will be saved to the `sqs.aln` object, a list of alignments. For each of the gene regions, we will have access to the original alignment (`Aln.Original`), the masked one (`Aln.Masked`), and information on the masking process. The raw and masked alignments are presented in `??fig:alnraw`) and `??fig:alncur`), respectively.

In total, code in this section took 4 minutes to run in my local machine (RAM: 8.5899346×10^9 , CPU: Intel(R) Core(TM) i5-8257U CPU @ 1.40GHz, cores: 8, platform: x86_64-apple-darwin17.0, R: R version 4.2.0 (2022-04-22)). The performance of *phruta* is largely affected by the complexity of the dataset that is being assembled. On top of that, it is important to remember that multiple functions in *phruta* are directly retrieving data from different websites.

Basic phylogenetics with phruta

Phylogenetic inference with phruta and RAxML

Phylogenetic inference in *phruta* is conducted using the `tree.raxml()` function. To use this function, we will have to export our sequence alignments locally. We will follow the same folder structure as if we were exporting everything locally ([NEED FIGURE]). Specifically, our sequence alignments will be located in `2.Alignments` and we will exclusively export the alignments that were masked.

We are now ready to run RAxML. Note that in `tree.raxml()`, we will need to indicate where the aligned sequences are located (`folder` argument), the patterns of the files in the same folder (`FilePatterns` argument; “Masked_” in our case), and the total of bootstrap replicates. The `outgroup` argument is optional but since we are interested in calibrating our tree afterwards, we will define it using all the species in *Polyplectron*.

```
outgroup <- sqs.curated$Taxonomy[sqs.curated$Taxonomy$genus ==
  "Polyplectron", ]

tree.raxml(folder = "2.Alignments", FilePatterns = "Masked_",
  raxml_exec = "raxmlHPC", Bootstrap = 100, outgroup = paste(outgroup$species_names,
  collapse = ","))
```

```
## [1] "raxmlHPC -T 4 -f a -p 1234 -x 1234 -m GTRGAMMA -o Polyplectron_inopinatum,Polyp1"
```

The resulting trees are saved in the `3.Phylogeny` folder. For many, the bipartitions tree generated in these runs, `RAxML_bipartitions.phruta`, will be the most relevant. Additionally, `3.Phylogeny` includes RAxML-related input and output files. The resulting phylogeny from these analyses is presented in `??fig:raxmlphylo`).

Users can also run partitioned analyses in RAxML within *phruta*. This approach is implemented by setting the `partitioned` argument in `tree.raxml` to `TRUE`. For now, partitions are based on the genes being analyzed. The same model is used to analyze each

partition. More details on partitioned analyses can be customized by passing arguments in `ips::raxml`.

```
tree.raxml(folder = "2.Alignments", FilePatterns = "Masked_",
  raxml_exec = "raxmlHPC", Bootstrap = 100, outgroup = paste(outgroup$species_names,
  collapse = ","), partitioned = TRUE)
```

Finally, we note that users will sometimes need to generate constrain trees to seed their searches or limit the tree sampling space. These constraints can be generated using other software including TACT: Taxonomic Addition for Complete Trees.

Tree dating in phruta

Finally, let's perform tree dating in our phylogeny using secondary calibrations extracted from Scholl and Wiens (2016). I am only using this study because it has a large phylogeny but I expect to replace it in the near future. Note that `tree.dating()` requires the user to specify where the `1.Taxonomy.csv` file is. This file is created automatically when sequences are curated using `sq.curate()` and results are exported to your local repository. However, since we were keeping results in the environment, we will have to export it before we can move forward.

```
dir.create("1.CuratedSequences")
write.csv(sqs.curated$Taxonomy, "1.CuratedSequences/1.Taxonomy.csv")
```

Tree dating is performed using the `tree.dating()` function in `phruta`. We have to provide the name of the folder containing the `1.Taxonomy.csv` file created in `sq.curate()`. We also have to indicate the name of the folder containing the `RaxML_bipartitions.phruta` file. We will scale our phylogeny using `treePL`.

```
tree.dating(taxonomyFolder = "1.CuratedSequences", phylogenyFolder = "3.Phylogeny",
  scale = "treePL")
```

Running this line will result in a new folder `4.Timetree`, including the different time-calibrated phylogenies obtained (if any) and associated secondary calibrations used in the analyses. The resulting time-calibrated tree is presented in `fig:timecaltree`.

Advanced methods with phruta

Curating taxonomic names

You can use `taxonomy.retrieve()`, a function implemented inside `sq.curate()` in `phruta` to curate species names regardless of the kingdom. For instance, the block of code below will curate taxonomic names using the `gbif` taxonomic backbone. Note that the `kingdom` argument in `taxonomy.retrieve()` can be set to `NULL`, meaning that there won't be indication on the kingdom when performing taxonomic searches.

```
phruta::taxonomy.retrieve(species_names = c("Felis_catus", "PREDICTED:_Vulpes",
      "Phoca_largha", "PREDICTED:_Phoca", "PREDICTED:_Manis", "Felis_silvestris",
      "Felis_nigripes"), database = "gbif")
```

	kingdom	phylum	class	order	family	genus	species
## 1	Animalia	Chordata	Mammalia	Carnivora	Felidae	Felis	Felis catus
## 2	incertae sedis	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
## 3	Animalia	Chordata	Mammalia	Carnivora	Phocidae	Phoca	Phoca largha
## 4	incertae sedis	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
## 5	incertae sedis	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
## 6	Animalia	Chordata	Mammalia	Carnivora	Felidae	Felis	Felis silvestris
## 7	Animalia	Chordata	Mammalia	Carnivora	Felidae	Felis	Felis nigripes

However, gbif is efficient for retrieving accurate taxonomy when we provide details on the kingdom. Given that all the species we're interested in are animals, we could just use the following block of code to curate taxonomic names.

```
phruta::taxonomy.retrieve(species_names = c("Felis_catus", "PREDICTED:_Vulpes",
      "Phoca_largha", "PREDICTED:_Phoca", "PREDICTED:_Manis", "Felis_silvestris",
      "Felis_nigripes"), database = "gbif", kingdom = "animals")
```

	kingdom	phylum	class	order	family	genus	species
## 1	Animalia	Chordata	Mammalia	Carnivora	Felidae	Felis	Felis catus
## 2	incertae sedis	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
## 3	Animalia	Chordata	Mammalia	Carnivora	Phocidae	Phoca	Phoca largha
## 4	incertae sedis	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
## 5	incertae sedis	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
## 6	Animalia	Chordata	Mammalia	Carnivora	Felidae	Felis	Felis silvestris
## 7	Animalia	Chordata	Mammalia	Carnivora	Felidae	Felis	Felis nigripes

Depending on your sampling, you could also do the same for plants by using plants in the kingdom argument instead of animals. Now, what if we were interested in following other databases to retrieve taxonomic information for the species in our database? The latest version of phruta allow users to select the desired database. The databases follow the taxize::classification() function. Options are: ncbi, itis, eol, tropicos, nbn, worms, natserv, bold, wiki, and pow. Please select only one. Note that the gbif option in taxize::classification() is replaced by the internal gbif in phruta.

```
phruta::taxonomy.retrieve(species_names = c("Felis_catus", "PREDICTED:_Vulpes",
      "Phoca_largha", "PREDICTED:_Phoca", "PREDICTED:_Manis", "Felis_silvestris",
      "Felis_nigripes"), database = "itis")
```


Running PartitionFinder in phruta

With the current version of *phruta*, users are able to run *PartitionFinder* v1 within R. For this, users should provide the name of the folder where the alignments are stored, a particular pattern in the file names (*Masked_* in our case), and which models will be run in *PartitionFinder*. This function will download *PartitionFinder*, generate the input files, and run it all within R. The output files will be in a new folder within the working directory.

```
sq.partitionfinderv1(folderAlignments = "2.Alignments", FilePatterns = "Masked_",  
  models = "all")
```

Unfortunately, the output files are not integrated with the current *phruta* pipeline. This will be part of a new release. However, users can still perform gene-based partitioned analyses within *RAxML* or can use *PartitionFinder*'s output files to inform their own analyses outside *phruta*.

Identifying rogue taxa

phruta can help users run *RogueNaRok* implemented in the *Rogue* R package. Users can then examine whether rogue taxa should be excluded from the analyses. `tree.roguetaxa()` uses the bootstrap trees generated using the `tree.raxml()` function along with the associated best tree to identify rogue taxa. Rogue taxa for the analyses presented in this paper are shown in [Table 1: rogue](#)

```
tree.roguetaxa(folder = "3.Phylogeny")
```

Reproducibility with phruta

One of the central points of developing *phruta* was related to increasing the reproducibility of simple phylogenetic analyses. By either compiling or calling alternative tools that are commonly used to assemble species-level molecular and phylogenetic datasets inside an R package, *phruta* allows users to generate a clear, structured, and reproducible workflow. In fact, *phruta* is conceived as a package that allow users to choose between at least two alternative options to share their workflow. First, users can simply provide access to their workflow in an R script. For instance, this file can be stored in GitHub along with all the intermediate files that are created each each given step. Alternatively, given that the information in databases is constantly changing, users can share their R script and associated workspace to assure that the versions of the retrieved files correspond to specific versions of the databases.

As an example, all the analyses and sequences analyzed in the current manuscript will be available to anyone in a RMarkdown file accompanied by a n R workspace `phruta_ms.RData`. The RMarkdown will allow users to replicate the analyses presented in the current paper. With the workspace, users can choose to use the stored objects instead of performing searches on GenBank and additional taxonomic databases. These two files confer full reproducibility to the analyses presented in this article.

```
save.image(file = "phruta_ms.RData")
```

Performance

Assessing the performance of phruta is intrinsically challenging. However, it is expected for functions that scrape and curate information stored in GenBank to show a slow performance on search terms (e.g. clades) with extensive genetic sampling. Below, I focus on comparing the distribution of durations for assembling molecular datasets in seven clades with different sample size. These estimates encompass all the steps outlined above between `gene.sampling.retrieve()` and `sq.curate`. Time estimates were estimated using the `microbenchmark` R package. [fig:benchmark](#) shows a summary of phruta performance in different clades.

Table 1: Top six genes sampled in GenBank for species in Odontophoridae and Polyplectron.

Gene	Sampled in N species	PercentOfSampledSpecies
NADH dehydrogenase subunit 2	92	98.92473
12S ribosomal RNA	28	30.10753
eukaryotic elongation factor 2	26	27.95699
NADH dehydrogenase subunit 5	19	20.43011
cytochrome b	16	17.20430
cytochrome oxidase subunit 1	10	10.75269



Figure 1: Quail placeholder. Phyto by Brent Myers

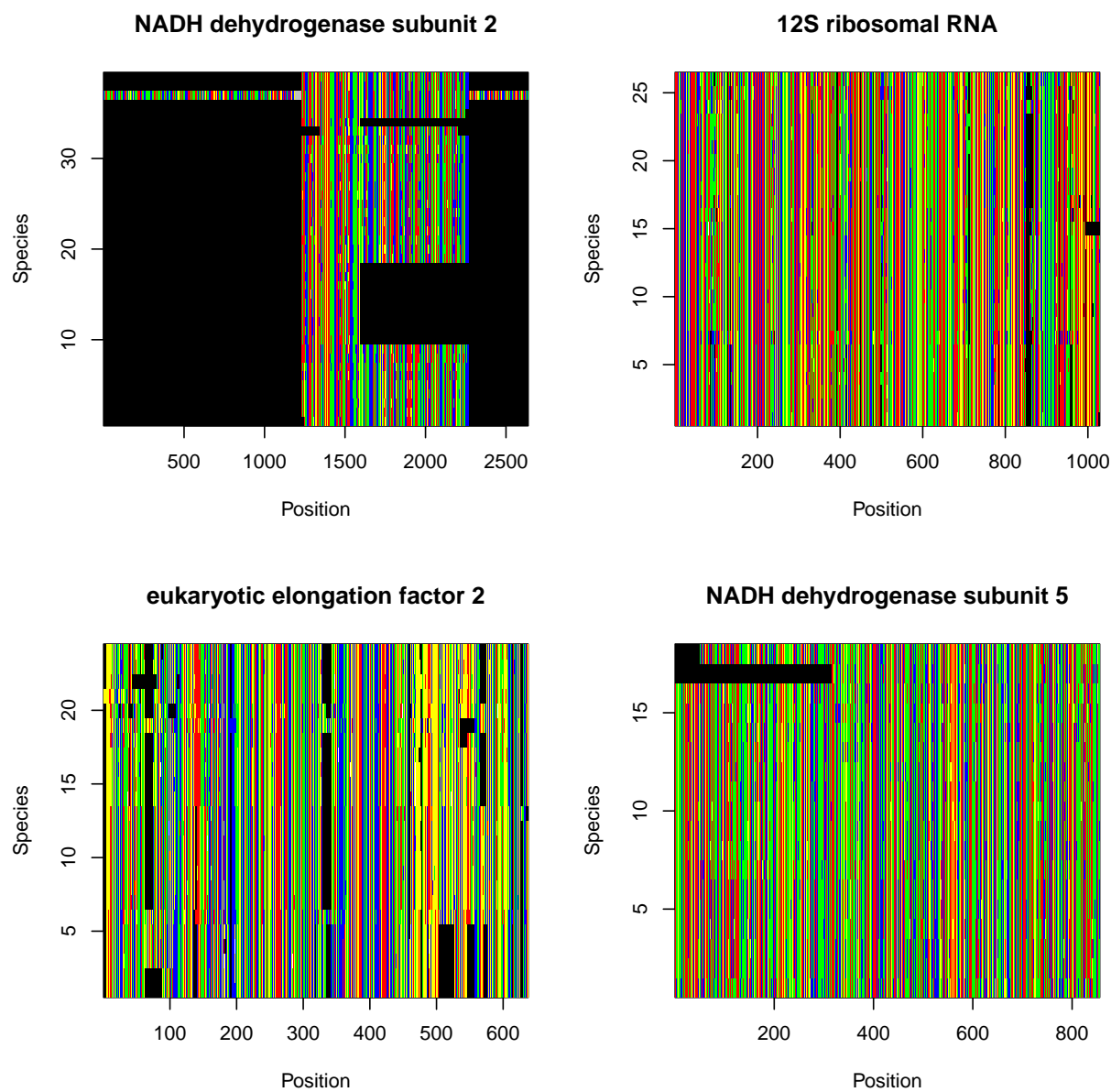


Figure 2: Raw alignments for gene regions sampled in more than 20% of the species in GenBank

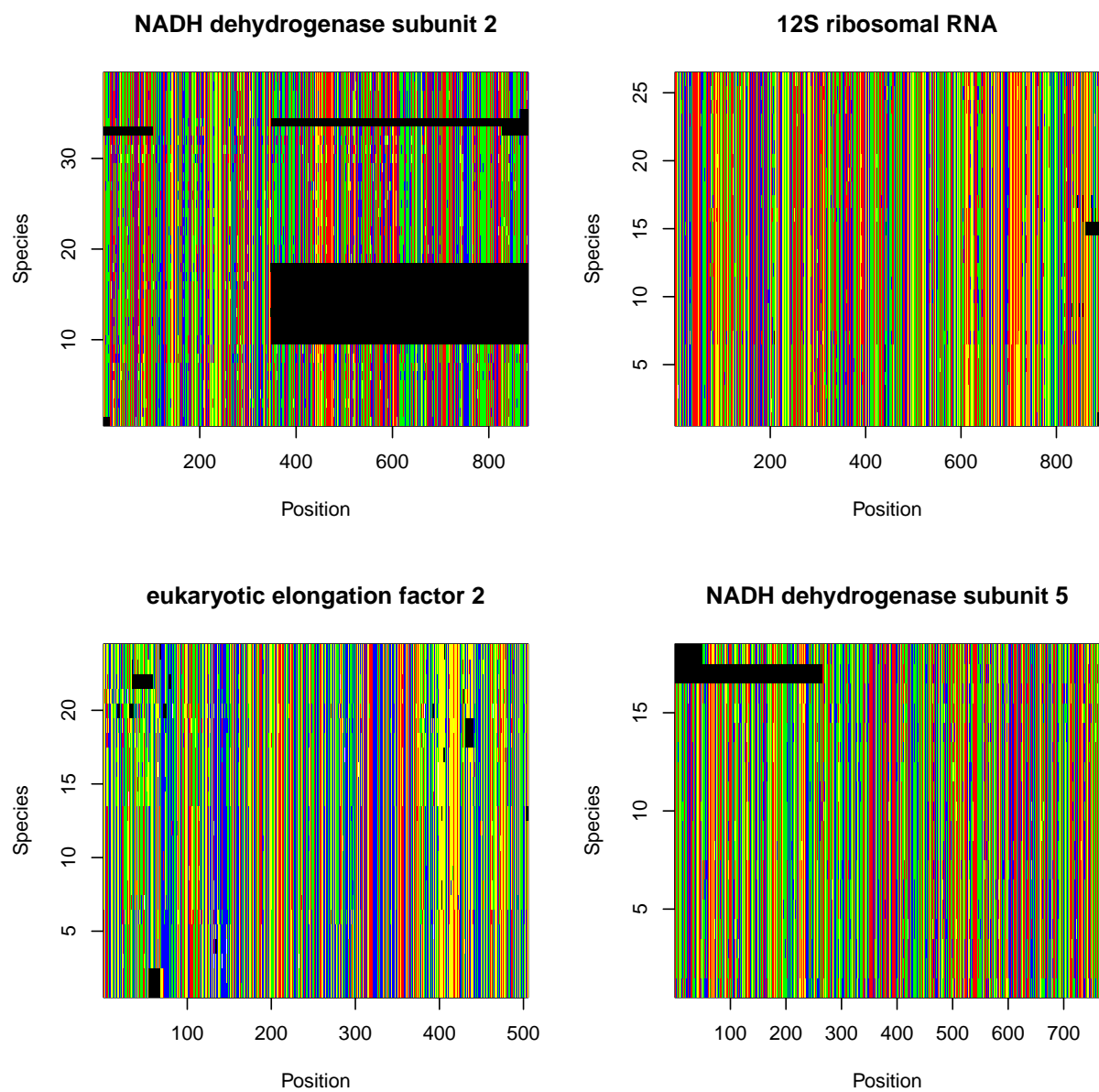


Figure 3: Curated alignments for gene regions sampled in more than 20% of the species in GenBank

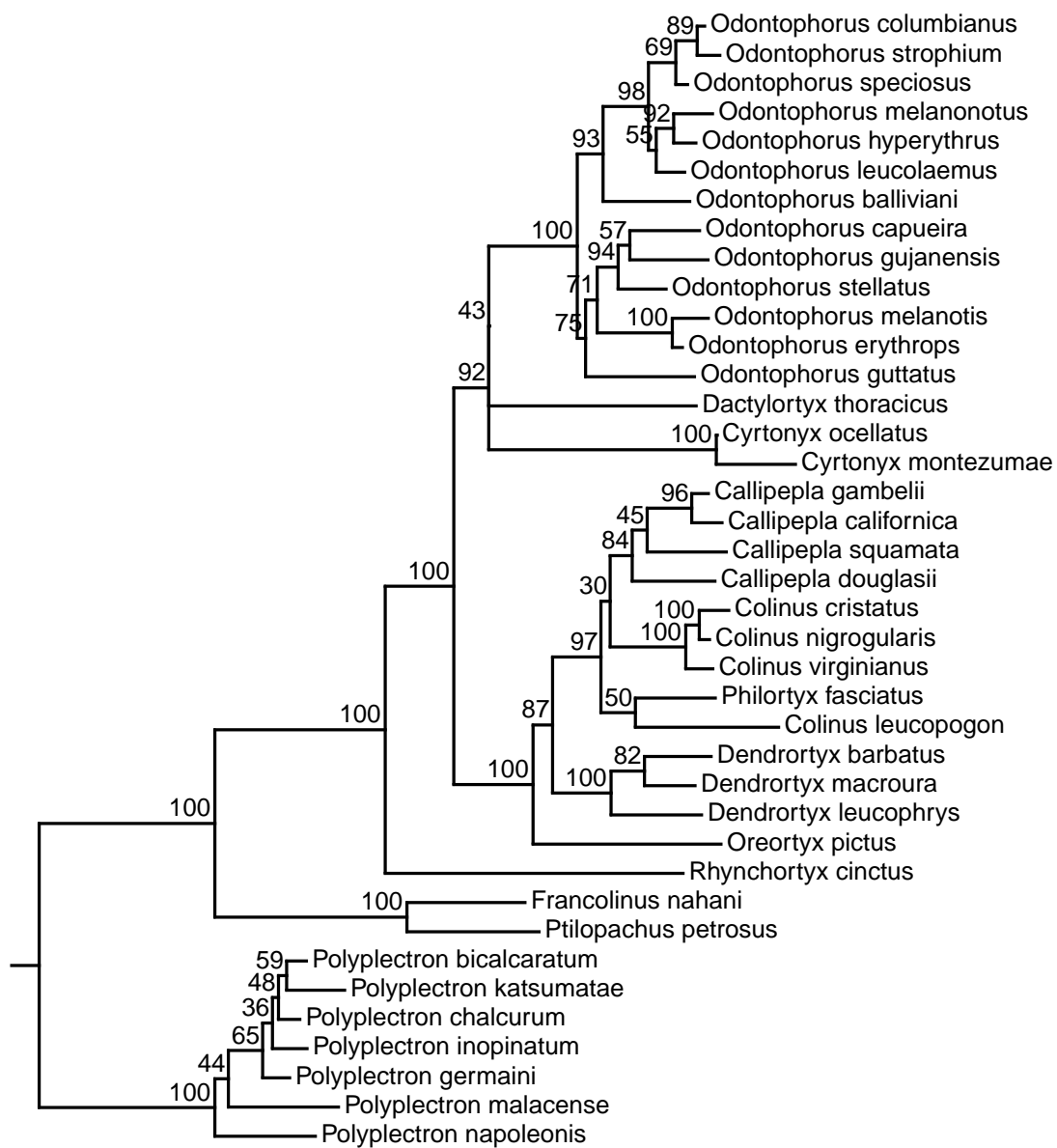


Figure 4: RAxML phylo

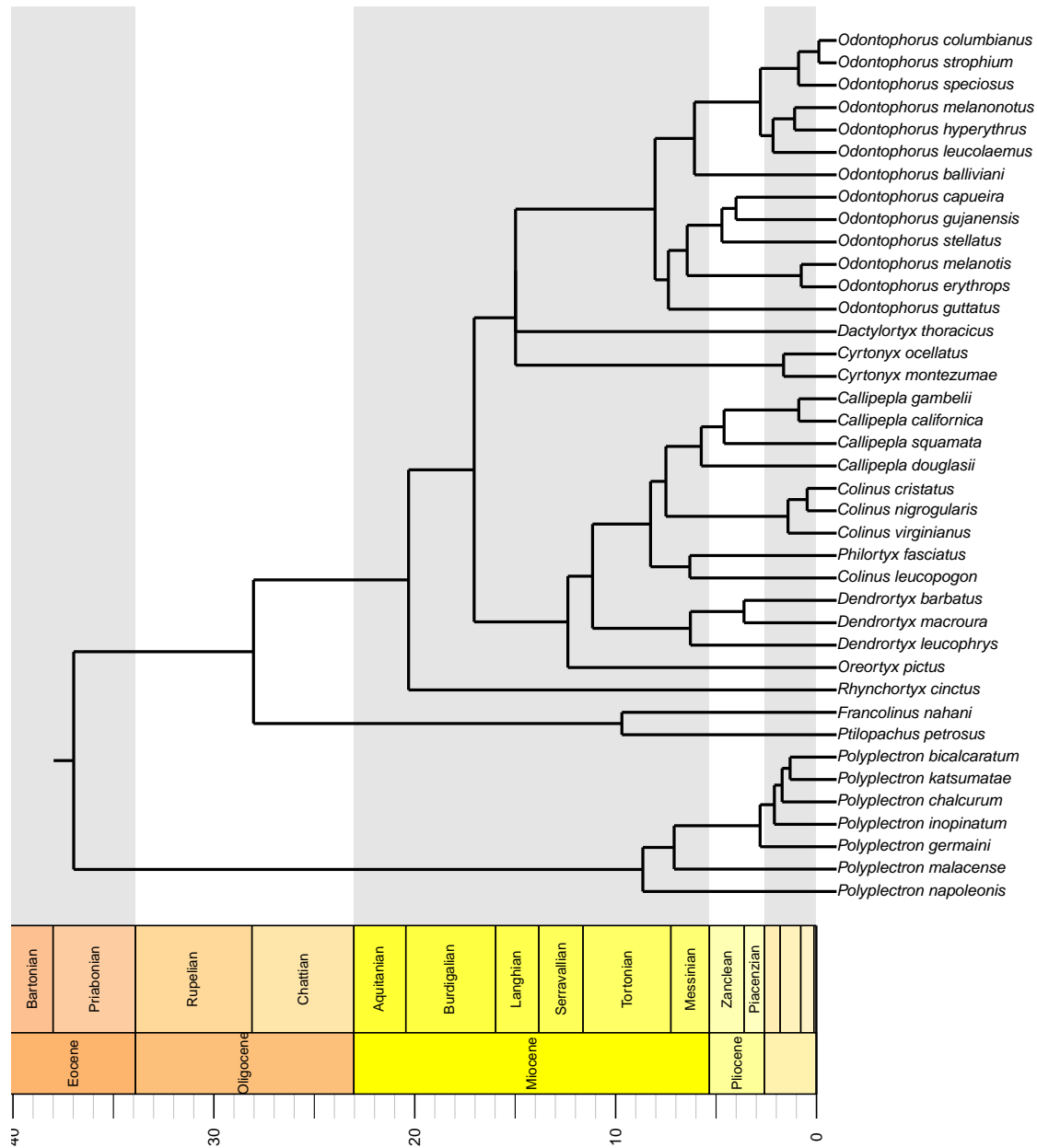


Figure 5: Time-calibrated phylogeny phylo

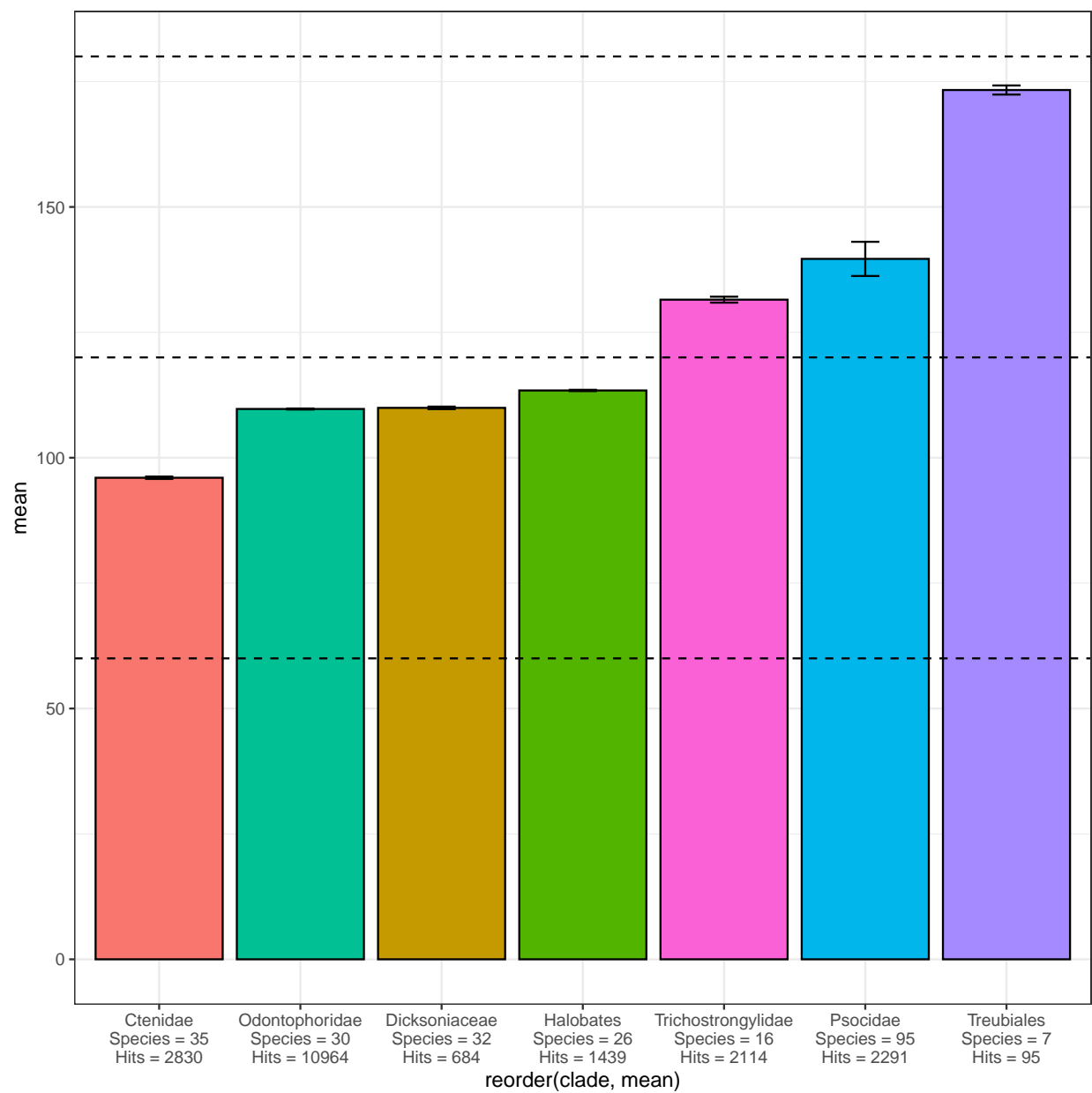


Figure 6: Time-calibrated phylogeny phylo

Table 2: Summary of potential accession numbers for the species in Odontophoridae, our ingroup, and Polyplectron, outgroup genus. This list of sequences has not been curated yet.

Species	Acc	gene
<i>Callipepla californica</i>	MZ476322	NADH dehydrogenase subunit 2
<i>Callipepla gambelii</i>	MZ476314	NADH dehydrogenase subunit 2
<i>Colinus virginianus</i>	EU166949	NADH dehydrogenase subunit 2
<i>Colinus cristatus</i>	AF222544	NADH dehydrogenase subunit 2
<i>Colinus nigrogularis</i>	KR732857	NADH dehydrogenase subunit 2
<i>Dendrortyx barbatus</i>	KR732856	NADH dehydrogenase subunit 2
<i>Philortyx fasciatus</i>	KR732855	NADH dehydrogenase subunit 2
<i>Rhynchortyx cinctus</i>	KR732850	NADH dehydrogenase subunit 2
<i>Cyrtonyx montezumae</i>	KR732849	NADH dehydrogenase subunit 2
<i>Oreortyx pictus</i>	KR732848	NADH dehydrogenase subunit 2
<i>Odontophorus leucolaemus</i>	KR732847	NADH dehydrogenase subunit 2
<i>Odontophorus speciosus</i>	KR732846	NADH dehydrogenase subunit 2
<i>Odontophorus erythrops</i>	KR732845	NADH dehydrogenase subunit 2
<i>Odontophorus guttatus</i>	KR732844	NADH dehydrogenase subunit 2
<i>Odontophorus gujanensis</i>	KR732843	NADH dehydrogenase subunit 2
<i>Odontophorus capueira</i>	KR732842	NADH dehydrogenase subunit 2
<i>Odontophorus stellatus</i>	KR732841	NADH dehydrogenase subunit 2
<i>Dactylortyx thoracicus</i>	KR732840	NADH dehydrogenase subunit 2
<i>Dendrortyx macroura</i>	KR732839	NADH dehydrogenase subunit 2
<i>Callipepla squamata</i>	KR732838	NADH dehydrogenase subunit 2
<i>Callipepla douglasii</i>	KR732837	NADH dehydrogenase subunit 2
<i>Colinus leucopogon</i>	KC556543	NADH dehydrogenase subunit 2
<i>Dendrortyx leucophrys</i>	KC556066	NADH dehydrogenase subunit 2
<i>Cyrtonyx ocellatus</i>	KC556060	NADH dehydrogenase subunit 2
<i>Odontophorus balliviani</i>	KC556524	NADH dehydrogenase subunit 2
<i>Odontophorus columbianus</i>	KC556517	NADH dehydrogenase subunit 2
<i>Odontophorus strophium</i>	KC556515	NADH dehydrogenase subunit 2
<i>Odontophorus melanonotus</i>	KC556513	NADH dehydrogenase subunit 2
<i>Odontophorus hyperythrus</i>	KC556512	NADH dehydrogenase subunit 2
<i>Odontophorus melanotis</i>	KC556507	NADH dehydrogenase subunit 2
<i>Ptilopachus nahani</i>	DQ768288	NADH dehydrogenase subunit 2
<i>Ptilopachus petrosus</i>	KR732851	NADH dehydrogenase subunit 2
<i>Polyplectron inopinatum</i>	EF569482	NADH dehydrogenase subunit 2
<i>Polyplectron napoleonis</i>	EF569481	NADH dehydrogenase subunit 2
<i>Polyplectron chalcum</i>	EF569480	NADH dehydrogenase subunit 2

Table 2: Summary of potential accession numbers for the species in Odontophoridae, our ingroup, and Polyplectron, outgroup genus. This list of sequences has not been curated yet. (*continued*)

Species	Acc	gene
Polyplectron bicalcaratum	EF569479	NADH dehydrogenase subunit 2
Polyplectron malacense	DQ768268	NADH dehydrogenase subunit 2
Polyplectron germaini	DQ768266	NADH dehydrogenase subunit 2
Polyplectron katsumatae	KC778823	NADH dehydrogenase subunit 2
Cyrtonyx montezumae	KR732830	12S ribosomal RNA
Dactylortyx thoracicus	KR732829	12S ribosomal RNA
Oreortyx pictus	KR732828	12S ribosomal RNA
Odontophorus erythrops	KR732827	12S ribosomal RNA
Odontophorus gujanensis	KR732826	12S ribosomal RNA
Odontophorus stellatus	KR732825	12S ribosomal RNA
Odontophorus capueira	KR732824	12S ribosomal RNA
Odontophorus speciosus	KR732823	12S ribosomal RNA
Odontophorus leucolaemus	KR732822	12S ribosomal RNA
Odontophorus balliviani	KR732821	12S ribosomal RNA
Dendrortyx macroura	KR732820	12S ribosomal RNA
Philortyx fasciatus	KR732819	12S ribosomal RNA
Callipepla gambelii	KR732818	12S ribosomal RNA
Callipepla californica	KR732817	12S ribosomal RNA
Callipepla douglasii	KR732816	12S ribosomal RNA
Callipepla squamata	KR732815	12S ribosomal RNA
Colinus cristatus	KR732814	12S ribosomal RNA
Colinus nigrogularis	KR732813	12S ribosomal RNA
Colinus virginianus	KR732812	12S ribosomal RNA
Ptilopachus petrosus	KR732832	12S ribosomal RNA
Polyplectron inopinatum	KC749467	12S ribosomal RNA
Polyplectron germaini	KC749466	12S ribosomal RNA
Polyplectron napoleonis	KC749465	12S ribosomal RNA
Polyplectron chalcum	KC749464	12S ribosomal RNA
Polyplectron bicalcaratum	KC749463	12S ribosomal RNA
Polyplectron katsumatae	KC778974	12S ribosomal RNA
Colinus cristatus	KR732895	eukaryotic elongation factor 2
Callipepla squamata	KR732894	eukaryotic elongation factor 2
Dendrortyx macroura	KR732893	eukaryotic elongation factor 2
Philortyx fasciatus	KR732892	eukaryotic elongation factor 2
Cyrtonyx montezumae	KR732891	eukaryotic elongation factor 2

Table 2: Summary of potential accession numbers for the species in Odontophoridae, our ingroup, and Polyplectron, outgroup genus. This list of sequences has not been curated yet. (*continued*)

Species	Acc	gene
Rhynchortyx cinctus	KR732889	eukaryotic elongation factor 2
Oreortyx pictus	KR732888	eukaryotic elongation factor 2
Colinus virginianus	KR732887	eukaryotic elongation factor 2
Callipepla douglasii	KR732886	eukaryotic elongation factor 2
Callipepla californica	KR732885	eukaryotic elongation factor 2
Callipepla gambelii	KR732884	eukaryotic elongation factor 2
Odontophorus erythrops	KR732883	eukaryotic elongation factor 2
Odontophorus guttatus	KR732882	eukaryotic elongation factor 2
Odontophorus speciosus	KR732881	eukaryotic elongation factor 2
Odontophorus leucolaemus	KR732880	eukaryotic elongation factor 2
Odontophorus balliviani	KR732879	eukaryotic elongation factor 2
Odontophorus stellatus	KR732878	eukaryotic elongation factor 2
Odontophorus capueira	KR732877	eukaryotic elongation factor 2
Ptilopachus petrosus	KR732890	eukaryotic elongation factor 2
Polyplectron inopinatum	KC749707	eukaryotic elongation factor 2
Polyplectron germaini	KC749706	eukaryotic elongation factor 2
Polyplectron chalcurum	KC749705	eukaryotic elongation factor 2
Polyplectron bicalcaratum	KC749704	eukaryotic elongation factor 2
Polyplectron katsumatae	KC778867	eukaryotic elongation factor 2
Philortyx fasciatus	KR732875	NADH dehydrogenase subunit 5
Callipepla douglasii	KR732874	NADH dehydrogenase subunit 5
Rhynchortyx cinctus	KR732872	NADH dehydrogenase subunit 5
Dactylortyx thoracicus	KR732871	NADH dehydrogenase subunit 5
Odontophorus leucolaemus	KR732870	NADH dehydrogenase subunit 5
Odontophorus balliviani	KR732869	NADH dehydrogenase subunit 5
Odontophorus guttatus	KR732868	NADH dehydrogenase subunit 5
Odontophorus erythrops	KR732867	NADH dehydrogenase subunit 5
Odontophorus capueira	KR732866	NADH dehydrogenase subunit 5
Odontophorus stellatus	KR732865	NADH dehydrogenase subunit 5
Odontophorus gujanensis	KR732864	NADH dehydrogenase subunit 5
Dendrortyx macroura	KR732863	NADH dehydrogenase subunit 5
Callipepla californica	KR732862	NADH dehydrogenase subunit 5
Callipepla gambelii	KR732861	NADH dehydrogenase subunit 5
Callipepla squamata	KR732860	NADH dehydrogenase subunit 5
Colinus virginianus	KR732859	NADH dehydrogenase subunit 5

Table 2: Summary of potential accession numbers for the species in Odontophoridae, our ingroup, and Polyplectron, outgroup genus. This list of sequences has not been curated yet. (*continued*)

Species	Acc	gene
Colinus nigrogularis	KR732858	NADH dehydrogenase subunit 5
Ptilopachus petrosus	KR732873	NADH dehydrogenase subunit 5

Table 3: Accession numbers for the retrieved sequences. This dataset has been curated.

	OriginalNames	AccN	Species	file
	MZ476322 Callipepla californica	MZ476322	Callipepla_californica	NADH dehydrogenase subunit 2
	MZ476314 Callipepla gambelii	MZ476314	Callipepla_gambelii	NADH dehydrogenase subunit 2
	EU166949 Colinus virginianus	EU166949	Colinus_virginianus	NADH dehydrogenase subunit 2
	AF222544 Colinus cristatus	AF222544	Colinus_cristatus	NADH dehydrogenase subunit 2
	KR732857 Colinus nigrogularis	KR732857	Colinus_nigrogularis	NADH dehydrogenase subunit 2
	KR732856 Dendrortyx barbatus	KR732856	Dendrortyx_barbatus	NADH dehydrogenase subunit 2
	KR732855 Philortyx fasciatus	KR732855	Philortyx_fasciatus	NADH dehydrogenase subunit 2
	KR732850 Rhynchortyx cinctus	KR732850	Rhynchortyx_cinctus	NADH dehydrogenase subunit 2
	KR732849 Cyrtonyx montezumae	KR732849	Cyrtonyx_montezumae	NADH dehydrogenase subunit 2
	KR732848 Oreortyx pictus	KR732848	Oreortyx_pictus	NADH dehydrogenase subunit 2
22	KR732847 Odontophorus leucolaemus	KR732847	Odontophorus_leucolaemus	NADH dehydrogenase subunit 2
	KR732846 Odontophorus speciosus	KR732846	Odontophorus_speciosus	NADH dehydrogenase subunit 2
	KR732845 Odontophorus erythrops	KR732845	Odontophorus_erythrops	NADH dehydrogenase subunit 2
	KR732844 Odontophorus guttatus	KR732844	Odontophorus_guttatus	NADH dehydrogenase subunit 2
	KR732843 Odontophorus gujanensis	KR732843	Odontophorus_gujanensis	NADH dehydrogenase subunit 2
	KR732842 Odontophorus capueira	KR732842	Odontophorus_capueira	NADH dehydrogenase subunit 2
	KR732841 Odontophorus stellatus	KR732841	Odontophorus_stellatus	NADH dehydrogenase subunit 2
	KR732840 Dactylortyx thoracicus	KR732840	Dactylortyx_thoracicus	NADH dehydrogenase subunit 2
	KR732839 Dendrortyx macroura	KR732839	Dendrortyx_macroura	NADH dehydrogenase subunit 2
	KR732838 Callipepla squamata	KR732838	Callipepla_squamata	NADH dehydrogenase subunit 2
	KR732837 Callipepla douglasii	KR732837	Callipepla_douglasii	NADH dehydrogenase subunit 2
	KC556543 Colinus leucopogon	KC556543	Colinus_leucopogon	NADH dehydrogenase subunit 2
	KC556066 Dendrortyx leucophrys	KC556066	Dendrortyx_leucophrys	NADH dehydrogenase subunit 2
	KC556060 Cyrtonyx ocellatus	KC556060	Cyrtonyx_ocellatus	NADH dehydrogenase subunit 2
	KC556524 Odontophorus balliviani	KC556524	Odontophorus_balliviani	NADH dehydrogenase subunit 2

Table 3: Accession numbers for the retrieved sequences. This dataset has been curated. (*continued*)

OriginalNames	AccN	Species	file
KC556517 Odontophorus columbianus	KC556517	Odontophorus_columbianus	NADH dehydrogenase subunit 2
KC556515 Odontophorus strophium	KC556515	Odontophorus_strophium	NADH dehydrogenase subunit 2
KC556513 Odontophorus melanonotus	KC556513	Odontophorus_melanonotus	NADH dehydrogenase subunit 2
KC556512 Odontophorus hyperythrus	KC556512	Odontophorus_hyperythrus	NADH dehydrogenase subunit 2
KC556507 Odontophorus melanotis	KC556507	Odontophorus_melanotis	NADH dehydrogenase subunit 2
DQ768288 Ptilopachus nahani	DQ768288	Francolinus_nahani	NADH dehydrogenase subunit 2
KR732851 Ptilopachus petrosus	KR732851	Ptilopachus_petrosus	NADH dehydrogenase subunit 2
EF569482 Polyplectron inopinatum	EF569482	Polyplectron_inopinatum	NADH dehydrogenase subunit 2
EF569481 Polyplectron napoleonis	EF569481	Polyplectron_napoleonis	NADH dehydrogenase subunit 2
EF569480 Polyplectron chalcurum	EF569480	Polyplectron_chalcurum	NADH dehydrogenase subunit 2
EF569479 Polyplectron bicalcaratum	EF569479	Polyplectron_bicalcaratum	NADH dehydrogenase subunit 2
DQ768268 Polyplectron malacense	DQ768268	Polyplectron_malacense	NADH dehydrogenase subunit 2
DQ768266 Polyplectron germaini	DQ768266	Polyplectron_germaini	NADH dehydrogenase subunit 2
KC778823 Polyplectron katsumatae	KC778823	Polyplectron_katsumatae	NADH dehydrogenase subunit 2
KR732830 Cyrtonyx montezumae	KR732830	Cyrtonyx_montezumae	12S ribosomal RNA
KR732829 Dactylortyx thoracicus	KR732829	Dactylortyx_thoracicus	12S ribosomal RNA
KR732828 Oreortyx pictus	KR732828	Oreortyx_pictus	12S ribosomal RNA
KR732827 Odontophorus erythrops	KR732827	Odontophorus_erythrops	12S ribosomal RNA
KR732826 Odontophorus gujanensis	KR732826	Odontophorus_gujanensis	12S ribosomal RNA
KR732825 Odontophorus stellatus	KR732825	Odontophorus_stellatus	12S ribosomal RNA
KR732824 Odontophorus capueira	KR732824	Odontophorus_capueira	12S ribosomal RNA
KR732823 Odontophorus speciosus	KR732823	Odontophorus_speciosus	12S ribosomal RNA
KR732822 Odontophorus leucolaemus	KR732822	Odontophorus_leucolaemus	12S ribosomal RNA
KR732821 Odontophorus balliviani	KR732821	Odontophorus_balliviani	12S ribosomal RNA
KR732820 Dendrortyx macroura	KR732820	Dendrortyx_macroura	12S ribosomal RNA

Table 3: Accession numbers for the retrieved sequences. This dataset has been curated. (*continued*)

OriginalNames	AccN	Species	file
KR732819 Philortyx fasciatus	KR732819	Philortyx_fasciatus	12S ribosomal RNA
KR732818 Callipepla gambelii	KR732818	Callipepla_gambelii	12S ribosomal RNA
KR732817 Callipepla californica	KR732817	Callipepla_californica	12S ribosomal RNA
KR732816 Callipepla douglasii	KR732816	Callipepla_douglasii	12S ribosomal RNA
KR732815 Callipepla squamata	KR732815	Callipepla_squamata	12S ribosomal RNA
KR732814 Colinus cristatus	KR732814	Colinus_cristatus	12S ribosomal RNA
KR732813 Colinus nigrogularis	KR732813	Colinus_nigrogularis	12S ribosomal RNA
KR732812 Colinus virginianus	KR732812	Colinus_virginianus	12S ribosomal RNA
KR732832 Ptilopachus petrosus	KR732832	Ptilopachus_petrosus	12S ribosomal RNA
KC749467 Polyplectron inopinatum	KC749467	Polyplectron_inopinatum	12S ribosomal RNA
KC749466 Polyplectron germaini	KC749466	Polyplectron_germaini	12S ribosomal RNA
KC749465 Polyplectron napoleonis	KC749465	Polyplectron_napoleonis	12S ribosomal RNA
KC749464 Polyplectron chalcum	KC749464	Polyplectron_chalcum	12S ribosomal RNA
KC749463 Polyplectron bicalcaratum	KC749463	Polyplectron_bicalcaratum	12S ribosomal RNA
KC778974 Polyplectron katsumatae	KC778974	Polyplectron_katsumatae	12S ribosomal RNA
KR732895 Colinus cristatus	KR732895	Colinus_cristatus	eukaryotic elongation factor 2
KR732894 Callipepla squamata	KR732894	Callipepla_squamata	eukaryotic elongation factor 2
KR732893 Dendrortyx macroura	KR732893	Dendrortyx_macroura	eukaryotic elongation factor 2
KR732892 Philortyx fasciatus	KR732892	Philortyx_fasciatus	eukaryotic elongation factor 2
KR732891 Cyrtonyx montezumae	KR732891	Cyrtonyx_montezumae	eukaryotic elongation factor 2
KR732889 Rhynchortyx cinctus	KR732889	Rhynchortyx_cinctus	eukaryotic elongation factor 2
KR732888 Oreortyx pictus	KR732888	Oreortyx_pictus	eukaryotic elongation factor 2
KR732887 Colinus virginianus	KR732887	Colinus_virginianus	eukaryotic elongation factor 2
KR732886 Callipepla douglasii	KR732886	Callipepla_douglasii	eukaryotic elongation factor 2
KR732885 Callipepla californica	KR732885	Callipepla_californica	eukaryotic elongation factor 2

Table 3: Accession numbers for the retrieved sequences. This dataset has been curated. (*continued*)

OriginalNames	AccN	Species	file
KR732884 Callipepla gambelii	KR732884	Callipepla_gambelii	eukaryotic elongation factor 2
KR732883 Odontophorus erythrops	KR732883	Odontophorus_erythrops	eukaryotic elongation factor 2
KR732882 Odontophorus guttatus	KR732882	Odontophorus_guttatus	eukaryotic elongation factor 2
KR732881 Odontophorus speciosus	KR732881	Odontophorus_speciosus	eukaryotic elongation factor 2
KR732880 Odontophorus leucolaemus	KR732880	Odontophorus_leucolaemus	eukaryotic elongation factor 2
KR732879 Odontophorus balliviani	KR732879	Odontophorus_balliviani	eukaryotic elongation factor 2
KR732878 Odontophorus stellatus	KR732878	Odontophorus_stellatus	eukaryotic elongation factor 2
KR732877 Odontophorus capueira	KR732877	Odontophorus_capueira	eukaryotic elongation factor 2
KR732890 Ptilopachus petrosus	KR732890	Ptilopachus_petrosus	eukaryotic elongation factor 2
KC749707 Polyplectron inopinatum	KC749707	Polyplectron_inopinatum	eukaryotic elongation factor 2
KC749706 Polyplectron germaini	KC749706	Polyplectron_germaini	eukaryotic elongation factor 2
KC749705 Polyplectron chalcum	KC749705	Polyplectron_chalcum	eukaryotic elongation factor 2
KC749704 Polyplectron bicalcaratum	KC749704	Polyplectron_bicalcaratum	eukaryotic elongation factor 2
KC778867 Polyplectron katsumatae	KC778867	Polyplectron_katsumatae	eukaryotic elongation factor 2
KR732875 Philortyx fasciatus	KR732875	Philortyx_fasciatus	NADH dehydrogenase subunit 5
KR732874 Callipepla douglasii	KR732874	Callipepla_douglasii	NADH dehydrogenase subunit 5
KR732872 Rhynchortyx cinctus	KR732872	Rhynchortyx_cinctus	NADH dehydrogenase subunit 5
KR732871 Dactylortyx thoracicus	KR732871	Dactylortyx_thoracicus	NADH dehydrogenase subunit 5
KR732870 Odontophorus leucolaemus	KR732870	Odontophorus_leucolaemus	NADH dehydrogenase subunit 5
KR732869 Odontophorus balliviani	KR732869	Odontophorus_balliviani	NADH dehydrogenase subunit 5
KR732868 Odontophorus guttatus	KR732868	Odontophorus_guttatus	NADH dehydrogenase subunit 5
KR732867 Odontophorus erythrops	KR732867	Odontophorus_erythrops	NADH dehydrogenase subunit 5
KR732866 Odontophorus capueira	KR732866	Odontophorus_capueira	NADH dehydrogenase subunit 5
KR732865 Odontophorus stellatus	KR732865	Odontophorus_stellatus	NADH dehydrogenase subunit 5
KR732864 Odontophorus gujanensis	KR732864	Odontophorus_gujanensis	NADH dehydrogenase subunit 5

Table 3: Accession numbers for the retrieved sequences. This dataset has been curated. (*continued*)

OriginalNames	AccN	Species	file
KR732863 Dendrortyx macroura	KR732863	Dendrortyx_macroura	NADH dehydrogenase subunit 5
KR732862 Callipepla californica	KR732862	Callipepla_californica	NADH dehydrogenase subunit 5
KR732861 Callipepla gambelii	KR732861	Callipepla_gambelii	NADH dehydrogenase subunit 5
KR732860 Callipepla squamata	KR732860	Callipepla_squamata	NADH dehydrogenase subunit 5
KR732859 Colinus virginianus	KR732859	Colinus_virginianus	NADH dehydrogenase subunit 5
KR732858 Colinus nigrogularis	KR732858	Colinus_nigrogularis	NADH dehydrogenase subunit 5
KR732873 Ptilopachus petrosus	KR732873	Ptilopachus_petrosus	NADH dehydrogenase subunit 5

Table 4: Taxonomic information for the retrieved species

kingdom	phylum	class	order	family	genus	species
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Callipepla	Callipepla californica
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Callipepla	Callipepla gambelii
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Colinus	Colinus virginianus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Colinus	Colinus cristatus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Colinus	Colinus nigrogularis
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Dendrortyx	Dendrortyx barbatus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Philortyx	Philortyx fasciatus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Rhynchortyx	Rhynchortyx cinctus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Cyrtonyx	Cyrtonyx montezumae
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Oreortyx	Oreortyx pictus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus leucolaemus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus speciosus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus erythrops
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus guttatus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus gujanensis
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus capueira
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus stellatus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Dactylortyx	Dactylortyx thoracicus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Dendrortyx	Dendrortyx macroura
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Callipepla	Callipepla squamata
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Callipepla	Callipepla douglasii
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Colinus	Colinus leucopogon
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Dendrortyx	Dendrortyx leucophrys
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Cyrtonyx	Cyrtonyx ocellatus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus balliviani

Table 4: Taxonomic information for the retrieved species (*continued*)

kingdom	phylum	class	order	family	genus	species
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus columbianus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus strophium
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus melanonotus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus hyperythrus
Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus	Odontophorus melanotis
Animalia	Chordata	Aves	Galliformes	Phasianidae	Francolinus	Francolinus nahani
Animalia	Chordata	Aves	Galliformes	Phasianidae	Ptilopachus	Ptilopachus petrosus
Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron	Polyplectron inopinatum
Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron	Polyplectron napoleonis
Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron	Polyplectron chalcurum
Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron	Polyplectron bicalcaratum
Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron	Polyplectron malacense
Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron	Polyplectron germaini
Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron	Polyplectron katsumatae

Table 5: Taxonomic sampling across gene regions

species_names	kingdom	phylum	class	order	family	genus
Callipepla_californica	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Callipepla
Callipepla_douglasii	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Callipepla
Callipepla_gambelii	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Callipepla
Callipepla_squamata	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Callipepla
Colinus_cristatus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Colinus
Colinus_leucopogon	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Colinus
Colinus_nigrogularis	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Colinus
Colinus_virginianus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Colinus
Cyrtonyx_montezumae	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Cyrtonyx
Cyrtonyx_ocellatus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Cyrtonyx
Dactylortyx_thoracicus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Dactylortyx
Dendrortyx_barbatus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Dendrortyx
Dendrortyx_leucophrys	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Dendrortyx
Dendrortyx_macroura	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Dendrortyx
Francolinus_nahani	Animalia	Chordata	Aves	Galliformes	Phasianidae	Francolinus
Odontophorus_balliviani	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_capueira	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_columbianus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_erythrops	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_gujanensis	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_guttatus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_hyperythrus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_leucolaemus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_melanonotus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_melanotis	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_speciosus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus

Table 5: Taxonomic sampling across gene regions (*continued*)

species_names	kingdom	phylum	class	order	family	genus
Odontophorus_stellatus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Odontophorus_strophium	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Odontophorus
Oreortyx_pictus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Oreortyx
Philortyx_fasciatus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Philortyx
Polyplectron_bicalcaratum	Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron
Polyplectron_chalcurum	Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron
Polyplectron_germaini	Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron
Polyplectron_inopinatum	Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron
Polyplectron_katsumatae	Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron
Polyplectron_malacense	Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron
Polyplectron_napoleonis	Animalia	Chordata	Aves	Galliformes	Phasianidae	Polyplectron
Ptilopachus_petrosus	Animalia	Chordata	Aves	Galliformes	Phasianidae	Ptilopachus
Rhynchortyx_cinctus	Animalia	Chordata	Aves	Galliformes	Odontophoridae	Rhynchortyx

Table 6: Results of RogueNaRock

num	taxNum	taxon	rawImprovement	IC
0	NA	NA	NA	414.9652
1	18	Colinus_leucopogon	11.05931	426.0245