2nd Rufford Small Grant
Project Update: June 2025

It has been one year since we initiated our fieldwork to establish passive acoustic monitoring for assessing the effects on forest biodiversity of the linear infrastructure of the "Tren Maya", a railway project traversing the Bacalar region and its surrounding areas. We have made significant progress in data collection and processing. With the acquisition of 25 new recording kits from Rainforest Connection as additional autonomous recording units (ARUs), consisting of Audiomoth recordings (AudioMoth Dev 1.0.0, Open Acoustic Devices) equipped with a battery supply, storage memory, and protective case, we have expanded the sampling area of study.

Acoustic monitoring sites

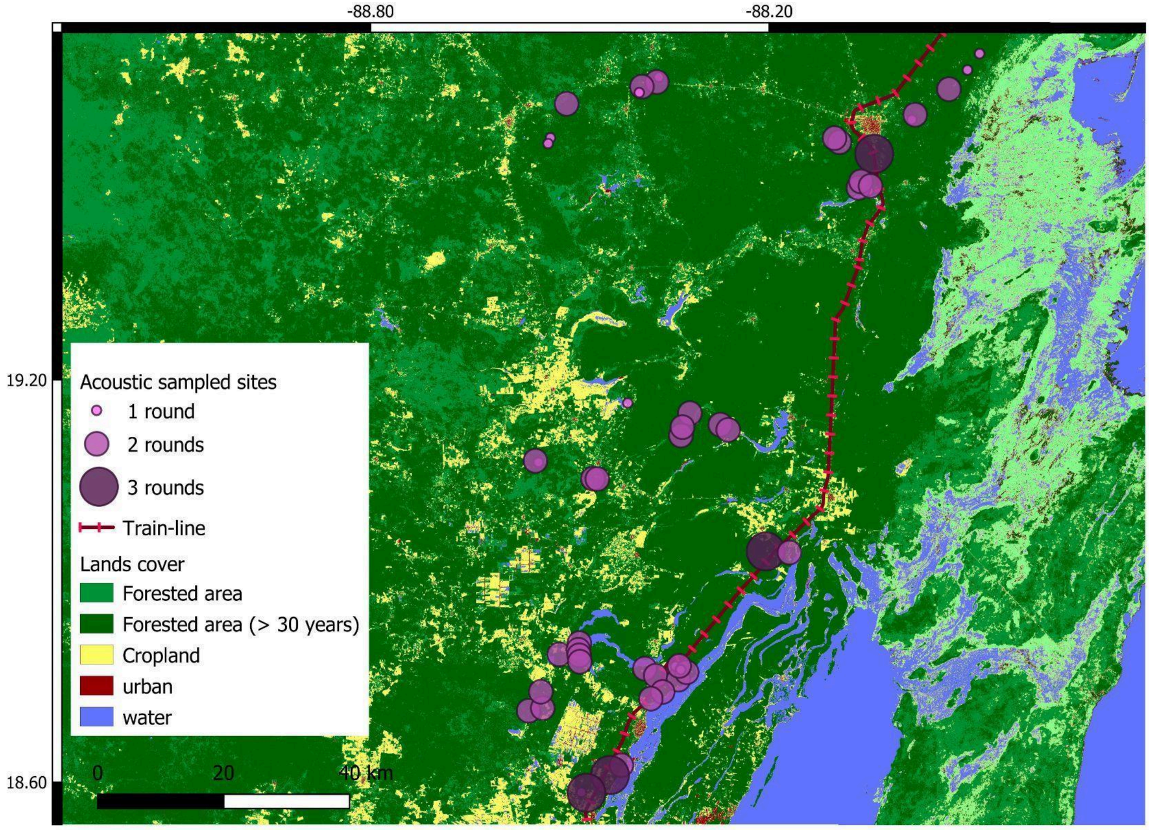

Figure 1. Acoustic sampling sites are distributed in the Selva Maya region of the southeastern Peninsula of Yucatán, in Quintana Roo, Mexico. The size of the circles indicating the location of the sampling sites reflects the number of ARU deployment rounds.

We planned to reinstall 15 sampling locations from our initial RSG in the Bacalar region and include 15 sampling locations monitored by collaboration during the pre-train-launch period in areas of the north of the Bacalar region, as well as 15-30 new sampling locations that have never been monitored before. We have deployed ARUs at 56 locations in the region. Despite our efforts to sample these locations, some sites were not monitored or were monitored only once due to access difficulties by environmental conditions that promote changes in the landscape, including floodable areas. Consequently, some of the sampling sites were replaced, and we have currently established 46 permanent monitoring locations, along with two additional new sites that we hope will sustain monitoring, for a total of 48 active monitoring locations. These locations include some of our previous monitoring sites, new locations previously monitored by colleagues, as well as new sites that have never been monitored before. Our sampling locations expanded the sampling area of our first RSG project, which was focused near Lake Bacalar. We now include locations at various distances along the train line, covering different land cover types and anthropogenic disturbances (Fig. 1), including forested and anthropized areas in Felipe Carrillo Puerto municipality, as well as further inland in the Bacalar municipality. Most sites are private properties, but there are also large community-owned areas (locally called “ejidos) used for conservation and forestry.

Monitoring protocol and fieldwork

We planned a 1-year acoustic monitoring program in our sample locations, with rounds of deployment of recordings conducted every three months. We have some differences in the rounds of deployment among ARUs due to changes in the establishment of sampling locations. Other sites had to be repeated due to failures with the ARUs (electronic problems caused by environmental moisture conditions or animal damage). However, most of the sites are nearing completion of their fourth round of deployment (expected in August 2025).

We defined an intensive monitoring protocol consisting of 1 minute of recording and 1 minute of non-recording. Each recorder was configured and recorded during an average of 18 days. We prepared and configured each ARU by loading the monitoring protocol into each recorder before visiting the sampling sites. This involved checking the equipment's functionality, verifying the available battery and memory, and configuring the recording cycles and local time (Fig. 2).

Figure 2. Preparation of the ARUs and configuration of the monitoring protocol.

During our first field trip in April 2024, we observed the changes in the landscape resulting from the construction of the “Tren Maya” with several areas deforested and prepared for the installation of the train rails, as well as patches for material benches for their construction, with high truck traffic activity on the roads. In subsequent visits to the field, we observed some progress in the construction of the train, including the establishment of rails, the installation of insulation and protection netting, and the installation of an electrical power supply (Fig. 3). At the end of 2024, the Tren Maya began operating in the section of the train line near our sampling locations, initially with a low number of runs, running two trains per day.

Figure 3. Progress on the construction of the Tren Maya railroad tracks during August 2024.

During our visits to the field, we presented our project to local communities to receive their support and obtain permits to establish sample locations on their community-owned forested land (Fig. 4). The people showed their interest and provided access to their forested areas under their management to deploy recorders.

Figure 4. A presentation of the project to the community from the Bacalar municipality.

Some members of the community have joined us on field trips, which have facilitated mutual learning that enables us to better understand the terrain. With their help, we identify the best places and habitats to deploy recorders. Meanwhile, we share the methods related to the installation of ARUs for acoustic monitoring of their lands (Fig. 5).

Figure 5. A member of a local community collaborates in the deployment of an autonomous recording unit (right), and an installed ARU in a forested area of the ejido of Dzulá in the Felipe Carrillo Puerto municipality (left).

Acoustic data collected and management

Most sampling locations have completed their third round of recording deployment. Four sites have achieved up to four successful deployments since they started monitoring following our first visit to the field during the logistics trip. Some sites reach only one round of deployment because we will no longer be monitoring the location for accessibility difficulties (8 sites). One site failed during its second round (R207), and we recently added two sites (R521 and R522) to the monitoring locations (Fig. 6). During each completed round, we collected an average of 180 hours of recordings, accumulating more than 500 hours across three complete rounds. In Figure 7, the three sites with the lowest cumulative hours represent one site where the ARU failed in its second round and two sites that we recently added to the monitoring. Four sites exceeding 600 hours represent locations that were able to initiate recording rounds during our first field logistics visit and reached up to four recording rounds. In August 2025, we will complete the fourth round of deployment for most locations. With the collected acoustic data as of the current date, we are processing audio to perform species detection and calculate the acoustic indices.

We processed all the audio files obtained in the field to carry out debugging (separating audio files with errors, removing empty files at the end of the recording period, and moving the first audio files with metadata information and sounds during deployment). Later, we made copies on hard drives. The audio files, originally in WAV format (a raw audio format), were converted to FLAC (a compressed lossless audio format). This audio format is used as an additional backup copy and supports fast uploading to the cloud. With the support of our collaborator, the nonprofit NGO Asociación GAICA, we gain access to cloud storage (Google Drive Workspace) to upload all FLAC audio files. Additionally, approximately 12,000 hours of acoustic data have been uploaded and added to our acoustic monitoring project on the

Arbimon platform (Rainforest Connection, https://arbimon.org), where we uploaded and analyzed the data from our first RSG project.

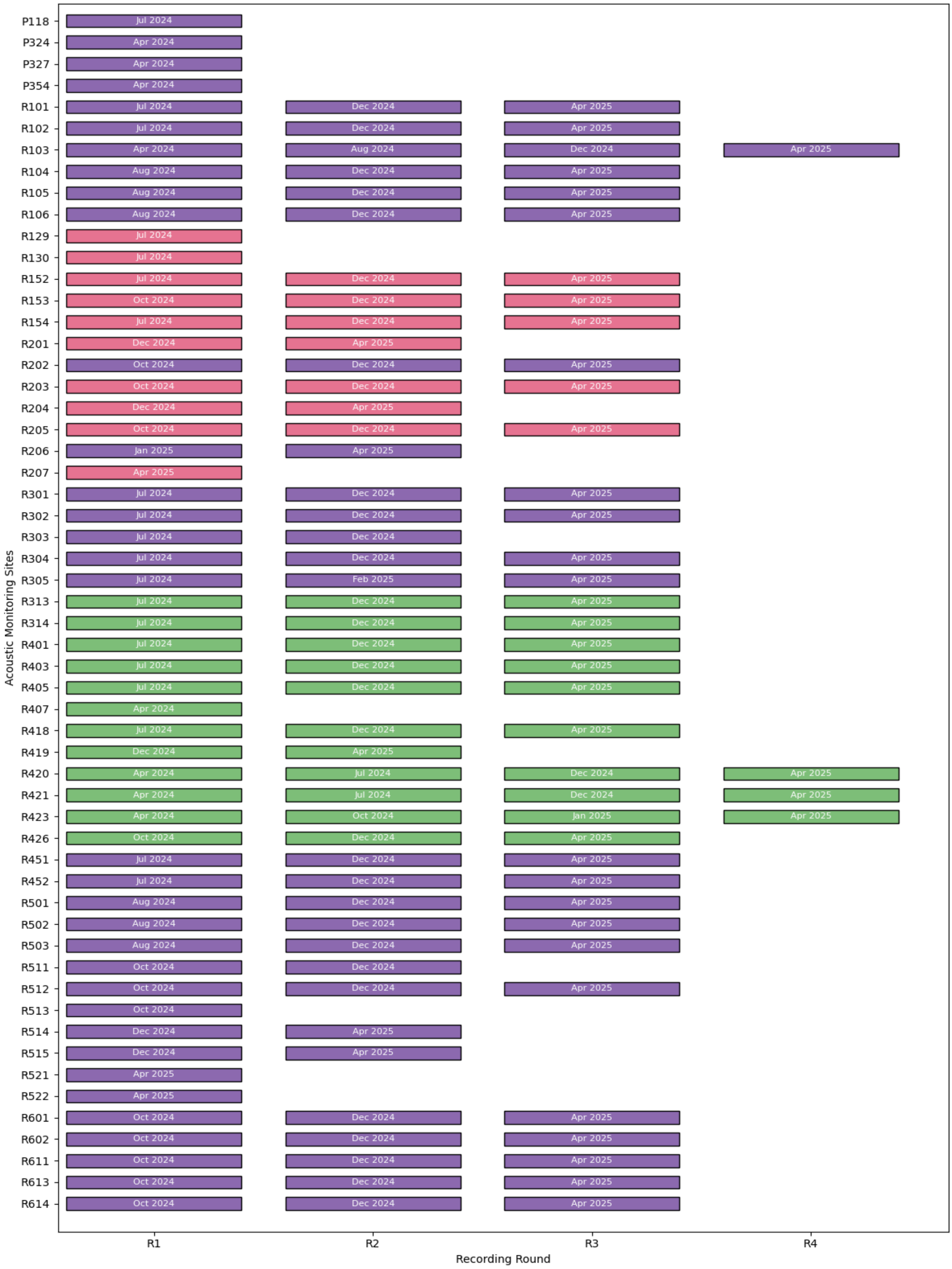

Figure 6. Recording rounds per sampling site, including locations that will no longer be monitored. Colored bars represent previous and new sites. Locations previously monitored in our first RSG in the Bacalar region are slightly green. Soft pink indicates new sampling locations included for this 2nd RSG, which collaborators in the north of the Bacalar region previously monitored. Purple indicates

new sampling locations included in this 2nd RSG, which have not been monitored before. The month and year of deployment are shown within the bar.

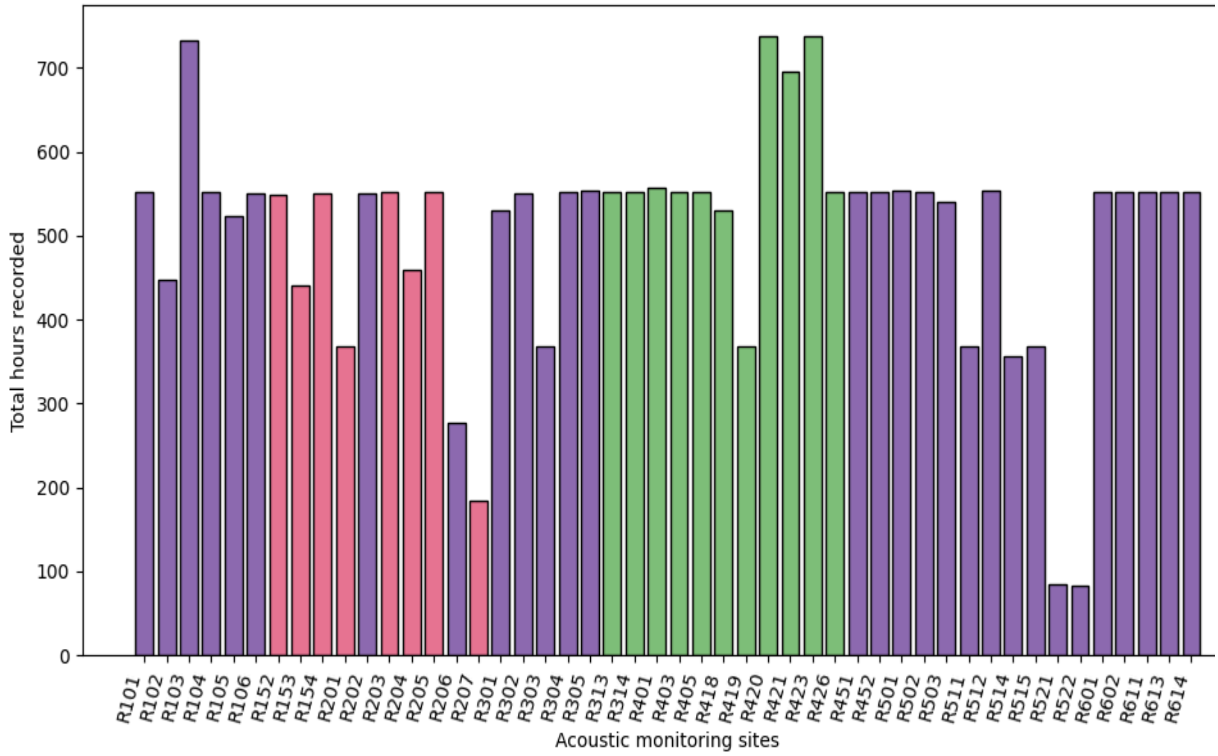

Figure 7. Total hours recorded for the 48 monitoring locations. Colored bars denote the previous and new sample sites.

Data processing of acoustic recordings

While continuing data collection, we are using the data collected up to February 2025 to run preliminary analyses that will be repeated with the entire dataset. One of these analyses consists in calculating acoustic indices that provide information about the acoustic characteristics of soundscapes related to biodiversity such as the Acoustic Complexity Index (ACI), the Bioacoustic Index (BI), the Number of Peaks (NP) and indices related to the anthropogenic sources such as the Normalized Difference Soundscape Index (NDSI).

Another type of analysis involves species detection based on their vocalizations at monitoring sites using semi-automated detection approach such as pattern matching, as well as machine learning solutions through the use of custom classifiers based on convolutional neural network (CNN) algorithms and data processing tool such as BirdNET-Analyzer (K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology and The Chemnitz University of Technology).

In collaboration with WildMon, a nonprofit organization that provides technology solutions for biodiversity and ecosystem conservation, we are utilizing customized CNN classifiers based on the pre-trained BirdNET algorithm. This approach enables us to optimize the detection of the species at our sampling sites, as the model is built and trained using acoustic data from our previous acoustic monitoring and additional sound clips from a public database. The first version of the custom model was trained using acoustic data from vocalizations of 20 species of birds and mammals from the Mesoamerican region (Table 1), including acoustic data collected during our first RSG project.

Table 1. Species included in the WildMon regional custom CNN model.

Taxon	**Species**	**Common name**
Bird	*Amazona albifrons*	White-fronted amazon
Mammal	*Ateles geoffroyi*	Geoffroy's spider monkey
Bird	*Attila spadiceus*	Bright-rumped attila
Bird	*Crypturellus cinnamomeus*	Thicket tinamou
Bird	*Cyanocorax yucatanicus*	Yucatan jay
Bird	*Cyclarhis gujanensis*	Rufous-browed peppershrike
Bird	*Dendrocolaptes sanctithomae*	Northern Barred-Woodcreeper
Bird	*Eucometis penicillata*	Grey-headed tanager
Bird	*Formicarius analis*	Black-faced Antthrush
Bird	*Micrastur semitorquatus*	Collared Forest Falcon
Mammal	*Nasua narica*	White-nosed coati
Bird	*Psilorhinus morio*	Brown jay
Bird	*Pteroglossus torquatus*	Collared aracari
Bird	*Ramphastos sulfuratus*	Keel-billed toucan
Bird	*Saltator atriceps*	Black-headed saltator
Bird	*Strix nigrolineata*	Black-and-white owl
Bird	*Thamnophilus doliatus*	Barred antshrike
Bird	*Tyrannus melancholicus*	Tropical Kingbird
Bird	*Uropsila leucogastra*	White-bellied wren
Bird	*Vireo griseus*	White-eyed Vireo

We are currently running the classifier in some monitoring localities to identify the presence of the species and its confidence scores (Fig. 8). For most species, the classifier showed a high rate of detections with a confidence score around 0.5 (i.e., the spider monkey, *Ateles geoffroyi*), other species show high confidence scores of classifications above 0.75 (i.e., the brown jay, *Psilorhinus morio*). In contrast, the classifier yields very low detections for some species, such as the black-headed saltator (*Saltator atriceps*). Despite these examples, the following steps involve translating the BirdNET confidence scores into prediction accuracy, which requires a manual evaluation of model performance on a subset of the data. Evaluation and validation of our custom classifier involves the local community and experts who collaborate with the project. This process enables us to determine whether future fine-tuning and additional training are necessary to achieve our research objectives.

The use of species detectors powered by machine learning approaches, combined with the measurement of acoustic indices at the monitoring locations, enables us to evaluate the environmental impact of the Tren Maya project on the occupancy of key conservation species and changes in acoustic activity and biodiversity.

For the activities of audio processing, we are developing a workflow to perform this analysis remotely and directly on the data in cloud storage, utilizing Command-Line Interface and Google Colab Notebooks we have written code examples in Python and R language that facilitate the processing and analysis of the acoustic data directly from cloud storage, which supports opportunities for collaboration and remote data processing. All the pipeline and code are hosted in a public repository and can be accessed at: https://github.com/ronaldehido/EchoSonR

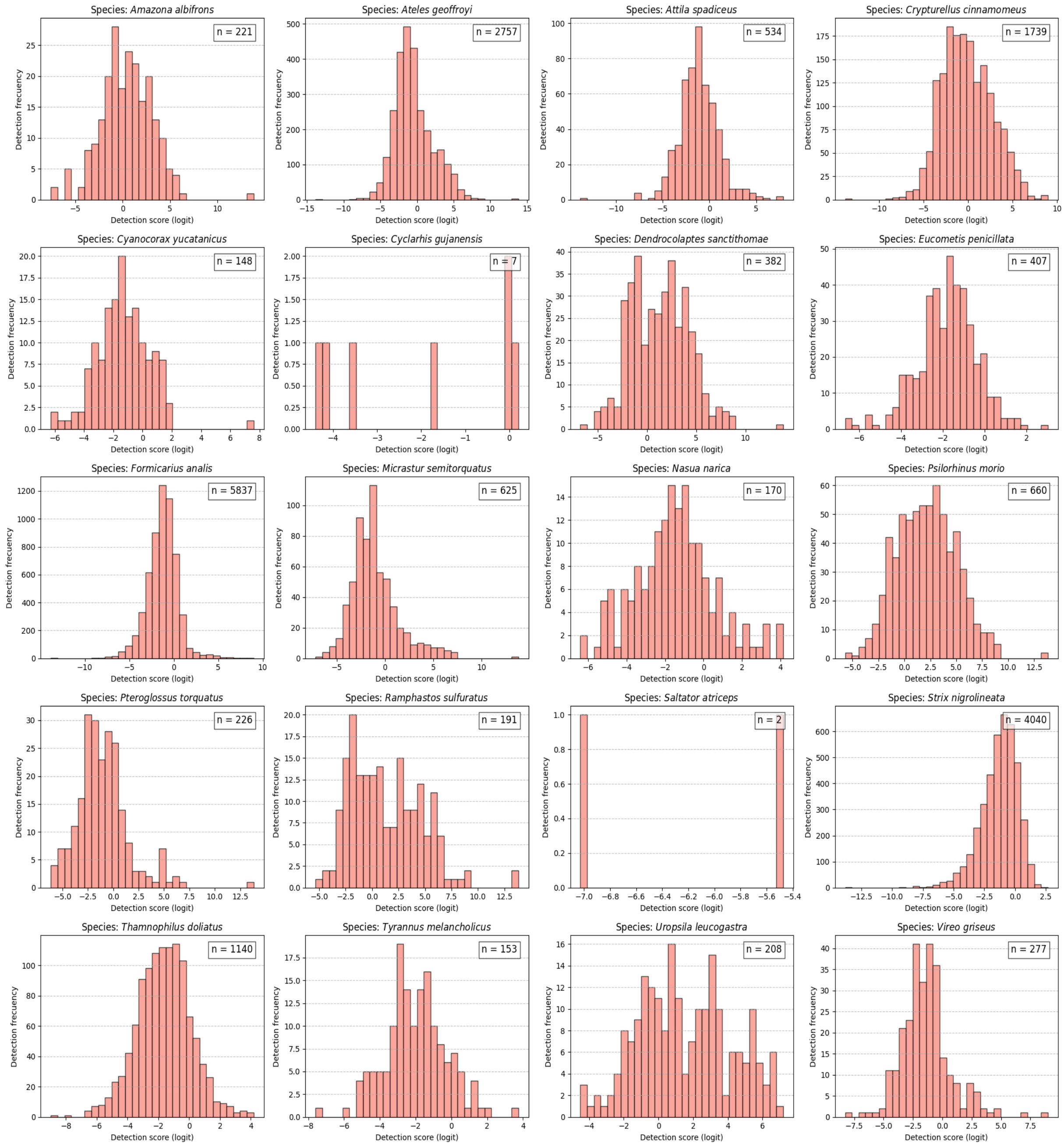

Figure 8. Example of BirdNET confidence scores from the detection of 20 species applying the WildMon regional custom classifier on data collected at one of our sampling sites (R101) in a forested area in the Felipe Carrillo Puerto municipality.