



HARRIET LEE-MERRISON

Spot the difference

Researchers are painstakingly poring over genomes, with a little help from artificial intelligence, to uncover thousands of new species – discoveries that could be key in the fight to rescue biodiversity, finds **Kate Douglas**



Bats are the most charismatic species out there,” says Bryan Carstens, an evolutionary biologist at The Ohio State University in Columbus. Here is an animal that evolved from the same ancestor as we did and yet most species now fly around, hunting down insects in the dark by shouting at them and listening for echoes. They live for decades, longer than other mammals their size, and they are extremely social.

For Carstens, bats are compelling for another reason, too: there are hundreds of unidentified species hiding in plain sight. By scrutinising the genes of the 1400 types of bats we know of, he and his colleagues have discovered 600 new ones. And these hidden species, which have been mistakenly lumped in with others, are just the tip of the iceberg, because Carstens and his team have identified thousands of missing mammals.

We shouldn’t be too surprised. It is estimated that between 90 and 99 per cent of all living species are yet to be identified. It is also becoming apparent that many new species have already been collected, but are languishing unrecognised in museums and other institutes. In other words, our knowledge of life on Earth is pitiful. Yet with ingenuity and hard graft, researchers have found a way to reveal the many species we have missed.

It comes not a moment too soon. Our activities are causing a mass extinction rivalling the one that killed off the dinosaurs, and we urgently need to know what species are out there if we are to have any hope of reversing this destruction of biodiversity.

Cataloguing nature is no mean feat. Taxonomists have described some 2 million species, but they only have a vague idea of the true number out there. “A lot of estimates about how many unknown species there might be are based on back-of-the-envelope calculations,” says Carstens. Then there is the issue of how you define a species. “The rule is that there aren’t really any rules,” he says.

Traditionally, taxonomists based their decisions on morphology: they would look at the physical attributes of organisms to decide whether they belonged together. But evolution tends to be a process of constant, gradual change and there is always variety within a population of organisms, so such judgements are inevitably subjective. In recent years, developments in genetics have led to

new ways of distinguishing between species. However, there is no consensus on which approach to take, and many genetic analyses ignore physical and behavioural traits, which are what evolution actually works on.

This year, one study aimed to avoid these pitfalls. Horseshoe bats are a rapidly expanding genus consisting of 106 known species, 18 of which have been identified since 2000. Alice Hughes at the University of Hong Kong and her colleagues examined hundreds of individuals from 11 of these. First, they compared the animals’ genes and then scrutinised two traits with evolutionary relevance: vocalisations and nose shape.

“The rule of how you define a species is that there aren’t really any rules”

Together, genes and morphology revealed that the 11 species contained 44 potential hidden species – indicating that there are around 40 per cent more Asian horseshoe bat species than we thought.

Such gold-standard studies are extremely laborious and still only give insights into a tiny fragment of biodiversity. What’s more, taxonomy lacks resources and people power. So how can we understand the bigger picture?

Enter Carstens and his team. With their usual fieldwork curtailed because of the pandemic, they tried to answer this question. Taking on the whole of nature would be too ambitious for a group of five: Carstens and three of his postgraduate students, Danielle Parsons, Jamin Wieringa and Drew Duckett, along with Tara Pelletier at Radford University in Virginia. Instead, they focused on mammals – but with the aim of creating a new way to explore hidden biodiversity that could be rolled out to other organisms.

Taxonomists have identified around 6500 species of mammal and the team looked at 4310 of these, representing all 27 mammalian orders (a group of related species). “We only included species for which we have solid genetic data,” says Parsons, the lead author of the recent paper describing

the work. In particular, the researchers needed information about two so-called barcoding genes. These are genes on DNA found in mitochondria, the powerhouses of cells where energy is produced. Barcoding genes have been shown to have very specific variations, says Parsons. The two they analysed are the most widely used to identify new mammal species. Ideally, you would want to look at more genes, she admits, but it is a trade-off. "There's less data, but also less cost and less time."

The researchers tested the barcoding data in two ways. The first involved measuring the genetic distance between pairs of individuals. This entails lining up the sequences of the barcoding genes to see how much they differ. If two entities are more than 5 per cent different from one another, they tend to be considered as different species, says Carstens.

The second method takes a more overtly evolutionary approach. It taps into the idea that, although genetic change tends to happen gradually and constantly within populations, when speciation occurs, many parents are passing novel gene variants to their offspring. This can be seen in the genetic data by looking



INAKEI RELANZON/NATUREPL.COM

There are probably about 40 per cent more horseshoe bat species than we thought

for lineages, where genetic changes are passed down generations without being lost, and plotting these against their time of occurrence. Gradual genetic change will generate few lineages, but when many lineages cluster, that indicates speciation. Although this approach tends to overestimate speciation, the genetic distance method tends to underestimate it. "So they balance each other," says Parsons.

Combining the two methods, the researchers found more than 2000 potential new species hidden within the genomes they examined. A conservative assessment indicated that one-third of the known species they scrutinised contained one or more hidden species. These were particularly common in three orders: bats, rodents and eulipotyphlans (see "Eu what?!", left). In total, the analysis indicated that about 20 per cent of mammals are yet to be identified.

But that was just the start. Next, the researchers wanted a way to predict where these hidden species might be found – crucial information for taxonomists trying to direct their time and resources most effectively. To do this, a massive amount of data had to be crunched: for each of the known species in the study, information had been collated on 117 traits, including aspects of morphology and life history, along with details of the climate, geography and environment they live in – 33 million facts in total. It wouldn't do to use a regular analysis to look for correlations because there are far too many variables and they may be interacting, says Carstens. Instead, they turned to artificial intelligence.

They used a technique called random forest analysis, similar to the approach employed by Netflix to predict which new programmes to recommend, based on your previous viewing. First, the researchers trained their AI on 80 per cent of the data. Then, they used the remaining 20 per cent to test it. This revealed that hidden species are most likely to be found in mammals with small bodies and large ranges that are living in the seasonal tropics, especially in South-East Asia.

Mammals are among the most studied groups of organisms. If our picture of them is so patchy, what does that tell us about estimates of how many of all species we have discovered so far? "I would guess it's closer to

Eu what?!

Eulipotyphla has a satisfying ring when it comes to names for groups of mammals. It translates as "truly fat and blind" and includes a ragtag bunch of hedgehogs, moles, shrews and other pointy-nosed insectivores, so odd that they have only recently been classified as a group. "They used to belong to a waste-bin taxa with all the other things taxonomists didn't really care to identify," says Danielle Parsons at The Ohio State University in Columbus. Nevertheless, to her, they are the most fascinating of mammals. "They're all very interesting,"

she says. "Some are venomous. They have a lot of weird stuff going on."

Parsons is particularly drawn to shrews. "They're amazing: their hearts beat at 1000 beats per minute, they eat three times their body weight a day, they barely sleep," she says. "North American water shrews are the world's smallest diving mammals and they can do really cool things like smell underwater."

Eulipotyphlans are also one of three mammalian orders that contain the most hidden species. In other words, there are many

unidentified new species hidden within the ones described to date, according to genetic analysis by Parsons and her colleagues. "The North American water shrew is being reclassified from a single species to four separate ones," she says. "And there are probably more."

The methods she helped pioneer (see main story) may yet lead to the discovery of new species in all mammal groups, from primates to pangolins, but it is these missing shrews that most excite Parsons. "They're just very charismatic," she says.

1 than 10 per cent,” says Carstens. “Think about mites. There are mites that live on mites. There are probably mites that live on the mites that live on mites.” The researchers believe their method could help predict where missing species are hiding. Admittedly, it can only reveal those lurking within already known specimens. Nevertheless, Carstens suspects there is a correlation with the bigger picture. “This might give a very rough estimate of how many undiscovered species are out there,” he says.

Mass extinction

Others are more cautious. One concern is that using genetic analysis to distinguish between species can lead to what is called taxonomic inflation. “Genetic diversity is higher in species with larger populations and high gene flow. So increasing genetic diversity does not mean more species,” says Mark Costello at Nord University, Norway. Stephen Garnett at Charles Darwin University, Australia, agrees. “Many taxonomic groups separated on the basis of a few mitochondrial genes do not stand up when examined closely,” he says. They both highlight the need to combine genetic analyses with studies of morphology. Even then, there are pitfalls. Hughes, who did just that in her study of horseshoe bats, points out that the most useful distinguishing characteristics may be quite idiosyncratic, like nose-leaf shape, the leaf-like structure on the nose of many bats which had been little studied before now. “If the wrong traits are measured, they will be less useful,” she says.

“My biggest issue with many high-volume molecular labs is the lack of morphological analysis and useful characteristics for distinguishing taxonomic groups,” says Scott Thomson at the Centre for the Study of Amazon Turtles in Brazil. Nevertheless, he sees the approach developed by Parsons and her colleagues as a “useful tool for focusing research” and encouraging much-needed investment. Hughes says the paper highlights the need for enhanced conservation efforts in critical regions.

Carstens and Parsons echo this view. “We are waking up to the fact that we have this giant mass extinction going on that we are responsible for. But people are still not fully



A waterfall in Lombok, Indonesia. Rainforests are thought to conceal many unknown species

“The analysis suggests about 20 per cent of mammals are yet to be identified”

on board with putting in the effort needed to correct it while we can,” says Parsons. “If we’re going to make intelligent decisions about conservation and have any hope of saving even a fraction of the biodiversity, we have to know what’s out there,” adds Carstens.

Their approach could also help with another 21st-century crisis. The covid-19 pandemic has alerted us to the dangers humanity faces from diseases that jump from other animals into humans. As Hughes and her colleagues point out, horseshoe bats are a major source of such pathogens, including coronaviruses like the one that causes covid-19. But they aren’t the only ones. To prevent future pandemics, we urgently need to know more about these disease-carrying species.

Ironically, without covid-19 restrictions, Parsons and her colleagues would never have had the time to do this research. “This was an emblematic pandemic science paper,” says Carstens. “It’s the kind of thing that wouldn’t have happened in normal times.” ■

Kate Douglas is a features editor at *New Scientist*