Most misidentified species on Inaturalist

Videocall #1

19 May 2020

Notes:

**Accessing data:**

* R package iNatTools seems to work to get data from the iNaturalist API.
  + <https://github.com/pjhanly/iNatTools>
* How do we find misidentified species?
  + **“Identifications” parameter can be set on:**
    - “Most agree”
    - “Some agree”
    - “Most disagree”

Couldn’t find a quantitative measure corresponding to these attributes, but API returns “num\_identification\_agreements” and “num\_identification\_disagreements” which must be linked to the “identifications” parameter above.

* + **API also returns “identifications\_count”**

**Idea:** Retrieve the “most disagree” observations for a time period / location and select the ones with “identification\_count > 1, in order to delete the observations where the identifier made a mistake and corrected him/herself.

**Problem:** The “identifications” parameter can change.

Ex: The first 2 identifiers give false identification A but the next 2 identifiers give true identification B. The identification parameter is set to ”most disagree”. However if the first 2 identifiers realize their mistake and correct their identification, the “identifications” parameter will change to most “most agree”. By only looking at the “most disagree” identifications we’ll miss the identifications that were misidentified but whose status has changed.

* + **Identifications quality**
    - <https://www.inaturalist.org/posts/10016-identification-quality-experiment-update>

|  |  |  |  |
| --- | --- | --- | --- |
| **Community identification** | **Leading identification** | **Improving identification** | **Maverick identification** |
| Consensus ID that emerges from everyone's IDs and dictates where an observation hangs on the tree of life | Taxon descends from community ID, but hasn't yet been corroborated by the community. | First suggestion of the taxon, that is corroborated by the community  Leading ID can become Improving ID | Identification that is in disagreement with the community.  Needs to be a lateral sibling to the community ID (rather than a descendant or ancestor) |
|  | | **Supporting identification** |  |
| Identification that corroborates a leading identification |

The “Maverick identification” could be a good indicator of a misidentified species.

API: identification\_category

Improving

Supporting

Leading

Maverick

I couldn’t find the identification category column in the data collected with the iNatTools R package but the following link seems to get the “Maverick” identifications.

<https://api.inaturalist.org/v1/identifications?current=true&category=maverick&order=desc&order_by=created_at>

* + **“Quality\_grade” parameter**

|  |  |  |
| --- | --- | --- |
| Casual | Needs\_Id | Research grade |
| Starting point of any observation | Observation is/has:   * dated * georeferenced * photos or sounds * isn’t captive/cultivated | Community agrees on species-level or lower.  2/3 of identifiers must agree on taxon |

Observation can revert from “Research” to “Needs\_id” if:

* the community ID shifts above the species-level
* the community votes that it needs more IDs

"Research Grade" at the genus level if the community agrees on a genus-level ID and votes that the observation does not need more IDs

iNat chooses the taxon with > 2/3 agreement, and if that's impossible, it walks up the taxonomic tree and chooses a taxon everyone agrees with, so if I say it's *Canis* and you say it's *Canis familiaris*, 2/2 identifications agree it's in *Canis* but only 1/2 think it's *Canis familiaris* so iNat goes with *Canis*.

Other ideas: Find a way to see the history / changes in identifications