# Inconsistencies in the AssignR database (v. 2.2.1):

# Dataset\_ID Dataset\_name

**1 Hobson et al. 1999 Oecologia**

* Sample sizes (n = 58 from 17 sites) do not match the original publication (n = 144 from 33 sites).
* Also, these data are the same wings/individuals that were analyzed in Hobson et al. (2019), although with different methods and additional isotopes. Therefore, the user could run into a situation of pseudoreplication by using both datasets together. It’s not entirely clear what these data are, because the sample sizes do not match what is reported in the publication, but our recommendation would be to remove this dataset.

**2 Hobson et al. 2012 PLOS**

* The Chabot et al. (2012) data for Loggerhead Shrikes *Lanius ludovicianus* (n = 556) are included within this dataset but were not used in Hobson et al. (2012). The biggest issue here is that although these are passerines, they are predators and more closely match the diet of raptors. In Maggozi et al. (2021) these are likely lumped with passerines.
* The paper is incorrectly cited and should be “Chabot AA, Hobson KA, Van Wilgenburg SL, McQuat GJ, Lougheed SC. 2012. Advances in linking wintering migrant birds to their breeding-ground origins using combined analyses of genetic and stable isotope markers. PLOS ONE 7:e43627.”.
* A Rufous-crowned Sparrow *Aimophila ruficeps* (Sample\_ID = 515) is listed as a waterbird but it should be a passerine. We first noticed this because in Fig. 3h of Maggozi et al. (2021) there is a single point from Hobson et al. (2012) in the waterfowl calibration and falls way off the predicted relationship.
* Sample sizes (n = 533) do not match the original publication (n = 544). The errors seem to be from the following species:

|  |  |  |
| --- | --- | --- |
| Species | AssignR n | Publication n |
| Gray Catbird *Dumetella carolinensis* | 1 | 2 |
| House Wren *Troglodytes aedon* | 2 | 1 |
| Rusty Blackbird *Euphagus carolinus* | 36 | 17 |
| Tree Swallow *Tachycineta bicolor* | 0 | 30 |

**3 Hobson and Wassenaar 1997 Oecologia**

* Least Flycatcher *Empidonax minimus* (n = 7) should be removed as we now know that they moult on the wintering grounds and the data are not known origin.

**4 Clark et al. 2006 Can J Zool**

* The data are listed as being ‘multiple waters’ which we assume refers to steam equilibration, but they were done under comparative equilibration method.
* Why are these presumed to be a known source? These are ducklings collected on the natal site. These should be clear case of known source.

**5 Hobson et al. 2004 Oecologia**

* Mallard spelling is “*Anas platyrhyncos*” – should be “*Anas platyrhynchos*” ending in ‘chos’.
* Black Woodpecker *Dryocopus maritus* (Sample\_ID = 1425) is listed as a passerine when it is a member of Piciformes (i.e., woodpeckers).
* There is a single sample from a ‘Turtle Dove’, which is missing from the original paper. Also, the Latin name for Turtle Dove should be provided in place of the common name (*Streptopelia turtur*, assuming this is a European Turtle Dove). Not clear where this sample came from.
* Eurasian curlew *Numenius arguata* could be listed as a ground bird, or at least a non-passerine like the Lapwing *Vanellus vanellus*. They are a grassland breeder although they do forage in mudflats. Grouping them with the other ducks may be inappropriate. Maybe waders would be a better grouping?
* Sample sizes (n = 132 from 65 sites) do not match the original publication (n = 141 from 38 sites). We realize that many of the missing samples are genus level Taxa (*Warbler sp.* n = 3; *tit sp.* n = 3, *thrush sp.* n = 3) which were omitted. Can these be added? Listing the taxa as sp. seems like a good option here. Also, adding the samples in Table 1 of the publication adds to 142, so there may be an error with that table (which we are using to compare the species-specific sample sizes). The following errors seem to be in the following species:

|  |  |  |
| --- | --- | --- |
| Species | AssignR n | Publication n |
| Common Wood Pigeon *Columba palumbus* | 19 | 20 |
| Fieldfare *Turdus pilaris* | 2 | 1 |
| Redwing *Turdus iliacus* | 3 | 5 |

* No source quality listed.

**6 Lott and Smith 2006 Auk**

* No issues found.
* No source quality listed.

**7 Hobson and Kohler 2015 Ecol Evol**

* Sample sizes (n = 105) do not match what was listed in the original publication (n = 104), but do match Appendix 1 (if you count the row), so we suspect it is an error with addition in the paper or an extra row was added to the Appendix. Either way, not an issue with AssignR.

**8 Thompson et al. 2010 Am J Phys Anthropol**

* Sample sizes (n = 25 from 25 sites) do not match the original publication (n = ? from 23 sites). It seems that these data are averaged to the site level, which should be clear in the database. Can a data field for ‘individual’ vs ‘site’ be added?
* Not a critique of the data stored in AssignR already, but it’s not clear why all of their data couldn’t be included here. The paper uses a subset of their data where they had at least two tap water measurements. This subset was necessary to correlate these values with tapwater, but the database could store all of these samples, because they are known origin human hair samples.

**9 Bowen et al. 2009 Am J Phys Anthropol**

* Didn’t verify.

**10 Ehleringer et al. 2008 PNAS**

* Didn’t verify.

**11 Wunder Plover**

* Didn’t verify. Three samples (Sample\_ID 2719-2721) have an ‘R’ at the end of their age. Is this meaningful or a typo?

**12 van Dijk et al. 2014 J Avian Biol**

* Sample sizes (n = 215 from 39 sites) do not match the original paper (n = 309). It seems like only the hatch-year birds (n = 215) were included here and the moulting adults were left out (n = 94). Is this on purpose? Also, we are surprised that the age data were not included here because age was an important consideration in their calibration.

**13 Neto et al. 2006 J Avian Biol**

* No issues found.

**16 Magozzi Towhee**

* Didn’t verify.
* Citation could be updated to reflect the publication details.

**17 Prochazka et al. 2013 J Avian Biol**

* No issues found.

**18 Langin et al. 2007 Oecologia**

* This one is a bit harder to interpret. From what we can tell, there are a handful of feathers (contour and primaries) sampled at one location. We think that the sample size is 93 (i.e., n = 42 P1 from adults, n = 13 contour from post-breeding adults, n = 34 contour from juveniles, n = 4 P1 from returning second year birds). Because multiple ages and feather types are being included here, can this information be included in the database?

**19 Bataille Canadian Human Hair**

* Didn’t verify.

**20 Hobson et al. 2019 FEE**

* No issues found.

# General Suggestions:

1. The ‘Source\_quality’ field is misleading as-is. Is this intended to be a filter where users can remove the ‘presumed’ known-origin data to increase confidence? Right now, it selects three datasets as being partially or completely ‘presumed’ known origin, but they are either incorrectly identified (see above) or are no more justified to be presumed than some of the other studies. This needs to be consistently applied across all datasets. Is the argument here that any flight-capable adult, regardless of ‘local’ non-migratory status (e.g., Hobson et al. (2004), Lott and Smith (2003)) or ‘returning breeders’ (e.g., Hobson and Kohler (2015), Hobson et al. (2012)) could be examples of ‘presumed known-origin’ because they are capable of movement? In other words, the only true known-origin samples juveniles sampled at natal sites (e.g., Prochazka et al. (2013), Clark et al. (2006))?