

Results of taxonomic harmonizations

```
d <- read_rds("biotime_results/harmonized.rds")
common <- read_csv("biotime_results/biotime_common.csv") %>%
  select(-class, -phylum) %>%
  mutate(common = modify(common, function(x)
    ifelse(is.na(x) | x == "mammals", "other", x)))

## Rows: 44382 Columns: 5

## -- Column specification -----
## Delimiter: ","
## chr (5): BioTIME, parsed, class, phylum, common

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

bogota <- d$bogota %>%
  filter(!is.na(binomial)) %>%
  transmute(binomial,
    common = modify(common, function(x)
      ifelse(is.na(x) | x == "mammals", "other", x))) %>%
  group_by(common) %>%
  tally() %>%
  pivot_wider(names_from = common, values_from = n) %>%
  add_column(workflow = "Bogota")

torino <- d$torino %>%
  filter(!is.na(binomial)) %>%
  transmute(binomial,
    common = modify(common, function(x)
      ifelse(is.na(x) | x == "mammals", "other", x))) %>%
  group_by(common) %>%
  tally() %>%
  pivot_wider(names_from = common, values_from = n) %>%
  add_column(workflow = "Torino")

gbif_prep <- d$gbif_preproc %>%
  filter(!is.na(binomial)) %>%
  mutate(common = modify(common, function(x)
    ifelse(is.na(x) | x == "mammals", "other", x))) %>%
  group_by(common) %>%
  tally() %>%
  pivot_wider(names_from = common, values_from = n) %>%
  add_column(workflow = "GBIF pre-processed")

gbif <- d$gbif %>%
  filter(!is.na(binomial)) %>%
  mutate(common = modify(common, function(x)
    ifelse(is.na(x) | x == "mammals", "other", x))) %>%
  add_column(workflow = "GBIF pre-processed")
```

```
group_by(common) %>%
tally() %>%
pivot_wider(names_from = common, values_from = n) %>%
add_column(workflow = "GBIF only")
```

```
res <- bogota %>%
  bind_rows(torino) %>%
  bind_rows(gbif_prep) %>%
  bind_rows(gbif) %>%
  pivot_longer(1:4, names_to = "common", values_to = "n") %>%
  pivot_wider(names_from = workflow, values_from = n)
```

In the *other* row, the 26 species are: 1 bird (*Turdus pallidus*), 24 vascular plants, and 1 fish (see below for the list).

```
wf_diff <- d$bogota %>%
  transmute(parsed, bogota = binomial,
            common = pmap(list(ebird, fishbase, lcvp), function(x, y, z) {
              if (!is.na(x))
                "birds"
              else if (!is.na(y))
                "fishes"
              else
                "plants"
            }) %>% unlist()) %>%
  distinct_all() %>%
  left_join(d$torino %>%
            transmute(parsed, torino = binomial) %>%
            distinct_all()) %>%
  mutate(bogota = modify(bogota, function(x) ifelse(is.na(x), "NA", x)),
         torino = modify(torino, function(x) ifelse(is.na(x), "NA", x))) %>%
  filter(bogota != torino)
```

```
## Joining, by = "parsed"
```

```
wf_diff %>%
  pull(common) %>%
  table()
```

```
## .
## birds fishes plants
##      1      1     24
```

```
wf_diff %>%
  pull(bogota)
```

```
## [1] "Aglaia oligophylla" "Aglaia rufibarbis" "Eremogone capillaris"
## [4] "Sabulina macra" "Polyscias biformis" "Isopogon tridens"
## [7] "Youngia longipes" "Eugenia confusa" "Eugenia patens"
## [10] "Eulalia aurea" "Ficus carica" "Ficus variegata"
## [13] "Fritillaria borealis" "Fritillaria haplostoma" "Oenothera anomala"
## [16] "Gaussia princeps" "Halophila gaudichaudii" "Inga laurina"
## [19] "Liparis atlanticus" "Swertia angustifolia" "Petitia amphophthalma"
## [22] "Polycarpaea violacea" "Spartina patens" "Tritonia gladiolaris"
## [25] "Turdus pallidus" "Strobocalyx arborea"
```

Unique names

Same procedure, but removing duplicate rows.

The first table shows the number of species names found and the second the number of unique species names found.

```
# I reassign them to the correct taxonomig group
res$Bogota[1] <- res$Bogota[1] + 1
res$Bogota[2] <- res$Bogota[2] + 1
res$Bogota[3] <- NA
res$Bogota[4] <- res$Bogota[4] + 24

res_unique$Bogota[1] <- res_unique$Bogota[1] + 1
res_unique$Bogota[2] <- res_unique$Bogota[2] + 1
res_unique$Bogota[3] <- NA
res_unique$Bogota[4] <- res_unique$Bogota[4] + 24

res %>% knitr::kable()
```

common	Bogota	Torino	GBIF pre-processed	GBIF only
birds	878	877	1113	1115
fishes	5414	5413	5587	5608
other	NA	NA	19780	19813
vascular plants	4569	4545	4703	4706

```
res_unique %>% knitr::kable()
```

common	Bogota	Torino	GBIF pre-processed	GBIF only
birds	878	877	1092	1093
fishes	5123	5122	5491	5496
other	NA	NA	19458	19466
vascular plants	4435	4412	4647	4649