From script 01-train-md5.R

> train.file %>% filter(md5 %in% names(duplicates))

filename species filesize md5

1 14864 11 3042 a51b1ffefa52fbe7b62aac22aa55b568

2 271 12 3042 a51b1ffefa52fbe7b62aac22aa55b568

All filenames are unique.

> train.dir[11:12,1:2]

directory species

11 11 chaetognath\_non\_sagitta

12 12 chaetognath\_other



