There was only one duplicate in the test and train sets:

filename species filesize md5

1 159128 8 2435 df7a547fc67ca47b38e83be45adf1c41

train\appendicularian\_straight



filename filesize md5

1 108193 2435 df7a547fc67ca47b38e83be45adf1c41

