

## entrez\_fetch returns multiple records from the nuccore database when given one ID

Thanks for making this package! It's a huge help. After a lot of successful use, I'm finding some very strange behavior. Basically when I try to fetch nuclotide data for one ID I'm getting many records back. Here's an example:

```
# get record
raw <- rentrez::entrez_fetch(db = 'popset', id = 'JN815406.1',
                             rettype = 'native', retmode = 'xml',
                             parsed = TRUE)

## No encoding supplied: defaulting to UTF-8.

# transform into a named vector (XML structure preserved in the names...
# there's probably a much better way to do this)
rawList <- unlist(XML::xmlToList(raw))
```

The ID should produce a result for the species *Drosophila murphyi*. But the query in fact produces many records, most of which are not for *Drosophila murphyi*:

```
# find all the times a species name is mentioned
allTax <- rawList[grepl('Org-ref_taxname', names(rawList))]
names(allTax) <- NULL

# there are a lot of times
length(allTax)
```

```
## [1] 212

# most are not for the expected species
sum(allTax == 'Drosophila murphyi')
```

```
## [1] 3
```

If I ask for the FASTA file it behaves as expected and gives me one record for the right species:

```
rawFASTA <- rentrez::entrez_fetch(db = 'nuccore', id = '403062956',
                                   rettype = 'fasta')
```

```
## No encoding supplied: defaulting to UTF-8.
```

```
cat(rawFASTA)
```

```
## >JN815406.1 Drosophila murphyi voucher M09059 elongation factor 1 gamma (EF-1g) gene, partial cds
## CAAATGTCTGACCGAGTCGAATGCCATTGCCCTACTTTTTGGCCAATGAGCAGCTGCGTGGCGGCAAATGT
## CCGCTGGTGCAGGCTCAGGTGCAGCAATGGATCTCATTGCGTGACAATGAAATCTGCCTGCGTCCTGCG
## CATGGGTGTTCCCACTGCTCGGCATAATGCCGCAGCAGAAGAATGCGAATGTGAAACGGGACGTTGAGGT
## TGTGCTGCAGCAGCTGAACAAGAAGCTGTTGGATGCCACTACCTCGCCGGTGAACGCATCACGTTGGCC
## GACATTGTTGTCTTCTGCACCCTGCTCCATTTGTATGAGCATGTRCTGGATTCAAGTGCACGCAGTGCGT
## ACGGCAATCTGAACCGTTGGTTCGTCACCATCCTCAATCAGCCGCAGGTGAAGGCTGTTGTCAAGGACTT
## TAAGCTGTGCGAAAAGCGCTCGTCTTTGATCCCAAGAAGTACGCCGAATTCCTGGCCAAGACTGGCGGT
## GCCAAGCCCCAGCAGGCGCCCAAGTCCAAGGATGAGAAAAAGGCCAAGAAGGAAGCGGCACCCGCACCCG
```

## AAMCCGAGGAGCTCGATGCTGCCGATGCCGCKTTGGCTATGGAGCCCAAGTCCAAGGATCCGTTTGATGC  
## CATGCCCAAGGGCACGTTCAATTTGATGACTTCAAGCGTGTCTATTCCAATGAGGAAGAGGCCAAGTCC  
## ATTCCCTATTTCTTTGAGAAATTCGATGCCGAGAACTATTCGATCTGGTTTGGCGAATACAAATACAACG  
## AAGAACTGACCAAGACTTTCATGTCCTGCAATCTGATCGGTGGCATG