

LAB 1 Bioinformatics

Hugo Morvan

2024-11-15

```
set.seed(12345)

library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

MM = 357
NN = 158
MN = 485
pop = 1000

p = sqrt(MM/pop)
q = sqrt(NN/pop)
pq = (MN/pop)/2

#assuming p is correct:
qe = 1-p
pqe_p = p*qe

chi_test_p = (q-qe)^2/qe + (pq-pqe_p)^2/pqe_p
print(chi_test_p)

## [1] 7.915616e-05

#assuming q is correct:
pe = 1-q
pqe_q = pe*q
chi_test_q = (p-pe)^2/pe + (pq-pqe_q)^2/pqe_q
print(chi_test_q)

## [1] 7.948755e-05

p_value_p <- pchisq(chi_test_p, df=2, lower.tail = FALSE)
print(p_value_p)

## [1] 0.9999604
```

```
p_value_q <- pchisq(chi_test_q, df=2, lower.tail = FALSE)
print(p_value_q)
```

```
## [1] 0.9999603
```

```
#2.2
```

```
MK465080.1 Branchipus schaefferi isolate ST2-2 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial TCTCCTAGGAGATGACCAACTTTATAACGTCATTGTTACTGCT-CACGCTTTTGTAAATGATTTTCTTCATA GTTATACCAATCCTTATTGGAGGATTTG-GAAATTGATTAGTCCCTTTAATACTAGGGGCTCCTGATATGG CTTTCCCCCGAC-TAAATAACTTAAGCTTTTGAATACTTCCTCCCTCATTAACTCTTCTAGTGGCCAGCTC-AATGGTAGAAAGAGGGGTAGGAACAGGATGAACAGTTTATCCACCCCTATCT-GCTGCTATTGCCCATGCT GGTCCTTCTGTTGATTTAGCAATCTTTTCACTTCAC-CTCGCAGGGATCTCTTCAATTTTAGGAGCTGTAA ATTTCATTACAACCTGTAAT-TAATATACGGCCTCATTCCATAAGATTAGACCGAATACCTTTATTTGCATG AGCG-GTAGTTATCACAGCAGTTCTTCTTCTCCTTTCTCTCCAGTATTAGCAG
```

```
From NLM /protein_id="QDG08361.1" /translation="LLGDDQLYNVIVTAHAFVMIFFMVMPILIGGFNGWLVPMLGLA-DMAFPRLNNSFWMLPPSLTLVASSMVESGVGTGWTVPPLSAAIAHAGPSVDLAIF SLHLAGIS-SILGAVNFITTVINMRPHSMSLDRMPLFAWAVVITAVLLLLSLPVLAX"
```

```
From transeq : >MK465080.1_2 Branchipus schaefferi isolate ST2-2 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial LLGDDQLYNVIVTAHAFVMIFFMVMPILIGGFNGWLVPMLL-GAPDMAFPRLNNSFWMLP PSLTLVASSMVESGVGTGWTVPPLSAAIAHAGPSVDLAIFSLH-LAGISSILGAVNFIT TVINMRPHSMSLDRMPLFAWAVVITAVLLLLSLPVLAX
```

```
from backtranseq:
```

```
Download Using codon table Caenorhabditis elegans (nematode)
```

```
MK465080.1_2 Branchipus schaefferi isolate ST2-2 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial CTTCTTGAGATGATCAACTTTATAATGT-TATTGTTACAGCTCATGCTTTCGTTATGATT TTCTTCATGGTTATGCCAATTCT-TATTGGAGGATTTCGGAATTGGCTTGTTCACCTTATG CTTGGAGCTCCAGATATG-GCTTTCCCAAGACTTAATAATCTTTCATTCTGGATGCTTCCA CCATCACTTA-CACCTTCTTGTTGCTTCATCAATGGTTGAATCAGGAGTTGGAACAGGATGG ACAGTT-TATCCACCACTTTCAGCTGCTATTGCTCATGCTGGACCATCAGTTGATCTTGCT-ATTTTCTCACTTCATCTTGCTGGAATTTTCATCAATTCTTGGAGCTGTTAATTTTCAT-TACA ACAGTTATTAATATGAGACCACATTCAATGTCACCTTGATAGAATGCCACTTTTCGCTTGG-GCTGTTGTTATTACAGCTGTTCTTCTTCTTCTTCACTTCCAGTTCTTGCTNNN
```

```
Using codon table drosophila melanogaster
```

```
MK465080.1_2 Branchipus schaefferi isolate ST2-2 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial CTGCTGGGCGATGATCAGCTGTACAACGTGATCGT-GACCGCCACGCCTTCGTGATGATC TTCTTCATGGTGATGCCCATCCTGATCGGCG-GCTTCGGCAACTGGCTGGTGCCCCCTGATG CTGGGCGCCCCCGATATGGCCTTCCC-CCGCCTGAACAACCTGAGCTTCTGGATGCTGCC CCCAGCCTGACCCTGCTGGTG-GCCAGCAGCATGGTGGAGAGCGGCGTGGGCACCGGCTGG ACCGTGTACCCCCC-CTGAGCGCCGCCATCGCCACGCCGGCCCCAGCGTGGATCTGGCC ATCTTCAGC-CTGCACCTGGCCGGCATCAGCAGCATCCTGGGCGCCGTGAACTTCATCACC ACCGT-GATCAACATGCGCCCCACAGCATGAGCCTGGATCGCATGCCCTGTTTCGCCTGG-GCCGTGGTGATCACCGCCGTGCTGCTGCTGCTGAGCCTGCCCGTGCTGGCCNNN
```

```
Using codon table drosophila melanogaster (high)
```

```
MK465080.1_2 Branchipus schaefferi isolate ST2-2 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial CTGCTGGGCGACGACCAGCTGTACAACGTGATCGT-
```

GACCGCCACGCCTTCGTGATGATC TTCTTCATGGTGATGCCCATCCTGATCGGCG-
 GCTTCGGCAACTGGCTGGTGCCCCTGATG CTGGGCGCCCCGACATGGCCTTCCC-
 CCGCCTGAACAACCTGTCTTCTGGATGCTGCC CCCTCCCTGACCCTGCTGGTG-
 GCCTCCTCCATGGTGGAGTCCGGCGTGGGCACCGGCTGG ACCGTGTACCCCCC-
 CTGTCCGCGCCATCGCCACGCGGCCCTCCGTGGACCTGGCC ATCTTCTCC-
 CTGCACCTGGCCGGCATCTCCTCCATCCTGGGCGCCGTGAACTTCATCACC ACCGT-
 GATCAACATGCGCCCCCACTCCATGTCCCTGGACCGCATGCCCTGTTGCGCTGG
 GCCGTGGTGATCACCGCCGTGCTGCTGCTGCTGTCCCTGCCCGTGCTGGCCNN

#2.3

Reverse (Caenorhabditis elegans codon table):

TTAGTATTGCTTTCGTACTCGACATTGTTATTGTAATATTTCAACTAGTAGAGGTTCTTC GTATTAC-
 CTTGTTTCGGTTAAAGGCTTAGGAGGTTATTCTTAACCGTATTGGTACTTCTT ACCTTCGTAG-
 GTCTTACTTTCTAATAATTCAGAACCCTTTCGGTATAGACCTCGAGGTTT GGTAGGACAAG-
 GTTGAGGACTAAGTTGGTAACTACTTCGTTGTTCTTCACATTCCTACC TCGTTCTAGTTGAC-
 TACCAGGTCGTACTCGTTATCGTCGACTTTCACACCTATTTGACA ACATTACTTTAATTGTC-
 GAGGTTCTTAACTACTTTAAGGTCGTTCTACTTCACTCTTTTA GGTTCGCTTTTCACCGTAA-
 GATAGTTCACTGTAACTTACACCAGAGTATAATTATTGACA NNNTCGTTCTTGACCTTCACTTTCTTCTTCTCT
 GACATTATTGTTGTCTG

Complement:

AATCATAACGAAAGCATGAGCTGTAACAATAACATTATAAAGTTGATCATCTCCAAGAAG
 CATAAGTGGAAACAAGCCAATTTCCGAATCCTCCAATAAGAATTGGCATAACCATGAAGAA TG-
 GAAGCATCCAGAATGAAAGATTATTAAGTCTTGGGAAAGCCATATCTGGAGCTCCAAG CCATC-
 CTGTTCCAACCTCCTGATTCAACCATTGATGAAGCAACAAGAAGTGTAAGTGATGG AGCAA-
 GATCAACTGATGGTCCAGCATGAGCAATAGCAGCTGAAAGTGGTGGATAAACTGT TGTAAT-
 GAAATTAACAGCTCCAAGAATTGATGAAATTCCAGCAAGATGAAGTGAGAAAAT CCAAGC-
 GAAAAGTGGCATTCTATCAAGTGACATTGAATGTGGTCTCATATTAATAACTGT NNNAGCAA-
 GAACTGGAAGTGAAAGAAGAAGAAGAACAGCTGTAATAACAACAGC

backtranseqmbig with invertebrate mitochondrial:

EMBOSS_001 YTNYTNGGNGAYGAYCARYTNTAYAAYGTNATYGTNACNGCNCAYGC-
 NTTYGTNATRATY TTYTTYATRGTNATRCCNATYYTNATYGGNGGNTTYGGNAAYT-
 GRYTNGTNCNNTNATR YTNNGGNGCNCNGAYATRGCNTTYCCNCGNYTNAAAYAYT-
 NWSNTTYTGRATRYTNCCN CCNWSNYTNACNYTNYTNGTNGCNWSNWSNATRGT-
 NGARWSNGGNGTNGGNACNGGNTGR ACNGTNTAYCCNCCNYTNWSNGCNGC-
 NATYGCNCAYGCNGGNCNWSNGTNGAYYTNGCN ATYTTYWSNYTNCAYYTNGC-
 NNGNATYWSNWSNATYYTNGGNGCNGTNAAYTTYATYACN ACNGTNATYAAYATR-
 CGNCCNCAYWSNATRWSNYTNGAYCGNATRCCNYTNTTYGCNTGR GCNGTNGT-
 NATYACNGCNGTNYTNYTNYTNYTNWSNYTNCCNGTNYTNGCNNNN

3.1

C. elegans was the first multicellular organism to have its whole genome sequenced, and in 2019[14] it was the first organism to have its connectome (neuronal “wiring diagram”) completed.