

Bioinformatics Report

Module 6

4624650

Anna Paula Pawlicka Maule

Exercise 1:

Measure the Dn / Ds ratio for CDS's (Coding Sequence) mtDNA of human, mouse and chimpanzee. The identification of sequences in GenBank are NC_012920, NC_005089, NC_001643 respectively.

Answer:

The following results were obtained by aligning the 3 genes using Clustal Omega and running the [SNAP online Tool](#)

| Compare | Sequence names | Sd | Sn | S | N | ps | pn | ds | dn | ds/dn | ps/pn |
|---------|-----------------------|-----------|-----------|-----------|------------|--------|--------|--------|--------|--------|--------|
| 0 | 1 NC_005089 NC_012920 | 1291.0000 | 3230.0000 | 3571.5000 | 12079.5000 | 0.3615 | 0.2674 | 0.4933 | 0.3307 | 1.4918 | 1.3518 |
| 0 | 2 NC_005089 NC_001643 | 1312.5000 | 3300.5000 | 3609.5000 | 12275.5000 | 0.3636 | 0.2689 | 0.4974 | 0.3330 | 1.4941 | 1.3524 |
| 1 | 2 NC_012920 NC_001643 | 452.1667 | 912.8333 | 3732.3333 | 12149.6667 | 0.1211 | 0.0751 | 0.1321 | 0.0792 | 1.6690 | 1.6125 |

Averages of all pairwise comparisons:

ds = 0.3743 dn = 0.2476 ds/dn = 1.5516 ps/pn = 1.4389

Averages of the first sequence compared to others:

ds = 0.4954 dn = 0.3318 ds/dn = 1.4929 ps/pn = 1.3521

The values for Dn/Ds are:

NC_005089 and NC_012920 : 0.6703831339955402

NC_005089 and NC_001643 : 0.669481302774427

NC_001643 and NC_012920 : 0.5995457986373959

Dn: is the average number of nucleotide differences between the sequences per non-synonymous site.

Ds: is the average number of nucleotide differences between sequences per synonymous site

A purifying selection (selection against deleterious non-synonymous substitutions) has definitely operated in all 3 genes. Some amino-acid substitutions have been may caused by selection, just not enough to overcome the effects of purifying selection.

Exercise 2:

Repeat the same exercise, this time using viral genomes (HIV or SARS, discussed in more detail in Chapter 7). Compare the results with those obtained in the exercise 1.

HIV:

| Compare | Sequence names | Sd | Sn | S | N | ps | pn | ds | dn | ds/dn | ps/pn |
|---------|-----------------------|-----------|-----------|-----------|------------|--------|--------|--------|--------|--------|--------|
| 0 | 1 AF033819. NC_005089 | 899.6667 | 3444.3333 | 1619.8333 | 6024.1667 | 0.5554 | 0.5718 | 1.0119 | 1.0777 | 0.9389 | 0.9714 |
| 0 | 2 AF033819. NC_012920 | 1035.5000 | 3647.5000 | 1795.8333 | 6328.1667 | 0.5766 | 0.5764 | 1.0984 | 1.0975 | 1.0009 | 1.0004 |
| 0 | 3 AF033819. NC_001643 | 973.3333 | 3484.6667 | 1681.8333 | 5986.1667 | 0.5787 | 0.5821 | 1.1076 | 1.1226 | 0.9867 | 0.9942 |
| 1 | 2 NC_005089 NC_012920 | 1136.3333 | 3341.6667 | 3527.6667 | 11976.3333 | 0.3221 | 0.2790 | 0.4209 | 0.3489 | 1.2063 | 1.1545 |
| 1 | 3 NC_005089 NC_001643 | 1140.3333 | 3429.6667 | 3574.5000 | 12157.5000 | 0.3190 | 0.2821 | 0.4155 | 0.3539 | 1.1742 | 1.1309 |
| 2 | 3 NC_012920 NC_001643 | 385.6667 | 972.3333 | 3708.5000 | 12035.5000 | 0.1040 | 0.0808 | 0.1119 | 0.0855 | 1.3097 | 1.2872 |

Averages of all pairwise comparisons:

ds = 0.6944 dn = 0.6810 ds/dn = 1.1028 ps/pn = 1.0898

Averages of the first sequence compared to others:

ds = 1.0726 dn = 1.0992 ds/dn = 0.9755 ps/pn = 0.9887

Dn/Ds values:

AF033819 and NC_005089 : 1.065026188358533

AF033819 and NC_001649 : 1.013542795232936

Since the Dn/Ds for the AF033819 and NC_005089, and AF033819 and NC_001649 are greater than 1, it is possible to conclude that the selection has caused some amino-acid substitution for the virus. Some substitutions also have been caused by genetic drift. A purifying selection also likely occurs, but is not strong enough to overcome the effects of the positive selection.

AF033819 and NC_012920 : 0.9991806263656227

A purifying selection has definitely operated at the HIV virus towards the human gene.

SARS:

| Compare | Sequence names | Sd | Sn | S | N | ps | pn | ds | dn | ds/dn | ps/pn |
|---------|-----------------------|-----------|-----------|-----------|------------|--------|--------|--------|--------|--------|--------|
| 0 | 1 AY274119. NC_005089 | 1503.3333 | 5497.6667 | 2621.1667 | 9759.8333 | 0.5735 | 0.5633 | 1.0852 | 1.0429 | 1.0406 | 1.0182 |
| 0 | 2 AY274119. NC_012920 | 1615.8333 | 5539.1667 | 2772.0000 | 9732.0000 | 0.5829 | 0.5692 | 1.1262 | 1.0669 | 1.0556 | 1.0241 |
| 0 | 3 AY274119. NC_001643 | 1621.1667 | 5583.8333 | 2728.6667 | 9586.3333 | 0.5941 | 0.5825 | 1.1783 | 1.1242 | 1.0481 | 1.0200 |
| 1 | 2 NC_005089 NC_012920 | 1123.3333 | 3359.6667 | 3485.3333 | 11982.6667 | 0.3223 | 0.2804 | 0.4212 | 0.3511 | 1.1998 | 1.1495 |
| 1 | 3 NC_005089 NC_001643 | 1124.1667 | 3455.8333 | 3534.3333 | 12161.6667 | 0.3181 | 0.2842 | 0.4139 | 0.3572 | 1.1587 | 1.1193 |
| 2 | 3 NC_012920 NC_001643 | 395.0000 | 958.0000 | 3685.5000 | 12019.5000 | 0.1072 | 0.0797 | 0.1157 | 0.0843 | 1.3725 | 1.3447 |

Averages of all pairwise comparisons:

ds = 0.7234 dn = 0.6711 ds/dn = 1.1459 ps/pn = 1.1126

Averages of the first sequence compared to others:

ds = 1.1299 dn = 1.0780 ds/dn = 1.0481 ps/pn = 1.0208

AY274119 and NC_005089: 0.9610210099520826

AY274119 and NC_012920: 0.9473450541644468

AY274119 and NC_001649 : 0.9540863956547569

It is possible to conclude that a purifying selection occurred at the SARS towards all the other 3 genes.