**Geometric Feature of DNA Sequences**

Hongjie Xu

**Supplementary Contents**

These contents offer extra information for this work.

* The first three data sets information is shown in Table 1-3.

**Table 1**. Information for the coding sequences (CDS) of the β-globin gene of 11 animals

|  |  |  |
| --- | --- | --- |
| Abbreviation | NCBI Accession Number | Length(bp) |
| Mouse | V00722 | 444 |
| Rat | X06701 | 444 |
| Opossum | J03643 | 444 |
| Bovine | X00376 | 438 |
| Goat | M15387 | 438 |
| Chimpanzee | X02345 | 376 |
| Gorilla | X61109 | 364 |
| Lemur | M15734 | 444 |
| Human | U01317 | 444 |
| Rabbit | V00882 | 444 |
| Gallus | V00409 | 444 |

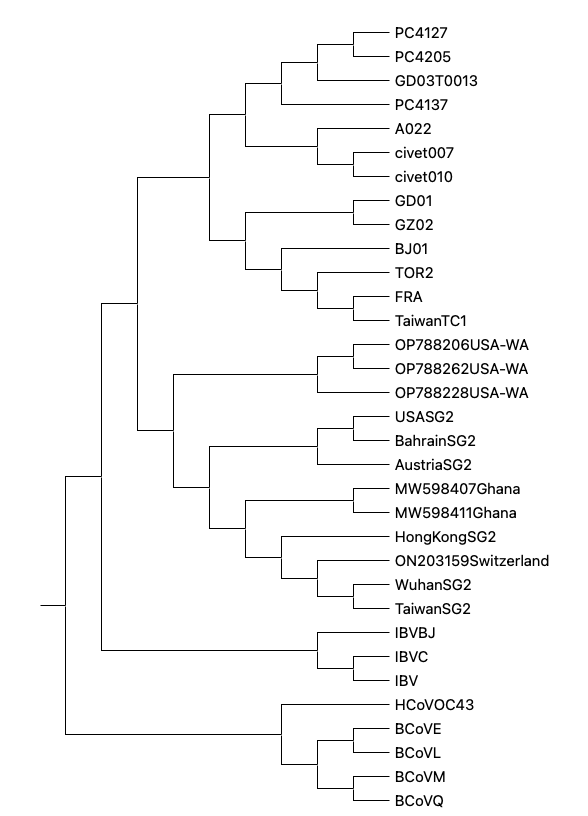
**Table 2**. Information for ribulose bisphosphate carboxylase small chain gene (rbcS) of 11 flowering plants

|  |  |  |  |
| --- | --- | --- | --- |
| Abbreviation | Gene ID | NCBI Accession Number | Length(bp) |
| Dura | 107460132 | NC\_029772.2 | 657 |
| Ipae | 107642447 | NC\_029785.2 | 663 |
| Rubber | 110661450 | NW\_018745897.1 | 552 |
| Cassava | 110616119 | NC\_035165.1 | 549 |
| Poplar | 105132993 | NW\_011499902.1 | 546 |
| Tobacco | 107802883 | NW\_015936331.1 | 543 |
| Tome | 104103374 | NW\_008893171.1 | 543 |
| Sunflower | 110889449 | NC\_035443.1 | 537 |
| Brome | 100840557 | NW\_014576701.1 | 528 |
| Sina | 102711124 | NW\_006267424.1 | 513 |
| Banana | 103976659 | NC\_025203.1 | 543 |

**Table 2**. Information for mitochondrial genome sequences of 11 mammals

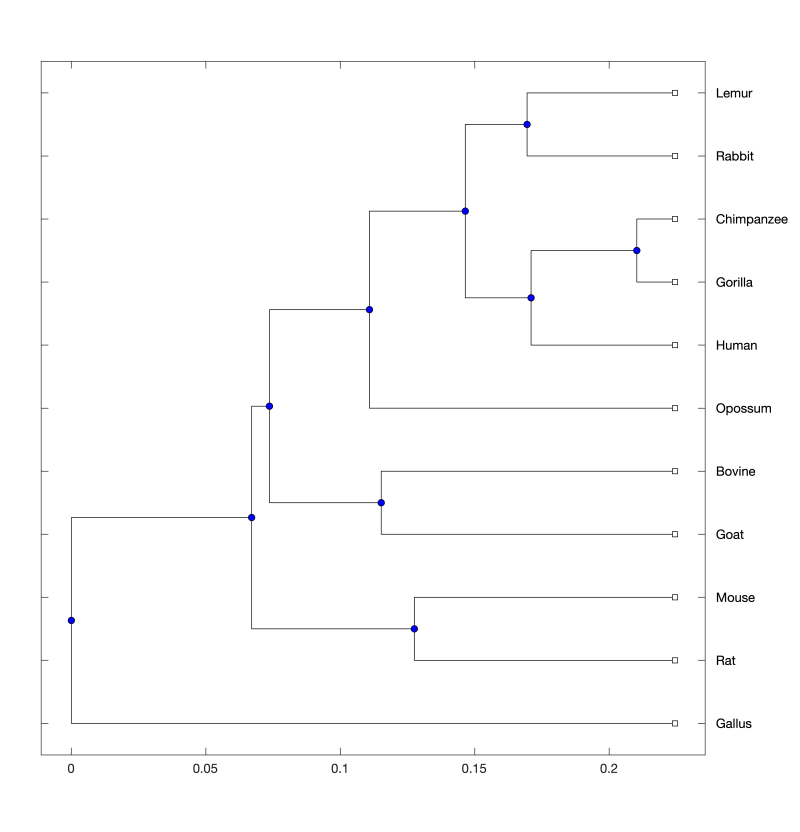
|  |  |  |
| --- | --- | --- |
| Abbreviation | NCBI Accession Number | Length(bp) |
| Pygmy chimpanzee | D38116 | 16563 |
| Common chimpanzee | D38113 | 16554 |
| Gorilla | D38114 | 16472 |
| Baboon | Y18001 | 16389 |
| Vervet monkey | AY863426 | 16586 |
| Cat | U20753 | 16364 |
| Dog | U96639 | 17009 |
| Wolf | EU442884 | 16355 |
| Tiger | EF551003 | 16774 |
| Leopard | EF551002 | 16990 |
| Macaca Thibet | NC\_002764 | 17447 |

* The phylogenetic tree generated by ClustalW algorithm which is performed on MEGA 11. Each group is assembled properly, which is similar to our result.

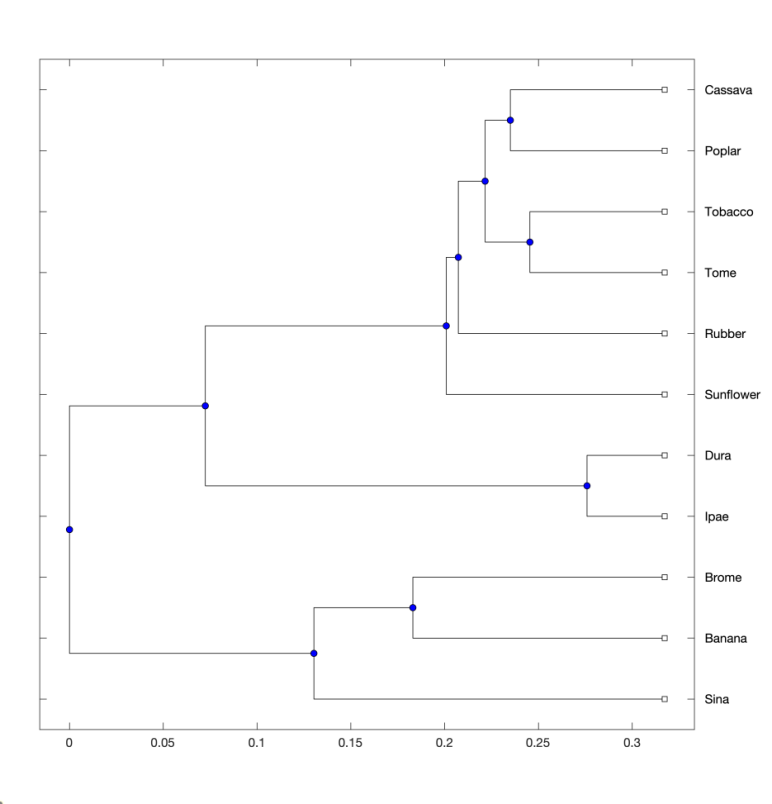


* The following 4 phylogenetic trees are obtained by Gong and Fan’s method.

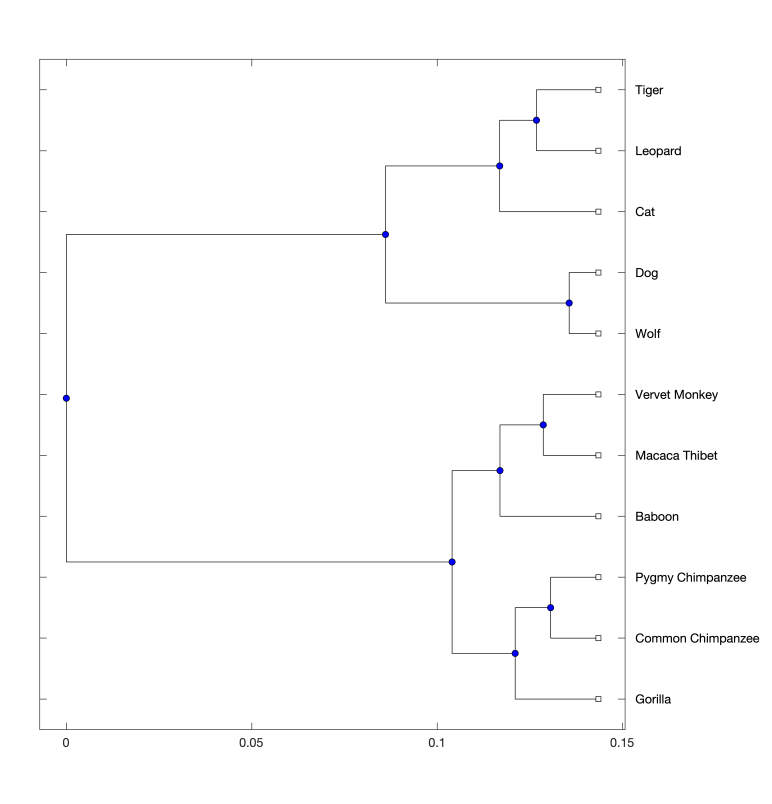
Phylogentic tree for the first data set



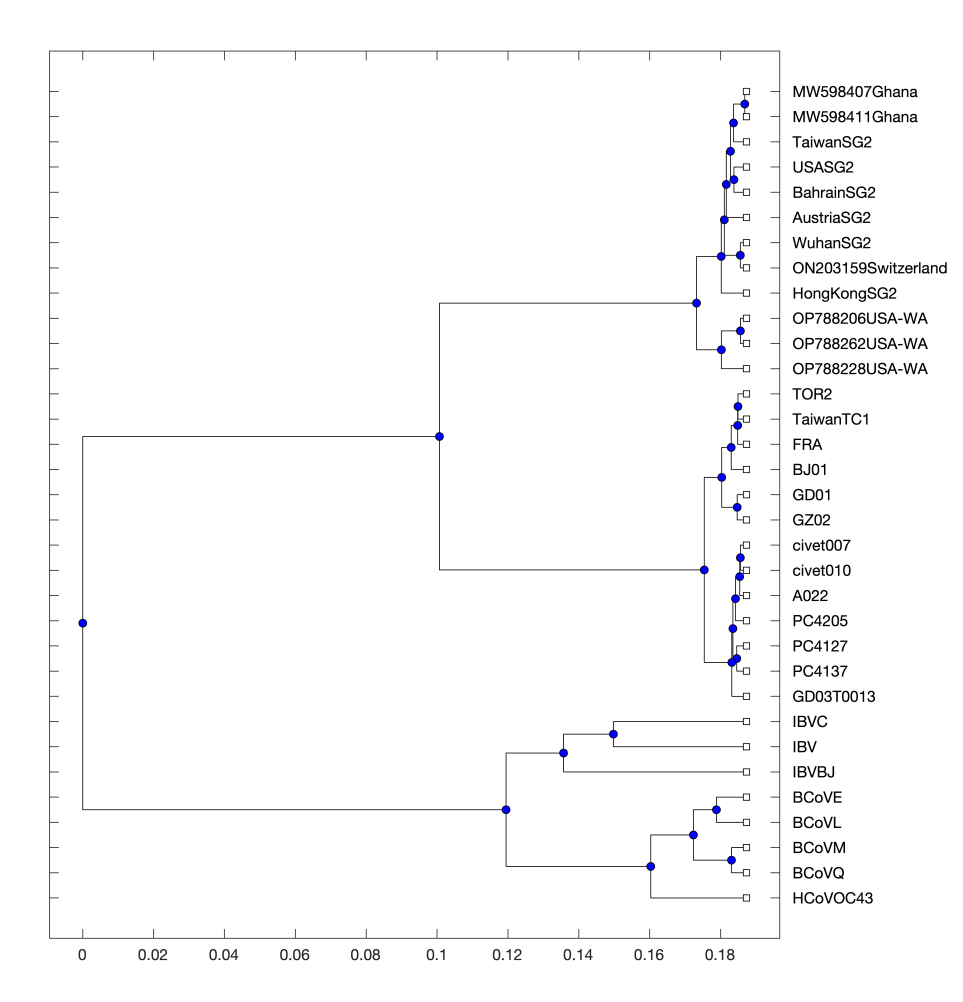
Phylogentic tree for the second data set



Phylogentic tree for the third data set



Phylogentic tree for the fourth data set



* The distance matrix for the first data set (Due to the big volume, it is divided into two tables.).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Bovine | Chimpanzee | Gallus | Goat | Gorilla |
| Bovine | 0 | 0.248282024 | 0.589737639 | 0.168874558 | 0.253913561 |
| Chimpanzee | 0.248282024 | 0 | 0.545246027 | 0.23494396 | 0.02576402 |
| Gallus | 0.589737639 | 0.545246027 | 0 | 0.535980204 | 0.537803758 |
| Goat | 0.168874558 | 0.23494396 | 0.535980204 | 0 | 0.231085544 |
| Gorilla | 0.253913561 | 0.02576402 | 0.537803758 | 0.231085544 | 0 |
| Human | 0.257471609 | 0.110776422 | 0.464586239 | 0.231094356 | 0.109015607 |
| Lemur | 0.247552179 | 0.112754123 | 0.532737812 | 0.179487745 | 0.101366563 |
| Mouse | 0.290873927 | 0.28814804 | 0.324981941 | 0.279334817 | 0.283544985 |
| Opossum | 0.25402689 | 0.271421789 | 0.553965441 | 0.267352613 | 0.265602258 |
| Rabbit | 0.230779569 | 0.149425239 | 0.560574784 | 0.1921186 | 0.141667945 |
| Rat | 0.280722472 | 0.314325502 | 0.463083771 | 0.342062833 | 0.316719466 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Human | Lemur | Mouse | Opossum | Rabbit | Rat |
| Bovine | 0.257471609 | 0.247552179 | 0.290873927 | 0.25402689 | 0.230779569 | 0.280722472 |
| Chimpanzee | 0.110776422 | 0.112754123 | 0.28814804 | 0.271421789 | 0.149425239 | 0.314325502 |
| Gallus | 0.464586239 | 0.532737812 | 0.324981941 | 0.553965441 | 0.560574784 | 0.463083771 |
| Goat | 0.231094356 | 0.179487745 | 0.279334817 | 0.267352613 | 0.1921186 | 0.342062833 |
| Gorilla | 0.109015607 | 0.101366563 | 0.283544985 | 0.265602258 | 0.141667945 | 0.316719466 |
| Human | 0 | 0.156675738 | 0.207452343 | 0.249380733 | 0.15904186 | 0.248824653 |
| Lemur | 0.156675738 | 0 | 0.285011832 | 0.239581844 | 0.116965263 | 0.336518329 |
| Mouse | 0.207452343 | 0.285011832 | 0 | 0.266239248 | 0.274387625 | 0.172820445 |
| Opossum | 0.249380733 | 0.239581844 | 0.266239248 | 0 | 0.187750909 | 0.219799288 |
| Rabbit | 0.15904186 | 0.116965263 | 0.274387625 | 0.187750909 | 0 | 0.285314441 |
| Rat | 0.248824653 | 0.336518329 | 0.172820445 | 0.219799288 | 0.285314441 | 0 |

* The distance matrix for the second data set (two tables below).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Dura | Ipae | Rubber | Cassava | Poplar | Tobacco |
| Dura | 0 | 0.059918767 | 0.582780827 | 0.584311228 | 0.550133701 | 0.478414394 |
| Ipae | 0.059918767 | 0 | 0.58140572 | 0.584146666 | 0.54637468 | 0.483038273 |
| Rubber | 0.582780827 | 0.58140572 | 0 | 0.190273731 | 0.1309312 | 0.22373488 |
| Cassava | 0.584311228 | 0.584146666 | 0.190273731 | 0 | 0.152142647 | 0.230182417 |
| Poplar | 0.550133701 | 0.54637468 | 0.1309312 | 0.152142647 | 0 | 0.158721911 |
| Tobacco | 0.478414394 | 0.483038273 | 0.22373488 | 0.230182417 | 0.158721911 | 0 |
| Tome | 0.474016829 | 0.479880307 | 0.193253297 | 0.242670659 | 0.159394289 | 0.123534921 |
| Sunflower | 0.56843802 | 0.565887855 | 0.207829542 | 0.195553068 | 0.147799271 | 0.246637571 |
| Brome | 1.074206942 | 1.066040193 | 0.615329392 | 0.534686368 | 0.574991474 | 0.680747427 |
| Sina | 1.231933525 | 1.220308112 | 0.71350927 | 0.682199499 | 0.696961651 | 0.798602072 |
| Banana | 1.02051774 | 1.014508289 | 0.51612038 | 0.489643264 | 0.50094794 | 0.619407377 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Tome | Sunflower | Brome | Sina | Banana |
| Dura | 0.474016829 | 0.56843802 | 1.074206942 | 1.231933525 | 1.02051774 |
| Ipae | 0.479880307 | 0.565887855 | 1.066040193 | 1.220308112 | 1.014508289 |
| Rubber | 0.193253297 | 0.207829542 | 0.615329392 | 0.71350927 | 0.51612038 |
| Cassava | 0.242670659 | 0.195553068 | 0.534686368 | 0.682199499 | 0.489643264 |
| Poplar | 0.159394289 | 0.147799271 | 0.574991474 | 0.696961651 | 0.50094794 |
| Tobacco | 0.123534921 | 0.246637571 | 0.680747427 | 0.798602072 | 0.619407377 |
| Tome | 0 | 0.238470047 | 0.700280669 | 0.817886777 | 0.621033907 |
| Sunflower | 0.238470047 | 0 | 0.541706661 | 0.699353084 | 0.457237147 |
| Brome | 0.700280669 | 0.541706661 | 0 | 0.272279103 | 0.193384896 |
| Sina | 0.817886777 | 0.699353084 | 0.272279103 | 0 | 0.310998149 |
| Banana | 0.621033907 | 0.457237147 | 0.193384896 | 0.310998149 | 0 |

* The distance matrix for the third data set (3 tables below).

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pygmy Chimpanzee | Common Chimpanzee | Baboon |
| Pygmy Chimpanzee | 0 | 0.021365574 | 0.052815478 |
| Common Chimpanzee | 0.021365574 | 0 | 0.058744179 |
| Baboon | 0.052815478 | 0.058744179 | 0 |
| Vervet Monkey | 0.06747094 | 0.075436244 | 0.058929783 |
| Macaca Thibet | 0.061443658 | 0.069290544 | 0.038757642 |
| Gorilla | 0.044335686 | 0.029455147 | 0.058102172 |
| Cat | 0.270732415 | 0.275035337 | 0.278809021 |
| Dog | 0.326806688 | 0.329391089 | 0.344935798 |
| Wolf | 0.331736127 | 0.334356412 | 0.349041446 |
| Tiger | 0.254485427 | 0.257639477 | 0.2618957 |
| Leopard | 0.253749545 | 0.255947397 | 0.260207516 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Vervet Monkey | Macaca Thibet | Gorilla | Cat |
| Pygmy Chimpanzee | 0.06747094 | 0.061443658 | 0.044335686 | 0.270732415 |
| Common Chimpanzee | 0.075436244 | 0.069290544 | 0.029455147 | 0.275035337 |
| Baboon | 0.058929783 | 0.038757642 | 0.058102172 | 0.278809021 |
| Vervet Monkey | 0 | 0.025728675 | 0.080001911 | 0.257278932 |
| Macaca Thibet | 0.025728675 | 0 | 0.072690238 | 0.263793599 |
| Gorilla | 0.080001911 | 0.072690238 | 0 | 0.2713788 |
| Cat | 0.257278932 | 0.263793599 | 0.2713788 | 0 |
| Dog | 0.32518886 | 0.333646519 | 0.326487947 | 0.108884662 |
| Wolf | 0.330782304 | 0.338533117 | 0.331238663 | 0.110415569 |
| Tiger | 0.250195605 | 0.253222547 | 0.251621899 | 0.053101377 |
| Leopard | 0.245636594 | 0.250102907 | 0.24846335 | 0.050048915 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Dog | Wolf | Tiger | Leopard |
| Pygmy Chimpanzee | 0.326806688 | 0.331736127 | 0.254485427 | 0.253749545 |
| Common Chimpanzee | 0.329391089 | 0.334356412 | 0.257639477 | 0.255947397 |
| Baboon | 0.344935798 | 0.349041446 | 0.2618957 | 0.260207516 |
| Vervet Monkey | 0.32518886 | 0.330782304 | 0.250195605 | 0.245636594 |
| Macaca Thibet | 0.333646519 | 0.338533117 | 0.253222547 | 0.250102907 |
| Gorilla | 0.326487947 | 0.331238663 | 0.251621899 | 0.24846335 |
| Cat | 0.108884662 | 0.110415569 | 0.053101377 | 0.050048915 |
| Dog | 0 | 0.012926506 | 0.113016448 | 0.112724116 |
| Wolf | 0.012926506 | 0 | 0.11177344 | 0.11292413 |
| Tiger | 0.113016448 | 0.11177344 | 0 | 0.024176388 |
| Leopard | 0.112724116 | 0.11292413 | 0.024176388 | 0 |

* The distance matrix for the fourth data set (Because it is too big, it is shown in an Excel table(double clicks)).

