# Supplemental File 1

**Supplemental File 1, Figure S1**

A picture containing text, device

Description automatically generated

**Supplemental Figure S1.** A phylogenetic tree of species in this study. Evolutionary history has been inferred from a multiple sequence alignment of the *rbcL* coding sequence in each species. The phylogeny is displayed as a cladogram for ease of visualisation, and species membership to different taxonomic groups (labelled) are highlighted by colour. Dark brown: red algae (*Rhodophyta*; *n* = 201). Light brown: SAR supergroup (*Stramenopiles*, *Alveolates*, and *Rhizaria*; *n* = 129). Lilac: bacteria (*Bacteria*; *n* = 78). Dark green: land plants (*Streptophyta*; *n* = 68). Light green: green algae (*Chlorophyta*; *n* = 12). Species at terminal nodes which are likely misplaced in this phylogeny are marked by an asterisk (\*).

**Supplemental File 1, Table S1**

**Table S1.** Summary of the percentile rate of nucleotide and protein evolution in each rubisco subunit in context of all other genes in each taxonomic group. Data have been calculated between each pairwise combination of species within a taxonomic group for which either a nuclear (land plants, green algae, red algae, SAR) or bacterial (bacteria) genome could be acquired. Statistics are rounded to three decimal places, and include the median, first quartile (Q1), third quartile (Q3), the interquartile range (IQR), the mean and the standard error (S.E.).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Taxonomic Group** | **Rubisco Subunit** | **Sequence** | **Percentile Rate** | | | | | |
| **Median** | **Q1** | **Q3** | **IQR** | **Mean** | **S.E.** |
| Bacteria | ***rbcL*** | **Nucleotide** | 2.583 | 0.856 | 5.641 | 4.785 | 4.711 | 0.315 |
| Land plants |  |  | 0.057 | 0.016 | 0.170 | 0.154 | 0.362 | 0.100 |
| Green algae |  |  | 0.042 | 0.019 | 0.088 | 0.069 | 0.052 | 0.016 |
| Red algae |  |  | 0.043 | 0.030 | 0.206 | 0.176 | 0.133 | 0.040 |
| SAR |  |  | 1.065 | 0.308 | 2.781 | 2.473 | 6.392 | 2.035 |
| Bacteria | **RbcL** | **Protein** | 2.647 | 0.960 | 5.122 | 4.162 | 3.983 | 0.244 |
| Land plants |  |  | 2.187 | 1.573 | 3.314 | 1.741 | 3.095 | 0.201 |
| Green algae |  |  | 0.979 | 0.476 | 1.695 | 1.219 | 1.173 | 0.343 |
| Red algae |  |  | 0.953 | 0.900 | 1.139 | 0.238 | 1.184 | 0.190 |
| SAR |  |  | 3.055 | 2.358 | 4.059 | 1.702 | 6.878 | 1.618 |
| Bacteria | ***rbcS*** | **Nucleotide** | 25.114 | 14.814 | 39.697 | 24.883 | 28.233 | 0.836 |
| Land plants |  |  | 64.057 | 52.487 | 76.510 | 24.023 | 61.859 | 1.089 |
| Green algae |  |  | 3.953 | 1.896 | 5.653 | 3.757 | 5.208 | 2.087 |
| Red algae |  |  | 1.253 | 1.008 | 1.963 | 0.955 | 1.392 | 0.204 |
| SAR |  |  | 7.048 | 1.269 | 15.865 | 14.595 | 20.456 | 5.489 |
| Bacteria | **RbcS** | **Protein** | 41.876 | 24.325 | 52.759 | 28.434 | 38.893 | 0.844 |
| Land plants |  |  | 52.024 | 36.134 | 67.519 | 31.384 | 52.231 | 1.002 |
| Green algae |  |  | 14.360 | 10.353 | 17.700 | 7.347 | 14.662 | 2.028 |
| Red algae |  |  | 8.708 | 7.083 | 16.700 | 9.617 | 11.424 | 1.698 |
| SAR |  |  | 16.717 | 4.676 | 20.884 | 16.208 | 21.167 | 3.874 |

**Supplemental File 1, Table S2**

**Supplemental File 1, Table S2.** Summary of the percentile rate of nucleotide and protein evolution in each rubisco subunit in context of all other enzyme-encoding genes in each taxonomic group. Data have been calculated between each pairwise combination of species within a taxonomic group for which either a nuclear (land plants, green algae, red algae, SAR) or bacterial (bacteria) genome could be acquired. Statistics are rounded to three decimal places, and include the median, first quartile (Q1), third quartile (Q3), the interquartile range (IQR), the mean and the standard error (S.E.).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Taxonomic Group** | **Rubisco Subunit** | **Sequence** | **Percentile Rate** | | | | | |
| **Median** | **Q1** | **Q3** | **IQR** | **Mean** | **S.E.** |
| Bacteria | ***rbcL*** | **Nucleotide** | 3.090 | 0.993 | 7.016 | 6.023 | 5.573 | 0.365 |
| Land plants |  |  | 0.070 | 0.031 | 0.174 | 0.143 | 0.384 | 0.108 |
| Green algae |  |  | 0.057 | 0.053 | 0.061 | 0.008 | 0.056 | 0.002 |
| Red algae |  |  | 0.092 | 0.076 | 0.161 | 0.085 | 0.152 | 0.037 |
| SAR |  |  | 0.618 | 0.302 | 2.871 | 2.569 | 5.885 | 1.952 |
| Bacteria | **RbcL** | **Protein** | 2.978 | 0.961 | 5.706 | 4.745 | 4.334 | 0.274 |
| Land plants |  |  | 2.688 | 1.868 | 4.146 | 2.278 | 3.802 | 0.234 |
| Green algae |  |  | 0.228 | 0.191 | 0.284 | 0.093 | 0.333 | 0.114 |
| Red algae |  |  | 0.510 | 0.452 | 0.887 | 0.435 | 0.923 | 0.262 |
| SAR |  |  | 2.778 | 1.874 | 3.709 | 1.835 | 6.657 | 1.684 |
| Bacteria | ***rbcS*** | **Nucleotide** | 30.412 | 18.002 | 46.773 | 28.771 | 32.932 | 0.944 |
| Land plants |  |  | 71.256 | 60.765 | 83.175 | 22.410 | 68.478 | 1.055 |
| Green algae |  |  | 2.862 | 1.017 | 5.632 | 4.615 | 4.291 | 1.815 |
| Red algae |  |  | 1.038 | 0.523 | 2.165 | 1.642 | 1.406 | 0.305 |
| SAR |  |  | 6.782 | 1.068 | 16.591 | 15.523 | 20.672 | 5.402 |
| Bacteria | **RbcS** | **Protein** | 47.314 | 29.298 | 56.569 | 27.271 | 43.035 | 0.887 |
| Land plants |  |  | 61.809 | 45.382 | 76.873 | 31.491 | 60.920 | 1.016 |
| Green algae |  |  | 18.046 | 12.067 | 23.666 | 11.599 | 17.880 | 2.612 |
| Red algae |  |  | 12.160 | 8.291 | 21.634 | 13.343 | 15.306 | 2.460 |
| SAR |  |  | 21.205 | 4.505 | 27.904 | 23.399 | 22.595 | 3.859 |

**Supplemental File 1, Table S3**

**Supplemental File 1, Table S3.** Summary of the rate of nucleotide and protein evolution in each Calvin-Benson-Bassham cycle enzyme as a percentage ratio (%) of that measured in the rubisco large subunit (*rbcL*/RbcL). Data have been calculated between each pairwise combination of species within the land plant clade for which a nuclear genome could be acquired. Statistics are rounded to three decimal places, and include the median, first quartile (Q1), third quartile (Q3), the interquartile range (IQR), the mean and the standard error (S.E.). RBCS: rubisco small subunit. PGK: phosphoglycerate kinase. GAPDH-A: glyceraldehyde-3-phosphate dehydrogenase A subunit. GAPDH-B: glyceraldehyde-3-phosphate dehydrogenase B subunit. TPI: triose phosphate isomerase. FBA: fructose-bisphosphate aldolase. FBP: fructose-1,6-bisphosphatase. TKL: transketolase. SBP: sedoheptulose-bisphosphatase. RPI: ribose 5-phosphate isomerase. RPE: ribulose-p-3-epimerase. PRK: phosphoribulokinase.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Enzyme** | **Sequence** | **% of *rbcL*/RbcL Molecular Evolution** | | | | | |
| **Median** | **Q1** | **Q3** | **IQR** | **Mean** | **S.E.** |
| RBCS | **Nucleotide** | 336.230 | 303.768 | 465.258 | 161.490 | 14215.472 | 6977.094 |
| PGK |  | 194.398 | 154.133 | 260.690 | 106.557 | 10222.926 | 8653.225 |
| GAPDH-A |  | 211.457 | 181.128 | 268.329 | 87.201 | 12074.030 | 9018.351 |
| GAPDH-B |  | 146.379 | 112.891 | 298.844 | 185.954 | 3293.398 | 1821.487 |
| TPI |  | 216.343 | 173.907 | 284.861 | 110.954 | 5679.185 | 3315.866 |
| FBA |  | 221.158 | 168.605 | 302.040 | 133.435 | 10306.953 | 8831.565 |
| FBP |  | 292.943 | 235.488 | 399.049 | 163.561 | 10201.657 | 5517.012 |
| TKL |  | 239.675 | 202.167 | 303.532 | 101.365 | 9195.482 | 5239.744 |
| SBP |  | 232.694 | 204.043 | 305.004 | 100.962 | 4667.535 | 3302.289 |
| RPI |  | 426.887 | 268.135 | 694.738 | 426.602 | 15006.493 | 11123.542 |
| RPE |  | 204.113 | 181.546 | 283.966 | 102.419 | 5462.525 | 4008.166 |
| PRK |  | 223.418 | 159.157 | 288.776 | 129.619 | 3818.743 | 2612.882 |
| RBCS | **Amino Acid** | 677.203 | 513.747 | 880.073 | 366.326 | 13630.350 | 6097.334 |
| PGK |  | 188.897 | 128.489 | 271.151 | 142.662 | 1478.183 | 1094.706 |
| GAPDH-A |  | 168.140 | 103.943 | 231.241 | 127.298 | 7539.044 | 5584.847 |
| GAPDH-B |  | 102.634 | 69.460 | 180.491 | 111.031 | 7341.694 | 5148.096 |
| TPI |  | 227.879 | 190.426 | 290.050 | 99.624 | 3948.902 | 2281.597 |
| FBA |  | 145.765 | 107.797 | 180.354 | 72.556 | 5257.912 | 5090.570 |
| FBP |  | 267.831 | 189.307 | 397.150 | 207.842 | 25149.320 | 19153.272 |
| TKL |  | 284.397 | 225.285 | 339.574 | 114.289 | 5429.319 | 3058.388 |
| SBP |  | 141.192 | 96.510 | 217.906 | 121.396 | 2784.952 | 2048.169 |
| RPI |  | 186.521 | 114.161 | 284.244 | 170.083 | 12075.717 | 11840.183 |
| RPE |  | 233.601 | 167.671 | 306.208 | 138.536 | 2686.190 | 1714.928 |
| PRK |  | 167.644 | 132.689 | 248.041 | 115.352 | 1892.755 | 1245.509 |

**Supplemental File 1, Table S4**

**Supplemental File 1, Table S4.** One-Sample Wilcoxon Signed Rank Test to assess significant differences in the rate of nucleotide and amino acid evolution between the rubisco large subunit (*rbcL*/RbcL) and each respective Calvin-Benson-Bassham cycle enzyme. A non-parametric test was used as data failed to conform to normality (Shapiro-Wilk test; *p* < 0.05). Statistics are rounded to three decimal places and corrected significance values are represented as α levels, where; α = 0.001 if *P* < 0.001, α = 0.01 if 0.001 < *P* < 0.01, α = 0.05 if 0.01 < *P* < 0.05, and α = ns if *P* > 0.05. Calvin-Bensen-Bassham cycle enzymes/subunits are abbreviated following the convention in table S3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Enzyme** | **Nucleotide** | | **Protein** | |
| **Statistic** | **α** | **Statistic** | **α** |
| **RBCS** | 2476425 | 0.001 | 2476417 | 0.001 |
| **PGK** | 51040 | 0.001 | 48937 | 0.001 |
| **GAPDH-A** | 55278 | 0.001 | 49267 | 0.001 |
| **GAPDH-B** | 2550 | 0.001 | 1689 | 0.01 |
| **TPI** | 59685 | 0.001 | 59570 | 0.001 |
| **FBA** | 8984 | 0.001 | 7727 | 0.001 |
| **FBP** | 60378 | 0.001 | 60261 | 0.001 |
| **TKL** | 60378 | 0.001 | 60378 | 0.001 |
| **SBP** | 59684 | 0.001 | 50188 | 0.001 |
| **RPI** | 37128 | 0.001 | 33896 | 0.001 |
| **RPE** | 45450 | 0.001 | 45030 | 0.001 |
| **PRK** | 50721 | 0.001 | 49118 | 0.001 |

**Supplemental File 1, Table S5**

**Supplemental File 1, Table S5.** The gene loci encoding the putative photosynthetic isoforms of Calvin-Benson-Bassham cycle enzymes in *Arabidopsis thaliana*. Accession numbers for each gene are provided as identification codes relating to The Arabidopsis Information Resource (TAIR; <https://www.arabidopsis.org/>) database. Phosphoglycerate kinase (PGK). Glyceraldehyde-3-phosphate dehydrogenase A subunit (GAPDH-A). Glyceraldehyde-3-phosphate dehydrogenase B subunit (GAPDH-B). Triose phosphate isomerase (TPI). Fructose-bisphosphate aldolase (FBA). Fructose-1,6-bisphosphatase (FBP). Transketolase (TKL). Sedoheptulose-bisphosphatase (SBP). Ribose 5-phosphate isomerase (RPI). Ribulose-p-3-epimerase (RPE). Phosphoribulokinase (PRK).

|  |  |
| --- | --- |
| **Gene** | **TAIR ID** |
| PGK | AT1G56190  AT3G12780 |
| GAPDH-A | AT1G12900  AT3G26650 |
| GAPDH-B | AT1G42970 |
| TPI | AT2G21170 |
| FBA | AT2G21330  AT4G38970 |
| FBP | AT3G54050 |
| TKL | AT3G60750 |
| SBP | AT3G55800 |
| RPI | AT3G04790 |
| RPE | AT5G61410 |
| PRK | AT1G32060 |