**Table S1. Selected statistics of genotypes at different divergence from five GFP sequences**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of amino acid mutations | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| amacGFP  | Number of genotypesMedian fluorescenceFraction wildtype-like\*Fraction dark\*\*Total possible genotypes Fraction sampled\*\*\* | 12143.9463.7% 9.0%476025.6% (84%) | 104393.944.7% 11.7%~107~0.1% | 63893.8531.2%19.6%~1.5 · 1010 ~10-7 | 30773.7521.6%27.9%~1.7 · 1013 ~10-10 | 12693.4415.2%38.4%~1.5 · 1016 ~10-13 | 5112.828.6%55.0%~1.1 · 1019 ~10-17 | 1882.748.5%63.3%~7 · 1021 ~10-20 |
| amacGFP:V12L | Number of genotypesMedian fluorescenceFraction wildtype-like\*Fraction dark\*\*Total possible genotypesFraction sampled\*\*\* | 10683.9365.0%5.8%476022.4% (74%) | 56213.8946.3%10.7%~107 ~0.06% | 30103.8533.0%16.8%~1.5 · 1010 ~10-7 | 13133.7824.1%27.3%~1.7 · 1013 ~10-10 | 5343.4616.7%42.1%~1.5 · 1016 ~10-14 | 2072.839.7%52.3%~1.1 · 1019 ~10-17 | 842.724.8%67.9%~7 · 1021 ~10-20 |
| cgreGFP | Number of genotypesMedian fluorescenceFraction WT-like\*Fraction dark\*\*Total possible genotypesFraction sampled\*\*\* | 11884.4344.4%14.5%470025.3% (84%) | 103474.3624.3%23.7%~9.9 · 106 ~0.1% | 65673.6211.4%45.2%~1.4 · 1010 ~10-7 | 36662.795.7%65.5%~1.6 · 1013 ~10-10 | 19592.772.0%81.4%~1.4 · 1016 ~10-13 | 10612.770.9%87.6%~1019 ~10-16 | 5462.770.2%92.3%~6.4 · 1021 ~10-20 |
| ppluGFP2 | Number of genotypesMedian fluorescenceFraction wildtype-like\*Fraction dark\*\*Total possible genotypesFraction sampled\*\*\* | 11634.266.1%4.4%444026.2% (87%) | 161344.1643.8%10.2%~8.9 · 106~0.18% | 89204.129.7%18.7%~1.2 · 1010 ~10-7 | 37103.9619.6%29.8%~1.2 · 1013 ~10-10 | 13703.6716.0%38.8%~1.1 · 1016 ~10-13 | 4563.3217.1%44.5%~7.3 · 1018 ~10-17 | 1862.8611.3%60.2%~4.3 · 1021 ~10-20 |
| avGFP | Number of genotypesMedian fluorescenceFraction wildtype-like\*Fraction dark\*\*Total possible genotypesFraction sampled\*\*\* | 11143.6468.9%9.4%4760 23.4% (90%) | 130103.5955.5%12.4%~107 ~0.13% | 126833.4939.1%27.0%~1.5 · 1010 ~10-7 | 97593.1423.3%47.9%~1.7 · 1013 ~10-10 | 72151.5313.1%68.4%~1.5 · 1016 ~10-13 | 46431.436.6%83.0%~1.1 · 1019 ~10-16 | 27831.362.3%92.0%~7 · 1021 ~10-19 |

\*Fraction wildtype-like refers to the fraction of genotypes displaying fluorescence levels within two standard deviations of the wildtype, or brighter.

\*\*Fraction dark refers to the fraction of fully non-functional genotypes (i.e. with fluorescence values falling within the darkest FACS gate). Remaining genotypes not accounted for in these two categories displayed a range of intermediate fluorescence levels.

\*\*\*The number of possible protein sequences of length L with N mutations, starting from a wildtype sequence with 19 possible amino acid substitutions at each site, can be calculated as $\frac{L}{N}· 19^{N}$. At N>1, it is not experimentally feasible to sample more than a tiny fraction of the theoretical space. For N=1, the percentage of protein genotypes requiring only a single nucleotide mutation are also shown, in parentheses.