**SUPPLEMENTARY MATERIALS**

**Table 1:** Tabular Representation of Cumulative Data Transfers from Source Directories to Output Directories

|  |  |  |
| --- | --- | --- |
| **Source Directories** | **Output Directories** | **Content** |
| directory\_1 | A | directory\_1 |
| directory\_2 | B | directory\_1, directory\_2 |
| directory\_3 | C | directory\_1, directory\_2, directory\_3 |
| directory\_4 | D | directory\_1, directory\_2, directory\_3, directory\_4 |
| directory\_5 | E | directory\_1, directory\_2, directory\_3, directory\_4, directory\_5 |
| directory\_6 | F | directory\_1, directory\_2, directory\_3, directory\_4, directory\_5, directory\_6 |
| directory\_7 | G | directory\_1, directory\_2, directory\_3, directory\_4, directory\_5, directory\_6, directory\_7 |
| directory\_8 | H | directory\_1, directory\_2, directory\_3, directory\_4, directory\_5, directory\_6, directory\_7, directory\_8 |
| directory\_9 | I | directory\_1, directory\_2, directory\_3, directory\_4, directory\_5, directory\_6, directory\_7, directory\_8, directory\_9 |
| directory\_10 | J | directory\_1, directory\_2, directory\_3, directory\_4, directory\_5, directory\_6, directory\_7, directory\_8, directory\_9, directory\_10 |
| directory\_11 | K | directory\_1, directory\_2, directory\_3, directory\_4, directory\_5, directory\_6, directory\_7, directory\_8, directory\_9, directory\_10, directory\_11 |

**Table 2:** Top 10 Gamma-proteobacterial Genes with Positive Association to CRISPR Presence

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Full Name** | **Odds Ratio** | **Naïve p-value** |
| bm3R1 | hypothetical protein | inf | 0.003348345 |
| ganB | Arabinogalactan endo-beta-1,4-galactanase | inf | 0.003359613 |
| blaI | hypothetical protein | inf | 0.005718477 |
| cas3 | CRISPR-associated nuclease/helicase Cas3 subtype I-F/YPEST | inf | 0.005913266 |
| ndvB | Cyclic beta-(1,2)-glucan synthase NdvB | inf | 0.005913266 |
| spmT | hypothetical protein | inf | 0.009948811 |
| group\_3747 | hypothetical protein | inf | 0.009948811 |
| cas6f | CRISPR-associated endonuclease Cas6/Csy4 | inf | 0.009948811 |
| iraP | Anti-adapter protein IraP | inf | 0.010053998 |
| yciF | Protein YciF | inf | 0.010053998 |

**Table 3:** Top 10Gamma-proteobacterial Genes with Negative Association to CRISPR Presence

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Full Name** | **Odds Ratio** | **Naïve p-value** |
| group\_12451 | hypothetical protein | 0 | 0.002471832 |
| group\_14047 | hypothetical protein | 0 | 0.002471832 |
| group\_14260 | hypothetical protein | 0 | 0.002471832 |
| group\_14576 | hypothetical protein | 0 | 0.002471832 |
| group\_14577 | hypothetical protein | 0 | 0.002471832 |
| group\_14587 | hypothetical protein | 0 | 0.002471832 |
| group\_14797 | hypothetical protein | 0 | 0.002471832 |
| group\_14799 | hypothetical protein | 0 | 0.002471832 |
| group\_14802 | hypothetical protein | 0 | 0.002471832 |
| group\_15455 | hypothetical protein | 0 | 0.002471832 |

**Table 4:** Top 10 Gamma-proteobacterial Genes with Positive Association to Cas Presence

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Full Name** | **Odds Ratio** | **Naïve p-value** |
| casD | hypothetical protein | inf | 5.34E-12 |
| casC | CRISPR system Cascade subunit CasC | inf | 5.34E-12 |
| casE | CRISPR system Cascade subunit CasE | inf | 2.24E-11 |
| casA | hypothetical protein | inf | 8.79E-08 |
| group\_5982 | hypothetical protein | inf | 1.23E-06 |
| group\_6979 | hypothetical protein | inf | 1.23E-06 |
| group\_7997 | hypothetical protein | inf | 5.84E-05 |
| group\_8220 | hypothetical protein | inf | 5.84E-05 |
| group\_7533 | hypothetical protein | inf | 0.000205845 |
| group\_8784 | hypothetical protein | inf | 0.000205845 |

**Table 5:** Top 10 Gamma-proteobacterial Genes with Negative Association to Cas Presence

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Full Name** | **Odds Ratio** | **Naïve p-value** |
| codB | Cytosine permease | 0 | 2.78E-07 |
| group\_3545 | hypothetical protein | 0 | 5.11E-07 |
| group\_3891 | hypothetical protein | 0 | 1.01E-06 |
| group\_4341 | hypothetical protein | 0 | 1.10E-06 |
| group\_4541 | hypothetical protein | 0 | 3.84E-06 |
| group\_4175 | hypothetical protein | 0 | 4.23E-06 |
| yfdG | Prophage bactoprenol-linked glucose translocase | 0 | 7.48E-06 |
| group\_5031 | hypothetical protein | 0 | 7.48E-06 |
| group\_5958 | hypothetical protein | 0 | 7.48E-06 |
| group\_4345 | hypothetical protein | 0 | 1.41E-05 |

**Table 6:** Summary Statistics of Genome Pan Analysis: Core, Soft Core, Shell, and Cloud Genes

|  |  |  |
| --- | --- | --- |
| **Genes** | **% Strains Present in** | **Number of Genes** |
| Core | (99% <= strains <= 100%) | 450 |
| Soft Core | (95% <= strains < 99%) | 188 |
| Shell | (15% <= strains < 95%) | 3232 |
| Cloud | (0% <= strains < 15%) | 60710 |
| Total | (0% <= strains <= 100%) | 64580 |

**Table 7:** Top 5 Genes with Lowest P-values for CRISPR Association

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Full Name** | **Odds Ratio** | **Naïve p-value** |
| pdhD | Dihydrolipoyl dehydrogenase | 0.159548751 | 1.55E-05 |
| ymdB | O-acetyl-ADP-ribose deacetylase | 14.11764706 | 0.000347844 |
| group\_8061 | hypothetical protein | 0.087619048 | 0.000631728 |
| group\_9271 | hypothetical protein | 0.054237288 | 0.000655585 |
| group\_6354 | hypothetical protein | 0.127906977 | 0.000872147 |

**Table 8:** Top 5 Genes with Lowest P-values for Cas Trait Association

|  |  |  |  |
| --- | --- | --- | --- |
| **Genes** | **Full Name** | **Odds Ratio** | **Naïve p-value** |
| csy3 | CRISPR-associated protein Csy3 | 80.46666667 | 5.56E-19 |
| cas3 | CRISPR-associated nuclease/helicase Cas3 subtype I-F/YPEST | 41.47058824 | 3.27E-15 |
| cas6f | CRISPR-associated endonuclease Cas6/Csy4 | 36.55555556 | 6.69E-14 |
| cas1 | CRISPR-associated endonuclease Cas1 | 29 | 9.99E-14 |
| csy1 | CRISPR-associated protein Csy2 | 48.57894737 | 1.60E-13 |