**Supporting Information**

**Supplementary Tables**

**Table S1.** The original values of the 10 physico-chemical properties for 16 dinucleotides.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dinucleotides** | **PC1** | **PC2** | **PC3** | **PC4** | **PC5** | **PC6** | **PC7** | **PC8** | **PC9** | **PC10** |
| AA | 3.18 | 7.0 | -0.08 | -1.27 | -0.8 | 31 | -6.82 | -18.4 | -13.7 | -0.9 |
| AC | 3.24 | 4.8 | 0.23 | -1.43 | 0.8 | 32 | -11.40 | -26.2 | -13.8 | -2.1 |
| AG | 3.3 | 8.5 | -0.04 | -1.50 | 0.5 | 30 | -10.48 | -19.2 | -14 | -1.7 |
| AT | 3.24 | 7.1 | -0.06 | -1.36 | 1.1 | 33 | -9.38 | -15.5 | -15.4 | -0.9 |
| CA | 3.09 | 9.9 | 0.11 | -1.46 | 1.0 | 31 | -10.44 | -27.8 | -14.4 | -1.8 |
| CC | 3.32 | 8.7 | -0.01 | -1.78 | 0.3 | 32 | -13.39 | -29.7 | -11.1 | -2.9 |
| CG | 3.3 | 12.1 | 0.3 | -1.89 | -0.1 | 27 | -10.64 | -19.4 | -15.6 | -2 |
| CT | 3.3 | 8.5 | -0.04 | -1.50 | 0.5 | 30 | -10.48 | -19.2 | -14.0 | -1.7 |
| GA | 3.38 | 9.4 | 0.07 | -1.70 | 1.3 | 32 | -12.44 | -35.5 | -14.2 | -2.3 |
| GC | 3.22 | 6.1 | 0.07 | -1.39 | 0.0 | 35 | -14.88 | -34.9 | -16.9 | -3.4 |
| GG | 3.32 | 12.1 | -0.01 | -1.78 | 0.3 | 32 | -13.39 | -29.7 | -11.1 | -2.9 |
| GT | 3.24 | 4.8 | 0.23 | -1.43 | 0.8 | 32 | -11.40 | -26.2 | -13.8 | -2.1 |
| TA | 3.26 | 10.7 | -0.02 | -1.45 | -0.2 | 32 | -7.69 | -22.6 | -16.0 | -1.1 |
| TC | 3.38 | 9.4 | 0.07 | -1.70- | 1.3 | 32 | -12.44 | -26.2 | -14.2 | -2.1 |
| TG | 3.09 | 9.9 | 0.11 | -1.46 | 1.0 | 31 | -10.44 | -19.2 | -14.4 | -1.7 |
| TT | 3.18 | 7.0 | -0.08 | -1.27 | -0.8 | 31 | -6.82 | -18.4 | -13.7 | -0.9 |

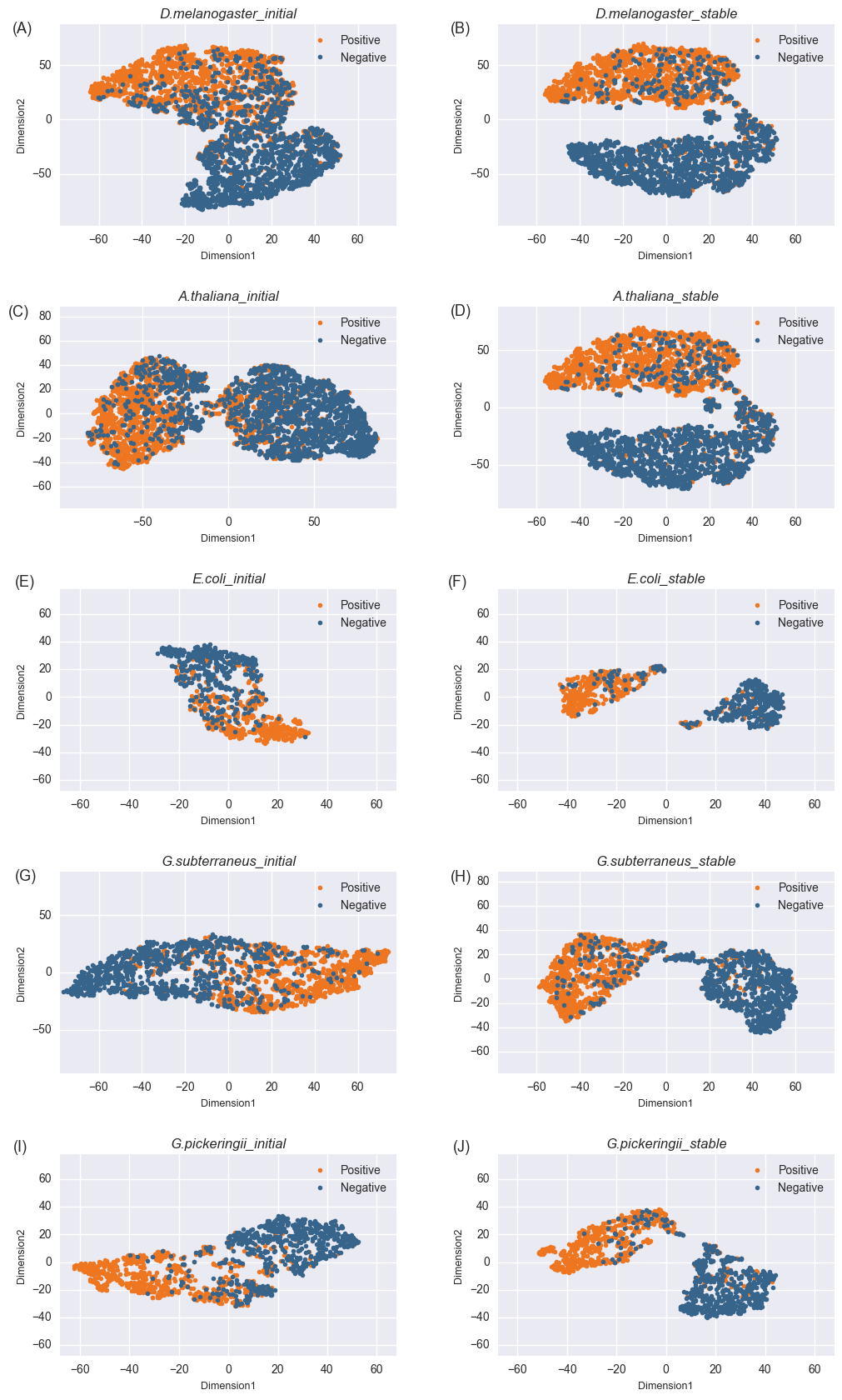
**Table S2.** The original values of the three physicochemical properties for 16 dinucleotides

|  |  |  |  |
| --- | --- | --- | --- |
| **Dinucleotides** | **Enthalpy** | **Entropy** | **Free energy** |
| GG | -12.2 | -29.7 | -3.26 |
| GA | -13.3 | -35.5 | -2.35 |
| GC | -14.2 | -34.9 | -3.42 |
| GT | -10.2 | -26.2 | -2.24 |
| AG | -7.6 | -19.2 | -2.08 |
| AA | -6.6 | -18.4 | -0.93 |
| AC | -10.2 | -26.2 | -2.24 |
| AT | -5.7 | -15.5 | -1.10 |
| CG | -8.0 | -19.4 | -2.36 |
| CA | -10.5 | -27.8 | -2.11 |
| CC | -12.2 | -29.7 | -3.26 |
| CT | -7.6 | -19.2 | -2.08 |
| TG | -7.6 | -19.2 | -2.11 |
| TA | -8.1 | -22.6 | -1.33 |
| TC | -10.2 | -26.2 | -2.35 |
| TT | -6.6 | -18.4 | -0.93 |

**Table S3.** EIIP values of nucleotides

|  |  |
| --- | --- |
| **Nucleotide** | **EIIP** |
| A | 0.1260 |
| C | 0.1340 |
| G | 0.0806 |
| T | 0.1335 |

**Supplementary figures**

****

**Figure S1. t-SNE visualization of the D.melanogaster, A.thaliana, E.coli, G.subterraneus and G.pickeringii datasets in a two-dimensional feature space.** The left column ((A) (C) (E) (G) (I)) The original feature space (before iterative algorithm), and the right column ((B) (D) (F) (H) (J)) the stable feature space (after iterative algorithm).