

Characterization of DNA methylation as a  
function of biological complexity via  
dinucleotide inter-distances  
SUPPLEMENTARY MATERIAL

Giulia Paci<sup>1</sup>, Giampaolo Cristadoro<sup>3</sup>, Barbara Monti<sup>4</sup>,  
Marco Lenci<sup>2,3</sup>, Mirko Degli Esposti<sup>3</sup>, Gastone C. Castellani<sup>1,2</sup> and  
Daniel Remondini<sup>\*1,2</sup>

<sup>1</sup>Department of Physics and Astronomy, University of Bologna,  
Viale B. Pichat 6/2, 40127 Bologna, Italy

<sup>2</sup>INFN, Bologna Unit, Viale B. Pichat 6/2, 40127 Bologna, Italy

<sup>3</sup>Department of Mathematics, University of Bologna,  
Piazza di Porta S. Donato 5, 40126 Bologna, Italy

<sup>4</sup>Department of Pharmacy and Biotechnology, University of  
Bologna, Via S. Donato 15, 40127 Bologna, Italy

---

\*Corresponding author

Table 1: List of the organisms and their DNA sequence repository website.

Organism	Repository
Human Adenovirus 54	<a href="http://www.ncbi.nlm.nih.gov/nuccore/253761974">www.ncbi.nlm.nih.gov/nuccore/253761974</a>
Apis mellifera (honey bee, release 4.5)	<a href="http://hymenopteragenome.org/beebase/q=download_sequences">hymenopteragenome.org/beebase/q=download_sequences</a>
Bos taurus (cow)	<a href="http://www.ensembl.org/Bos_taurus/Info/Index">www.ensembl.org/Bos_taurus/Info/Index</a>
Caenorhabditis elegans (round worm)	<a href="http://www.ensembl.org/Caenorhabditis_elegans/Info/Index">www.ensembl.org/Caenorhabditis_elegans/Info/Index</a>
Canis familiaris (dog)	<a href="http://www.ensembl.org/Canis_familiaris/Info/Index">www.ensembl.org/Canis_familiaris/Info/Index</a>
Ciona intestinalis (sea vase)	<a href="http://www.ensembl.org/Ciona_intestinalis/Info/Index">www.ensembl.org/Ciona_intestinalis/Info/Index</a>
Danio rerio (Zebrafish)	<a href="http://www.ensembl.org/Danio_rerio/Info/Index">www.ensembl.org/Danio_rerio/Info/Index</a>
Drosophila Melanogaster (Fruit Fly)	<a href="http://www.ensembl.org/Drosophila_melanogaster/Info/Index">www.ensembl.org/Drosophila_melanogaster/Info/Index</a>
Equus caballus (horse)	<a href="http://www.ensembl.org/Equus_caballus/Info/Index">www.ensembl.org/Equus_caballus/Info/Index</a>
Escherichia Coli	<a href="http://www.genome.wisc.edu">www.genome.wisc.edu</a>
Homo Sapiens (man, release hg19)	<a href="http://hgdownload.cse.ucsc.edu/downloads.html#human">hgdownload.cse.ucsc.edu/downloads.html#human</a>
Macaca mulatta (rhesus monkey)	<a href="http://www.ensembl.org/Macaca_mulatta/Info/Index">www.ensembl.org/Macaca_mulatta/Info/Index</a>
Monodelphis domestica (opossum)	<a href="http://www.ensembl.org/Monodelphis_domestica/Info/Index">www.ensembl.org/Monodelphis_domestica/Info/Index</a>
Mus musculus (mouse)	<a href="http://www.ensembl.org/Mus_musculus/Info/Index">www.ensembl.org/Mus_musculus/Info/Index</a>
Oikopleura diotica (tunicate)	<a href="http://www.genoscope.cns.fr/externe/GenomeBrowser/Oikopleura">www.genoscope.cns.fr/externe/GenomeBrowser/Oikopleura</a>
Ornithorhynchus anatinus (platypus)	<a href="http://www.ensembl.org/Ornithorhynchus_anatinus/Info/Index">www.ensembl.org/Ornithorhynchus_anatinus/Info/Index</a>
Pan troglodytes (chimpanzee)	<a href="http://www.ensembl.org/Pan_troglodytes/Info/Index">www.ensembl.org/Pan_troglodytes/Info/Index</a>
Rattus norvegicus (rat)	<a href="http://www.ensembl.org/Rattus_norvegicus/Info/Index">www.ensembl.org/Rattus_norvegicus/Info/Index</a>
Saccharomyces cerevisiae R64-1-1	<a href="http://www.ensembl.org/Saccharomyces_cerevisiae">www.ensembl.org/Saccharomyces_cerevisiae</a>
Tetraodon nigroviridis (puffer fish)	<a href="http://www.ensembl.org/Tetraodon_nigroviridis/Info/Index">www.ensembl.org/Tetraodon_nigroviridis/Info/Index</a>
Tribolium castaneum (beetle)	<a href="http://metazoa.ensembl.org/Tribolium_castaneum/Info/Index">metazoa.ensembl.org/Tribolium_castaneum/Info/Index</a>

Table 2: Power-law fit of all human dinucleotide distributions. For each dinucleotide, the fit parameters  $b$ , the goodness of fit  $r^2$ , the P-value of the normalized Chi-square test  $P(\chi^2)$  are shown. All errors are expressed as 95% confidence intervals, and rounded to the first significant digit. Only the dinucleotide CG distribution is significantly non compatible with a power-law distribution.

Dinucleotide	$b$	$r^2$	$P(\chi^2)$
AA	$-3.1 \pm 0.2$	0.98	1
AC	$-3.7 \pm 0.2$	0.94	1
AG	$-2.9 \pm 0.2$	0.94	1
AT	$-3.5 \pm 0.1$	0.99	1
CA	$-3.1 \pm 0.2$	0.93	1
CC	$-3.6 \pm 0.2$	0.98	0.99
CG	$-2.7 \pm 0.4$	0.83	0.00082
CT	$-3.0 \pm 0.2$	0.96	1
GA	$-3.2 \pm 0.2$	0.96	1
GC	$-4.1 \pm 0.2$	0.98	0.99
GG	$-3.6 \pm 0.2$	0.97	0.99
GT	$-3.8 \pm 0.3$	0.95	1
TA	$-3.6 \pm 0.2$	0.98	0.99
TC	$-3.2 \pm 0.2$	0.96	1
TG	$-3.3 \pm 0.2$	0.96	1
TT	$-2.9 \pm 0.1$	0.98	1

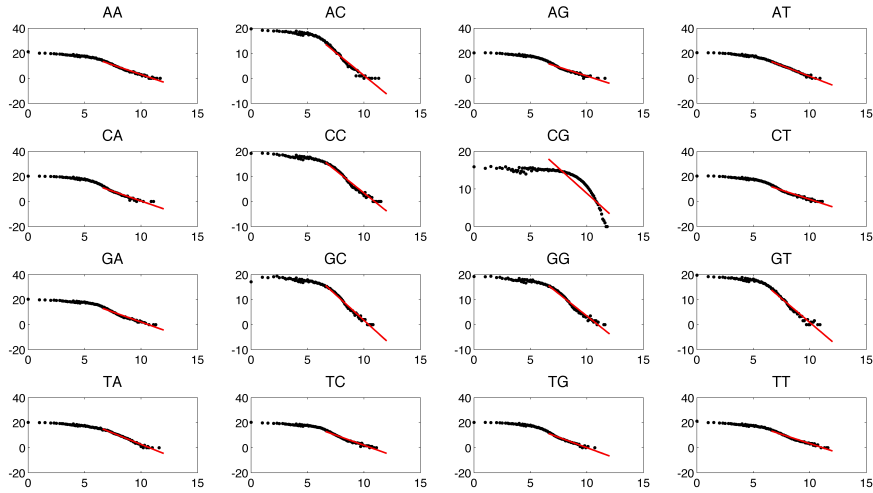


Figure 1: Double logarithmic plot of the dinucleotide distance distributions for human, together with the power-law fit (red line). The curves were fitted in the tails ( $d > 90$ , corresponding to  $x = 6.5$  in logarithmic scale).

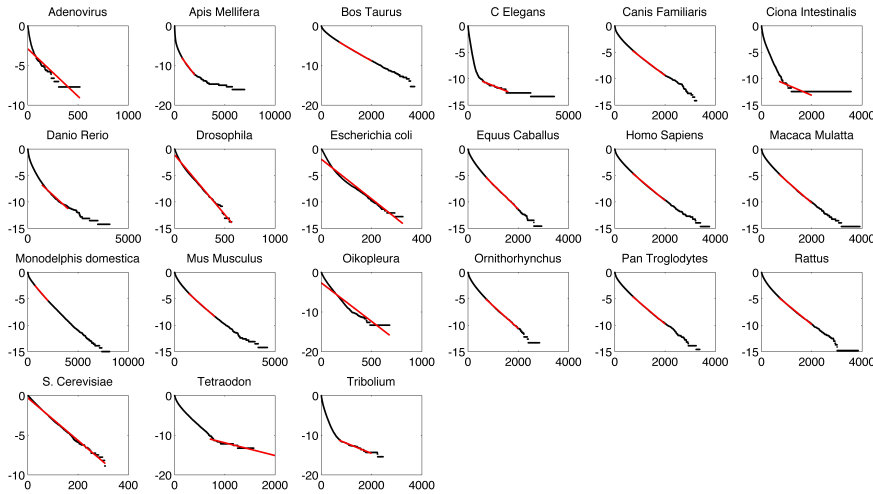


Figure 2: Plot of the cumulative distributions for all the studied organisms in semi-logarithmic scale, together with the exponential fit (red line). The curves were fitted in the interval  $700 < d < 2000$ .