Supplementary Materials for

Trinucleotide’s quadruplet symmetries and natural symmetry law of
DNA creation ensuing Chargaff’s second parity rule

Marija Rosandić a, Ines Vlahović b*, Matko Glunčić b and Vladimir Paar a, b

a Croatian Academy of Sciences and Arts, 10000 Zagreb, Croatia;
b Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia.

*Corresponding author. Email: ines@phy.hr

Full addresses of all authors:

Marija Rosandić,
Croatian Academy of Sciences and Arts, HAZU, Bioinformatics and Biological
Physics, Zrinski trg 11, 10000 Zagreb, Croatia
--385 99 2420 147
rosandic@hazu.hr

Ines Vlahović,
Faculty of Science, University of Zagreb, Bijenicka 32, 10000 Zagreb, Croatia
--385 99 8043 837
ines@phy.hr

Matko Glunčić,
Faculty of Science, University of Zagreb, Bijenicka 32, 10000 Zagreb, Croatia
--385 98 9033 926
matko@phy.hr

Vladimir Paar
Croatian Academy of Sciences and Arts, HAZU, Bioinformatics and Biological
Physics, Zrinski trg 11, 10000 Zagreb, Croatia
--385 95 9150 109
paar@hazu.hr
This file includes:

Materials and Methods

Table S1

1 Materials and Methods

1.1 Equivalence class of quadruplets composed of trinucleotides

Each set of four trinucleotides $\tau, RC(\tau), C(\tau), R(\tau)$ is defined as an equivalence class (14) of trinucleotides, where $\tau$ denotes any of the 43 = 64 trinucleotides, RC($\tau$) its reverse complement, C(\tau) its complement and R(\tau) reverse. Such a set is here referred to as trinucleotide’s quadruplet.

According to CSPR, for any trinucleotide holds a relation among frequencies $f$ of trinucleotides within its quadruplet

$$f(\tau) \approx f(RC(\tau)) and f(C(\tau)) \approx f(R(\tau)) .$$

Thus to every quadruplet correspond two frequency subgroups, $f(\tau) \approx f(RC(\tau))$ and $f(C(\tau)) \approx f(R(\tau))$. In most cases, for actual genomic sequences the frequencies of these two subgroups mutually differ.

In a random sequence of nucleotides, all four frequencies within a quadruplet are similar if the frequency of appearance of all four nucleotides are the same:

$$f(\tau) \approx f(RC(\tau)) \approx f(C(\tau)) \approx f(R(\tau)) .$$

Thus, for actual DNA sequences symmetry is weaker than for random sequence.

Any single trinucleotide from a given quadruplet is sufficient to generate the other three trinucleotides from the same quadruplet, i.e. it may represent its quadruplet.
The 43 = 64 trinucleotides can be classified into 20 quadruplets. However, their 20 representative trinucleotides can be mathematically chosen in numerous different ways.

1.2 Start/stop codon like trinucleotide matrix

The novel classification scheme defined in Table 1 is referred to as start/stop codon like trinucleotide matrix. In the first four quadruplets in Table 1 the representative trinucleotides are the start/stop codon like trinucleotides (Rosandić et al., 2013b). Trinucleotides in subgroups Ib,c and IIb,c, consisting of at most two different nucleotides, are also organized into quadruplets as shown in the matrix.

There are 20 possible quadruplets instead of 16 expected for 64 different trinucleotides (64:4=16). The reason is that 48 of 64 trinucleotides have nonsymmetrical forms (like ATG, TGA etc.) and 16 have symmetric forms (like ACA, GAG etc.). Each quadruplet with nonsymmetric trinucleotides contains four different trinucleotides. There are 48:4 = 12 nonsymmetrical quadruplets (6 A+T rich and 6 C+G rich) (first six rows in Table 1). For each of 16 symmetric trinucleotides, a direct trinucleotide is equal to its reverse and complement to its reverse complement. Therefore each of the corresponding quadruplets contains only two different trinucleotides and there are 16:2 = 8 different symmetric quadruplets (4 A+T rich and 4 C+G rich) (last four rows in Table 1). Classifying trinucleotides and their quadruplets into A+T rich and C+G rich, they are automatically characterized by symmetry.

1.3 Double stranded combined frequency

In order to elucidate the origin of second Chargaff’s parity rule, we consider simultaneously both strands of DNA because it appears in all double stranded DNAs but not in single stranded genomes. Starting from the entrance of a single nucleotide into one strand of DNA, for example the nucleotide A into top strand, at the corresponding
position in the bottom strand enters its Watson-Crick pair T. We argue that in this case
the same type of nucleotide which entered into top strand (A) should also enter at some
position somewhere in the bottom strand, followed by the entrance of its Watson-Crick
pair (T) into the top strand (Fig. 3, left panel). Thus the balance of A/T, T/A nucleotides
between the two strands of DNA duplex is established, independently of position of two
Watson-Crick partners. In order to describe schematically this pattern, we use the
concept of Q-box (Fig. 3, right panel). To each of four elements of Q(box) we assign the
frequency $f$ of its appearance in genomic sequence. This frequency has the same value
for nucleotides A in the top and bottom strand due to the symmetry law of DNA
creation and their paired nucleotides T due to Watson-Crick pairing.

The combined frequency of both strands has the same value $f_q$ for each of four
trinucleotides belonging to the same quadruplet in Q-dibox(D):

$$f_q(D) = f_1(D/ATG \ t.s.) + f_2(D/ATG \ b.s.) =$$

$$f_1(RC(D)/CAT \ t.s.) + f_2(RC(D)/CAT \ b.s.) =$$

$$f_1(C(D)/TAC \ b.s.) + f_2(C(D)/TAC \ t.s.) =$$

$$f_1(R(D)/GTA \ b.s.) + f_2(R(D)/GTA \ t.s.) =$$

$$f_1 + f_2.$$

Notations t.s. and b.s. refer to top and bottom strand, respectively. In each sum
the first term is arising from QD-RC-box and the second from QC-R-box.

Each of 20 Q-diboxes for the whole DNA is characterized by its combined
frequency (Fig. 5).
Table S1.

Frequencies of trinucleotides of genomes included in diagram from Fig. 6.

**E. coli**

Downloaded from http://www.ncbi.nlm.nih.gov/nuccore/545778205?report=fasta (28.1.2015.).

### A+T rich group (I)

| D   | f   | %   | RC  | f   | %   | C   | f   | %   | R   | f   | %   | Sub-group |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|
| ATG | 76282 | 1.64 | CAT | 77041 | 1.66 | TAC | 52611 | 1.13 | GTA | 52688 | 1.14 | Ia         |
| TGA | 83532 | 1.80 | TCA | 84101 | 1.81 | ACT | 49886 | 1.07 | AGT | 49792 | 1.07 |           |
| TAG | 27254 | 0.59 | CTA | 26770 | 0.58 | ATC | 86523 | 1.86 | GAT | 86587 | 1.87 |           |
| TAA | 68858 | 1.48 | TTA | 68845 | 1.48 | ATT | 83424 | 1.80 | AAT | 83021 | 1.79 | Ib         |
| AAC | 82616 | 1.78 | GTT | 82622 | 1.78 | TTTG| 76998 | 1.66 | CAA | 76654 | 1.65 |           |
| AAG | 63405 | 1.37 | CTT | 63676 | 1.37 | TTC | 83879 | 1.81 | GAA | 83530 | 1.80 | Ic         |

### C+G rich group (II)

| D   | f   | %   | RC  | f   | %   | C   | f   | %   | R   | f   | %   | Sub-group |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|
| CGT | 73184 | 1.58 | ACG | 73288 | 1.58 | GCA | 96071 | 2.07 | TGC | 95270 | 2.05 | Ila        |
| GTC | 54247 | 1.17 | GAC | 54764 | 1.18 | CAG | 104850 | 2.26 | CTG | 102957 | 2.22 |           |
| GCT | 80333 | 1.73 | AGC | 80909 | 1.74 | CGA | 70971 | 1.53 | TCG | 71759 | 1.55 |           |
| GCC | 93028 | 2.00 | GGC | 92189 | 1.99 | CGG | 86904 | 1.87 | CCG | 87076 | 1.88 | IIb        |
| CCA | 86491 | 1.86 | TGG | 85180 | 1.84 | GGT | 74326 | 1.60 | ACC | 74935 | 1.61 |           |
| CCT | 50447 | 1.09 | AGG | 50653 | 1.09 | GGA | 56222 | 1.21 | TCC | 56051 | 1.21 |           |
| CAC | 66782 | 1.44 | GTG | 66142 | 1.42 | GTG*| 65296 | 1.38 | CAC*| 64404 | 1.36 | IIc        |
| CGC | 115734 | 2.49 | GCG | 114670 | 2.47 | GCG*| 113576 | 2.38 | CGC*| 112689 | 2.36 |           |
| CTC | 42746 | 0.92 | GAG | 42503 | 0.92 | GAG*| 42214 | 0.91 | CTC*| 41809 | 0.91 |           |
| CCC | 47807 | 1.03 | GGG | 47515 | 1.02 | GGG*| 46821 | 0.99 | CCC*| 46129 | 0.98 |           |

### Base No. Percentage

| Base | No.  | Percentage |
|------|------|------------|
| A    | 1142742 | 24.62%     |
| C    | 1180091 | 25.42%     |
| G    | 1177437 | 25.37%     |
| T    | 1141382 | 24.59%     |
| N    | 0     | 0.00%      |
| **Sum:** | 4641652 | 100.00%    |
**Helicobacter pylori**

Downloaded from http://www.ncbi.nlm.nih.gov/nuccore/15611071?report=fasta (14.1.2015.).

### A+T rich group (I)

| Sub-group | D   | f   | %   | RC  | f   | %   | C   | f   | %   | R   | f   | %   |
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ia        | ATG | 27697 | 1.68 | CAT | 27645 | 1.68 | TAC | 11718 | 0.71 | GTA | 11731 | 0.71 |
|           | TGA | 29544 | 1.80 | TCA | 30181 | 1.84 | ACT | 15856 | 0.96 | AGT | 15369 | 0.93 |
|           | TAG | 24786 | 1.51 | CTA | 25073 | 1.53 | ATC | 28737 | 1.75 | GAT | 28566 | 1.74 |
| Ib        | TAA | 47294 | 2.88 | TTA | 46844 | 2.85 | ATT | 46048 | 2.80 | AAT | 46421 | 2.82 |
|           | AAC | 27948 | 1.70 | GTT | 27400 | 1.67 | TTG | 38353 | 2.33 | CAA | 38946 | 2.37 |
|           | AAG | 40949 | 2.49 | CTT | 42273 | 2.57 | TTC | 30523 | 1.86 | GAA | 29078 | 1.77 |

### C+G rich group (II)

| Sub-group | D   | f   | %   | RC  | f   | %   | C   | f   | %   | R   | f   | %   |
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ia        | CGT | 13903 | 0.85 | ACG | 14121 | 0.86 | GCA | 21564 | 1.31 | TGC | 20971 | 1.28 |
|           | GTC | 9052  | 0.55 | GAC | 8971  | 0.55 | CAG | 10416 | 0.63 | CTG | 10246 | 0.62 |
|           | GCT | 31651 | 1.93 | AGC | 31588 | 1.92 | CGA | 12281 | 0.75 | TCG | 12409 | 0.75 |
| Ib        | GCC | 19490 | 1.19 | GGC | 19678 | 1.20 | CGG | 8403  | 0.51 | CGC | 8525  | 0.52 |
|           | CCA | 23383 | 1.42 | TGG | 23355 | 1.42 | GGT | 16485 | 1.00 | ACC | 16781 | 1.02 |
|           | CCT | 19881 | 1.21 | AGG | 19707 | 1.20 | GGA | 15302 | 0.93 | TCC | 15518 | 0.94 |

### Base Counts

| Base | No.  | Percentage  |
|------|------|-------------|
| A    | 498514 | 30.33%      |
| C    | 323770 | 19.70%      |
| G    | 320426 | 19.49%      |
| T    | 501121 | 30.48%      |
| N    | 0     | 0.00%       |

**Sum:** 1643831 100.00%
Saccharomyces cerevisiae

Downloaded from http://www.ensembl.org/Saccharomyces_cerevisiae/Info/Index (28.1.2015.).

Chromosome 1

### A+T rich group (I)

| D   | f  | %  | RC | f  | %  | C   | f  | %  | R   | f  | %  | Sub-group |
|-----|----|----|----|----|----|-----|----|----|-----|----|----|-----------|
| ATG | 4297 | 1.87 | CAT | 4266 | 1.85 | TAC | 3281 | 1.43 | GTA | 3494 | 1.52 | Ia        |
| TGA | 4799 | 2.08 | TCA | 4610 | 2.00 | ACT | 3534 | 1.54 | AGT | 3701 | 1.61 |           |
| TAG | 2929 | 1.27 | CTA | 2754 | 1.20 | ATC | 3846 | 1.67 | GAT | 4007 | 1.74 |           |
| TAA | 4784 | 2.08 | TTA | 4694 | 2.04 | ATT | 6385 | 2.77 | AAT | 6307 | 2.74 | Ib        |
| AAC | 4105 | 1.78 | GTT | 4193 | 1.82 | TTG | 5452 | 2.37 | CAA | 5150 | 2.24 |           |
| AAG | 4961 | 2.15 | CTT | 4729 | 2.05 | TTC | 5161 | 2.24 | CAA | 5439 | 2.36 |           |
| ACA | 3925 | 1.70 | TGT | 4184 | 1.82 | TGT* | 1.07 | ACG | 2183 | 0.95 | IIa       |
| ATA | 5241 | 2.28 | TAT | 5189 | 2.74 | TAT* | 1.04 | GAC | 2384 | 1.04 |           |
| AGA | 4538 | 1.97 | CTT | 4729 | 2.05 | TTC | 5161 | 2.24 | GAA | 5439 | 2.36 |           |
| AAG | 8575 | 3.72 | TTT | 8848 | 3.84 | TTT* | 1.07 | GCC | 2700 | 1.17 | IIC       |

### C+G rich group (II)

| D   | f  | %  | RC | f  | %  | C   | f  | %  | R   | f  | %  | Sub-group |
|-----|----|----|----|----|----|-----|----|----|-----|----|----|-----------|
| CGT | 2104 | 0.91 | ACG | 2183 | 0.95 | GCA | 2997 | 1.30 | TGC | 2951 | 1.28 | IIa       |
| GTC | 2455 | 1.07 | GAC | 2384 | 1.04 | CAG | 3094 | 1.34 | CTG | 3075 | 1.34 |           |
| GCT | 2700 | 1.17 | AGC | 2680 | 1.16 | CGA | 2156 | 0.94 | TCG | 2199 | 0.96 |           |
| GCC | 1963 | 0.85 | GGC | 1908 | 0.83 | CGG | 1444 | 0.63 | CCG | 1443 | 0.63 | IIb       |
| CCA | 3699 | 1.61 | TGG | 3688 | 1.60 | GGT | 2951 | 1.28 | ACC | 2849 | 1.24 |           |
| CCT | 2454 | 1.07 | AGG | 2700 | 1.18 | GGA | 2981 | 1.29 | TCC | 2783 | 1.21 |           |
| CAC | 2721 | 1.18 | GTG | 2798 | 1.22 | GTG* | 1.07 | GCG | 1259 | 0.55 | IIC       |
| CGC | 1380 | 0.60 | GCG | 1259 | 0.55 | GCG* | 1.07 | GCG | 1259 | 0.55 |           |
| CTC | 2553 | 1.11 | GAG | 2644 | 1.15 | GAG* | 1.07 | CTC | 2644 | 1.15 |           |
| CCC | 1619 | 0.70 | GGG | 1591 | 0.69 | GGG* | 1.07 | CCC | 1591 | 0.69 |           |

### Base Distribution

| Base | No.  | Percentage |
|------|------|------------|
| A    | 69836| 30.33%     |
| C    | 44641| 19.39%     |
| G    | 45766| 19.88%     |
| T    | 69975| 30.40%     |
| N    | 0    | 0.00%      |
| Sum  | 230218 | 100.00%   |
**Oryza sativa**

Downloaded from [http://plants.ensembl.org/Oryza_sativa/Info/Index](http://plants.ensembl.org/Oryza_sativa/Info/Index), ftp://ftp.ensemblgenomes.org/pub/plants/release-25/fasta/oryza_sativa/dna/ (27.1.2015).

**Chromosome 1**

### A+T rich group (I)

| D   | f     | %    | RC  | f     | %    | C   | f     | %    | R   | f     | %    | Sub-group |
|-----|-------|------|-----|-------|------|-----|-------|------|-----|-------|------|-----------|
| ATG | 805849| 1.86 | CAT | 807372| 1.87 | TAC | 534276| 1.23 | GTA | 533403| 1.23 | Ia        |
| TGA | 751755| 1.74 | TCA | 752554| 1.74 | ACT | 612537| 1.42 | AGT | 611186| 1.41 |           |
| TAG | 573339| 1.33 | CTA | 571536| 1.32 | ATC | 692018| 1.60 | GAT | 692170| 1.60 |           |
| TAA | 823530| 1.90 | TTA | 821867| 1.90 | ATT | 1027767| 2.38 | AAT | 1030430| 2.38 | Ib        |
| AAC | 683003| 1.58 | GTT | 681300| 1.57 | TTG | 858389| 1.98 | CAA | 861917| 1.99 |           |

### C+G rich group (II)

| D   | f     | %    | RC  | f     | %    | C   | f     | %    | R   | f     | %    | Sub-group |
|-----|-------|------|-----|-------|------|-----|-------|------|-----|-------|------|-----------|
| CGT | 400424| 0.93 | ACG | 399896| 0.92 | GCA | 650036| 1.50 | TGC | 650435| 1.50 | Ila       |
| GTC | 472370| 1.09 | GAC | 472250| 1.09 | CAG | 562383| 1.30 | CTG | 562969| 1.30 |           |
| GCT | 584770| 1.35 | AGC | 586258| 1.36 | CGA | 453963| 1.05 | TCG | 454675| 1.05 |           |
| GCC | 591914| 1.37 | GGC | 591574| 1.37 | CGG | 490433| 1.13 | CCG | 488702| 1.13 | Ilb       |
| CCA | 677747| 1.57 | TGG | 674006| 1.56 | GGT | 524453| 1.21 | ACC | 525761| 1.22 |           |
| CCT | 591547| 1.37 | AGG | 591424| 1.37 | GAA | 639834| 1.48 | TCC | 640319| 1.48 |           |
| CAC | 604160| 1.40 | GTG | 602957| 1.39 | GTG*| 602957| 1.39 | CAC*| 602957| 1.39 | Ilc       |
| CGC | 484037| 1.12 | GCG | 485584| 1.12 | GCG*| 485584| 1.12 | CGC*| 485584| 1.12 |           |
| CTC | 664396| 1.54 | GAG | 665079| 1.54 | GAG*| 665079| 1.54 | CTC*| 665079| 1.54 |           |
| CCC | 501746| 1.16 | GGG | 500944| 1.16 | GGG*| 500944| 1.16 | CCC*| 500944| 1.16 |           |

### Base

| Base | No.   | Percentage |
|------|-------|------------|
| A    | 12172333| 28.13%     |
| C    | 9471278 | 21.89%     |
| G    | 9463081 | 21.87%     |
| T    | 12155831| 28.09%     |
| N    | 8400   | 0.02%      |
| Sum: | 43270923| 100.00%    |
Zea mays

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Zea_mays/Assembled_chromosomes/seq/ (28.1.2015.).

Chromosome 1

### A+T rich group (I)

| D   | f   | %  | RC f | %  | C   | f   | %  | R   | f   | %  | Sub-group |
|-----|-----|----|------|----|-----|-----|----|-----|-----|----|-----------|
| ATG | 5266410 | 1.76 | CAT  | 5268665 | 1.76 | TAC  | 3165301 | 1.06 | GTA  | 3154749 | 1.05 | Ia        |
| TGA | 5354920 | 1.79 | TCA  | 5368394 | 1.79 | ACT  | 4263244 | 1.42 | AGT  | 4255804 | 1.42 |           |
| TAG | 3837307 | 1.28 | CTA  | 3844580 | 1.28 | ATC  | 4593296 | 1.53 | GAT  | 4583220 | 1.53 |           |
| TAA | 4541535 | 1.52 | TTA  | 4544313 | 1.52 | ATT  | 5641218 | 1.88 | AAT  | 5643633 | 1.88 | lb        |
| AAC | 4527101 | 1.51 | GTT  | 4521409 | 1.51 | TGG  | 6251277 | 2.09 | CAA  | 6268490 | 2.09 |           |
| AAG | 7530795 | 2.51 | TTT  | 7547966 | 2.52 | TTT  | 7547966 | 2.52 | AAA  | 7510535 | 2.50 |           |

### C+G rich group (II)

| D   | f   | %  | RC f | %  | C   | f   | %  | R   | f   | %  | Sub-group |
|-----|-----|----|------|----|-----|-----|----|-----|-----|----|-----------|
| CGT | 2565920 | 0.86 | ACG  | 2570286 | 0.86 | GCA  | 4571677 | 1.53 | TGC  | 4559427 | 1.52 | Iia       |
| GTC | 3967936 | 1.32 | GAC  | 3970338 | 1.33 | CAG  | 3904399 | 1.30 | CTG  | 3895560 | 1.30 |           |
| GCT | 4553893 | 1.52 | AGC  | 4558010 | 1.52 | CGA  | 3702810 | 1.24 | TCG  | 3705369 | 1.24 |           |
| GCC | 4453872 | 1.49 | GGC  | 4447794 | 1.48 | CGG  | 3705588 | 1.24 | CCG  | 3712858 | 1.24 | Ib        |
| CCA | 4969577 | 1.66 | TGG  | 4949877 | 1.65 | GGT  | 4279650 | 1.43 | ACC  | 4297618 | 1.43 |           |
| CCT | 4973418 | 1.66 | AGG  | 4957132 | 1.65 | GGA  | 4885299 | 1.63 | TCC  | 4904515 | 1.64 |           |

### Base Number Percentage

| Base | No.   | Percentage |
|------|-------|------------|
| A    | 79447048 | 26.36%     |
| C    | 70418884 | 23.36%     |
| G    | 70260228 | 23.31%     |
| T    | 79449346 | 26.36%     |
| N    | 1857874  | 0.62%      |
| Sum: | 301433380 | 100.00%    |
**Arabidopsis thaliana**

Downloaded from ftp://ftp.ensemblgenomes.org/pub/plants/release-25/fasta/arabidopsis_thaliana/dna/ (29.1.2015.).

Chromosome 1

### A+T rich group (I)

|   |   | D   | f  | %   | RC | f  | %   | C   | f  | %   | R   | f  | %   | Sub-group |
|---|---|-----|----|-----|----|----|-----|-----|----|-----|-----|----|-----|-----------|
| ATG | 565975 | 1.87 | CAT | 567650 | 1.88 | TAC | 380315 | 1.26 | GTA | 380348 | 1.26 | Ia |
| TGA | 627384 | 2.07 | TCA | 624595 | 2.06 | ACT | 466633 | 1.54 | AGT | 467752 | 1.55 |
| TAG | 397027 | 1.31 | CTA | 403610 | 1.33 | ATC | 555283 | 1.83 | GAT | 554942 | 1.83 |
| TAA | 780160 | 2.58 | TTA | 772517 | 2.55 | ATT | 906155 | 2.99 | AAT | 905850 | 2.99 | lb |
| AAC | 604682 | 2.00 | GTT | 598859 | 1.98 | TGT | 718653 | 2.37 | CAA | 715447 | 2.36 |
| AAG | 565210 | 1.97 | CTT | 666484 | 2.20 | TTC | 680133 | 2.25 | GAA | 680133 | 2.25 |
| ACA | 595543 | 1.97 | TGT | 589534 | 1.95 | TGT* |  |   | ACA* |  |   |
| ATA | 779520 | 2.58 | TAT | 778490 | 2.57 | TAT* |  |   | ATA* |  |   |
| AGA | 700267 | 2.31 | TCT | 701635 | 2.32 | TCT* |  |   | AGA* |  |   |
| AAA | 1343856 | 4.44 | TTT | 1340726 | 4.43 | TTT* |  |   | AAA* |  |   |

### C+G rich group (II)

|   |   | D   | f  | %   | RC | f  | %   | C   | f  | %   | R   | f  | %   | Sub-group |
|---|---|-----|----|-----|----|----|-----|-----|----|-----|-----|----|-----|-----------|
| CGT | 203265 | 0.67 | ACG | 202297 | 0.67 | GCA | 309004 | 1.02 | TGC | 307996 | 1.02 | Ia |
| GTC | 277102 | 0.92 | GAC | 277353 | 0.92 | CAG | 314849 | 1.04 | CTG | 311985 | 1.03 |
| GCT | 322078 | 1.06 | AGC | 322883 | 1.07 | CGA | 241733 | 0.80 | TCG | 242103 | 0.80 |
| GCC | 163714 | 0.54 | GGC | 163961 | 0.54 | CGG | 144879 | 0.48 | CCG | 145433 | 0.48 | Ib |
| CCA | 394681 | 1.30 | TGG | 395159 | 1.31 | GTG | 319217 | 1.05 | ACC | 323744 | 1.07 |
| CCT | 307752 | 1.02 | AGG | 303927 | 1.00 | GGA | 360788 | 1.19 | TCC | 360406 | 1.19 |
| CAC | 325871 | 1.08 | GTC | 323462 | 1.07 | GTG* |  |   | CAC* |  |   |
| CGC | 107492 | 0.36 | GCC | 107536 | 0.36 | GCG* |  |   | CGC* |  |   |
| CTC | 416020 | 1.37 | GAG | 417745 | 1.38 | GAG* |  |   | CTC* |  |   |
| CCC | 168169 | 0.56 | GGG | 164885 | 0.54 | GGG* |  |   | CCC* |  |   |

**Base**  | **No.**  | **Percentage**
---|---|---
A  | 9709674 | 31.91%
G  | 5435374 | 17.86%
C  | 5421151 | 17.82%
T  | 9697113 | 31.87%
N  | 163958 | 0.54%
**Sum:**  | 30427270 | 100.00%
**Tribolium castaneum**

Downloaded from http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AAJJ01 (21.1.2014.).

Chromosome 1

### A+T rich group (I)

| Df  | %  | RC | f  | %  | C  | f  | %  | R  | f  | %  | Subgroup |
|-----|----|----|----|----|----|----|----|----|----|----|----------|
| ATG | 104238 | 1.49 | CAT | 103992 | 1.48 | TAC | 92365 | 1.32 | GTA | 91912 | 1.31 | la |
| TGA | 120336 | 1.72 | TCA | 119652 | 1.71 | ACT | 106720 | 1.52 | AGT | 106767 | 1.52 |
| TAG | 76733 | 1.09 | CTA | 76670 | 1.09 | ATC | 96247 | 1.37 | GAT | 96697 | 1.38 |
| TAA | 217965 | 3.11 | TTA | 219193 | 3.13 | ATT | 262343 | 3.74 | AAT | 261332 | 3.73 | lb |
| AAC | 138990 | 1.98 | GAT | 141616 | 2.02 | TTG | 162347 | 2.32 | CAA | 158825 | 2.27 |
| AAG | 117470 | 1.68 | CTT | 118438 | 1.69 | TCC | 140832 | 2.01 | GAA | 140965 | 2.01 |
| ACA | 135031 | 1.93 | TGT | 136915 | 1.95 | TGT* | 140508 | 2.00 | ACA* | 140965 | 2.01 |
| ATA | 165093 | 2.35 | TAT | 165897 | 2.37 | TAT* | 140965 | 2.01 | ATA* | 140965 | 2.01 |
| AGA | 96496 | 1.38 | TCT | 98106 | 1.40 | TCT* | 140965 | 2.01 | AGA* | 140965 | 2.01 |
| AAA | 365082 | 5.21 | TTA | 370273 | 5.28 | TTA* | 140965 | 2.01 | AAA* | 140965 | 2.01 |

### C+G rich group (II)

| Df  | %  | RC | f  | %  | C  | f  | %  | R  | f  | %  | Subgroup |
|-----|----|----|----|----|----|----|----|----|----|----|----------|
| CGT | 73875 | 1.05 | ACG | 73207 | 1.04 | GCA | 84694 | 1.21 | TGC | 85518 | 1.22 | la |
| GTC | 69343 | 0.99 | GAC | 68236 | 0.97 | CAG | 73141 | 1.04 | CTG | 73529 | 1.05 |
| GCT | 71690 | 1.02 | AGC | 70761 | 1.01 | CGA | 83413 | 1.19 | TCG | 84161 | 1.20 |
| GCC | 54674 | 0.78 | GGC | 54753 | 0.78 | CGG | 55084 | 0.79 | CGG | 55033 | 0.78 | lb |
| CCA | 79103 | 1.13 | TGG | 81012 | 1.16 | GGT | 68994 | 0.98 | ACC | 67166 | 0.96 |
| CCT | 59697 | 0.85 | AGG | 60288 | 0.86 | GGA | 72628 | 1.04 | TCC | 71991 | 1.03 |
| CAC | 82532 | 1.18 | GTG | 83677 | 1.19 | GTG* | 86994 | 1.26 | CAC* | 86994 | 1.26 |
| CGC | 57267 | 0.82 | GCG | 57235 | 0.82 | GCG* | 86994 | 1.26 | CGC* | 86994 | 1.26 |
| CTC | 67478 | 0.96 | GAG | 66966 | 0.96 | GAG* | 86994 | 1.26 | CTC* | 86994 | 1.26 |
| CCC | 48138 | 0.69 | GGG | 50138 | 0.72 | GGG* | 86994 | 1.26 | CCC* | 86994 | 1.26 |

### Base No. Percentage

| Base | No. | Percentage |
|------|-----|------------|
| A    | 2227444 | 20.48%    |
| C    | 1266370 | 11.64%    |
| G    | 1274313 | 11.71%    |
| T    | 2243557 | 20.63%    |
| N    | 3865951 | 35.54%    |
| Sum: | 10877635 | 100.00%   |
Danio rerio

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/D_rerio/Assembled_chromosomes/seq/ (28.1.2015.).

Chromosome 1

A+T rich group (I)

| Base | No.      | Percentage |
|------|----------|------------|
| A    | 18705573 | 31.77%     |
| C    | 10702228 | 18.18%     |
| T    | 18698363 | 31.76%     |
| N    | 64150    | 0.11%      |
| Sum: | 58871917 | 100.00%    |

C+G rich group (II)

| Base | No.      | Percentage |
|------|----------|------------|
| A    | 18705573 | 31.77%     |
| C    | 10702228 | 18.18%     |
| T    | 18698363 | 31.76%     |
| N    | 64150    | 0.11%      |
| Sum: | 58871917 | 100.00%    |
**Gallus gallus**

Downloaded from
ftp://ftp.ncbi.nih.gov/genomes/Gallus_gallus/Assembled_chromosomes/seq/
(28.1.2015.).

Chromosome 1

### A+T rich group (I)

| D | f   | %   | RC  | f   | %   | C  | f   | %   | R  | f   | %   | Sub-group |
|---|-----|-----|-----|-----|-----|----|-----|-----|----|-----|-----|-----------|
| ATG | 3672156 | 1.90 | CAT | 3651251 | 1.89 | TAC | 2379288 | 1.23 | GTA | 2402979 | 1.24 | Ia        |
| TGA | 3995202 | 2.07 | TCA | 3962713 | 2.05 | ACT | 3321136 | 1.72 | AGT | 3346967 | 1.73 |           |
| TAG | 2405263 | 1.25 | CTA | 2382364 | 1.23 | ATC | 2655825 | 1.38 | GAT | 2673303 | 1.38 |           |
| TAA | 3939932 | 2.04 | TTA | 3952768 | 2.05 | ATT | 4619998 | 2.39 | AAT | 4609800 | 2.39 | Ib        |
| AAC | 3038786 | 1.57 | GTT | 3094439 | 1.60 | TTG | 3775778 | 1.96 | CAA | 3701405 | 1.92 |           |
| AAG | 4196818 | 2.17 | CTT | 4233911 | 2.19 | TTC | 4219780 | 2.19 | GAA | 4204058 | 2.18 |           |
| ACA | 4186458 | 2.17 | TGT | 4283764 | 2.22 | TGT* |       |      | ACA* |       |      | Ic        |
| ATA | 3669695 | 1.90 | TAT | 3683331 | 1.91 | TAT* |       |      | ATA* |       |      |           |
| AGA | 4414058 | 2.29 | TCT | 4444813 | 2.30 | TCT* |       |      | AGA* |       |      |           |
| AAA | 7259471 | 3.76 | TTT | 7353172 | 3.81 | TTT* |       |      | AAA* |       |      |           |

### C+G rich group (II)

| D | f   | %   | RC  | f   | %   | C  | f   | %   | R  | f   | %   | Sub-group |
|---|-----|-----|-----|-----|-----|----|-----|-----|----|-----|-----|-----------|
| CGT | 541504 | 0.28 | ACG | 537597 | 0.28 | GCA | 3295938 | 1.71 | TGC | 3313702 | 1.72 | Iia       |
| GTC | 1806040 | 0.94 | GAC | 1795630 | 0.93 | CAG | 4368678 | 2.26 | CTG | 4397326 | 2.28 |           |
| GCT | 3074240 | 1.59 | AGC | 3054599 | 1.58 | CGA | 356092  | 0.18 | TCG | 356180  | 0.18 |           |
| GCC | 1753171 | 0.91 | GGC | 1758976 | 0.91 | CGG | 390075  | 0.20 | CCG | 389954  | 0.20 | Ib        |
| CCA | 3060132 | 1.58 | TGG | 3101519 | 1.61 | GTT | 1980119 | 1.03 | ACC | 1952397 | 1.01 |           |
| CCT | 3096325 | 1.60 | AGG | 3100379 | 1.61 | GGA | 2852854 | 1.48 | TCC | 2840866 | 1.47 |           |
| CAC | 2783895 | 1.44 | GTG | 2848906 | 1.48 | GTG* |       |      | CAC* |       |      | Iic       |
| CGC | 369321 | 0.19 | GCG | 373267  | 0.19 | GCG* |       |      | CGC* |       |      |           |
| CTC | 2922914 | 1.51 | GAG | 2945244 | 1.53 | GAG* |       |      | CTC* |       |      |           |
| CCC | 1953331 | 1.01 | GGG | 1970630 | 1.02 | GGG* |       |      | CCC* |       |      |           |

| Base | No.   | Percentage |
|------|-------|------------|
| A    | 57637041 | 29.52%     |
| C    | 38599597 | 19.77%     |
| G    | 38830919 | 19.89%     |
| T    | 58008993 | 29.71%     |
| N    | 2200200  | 1.13%      |
| Sum: | 195276750 | 100.00%    |
**Equus caballus**

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Equus_caballus/Assembled_chromosomes/seq/ (27.1.2015.).

**Chromosome 1**

### A+T rich group (I)

| Base | No. | Percentage |
|------|-----|------------|
| A    | 53722912 | 28.91% |
| C    | 38141705 | 20.52% |
| G    | 38148280 | 20.53% |
| T    | 53548958 | 28.82% |
| N    | 2276254  | 1.22% |
| Sum: | 185838109 | 100.00% |

### C+G rich group (II)

| Base | No. | Percentage |
|------|-----|------------|
| A    | 53722912 | 28.91% |
| G    | 4658908  | 25.52% |
| T    | 23841705 | 13.45% |
| N    | 2276254  | 1.22% |
| C    | 5064812  | 27.72% |
| Sum: | 185838109 | 100.00% |
Bos taurus

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Bos_taurus/Assembled_chromosomes/seq/ (28.1.2015.).

Chromosome 1

A+T rich group (I)

|     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|
| D   | f   | %   | RC  | f   | %   | C   | f   | %   | R   | f   | %   |
| ATG | 3098824 | 1.97 | CAT | 3094866 | 1.97 | TAC | 1736682 | 1.10 | GTA | 1724095 | 1.10 |
| TGA | 3296960 | 2.10 | TCA | 3301493 | 2.10 | ACT | 2839801 | 1.81 | AGT | 2830043 | 1.80 |
| TAG | 1951590 | 1.24 | CTA | 1967086 | 1.25 | ATC | 2265678 | 1.44 | GAT | 2263024 | 1.44 |
| TAA | 3203673 | 2.04 | TTA | 3189004 | 2.03 | ATT | 3908072 | 2.48 | AAT | 3926419 | 2.50 |
| AAC | 2368770 | 1.51 | GTT | 2343089 | 1.49 | CAA | 3045531 | 1.94 | AAA | 3045531 | 1.94 |
| AAG | 3415010 | 2.17 | CTT | 3392175 | 2.16 | TTC | 3452958 | 2.20 | GAA | 3461030 | 2.20 |

C+G rich group (II)

|     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|
| D   | f   | %   | RC  | f   | %   | C   | f   | %   | R   | f   | %   |
| CGT | 361392 | 0.23 | ACG | 363162 | 0.23 | GCA | 2197851 | 1.40 | TGC | 2190470 | 1.39 |
| GTC | 1710515 | 1.09 | GAC | 1714659 | 1.09 | CAG | 3212413 | 2.04 | CTC | 3202447 | 2.04 |
| GCT | 2021773 | 1.29 | AGC | 2028653 | 1.29 | CGA | 339045 | 0.22 | TCG | 337160 | 0.21 |
| GCC | 1464379 | 0.93 | GGC | 1465058 | 0.93 | CGG | 318667 | 0.20 | CCG | 318613 | 0.20 |
| CCA | 2935521 | 1.87 | TGC | 2922156 | 1.86 | GGT | 1693250 | 1.08 | ACC | 1699266 | 1.08 |
| CCT | 2547351 | 1.62 | AGG | 2554332 | 1.62 | GGA | 2636787 | 1.68 | TCC | 2637775 | 1.68 |
| CAC | 2165606 | 1.38 | GTG | 2151124 | 1.37 | GTG* | 2151124 | 1.37 | CAC* | 2151124 | 1.37 |
| CGC | 264581 | 0.17 | GCC | 264757 | 0.17 | GCG* | 264757 | 0.17 | CGC* | 264757 | 0.17 |
| CTC | 2498764 | 1.59 | GAG | 2499634 | 1.59 | GAG* | 2499634 | 1.59 | CTC* | 2499634 | 1.59 |
| CCC | 1877665 | 1.19 | GGG | 1873062 | 1.19 | GGG* | 1873062 | 1.19 | CCC* | 1873062 | 1.19 |

Base   | No.   | Percentage
-------|-------|--------------
A       | 47238186 | 29.83%
C       | 31543258 | 19.92%
G       | 31485913 | 19.89%
T       | 47026892 | 29.70%
N       | 1042818 | 0.66%
Sum:   | 158337067 | 100.00%
### Canis Lupus

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Canis_lupus_familiaris/Assembled_chromosomes/seq/ (27.1.2015.,28.1.2015.).

#### Chromosome 1

**A+T rich group (I)**

|    | f   | %   | RC | f   | %   | C  | f   | %   | R  | f   | %   | Sub-group |
|----|-----|-----|----|-----|-----|----|-----|-----|----|-----|-----|-----------|
| ATG| 2218053 | 1.82 | CAT | 2217341 | 1.82 | TAC | 1272802 | 1.04 | GTA | 1268603 | 1.04 | Ia        |
| TGA| 2341410 | 1.92 | TCA | 2349594 | 1.92 | ACT | 1849390 | 1.52 | AGT | 1842507 | 1.51 |           |
| TAG| 1466400 | 1.20 | CTA | 1469685 | 1.20 | ATC | 1684971 | 1.38 | CAT | 1681671 | 1.38 |           |
| TAA| 2534532 | 2.08 | TTA | 2520078 | 2.06 | ATT | 2949780 | 2.42 | AAT | 2969174 | 2.43 | Ib        |
| AAC| 1699875 | 1.39 | GTT | 1678514 | 1.38 | TAC | 1272802 | 1.04 | GTA | 1268603 | 1.04 |           |
| AAG| 2495445 | 2.04 | CTT | 2478017 | 2.03 | TTC | 2446278 | 2.00 | GAA | 2466259 | 2.02 |           |
| ACA| 2392076 | 1.96 | TGT | 2365191 | 1.94 | TGT | 2949780 | 2.42 | AAT | 2969174 | 2.43 | Ic        |
| ATA| 2465435 | 2.02 | TAT | 2450060 | 2.01 | TAT | 2139977 | 1.75 | CAA | 2163724 | 1.77 |           |
| AGA| 2914527 | 2.39 | TCT | 2894594 | 2.37 | TCT | 2446278 | 2.00 | GAA | 2466259 | 2.02 |           |
| AAA| 4676494 | 3.83 | TTT | 4610839 | 3.78 | TTT | 2949780 | 2.42 | AAT | 2969174 | 2.43 |           |

**C+G rich group (II)**

|    | f   | %   | RC | f   | %   | C  | f   | %   | R  | f   | %   | Sub-group |
|----|-----|-----|----|-----|-----|----|-----|-----|----|-----|-----|-----------|
| CGT| 344686 | 0.28 | ACG | 347357 | 0.28 | GCA | 1709722 | 1.40 | TGC | 1703365 | 1.40 | Ila       |
| GTC| 1222697 | 1.00 | GAC | 1228272 | 1.01 | CAG | 2454382 | 2.01 | CTG | 2443556 | 2.00 |           |
| GCT| 1639240 | 1.34 | AGC | 1643365 | 1.35 | CGA | 308138 | 0.25 | TCG | 305614 | 0.25 |           |
| GCC| 1406891 | 1.15 | GGC | 1408336 | 1.15 | CGG | 430394 | 0.35 | CCG | 431023 | 0.35 | IIb       |
| CCA| 2169034 | 1.78 | TGG | 2164567 | 1.77 | GGT | 1390370 | 1.14 | ACC | 1397119 | 1.14 |           |
| CCT| 2210674 | 1.81 | AGG | 2217945 | 1.82 | GGA | 2014210 | 1.65 | TCC | 2006748 | 1.64 |           |
| CAC| 1784982 | 1.46 | GTG | 1772962 | 1.45 | GTG | 1681671 | 1.38 | GGT | 1678514 | 1.38 | IIc       |
| CGC| 336705 | 0.28 | GCG | 335930 | 0.28 | GCG* | 305614 | 0.25 | CCG* | 30614 | 0.25 |           |
| CTC| 2202645 | 1.80 | GAG | 2202111 | 1.80 | GAG* | 305614 | 0.25 | CCG* | 30614 | 0.25 |           |
| CCC| 1928187 | 1.58 | GGG | 1933856 | 1.58 | GGG* | 305614 | 0.25 | CCG* | 30614 | 0.25 |           |

#### Base Distribution

| Base | No. | Percentage |
|------|-----|------------|
| A    | 35764008 | 29.15% |
| C    | 25373843 | 20.68% |
| G    | 25360206 | 20.67% |
| T    | 35572566 | 29.00% |
| N    | 608162  | 0.50% |
| Sum: | 122678785 | 100.00% |
**Felis catus**

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Felis_catus/Assembled_chromosomes/seq/ (29.1.2015.).

Chromosome A1

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| D | f | % | RC | f | % | C | f | % | R | f | % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ATG | 4297488 | 1.84 | CAT | 4302407 | 1.85 | TAC | 2621284 | 1.12 | GTA | 2627603 | 1.13 |
| TGA | 4670416 | 2.00 | TCA | 4657489 | 2.00 | ACT | 3699789 | 1.59 | AGT | 3708306 | 1.59 |
| TAG | 2873807 | 1.23 | CTA | 2860395 | 1.23 | ATC | 3116765 | 1.34 | GAT | 3123617 | 1.34 |
| TAA | 5186210 | 2.22 | TTA | 5209055 | 2.23 | ATT | 6002473 | 2.57 | AAT | 5974856 | 2.56 |
| AAC | 3533094 | 1.52 | GTT | 3569706 | 1.53 | TTG | 4333536 | 1.86 | CAA | 4277284 | 1.83 |
| AAG | 4776333 | 2.05 | CTT | 4824387 | 2.07 | TTC | 4853505 | 2.08 | GAA | 4820144 | 2.07 |
| ACA | 4790443 | 2.05 | TGT | 4841695 | 2.08 | TGT* | ACA* |  |  |  |  |
| ATA | 5047953 | 2.17 | TAT | 5063719 | 2.17 | TAT* | ATA* |  |  |  |  |
| AGA | 5607927 | 2.41 | TCT | 5650018 | 2.42 | TCT* | AGA* |  |  |  |  |
| AAA | 9194893 | 3.94 | TTC | 9294553 | 3.99 | TTC* | AAA* |  |  |  |  |
|   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |

**Base No.**

|   |   |   |
|---|---|---|
| A | 69877173 | 29.20% |
| C | 46531955 | 19.44% |
| G | 46529589 | 19.44% |
| T | 70220808 | 29.34% |
| N | 6143378 | 2.57% |
| Sum: | 239302903 | 100.00% |

C+G rich group (II)

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| D | f | % | RC | f | % | C | f | % | R | f | % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CGT | 747879 | 0.32 | ACG | 743882 | 0.32 | GCA | 3003000 | 1.29 | TGC | 3016056 | 1.29 |
| GTC | 2332485 | 1.00 | GAC | 2323304 | 1.00 | CAG | 4398237 | 1.89 | CTG | 4414451 | 1.89 |
| GCT | 3065511 | 1.31 | AGC | 3055762 | 1.31 | CGA | 631856 | 0.27 | TCG | 633754 | 0.27 |
| GCC | 2296204 | 0.98 | GGC | 2292785 | 0.98 | CGG | 687455 | 0.29 | CGG | 689731 | 0.30 |
| CCA | 3826039 | 1.64 | TGG | 3834184 | 1.64 | GGT | 2548922 | 1.09 | ACC | 2542659 | 1.09 |
| CCT | 3792381 | 1.63 | AGG | 3774189 | 1.62 | GGA | 3454109 | 1.48 | TCC | 3469315 | 1.49 |
| CAC | 3299270 | 1.42 | GTG | 3317189 | 1.42 | GTG* | CAC* |  |  |  |  |
| CGC | 541102 | 0.23 | GCC | 540963 | 0.23 | GCC* | CGC* |  |  |  |  |
| CTC | 4108372 | 1.76 | GAG | 4097734 | 1.76 | GAG* | CTC* |  |  |  |  |
| CCC | 3123774 | 1.34 | GGG | 3108461 | 1.33 | GGG* | CCC* |  |  |  |  |
|   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |
Mus Musculus

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Mus_musculus/Assembled_chromosomes/seq/ (27.1.2015.).

Chromosome 1

A+T rich group (I)

|     | f   | %   |     | f   | %   |     | f   | %   |     | f   | %   | Sub-group |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----------|
| ATG | 3645710 | 1.90 | CAT | 3649283 | 1.90 | TAC | 2242076 | 1.17 | GTA | 2236237 | 1.17 | Ia        |
| TGA | 3694871 | 1.93 | TCA | 3701876 | 1.93 | ACT | 3146549 | 1.64 | AGT | 3137538 | 1.63 |           |
| TAG | 2641989 | 1.38 | CTA | 2648682 | 1.38 | ATC | 2623296 | 1.37 | GAT | 2615114 | 1.36 |           |
| TAA | 3715657 | 1.94 | TTA | 3711121 | 1.93 | ATT | 4249500 | 2.21 | AAT | 4257794 | 2.22 | Ib        |
| AAC | 2931824 | 1.53 | GTT | 2919516 | 1.52 | TTG | 3614818 | 1.88 | CAA | 3635711 | 1.89 |           |
| AAG | 4124669 | 2.15 | CTT | 4108026 | 2.14 | TTC | 3951101 | 2.06 | GAA | 3962916 | 2.07 |           |
| ACA | 4435786 | 2.31 | TGT | 4414730 | 2.30 | TGT*| 412731  | 0.22 | ACG | 414584  | 0.22 | IIa       |
| ATA | 3931770 | 2.05 | TAT | 3928089 | 2.05 | TAT*| 1984631 | 1.03 | GAC | 1985936 | 1.03 | IIb       |
| AGA | 4752169 | 2.48 | TCT | 4742704 | 2.47 | TCT*| 2719184 | 1.42 | AGC | 2720661 | 1.42 |           |
| AAA | 6513500 | 3.39 | TTT | 6480119 | 3.38 | TTT*| 2719184 | 1.42 | GCC | 2720661 | 1.42 |           |

C+G rich group (II)

|     | f   | %   |     | f   | %   |     | f   | %   |     | f   | %   | Sub-group |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----------|
| CGT | 412731 | 0.22 | ACG | 414584 | 0.22 | GCA | 2651333 | 1.38 | TGC | 2649216 | 1.38 | Ila       |
| GTC | 1984631 | 1.03 | GAC | 1985936 | 1.03 | CAG | 3908019 | 2.04 | CTG | 3909605 | 2.04 |           |
| GCT | 2719184 | 1.42 | AGC | 2720661 | 1.42 | CGA | 366838  | 0.19 | TCG | 364620  | 0.19 |           |
| GCC | 1936358 | 1.01 | GGC | 1938068 | 1.01 | CGG | 387914  | 0.20 | CCG | 387208  | 0.20 | IIb       |
| CCA | 3469405 | 1.81 | TGG | 3461775 | 1.80 | GGT | 2225845 | 1.16 | ACC | 2228305 | 1.16 |           |
| CCT | 3409546 | 1.78 | AGG | 3416509 | 1.78 | GGA | 3102282 | 1.62 | GCC | 3101500 | 1.62 |           |
| CAC | 3065392 | 1.60 | GTG | 3050463 | 1.59 | GTG*| 3050463 | 1.59 | CAC | 3050463 | 1.59 | IIc       |
| CGC | 310078  | 0.16 | GGC | 311148  | 0.16 | GGC*| 311148  | 0.16 | CGC | 311148  | 0.16 |           |
| CTC | 3351673 | 1.75 | GAG | 3352196 | 1.75 | GAG*| 3352196 | 1.75 | CTC | 3352196 | 1.75 |           |
| CCC | 2475178 | 1.29 | GGG | 2476152 | 1.29 | GGG*| 2476152 | 1.29 | CCC | 2476152 | 1.29 |           |

| Base | No.     | Percentage |
|------|---------|------------|
| A    | 56530182 | 28.92%     |
| C    | 39495313 | 20.21%     |
| G    | 39467408 | 20.19%     |
| T    | 56416289 | 28.86%     |
| N    | 3562779  | 1.82%      |
| Sum: | 195471971| 100.00%    |
Sus Scrofa

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Sus_scrofa/Assembled_chromosomes/seq/ (27.1.2015.).

Chromosome 1

A+T rich group (I)

| D   | f   | %   | RC | f   | %   | C   | f   | %   | R   | f   | %   | Sub-group |
|-----|-----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----------|
| ATG | 5259577 | 1.88 | CAT | 5260315 | 1.88 | TAC | 3102429 | 1.11 | GTA | 3115096 | 1.11 | Ia        |
| TGA | 5344828 | 1.91 | TCA | 5336364 | 1.91 | ACT | 4410387 | 1.58 | AGT | 4426794 | 1.58 |           |
| TAG | 3748809 | 1.34 | CTA | 3736508 | 1.34 | ATC | 3911100 | 1.40 | GAT | 3916636 | 1.40 |           |
| TAA | 5951758 | 2.13 | TTA | 5973566 | 2.13 | ATT | 7003291 | 2.50 | AAT | 6974788 | 2.49 | lb        |
| AAC | 4294246 | 1.53 | GTC | 4328906 | 1.55 | TGT | 5267067 | 1.88 | CAA | 5223159 | 1.87 |           |
| AAG | 5652297 | 2.02 | TGT | 5701906 | 2.04 | TGT | ACA |     |     |     |     | Ic        |
| ACA | 5652297 | 2.02 | TGT | 5701906 | 2.04 | TGT | ACA |     |     |     |     | Ic        |
| ATA | 5982705 | 2.14 | TAT | 6004880 | 2.15 | TAT | ATA |     |     |     |     |          |
| AGA | 6123571 | 2.19 | TCT | 6153378 | 2.20 | TCT | AGA |     |     |     |     |          |
| AAA | 1.1E+07 | 4.00 | TTT | 1.1E+07 | 4.03 | TTT | AAA |     |     |     |     |          |

C+G rich group (II)

| D   | f   | %   | RC | f   | %   | C   | f   | %   | R   | f   | %   | Sub-group |
|-----|-----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----------|
| CGT | 707762 | 0.25 | ACG | 706870 | 0.25 | GCA | 3912061 | 1.40 | TGC | 3917238 | 1.40 | Iia       |
| GTC | 2617484 | 0.94 | GAC | 2608773 | 0.93 | CAG | 5541622 | 1.98 | CTG | 5549981 | 1.98 |           |
| GCT | 3884893 | 1.39 | AGC | 3874696 | 1.38 | CGA | 635563 | 0.23 | TCG | 637012 | 0.23 |           |
| GCC | 3005556 | 1.07 | GGC | 3010735 | 1.08 | CGG | 772438 | 0.28 | CCG | 771640 | 0.28 | llb       |
| CCA | 5101436 | 1.82 | TGG | 5114857 | 1.83 | GGT | 3227460 | 1.15 | ACC | 3213244 | 1.15 |           |
| CCT | 4794177 | 1.71 | AGG | 4789740 | 1.71 | GGA | 4438784 | 1.59 | TCC | 4448525 | 1.59 |           |
| CAC | 3977317 | 1.42 | GTG | 4002359 | 1.43 | GTG | CAC |     |     |     |     | Iic       |
| CGC | 999901 | 0.21 | GCG | 600107 | 0.21 | GCG | CGC |     |     |     |     |          |
| CTC | 4271686 | 1.53 | GAG | 4272163 | 1.53 | GAG | CTC |     |     |     |     |          |
| CCC | 3654448 | 1.31 | GGG | 3662061 | 1.31 | GGG | CCC |     |     |     |     |          |

| Base | No. | Percentage |
|------|-----|------------|
| A    | 83470214 | 26.47%     |
| C    | 56290309 | 17.85%     |
| G    | 56356051 | 17.87%     |
| T    | 83780365 | 26.57%     |
| N    | 35424383 | 11.24%     |
| Sum: | 315321322 | 100.00%    |
**Macaca mulatta**

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Macaca_mulatta/Assembled_chromosomes/seq/ (28.1.2015.).

**Chromosome 1**

### A+T rich group (I)

| D    | f    | %   | RC  | f    | %   | C   | f    | %   | R   | f    | %   | Sub-group |
|------|------|-----|-----|------|-----|-----|------|-----|-----|------|-----|------------|
| ATG  | 3785230 | 1.77 | CAT | 3793148 | 1.78 | TAC | 2405319 | 1.13 | GTA | 2400047 | 1.12 | Ia         |
| TGA  | 4153329 | 1.95 | TCA | 4166283 | 1.95 | ACT | 3480827 | 1.63 | AGT | 3473784 | 1.63 |            |
| TAG  | 2733159 | 1.28 | CTA | 2738232 | 1.28 | ATC | 2824135 | 1.32 | CAT | 2814077 | 1.32 |            |
| TAA  | 4252408 | 1.99 | TTA | 4244270 | 1.99 | ATT | 5029348 | 2.36 | AAT | 5039825 | 2.36 | lb         |
| AAC  | 3083176 | 1.44 | GTT | 3064074 | 1.44 | TTG | 3947522 | 1.85 | CAA | 3973559 | 1.86 |            |
| AAG  | 4257659 | 2.00 | CTT | 4246248 | 1.99 | TTC | 4147817 | 1.94 | GAA | 4154601 | 1.95 |            |
| ACA  | 4249722 | 1.99 | TGT | 4214141 | 1.97 | TGT*| 4145825 | 1.87 | ACA*| 4154601 | 1.95 | Ic         |
| ATA  | 4101775 | 1.92 | TAT | 4093434 | 1.92 | TAT*| 4032124 | 1.90 | ATA*| 4032124 | 1.90 |            |
| AGA  | 4752049 | 2.23 | TCT | 4741699 | 2.22 | TCT*| 4684079 | 2.08 | AGA*| 4684079 | 2.08 |            |
| AAA  | 7969347 | 3.73 | TTT | 7928581 | 3.72 | TTT*| 7872120 | 3.57 | AAA*| 7872120 | 3.57 |            |

### C+G rich group (II)

| D    | f    | %   | RC  | f    | %   | C   | f    | %   | R   | f    | %   | Sub-group |
|------|------|-----|-----|------|-----|-----|------|-----|-----|------|-----|------------|
| CGT  | 564744 | 0.26 | ACG | 567612 | 0.27 | GCA | 3098167 | 1.45 | TGC | 3093619 | 1.45 | Ila        |
| GTC  | 2083532 | 0.98 | GAC | 2085876 | 0.98 | CAG | 4495719 | 2.11 | CTG | 4486634 | 2.10 |            |
| GCT  | 3062190 | 1.43 | AGC | 3066603 | 1.44 | CGA | 5056730 | 2.47 | TCG | 504929 | 0.24 |            |
| GCC  | 2677211 | 1.25 | GGC | 2674336 | 1.25 | CGG | 663374 | 0.31 | CCG | 664289 | 0.31 | Iib        |
| CCA  | 4017178 | 1.88 | GTG | 3999807 | 1.87 | GGT | 2536560 | 1.19 | ACC | 2547842 | 1.19 |            |
| CCT  | 3951408 | 1.85 | AGG | 3945785 | 1.85 | GGA | 3397988 | 1.59 | TCC | 3407730 | 1.60 |            |
| CAC  | 3268935 | 1.53 | GTG | 3241531 | 1.52 | GTG*| 3192517 | 1.49 | CAC*| 3192517 | 1.49 | Ilc        |
| CGC  | 556389 | 0.26 | GCG | 553352 | 0.26 | GCC | 558122 | 0.27 | CGC*| 558122 | 0.27 |            |
| CTC  | 3765037 | 1.76 | GAG | 3751687 | 1.76 | GAG*| 3703485 | 1.76 | CTC*| 3703485 | 1.76 |            |
| CCC  | 2964086 | 1.39 | GGG | 2955556 | 1.38 | GGG*| 2908349 | 1.34 | CCC*| 2908349 | 1.34 |            |

### Base No. Percentage

| Base | No.    | Percentage |
|------|--------|------------|
| A    | 62182780 | 27.24%   |
| C    | 44661041 | 19.57%   |
| G    | 44560320 | 19.52%   |
| T    | 62042279 | 27.18%   |
| N    | 14805795 | 6.49%     |
| Sum: | 228252215 | 100.00%  |
Pan troglodytes

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Pan_troglodytes/Assembled_chromosomes/seq/ (27.1.2015.).

Chromosome 1

### A+T rich group (I)

| D  | f          | %   | RC | f          | %   | C  | f          | %   | R  | f          | %   | Sub-group |
|----|------------|-----|----|------------|-----|----|------------|-----|----|------------|-----|-----------|
| ATG| 3893943    | 1.80| CAT| 3892901    | 1.80| TAC| 2422287   | 1.12| GTA| 2423017   | 1.12| Ia        |
| TGA| 4218758    | 1.95| TCA| 4214102    | 1.95| ACT| 3492454   | 1.61| AGT| 3495375   | 1.61|           |
| TAG| 2790814    | 1.29| CTA| 2790300    | 1.29| ATC| 2863229   | 1.32| GAT| 2866671   | 1.32|           |
| TAA| 4365019    | 2.02| TTA| 4373027    | 2.02| ATT| 5207336   | 2.41| AAT| 5196812   | 2.40| Ib        |
| AAC| 3093401    | 1.43| GTT| 3108062    | 1.44| TTG| 4056271   | 1.87| CAA| 4038400   | 1.87|           |
| AAG| 4314477    | 1.99| CTT| 4327717    | 2.00| TTC| 4213971   | 1.95| GAA| 4201444   | 1.94|           |
| ACA| 4264450    | 1.97| TGT| 4285881    | 1.98| TGT*|          |     | ACA*|          |     |           |
| ATA| 4268730    | 1.97| TAT| 4276931    | 1.98| TAT*|          |     | ATA*|          |     |           |
| AGA| 4789840    | 2.21| TCT| 4808530    | 2.22| TCT*|          |     | AGA*|          |     |           |
| AAA| 8047279    | 3.72| TTT| 8090699    | 3.74| TTT*|          |     | AAA*|          |     |           |

### C+G rich group (II)

| D  | f          | %   | RC | f          | %   | C  | f          | %   | R  | f          | %   | Sub-group |
|----|------------|-----|----|------------|-----|----|------------|-----|----|------------|-----|-----------|
| CGT| 533808     | 0.25| ACG| 532137     | 0.25| GCA| 3151821   | 1.46| TGC| 3158491   | 1.46| Ila       |
| GTC| 2076163    | 0.96| GAC| 2071108    | 0.96| CAG| 4517731   | 2.09| CTG| 4528331   | 2.09|           |
| GCT| 3120499    | 1.44| AGC| 3115233    | 1.44| CGA| 482126    | 0.22| TCG| 484356    | 0.22|           |
| GCC| 2712004    | 1.25| GGC| 2709048    | 1.25| CGG| 610938    | 0.28| CCG| 611884    | 0.28| IIb       |
| CCA| 3015386    | 1.59| TGG| 3003473    | 1.39| GGG| 3424441   | 1.58| TCC| 3426302   | 1.58|           |
| CAC| 3279901    | 1.52| GTG| 3279901    | 1.52| GTG*|          |     | CAC*|          |     |           |
| CGC| 533808     | 0.25| ACG| 532137     | 0.25| GCA| 3151821   | 1.46| TGC| 3158491   | 1.46| Ila       |
| GCC| 2712004    | 1.25| GGC| 2709048    | 1.25| CGG| 610938    | 0.28| CCG| 611884    | 0.28| IIb       |
| CCA| 3015386    | 1.59| TGG| 3003473    | 1.39| GGG| 3424441   | 1.58| TCC| 3426302   | 1.58|           |

### Base No. Percentage

| Base | No.    | Percentage |
|------|--------|------------|
| A    | 63148537 | 27.66%     |
| C    | 4501700  | 19.71%     |
| G    | 45025109 | 19.72%     |
| T    | 62287824 | 27.72%     |
| N    | 11860701 | 5.19%      |
| Sum: | 228333871 | 100.00%    |
Neanderthal

Downloaded from [http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/VCF/](http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/VCF/).

Chromosome 1

### A+T rich group (I)

| D   | f   | %     | RC | f   | %     | C   | f   | %     | R   | f   | %     | Sub-group |
|-----|-----|-------|----|-----|-------|-----|-----|-------|-----|-----|-------|-----------|
| ATG | 4033513 | 1.79  | CAT | 4031747 | 1.79  | TAC | 2508730 | 1.11 | GTA | 2505803 | 1.11  | Ia        |
| TGA | 4382479 | 1.95  | TCA | 4376198 | 1.94  | ACT | 3626288 | 1.61 | AGT | 3625965 | 1.61  |          |
| TAG | 2891122 | 1.28  | CTA | 2891550 | 1.28  | ATC | 2975435 | 1.32 | GAT | 2972872 | 1.32  |          |
| TAA | 4495741 | 2     | TTA | 4503461 | 2     | ATT | 5364622 | 2.38 | AAT | 5363729 | 2.38  | Ib        |
| AAC | 3212047 | 1.43  | GTT | 3225909 | 1.43  | TTG | 4206294 | 1.87 | CAA | 4195901 | 1.86  |          |
| AAG | 4480227 | 1.99  | CTT | 4491979 | 1.99  | TTC | 4372755 | 1.94 | GAA | 4364361 | 1.94  |          |
| ACA | 4430760 | 1.97  | TGT | 4447034 | 1.97  | TGT*| 3285700 | 1.46 | ACA*| 3289439 | 1.46  | Ic        |
| ATA | 4379418 | 1.94  | TAT | 4384640 | 1.95  | TAT*| 4206294 | 1.87 | ATA*| 4206294 | 1.87  |          |
| AGA | 4979482 | 2.21  | TCT | 4992378 | 2.22  | TCT*| 3285700 | 1.46 | AGA*| 3285700 | 1.46  |          |
| AAA | 8328520 | 3.7   | TTT | 8348742 | 3.71  | TTT*| 3285700 | 1.46 | AAA*| 3285700 | 1.46  |          |

### C+G rich group (II)

| D   | f   | %     | RC | f   | %     | C   | f   | %     | R   | f   | %     | Sub-group |
|-----|-----|-------|----|-----|-------|-----|-----|-------|-----|-----|-------|-----------|
| CGT | 564292 | 0.25  | ACG | 562379 | 0.25  | GCA | 3285700 | 1.46 | TGC | 3289439 | 1.46  | Iia       |
| GTC | 2169347 | 0.96  | GAC | 2165877 | 0.96  | CAG | 4726068 | 2.1  | CTG | 4733685 | 2.1   |          |
| GCT | 3257195 | 1.45  | AGC | 3257760 | 1.45  | CGA | 509832  | 0.23 | TCG | 501345  | 0.23  |          |
| GCC | 2859033 | 1.27  | GGC | 2854237 | 1.27  | CGG | 660573  | 0.29 | CCG | 662465  | 0.29  | Iib       |
| CCA | 4288322 | 1.9   | TGG | 4286513 | 1.9   | GTG | 2695741 | 1.2  | ACC | 2694491 | 1.2   |          |
| CCT | 4206350 | 1.87  | AGG | 4193550 | 1.86  | GGA | 3590657 | 1.59 | TCC | 3603613 | 1.6   |          |
| CAC | 3427264 | 1.52  | GTG | 3431973 | 1.52  | GTG*| 3285700 | 1.46 | CAC*| 3285700 | 1.46  | Iic       |
| CGC | 567304  | 0.25  | GCG | 566812 | 0.25  | GCG*| 3285700 | 1.46 | CGC*| 3285700 | 1.46  |          |
| CTC | 3964997 | 1.76  | GAG | 3959340 | 1.76  | GAG*| 3285700 | 1.46 | CTC*| 3285700 | 1.46  |          |
| CCC | 3164177 | 1.4   | GGG | 3156887 | 1.4   | GGG*| 3285700 | 1.46 | CCC*| 3285700 | 1.46  |          |

### Base No. Percentage

| Base | No.     | Percentage |
|------|---------|------------|
| A    | 65508186 | 29.08      |
| C    | 47086506 | 20.90      |
| G    | 47061746 | 20.90      |
| T    | 65599484 | 29.12      |
| N    | 0        | 0.00       |
| Sum: | 225255922|            |
**Homo sapiens**

Downloaded from ftp://ftp.ncbi.nih.gov/genomes/Homo_sapiens/Assembled_chromosomes/seq/ (26.1.2015.).

### Chromosome 1

#### A+T rich group (I)

|       | D  | f     | %   | RC | f     | %   | C  | f     | %   | R   | f     | %   | Subgroup |
|-------|----|-------|-----|----|-------|-----|----|-------|-----|-----|-------|-----|----------|
| ATG   | 4110209 | 1.78 | CAT | 4120946 | 1.79 | TAC | 2542958 | 1.10 | GTA | 2566721 | 1.11 | Ia       |
| TAG   | 2960898 | 1.28 | CTA | 2941433 | 1.28 | ACT | 3035996 | 1.32 | GAT | 3056974 | 1.33 |          |
| TAA   | 4577976 | 1.99 | TTA | 4571528 | 1.98 | ATT | 5500468 | 2.39 | AAT | 5470905 | 2.37 | Ib       |
| AAC   | 3332435 | 1.45 | GTT | 3329970 | 1.44 | TGT | 4344678 | 1.89 | CAA | 4288540 | 1.86 |          |
| AAG   | 4581648 | 1.99 | CTT | 4634644 | 2.01 | TTC | 4548877 | 1.97 | GTA | 4518460 | 1.96 |          |
| ACA   | 4553751 | 1.98 | TGT | 4584113 | 1.99 | TGT*|       |      | ACA*|        |      |          |
| ATA   | 4475100 | 1.94 | TAT | 4472951 | 1.94 | TAT*|       |      | ATA*|        |      |          |
| AGA   | 5150760 | 2.23 | TCT | 5129424 | 2.23 | TCT*|       |      | AGA*|        |      |          |
| AAA   | 8516543 | 3.70 | TTT | 8583142 | 3.72 | TTT*|       |      | AAA*|        |      |          |

#### C+G rich group (II)

|       | D  | f     | %   | RC | f     | %   | C  | f     | %   | R   | f     | %   | Subgroup |
|-------|----|-------|-----|----|-------|-----|----|-------|-----|-----|-------|-----|----------|
| CGT   | 597422 | 0.26 | ACG | 586276 | 0.25 | GCA | 3361131 | 1.46 | TGC | 3357313 | 1.46 | Iia      |
| GTC   | 2202280 | 0.96 | GAC | 2216474 | 0.96 | CAG | 4852390 | 2.11 | CTG | 4811169 | 2.09 |          |
| GCT   | 3309131 | 1.44 | AGC | 3317232 | 1.44 | CGA | 523798  | 0.23 | TCG | 535651  | 0.23 |          |
| GCC   | 2891387 | 1.25 | GGC | 2891408 | 1.25 | CGG | 674618  | 0.29 | CCG | 669612  | 0.29 | Iib      |
| CCA   | 4330820 | 1.88 | GTG | 4368306 | 1.90 | GGT | 2728078 | 1.18 | ACC | 2725309 | 1.18 |          |
| CCT   | 4273302 | 1.85 | AGG | 4265068 | 1.85 | GGA | 3684403 | 1.60 | TCC | 3657040 | 1.59 |          |
| CAC   | 3506405 | 1.5 | GTG | 3530308 | 1.53 | GTG*|       |      | CAC*|        |      | Iic      |
| CGC   | 579316  | 0.3 | GCG | 583618  | 0.25 | GCG*|       |      | CGC*|        |      |          |
| CTC   | 4057534 | 1.8 | GAG | 4053693 | 1.76 | GAG*|       |      | CTC*|        |      |          |
| CCC   | 3195020 | 1.4 | GGG | 3187515 | 1.38 | GGG*|       |      | CCC*|        |      |          |

| Base | No.  | Percentage |
|------|------|------------|
| A    | 67070277 | 26.94%     |
| C    | 48055043 | 19.30%     |
| G    | 4811528  | 19.33%     |
| T    | 67244164 | 27.01%     |
| N    | 18475408 | 7.42%      |
| Sum: | 248956420 | 100.00%    |